A new hidden Markov-switching volatility model

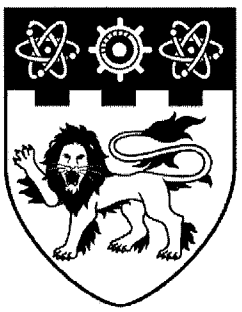
Liu, Xin Yi

2009

Liu, X. Y. (2009). A new hidden markov-switching volatility model. Doctoral thesis, Nanyang Technological University, Singapore.

https://hdl.handle.net/10356/20675

https://doi.org/10.32657/10356/20675



NANYANG TECHNOLOGICAL UNIVERSITY

A NEW HIDDEN MARKOV-SWITCHING VOLATILITY MODEL

A NEW HIDDEN MARKOV-SWITCHING VOLATILITY MODEL

LIU XINYI

LIU XINYI

DIVISION OF ECONOMICS

SCHOOL OF HUMANITY & SOCIAL SCIENCE

2009

2009

A New Hidden Markov-Switching Volatility Model

Liu Xinyi

School of Humanity & Social Science

A thesis submitted to the Nanyang Technological University
in fulfillment of the requirement for the degree of
Doctor of Philosophy

2009

A New Hidden Markov-Switching Volatility Model

Oct/2008

in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Economics
Nanyang Technological University

Author Liu Xinyi
Division of Economics, the School of Humanity and Social Science

Thesis Supervisor Wang Peiming

Abstract: The thesis proposes and applies a two-state hidden Markov-switching model for financial time series featured with periodic structure breaks in volatility. The expected return, volatility and state transition probability are determined by three link functions respectively, whose coefficients are further governed by the hidden state. The proposed model particularly emphasizes on the *parallel* structure of the two states. The *parallel* structure separates the INTER-state and INTRA-state dynamics, enhances greater transparency, balances the memory of both recent and distant history, provides more consistent economic implication, and greatly simplifies and stabilizes the EM algorithm. We further discuss its estimation, inference, standard errors of the parameter estimate, forecasting, model selection and implementation, especially our innovations in those issues. The Monte Carlo experiments suggest that the proposed estimation method is accurate and reliable, the choice of the initial state probability has little effect on proposed model, and the information matrix calculated numerically is stable and reliable.

We then apply the proposed model to forecast the conditional distribution of the weekly return of the S&P 500 index. It is found that the volatility at the tranquil state is best described as a constant while the volatility at the turbulent state is more closely correlated with the lagged intraweek ranges. Eventually, we give the economic implication on the clustered volatility and its interplay with the expected return.

KEYWORDS: HIDDEN MARKOV-SWITCHING MODEL, PARALLEL STRUCTURE, VOLATILITY, LINK FUNCTION, EM ALGORITHM, FORWARD-BACKWARD ALGORITHM, RESCALING, INFORMATION MATRIX OF THE MISSING DATA, SUPPLEMENTED EM ALGORITHM, RANGE.

Acknowledgement

I would like to express my gratitude to all those who gave me the possibility to complete this thesis.

I am deeply indebted to my supervisor Prof. WANG Peiming for the stimulating advice on logical thinking, economic reasoning, statistical method, computational technique, and writing. And most of all, I want to thank him for his continuous passionate encouragement for the attempt of innovations. His passion in Econometric modeling has greatly influence me.

I want to thank the professors in Division of Economics, in particular the coursework lecturers. Because of them, my passion for economics is not a coincidence and I consider myself lucky. I'd like to express my gratitude to the committee member Prof. TAN Kim Heng, Prof. RAHMAN Shahidur, and Prof. TAN Randolph. I'd like to thank Prof. Chen Kang, who encouraged me to apply for PhD program of NTU during his visit to Xiamen University in 2002. I'd like to thank Prof. Gu Qingyang, who helped enormously at the last stage the PhD program.

I want to thank my classmate for all their help, support, interest and valuable hints. Especially I am obliged to Huang Huamei, Li Ting, Gao Zhangpeng, Yu Lei, Zhu Xiaoneng, Zhang Yang, Ho Woon Yee and Huang Hanming.

Finally, I am forever indebted to my dear parents for their understanding and endless patience, especially when it is most needed.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACF	Autoregressive function
AIC	Akaike information criterion
BIC	Bayesian information criterion
CARR	Conditional autoregressive range
DGP	Sata generating process
EM	The expectation and maximization algorithm
FX	Foreign exchange rate
[G]ARCH	[Generalized] autoregressive conditional heteroskedasity
EGARCH	Exponential GARCH
IGARCH	Integrated GARCH
GARCH-MixN	GARCH model with mixture normal innovations
GARCH-T	GARCH model whose innovation follows Student's distribution
HMS-V	The proposed hidden Markov-switching volatility model
J-B test	Jarque-Bera test
KL	Kullback-Leibler information measure
MAD	Mean absolute deviation
MS (HMM)	Hidden Markov-switching model
MLE	Maximum likelihood estimator

MS-[G]ARCH	Hidden Markov-switching [G]ARCH model
MSE	Mean square error
MixN	Mixture normal diffusion
SEM	The supplemented expectation and maximization algorithm
XCF	Cross-sectional correlation function

LIST OF SYMBOLS

Vectors and matrices will be denoted with boldfaces (e.g. $\mathbf{C}$ and $\boldsymbol{\beta}$); a parameter vector is always a column vector; and sets and spaces will be denoted with calligraphic letters (e.g. $\mathcal{V}$).

$I(x)$	Indicator function, which equals to 1 if the statement x is true and 0 other wise
$\mathbf{T}$	A vector of a time series indices, $\mathbf{T} = \{t : t = 1, \cdots, T\}$
Δ_t	The state at time t
$\boldsymbol{s}_{\langle i \rangle, t}$	An indicator function of the state variable Δ_t, $\boldsymbol{s}_{\langle i \rangle, t} := I(\Delta_t = i)$, that is, the realization of a Bernoulli trial; $\boldsymbol{s}_{\langle i \rangle, t} = 1$ if $\Delta_t = i$, and $\boldsymbol{s}_{\langle i \rangle, t} = 0$ if $\Delta_t \neq i$
$\widetilde{\boldsymbol{s}_{\langle i \rangle, t}}$	The "true" rate of a successful Bernoulli trial $\boldsymbol{s}_{\langle i \rangle, t}$, $\widetilde{\boldsymbol{s}_{\langle i \rangle, t}} := E\left(\boldsymbol{s}_{\langle i \rangle, t}\right)$
$\widehat{\boldsymbol{s}_{\langle i \rangle, t}}$	The estimated rate of a successful Bernoulli trial $\boldsymbol{s}_{\langle i \rangle, t}$, $\widehat{\boldsymbol{s}_{\langle i \rangle, t}} := E\left(\widetilde{\boldsymbol{s}_{\langle i \rangle, t}}\right) = E\left(\boldsymbol{s}_{\langle i \rangle, t}\right)$
$\Re^m$	The m-dimensional Euclidean space.
$\widehat{\boldsymbol{\beta}}$	Estimation of the parameter vector $\boldsymbol{\beta}$
$\widehat{\beta^{(l+1)}}$	The estimate of the parameter β at the $(l+1)^{th}$ iteration.
$\left\|\widehat{\boldsymbol{\beta}^{(l+1)}} - \widehat{\boldsymbol{\beta}^{(l)}}\right\|$	The distance between the two matrices, $\widehat{\boldsymbol{\beta}^{(l+1)}}$ and $\widehat{\boldsymbol{\beta}^{(l)}}$, for the

sizes of two matrices are both $i \times j$,

$$\left\| \widehat{\boldsymbol{\beta}^{(l+1)}} - \widehat{\boldsymbol{\beta}^{(l)}} \right\| = \sum_{k=1}^{i} \sum_{h=1}^{j} \left| \widehat{\boldsymbol{\beta}_{k,h}^{(l+1)}} - \widehat{\boldsymbol{\beta}_{k,h}^{(l)}} \right|$$

$\boldsymbol{\theta}^{(l+1)} = \mathcal{M}\left(\boldsymbol{\theta}^{(l)}\right)$	A mapping from $\boldsymbol{\theta}^{(l)}$ to $\boldsymbol{\theta}^{(l+1)}$, where $\mathcal{M}$ is the rule
L	The lag operator
$\mathbf{X_T}$	A matrix with the index $\mathbf{T}$, $\mathbf{T} = 1, \cdots, T$
$\mathbf{x}'$	Transpose of the vector $\mathbf{x}$
$\mathbf{x} > 0$	All elements of vector $\mathbf{x}$ are positive elementwise
$\mathbf{0}_{a \times b}$	The zero matrix with size $a \times b$
$\mathbf{1}_{a \times b}$	The unity matrix with size $a \times b$
$diag\left(c_1, \cdots, c_m\right)$	The diagonal matrix with diagonal elements $\left(c_1, \cdots, c_m\right)$
$tr\left(\mathbf{S}\right)$	The trace of the square matrix $\mathbf{S}$
$\mathbf{I}$	The identity matrix
ψ_t	All the available information up to, and including, time t
$E_t\left(\varepsilon \mid x\right)$	The expectation of the variable ε at time t based on the information x
$Var_t\left(\varepsilon \mid x\right)$	The variance of the variable ε at time t based on the information x
$\phi\left(\bullet \middle\| \mu, \sigma^2\right)$	Probability density function of the normal distribution $N\left(\mu, \sigma^2\right)$
$\Phi\left(\bullet \middle\| \mu, \sigma^2\right)$	Cumulative density function of the normal distribution $N\left(\mu, \sigma^2\right)$
$\mathbf{I}_{com}$	The information matrix of the complete data
$\mathbf{I}_{obs}$	The information matrix of the observable data

$\mathbf{I}_{miss}$	The information matrix of the missing data
$\mathcal{I}$	The Fisher information matrix
$f_{\langle i \rangle}(\bullet \mid \bullet, \bullet)$	A function $f_{\langle i \rangle}$ with 2 inputs
$LF(\boldsymbol{\theta})$	A likelihood function whose parameter is $\boldsymbol{\theta}$
$LLF(\boldsymbol{\theta})$	A log likelihood function whose parameter is $\boldsymbol{\theta}$
$LLF_{obs}(\boldsymbol{\theta})$	The log likelihood function of the observed data
$LLF_{com}(\boldsymbol{\theta})$	The log likelihood function of the complete data
$\nabla LLF(\boldsymbol{\theta})$	The Gradient of the log likelihood function
$\nabla^2 LLF(\boldsymbol{\theta})$	The Hessian of the log likelihood function

1. INTRODUCTION

Volatility, the most extensively used measure of uncertainty, is fundamental to much of modern finance theory. Quantifying and forecasting volatility is essential to financial asset pricing, portfolio optimization and risk management. Over the past two decades, a diverse range of theoretical and empirical research has been carried out on modeling the volatility of time series. Following the widespread popularity of a number of financial instruments, whose values are intrinsically linked to their corresponding volatilities, the development of volatility models and estimation frameworks has emerged as active research areas.

The proposed model is a further extension of those presented in existing literature. It uniquely emphasizes the parallel structure of the Markov states. This parallel structure separates the INTER-state and INTRA-state dynamics, allows for greater transparency, balances the memory of both recent and distant history, provides more consistent economic implication as well as greatly simplifying and stabilizing the EM algorithm.

1.1. Volatility Observations

It is widely recognized that the volatilities of most financial asset returns are time-varying and typically display, at least, the following six stylized characteristics:-

(Obs. 1) ***Clustering:*** Mandelbrot (1963b) and Fama (1965), among the others, found that large (small) variances tended to be followed by large (small) variances. Statistically, the squared return has significant positive autocorrelations, as does the absolute return.

(Obs. 2) ***Structure breaks:*** The pattern of many financial asset returns often exhibits dramatic and sudden changes. Those changes are often event-triggered. They are often associated with financial crisis (Cerra, 2005; Hamilton, 2005) or abrupt government policy changes (Hamilton, 1988; Sims and Tao, 2004; Davig, 2004). Therefore, given the same volatility functional form, different coefficients can be used to describe the time series before and after a structure break.

(Obs. 3) ***Reoccurrence of distant history:*** Although the pattern of asset return is likely to follow recent history, a seemingly unlikely scenario that resembles that of distant history should not be considered irrelevant.

(Obs. 4) ***Long memory:*** Traditionally, long memory (long-term dependence) has been defined in both the time domain, in terms of long lag *linear* autocorrelation, and the frequency domain, in terms of the explosion of low frequency spectra. An abundance of publications demonstrate the existence of long memory in financial economics. Baillie, Chung and Tieslau (1996) found long memory in the volatility of the Deutschemark to U.S. Dollar foreign exchange rate; and long term dependence in the German DAX was found by Lux (1996). While the typical research focus is on developed financial markets, there are also numerous reported studies on smaller and less developed markets. For example, the stock market in Finland was analyzed by Tolvi (2003) , Madhusoodanan (1998) provided evidence of long memory on individual stocks on the Indian Stock Exchange, similar evidence on the Greek financial market was

given by Barkoulas and Baum (2000). It is worth noting that that for most long memory studies, a linear auto-correlation is assumed[1].

(Obs. 5) ***Interplay between volatility and expected return:*** There is a realistic probability that volatility and expected return are closely related. For example, in the case of the stock market, Hamilton and Lin (1996) as well as Maheu and Mccurdy (2000), documented that volatility is higher during a bear market (e.g. during the great depression and the stagflation age in the 1970s, etc). Findings on the relationships between volatility and return can also be found in credit spreads (see Alexander and Kaeck, 2008), foreign exchange rates (see Engle and Hamilton, 1990), and other asset classes.

(Obs. 6) ***Leptokurtosis and asymmetry:*** The unconditional distribution of a financial asset return has thicker tails than those of the normal distribution (Mandelbrot, 1963a, 1993b). Compared to the classical normal distribution assumption, the Leptokurtosis suggests greater upside and downside risks. Significant skewness of a financial time series has been either inherently assumed or declared by a number of researchers (e.g. Engle and Patton, 2001; Cont, 2001; Chen, Hong and Stein, 2001).

As shown later in Section 5 of the thesis, following the results of an empirical study of the S&P 500 stock index, two more additional observations are proposed:-

(Obs. 7) ***Asymmetric cycle and lasting tranquility:*** The tranquil regime is more lasting and is related to the asymmetric business cycle.

(Obs. 8) ***Time varying sensitivity to recent shocks:*** In a turbulent regime, volatility is more sensitive to the recent shocks.

[1] The assumed linear relationship could be a potential reason.

While some of these observations have been intensively addressed in the literature, there are a number of outstanding oversights which require further attention. This thesis aims to address these observations and intently focus on the oversights exposed in the following section.

1.2. Motivations and Oversights

This thesis proposes a *parallel* two-state hidden Markov-switching volatility (HMS-V) model. The model is *"parallel"* in the sense that the dynamics of the two states are less dependent on each other. The model is proposed to explicitly address the oversights of both the existing single-regime and the Markov-switching volatility model. As exposed in the literature review of Section 3, a linear autoregressive relationship in variance coupled with various conditional distributions are commonly employed to describe the volatility of asset returns. However, such a 'one-size-fits-all' model may prove unsatisfactory as, for instance, it may be unable describe a time series with structure breaks.

For practitioners, if the parameters are to be adjusted according to recent and hence the most relevant economic context, a popular compromise is to focus only on the most recent observations. However, such a compromise is based on the assumption that only recent history will repeat itself, and disregards distant history. This assumption is not satisfactory since a seemly remote event has a finite probability of recurrence, the consequences of which can often be far reaching[2]. Popular wisdom commonly adopts a

[2] This is an important lesson learned from the recent credit market turmoil: if a modeler assumes that the relevance of a scenario dwindles linearly, then based on the post WWII nominal U.S. housing statistics, he/she would seek comfort and preclude the possibility of a housing Armageddon that happened about seven decades ago. The preclusion of such

rule that "the further backward you look, the further forward you can see."[3]. It is therefore dangerous to extrapolate the future based only on recent history. For example, in the case of the S&P 500 stock index, based only on the tranquility from 2004 to 2006, the volatility forecast can be downward biased. In this case, excessive emphasis has been placed on the recent tranquility, resulting in a model that suggests a deceptive sense of security.

This thesis proposes the consistent incorporation of longer history. The proposed parallel Markov-switching model is able to estimate the changing volatility behavior through explicit consideration of the long history. It is most suited to a time series, whose recent behavior is dissimilar to its distant history, but whose distant history is still relevant.

Various Markov-Switching volatility models have been proposed (see Section 3.4) to describe the changing behavior of volatility. This thesis emphasises that existing regime-switching volatility models are *not* parallel, and possess the following oversights:-

(Oversight. 1) ***Same volatility link function for both INTER-state and INTRA-state dynamics:*** For some popular Markov-switching volatility model, a same volatility link function is adopted to explain both INTER-state and INTRA-state dynamics[4]. This oversight is directly addressed by the proposed model in which the INTER-state dynamics are better explained by the switching of regimes and the volatility link function focuses only on the INTRA-state dynamics.

---

an "impossible" scenario has led to the excessive credit expansion and mis-pricing of U.S. mortgage backed securities in the mid 2000s.

[3] Winston Churchill

[4] For example, it assumes that volatility link function at the turbulent state should address: (1) the INTER-state jump of volatility from the tranquil state to the turbulent state, and (2) the INTRA-state evolution of volatility form the turbulent state to the turbulent state.

(Oversight. 2) ***Less tractable dynamics:***[5] Analytical tractability is reduced since the two states are intertwined with each other in a complex manner.

(Oversight. 3) ***Possibly excessive sensitivity to recent observations:*** According to classical non-parallel model specifications, it is not unlikely that the volatility link functions of both states greatly absorb and adapt to recent observations.

(Oversight. 4) ***Passive adaptation and hence possibly missing memory:*** If both states greatly adapt to recent history, the memory of the distant history may diminish. Only when there is news resembling that of the distant history, will the memory be passively invoked[6].

(Oversight. 5) ***Reduced appeal of regime-switching:*** The volatility at the two states can be less distinguishable if there is excessive adaptation to recent history. In this case, the clustering of volatility is unlikely to be explained by the clustering of Markov states[7] and the regime-switching model becomes less appealing.

(Oversight. 6) ***Long memory not addressed by regime-switching:*** If the two states are less identifiable, then volatility clustering and hence "long memory" are primarily addressed by the volatility link function, instead of by the clustering / switching of regimes.

(Oversight. 7) ***Possibly inconsistent economic implication of regime switching:*** If the scenarios described by the two states are homogeneous, the economic implication of a regime-switching model is probably inconsistent.

(Oversight. 8) ***Excessive computational burden:*** As detailed in Section 4.1.3, the non-parallel structure of the Markov-switching renders the estimation process

---

[5] "Things should be made as simple as possible -- but no simpler." - A. Einstein

[6] For example, both states may excessively adapt to the recent tranquility, and the volatilities at both states are very low. In this case, the turbulent memory will only be invoked when there is a significant shock.

[7] It may largely be explained by the volatility link function instead.

non-parallel. It increases the computational burden and reduces the robustness of the parameter estimation.

1.3. Thesis Contributions

The proposed parallel Markov-switching model, as is outlined in this section, addresses the empirical observations while simultaneously addressing the oversights of existing models exposed in the previous section.

1.3.1. Addressing the observations

(Obs. 1) ***Clustering:*** Statistically, it is assumed that the volatility of a time series at different Markov states exhibits different patterns. Therefore, if the Markov states are clustered, the volatility is also clustered.

(Obs. 2) ***Structure breaks:*** When the Markov state switches, the parameters of the dominant state will describe the time series after the structure break.

(Obs. 3) ***Reoccurrence of distant history:*** Statistically, the distant history is memorized and stored in a hidden and recessive Markov state, hence its relevancy is consistently considered[8].

(Obs. 4) ***Long memory:*** The proposed model has the ability to describe a periodic structure break. The long memory will be explained if the periodic structure breaks is indeed in the data.

[8] It is not dominant but the relevancy always exists.

(Obs. 5) ***Interplay between volatility and expected return:*** Apart from volatility, for different Markov states, if the expected return also exhibits a different pattern, the interplay between volatility and the expected return is also described.

(Obs. 6) ***Leptokurtosis and asymmetry:*** Within each state, the volatility is clustered. The Markov state is a latent variable and has to be estimated. In this case, the unconditional distribution will be leptokurtosis, and possibly asymmetric.

(Obs. 7) ***Lasting tranquility:*** If more observations are clustered in the tranquil state than in the turbulent state, then the lasting tranquility is described. It is related to the asymmetric business cycle.

(Obs. 8) ***Time varying sensitivity to recent shocks:*** The proposed parallel model suggests that a volatility link function adapts only to more recent shocks. As a result, it is easy to compare the sensitivity of volatility to the recent shocks at different Markov states.

1.3.2. Addressing the Oversights

The proposed model suggests that longer history be explicitly considered[9]. With a longer history, (i) recent history can be incorporated through the volatility function of the dominant state; and (ii) the distant history can be stored in the recessive state.

The proposed model does not need to use a GARCH-type recursive volatility link function. This is because (i) the Markov state is already a recursive process and able to describe clustering; (ii) the GARCH-type recursive volatility link function may encourage

[9] In a previous version, for the empirical study on the S&P 500, only the data from the year 1990 is used. The current version uses the data from 1983 onwards, as the range data prior to 1983 is somewhat inconsistent.

both states to excessively emphasise very recent observations.

The parallel structure improves upon existing non-parallel Markov volatility models:-

(Addressing Oversight. 1) ***Volatility link function only for INTRA-state dynamics:*** According to the proposed model, the INTER-state dynamics are addressed by the switching of regimes, and each volatility link function focuses only on the INTRA-state dynamics.

(Addressing Oversight. 2) ***Greater tractability:*** Since the two states are less intertwined, the proposed model is more tractable.

(Addressing Oversight. 3) ***Balanced adaptation to recent signals:*** Note that only the most (not all) recent information (e.g. limited lag terms) is used for each volatility link function, and according to the non-parallel model specification, it is likely that one of the volatility link functions will focus on the volatility pattern of recent history, while the other volatility link function will focus on the pattern of distant history[10].

(Addressing Oversight. 4) ***Distant memory still alive:*** The proposed model emphasises the independence of the two states. One of the states is less affected by the recent information, and keeps memorising the distant history. It reminds the modeler of the danger of extrapolating the future based only on the recent history.

(Addressing Oversight. 5) ***Identifiable states:*** The parallel structure makes the Markov model easier to identify, in particular if the volatility of one state is almost always greater than the other state. A sufficient condition of identification is discussed in Appendix I.

[10] Particularly if the recent history and distant history exhibit different pattern.

(Addressing Oversight. 6) ***Long memory addressed by regime-switching:*** The long memory is considered more by the switching of regimes as opposed to by the auto-regressive volatility link functions.

(Addressing Oversight. 7) ***More consistent economic implication of regime switching:*** The parallel structure aims to emphasise the distinctive differences of the two states. With consistently distinctive different state profiles, the economic implications for the Markov states are likely to be more clear-cut and consistent.

(Addressing Oversight. 8) ***Reduced computational burden:*** According to section 4.1, the Expectation Maximization (EM) algorithm for the proposed model will be more stable. Thos is because in the M step of EM algorithm, the coefficients for each state can be independently estimated.

1.3. Thesis Outline

Inclusive of this introduction, this thesis is comprised of 7 chapters. Chapter 2 briefly presents the proposed hidden Markov-switching volatility model framework. A model overview is provided, including model assumptions, setup, some special cases as well as the proposed likelihood function.

Chapter 3 reviews existing literature on classical volatility models, including hidden Markov switching models. Namely, the chapter shows:-

- that classical volatility models are the building blocks of the proposed hidden Markov-switching volatility model
- a review and comparison of the proposed model with two popular Markov-switching volatility models

- some special cases to illustrate the overviews and pitfalls of existing models while highlighting the enhancements of the proposed model.

Chapter 4 formally discusses the proposed models identifiability condition, estimation and forecasting, including:-

- a sufficient identifiability condition
- the estimation procedure and computational advantages (e.g. the Expectation, Maximization and Quasi-Newton)
- various error estimation numerical methods to evaluate the accuracy of the estimated parameters, in particular, the Supplemented Expectation Maximization (SEM) algorithm
- Monte Carlo simulations, which show (i) the asymptotic normal property of the maximum likelihood estimate (MLE) is reflected in a good finite-sample property (i.e., *approximate normal and small bias*); (ii) the numerical method is adequately stable and accurate to obtain the standard errors of the MLE; (iii) the impact of the initial state probabilities is negligible, given that the time series is long enough and
- forecasting volatility with the proposed model.

Chapter 5 adopts the proposed hidden Markov-switching volatility model to measure and forecast the weekly return of the S&P 500 stock index. The intraweek range is also used to construct both the covariate of the Logit and volatility link functions. It is shown that the proposed model demonstrates impressive performance under certain selection criteria conditions.

Chapter 6 presents a number of economic interpretations based on the empirical study:-

- Asymmetric business cycle: The business cycle and hence the velocity of economic time are asymmetric. Therefore to scale the calendar time to economic time, more than one set of parameters are required to describe the volatility. One set of parameters is necessary when the economic time speeds up, while another set is needed when the economic time slows down.
- Surprises: In the tranquil state, there are fewer surprises and challenges to the bullish assumptions on asset pricing, while the opposite is true for the turbulent state. Furthermore, volatile assumptions lead to a more volatile price.
- Sensitivities: In the turbulent state, the price is more sensitive to unexpected surprises, while the opposite is true for the tranquil state. For a given amount of surprise, greater sensitivity induces greater volatility.
- Correlations: In the turbulent state, the correlations of stock returns are higher, thus the diversification benefit of a balanced portfolio is less pronounced. For the tranquil state, the opposite is true. Greater correlations cause greater volatility for a portfolio (hence a stock index).
- Risk Premia: For risk-averse investors, an unexpected increase in volatility must be compensated with greater expected return. A lower current price is necessary to achieve greater expected future returns.

Finally, Chapter 7 discusses some drawbacks and limitations of the proposed model, as well as providing numerous interesting avenues of further research

2. THE PROPOSED HIDDEN MARKOV-SWITCHING VOLATILITY MODEL

2.1. Model Overview

The proposed hidden markov switching volatility (HMS-V) model analyses the time series of financial asset return through the following assumptions:-

1) The model of return from a time series takes the following form:

$$\begin{aligned} y_t &= \mu(\boldsymbol{\alpha}, \psi_t) + \sigma(\boldsymbol{\beta}, \psi_t) \bullet \eta_t \\ &= \mu_t + \sigma_t \bullet \eta_t \end{aligned} \tag{2.1}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are vectors of unknown parameters, ψ_t is the information up to time *t*, and η_t's are distributed according to

$$\eta_t \overset{i.i.d}{\sim} N(0,1), \tag{2.2}$$

which represent independent and identically distributed (i.i.d) Gaussians densities of zero mean and unit variance.

Let $\phi\left(\bullet \middle| \mu, \sigma^2\right)$ be the probability density function *(pdf)* of a normal distribution with mean μ and standard deviation σ, and let $\Phi\left(\bullet \middle| \mu, \sigma^2\right)$ be its cumulative density function *(cdf)*, that is,

$$\phi\left(v \middle| \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi}\sigma} \bullet \exp\left[-\frac{(v-\mu)^2}{2\sigma^2} \right], \text{ and,} \tag{2.3}$$

$$\Phi\left(v\middle|\mu,\sigma^2\right)=\int_{-\infty}^{v}\phi\left(v\middle|\mu,\sigma^2\right)dv\,. \tag{2.4}$$

2) There are two hidden discrete states. At time, t, the observation of the time series is generated by one of the states:

$$\Delta_t=1 \text{ or } \Delta_t=2, \tag{2.5}$$

where Δ_t is the state in which that the time series resides.

Therefore,

$$y_t=\begin{cases}\mu_{\langle 1\rangle}\left(\boldsymbol{\alpha}_{\langle 1\rangle},\ \psi_t\right)+\sigma_{\langle 1\rangle}\left(\boldsymbol{\beta}_{\langle 1\rangle},\ \psi_t\right)\bullet\eta_t & \text{if } \Delta_t=1\\ \mu_{\langle 2\rangle}\left(\boldsymbol{\alpha}_{\langle 2\rangle},\ \psi_t\right)+\sigma_{\langle 2\rangle}\left(\boldsymbol{\beta}_{\langle 2\rangle},\ \psi_t\right)\bullet\eta_t & \text{if } \Delta_t=2\end{cases} \tag{2.6}$$

where $\mu_{\langle i\rangle,t}$ and $\sigma_{\langle i\rangle,t}$ are the mean and standard deviation of the time series in state, i, respectively at time t.[11]

It is worth noting that this section gives a very general model specification, since ψ_t encompasses a broad range of potential model specifications. Given the information set, ψ_t, the relevant information for the expected return and volatility consists of covariates[12] $\mathbf{c}_t$ and $\mathbf{z}_t$:-

- $\mathbf{c}_t$ is the covariate for the link function $\mu_{\langle 1\rangle}\left(\boldsymbol{\alpha}_{\langle 1\rangle},\ \psi_t\right)$ and $\mu_{\langle 2\rangle}\left(\boldsymbol{\alpha}_{\langle 2\rangle},\ \psi_t\right)$, which is either fixed or strictly exogenous, and
- $\mathbf{z}_t$ is the covariate for the function $\sigma_{\langle 1\rangle}\left(\boldsymbol{\beta}_{\langle 1\rangle},\ \psi_t\right)$ and $\sigma_{\langle 2\rangle}\left(\boldsymbol{\beta}_{\langle 2\rangle},\ \psi_t\right)$, which is also either fixed or strictly exogenous.

[11] For visual clarity, state notation is usually expressed at the lower-right side and embraced with angle brackets

[12] ψ_t can also be allowed to consist of information other than covariates.

The values of $\mu_{\langle 1\rangle,t}$, $\mu_{\langle 2\rangle,t}$, $\sigma_{\langle 1\rangle,t}$ and $\sigma_{\langle 2\rangle,t}$ in equation (2.6) are then determined by the covariates $(\mathbf{c}_t, \mathbf{z}_t)$ and the corresponding parameters.

Given the above framework, let $f_{\langle i\rangle,t}$ be the *pdf* of y_t. Given $\Delta_t = i$, as well as strictly exogenous $\mathbf{c}_t$ and $\mathbf{z}_t$, $f_{\langle i\rangle,t}$ can be calculated though a function $f_{\langle i\rangle}(\bullet|\bullet,\bullet)$,

$$\begin{aligned} f_{\langle i\rangle,t} &= f_{\langle i\rangle}\left(y_t \mid \vartheta_{\langle i\rangle}, \psi_t\right) \\ &= \phi\left(y_t \middle| \mu_{\langle i\rangle,t}, \sigma^2_{\langle i\rangle,t}\right), \end{aligned} \tag{2.7}$$

where $\vartheta'_{\langle i\rangle} = \left(\alpha'_{\langle i\rangle}, \beta'_{\langle i\rangle}\right)$. For notation convenience, the parameter vectors are further grouped as follows, $\vartheta'_{\langle i\rangle} = \left(\alpha'_{\langle i\rangle}, \beta'_{\langle i\rangle}\right)$, $\alpha' = \left(\alpha'_{\langle 1\rangle}, \alpha'_{\langle 2\rangle}\right)$, $\beta' = \left(\beta'_{\langle 1\rangle}, \beta'_{\langle 2\rangle}\right)$, and $\vartheta' = \left(\vartheta'_{\langle 1\rangle}, \vartheta'_{\langle 2\rangle}\right)$.

3) The dynamic sequence of the hidden states is described by a first-order Markov-switching process, i.e., the current state depends only on the previous state and the transition probability. For clarity of exposition, higher order models such as a Markov random field are not considered.

Let $p_{\langle ij\rangle,t}$ be the state transition probability:

$$\begin{aligned} p_{\langle ij\rangle,t} &= \Pr\left(\Delta_t = j \mid \Delta_{t-1} = i\right) \\ &= p_{\langle ij\rangle}\left(\varphi_{\langle i\rangle}, \psi_t\right), \end{aligned} \tag{2.8}$$

where $\varphi_{\langle i\rangle}$ is the parameter vector.

When $t=1$, the probability that y_t resides in state i is determined by the initial state probability:

$$\Pr(\Delta_1 = i); \quad (i=1,2), \tag{2.9}$$

and $\Pr(\Delta_1 = 2) = 1 - \Pr(\Delta_1 = 1)$.

Suppose that for a given information set ψ_t, the relevant information for the state transition probability $p_{(ij),t}$ consists of covariate $\mathbf{w}_t$. Let $\mathbf{x}_t = (\mathbf{w}_t, \mathbf{c}_t, \mathbf{z}_t)$ be the covariate of y_t and,

$$\mathbf{X}_T = (\mathbf{W}_T, \mathbf{C}_T, \mathbf{Z}_T) \tag{2.10}$$

$\mathbf{W}_T = \{\mathbf{w}_t;\ t = 2, \cdots, T\}$,[13]

$$\mathbf{C}_T = \{\mathbf{c}_t;\ t = 1, \cdots, T\}, \tag{2.11}$$

and $\mathbf{Z}_T = \{\mathbf{z}_t;\ t = 1, \cdots, T\}$.

The dimensions of the covariate matrices, $\mathbf{X}_T$, $\mathbf{W}_T$, $\mathbf{C}_T$ and $\mathbf{Z}_T$ are therefore $T \times D_X$, $(T-1) \times D_W$, $T \times D_C$, an $T \times D_Z$ respectively, where $D_X = D_W + D_C + D_Z$.

The HMS-V thus embeds three link functions, of which equation (2.6) defines both the link functions of the expected return and volatility, and equation (2.8) defines the link function of the transition probability. Each of the three link functions has two sets of coefficients respectively. If the covariate of a certain link function is only a column of 1s, the corresponding link function refers to a constant at each state.

[13] Note that $\mathbf{w}_t$ is the covariate of the transition probability from time $t-1$ to t. Since the transition probability from time 0 to time 1 is not needed, $\mathbf{w}_1$ is void.

Let $\boldsymbol{\theta}' = \left(\boldsymbol{\alpha}', \boldsymbol{\beta}', \boldsymbol{\varphi}', \Pr\left(\Delta_1 = 1\right)\right)$, which denotes the entire set of parameters of the two-state discrete HMS-V model. A finite possibility of a parameter being zero[14] is also considered. Therefore, the above framework presents a general framework of hidden Markov- switching volatility models (HMS-V).

2.3. Special cases and relationships with existing models

This section presents some special cases for the general form of the proposed HMS-V model. While the covariate of the HMS-V model assumes that the covariate is either fixed or strictly exogenous, this is not the case for some of the following cases.

- When the covariate matrices $\mathbf{W}_T$, $\mathbf{C}_T$ and $\mathbf{Z}_T$ contain only intercept terms, the model is equivalent to that of Engel and Hamilton (1990) and Engle (1994), who use the regime-switching model to describe the long swing of the U.S. Dollar. Note that if the volatilities in both states are constants, then the switching of parameters is equivalent to switching of the underlying processes.

- When both, $\mathbf{W}_T$ and $\mathbf{Z}_T$, contain only intercept terms, and the element of $\mathbf{C}_T$ equals to $\mathbf{c}_t = \{1, ads_t, y_{t-1}\}$[15], the proposed model is equivalent to the Markov-switching advertising model (e.g. Feichtinger, Hartl and Sethi, 1994; Naik and Raman, 2003).

[14] Examples can be found in chapter 5 of the thesis.

[15] where ads_t is the spending in advertising at time t, and y_t is the brand sales at time t.

- When both of the covariate matrices, $\mathbf{W_T}$ and $\mathbf{C_T}$, are only intercept terms, and the element of covariate $\mathbf{Z}_T$ equals to $\mathbf{z}_t = \left\{1, \left(y_{t-1} - \mu_{t-1}\right)^2\right\}$ [16], it is equivalent to the MS-ARCH model with only one lag term by Hamilton and Susmel (1994). However, Hamilton and Susmel (1994) suggest that the volatility at each state is always proportional to each other, which is different from our model.

- By imposing the following specific conditions on the covariate, the general form is equivalent to Gray's (1996) and Klaassen's (2002) Markov-switching volatility model,

$\mathbf{z}_t = \left(\varepsilon_t^2, \sigma_{t-1}^2\right)$, and

$$\varepsilon_t = y_t - E(\mu_t)$$
$$E(\mu_t) = \Pr(\Delta_t = 1)\bullet\mu_{\langle 1\rangle,t} + \Pr(\Delta_t = 2)\bullet\mu_{\langle 2\rangle,t} \qquad (2.12)$$
$$\sigma_{t-1}^2 = \Pr(\Delta_{t-1} = 1)\bullet\left(\sigma_{\langle 1\rangle,t-1}^2 + \mu_{\langle 1\rangle,t-1}^2\right) + \Pr(\Delta_{t-1} = 2)\bullet\left(\sigma_{\langle 2\rangle,t-1}^2 + \mu_{\langle 2\rangle,t-1}^2\right) - \left[\Pr(\Delta_{t-1} = 1)\bullet\mu_{\langle 1\rangle,t-1} + \Pr(\Delta_{t-1} = 2)\bullet\mu_{\langle 2\rangle,t-1}\right]^2$$

Note that in Gray/Klaassen's model, ψ_t includes the estimated volatility at time $t-1$, while the estimated volatility of a state is not used as the covariate of the other state. This is discussed in detail later in section 3.4 and section 3.5.

- By imposing the following conditions on the covariate, the proposed model is equivalent to Haas/ Mittnik/Paolella's (2004b) Markov-switching volatility model:-

$\mathbf{z}_t = \left(\varepsilon_t^2, \sigma_{\langle 1\rangle,t-1}^2, \sigma_{\langle 2\rangle,t-1}^2\right)$, and

$\varepsilon_t = y_t$,

$$\sigma_{\langle 1\rangle,t}^2 = \beta_{\langle 1\rangle,0} + \beta_{\langle 1\rangle,1}\bullet\varepsilon_t^2 + \beta_{\langle 1\rangle,2}\bullet\sigma_{\langle 1\rangle,t}^2 + 0\bullet\sigma_{\langle 2\rangle,t}^2, \qquad (2.13)$$

[16] Note that Hamilton and Susmel (1994) assume that the covariate is random.

$$\sigma^2_{\langle 2\rangle,t} = \beta_{\langle 2\rangle,0} + \beta_{\langle 2\rangle,1} \bullet \varepsilon_t^2 + 0 \bullet \sigma^2_{\langle 1\rangle,t} + \beta_{\langle 2\rangle,3} \bullet \sigma^2_{\langle 2\rangle,t}.$$

As discussed in the introduction, for the volatility link function, the proposed approach differs in its use of a very limited lag term instead of the ARCH (∞) process. This is also discussed in section 3.4 and section 3.5.

2.4. The link functions

To limit the scope of this thesis, the proposed HMS-V model only adopts the following specific link functions:

1) The logit function is a natural choice of the state transition link function[17],

$$\begin{aligned} p_{\langle ii\rangle,t} &= p_{\langle ii\rangle}\left(\boldsymbol{\varphi}_{\langle i\rangle}, \psi_t\right) \\ &\equiv LOGIT\left(\mathbf{w}_t \boldsymbol{\varphi}_{\langle i\rangle}\right), \qquad (i=1,2; t=2,\cdots T) \end{aligned} \tag{2.14}$$

so that $0 < p_{\langle ij\rangle,t} < 1$ is always satisfied.

2) The expected return of y_t, given that $\Delta_t = i$, is a linear function,

$$\mu_{\langle i\rangle,t} = \mu_{\langle i\rangle}\left(\boldsymbol{\alpha}_{\langle i\rangle}, \psi_t\right) \equiv \mathbf{c}_t \boldsymbol{\alpha}_{\langle i\rangle}, \quad (i=1,2; t=1,\cdots T). \tag{2.15}$$

3) The link function of the volatility of y_t, given that $\Delta_t = i$, takes an exponential form,

$$\sigma_{\langle i\rangle,t} = \sigma_{\langle i\rangle,t}\left(\boldsymbol{\beta}_{\langle i\rangle}, \psi_t\right) \equiv \exp\left(\mathbf{z}_t \boldsymbol{\beta}_{\langle i\rangle}\right), \quad (i=1,2; t=1,\cdots T) \tag{2.16}$$

so that $\sigma_{\langle i\rangle,t} > 0$ is always satisfied. No restriction on the parameter is therefore required[18].

[17] Note that $\mathbf{w}_1$ is void.

[18] Such functional form also makes the numerical estimation method more robust.

2.5. The likelihood function

The log likelihood function of the two-state HMS-V model described in section 2.1 is[19]

$$LLF_{obs}(\boldsymbol{\theta}) = \log\left(LF_{obs}\left(\boldsymbol{\theta} \mid \mathbf{X}_T, \mathbf{Y}_T\right)\right)$$
$$= \log\left\{\sum_{i=1}^{2}\sum_{\Delta_2=1}^{2}\cdots\sum_{\Delta_T=1}^{2}\left[\Pr(\Delta_1 = i)\cdot f_{\langle i\rangle,1}\cdot\prod_{t=2}^{T}\left(p_{\langle\Delta_{t-1}\Delta_t\rangle,t}\cdot f_{\langle i\rangle,t}\right)\right]\right\} \qquad (2.17)$$

where

$\boldsymbol{\theta}' = \left(\boldsymbol{\varphi}'_{\langle 1\rangle}, \boldsymbol{\varphi}'_{\langle 2\rangle}, \boldsymbol{\alpha}'_{\langle 1\rangle}, \boldsymbol{\beta}'_{\langle 1\rangle}, \boldsymbol{\alpha}'_{\langle 2\rangle}, \boldsymbol{\beta}'_{\langle 2\rangle}, \Pr\left(\Delta_1 = 1\right)\right)$ is the unknown parameter vector,

$\Pr(\Delta_1 = i)$ is defined by equation (2.9),

$p_{\langle\Delta_{t-1}\Delta_t\rangle,t} = p_{\langle\Delta_{t-1}\Delta_t\rangle}\left(\boldsymbol{\varphi}_{\langle i\rangle}, \psi_t\right)$ is defined by equation (2.8), and

$f_{\langle i\rangle,t} = f_{\langle i\rangle}\left(y_t \mid \boldsymbol{\vartheta}_{\langle i\rangle}, \psi_t\right)$ is defined by equation (2.7).

2.6. Initial state probabilities

According to the above framework, the initial state probability, $\Pr\left(\Delta_1 = 1\right)$, is an unknown. However, with covariates, Wang and Puterman (1999a, 1999b, and 2001) presented results from some Monte Carlo experiments and showed that for a Markov-switching Poisson regression, the effect of the initial state probability quickly becomes negligible as the number of observations increases.

It is believed that this observation of the influence of the initial state probability is also applicable to the proposed model. The sample size of a financial time series is typically

[19] We use the notation LLF_{obs} because the log likelihood function is based on the observable information. Please refer to section 4.1 for a more detailed explanation.

far greater than that of the clinical data studied by Wang and Puterman[20]. Therefore, a non-informative initial state probability is assumed,

$$\Pr(\Delta_1 = 1) = 0.5. \tag{2.18}$$

Given $\Pr(\Delta_1 = 1)$, the parameter vector $\boldsymbol{\theta}$ reduces to

$$\boldsymbol{\theta}' = (\boldsymbol{\alpha}', \boldsymbol{\beta}', \boldsymbol{\varphi}'). \tag{2.19}$$

However, due to the inherent differences between the proposed model and that of Wang and Puterman, further studies the influence of initial state probabilities are carried out through Monte Carlo experiments in section 4.5.

[20] For example, chapter 5 presents studies of a time series, whose in-sample number of observations is 989.

3. LITERATURE REVIEW AND MODEL COMPARISONS

Following the brief introduction of the proposed framework presented in chapter 2, this chapter reviews prominent literature related to hidden Markov-switching models (HMM) and outlines how the HMM is adapted to econometrics and consequently used for volatility modeling. This chapter introduces the primary building blocks of a HMM volatility model, namely the classical normal distribution[21], the finite mixture model, the Generalized Autoregressive Conditional Heteroskedasticity (GARCH)[22] model and the Markov switching GARCH models. In particular, this chapter reviews the following:-

1) A history of HMM and its initial application to econometrics problems: This is thoroughly addressed by the seminal work of Hamilton (1989, 1990).
2) The finite mixture model: A review of a two-component *i.i.d.* mixture normal distribution is presented, one whose state probabilities are fixed and one whose volatilities at each state are also constant.
3) Volatility link function and appropriate [G]ARCH models: This includes the classical GARCH model (Bollerslev, 1986), the absolute return GARCH model (Taylor, 1986), the Exponential GARCH model (Nelson, 1991), GARCH with Student's *t* distribution (Bollerslev, 1987) and empirical works on the Integrated GARCH model (Baillie, 1996).
4) GARCH-MixN: This is a GARCH model whose innovation is a two-component mixture normal.
5) MS-[G]ARCH: Two classes of MS-[G]ARCH models are then specifically

[21] or perhaps alternative distributions like the student's *t* distribution
[22] The GARCH feature is optional.

discussed, namely the approach of Hamilton/Gray/Klaassen and that of Haas/Mittnik/Paolella (2004b).

6) Parallel vs. Non-parallel: The strengths and limitations of existing models are then discussed in the context of the proposed parallel model.

3.1. Hidden Markov-switching model: A brief history

The Hidden Markov-switching model (HMM) was introduced in its full generality in 1966 by Baum and Petire. Within the HMM framework, the observation sequence depends on the Markov chain probabilistic structure. In 1970, Baum, Petrie, Soules and Weiss developed forward-backward recursions for calculating the conditional probability of a state given an observation sequence from a general HMM.

A HMM is a discrete-time finite-state Markov chain, observed through a discrete memoryless channel. The channel is characterized by a finite set of transition densities indexed by the states of the Markov chain. A HMM is comprised of a rich family of parametric processes and has found uses in numerous econometrics applications.

3.2. Seminal work on HMM in Econometrics

In an economy, the states before and after a structure break tend to repeat themselves, that is, if it occurs once, it is likely to recur. For example, a financial market may shift from the (relatively) turbulent state to a tranquil state, only to return to the turbulent state at a later date. It is desirable to model these states as stochastic, dynamic, cyclical and

unobservable processes in order to mimic real world time series. The regime-switching model provides an excellent analytical framework to capture such assumed probabilistic state transitions over time.

The regime-switching model was introduced to the econometrics mainstream by Hamilton (1989), who proposed the model to investigate postwar U.S. real GNP growth[23]. According to the presented empirical results, it was shown that the growth rates of real U.S. GNP are subject to autocorrelated discrete shifts. Furthermore, the results illustrated that the business cycle is better characterized by a recurrent pattern of such shifts between a recessionary state and a growth state, rather than by positive coefficients at low lags in an autoregressive model.

In 1990, Hamilton formally generalized the idea of his 1989 paper and suggested that the EM algorithm can also be employed to estimate the parameters of his models. An attractive merit of the EM algorithm lies in its ability to simplify a complex computational problem and stabilize the estimation process. The proposed HMS-V of this thesis also adopts the EM algorithm, however it is used in a far more efficient manner than in previous implementations (as presented later in section 4.1).

Since the 1990s, the regime-switching models based on Hamilton's framework have gained in popularity. They have been adopted to study various economic time series as well as to endogenize the ergodic structure breaks (e.g. the structure breaks of volatility).

[23] Since then, Markov states are referred to as "regimes" in econometrics, and a HMM is therefore also known as a "regime-switching" model accordingly. In this thesis, the two words "regime" and "state" are often used interchangeably.

3.3. Prominent volatility models

A regime-switching volatility model is a complex mixture regression model. Apart from the basic HMM model, to enhance the understanding of each building block, some relevant volatility models are discussed here.

3.3.1. The i.i.d. mixture distribution

Researchers have used various unconditional and conditional fat-tailed/skewed distribution assumptions, for example, the Student's t distribution, skewed-Student's t distribution, general error distribution, hyperbolic approach, and the mixture normal distribution amongst others.

The two-component mixture normal distribution is a natural evolution of the classical normal distribution. The *p.d.f.* of a mixture normal distribution is

$$MixN\left(v\middle|\lambda_1,\mu_1,\sigma_1^2,\mu_2,\sigma_2^2\right)=\lambda_1\bullet\phi\left(v\middle|\mu_1,\sigma_1^2\right)+\left(1-\lambda_1\right)\bullet\phi\left(v\middle|\mu_2,\sigma_2^2\right), \tag{3.1}$$

where $\phi\left(\bullet\middle|\mu,\sigma^2\right)$ is the *p.d.f.* of a normal distribution, λ_1 is the probability that v is generated by the first component, and $\lambda_2=\left(1-\lambda_1\right)$ is the probability that v is generated by the second component. For an *i.i.d.* two-component mixture normal distribution with constant weight, the two variables $\left(v,\ \lambda_1\right)$ are pair-wise independent. The first four moments of the mixture normal distribution are:-

$$E\left(v\right)=\mu=\sum_{i=1}^{2}\lambda_i\bullet\mu_i\ , \tag{3.2}$$

$$Var\left(v\right)=\sigma^2=\sum_{i=1}^{2}\lambda_i\bullet\left[\sigma_i^2+\mu_i^2\right]-\mu^2\ , \tag{3.3}$$

$$Skewness(v) = \frac{1}{\sigma^3} \bullet \sum_{i=1}^{2} \lambda_i \bullet \left[\mu_i - \mu \right] \bullet \left[3\sigma^3 + \left(\mu_i - \mu \right)^2 \right], \text{ and} \tag{3.4}$$

$$Kurtosis(v) = \frac{1}{\sigma^4} \bullet \sum_{i=1}^{2} \lambda_i \bullet \left[3\sigma_i^4 + 6\left(\mu_i - \mu \right)^2 \bullet \sigma_i^2 + \left(\mu_i - \mu \right)^4 \right]. \tag{3.5}$$

These moments can take various values and the density function (3.1) is capable of describing a great spectrum of fat-tailed and/or skewed distributions[24], as shown in *figure 1*. The model is referred to as the "Normal Mixture Diffusion" (MixN).

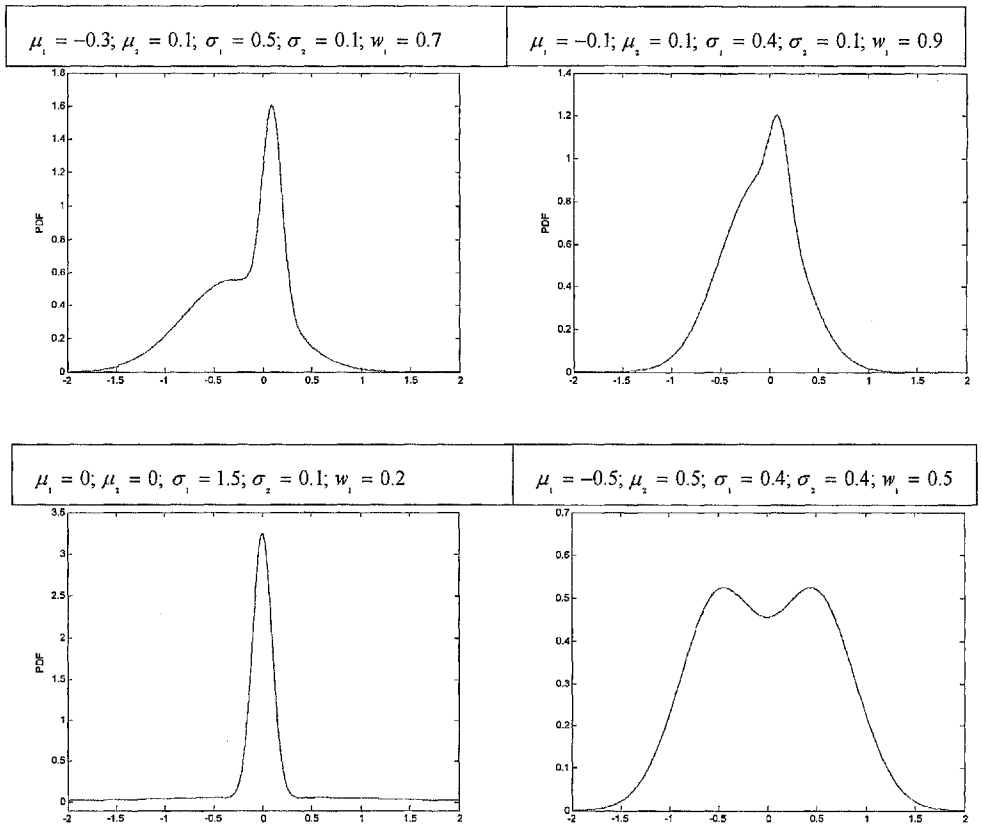


Figure 1 *p.d.f. of MixN with different parameter settings*

***upper-left panel**: highly skewed and fat-tailed*	***upper-right panel**: slightly skewed and fat-tailed*
***lower-left panel**: zero skewness but highly fat-tailed*	***lower-right panel**: multi-mode mixture*

[24] Please refer to **McLachlan G. and Peer D.,** *Finite Mixture Models*, 2000, *New York: Wiley.*

3.3.2. Classic [G]ARCH models

Apart from leptokurtosis and skewness, the simple *i.i.d.* MixN still fails to account for other stylized facts of financial returns, most notably, clustered volatility. Perhaps one of the most popular models which explicitly address clustered volatility is the [generalized] autoregressive conditional heteroscedasticity ([G]ARCH). The [G]ARCH family was pioneered by Engle (1982), generalized by Bollerslev (1986) and subsequently extended by a number of researchers. Furthermore, the model accounts for some of the stylized facts of financial time series returns outlined in the introduction, but from a significantly differing perspective than that of the proposed model. For a GARCH model, the volatility is a deterministic, continuous and smooth function of the past innovations. It is therefore more suited to describe *smooth and linear* changes in variance. On the other hand, the proposed HMS-V model, allows for radical and nonlinear structure breaks in variance.

[G]ARCH model, Bollerslev (1986)

The GARCH (k,h) model proposed by Bollerslev (1986) takes the following form,

$y_t = \mu + \varepsilon_t$,

$\varepsilon_t = \sigma_t \bullet \eta_t$, and $\eta_t \overset{i.i.d}{\sim} N(0,1)$,

$$\sigma_t^2 = \beta_0 + \sum_{i=1}^{k} \beta_i \bullet \varepsilon_{t-i}^2 + \sum_{j=1}^{h} \beta_{k+j} \bullet \sigma_{t-j}^2 \qquad (3.6)$$

where $\beta_i \geq 0$ for $i = 0,1,\cdots,k+h$ is the non-negative condition. The conditional variance σ_t^2 is thus linearly dependent on ε_{t-i}^2 and σ_{t-j}^2.

Since a vast number of alternative GARCH formulae exist in the literature, only those most relevant to the proposed HMS-V model are discussed here.

Modeling the absolute return

Taylor (1986), amongst others, suggests modeling the autocorrelation of absolute return directly,

$$\sigma_t = \beta_0 + \sum_{i=1}^{k} \beta_i \bullet |\varepsilon_{t-i}| + \sum_{j=1}^{h} \beta_{k+j} \bullet \sigma_{t-j}, \qquad (3.7)$$

where $\beta_i \geq 0$ for $i = 0,1,2$. The functional form of equation (3.7) is the same as GARCH except that the linear relationship is imposed on the standard deviation instead of the variance.

Standard deviation based model as opposed to variance-based the variance possesses many potential advantages. One motivating factor for adopting standard deviation-based models can be linked to the least absolute deviation versus least squared approach, of which the former is more robust to the presence of outliers. Another factor could be that the absolute return is a more direct measure of risk than the squared return, which was studied by Granger and Ding (1993).

EGARCH (Exponential GARCH)

One of the difficulties encountered by the linear [G]ARCH model of equation (3.6) is that without non-negative restrictions, the MLE coefficients for the innovation terms are often found to be negative. This same critic also applies to equation (3.7).

To relax the non-negativity restriction and to emphasize asymmetry of volatility in response to news, Nelson (1991) presented an exponential GARCH model (EGARCH),

$$\log(\sigma_t) = \beta_0 + \sum_{i=1}^{k} \beta_i \bullet g(\nu_{t-i}) + \sum_{j=1}^{h} \beta_{k+j} \bullet \log(\sigma_{t-j}), \qquad (3.8)$$

where $g(v_t) = a \bullet \eta_t + b \bullet \left[|\eta_t| - E|\eta_t| \right]$. Note that due to the properties of the exponential distribution, non-negative constraints are no longer required. Similarly, the proposed HMS-V model also does not require non-negative constraints. A further consequence of using an exponential form is that it is identical to direct standard deviation or variance modeling (e.g. Taylor, 1986)[25].

GARCH-T (GARCH with Student's t distributed innovation)

An attractive feature of the [G]ARCH process is that while the conditional distribution of the error is normal, the unconditional distribution is fat-tailed. In spite of this attractive feature, empirical results using high frequency data often indicate that the implied unconditional distributions of the estimated [G]ARCH models are insufficiently leptokurtotic to represent the returns distribution. Bollerslev (1987) therefore adopted a conditional Student's *t* distribution,

$$y_t = \mu + \varepsilon_t ,$$

$$\varepsilon_t = \sigma_t \bullet \eta_t ,$$

$$\sigma_t^2 = \beta_0 + \sum_{i=1}^{k} \beta_i \bullet \varepsilon_{t-i}^2 + \sum_{j=1}^{h} \beta_{k+j} \bullet \sigma_{t-j}^2 \qquad (3.9)$$

where η_t follows an *i.i.d.* Student's *t* distribution with unity variance and DoF degrees of freedom, whose density function is,

$$\frac{\Gamma\left(\frac{DoF+1}{2}\right)}{\sqrt{\pi(DoF-2)} \bullet \Gamma\left(\frac{DoF}{2}\right)} \bullet \left(1 + \frac{\eta^2}{(DoF-2)}\right)^{-(DoF+1)/2} \qquad (3.10)$$

[25] Specifying $\sigma_{(i),t} = \exp\left(\mathbf{z}_t \boldsymbol{\beta}_{(i)}\right)$ is equivalent to specifying $\sigma_{(i),t}^2 = \exp\left(2 \bullet \mathbf{z}_t \boldsymbol{\beta}_{(i)}\right)$.

IGARCH (Integrated GARCH)

According to the [G]ARCH model of equation (3.6), the impact of current innovation on future volatilities is measured by persistence where,

$$persistence = \sum_{i=1}^{k} \beta_i + \sum_{j=1}^{h} \beta_{k+j} \,. \qquad (3.11)$$

For a GARCH(1,1) model, equation (3.11) can be simplified to $\beta_1 + \beta_2$. If $\beta_1 + \beta_2 = 1$, the conditional variance grows linearly with the forecast horizon, in which case the GARCH(1,1) model is referred to as the Integrated GARCH (IGARCH). The IGARCH model assumes:

$y_t = \mu + \varepsilon_t$,

$\varepsilon_t = \sigma_t \bullet \eta_t$, and $\eta_t \overset{i.i.d}{\sim} N(0,1)$,

$\sigma_t^2 = \beta_0 + \beta_1 \bullet \varepsilon_{t-1}^2 + \beta_2 \bullet \sigma_{t-1}^2$,

$\beta_1, \beta_2 > 0$ *and* $\beta_1 + \beta_2 = 1$.

Following a review of empirical results in the literature, it can be notices that the high persistence level of GARCH is not at all ubiquitous (Baillie, 1996). These empirical findings suggest that volatility is also perhaps fractionally integrated and has a long memory. This observation was also reflected in the fractionally integrated GARCH model (FIGARCH by Baillie, Chung and Tieslau, 1996).

3.3.3. [G]ARCH with mixture normal innovation

The volatility at each component of the MixN model discussed previously in section 3.3.1

is fixed, however the volatility of a [G]ARCH model in section 3.3.2, with only one component, is dynamic. MixN and GARCH can be integrated in such way that the volatility of the MixN model is embedded in a GARCH process.

Vlaar and Palm (1993) are likely the first to suggest the use of a mixture normal innovation in a GARCH context (GARCH-MixN). The model permits non-zero skewness by allowing the component means to be non-zero. Bauwens, Bos, and Dijk (1999) considered a GARCH-MixN model with two components, where the component variances are proportional to each other. Wong and Li (2000, 2001) proposed a model that allows for nonlinear dynamics in the mean.

Recently, Haas, Mittnik, and Paolella (2004a) also considered a GARCH-MixN for a univariate time series coupled with a multivariate GARCH type structure. Employing daily return data of the NASDAQ index, the GARCH-MixN generated a plausible disaggregation of the conditional variance process in which a component's volatility dynamics had a clearly distinct and meaningful behavior.

Unfortunately, as was the case for the GARCH-T model, the GARCH-MixN is unable to account for the long memory of financial time series. The long memory behavior of financial time series can be better incorporated by a regime switching volatility model.

3.3.4. Regime-switching volatility models

A more dynamic approach is to allow both the component weight of a MixN and each components volatility to be time-varying. The proposed model and the MS-[G]ARCH

models both belong to this category. Such enriched dynamics helps to explain the long memory process of an IGARCH or an FIGARCH model reviewed in section 3.3.2. The regime-switching model provides an alternative method of considering structure breaks and the long memory of a financial time series.

The consistent observation of very large persistence in the variance of financial time series is perplexing. Diebold (1986) suggested that the failure to accommodate shifts in monetary policy regimes, reflected by changes in the constant term of the conditional variance equation, might result in an erroneous estimate of the integrated interest rate volatility. Diebold and Inoue (2001) showed analytically that stochastic regime switching is easily confused with the long memory process. Granger and Hyung (1999) also analyzed how high persistence can be incorrectly found in the presence of breaks.

To illustrate the above argument, a simple Monte Carlo experiment is carried out to demonstrate an example of the relationship between the ergodic structure breaks, long memory and a potentially mis-specified IGARCH process. Suppose that:-

- there are only occasional structure breaks and the differences between the two regimes are significant; the first states follows $i.i.d.\ N(0,1)$ and the second state follows $i.i.d.\ N(0,2^2)$;
- the sample size of the time series is 10,000 with two underlying states, and
- a regime shift takes place for every 1,000 observations.

Figure 2 plots the simulated time series and the sample autocorrelation functions (ACF) of the squared return up to 700 lags. The sample ACF does not die down exponentially[26],

[26] Theoretically, a short memory GARCH model indicates exponential decaying ACF.

and it diminishes to 0 only after a few hundred lags. Such a pattern of sample ACF is exhibits similarities to an IGARCH or a FIGARCH process.

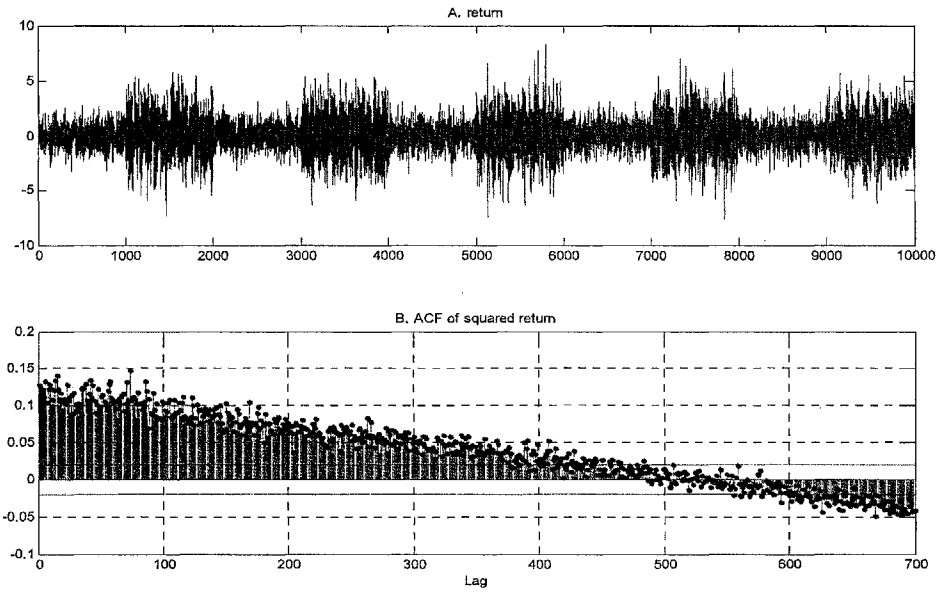


Figure 2
Top panel: the simulated time series with 10,000 observations

Bottom panel: the sample ACFs (long memory) of the squared return

Fitting the plotted data to the GARCH(1,1) –Normal model, an almost unity persistence (IGARCH process)[27] can be found,

$$\sigma_t^2 = \underset{(0.001)}{4.46 \bullet 10^{-3}} + \underset{(0.003)}{0.032} \bullet \varepsilon_{t-1}^2 + \underset{(0.003)}{0.967} \bullet \sigma_{t-1}^2, \text{ and} \qquad (3.12)$$

$LLF = -17868.2$.

[27] $0.032 + 0.967 = 0.999$

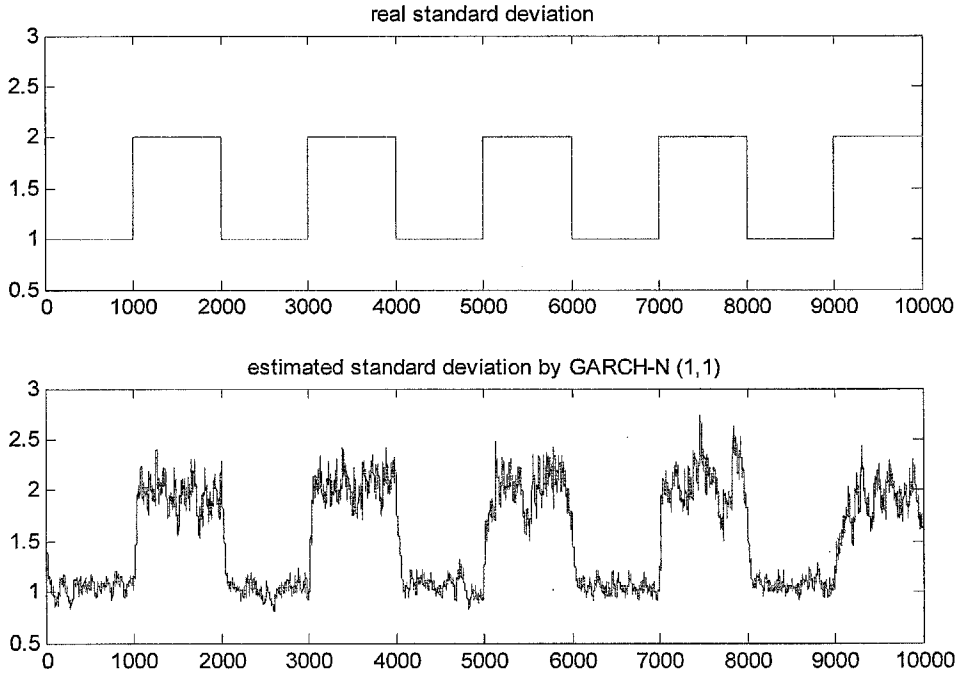


Figure 3
Top panel: the true standard deviation

Bottom panel: the estimated standard deviation by the GARCH(1,1)-Normal model.

In equation (3.12), the coefficient of the ARCH term, which absorbs innovations, is close to zero, while the GARCH term, which reserves the accumulated information, is close to 1. Note that a GARCH model can also be regarded as a state updating process on volatility. In equation (3.12), σ_{t-1} can be regarded as the state variable at time t, and the GARCH term, $\widehat{\beta}_2 = 0.967$, *linearly* disposes the state information. The ARCH term, $\widehat{\beta}_1 = 0.032$, then *linearly* learns the new innovations, ε_t.

Figure 3 suggests that:-

- when the underlying regime does not shift and the "true" volatility does not change, the estimated volatility by the GARCH model fluctuates largely. This is due to the

inability of the GARCH to cease *absorbing* the innovations in a *linear* speed, and furthermore its inability to distinguish the importance (relevance) of the new shock

- when the underlying regime suddenly changes, the GARCH model also absorbs the shock *linearly*, which is not sufficiently fast enough since the true volatility should instantaneously switch to a completely different state.

In this example, the "long memory" indicated by sample ACF is caused by the periodic structure breaks. The GARCH model tries to *linearly* fit the occasional abrupt changes with continuous gradual changes, while the estimated parameters suggest an IGARCH process. In order to rigorously address the non-linear structure breaks more, researchers have developed models where the coefficients of the [G]ARCH process are dynamic and governed by a hidden Markov process, i.e., Markov-switching [G]ARCH models.

3.4. Two classes of popular Markov-switching [G]ARCH approaches

3.4.1. The Hamilton/Gray/Klaassen approach

3.4.1.1. Model Overview

Hamilton and Susmel (1994) proposed that the spurious high persistence obtained by classic GARCH models can be mitigated by a hidden Markov-switching ARCH process. For a two-state Markov-switching GARCH (1,1) model, Hamilton and Susmel (1994) assumed the following data generating process for $\{\varepsilon_t\}$:

$$\varepsilon_t = \eta_t \bullet \sigma_t, \qquad \eta_t \overset{i.i.d}{\sim} N(0,1),$$

$$\sigma_t^2 = \sum_{i=1}^{2} I\left(\Delta_{t-1} = i\right) \bullet \left(\beta_{\langle i \rangle,0} + \beta_{\langle i \rangle,1} \bullet \varepsilon_{t-1}^2 + \beta_{\langle i \rangle,2} \bullet \sigma_{t-1}^2\right) \tag{3.13}$$

where $I\left(\Delta_t = i\right)$ is an indicator function. Iteratively expanding equation (3.13) yields

$$\begin{aligned} \sigma_t^2 = \sum_{k=0}^{t-1} \Bigg\{ \left[\sum_{i=1}^{2} I\left(\Delta_{t-k} = i\right) \bullet \left(\beta_{\langle i \rangle,0} + \beta_{\langle i \rangle,1} \bullet \varepsilon_{t-1-k}^2\right) \right] \bullet \prod_{g=0}^{k-1} \left[\sum_{i=1}^{2} I\left(\Delta_{t-g} = i\right) \bullet \beta_{\langle i \rangle,2} \right] \\ + \sigma_0^2 \bullet \prod_{k=0}^{t-1} \left[\sum_{i=1}^{2} I\left(\Delta_{t-k} = i\right) \bullet \beta_{\langle i \rangle,2} \right] \Bigg\}, \end{aligned} \tag{3.14}$$

such that σ_t^2 depends on the entire regime history. The evaluation of the likelihood function for a sample of length T requires the integration over all 2^T possible (unobservable) regime paths, rendering the estimation of equation (3.14) computationally intractable in practice. Therefore, Hamilton and Susmel (1994) restricted the model specification to MS-ARCH models. Only the four most recent innovations were taken into account (from ε_{t-1}^2 to ε_{t-4}^2), i.e., an MS-ARCH(4) model. To avoid over-parameterization, the volatilities at different regimes were assumed always proportional,

$$\sigma_t = \sqrt{b_{\langle i \rangle,t}} \bullet \tilde{\tilde{\sigma}}_t \tag{3.15}$$

The underlying ARCH (4) variable $\tilde{\tilde{\sigma}}_t$ is then multiplied by the constant $\sqrt{b_{\langle 1 \rangle}}$ if $\Delta_t = 1$, and by $\sqrt{b_{\langle 2 \rangle}}$ if $\Delta_t = 2$. As a result, σ_t^2 is a function of 4 lagged terms of state probabilities,

$$\begin{aligned} \sigma_t^2 &= E\left\{\varepsilon_t^2 \middle| \Delta_t, \Delta_{t-1}, \cdots \Delta_{t-4}, \varepsilon_{t-1}, \cdots \varepsilon_{t-4}\right\} \\ &= b_{\langle \Delta t \rangle} \left\{\alpha_0 + \alpha_1 \bullet \sigma_{\langle \Delta_{t-1} \rangle,t}^2 + \cdots + \alpha_4 \bullet \sigma_{\langle \Delta_{t-4} \rangle,t}^2\right\}, \\ &= \sigma_t^2\left(\Delta_t, \Delta_{t-1}, \cdots \Delta_{t-4}\right). \end{aligned} \tag{3.16}$$

$\tilde{\tilde{\sigma}}_t = \tilde{\tilde{\sigma}}_t\left(\Delta_t, \Delta_{t-1}, \cdots \Delta_{t-4}\right)$;

$$\sigma_{\langle 1\rangle,t} = b_{\langle 1\rangle,t} \bullet \tilde{\tilde{\sigma}}_t\left(\Delta_t, \Delta_{t-1}, \cdots \Delta_{t-4}\right);$$

$$\sigma_{\langle 2\rangle,t} = b_{\langle 2\rangle,t} \bullet \tilde{\tilde{\sigma}}_t\left(\Delta_t, \Delta_{t-1}, \cdots \Delta_{t-4}\right).$$

Gray (1996)

To circumvent the path dependence and generalize the Markov-switching ARCH (MS-ARCH) model into a Markov-switching GARCH (MS-GARCH) model, Gray (1996) proposed a simple non-path dependence collapsing scheme. The basic idea of the collapsing scheme is to calculate the expected value of the volatility at time t based on information up to $t-1$,

$$\begin{aligned}\sigma_t^2 &= E\left[\varepsilon_t^2\right] - E^2\left[\varepsilon_t\right] \\ &= \sum_{i=1}^{2} \Pr\left(\Delta_t = i \middle| \psi_{t-1}\right) \bullet \left(\mu_{\langle i\rangle,t}^2 + \sigma_{\langle i\rangle,t}^2\right) - \left(\sum_{i=1}^{2} \Pr\left(\Delta_t = i \middle| \psi_{t-1}\right) \bullet \mu_{\langle i\rangle,t}\right)^2\end{aligned} \tag{3.17}$$

Note that equation (3.17) is essentially derived from equation (3.14) by taking the following expectation,

$$\begin{aligned}\sigma_t^2 &= E\left[\sigma_t^2 \middle| \psi_{t-1}\right] \\ &= \sum_{i=1}^{2} \left\{ E\left[\Delta_t = i \middle| \psi_{t-1}\right] \bullet \left(E^2\left[\mu_{\langle i\rangle,t} \middle| \psi_{t-1}\right] + E\left[\sigma_{\langle i\rangle,t}^2 \middle| \psi_{t-1}\right] \right) \right\} \\ &\qquad - \left\{ \sum_{i=1}^{2} E\left[\Delta_t = i \middle| \psi_{t-1}\right] \bullet E\left[\mu_{\langle i\rangle,t} \middle| \psi_{t-1}\right] \right\}^2\end{aligned} \tag{3.18}$$

Gray (1996) used equation (3.17) to replace equation (3.16) with a GARCH process for the volatility at both states,

$$\sigma_{\langle i\rangle,t}^2 = \beta_{\langle i\rangle,0} + \beta_{\langle i\rangle,1} \bullet \varepsilon_{\langle i\rangle,t-1}^2 + \beta_{\langle i\rangle,2} \bullet \sigma_{t-1}^2 \qquad i = 1,2 \tag{3.19}$$

where $\varepsilon_{\langle i\rangle,t-1} = y_{t-1} - \mu_{\langle i\rangle,t-1}$.

equation (3.19) suggests that:

$$\begin{aligned}\sigma^2_{\langle i\rangle,t} &= E\left[\sigma^2_{\langle i\rangle,t}\middle|\psi_{t-1}\right] \\ &= \beta_{\langle i\rangle,0} + \beta_{\langle i\rangle,1}\bullet\varepsilon^2_{\langle i\rangle,t-1} + \beta_{\langle i\rangle,2}\bullet E\left[E\left(\sigma^2_{t-1}\middle|\psi_{t-2}\right)\middle|\psi_{t-1}\right] \\ &= \beta_{\langle i\rangle,0} + \beta_{\langle i\rangle,1}\bullet\varepsilon^2_{\langle i\rangle,t-1} + \beta_{\langle i\rangle,2}\bullet E\left[\sigma^2_{t-1}\middle|\psi_{t-2}\right]\end{aligned} \tag{3.20}$$

Gray therefore managed to obtain the expected volatility iteratively from equation (3.18) and equation (3.20) and he generalized the ARCH structure in equation (3.16) to the GARCH structure. In addition, with the iterative collapsing scheme the proportional relationship between the volatility at State 1 and the contemporary volatility at State 2[28] need not be assumed. Such model specification further enriches the dynamics of the MS-GARCH model. According to Gray's empirical study on the US short-term T-bill rate, when the interest rate is high, as in the case of an expected oil price shock and the Fed experiment during 1979 and 1982[29], the short-term US T-bill rate resembles the GARCH model.

To gain a more intuitive understanding of Gray's approach, suppose there are only two states and $\mu = \mu_{\langle 1\rangle} = \mu_{\langle 2\rangle} = 0$. Equation (3.18) and equation (3.20) can then be simplified to:

$$\begin{aligned}\sigma^2_t &= E\left[\sigma^2_t\middle|\psi_{t-1}\right] \\ &= \sum_{i=1}^{2}\left\{E\left[\Delta_t = i\middle|\psi_{t-1}\right]\bullet E\left[\sigma^2_{\langle i\rangle,t}\middle|\psi_{t-1}\right]\right\}\end{aligned} \tag{3.21}$$

$$\begin{aligned}\sigma^2_{\langle i\rangle,t} &= E\left[\sigma^2_{\langle i\rangle,t}\middle|\psi_{t-1}\right] \\ &= \beta_{\langle i\rangle,0} + \beta_{\langle i\rangle,1}\bullet y^2_{t-1} + \beta_{\langle i\rangle,2}\bullet E\left[E\left(\sigma^2_{t-1}\middle|\psi_{t-2}\right)\middle|\psi_{t-1}\right] \\ &= \beta_{\langle i\rangle,0} + \beta_{\langle i\rangle,1}\bullet y^2_{t-1} + \beta_{\langle i\rangle,2}\bullet E\left[\sigma^2_{t-1}\middle|\psi_{t-2}\right]\end{aligned} \tag{3.22}$$

The volatilities at the two states are intertwined because:-

[28] The proportional relationship in volatilities at different states is specified by Hamilton and Susmel (1994)

[29] The interested reader may refer to "The Incredible Volcker disinflation." by Goodfriend and Robert (2005) for a detailed background of the Fed monetary experiments from 1979 to 1982.

- in equation (3.22), σ_{t-1} is assumed to be equal to the volatility at time $t-1$ for both states, since the estimated value of σ_{t-1} is shared by both states to update $\sigma_{\langle 1\rangle,t}$ and $\sigma_{\langle 2\rangle,t}$
- in equation (3.21), σ_{t-1}^2 explicitly depends on $\sigma_{\langle 1\rangle,t-1}^2$ and $\sigma_{\langle 2\rangle,t-1}^2$.

Therefore, by the above recursive relationships, such an intertwined model specification reduces transparency.

Klaassen (2002)

Klaassen (2002) improved upon Gray's proposal by using more relevant information to forecast volatility. It was discovered that if Δ_t is highly auto-correlated, then the information at time *t* (e.g. the observation y_t), also provides valuable information for the state probability at time *t*, whereas the approach of Gray only exploited the information at time $t-1$..

Again, let a simplistic scenario assume,

- $\mu = \mu_{\langle 1\rangle} = \mu_{\langle 2\rangle} = 0$, and
- that at time *t*, the volatilities at both states for both Gray's and Klaassen's model are identical, that is, $E_{\text{Gray}}\left[\sigma_{\langle i\rangle,t}^2\right] = E_{\text{Klaassen}}\left[\sigma_{\langle i\rangle,t}^2\right] = E_{\text{Start}}\left[\sigma_{\langle i\rangle,t}^2\right]$.

The following table examines the primary differences between the approaches of Gray and Klaassen.

Gray (1996)	Klaassen (2002)
estimate σ_t based on observations $(y_1, y_2, \cdots, y_t)$:	
simply use $E\left[\Delta_t = i \middle\| \psi_{t-1}\right]$ to estimate σ_t^2 $E_{\text{Gray}}\left[\sigma_t^2 \middle\| \psi_t\right]$ $= \sum_{i=1}^{2}\left\{E\left[\Delta_t = i \middle\| \psi_{t-1}\right] \bullet E_{\text{Start}}\left[\sigma_{\langle i\rangle,t}^2\right]\right\}$	use all the information of ψ_t to calculate $E\left[\Delta_t = i \middle\| \psi_t\right]$ first; then $E_{\text{Klaassen}}\left[\sigma_t^2 \middle\| \psi_t\right]$ $= \sum_{i=1}^{2}\left\{E\left[\Delta_t = i \middle\| \psi_t\right] \bullet E_{\text{Start}}\left[\sigma_{\langle i\rangle,t}^2\right]\right\}$ Note that $E_{\text{Klaassen}}\left[\sigma_t^2 \middle\| \psi_t\right]$ is now more precise than $E_{\text{Gray}}\left[\sigma_t^2 \middle\| \psi_t\right]$.
Forecast $\sigma_{\langle i\rangle,t+1}$ based on observations $(y_1, y_2, \cdots, y_t)$ and $E\left[\sigma_t^2\right]$:	
$E_{\text{Gray}}\left[\sigma_{\langle i\rangle,t+1}^2 \middle\| \psi_t\right]$ $= \beta_{\langle i\rangle,0} + \beta_{\langle i\rangle,1} \bullet y_t^2 + \beta_{\langle i\rangle,2} \bullet E_{\text{Gray}}\left[\sigma_t^2 \middle\| \psi_t\right]$	$E_{\text{Klaassen}}\left[\sigma_{\langle i\rangle,t+1}^2 \middle\| \psi_t\right]$ $= \beta_{\langle i\rangle,0} + \beta_{\langle i\rangle,1} \bullet y_t^2 + \beta_{\langle i\rangle,2} \bullet E_{\text{Klaassen}}\left[\sigma_t^2 \middle\| \psi_t\right]$
Forecast σ_{t+1} based on observations $(y_1, y_2, \cdots, y_t)$ and $E\left[\sigma_{\langle i\rangle,t+1}^2 \middle\| \psi_t\right]$	
$E_{\text{Gray}}\left[\sigma_{t+1}^2 \middle\| \psi_t\right]$ $= \sum_{i=1}^{2} E\left[\Delta_{t+1} = i \middle\| \psi_t\right] \bullet E_{\text{Gray}}\left[\sigma_{\langle i\rangle,t+1}^2 \middle\| \psi_t\right]$	$E_{\text{Klaassen}}\left[\sigma_{t+1}^2 \middle\| \psi_t\right]$ $= \sum_{i=1}^{2} E\left[\Delta_{t+1} = i \middle\| \psi_t\right] \bullet E_{\text{Klaassen}}\left[\sigma_{\langle i\rangle,t+1}^2 \middle\| \psi_t\right]$
with the realization of y_{t+1}, estimate $E\left[\sigma_{t+1}^2 \middle\| \psi_{t+1}\right]$	
$E_{\text{Gray}}\left[\sigma_{t+1}^2 \middle\| \psi_{t+1}\right]$ $= \sum_{i=1}^{2}\left\{E\left[\Delta_{t+1} = i \middle\| \psi_t\right] \bullet E_{\text{Gray}}\left[\sigma_{\langle i\rangle,t+1}^2 \middle\| \psi_t\right]\right\}$	use ALL the information of ψ_{t+1} to calculate $E\left[\Delta_{t+1} = i \middle\| \psi_{t+1}\right]$ first; then $E_{\text{Klaassen}}\left[\sigma_{t+1}^2 \middle\| \psi_{t+1}\right]$ $= \sum_{i=1}^{2}\left\{E\left[\Delta_{t+1} = i \middle\| \psi_{t+1}\right] \bullet E_{\text{Klaassen}}\left[\sigma_{\langle i\rangle,t+1}^2 \middle\| \psi_t\right]\right\}$

Table 1

This table shows a simplified comparison between Gray's (1996) and Klaassen's (2002) Markov-switching GARCH model. Note that Klaassen's approach uses more relevant information to update the volatility.

As suggested by Klaassen (2002), the value of y_t provides highly valuable information to estimate the state probability at time *t*. *Table 1* summarizes how this is integrated into the Markov-switching GARCH model of Gray (1996). However, despite the differences,

the overall frameworks of the two Markov-switching GARCH models maintain some similarity.

3.4.1.2. Non-parallel structure and potential limitations

The Hamilton/Gray/Klaassen MS-GARCH models are elegant and versatile. Despite this, a few oversights are notable:-

(Oversight. 1) ***Same volatility link function for both INTER-state and INTRA-state dynamics:***

To portray this simplification, an analysis is performed where the Markov state transits from the tranquil state to the turbulent state, then remains in the turbulent state. *Figure 4* demonstrates an illustration of the oversight.

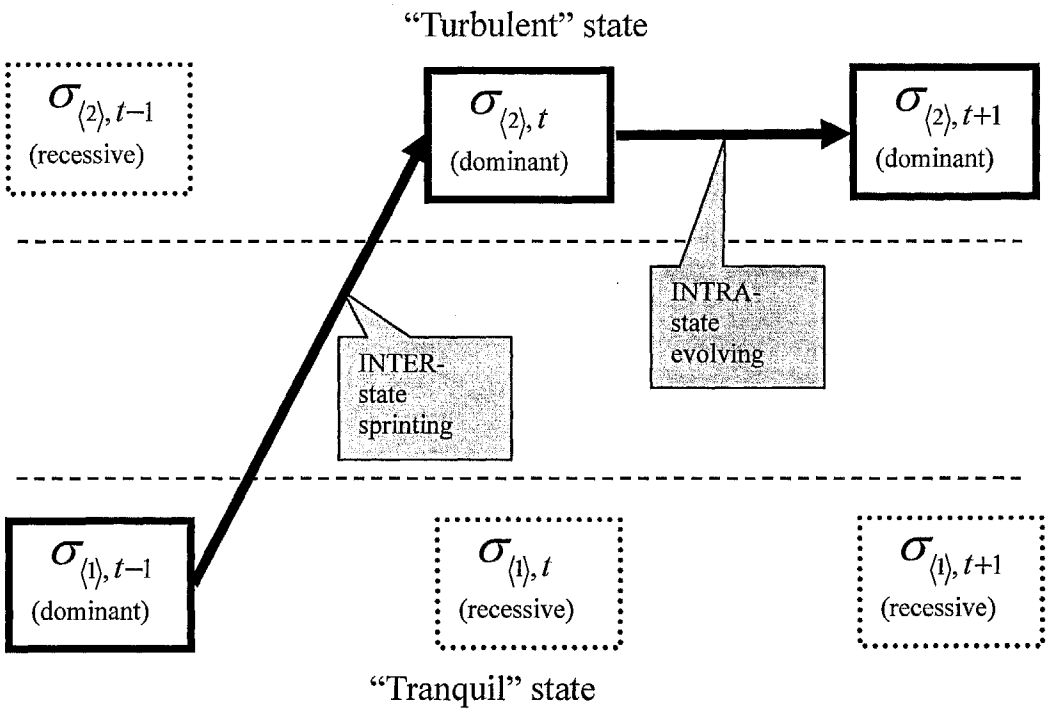


Figure 4

INTER-state and INTRA-state dynamics of Hamilton/Gray/Klaassen models (when the underlying hidden state is known). Suppose it is known that state(t-1) is tranquil, state(t) is turbulent, and

state(t+1) is turbulent. The textbox with the solid (dotted) boundary means the state is dominant (recessive).

As shown in *Figure 4*:

- from time $t-1$ to t, the volatility link function for the turbulent state is responsible for the INTER-state dynamics. It acts as a sprinter to describe the jump in volatility from the low volatility state to a high volatility state. The shock is completely absorbed by the volatility link function.
- from t to $t+1$, the volatility link function for the turbulent state is responsible for the INTRA-state dynamics. It describes the evolution within the turbulent state.

Therefore, the Gray/Klaassen model is more appropriate in the case of the INTER-state dynamics being the same or similar to the INTRA-state dynamics.

(Oversight. 2) ***Less transparent dynamics:***

Note that *Figure 4* only depicts the situation where the underlying state is known, whereas in reality the state probability has to be estimated, thus further reducing the tractability. *Figure 5* illustrates a similar situation as that of *figure 4*, however in this case the state is unknown and has to be estimated.

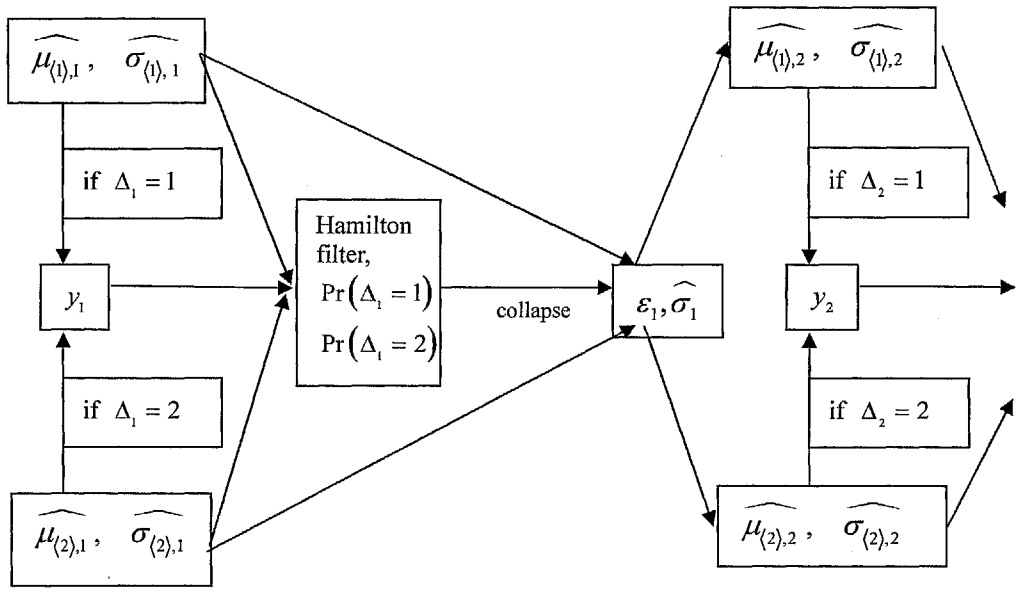


Figure 5
Gray/Klaassen Model (the underlying process is unknown and has to be estimated) .

(Oversight. 3) ***<u>Possibly excessive adaptation to recent signals :</u>***

According to the non-parallel model specification, it is likely that the volatility link functions of both states excessively absorb and adapt to recent history.

For example, to describe the dynamics of a stock index with the Gray/Klaassen model, a modeler often refers to a "tranquil state" (say State 1) and a "turbulent state" (say State 2). However, according to equation (3.19), the volatility of both states, $\sigma_{\langle 1\rangle,t}$ and $\sigma_{\langle 1\rangle,t}$, is based on the same value of σ_{t-1}.

$$\sigma^2_{\langle 1\rangle,t} = k_{\langle 1\rangle} + a_{\langle 1\rangle} \bullet \varepsilon^2_{\langle 1\rangle,t-1} + b_{\langle 1\rangle} \bullet \sigma^2_{t-1}$$
$$\sigma^2_{\langle 2\rangle,t} = k_{\langle 2\rangle} + a_{\langle 2\rangle} \bullet \varepsilon^2_{\langle 2\rangle,t-1} + b_{\langle 2\rangle} \bullet \sigma^2_{t-1} \qquad (3.23)$$

In order for the volatility of the second state to always exceed that of the first state, it is required that $k_{\langle 1\rangle} \le k_{\langle 2\rangle}$, $a_{\langle 1\rangle} \le a_{\langle 2\rangle}$, and $b_{\langle 1\rangle} \le b_{\langle 2\rangle}$. Otherwise, in the case of a prolonged tranquil period, the volatility of the turbulent state could excessively adapt to the

tranquility. The volatilities of the two states would hence become homogenous.

(Oversight. 4) ***Passive adaptation and potential missing memory:***

If both states excessively adapt to recent history, the relevance of distant history could be insufficiently represented. As the case study in chapter 5 shows, to invoke the memory of distant history, the model must passively absorb new "shocks". Prior to the new "shock", a long lasting tranquility can give the user of the model a deceptive sense of security.

In addition, as shown in *Figure 4*, the volatility link function at the turbulent state is required to describe both the INTER-state jump and INTRA-state evolution. For the INTER-state jump, the shock is completely absorbed by the volatility link function. Therefore, it is possible that the distant turbulence is memorized not as turbulence, but as a fast-adapting (jumping) process. This is explained in further detail in chapter 5.

(Oversight. 5) ***Reduced appeal of regime-switching:***

For the Gray/Klaassen model, according to equation (3.19) and *Figure 4*, it is possible that the volatility at the two states becomes less distinguishable and the clustering of Markov states fails to account for the clustering of volatility. The clustered volatility remains primarily explained by the volatility link functions instead of by the clustered Markov states.

In addition, statistically, if $p_{\langle 11 \rangle} + p_{\langle 22 \rangle} \approx 1$, then the Markov-switching model is similar to a single-regime GARCH-MixN model. If this is the case, regime-switching is less appealing, and the corresponding economic reasoning also becomes less relevant.

(Oversight. 6) ***Long memory not addressed by regime-switching:***

The volatility clustering and possibly long memory are primarily explained by the ARCH (∞) recursive volatility link functions, not by the clustering and switching of regimes.

(Oversight. 7) ***Possibly inconsistent economic implication for some time series:***

For some time series, the Gray/Klaassen may not provide a consistent economic implication. In practice, both the INTER-state and INTRA-state dynamics, and hence the economic implication may be significantly different, while Gray/Klaassen's approach suggests that they are the same (see *figure 4*).

As previously analysed in section 3.6.4, for the stock market, Maheu and Mccurdy (2000) and Bhar and Hamori (2004) amongst the others, suggested that the return data is generated by either a high-return tranquil state (say State 1) or a low-return turbulent state (say State 2). However, although such relationships are implicitly present in the data, the empirical study using the Gray/Klaassen approach may indicate otherwise. If there is prolonged tranquility, which has been excessively adapted to by the turbulent state, it is not unlikely that the volatility at the "tranquil" state is higher than the volatility at the "turbulent" state. Under such circumstances, the volatilities at the two states are no longer inline with the economic implication (e.g. "low volatility state" and 'high volatility state"). In section 5.6.2, an example of this is provided. The induced complication is more acute if it has been described that the "low volatility state" is associated with higher return.

(Oversight. 8) ***Excessive computational burden:***

As detailed in section 4.1.3, the non-parallel structure of the Markov-switching makes the estimation process non-parallel. With a non-parallel structure, it is difficult to decompose the complex computational process into more straightforward steps, and all the parameters for both the volatility link functions have to be simultaneously estimated. The computational burden increases, while the robustness of parameter estimation is reduced.

3.4.2. The Haas, Mittnik and Paolella approach

3.4.2.1 Model Overview

Haas, Mittnik and Paolella (2004b) are the first to propose a more parallel MS-GARCH model for volatility (but not for the expected return). To simplify the deduction and calculation, they initially fit the time series to an AR model, and then separated study the residual of the AR process. In this approach, they assume that the first and third moments of the adjusted time series equal to zero:-

$y_t = \varepsilon_t$,

$$E(y_t) = E(\varepsilon_t) = 0, \tag{3.24}$$

and $E^3(y_t) = E^3(\varepsilon_t) = 0$.

It is further assumed that:-

$$y_t = \varepsilon_t = \sum_{i=1}^{2} \eta_t \bullet \sigma_{\langle i \rangle,t} \bullet I(\Delta_t = i), \text{ and}$$

$$\sigma^2_{\langle i \rangle,t} = \beta_{\langle i \rangle,0} + \beta_{\langle i \rangle,1} \bullet \varepsilon^2_{t-1} + \beta_{\langle i \rangle,2} \bullet \sigma^2_{\langle i \rangle,t-1}; \qquad i = 1,2; t = 2,\cdots,T. \tag{3.25}$$

Equation (3.25) indicates that volatility within each state does not explicitly depend on the coefficients of the other state. The volatilities at both states interact with each other implicitly through the shared innovation and shock, ε_t. Hence the volatility link function is only used for the INTRA-state but not the INTER-state dynamics (see *Figure 6*), which

differs to that shown previously in *Figure 4*.

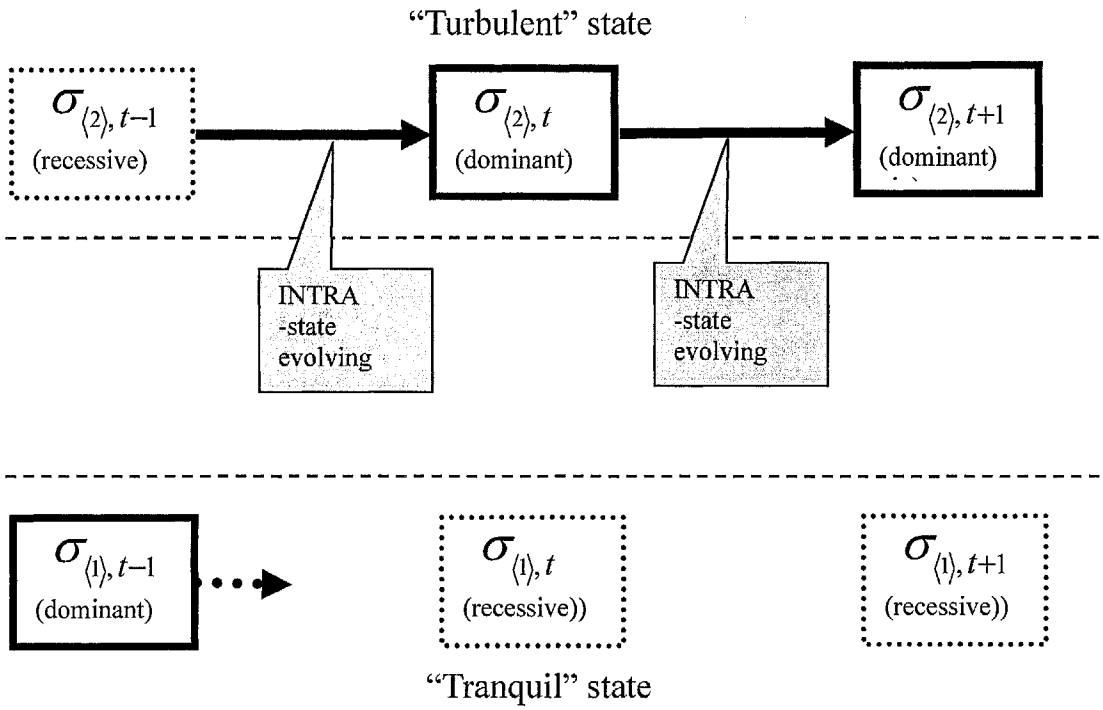


Figure 6

INTER-state and INTRA-state dynamics of Haas/Mittnik/Paolella (2004b), (when the underlying hidden state is known). Suppose it is known that State(t-1) is tranquil, State(t) is turbulent, and State(t+1) is turbulent. The textbox with solid (dotted) boundary means that the state is dominant (recessive). Therefore, from $t-1$ to t and from t to $t+1$, the volatility link function for the turbulent state describes the INTRA-state evolution.

As shown in *Figure 6*:-

- From time $t-1$ to time t, the switching of the regime is responsible for the INTER-state dynamics. The sudden jump in volatility can be mostly absorbed by the switching of the underlying process, as opposed to only by the switching of the parameters.
- From time t to time $t+1$, the volatility link function for the turbulent state is responsible for the INTRA-state dynamics.

Therefore, the Haas/Mittnik/Paolella (2004b) approach is more theoretically appropriate than that of Gray/Klaassen if the INTER-state and INTRA-state volatility dynamics are dissimilar.

3.4.2.2. Improvements/Limitations

The improvements and limitations of Haas/Mittnik/Paolella's (2004b) approach are discussed from two aspects: (1) the limitation imposed by assuming zero mean for each state; (2) the improvements and limitations in addressing the oversights listed in section 1.2.

According to the assumptions of Haas/Mittnik/Paolella (2004b), the innovation y_t always has zero mean implying that filtering of the residuals is unnecessary. However, to obtain such a time series, they run an autoregressive (AR) regression independently before using the MS-GARCH to study the residuals. Although such an approach greatly simplifies the estimation, the AR parameter is not estimated together with the MS-GARCH parameters by a MLE. Such an *ad hoc* treatment of the expected return is yet to be justified. In addition, the approach does not address the different patterns of expected return of a financial time series during a turbulent and tranquil state.

(Oversight. 1) ***Same volatility link function for both INTER-state and INTRA-state dynamics:***

As illustrated in *Figure 6*, Haas/Mittnik/Paolella's (2004b) approach addresses this shortcoming quite well: the INTER-state dynamics are addressed by the switching of the regime, while the INTRA-state dynamics are addressed by the volatility link function of each state. However, note that because Haas/Mittnik/Paolella's (2004b) assumes the same mean for both states, there is no INTER-state dynamics for the mean.

(Oversight. 2) ***Less tractable dynamics:***

According to Haas/Mittnik/Paolella's (2004b) approach, a volatility link function of each state only describes the INTRA-state dynamics, and the complexity has been reduced. *Figure 6* depicts a scenario in which the underlying state is known. For the case of the state probability being unknown, *Figure 7* provides the illustration.

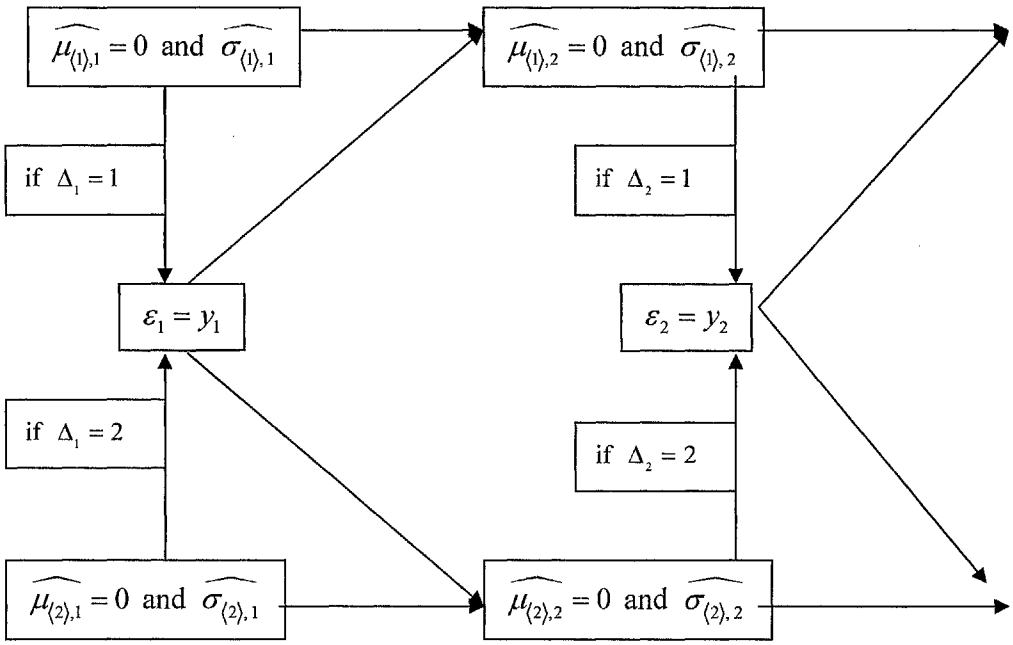


Figure 7
The model proposed by Haas, Mittnik, and Paolella's (2004b) (when the underlying process is unknown and has to be estimated)

(Oversight. 3) ***Possible excessive adaptation to recent signals:***

Haas/Mittnik/Paolella's (2004b) approach remains insufficiently parallel and the $\mathrm{ARCH}(\infty)$ [30] volatility link function(s) is likely to let both state excessively adapt to recent signals. There lacks an independent state to specifically address a scenario that resembles the distant history, in particular if there is prolonged tranquility or turbulence.

[30] Thus it is proposed to use only the most recent information instead of an ARCH (∞) process.

(Oversight. 4) ***Passive adaptation and potential missing memory:***

A relative independent state that emphasizes the distant history is also disregarded. Hence the memory of the distant history may be discarded and will only be passively invoked by new shocks.

(Oversight. 5) ***Reduced appeal of regime-switching:***

From comparison of *Figure 4* and *Figure 6*, a regime switch indicates a process switch, which increases the appeal of regime-switching when the jump in volatility is severe/ However, the Haas/Mittnik/Paolella's (2004b) model is insufficiently parallel due to the ARCH (∞) volatility link function. Note that according to the following equation:

$$\sigma^2_{\langle i\rangle,t} = \beta_{\langle i\rangle,0} + \beta_{\langle i\rangle,1} \bullet \varepsilon^2_{t-1} + \beta_{\langle i\rangle,2} \bullet \sigma^2_{\langle i\rangle,t-1}; \qquad i = 1,2; t = 2,\cdots,T. \qquad (3.26)$$
$$\sigma_{\langle 1\rangle,t} > \sigma_{\langle 2\rangle,t}$$

for (3.26) to be true, it is required that

$$\beta_{\langle 1\rangle,0} \geq \beta_{\langle 2\rangle,0}; \beta_{\langle 1\rangle,1} \geq \beta_{\langle 2\rangle,1}, \text{ or } \beta_{\langle 1\rangle,2} \geq \beta_{\langle 2\rangle,2} \qquad (3.27)$$

However, the empirical study does not always satisfy (3.27) or consequently(3.26). For example, Haas/Mittnik/Paolella (2004b) use their model to study the exchange rate of JPY/USDP:-

$$\mu_{\langle 1\rangle,t} = \mu_{\langle 2\rangle,t} = \mu_t;$$

$$\sigma^2_{\langle 1\rangle,t} = 0.003 + 0.023 \bullet \varepsilon^2_{t-1} + 0.945 \bullet \sigma^2_{\langle 1\rangle,t-1};$$
$$\sigma^2_{\langle 2\rangle,t} = 0.097 + 0.227 \bullet \varepsilon^2_{t-1} + 0.818 \bullet \sigma^2_{\langle 2\rangle,t-1}; \qquad (3.28)$$
$$\widehat{p_{\langle 11\rangle}} = 0.744; \qquad \widehat{p_{\langle 22\rangle}} = 0.285.$$

In this case, it is not known which state has a higher volatility and which state has a lower volatility. In addition, $p_{\langle 11\rangle} + p_{\langle 22\rangle} = 1.029 \approx 1$, which suggests that it is somewhat similar to a GARCH-MixN model.

(Oversight. 6) ***Long memory not addressed by regime-switching:***

Similar to the Gray/Klaassen model, if the difference between the two states is less distinctive, and $p_{\langle 11 \rangle} + p_{\langle 22 \rangle} \approx 1$, it is most likely that the state transition probability matrix is unable to address the clustering of hidden Markov states (see equation (3.28)). In this case, the volatility clustering and "long memory" is primarily addressed by the recursive ARCH (∞) volatility link functions, instead of the switching of regimes.

(Oversight. 7) ***Possible inconsistent economic implication for some time series:***

If the MS-GARCH model is more similar to a GARCH-MixN than to a regime-switching model, potential inconsistencies could be introduced in the economic implication of "regime-switching", especially for the interplay between the volatility and mean.

(Oversight. 8) ***Excessive computational burden:***

As shown in section 4.1.3, for Haas/Mittnik/Paolella's (2004b) approach, if the mean of the time series is zero, then the computational burden is far less than that of the Gray/Klaassen model. However, if the mean is non-zero and must be estimated together with other parameters, then the parameters of the two volatility link functions cannot be solved in a parallel manner. Such structure increases the computational burden.

3.5. Strengths of the proposed parallel hidden Markov-switching volatility model

To demonstrate the strengths of the proposed HMS volatility model, the following

discussion indicates how it specifically addresses some of the previously outlined oversights of previous models.

(Addressing Oversight 1) ***Same volatility link function for both INTER-state and INTRA-state dynamics:***

Both the proposed model and Haas/Mittnik/Paolella's (2004b) approach address this oversight quite well: the INTER-state dynamics are described by the switching of the regime, while the INTRA-state dynamics are described by the volatility link function of each state. The proposed model however, further improves upon Haas/Mittnik/Paolella's (2004b) by addressing the INTER-state dynamics of the mean.

(Addressing Oversight 2) ***Less tractable dynamics:***

As is the case with Haas/Mittnik/Paolella's (2004b) approach, the parallel structure enhances tractability. *Figure 8* plots the scenario when the hidden state has to be estimated, where it is evident that it is more tractable than the reviewed MS-GARCH model (see *Figure 4* and *Figure 6*)

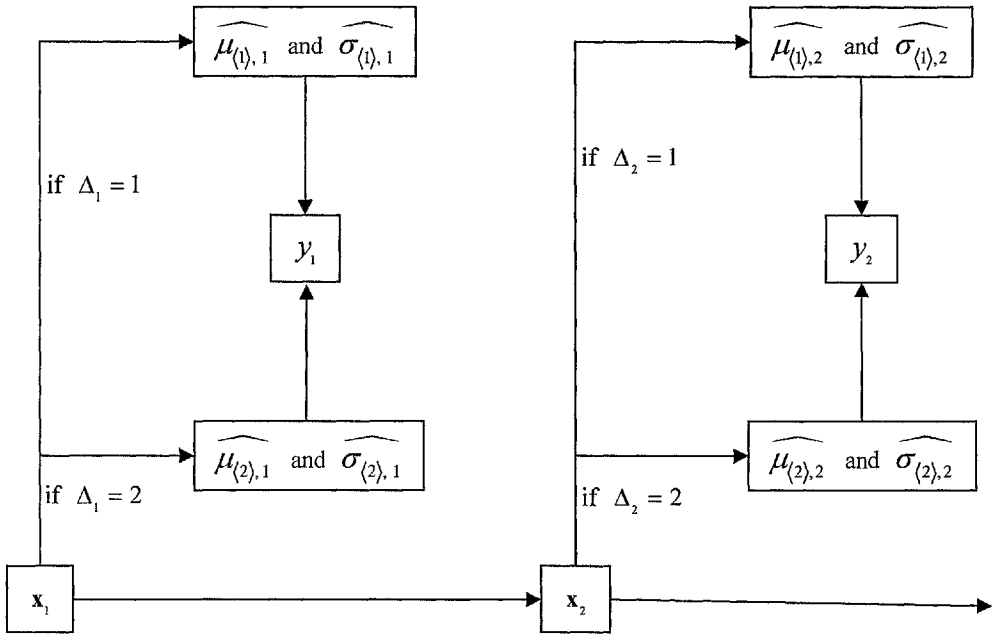


Figure 8
The proposed model (when the underlying process is unknown and has to be estimated)

(Addressing Oversight 3) ***Possible excessive adaptation to recent signals:***

The proposed parallel structure emphasizes a greater independency of the two regimes. Note that instead of an ARCH (∞) structure, the volatility link function only employs limited recent information (e.g. limited lag terms), hence the turbulent state will not adapt excessively to the observations at the tranquil state and the memory of turbulence is maintained[31].

(Addressing Oversight 4) ***Passive adaptation and hence possible missing memory:***

The proposed model has an increased ability to capture the distinctive differences between the two states. Despite one of the states aggressively adapting to recent history, the other state remains relatively independent and stores the memory of the distant history.

[31] There are a few alternative approaches based on MS-GARCH. One can impose constant volatility for the tranquil state, which does not prevent the turbulent state excessively learning from the tranquility. One can impose constant volatility for the turbulent state, but the description of turbulence would be too rigid and also have a much lower likelihood. One solution is a truncated GARCH model that was previously proposed by the author (not in this thesis) to describe the daily S&P 500 return.

The memory of the distant history will be always maintained, even without any shock. The case study in Chapter 5 further illustrates this enhancement.

(Addressing Oversight 5) ***Reduced appeal of regime-switching:***

The proposed parallel structure emphasizes the independence and distinctive differences of the two states. If the differences between two states are distinctive, a greater portion of volatility clustering will likely be explained by the clustering of regimes, and regime switching is more appealing.

(Addressing Oversight 6) ***Long memory not addressed by regime-switching:***

A long memory can be caused by the switching of a regime (e.g. see *Figure 2*), and/or an ARCH (∞) process with near unity persistence (see section 3.3). Since an ARCH (∞) link function is not adopted by the proposed model, the long memory will be primarily addressed by regime-switching. This form is comparable to portions of the ARCH (∞) structure of a GARCH-MixN model.

(Addressing Oversight 7) ***Possible inconsistent economic implication for some time series:***

Compared with the reviewed MS-GARCH models, the proposed model gives a more consistent economic implication. This is because (i) the INTER-state and INTRA-state dynamics, and hence their economic implication, have been described by the switching of regimes and the volatility link functions respectively; (ii) the volatility link function is more parallel and the two states are more likely to describe different volatility dynamics as well as economic implications; and (iii) the two states need not share the same mean,

and the economic implication of both the mean and volatility within different economic contexts can be investigated.

(Addressing Oversight 8) ***<u>Excessive computational burden:</u>***

As detailed in section 4.1.3, a parallel structure suggests independent optimization for parameter estimation, which requires less computational effort and provides greater stability.

Finally, few published articles on the MS-GARCH models discuss the parameter identifiability condition. A Student's *t* distribution possesses great advantages in modeling fat-tailed financial time series, but the identifiability condition for the mixture Student's *t* distribution within such a complex context is more challenging than for a mixture normal. For generality, only a normal distribution is considered in this work, and more complex conditional distributions are not analyzed. Appendix I analyses a sufficient condition of the model identifiability condition.

3.6. HMM empirical studies

In recent years, the HMM has been widely for solving problems in applied econometrics, in particular the Markov- switching volatility models. The model has been applied to describe the dynamics of business cycles, interest rates, yield curves, foreign exchange rates, stock index returns, credit and default, and commodity price.

3.6.1. Business Cycles

The notion that business cycles exhibit asymmetries is well established in economics. Recessions tend to be more pronounced and of a shorter duration than expansions, and recoveries appear to take a more moderate course than contractions.

Following the work by Hamilton (1989) on applying the Markov-switching model to non-linear business cycles, Clements & Krolzig (2003) proposed parametric tests for asymmetries based on Markov-switching processes, which are widely regarded as suitable for the investigation of variables subject to business cycle fluctuations.

Krolzig (2001) addressed the issue of identification and dating of the Euro-zone business cycle with the Markov-switching approach innovated by Hamilton (1989). Regime shifts in the stochastic process of economic growth in the Euro-zone were identified by fitting Markov-switching models to aggregated and single-country Euro-zone real GDP growth data over the last two decades. The models were found to be statistically congruent and economically meaningful and the resulting evidence of the presence of a common Euro-zone cycle was convincing.

3.6.2. Interest Rates and Yield Curves

Gray's (1996) work examined the short-term U.S. treasury interest rate. Christiansen (2004) used the MS-GARCH to study the yield curve. A bivariate two-state MS-GARCH model was proposed to investigate the relationship between the short rate changes and the yield curve slope. The two states were characterized by the variance of the short-rate changes, into low and high variance states. In the high-variance regime the yield curve

became steeper with the interest-rate variance whereas in the low-variance regime the slope was independent thereof.

3.6.3. Foreign Exchange Rates

Engle and Hamilton (1990) discovered that the value of the US Dollar (USD) appeared to move in one direction (resides in one regime) for long periods of time. Therefore they used a Markov-switching approach to model the segmented time trends. They rejected the null hypothesis that exchange rates follow a random walk behavior in favor of a regime-switching model.

Bollen, Gray and Whaley (2000) examined the ability of regime-switching models to encapsulate the dynamics of foreign exchange rates. They reported that a regime-switching model with independent shifts in mean and variance exhibited a closer fit and more accurate variance forecasts when compared to a range of other models. They also established that a simulated trading strategy based on regime-switching option valuation generated higher profits than standard single-regime alternatives.

The relationship between the exchange rate and central bank interventions was investigated by Beine, Laurent and Lecourt (2003). They used the MS-GARCH process to investigate whether official interventions can account for the observed volatility regime-switches. They determined that depending on the prevailing volatility level, coordinated central bank interventions can result in either a stabilizing or a destabilizing effect.

Dueker and Neely (2007) merged the literature on technical trading rules with the literature on Markov switching to develop economically useful trading rules. The Markov modes' out-of-sample returns modestly exceeded those of standard technical rules and were profitable over the most recent subsample.

3.6.4. Stock Index

Turner, Startz and Nelson (1989) investigated the risk and learning of market participants when the heteroskedasticity is caused by a hidden Markov process. They examined a variety of models in which the variance of a portfolio's excess return depended on a state variable generated by a first-order Markov process. Firstly, a model in which the state was known to the economic agent was estimated. It suggested that the mean excess return moved inversely to the level of risk. They then estimated a model in which agents are uncertain of the state. The model indicated that agents are consistently surprised by high-variance period, there is consequently a negative correlation between movements in volatility and excess returns.

Hamilton and Lin (1996) investigated the joint time series behavior or monthly stock returns and growth in industrial production. They found that economic recessions were the primary factor that drove fluctuations in the volatility of stock returns.

Maheu and Mccurdy (2000) studied the duration of different stock market sentiment. They developed a Markov-switching model that incorporated duration dependence to capture the non-linear structure in both the conditional mean and the conditional variance of stock returns. They assumed that the return data was generated by either a high-return

tranquil state or by a low-return turbulent state, which were labeled as bull and bear markets respectively. Their method identified all major stock-market downturns in over 160 years of monthly data. According to their findings, the best market gains come at the start of a bull market, which also has a declining hazard function. Furthermore, it was observed that volatility increases with the duration of the bear market.

Similarly to the duration argument, Bhar and Hamori (2004) used the hidden Markov-switching model to decompose stock returns into permanent and transitory components. According to their empirical study of four major economies, there was a substantial variation in the duration of the volatility states of the transitory component in these markets. The U.S. market was also reported to have a positive correlation with other markets for the permanent component, whereas it displayed a negative correlation with the temporal component.

Marcucci (2005) compared different GARCH models in terms of their ability to describe and forecast volatility of stock indices. To take into account the excessive persistence usually found in GARCH models (implying too smooth and too high volatility forecasts), he studied the MS-GARCH models and found that its parameters were allowed to switch between a low and high volatility regime. The empirical analysis demonstrated that MS-GARCH models substantially outperform standard GARCH models in forecasting volatility at short horizons.

3.6.5. Default probabilities and credit spread

Recently, Reinhart and Rogoff (2008) studied the history of financial crises dating from England's fourteenth-century default to the current United States sub-prime financial crisis. They found that major default episodes are typically spaced some years (or decades) apart, creating an illusion that "this time is different" among policymakers and investors. Although the Markov-switching model was not mentioned, the authors qualitatively articulated two distinctively differing episodes, i.e., an episode with high and low serial default.

A number of researchers have been using the hidden Markov concept to model the default rate time series. Giampieri, Davis and Crowder (2005) modeled the occurrence of defaults within a bond portfolio as a hidden Markov process. The hidden variable represented the risk state, which was assumed to be common to all bonds within one particular sector and region. After describing the model and recalling the basic properties of hidden Markov chains, it was shown how to apply the model to a simulated sequence of default events. They also addressed the issue of global versus industry-specific risk factors. By extending their model to include independent hidden risk sequences, they were able to disentangle the risk associated with the business cycle from that specific to the individual sector.

Banachewicz, Lucas and Vaart (2007) extended this Hidden Markov model for defaults to include covariates. Using empirical U.S. default data, it was found that GDP growth, the term structure of interest rates and stock market returns, impacted the state transition probabilities. The impact, however, was not uniform across industries.

The expected default rate plays a pivotal role not only in the price of credit, but also in the

insurance against the credit, for example, the credit default swap (CDS) spread. Alexander and Kaeck (2008) investigated the determinants of the iTraxx CDS Europe indices and observed strong evidence of regime dependence. It was found that during volatile periods credit spreads become highly sensitive to stock volatility whereas during tranquil periods they are more sensitive to stock return.

3.6.6. Commodity Prices

Fong and Kim (2002) examined the temporal behavior of the volatility of daily returns on crude oil futures using a Markov switching model that allowed for abrupt changes in mean and variance, GARCH dynamics, basis-driven time-varying transition probabilities and conditional leptokurtosis. They showed that regime shifts were clearly present in the data and dominated GARCH effects. It was seen that within the high volatility state, a negative basis was more likely to increase regime persistence than a positive basis, which is consistent with previous empirical research on the theory of storage. The volatility regimes identified by their model correlated well with major events affecting supply and demand for oil. It was concluded that regime switching models provide a useful framework for the financial historian interested in studying factors behind the evolution of volatility and to oil futures traders interested short-term volatility forecasts.

Alizadeh, Nomikos and Pouliasis (2008) used a Markov regime switching vector error correction model to estimate constant and dynamic hedge ratios in the New York Mercantile Exchange oil futures markets and examine their hedging performance. They linked the concept of disequilibrium to that of uncertainty across high and low volatility

regimes. Overall, in and out-of-sample tests indicated that state dependent hedge ratios were able to provide a significant reduction of portfolio risk.

4. MODEL ESTIMATION AND FORECAST

This chapter firstly discusses estimation methods for the proposed HMS-V model, including comparisons with the reviewed MS-GARCH model. The application of the proposed model to forecasting time series is then presented.

The complete proof of the consistency of the Maximum Likelihood Estimator (MLE) is beyond this thesis. A theoretical consistency study on the exact MLE is a difficult problem and has been investigated in the following cases. For stationary Markov chains, Leroux (1992) proved the consistency of the MLE when the unobserved Markov chain had a finite state space, with the asymptotic normality being proven by Bickel, Ritov and Ryden (1998). In the case of the Markov chain being time-varying, with both the transition kernel of the hidden chain and the conditional distribution of the observations depending on a parameter vector θ, Douc and Matias (2001) gave the identifiability condition as well as proving the consistency and asymptotic normality of the MLE. Their proof was shown to follow from the exponential memorylessness properties of the state prediction filter and geometric ergodicity of suitably extended Markov chains. LeGland and Mevel (2000) proved the consistency and asymptotic normality of the MLE for hidden Markov models with a finite hidden state space. This work was based on the observation that the likelihood could be expressed as an additive function of an extended Markov chain. The key to the proof resided in the fact that, under appropriate conditions, the extended Markov-chain was shown to be geometrically ergodic.

4.1. The EM algorithm

4.1.1. Overview

Despite the complexity of the log likelihood function of equation (2.17), tractable solutions can be obtained using the well-known EM algorithm (Dempster, Laird, and Rubin, 1977). Wang and Puterman (2001) provided detailed steps in applying the EM algorithm to estimate a hidden Markov-switching Poisson regression model with time-varying transition probabilities and fixed covariate. For the proposed model, the EM algorithm is applied by treating the hidden binary random state variables Δ_t as the missing information and representing a complete data set for the model. The complete data is,

$$\text{DATA}_{com} = \left(\mathbf{Y}_\mathbf{T}, \mathbf{X}_\mathbf{T}, \boldsymbol{s}_\mathbf{T}, \boldsymbol{\mathcal{s}}_\mathbf{T}\right),$$

and the hidden states are the missing data,

$$\text{DATA}_{miss} = \left(\boldsymbol{s}_\mathbf{T}, \boldsymbol{\mathcal{s}}_\mathbf{T}\right),$$

where,

$$\begin{aligned} \boldsymbol{s}_\mathbf{T} &= \left\{ s_{\langle i \rangle, t};\ i = 1,2; t = 1, \cdots, T \right\}, \\ \boldsymbol{\mathcal{s}}_\mathbf{T} &= \left\{ \mathcal{s}_{\langle ij \rangle, t};\ i, j = 1,2; t = 2, \cdots, T \right\}, \end{aligned} \tag{4.1}$$

$$s_{\langle i \rangle, t} = \begin{cases} 1 & \text{if} \quad \Delta_t = i; \\ 0 & \text{otherwise;} \end{cases} \tag{4.2}$$

$$\mathcal{s}_{\langle ij \rangle, t} = \begin{cases} 1 & \textit{if} \quad \Delta_{t-1} = i \quad \text{and} \quad \Delta_t = j; \\ 0 & \text{otherwise.} \end{cases} \tag{4.3}$$

For convenience of matrix operation notation, we also define

$$\mathbf{S}_t' = \left[s_{\langle 1 \rangle, t}, s_{\langle 2 \rangle, t} \right]. \tag{4.4}$$

Each $s_{\langle i\rangle,t}$ is the realization of a Bernoulli trial at time t, whose rate of success is,

$$\begin{aligned} &E\left(s_{\langle i\rangle,t}\right)=\widetilde{s_{\langle i\rangle,t}}=\Pr\left(\Delta_t=i\right);\quad \widetilde{s_{\langle 1\rangle,t}}+\widetilde{s_{\langle 2\rangle,t}}=1;\\ &\widetilde{\mathbf{S}}_t'=\left[\widetilde{s_{\langle 1\rangle,t}},\widetilde{s_{\langle 2\rangle,t}}\right];\qquad i=1,2. \end{aligned} \tag{4.5}$$

Equipped with the information of the hidden state, the log-likelihood function LLF_{com} of the complete data can be written as,

$$\begin{aligned} LLF_{com}(\boldsymbol{\theta}) &= \log\Pr\left(\mathbf{DATA}_{com}\,|\,\boldsymbol{\theta}\right)\\ &= \log\Pr\left(\mathbf{Y}_{\mathbf{T}},\mathbf{X}_{\mathbf{T}},\boldsymbol{s}_{\mathbf{T}},\boldsymbol{\mathcal{s}}_{\mathbf{T}}\,|\,\boldsymbol{\theta}\right)\\ &= \log\Pr\left(\Delta_1=i\right)\\ &\quad+\sum_{t=2}^{T}\left[\boldsymbol{\mathcal{s}}_{\langle 11\rangle,t}\bullet\log p_{\langle 11\rangle,t}+\boldsymbol{\mathcal{s}}_{\langle 12\rangle,t}\bullet\log p_{\langle 12\rangle,t}\right]\\ &\quad+\sum_{t=2}^{T}\left[\boldsymbol{\mathcal{s}}_{\langle 22\rangle,t}\bullet\log p_{\langle 22\rangle,t}+\boldsymbol{\mathcal{s}}_{\langle 21\rangle,t}\bullet\log p_{\langle 21\rangle,t}\right]\\ &\quad+\sum_{t=1}^{T}s_{\langle 1\rangle,t}\bullet\log f_{\langle 1\rangle,t}+\sum_{t=1}^{T}s_{\langle 2\rangle,t}\bullet\log f_{\langle 2\rangle,t}\\ &= \log\Pr(\Delta_1=i)\\ &\quad+\sum_{t=2}^{T}\left[\boldsymbol{\mathcal{s}}_{\langle 11\rangle,t}\bullet\log p_{\langle 11\rangle}\left(\boldsymbol{\varphi}_{\langle 1\rangle},\psi_t\right)+\boldsymbol{\mathcal{s}}_{\langle 12\rangle,t}\bullet\log\left(1-p_{\langle 11\rangle}\left(\boldsymbol{\varphi}_{\langle 1\rangle},\psi_t\right)\right)\right]\\ &\quad+\sum_{t=2}^{T}\left[\boldsymbol{\mathcal{s}}_{\langle 22\rangle,t}\bullet\log p_{\langle 22\rangle}\left(\boldsymbol{\varphi}_{\langle 2\rangle},\psi_t\right)+\boldsymbol{\mathcal{s}}_{\langle 21\rangle,t}\bullet\log\left(1-p_{\langle 22\rangle}\left(\boldsymbol{\varphi}_{\langle 2\rangle},\psi_t\right)\right)\right]\\ &\quad+\sum_{t=1}^{T}s_{\langle 1\rangle,t}\bullet\log f_{\langle 1\rangle}\left(y_t\,|\,\boldsymbol{\vartheta}_{\langle 1\rangle},\psi_t\right)\\ &\quad+\sum_{t=1}^{T}s_{\langle 2\rangle,t}\bullet\log f_{\langle 2\rangle}\left(y_t\,|\,\boldsymbol{\vartheta}_{\langle 2\rangle},\psi_t\right)\\ &= \log\Pr(\Delta_1=i)+LLF_1\left(\boldsymbol{\varphi}_{\langle 1\rangle}\right)+LLF_2\left(\boldsymbol{\varphi}_{\langle 2\rangle}\right)+LLF_3\left(\boldsymbol{\vartheta}_{\langle 1\rangle}\right)+LLF_4\left(\boldsymbol{\vartheta}_{\langle 2\rangle}\right), \end{aligned} \tag{4.6}$$

where,

$$\begin{aligned} LLF_1\left(\boldsymbol{\varphi}_{\langle 1\rangle}\right)=\sum_{t=2}^{T}\Big[&\boldsymbol{\mathcal{s}}_{\langle 11\rangle,t}\bullet\log p_{\langle 11\rangle}\left(\boldsymbol{\varphi}_{\langle 1\rangle},\psi_t\right)\\ &+\boldsymbol{\mathcal{s}}_{\langle 12\rangle,t}\bullet\log\left(1-\log p_{\langle 11\rangle}\left(\boldsymbol{\varphi}_{\langle 1\rangle},\psi_t\right)\right)\Big], \end{aligned} \tag{4.7}$$

$$LLF_2\left(\boldsymbol{\varphi}_{\langle 2\rangle}\right)=\sum_{t=2}^{T}\Big[\boldsymbol{\varepsilon}_{\langle 22\rangle,t}\bullet\log p_{\langle 22\rangle}\left(\boldsymbol{\varphi}_{\langle 2\rangle},\ \psi_t\right) + \boldsymbol{\varepsilon}_{\langle 21\rangle,t}\bullet\log\left(1-\log p_{\langle 22\rangle}\left(\boldsymbol{\varphi}_{\langle 2\rangle},\ \psi_t\right)\right)\Big], \tag{4.8}$$

$$LLF_3\left(\boldsymbol{\vartheta}_{\langle 1\rangle}\right)=\sum_{t=1}^{T}s_{\langle 1\rangle,t}\bullet\log f_{\langle 1\rangle}\left(y_t\mid\boldsymbol{\vartheta}_{\langle 1\rangle},\psi_t\right),\ \text{and} \tag{4.9}$$

$$LLF_4\left(\boldsymbol{\vartheta}_{\langle 2\rangle}\right)=\sum_{t=1}^{T}s_{\langle 2\rangle,t}\bullet\log f_{\langle 2\rangle}\left(y_t\mid\boldsymbol{\vartheta}_{\langle 2\rangle},\psi_t\right). \tag{4.10}$$

The EM algorithm iteratively finds the MLE through two steps: the E-step (expectation) and the M-step (maximization). If $\boldsymbol{\theta}^{(l)}$ and $\boldsymbol{\theta}^{(l+1)}$ are the parameter estimates at the l^{th} and $(l+1)^{th}$ step respectively, then the function can be recursively written, $\boldsymbol{\theta}^{(l+1)}=\mathcal{M}\left(\boldsymbol{\theta}^{(l)}\right)$, where $\mathcal{M}$ is the mapping function consisting of an E step and an M step (see section 4.1.2 to section 4.1.3 for further details).

4.1.2. The E-Step

Given the observable data $(\mathbf{Y}_\mathbf{T},\mathbf{X}_\mathbf{T})$ and the estimated parameters at the l^{th} iteration, $\boldsymbol{\theta}^{(l)}$, the value of $\left\{p^{(l)}_{\langle ij\rangle,t};\quad i,j=1,2;\,t=2,\cdots,T\right\}$ and $\left\{\left(\mu^{(l)}_{\langle i\rangle,t}\ ,\sigma^{(l)}_{\langle i\rangle,t}\right);\,i=1,2;\ t=1,\cdots,T\right\}$ can be obtained. Re-obtaining these values again at the beginning of the $(l+1)^{th}$ iteration may sound redundant, as they are readily available at the end of the M-step at the l^{th} iteration, however, it is very important for the later introduced Supplemented EM (SEM) algorithm.

The essential operation of the E-step is to replace the missing data $\boldsymbol{s}_\mathbf{T}$ and $\boldsymbol{\varepsilon}_\mathbf{T}$ by the

conditional expectations $\hat{\boldsymbol{s}}_{\mathbf{T}}^{(l+1)}$ and $\hat{\boldsymbol{\mathcal{S}}}_{\mathbf{T}}^{(l+1)}$ respectively where,

$$\hat{s}_{\langle i\rangle,t}^{(l+1)} = E\left\{s_{\langle i\rangle,t} \mid \mathbf{Y}_{\mathbf{T}}, \mathbf{X}_{\mathbf{T}}, \boldsymbol{\theta}^{(l)}\right\} \tag{4.11}$$

$$\hat{\mathcal{s}}_{\langle ij\rangle,t}^{(l+1)} = E\left\{\mathcal{s}_{\langle ij\rangle,t} \mid \mathbf{Y}_{\mathbf{T}}, \mathbf{X}_{\mathbf{T}}, \boldsymbol{\theta}^{(l)}\right\}. \tag{4.12}$$

Therefore, the log-likelihood function of the complete-data can be found at the l^{th} step by treating $\boldsymbol{\theta}^{(l)}$ as $\boldsymbol{\theta}$,

$$LLF_{com}\left(\boldsymbol{\theta}\middle|\boldsymbol{\theta}^{(l)}\right) = \int f\left(\mathbf{Y}_{\mathbf{T}}\middle|\mathbf{X}_{\mathbf{T}}, \boldsymbol{\theta}\right) \bullet f\left(\mathbf{DATA}_{miss}\middle|\mathbf{Y}_{\mathbf{T}}, \mathbf{X}_{\mathbf{T}}, \boldsymbol{\theta} = \boldsymbol{\theta}^{(l)}\right) d\,\mathbf{DATA}_{miss}\,.$$

The estimation of the missing information of the proposed model relies on the following backward and forward recursive algorithm.

Forward and backward algorithm

Given the link functional form, the covariates and the parameters, the forward and backward algorithm is used to estimate the state probabilities. The forward and backward algorithm was firstly proposed by Baum, Petrie, Soules and Weiss (1970) resulting in impressive simplification the computation. Furthermore, the forward recursion also provides the value of the log likelihood function of the observed data defined previously by equation (2.17).

The forward algorithm

The forward probability is defined as,

$$A_{\langle i\rangle,t} = f\left(y_1, \cdots, y_t, \Delta_t = i \mid \mathbf{X}_{\mathbf{T}}, \boldsymbol{\theta}^{(l)}\right); \quad (i = 1,2; t = 1, \cdots T). \tag{4.13}$$

Initialization:

$$A_{\langle i\rangle,1} = f\left(\Delta_1 = i\right)\bullet f_{\langle i\rangle,1};\ \left(i = 1,2\right). \tag{4.14}$$

Induction:

$$A_{\langle i\rangle,t+1} = \left[\sum_{j=1}^{2} A_{\langle j\rangle,t}\bullet p^{(l)}_{\langle ji\rangle,t+1}\right]\bullet f_{\langle i\rangle,t+1};\quad \left(i,j = 1,2;\ t = 1,\cdots,T-1\right). \tag{4.15}$$

Termination:

$$\Pr\left(\mathbf{Y_T} \mid \mathbf{X_T}, \boldsymbol{\theta}^{(l)}\right) = \sum_{i=1}^{2} f\left(\mathbf{Y_T}, \Delta_T = i \mid \mathbf{X_T}, \boldsymbol{\theta}^{(l)}\right) = \sum_{i=1}^{2} A_{\langle i\rangle,T}\,, \tag{4.16}$$

and,

$$LLF_{obs}\left(\boldsymbol{\theta}^{(l)}\right) = \log f\left(\mathbf{Y_T} \mid \mathbf{X_T}, \boldsymbol{\theta}^{(l)}\right), \tag{4.17}$$

where $LLF_{obs}\left(\boldsymbol{\theta}^{(l)}\right)$ is the log likelihood function of the observed data at the l^{th} iteration.

The backward algorithm

The backward probability is defined as

$$B_{\langle i\rangle,t} = f\left(y_{t+1},\ldots,y_T \mid \Delta_t = i\right);\ \left(i = 1,2; t = 1,\cdots T\right). \tag{4.18}$$

Initialization:

$$B_{\langle i\rangle,T} = 1; \qquad \left(i = 1,2\right). \tag{4.19}$$

Induction:

$$B_{\langle i\rangle,t} = \sum_{j=1}^{2} p^{(l)}_{\langle ij\rangle,t} \bullet B_{\langle j\rangle,t+1} \bullet f_{\langle j\rangle,t+1}; \quad (i,j = 1,2; t = 1,\cdots,T-1). \tag{4.20}$$

Using the forward and backward probabilities, $A_{\langle i\rangle,t}$ and $B_{\langle i\rangle,t}$, the Markov state probabilities can be estimated according to,

$$\begin{aligned} \hat{s}^{(l+1)}_{\langle i\rangle,t} &= f\left(\Delta_t = i \middle| \mathbf{Y_T}, \mathbf{X_T}, \boldsymbol{\theta}^{(l)}\right) \\ &= \frac{f\left(\mathbf{Y_T}, \Delta_t = i \middle| \mathbf{X_T}, \boldsymbol{\theta}^{(l)}\right)}{f\left(\mathbf{Y_T} \mid \mathbf{X_T}, \boldsymbol{\theta}^{(l)}\right)} \\ &= \frac{f\left(y_1,\ldots,y_t, \Delta_t = i \middle| \mathbf{X_T}, \boldsymbol{\theta}^{(l)}\right) \bullet f\left(y_{t+1},\ldots,y_T \middle| \Delta_t = i, \mathbf{X_T}, \boldsymbol{\theta}^{(l)}\right)}{f\left(\mathbf{Y_T} \middle| \mathbf{X_T}, \boldsymbol{\theta}^{(l)}\right)} \\ &= \frac{A_{\langle i\rangle,t} \bullet B_{\langle i\rangle,t}}{A_{\langle 1\rangle,T} + A_{\langle 2\rangle,T}}. \end{aligned} \tag{4.21}$$

and,

$$\begin{aligned} \hat{s}^{(l+1)}_{\langle ij\rangle,t} &= f\left(\Delta_{t-1} = i, \Delta_t = j \middle| \mathbf{Y_T}, \mathbf{X_T}, \boldsymbol{\theta}^{(l)}\right) \\ &= \frac{f(\mathbf{Y_T}, \Delta_{t-1} = i, \Delta_t = j)}{f(\mathbf{Y_T})} \\ &= \frac{A_{\langle i\rangle,t-1} \bullet p^{(l)}_{\langle ij\rangle,t} \bullet f_{\langle j\rangle,t} \bullet B_{\langle j\rangle,t}}{A_{\langle 1\rangle,T} + A_{\langle 2\rangle,T}} \end{aligned} \tag{4.22}$$

Further details are outlined in Appendix I.

Numerical instability

Both the forward and backward routines are susceptible to either overflow or underflow as the sample size grows. Leroux and Puterman (1992) documented the underflow and recommended a rescaling solution. In the work of this these, overflow of the forward-backward probabilities for financial time series data was also observed. The possibility of overflow and appropriate numerical solutions are discussed in Appendix III.

4.1.3. The parallel M-Step

Given the observed $(\mathbf{Y}_\mathbf{T}, \mathbf{X}_\mathbf{T})$ and missing $\left(\hat{\boldsymbol{s}}_\mathbf{T}^{(l+1)}, \hat{\boldsymbol{s}}_\mathbf{T}^{(l+1)}\right)$ data, parameter estimates can be readily obtained by maximizing each one of the following four functions independently,

$$LLF_1^{(l+1)}\left(\boldsymbol{\varphi}_{\langle 1\rangle}\right) = \sum_{t=2}^{T}\left[\hat{s}_{\langle 11\rangle,t}^{(l+1)} \bullet \log p_{\langle 11\rangle}\left(\boldsymbol{\varphi}_{\langle 1\rangle}, \psi_t\right) + \hat{s}_{\langle 12\rangle,t}^{(l+1)} \bullet \log\left(1 - p_{\langle 11\rangle}\left(\boldsymbol{\varphi}_{\langle 1\rangle}, \psi_t\right)\right)\right], \tag{4.23}$$

$$LLF_2^{(l+1)}\left(\boldsymbol{\varphi}_{\langle 2\rangle}\right) = \sum_{t=2}^{T}\left[\hat{s}_{\langle 22\rangle,t}^{(l+1)} \bullet \log p_{\langle 22\rangle}\left(\boldsymbol{\varphi}_{\langle 2\rangle}, \psi_t\right) + \hat{s}_{\langle 21\rangle,t}^{(l+1)} \bullet \log\left(1 - p_{\langle 22\rangle}\left(\boldsymbol{\varphi}_{\langle 2\rangle}, \psi_t\right)\right)\right], \tag{4.24}$$

$$LLF_3^{(l+1)}\left(\boldsymbol{\vartheta}_{\langle 1\rangle}\right) = \sum_{t=1}^{T} \hat{s}_{\langle 1\rangle,t}^{(l+1)} \bullet \log f_{\langle 1\rangle}\left(y_t \mid \boldsymbol{\vartheta}_{\langle 1\rangle}, \psi_t\right), \tag{4.25}$$

and,

$$LLF_4^{(l+1)}\left(\boldsymbol{\vartheta}_{\langle 2\rangle}\right) = \sum_{t=1}^{T} \hat{s}_{\langle 2\rangle,t}^{(l+1)} \bullet \log f_{\langle 2\rangle}\left(y_t \mid \boldsymbol{\vartheta}_{\langle 2\rangle}, \psi_t\right). \tag{4.26}$$

For clarity of exposition, let the above four sub-M steps be referred to as sub-M1, sub-M2, sub-M3, and sub-M4 respectively. The implementation of sub-M1 and sub-M3 are selected as examples and outlined here. The implementation of the remaining sub-M steps follows a similar process.

To maximize $LLF_1^{(l+1)}\left(\boldsymbol{\varphi}_{\langle 1\rangle}\right)$ with respect to $\boldsymbol{\varphi}_{\langle 1\rangle}$, the estimated value $\boldsymbol{\varphi}_{\langle 1\rangle}^{(l+1)}$ should satisfy the following equation,

$$\left.\frac{\partial LLF_1^{(l+1)}\left(\boldsymbol{\varphi}_{\langle 1\rangle}\right)}{\partial \boldsymbol{\varphi}_{\langle 1\rangle}}\right|_{\boldsymbol{\varphi}_{\langle 1\rangle}^{(l+1)}} = 0. \tag{4.27}$$

If the state transition probability function $p_{\langle 11\rangle}\left(\boldsymbol{\varphi}_{\langle 1\rangle}, \psi_t\right)$ is a constant, equation (4.27) has the following closed form solution,

$$\boldsymbol{\varphi}_{\langle 1\rangle,0}^{(l+1)} = LOGIT^{-1}\left(\frac{\sum_{t=2}^{T} \hat{\boldsymbol{\varepsilon}}_{\langle 11\rangle,t}^{(l+1)}}{\sum_{t=2}^{T} \hat{\boldsymbol{s}}_{\langle 1\rangle,t}^{(l+1)}}\right) \tag{4.28}$$

where $\boldsymbol{\varphi}_{\langle 1\rangle,0}^{(l+1)}$ is the first element of $\boldsymbol{\varphi}_{\langle 1\rangle}^{(l+1)}$.

In the case of a non-constant state transition probability, the solution can be obtained using the quasi-Newton algorithm in the MatLab 7.0 Optimization Toolbox. The closed form gradient and Hessian of (4.27) are provided in Appendix IV and Appendix V.

Similarly, to maximize $LLF_3^{(l+1)}\left(\boldsymbol{\vartheta}_{\langle 1\rangle}\right)$ with respect to $\boldsymbol{\vartheta}_{\langle 1\rangle}$, the estimated value $\boldsymbol{\vartheta}_{\langle 1\rangle}^{(l+1)}$ should satisfy the following equation,

$$\left.\frac{\partial LLF_3^{(l+1)}\left(\boldsymbol{\vartheta}_{\langle 1\rangle}\right)}{\partial \boldsymbol{\vartheta}_{\langle 1\rangle}}\right|_{\boldsymbol{\vartheta}_{\langle 1\rangle}^{(l+1)}} = 0. \tag{4.29}$$

If volatility at the first state is a constant, equation (4.29) has a closed form solution through the general least square (GLS) formula,

$$\boldsymbol{\alpha}_{\langle 1\rangle}^{(l+1)} = \left(\mathbf{C}_{\mathbf{T}}'\boldsymbol{\Omega}_{\langle 1\rangle,\mathbf{T}}\mathbf{C}_{\mathbf{T}}\right)^{-1}\mathbf{C}_{\mathbf{T}}\mathbf{Y}_{\mathbf{T}}, \tag{4.30}$$

$$\boldsymbol{\beta}_{\langle 1\rangle,0}^{(l+1)} = \frac{1}{2}\log\left\{\left(\mathbf{Y}_{\mathbf{T}} - \mathbf{C}_{\mathbf{T}}\boldsymbol{\alpha}_{\langle 1\rangle}^{(l+1)}\right)' \boldsymbol{\Omega}_{\langle 1\rangle,\mathbf{T}}\left(\mathbf{Y}_{\mathbf{T}} - \mathbf{C}_{\mathbf{T}}\boldsymbol{\alpha}_{\langle 1\rangle}^{(l+1)}\right) \Big/ \boldsymbol{\upsilon}_{\langle 1\rangle}\right\}, \tag{4.31}$$

$$\vartheta_{\langle 1\rangle}^{(l+1)} = \left(\alpha_{\langle 1\rangle}^{(l+1)}, \beta_{\langle 1\rangle}^{(l+1)}\right)',$$

where $\Omega_{\langle 1\rangle,\mathbf{T}} = diag\left(\hat{s}_{\langle 1\rangle,1}^{(l+1)}, \cdots, \hat{s}_{\langle 1\rangle,T}^{(l+1)}\right)$, $\upsilon_{\langle 1\rangle} = tr\left(\Omega_{\langle 1\rangle,\mathbf{T}}\right)$ and $\beta_{\langle 1\rangle,0}^{(l+1)}$ is the first element (intercept term) of $\beta_{\langle 1\rangle}^{(l+1)}$. Again, for a non-constant volatility, the quasi-Newton algorithm can be used.

Let $\mathbf{I}_1$, $\mathbf{I}_2$, $\mathbf{I}_3$ and $\mathbf{I}_4$ be defined as the independent information matrices of the four sub-M-steps respectively. Therefore, the information matrix of the complete data has the following block diagonal form,

$$\mathbf{I}_{com}^{(l+1)} = \begin{bmatrix} \mathbf{I}_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_4 \end{bmatrix}, \tag{4.32}$$

where the bold $\mathbf{0}$ stands for a matrix whose elements all equal to 0. Given the closed form Hessian of all the sub-M steps, it is straight forward to obtain the information matrix of the complete data. Therefore, when applying the EM algorithm to the proposed model, instead of dealing with all the elements of $\mathbf{I}_{com}^{(l+1)}$ simultaneously, each of the four diagonal blocks $\mathbf{I}_1$, $\mathbf{I}_2$, $\mathbf{I}_3$ and $\mathbf{I}_4$ are dealt with independently. Such a novel decomposition of the M-step further reduces the tractability and enhances the computational stability.

It is worth noting that the parallel structure of the proposed model enables a parallel solution for each sub-M step. Although both the proposed and MS-GARCH models can be estimated through the EM algorithm, the outlined simplification of the M-step plays a vital role in the superiority of the proposed algorithm. In practice, the M-step often

becomes a computational bottle neck, whose complexity is determined by the following three factors: (1) the availability of closed form solutions of the sub-M-steps; (2) the dimensions and sparse structure of the complete data information matrix; and (3) the availability of the closed form gradient and Hessian for a numerical algorithm.

Because of the parallel structure, the proposed model possess the following three merits: (1) a increased likelihood of having a closed form solution for a sub-M-step; (2) the diagonal structure of the information matrix of the complete data is easier to handle and each block has a lower dimensions; (3) each sub-M-step can be solved independently (and thus in parallel); and (4) both the closed form gradient and Hessian of the M-step are trivial and available. Previous MS-[G]ARCH models, generally do not display such simplicity and flexibility.

If the two GARCH processes are not parallel[32], the information matrix of the completed data is far more complex, and the two sets of GARCH coefficients have to be jointly, as opposed to independently, estimated. With such a complex M-step, the practical applicability of the EM algorithm is questionable.

4.1.4. Algorithm Summary

The following summarizes the implementation of the EM algorithm for the proposed HMS-V model.

[32] The two GARCH processes of Gray/Klaassen are naturally intertwined (non-parallel). The two GARCH link functions at the two Markov states of Haas, Mittnik and Paolella (2004) are also intertwined through the shared mean.

INPUT: a set of intuitive[33] starting values $\boldsymbol{\theta}^{(0)}$ and a tolerance level for the EM iteration, $\in_{EM}$.

Step 1.(E-step) At the $(l+1)^{th}$ iteration, compute $\hat{s}_{\langle i\rangle,t}^{(l+1)}$ and $\hat{s}_{\langle ij\rangle,t}^{(l+1)}$ for $t=2,\cdots,T$ and $i,j=1,2$; where $LLF_{obs}^{(l)}$ (the observed log likelihood at the l^{th} step) is a by-product (Appendix II);

Step 2.(M-step) Find the value of $\boldsymbol{\theta}^{(l+1)}$ by the four parallel sub-M-steps, where the closed form solution should always be used if possible; otherwise, use the quasi-Newton algorithm;

Step 3. If $\left\|\boldsymbol{\theta}^{(l+1)}-\boldsymbol{\theta}^{(l)}\right\| \geq \in_{EM}$, go to step 1, and start the $(l+2)^{th}$ iteration; Otherwise, terminate.

OUTPUT: $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{s}}_{\mathbf{T}}$, $\mathbf{I}_{com}\left(\hat{\boldsymbol{\theta}}\right)$ and the convergence path $\left(\boldsymbol{\theta}^{(0)},\boldsymbol{\theta}^{(1)},\cdots,\hat{\boldsymbol{\theta}}\right)$.

Despite its elegancy and numerical stability, the EM algorithm has a few limitations.

Limitation 1: Like most other iterative numerical methods, the EM algorithm converges to a local maximum, which may or may not be the global optimum. For the proposed model, $LLF_{com}^{(l+1)}(\boldsymbol{\theta}^{(l+1)}\,|\,\boldsymbol{\theta}^{(l)})$ and its first order partial derivatives are continuous both in $\boldsymbol{\theta}^{(l+1)}$ and $\boldsymbol{\theta}^{(l)}$. Applying Wu's general result (1983), the sequence of the observed log likelihood function (4.17) converges to a local maximum or saddle point regardless of the starting values. Note that (4.17) needs not be globally concave. Therefore, one needs to carefully choose sets of starting values to increase the chance of converging to the global maximum.

[33] It can be an educated guess or a random guess.

Limitation 2: The EM algorithm does not quantify the MLE error, such as for other Bayesian estimation techniques.

Limitation 3: The EM algorithm always converges linearly, therefore compared with the Newton type method, convergence is much slower in the vicinity of a local maximum. This drawback is further compounded when the M-step has no closed form solution. If speed of convergence is critical requirement, the quasi-Newton algorithm is a better alternative. Section 4.2 discusses the implementation of the quasi-Newton algorithm to optimize the log likelihood function of the observed data.

4.2. The quasi-Newton algorithm

The likelihood function of the observed data, written in the form of equation (2.17), is highly complex. Fortunately, the forward algorithm provides a simplified and tractable method of reformulating the log likelihood function, which is also used as the target function of the quasi-Newton algorithm. In addition, the same rescaling treatment developed for the E-step must be adopted to prevent underflow or overflow.

To maximize $LLF_{obs}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, the estimate $\hat{\boldsymbol{\theta}}$ must satisfy

$$\left.\frac{\partial \widehat{LLF_{obs}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}} = 0 \qquad (4.33)$$

for which neither a closed form gradient nor Hessian is available and should be obtained numerically. The quasi-Newton algorithm uses the observed behavior of $LLF_{obs}(\boldsymbol{\theta})$ and

$\nabla LLF_{obs}(\boldsymbol{\theta})$ to establish curvature information to formulate a new quadratic problem. In this work, the routine of MatLab 7.0 Optimized Toolbox is followed, and the popular BFGS symmetric rank 2 update is used to approximate the Hessian.

The quasi-Newton algorithm converges superlinearly, while the EM algorithm converges linearly. Therefore, if the initial estimates are in the vicinity of the global maximum, the quasi-Newton often converges significantly faster than the EM algorithm. However, quasi-Newton also possesses complications which need to be carefully considered.

Limitation 1 and the solution: The quasi-Newton algorithm is far more sensitive to the initial conditions than the EM algorithm, whereas the later is more robust to poorly chosen the initialisation parameters. There are two solutions to mitigate this "sensitivity". The first is to increase the number of sets of starting values. For example, Cheung and Erlandsson (2005) used up to 250 sets of starting values, which were randomly sampled from a uniform distribution. However, as the dimension of the parameter vector grows, increasing the number of starting values can lead to tractability problems. The other solution is to exploit an *a priori* data analysis to provide more intuitive initial values. This is further addressed in the second solution of section 4.3.

Limitation 2 and the solution: Due to the complexity of the target function, the recursive iteration often displays instability, which easily leads to underflow or overflow and potential crashing of the optimization process for various reasons. In practice, brute force methods are typically employed and estimation bounds are imposed. The technical details are overlooked here as this issue does not constitute a contribution of this thesis.

4.3. A hybrid approach

The EM algorithm is robust to poorly chosen starting values and commonly reports the local maximum after just a small number of iterations. On the other hand, in the vicinity of the local maximum, with some brute force methods, the quasi-Newton algorithm is often adequately stable and converges to the global maximum with greater speed. Various authors, including Redner and Walker (1984), proposed a hybrid approach to the computation of the MLE that would dually adopt the EM algorithm and a Newton-Raphson or some quasi-Newton algorithm.

In this work, a lenient convergence criterion is chosen for the EM algorithm, whose converged estimate may not be global optimum but instead provides the starting values (assumed to be well-chosen) to the quasi-Newton algorithm. The quasi-Newton algorithm then continues the estimation process and converges to the global (or a local) optimum. The EM algorithm hence is well suited as a preliminary data analysis tool for the quasi-Newton algorithm. Following the convergence of the quasi-Newton method, the EM algorithm is again applied to estimate the state probabilities $\hat{\boldsymbol{s}}_T$. *Figure 9* summarizes the process of the proposed hybrid approach.

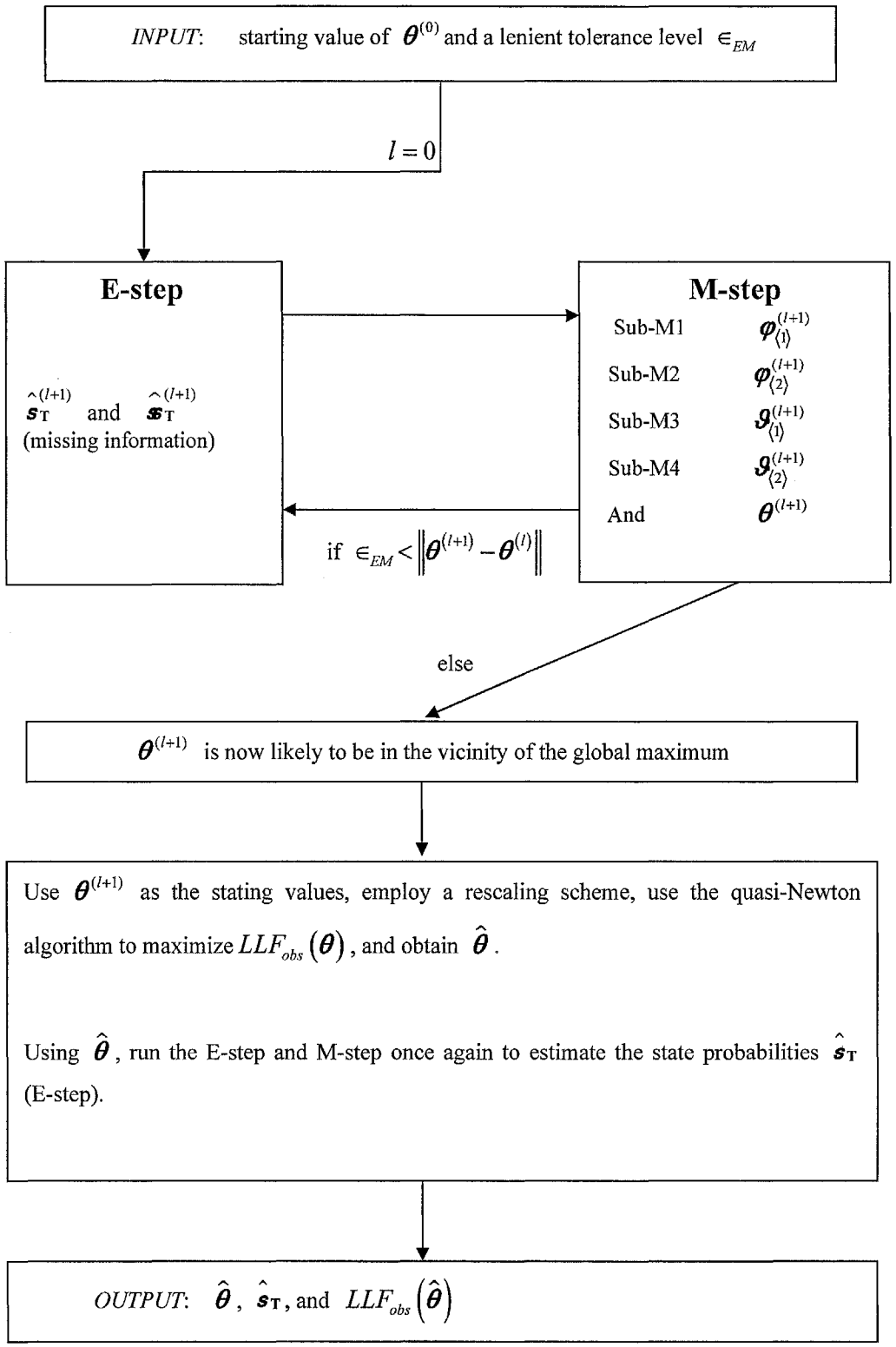


Figure 9

A summary of the hybrid EM and quasi-Newton approach.

4.4. Standard errors of the estimated parameters

One of the main drawbacks of the EM algorithm is that the standard errors of the parameters are not estimated. Unfortunately, the quasi-Newton algorithm, which can be used as the last step for the MLE, also skips the calculation of the "true" Hessian. Therefore, the estimation of the standard MLE errors must be addressed.

4.4.1. Information matrix by numerical difference

Section 4.3 suggests using the EM algorithm to provide suitable starting values, and the quasi-Newton algorithm to obtain the MLE. After optimization, the one-sided approximation of the Taylor series expansion can be adopted to calculate the "true" Hessian. Numerical approximation of the information matrix and the standard errors of the MLE can then be obtained accordingly. According to the Monte Carlo study of section 4.5 and empirical study in chapter 5 of the thesis, the majority of the estimated information matrices by the numerical difference method are positive definite.

Despite its convenience, the numerical difference method may be unstable, especially when the Hessian is not smooth. However, as a consequence of the recursive structure, the gradient and Hessian are both continuous and bounded. The following gives a very brief discussion.

Inserting the global maximizer $\boldsymbol{\theta}$ into the last step of the forward recursion,

$$f\left(\mathbf{Y_T} \mid \mathbf{X_T}, \boldsymbol{\theta}\right) = \sum_{i=1}^{2} f\left(\mathbf{Y_T}, \Delta_T = i \mid \mathbf{X_T}, \boldsymbol{\theta}\right) = \sum_{i=1}^{2} A_{\langle i \rangle, T} \,,$$

the observed log likelihood function of the whole sample becomes,

$$LLF_{obs}(\boldsymbol{\theta}) = \log f(\mathbf{Y_T} \mid \mathbf{X_T}, \boldsymbol{\theta}).$$

The gradient is then,

$$\begin{aligned} &\frac{\partial \log f(\mathbf{Y_T} \mid \mathbf{X_T}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \\ &= \frac{1}{f(\mathbf{Y_T} \mid \mathbf{X_T}, \boldsymbol{\theta})} \frac{\partial f(\mathbf{Y_T} \mid \mathbf{X_T}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}, \end{aligned} \tag{4.34}$$

where,

$$\begin{aligned} \frac{\partial f(\mathbf{Y_T} \mid \mathbf{X_T}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} &= \frac{\partial \sum_{i=1}^{2} A_{\langle i \rangle, T}}{\partial \boldsymbol{\theta}} \\ &= \frac{\partial \left(A_{\langle 1 \rangle, T-1} \bullet p_{\langle 11 \rangle, T} \bullet f_{\langle 1 \rangle, T} \right)}{\partial \boldsymbol{\theta}} + \frac{\partial \left(A_{\langle 2 \rangle, T-1} \bullet p_{\langle 21 \rangle, T} \bullet f_{\langle 1 \rangle, T} \right)}{\partial \boldsymbol{\theta}} \\ &\quad + \frac{\partial \left(A_{\langle 1 \rangle, T-1} \bullet p_{\langle 12 \rangle, T} \bullet f_{\langle 2 \rangle, T} \right)}{\partial \boldsymbol{\theta}} + \frac{\partial \left(A_{\langle 2 \rangle, T-1} \bullet p_{\langle 22 \rangle, T} \bullet f_{\langle 2 \rangle, T} \right)}{\partial \boldsymbol{\theta}}. \end{aligned} \tag{4.35}$$

Expanding the first element at the right hand side of (4.35),

$$\begin{aligned} &\frac{\partial \left(A_{\langle 1 \rangle, T-1} \bullet p_{\langle 11 \rangle, T} \bullet f_{\langle 1 \rangle, T} \right)}{\partial \boldsymbol{\theta}} \\ &= \frac{\partial A_{\langle 1 \rangle, T-1}}{\partial \boldsymbol{\theta}} \bullet p_{\langle 11 \rangle, T} \bullet f_{\langle 1 \rangle, T} + \frac{\partial p_{\langle 11 \rangle, T}}{\partial \boldsymbol{\theta}} \bullet A_{\langle 1 \rangle, T-1} \bullet f_{\langle 1 \rangle, T} + \frac{\partial f_{\langle 1 \rangle, T}}{\partial \boldsymbol{\theta}} \bullet A_{\langle 1 \rangle, T-1} \bullet p_{\langle 11 \rangle, T}. \end{aligned} \tag{4.36}$$

Therefore, $\frac{\partial A_{\langle i \rangle, T}}{\partial \boldsymbol{\theta}}$ is a function of $\frac{\partial A_{\langle i \rangle, T-1}}{\partial \boldsymbol{\theta}}$, $i = 1, 2$. Theoretically, $\frac{\partial A_{\langle 1 \rangle, T-1}}{\partial \boldsymbol{\theta}}$ can be further expanded and recursively traced all the way back to $\frac{\partial A_{\langle i \rangle, 1}}{\partial \boldsymbol{\theta}}$. If the number of observations is finite, the gradient implied by equation (4.35) is a polynomial with finite terms, whose components are all finite products of the elements of the following vectors,

$$\frac{\partial p_{\langle ij \rangle, t}}{\partial \boldsymbol{\theta}}, \quad \frac{\partial f_{\langle 1 \rangle, t}}{\partial \boldsymbol{\theta}}, \text{ for } i, j = 1, 2 \text{; and } t = 1. \cdots, T. \tag{4.37}$$

As shown in Appendix IV and Appendix V, all of the elements of $\frac{\partial p_{\langle ij \rangle, t}}{\partial \boldsymbol{\theta}}$ and $\frac{\partial f_{\langle 1 \rangle, t}}{\partial \boldsymbol{\theta}}$ are

continuous and bounded. In addition, $\frac{1}{f(\mathbf{Y}_T \mid \mathbf{X}_T, \boldsymbol{\theta})}$ is positive. Therefore, the gradient is continuous and bounded.

The Hessian matrix is

$$\begin{aligned}&\frac{\partial^2 \log f(\mathbf{Y}_T \mid \mathbf{X}_T, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}'} \\ &= -\left[\frac{1}{f(\mathbf{Y}_T \mid \mathbf{X}_T, \boldsymbol{\theta})}\right]^2 \frac{\partial f(\mathbf{Y}_T \mid \mathbf{X}_T, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial f(\mathbf{Y}_T \mid \mathbf{X}_T, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \\ &\quad + \frac{1}{f(\mathbf{Y}_T \mid \mathbf{X}_T, \boldsymbol{\theta})} \frac{\partial^2 f(\mathbf{Y}_T \mid \mathbf{X}_T, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}'}.\end{aligned} \tag{4.38}$$

Similarly, $\frac{\partial^2 A_{\langle i \rangle, T}}{\partial \boldsymbol{\theta}^2}$ is a function of $\frac{\partial^2 A_{\langle i \rangle, T-1}}{\partial \boldsymbol{\theta}^2}$, $i = 1, 2$. Recursively, the second order derivative $\frac{\partial^2 f(\mathbf{Y}_T \mid \mathbf{X}_T, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}\, \partial \boldsymbol{\theta}'}$ is also a polynomial with finite terms, whose components are finite products of the elements of the following matrices,

$$\frac{\partial^2 p_{\langle ij \rangle, t}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}, \quad \frac{\partial^2 f_{\langle 1 \rangle, t}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}, \text{ for } i, j = 1, 2 \text{; and } t = 1, \cdots, T \tag{4.39}$$

which are also all continuous and bounded. Because $\frac{1}{f(\mathbf{Y}_T \mid \mathbf{X}_T, \boldsymbol{\theta})}$ is positive, the Hessian matrix is also almost completely smooth and continuous. In section 4.5 and section 5.3, Monte Carlo experiments are carried out to investigate the reliability of the numerical difference.

4.4.2. The Supplemented EM algorithm

The supplemented EM (SEM) algorithm was proposed by Meng and Rubin (1991). It

provided a deep insight and peculiar numerical method to obtain the information matrix through a forced EM algorithm.

Section 4.1.2 to section 4.1.3 gave the details of a mapping process from $\boldsymbol{\theta}^{(l)}$ to $\boldsymbol{\theta}^{(l+1)}$,

$$\boldsymbol{\theta}^{(l+1)} = \mathcal{M}\left(\boldsymbol{\theta}^{(l)}\right)^{34}. \tag{4.40}$$

Suppose that $\boldsymbol{\theta}^{(l)}$ converges to some point $\hat{\boldsymbol{\theta}}$ and $\mathcal{M}(\boldsymbol{\theta})$ is continuous, then $\hat{\boldsymbol{\theta}}$ maps to itself,

$$\hat{\boldsymbol{\theta}} = \mathcal{M}\left(\hat{\boldsymbol{\theta}}\right). \tag{4.41}$$

Meng and Rubin noticed that in the neighborhood of $\hat{\boldsymbol{\theta}}$, a Taylor series expansion yields,

$$\boldsymbol{\theta}^{(l+1)} - \hat{\boldsymbol{\theta}} \approx \left(\boldsymbol{\theta}^{(l)} - \hat{\boldsymbol{\theta}}\right) \bullet DM \tag{4.42}$$

where,

$$DM = \left(\frac{\partial \mathcal{M}_j(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i} \right)\Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \tag{4.43}$$

is the $D_{\mathbf{X}} \times D_{\mathbf{X}}$ Jacobian matrix for $\mathcal{M}(\boldsymbol{\theta}) = \left(\mathcal{M}_1(\boldsymbol{\theta}), \cdots, \mathcal{M}_{D_{\mathbf{X}}}(\boldsymbol{\theta})\right)$ evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. According to the factorization,

$$f\left(\mathbf{Y}_{\mathbf{T}}, \mathrm{DATA}_{miss} \mid \mathbf{X}_{\mathbf{T}}, \theta\right) = f\left(\mathbf{Y}_{\mathbf{T}} \mid \mathbf{X}_{\mathbf{T}}, \theta\right) \bullet f\left(\mathrm{DATA}_{miss} \mid \mathbf{X}_{\mathbf{T}}, \mathbf{Y}_{\mathbf{T}}, \theta\right) \tag{4.44}$$

it follows that the information matrices satisfy,

$$\mathbf{I}_{com} = \mathbf{I}_{obs} + \mathbf{I}_{miss} \tag{4.45}$$

which implies that the information matrix of the complete data equals the summation of the information matrix of the observed data and the information matrix of the missing data.

---

[34] Refer to section 4.1 for the definition of the mapping function $\mathcal{M}$.

Smith (1977) developed the following simple relationship between $\mathbf{V}$, the asymptotic variance of the observed data, and $\mathbf{V}_{com}$, the asymptotic variance of the complete data:

$$V = V_{com}/(1-r) \tag{4.46}$$

where r is the rate of convergence of the EM algorithm. Meng and Rubin (1991) further observed that a more statistically appealing representation of equation (4.46) is:

$$V = V_{com} + \Delta V\,, \tag{4.47}$$

where,

$$\Delta V = \left[r/(1-r)\right] \bullet V_{com} \tag{4.48}$$

is the increase in variance due to missing data. Meng and Rubin's SEM suggests:

$$\mathbf{V}_{SEM} = \mathbf{I}_{com}^{-1} + \Delta\mathbf{V} \tag{4.49}$$

$$\Delta\mathbf{V}_{SEM} = \mathbf{I}_{com}^{-1} \bullet DM \bullet (\mathbf{I} - DM)^{-1} \tag{4.50}$$

or,

$$\mathbf{V}_{SEM} = \mathbf{I}_{com}^{-1} + \mathbf{I}_{com}^{-1} \bullet DM \bullet (\mathbf{I} - DM)^{-1} \tag{4.51}$$

where $\mathbf{I}_{com}$ is readily available from the M step of the previous EM iteration. The difficulty lies in calculating the DM matrix.

Computation of the DM matrix

SEM indicates that each element of DM is the component-wise rate of convergence of a "forced" EM iteration. Let r_{ij} be the $(i,j)^{th}$ element of the DM and define $\boldsymbol{\theta}_i^{(t)}$ to be,

$$\boldsymbol{\theta}_i^{(t)} = \left(\hat{\boldsymbol{\theta}}_1, \cdots, \hat{\boldsymbol{\theta}}_{i-1}, \boldsymbol{\theta}_i^{(t)}, \hat{\boldsymbol{\theta}}_{i+1}, \cdots, \hat{\boldsymbol{\theta}}_{D_X}\right) \tag{4.52}$$

where only the i^{th} component in $\boldsymbol{\theta}_i^{(t)}$ is replaced, while the other components are fixed

at their MLE's. From the definition of r_{ij},

$$\begin{aligned} r_{ij} &= \frac{\partial M_j\left(\hat{\boldsymbol{\theta}}\right)}{\partial \theta_i} \\ &= \lim_{\theta_i \to \hat{\theta}_i} \frac{M_j\left(\hat{\theta}_1, \cdots, \hat{\theta}_{i-1}, \theta_i, \hat{\theta}_{i+1}, \cdots, \hat{\theta}_{d_X}\right) - M_j\left(\hat{\boldsymbol{\theta}}\right)}{\theta_i - \hat{\theta}_i} \\ &= \lim_{t \to \infty} \frac{M_j\left(\theta_i^{(t)}\right) - \hat{\theta}_j}{\theta_i^{(t)} - \hat{\theta}_i} = \lim_{t \to \infty} r_{ij}^{(t)} \end{aligned} \tag{4.53}$$

SEM is required to store all the iteration history of the EM algorithm, which leaves the footprint of the element-wise convergence rate. The following gives the detailed implementation steps.

INPUT: $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^{(t)}$

Step 1. Run the usual E- and M-steps to obtain $\boldsymbol{\theta}^{(t+1)}$.

Repeat step 2-3 for $i = 1, \cdots, D_X$.

Step 2. Calculate $\theta_i^{(t)}$ from (4.52), and treating it as the current estimate of $\boldsymbol{\theta}$, run one EM iteration to obtain $\widetilde{\theta_i^{(t+1)}}$.

Step 3. Obtain the ratio,

$$r_{ij}^{(t)} = \frac{\widetilde{\theta_i^{(t+1)}} - \hat{\theta}_j}{\theta_i^{(t)} - \hat{\theta}_i}, \quad \text{for } j = 1, \cdots, D_X \tag{4.54}$$

OUTPUT: $\theta_i^{(t+1)}$ and $\left\{r_{ij}^{(t)}; i, j = 1, \cdots, D_X\right\}$.

Note that theoretically, the closer $\boldsymbol{\theta}^{(t)}$ is to the global optimum, $\hat{\boldsymbol{\theta}}$, the better the estimate of DM. In practice, if it is too close, the round-off errors would dominate and

influence the precision of DM. Therefore, Meng and Rubin (1991) suggested obtaining r_{ij} when the sequence of $r_{ij}^{(l)}, r_{ij}^{(l+1)}, \cdots$ is stable for some $\hat{l}$.

The SEM algorithm exposes the underlying relationship between the rate of convergence, the observed information, the missing information and the complete information matrices. The amount of missing information is determined primarily by the greatest Eigen-value of the DM matrix[35]. The SEM algorithm has a superior theoretical appeal but can be computationally intensive if the M-step has no closed form solution. In practice, when the M-step has to be obtained numerically, and when the rate of convergence is slow, the SEM is becomes excessively expensive. In addition, if high speed of convergence is highly desired and the quasi-Newton algorithm is used, there is no footprint left for the SEM to trace. An illustration of the SEM algorithm and comparison with other widely adopted methods in presented in section 5.3.

4.4.3. Parametric bootstrap

When the covariate of the HMS-V model is fixed[36], $\mathbf{Y_T} \in \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_T$ can be simulated and the standard errors of the MLE can be determined by parametric bootstrap. Note that the input to the Monte Carlo experiment is the "true" parameter vector, whereas the input to the parametric bootstrap is the parameters MLE. Although conceptually different, they are implemented in the same way. The following section outlines the implementation of the parametric bootstrap.

[35] When the M-step has an analytical solution and the amount of missing information is moderate, the SEM provides an excellent algorithm to calculate the variance-covariance matrix of the MLE. The lack of (numerical) symmetry in $\Delta \mathbf{V}_{SEM}$ also helps to highlight some programming mistake.

[36] Note that this does not apply when the covariate is random, the series then has to be simulated according to the joint probability.

4.5. Monte Carlo experiments

4.5.1. Objectives

This section outlines two separate Monte Carlo experiments, which simulate a time series with the fixed flat initial state probabilities, a second with stochastic initial state probabilities. Through the Monte Carlo experiments, it is desired to establish:

- if the asymptotic properties are reflected in good infinite-sample properties. Experimentation is chosen to demonstrate the presence of this property, since its full proof is beyond the scope of this thesis. This aids in understanding the limitations and pitfalls of the MLE.
- the stability of the Hessian matrix by one-sided finite numerical difference,
- the impact of the initial state probabilities, and
- the reliability of the proposed algorithm along with the computational code.

4.5.2. The choice of covariate and "true" coefficients

It is assumed that there is such a time series whose hidden state probabilities are governed by a homogeneous two-state Markov chain. At both states, the volatility is a function of the lagged daily VIX index[37] by CBOE from Jan 02, 1990 to September 15, 2006; the state transition probability is a function of the lagged daily Fed target rate. *Figure 10* and *Figure 11* provide a visual summarization of the VIX and Fed target rate respectively.

[37] **VIX** is the ticker symbol for the **Chicago Board Options Exchange Volatility Index**, a popular measure of the implied volatility of S&P 500 index options. Referred to by some as the *fear index*, it represents one measure of the market's expectation of volatility over the following 30 day period. http://en.wikipedia.org/wiki/VIX

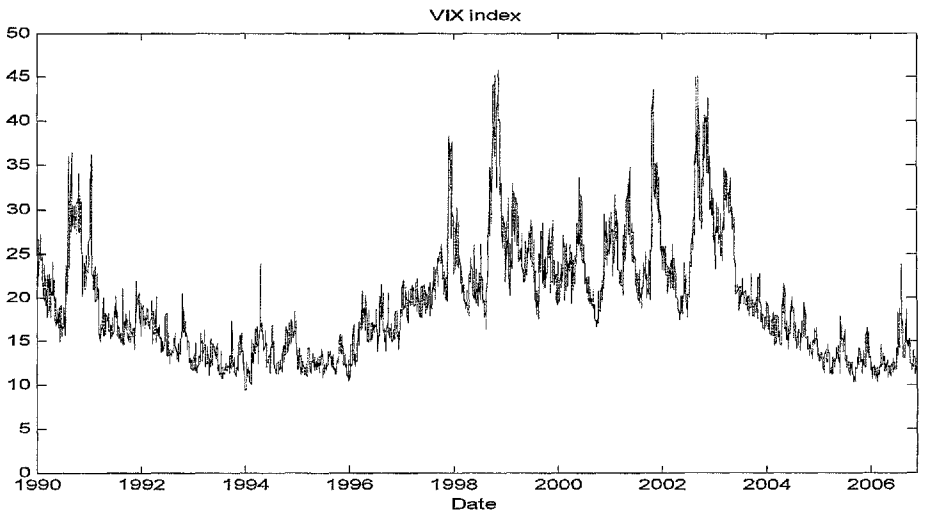


Figure 10
The VIX index by CBOE from Jan 02, 1990 to Sep 15, 2006

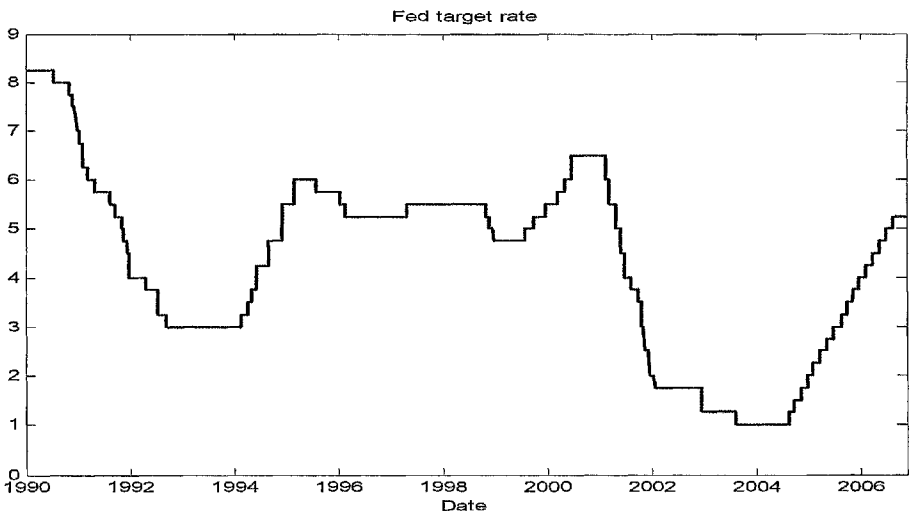


Figure 11
The Fed Target rate from Jan 02, 1990 to Sep 15, 2006

The covariate and the "true" parameter vector of the HMS-V model are designed as follows:-

$\mathbf{c}_t = [1]$;, so that the expected return at each state is a constant: $\mu_{(i),t} = \mu_{(i)}$; $i = 1, 2$;

$$\mathbf{z}_t = [1, \quad \log VIX_{t-1}]; \tag{4.55}$$

$$\mathbf{w}_t = [1, \quad 100 \bullet Fed_{t-1}].$$

To closely align with real world scenarios[38], the covariate and coefficient are designed as follows:

There are two Markov states. The first state is tranquil, the volatility is less sensitive to the lagged VIX index, and the expected return is higher:

$$\alpha_{\langle 1 \rangle, 0} = \mu_{\langle 1 \rangle} = 0.2;$$

$$\begin{aligned} \sigma_{\langle 1 \rangle, t} &= \exp\left(\mathbf{z}_t \boldsymbol{\beta}_{\langle 1 \rangle}\right) \\ &= \exp\left(\boldsymbol{\beta}_{\langle 1 \rangle, 0} + \boldsymbol{\beta}_{\langle 1 \rangle, 1} \bullet VIX_{t-1}\right) \\ &= \exp\left(-1.8 + 0.8 \ \bullet VIX_{t-1}\right). \end{aligned} \tag{4.56}$$

The second state is turbulent, the volatility is more sensitive to the lagged VIX index, and the expected return is lower:

$$\alpha_{\langle 2 \rangle, 0} = \mu_{\langle 2 \rangle} = -0.5;$$

$$\begin{aligned} \sigma_{\langle 2 \rangle, t} &= \exp\left(\mathbf{z}_t \boldsymbol{\beta}_{\langle 2 \rangle}\right) \\ &= \exp\left(\boldsymbol{\beta}_{\langle 2 \rangle, 0} + \boldsymbol{\beta}_{\langle 2 \rangle, 1} \bullet VIX_{t-1}\right) \\ &= \exp\left(-2.0 + 1.2 \ \bullet VIX_{t-1}\right). \end{aligned} \tag{4.57}$$

[38] Please refer to chapter 5 of this thesis for the result and explanation of the empirical result using the S&P 500 stock index. Similar findings are also observed using other famous stock indices, namely, the FTSE 100, NASDAQ, and the Hang Seng Index.

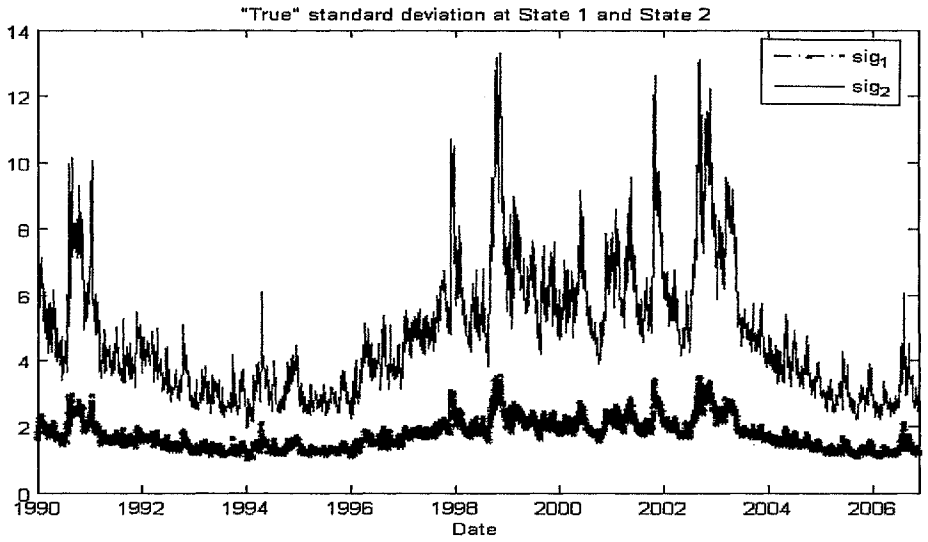


Figure 12
The "true" standard deviations at each state

Given that at time $t-1$, the time series resides at state 1 (i.e. $\Delta_{t-1}=1$), if the FED target rate at time $t-1$ is lower, then the time series at time t is more likely to reside at state 1. The parameters for the staying probability at state 1 are derived as:

$$\begin{aligned} p_{\langle 11\rangle} &= LOGIT\left(\mathbf{w}_t\varphi_{\langle 1\rangle}\right) \\ &= LOGIT\left(\varphi_{\langle 1\rangle,0}+\varphi_{\langle 1\rangle,1}\bullet 100\bullet Fed_{t-1}\right) \\ &= LOGIT\left(2 \quad -0.4 \bullet 100\bullet Fed_{t-1}\right). \end{aligned} \tag{4.58}$$

On the other hand, given that at time $t-1$, the time series resides at state 2 (i.e. $\Delta_{t-1}=2$), if the FED target rate at time $t-1$ is lower, then the time series at time t will be less likely to reside at state 2. The parameters for the staying probability at state 2 are derived as:

$$\begin{aligned} p_{\langle 22\rangle} &= LOGIT\left(\mathbf{w}_t\varphi_{\langle 2\rangle}\right) \\ &= LOGIT\left(\varphi_{\langle 2\rangle,0}+\varphi_{\langle 2\rangle,1}\bullet 100\bullet Fed_{t-1}\right) \\ &= LOGIT\left(-4 \quad +0.8 \bullet 100\bullet Fed_{t-1}\right). \end{aligned} \tag{4.59}$$

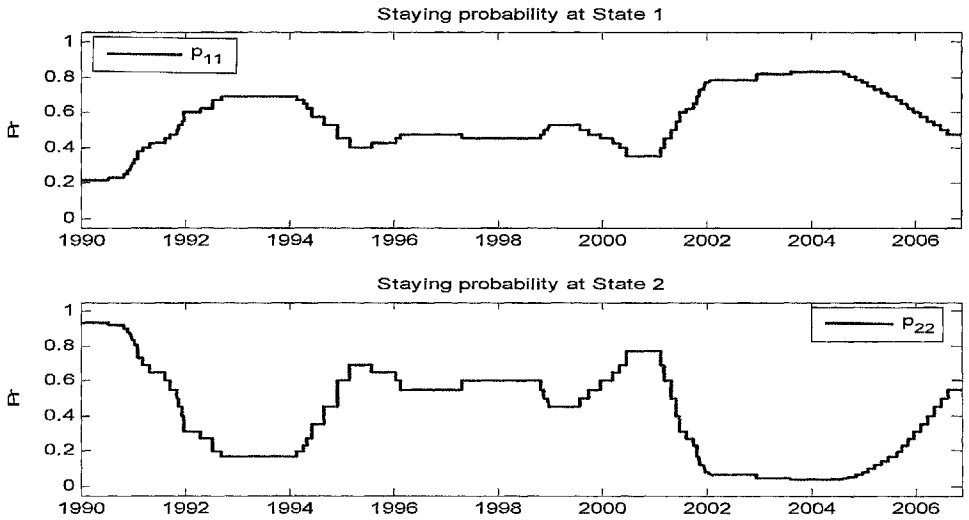


Figure 13

The "true" state transition probabilities at each state

1,000 Monte Carlo simulated time series, $\mathbf{Y_T} \in \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_T$ are generated for both experiment 1 and experiment 2. *Figure 14* plots a sample simulated time series with real world series being shown in *Figure 15* for comparison purposes. The aptly chosen parameters result in a pattern of simulated time series that realistically resemble actual real time series.

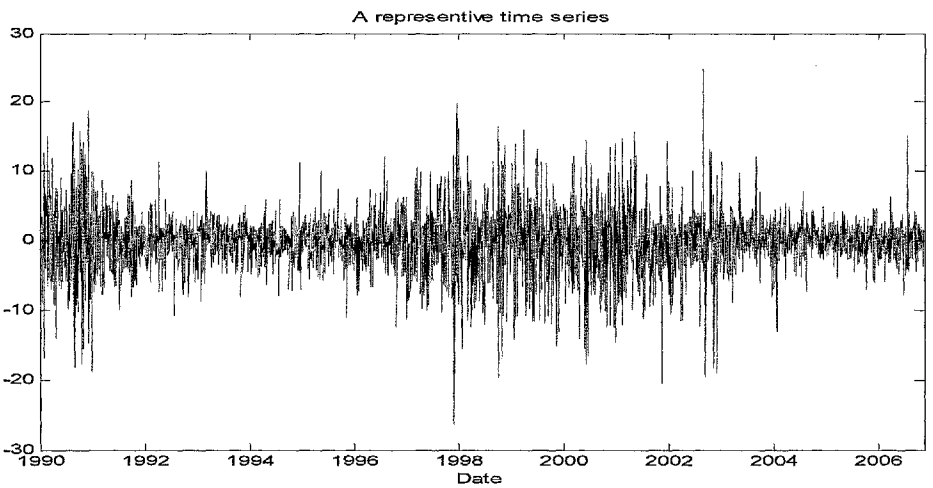


Figure 14

A representative time series simulated according to the model specified in equations (4.55)- (4.59)

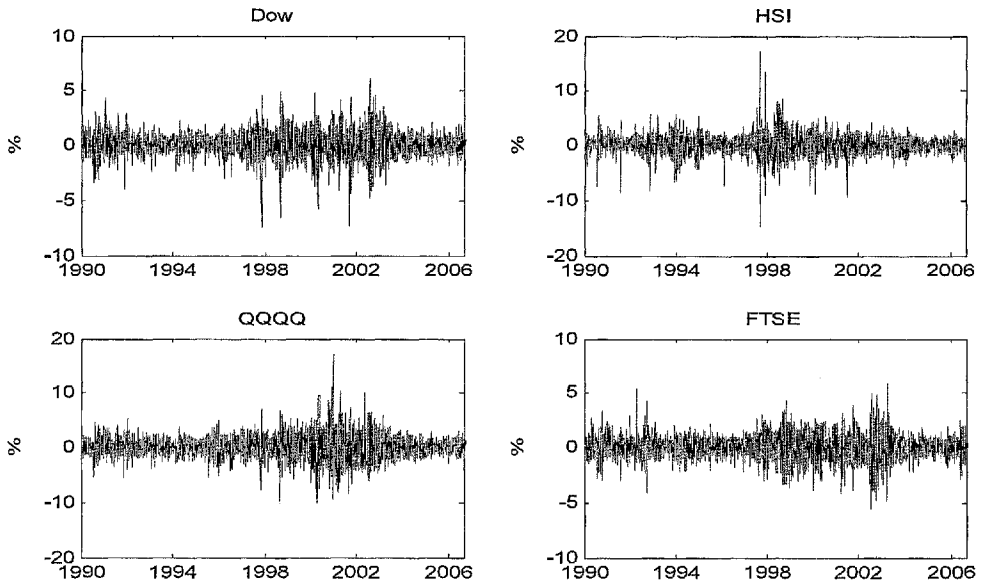


Figure 15
Some real world time series

Note that the representative simulated time series shown in Figure 14 resembles the volatility clustering pattern of real world time series.

***Upper-left panel**: the return data of DOW JONES INDUSTRIAL AVERAGE INDEX (^DJI, Diamond)*	***upper-right panel**: the return data of HANG SENG INDEX (^HSI)*
***lower-left panel**: the return data of NASDAQ-100 INDEX (^NDX, QQQQ)*	***lower-right panel**: the return data of FTSE 100 INDEX (^FTSE)*

4.5.3. Experiment 1: Flat initial state probabilities

Experiment design

INPUT: "True" parameters $\boldsymbol{\theta}$ and covariate $\mathbf{X}_T$, simulate hidden state for $t=1$.

Step 1. Sample a random variable from a uniform distribution $\upsilon_1 \overset{i.i.d}{\sim} UNIF[0,1]$, if $\upsilon_1 \le \Pr(\Delta_1 = 1)$, then $\Delta_1 = 1$, else, $\Delta_1 = 2$. Repeat step 2 for $t = 2,\cdots,T$, and simulate the hidden states.

Step 2. Suppose $\Delta_{t-1} = i$, sample a random variable from a uniform distribution $\upsilon_t \overset{i.i.d}{\sim} UNIF[0,1]$, if $\upsilon_t \le p_{\langle ii \rangle,t} = LOGIT\left(\mathbf{w}_t \boldsymbol{\varphi}_{\langle i \rangle}\right)$, then it remains at its current state i.e. $\Delta_t = i$, else, the state switches to $\Delta_t = j$ and $i \neq j$. Repeat step 3-4 for $t = 1, \cdots, T$, and simulate the time series

Step 3. Suppose $\Delta_t = i$, (which is simulated by step 2), gets the following values of the expected return and the standard deviation respectively,

$$\mu_{\langle i \rangle,t} = \mathbf{c}_t \boldsymbol{\alpha}_{\langle i \rangle}$$

$$\sigma_{\langle i \rangle,t} = \exp\left(\mathbf{z}_t \boldsymbol{\beta}_{\langle i \rangle}\right).$$

Step 4. Simulate a random variable, $\eta_t \overset{i.i.d}{\sim} N(0,1)$, and obtain the value of $y_t \in \mathcal{Y}_i$,

$$y_t = \mu_{\langle i \rangle,t} + \sigma_{\langle i \rangle,t} \bullet \eta_t .$$

Estimate the simulated time series using MLE

Step 5. Use $\boldsymbol{\theta}$ as the starting values, set the initial state probability at 0.5 and use the quasi-Newton algorithm to obtain the MLE of the simulated time series, $\hat{\boldsymbol{\theta}}$.

Step 6. Using $\hat{\boldsymbol{\theta}}$ as the initial parameters, run one E-step of the EM algorithm, and obtain $\widehat{\boldsymbol{s}_{\mathbf{T}}}$. Estimate the standard errors of the estimated parameters by numerical difference.

Step 7. Obtain the Hessian of the MLE by one-sided numerical difference through a Taylor series expansion and calculate the standard errors of the MLE, $\widehat{\mathbf{V}_{num}}$.

OUTPUT: $\hat{\boldsymbol{\theta}}$, $\widehat{\boldsymbol{s}_{\mathbf{T}}}$, $\mathbf{Y}_{\mathbf{T}} \in \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_T$, $\mathbf{q}_{\mathbf{T}} = \{\Delta_1, \cdots, \Delta_T\}$, $\widehat{\mathbf{V}_{num}}$.

Using the same inputs, the above process is repeated 1,000 times to obtain 1,000 sets of outputs.

4.5.4. Experiment 2: Stochastic initial state probabilities

Without covariate, as previously outlined, Leroux and Puterman (1992) showed that the effect of the initial state probability on a hidden Markov-switching Poisson model diminished quickly as the sample size grew. With covariate, Wang and Puterman (1999a, 1999b, and 2001) discovered the same behaviour in an MS-Poisson regression. This experiment investigates whether such a phenomenon also holds true for the proposed HMS-V model by simulating stochastic initial state probabilities.

Experiment design

INPUT: $\boldsymbol{\theta}$ and $\mathbf{X}_T$

Step 0. Sample a random variable $\upsilon_0 \overset{i.i.d}{\sim} UNIF[0,1]$, replace the non-informative initial state probability $\Pr(\Delta_1 = i)$ in $\boldsymbol{\theta}$ by υ_0.

Steps 1-8 are the same as the previous experiment.

OUTPUT: υ_0, $\hat{\boldsymbol{\theta}}$, $\widehat{\boldsymbol{s}_T}$, $\mathbf{Y}_T$, $\mathbf{q}_T = \{\Delta_1, \cdots, \Delta_T\}$ and $\widehat{\mathbf{V}_{nume}}$

1,000 sets of outputs are obtained.

Further extensive study on the resulting sets of $\widehat{\mathbf{V}_{nume}}$ is also carried out. To verify that the numerical method is reliable, the standard errors provided by each set of $\widehat{\mathbf{V}_{nume}}$ are required to be similar to the standard errors from the Monte Carlo study.

4.5.5. Results and Observations

The influence of the initial state probabilities

Table 2 to *Table 5* show that the asymptotic property of MLE is well reflected in good finite-sample properties (i.e., small bias and approximate normality). Both the mean and median of the MLE parameters of the simulated time series closely track the true parameters. The standard errors of the parameter estimate obtained by the two Monte Carlo studies are almost identical. It can be seen that the initial state probability has little impact on the model inference.

The numerical stability of the calculated information matrix

The standard errors estimated by the one-sided numerical difference for each obtained parameter estimate are also analyzed. For experiment 1, the 1,000 sets of standard errors calculate by numerical difference are compared with the distribution of the 1,000 sets of MLE. It can be concluded that the numerical difference is stable and adequately precise. The same conclusion can be drawn from the section experiment.

The asymptotic property of MLE

If it can be confirmed that the asymptotic property of MLE is well reflected in good finite-sample properties, the parametric bootstrap in section 4.4.3 can also be implemented in a similar manner. The only significant difference is that for the Monte Carlo simulation, the input is the true parameter $\boldsymbol{\theta}$, while for the bootstrap, the input is the estimated parameters, $\hat{\boldsymbol{\theta}}$. Therefore 1,000 sets of $\hat{\hat{\boldsymbol{\theta}}}$ are obtained as opposed to 1,000 sets of $\hat{\boldsymbol{\theta}}$. According the variance-covariance matrix of $\hat{\hat{\boldsymbol{\theta}}}$, the standard errors of the parameter estimate can be accordingly determined.

Results of Monte Carlo study 1 (*fixed flat initial state probabilities*) (to be continued)

Parameter	True value	Mean	Median	Standard deviation
$\mu_{\langle 1 \rangle}$	2.00	2.041	2.040	(0.295)
$\beta_{\langle 1 \rangle,0}$	-0.40	-0.410	-0.408	(0.076)
$\beta_{\langle 1 \rangle,1}$	-4.00	-4.174	-4.065	(0.889)
$\mu_{\langle 2 \rangle}$	0.80	0.830	0.813	(0.164)
$\beta_{\langle 2 \rangle,0}$	0.20	0.199	0.199	(0.046)
$\beta_{\langle 2 \rangle,1}$	-1.80	-1.795	-1.795	(0.204)
$\varphi_{\langle 1 \rangle,0}$	0.80	0.798	0.800	(0.070)
$\varphi_{\langle 1 \rangle,1}$	-0.50	-0.503	-0.504	(0.106)
$\varphi_{\langle 2 \rangle,0}$	-2.00	-1.995	-1.993	(0.168)
$\varphi_{\langle 2 \rangle,1}$	1.20	1.198	1.197	(0.057)

For the *i*th simulation, for the expected return at the first state,

the MLE estimate = $\widehat{\mu_{\langle 1 \rangle,i}}$

The mean of the estimate $\widehat{\mu_{\langle 1 \rangle}} = \sum_{j=i}^{1000} \widehat{\mu_{\langle 1 \rangle,i}}$.

Median and standard deviation can be derived in the same manner. Note that to calculate the standard error, the true parameter value $\mu_{\langle 1 \rangle}$ is used.

(continued) Stability of the standard errors by numerical difference

	Standard error of the 1,000 estimates	The summary of the 1,000 numerical differences		
	Standard error of the estimates	Mean of the standard error	Median of the standard error	Standard error of the standard error
$STD\left(\widehat{\mu_{\langle 1\rangle}}\right)$	(0.295)	0.293	2.040	((0.0177))
$STD\left(\widehat{\beta_{\langle 1\rangle,0}}\right)$	(0.076)	0.074	-0.408	((0.0059))
$STD\left(\widehat{\beta_{\langle 1\rangle,1}}\right)$	(0.889)	0.837	-4.065	((0.2060))
$STD\left(\widehat{\mu_{\langle 2\rangle}}\right)$	(0.164)	0.155	0.813	((0.0376))
$STD\left(\widehat{\beta_{\langle 2\rangle,0}}\right)$	(0.046)	0.045	0.199	((0.0017))
$STD\left(\widehat{\beta_{\langle 2\rangle,1}}\right)$	(0.204)	0.199	-1.795	((0.0093))
$STD\left(\widehat{\varphi_{\langle 1\rangle,0}}\right)$	(0.070)	0.068	0.800	((0.0032))
$STD\left(\widehat{\varphi_{\langle 1\rangle,1}}\right)$	(0.106)	0.106	-0.504	((0.0054))
$STD\left(\widehat{\varphi_{\langle 2\rangle,0}}\right)$	(0.168)	0.170	-1.993	((0.0065))
$STD\left(\widehat{\varphi_{\langle 2\rangle,1}}\right)$	(0.057)	0.057	1.197	((0.0022))

Given the *i*th simulation, for the expected return at the first state,

the standard error of the estimates with numerical estimate is $STD_i\left(\widehat{\mu_{\langle 1\rangle}}\right)$

The mean of the standard error = $\sum_{j=i}^{1000} STD_i\left(\widehat{\mu_{\langle 1\rangle}}\right)$.

Similarly, the median of the standard error and standard error of the standard error can be derived.

Table 2

This table summarizes the results of Monte Carlo study 1 with fixed flat initial state probabilities.
The upper panel gives a summary of the 1,000 MLEs of the 1,000 simulated time series.
The lower panel gives a summary of the 1,000 standard errors of each of the MLEs by numerical difference, including comparison with the standard errors estimated from the 1,000 MLEs.

Property of $\left(\widehat{\theta}_i - \theta_i\right) / SE\left(\widehat{\theta}_i\right)$

Monte Carlo study 1 (*fixed flat initial state probabilities*)

Parameter	Mean	Standard deviation	Skewness	Kurtosis
$\widehat{\mu_{\langle 1\rangle}} - \mu_{\langle 1\rangle}$	0.14	1.01	0.11	3.66
$\widehat{\beta_{\langle 1\rangle,0}} - \beta_{\langle 1\rangle,0}$	-0.13	1.03	-0.23	3.76
$\widehat{\beta_{\langle 1\rangle,1}} - \beta_{\langle 1\rangle,1}$	-0.21	1.06	-0.85	4.86
$\widehat{\mu_{\langle 2\rangle}} - \mu_{\langle 2\rangle}$	0.20	1.06	0.82	4.77
$\widehat{\beta_{\langle 2\rangle,0}} - \beta_{\langle 2\rangle,0}$	-0.03	1.03	-0.10	3.04
$\widehat{\beta_{\langle 2\rangle,1}} - \beta_{\langle 2\rangle,1}$	0.02	1.03	0.19	3.40
$\widehat{\varphi_{\langle 1\rangle,0}} - \varphi_{\langle 1\rangle,0}$	-0.03	1.02	-0.18	3.29
$\widehat{\varphi_{\langle 1\rangle,1}} - \varphi_{\langle 1\rangle,1}$	-0.03	1.00	-0.05	2.87
$\widehat{\varphi_{\langle 2\rangle,0}} - \varphi_{\langle 2\rangle,0}$	0.03	0.99	-0.10	2.89
$\widehat{\varphi_{\langle 2\rangle,1}} - \varphi_{\langle 2\rangle,1}$	-0.04	0.99	0.09	2.84

Table 3

This table summarizes the results of Monte Carlo study 1 with fixed flat initial state probabilities. This is to check whether the t-statistic $\left(\hat{\theta}_i - \theta_i\right) \Big/ SE\left(\hat{\theta}_i\right)$ *using different standard errors are approximately standard normal and, in particular, whether they have a unit variance.*

Results of Monte Carlo study 2 (*random initial state probabilities*)

Parameter	True value	Mean	Median	Standard deviation
$\mu_{\langle 1\rangle}$	2.00	2.025	2.023	(0.297)
$\beta_{\langle 1\rangle,0}$	-0.40	-0.406	-0.407	(0.077)
$\beta_{\langle 1\rangle,1}$	-4.00	-4.165	-4.095	(0.868)
$\mu_{\langle 2\rangle}$	0.80	0.830	0.815	(0.160)
$\beta_{\langle 2\rangle,0}$	0.20	0.198	0.198	(0.046)
$\beta_{\langle 2\rangle,1}$	-1.80	-1.790	-1.790	(0.197)
$\varphi_{\langle 1\rangle,0}$	0.80	0.796	0.796	(0.068)
$\varphi_{\langle 1\rangle,1}$	-0.50	-0.502	-0.501	(0.108)
$\varphi_{\langle 2\rangle,0}$	-2.00	-1.996	-1.998	(0.168)
$\varphi_{\langle 2\rangle,1}$	1.20	1.198	1.198	(0.056)

Stability of the standard errors by numerical difference

	Standard error of the 1,000 MLEs	A summary of the 1,000 numerical difference		
	Standard error of MLEs	Mean	Median	Standard error of the standard error
$STD\left(\widehat{\mu_{\langle 1\rangle}}\right)$	(0.297)	0.292	2.023	((0.0174))
$STD\left(\widehat{\beta_{\langle 1\rangle,0}}\right)$	(0.077)	0.074	-0.407	((0.0058))
$STD\left(\widehat{\beta_{\langle 1\rangle,1}}\right)$	(0.868)	0.829	-4.095	((0.1801))
$STD\left(\widehat{\mu_{\langle 2\rangle}}\right)$	(0.160)	0.153	0.815	((0.0325))
$STD\left(\widehat{\beta_{\langle 2\rangle,0}}\right)$	(0.046)	0.045	0.198	((0.0016))
$STD\left(\widehat{\beta_{\langle 2\rangle,1}}\right)$	(0.197)	0.199	-1.790	((0.0093))
$STD\left(\widehat{\varphi_{\langle 1\rangle,0}}\right)$	(0.068)	0.068	0.796	((0.0032))
$STD\left(\widehat{\varphi_{\langle 1\rangle,1}}\right)$	(0.108)	0.106	-0.501	((0.0054))
$STD\left(\widehat{\varphi_{\langle 2\rangle,0}}\right)$	(0.168)	0.170	-1.998	((0.0063))
$STD\left(\widehat{\varphi_{\langle 2\rangle,1}}\right)$	(0.056)	0.057	1.198	((0.0021))

Table 4

The results of Monte Carlo study 2 with random initial state probabilities (following a similar legend to that of table 2). Note that the two tables give almost identical results.

Property of $\left(\hat{\theta}_i - \theta_i\right) / SE\left(\hat{\theta}_i\right)$

Monte Carlo study 2 (*random initial state probabilities*)

Parameter	Mean	Standard deviation	Skewness	Kurtosis
$\widehat{\mu_{\langle 1\rangle}} - \mu_{\langle 1\rangle}$	0.09	1.02	0.15	3.27
$\widehat{\beta_{\langle 1\rangle,0}} - \beta_{\langle 1\rangle,0}$	-0.08	1.04	-0.15	3.15
$\widehat{\beta_{\langle 1\rangle,1}} - \beta_{\langle 1\rangle,1}$	-0.20	1.05	-0.68	4.30
$\widehat{\mu_{\langle 2\rangle}} - \mu_{\langle 2\rangle}$	0.20	1.04	0.59	3.94
$\widehat{\beta_{\langle 2\rangle,0}} - \beta_{\langle 2\rangle,0}$	-0.05	1.01	-0.01	3.01
$\widehat{\beta_{\langle 2\rangle,1}} - \beta_{\langle 2\rangle,1}$	0.05	0.99	-0.06	3.16
$\widehat{\varphi_{\langle 1\rangle,0}} - \varphi_{\langle 1\rangle,0}$	-0.06	0.99	0.05	3.16
$\widehat{\varphi_{\langle 1\rangle,1}} - \varphi_{\langle 1\rangle,1}$	-0.02	1.02	-0.11	3.08
$\widehat{\varphi_{\langle 2\rangle,0}} - \varphi_{\langle 2\rangle,0}$	0.02	0.99	0.00	2.87
$\widehat{\varphi_{\langle 2\rangle,1}} - \varphi_{\langle 2\rangle,1}$	-0.03	0.98	0.02	2.88

Table 5

This table summarizes the result of Monte Carlo study 2 with random flat initial state probabilities. This is to check whether the t-statistic $\left(\hat{\theta}_i - \theta_i\right) \Big/ SE\left(\hat{\theta}_i\right)$ *using different standard errors are approximately standard normal and, in particular, whether they have a unit variance.*

4.6. Forecasting with HMS-V

According to the derivation of the proposed HMS-V model, if the covariate is strictly exogenous, the value of $\widetilde{s_{\langle 1\rangle,t}}$, $\mu_{\langle 1\rangle,t}$, $\sigma^2_{\langle 1\rangle,t}$, $\mu_{\langle 2\rangle,t}$ and $\sigma^2_{\langle 2\rangle,t}$ can be forecasted, where the probability that y_t is generated by the first component is $\widetilde{s_{\langle 1\rangle,t}}$, which is the probability of success of the Bernoulli trial, $\widetilde{s_{\langle 1\rangle,t}} = E\left(s_{\langle 1\rangle,t}\right) = E\left(\Delta_t = 1\right) = \Pr\left(\Delta_t = 1\right)$.

Given the observed (NOT hidden) information up to time t, ψ_t, that is $\left(\mathbf{Y_t}, \mathbf{X_t} : \mathbf{t} = \{1, \cdots, t\}\right)$ and the parameter, $\boldsymbol{\theta}$, the conditional distribution of $y_{t+\tau}$, $\tau = 1, 2, \cdots$, can be forecasted as follows,

Step1. Using $\boldsymbol{\theta}$, run a single E-step of the EM algorithm and estimate the state probabilities at time t, $\left(\widehat{s_{\langle i\rangle,t}} \middle| \psi_t\right) = E\left(\widetilde{s_{\langle i\rangle,t}} \middle| \psi_t\right) = E\left(s_{\langle i\rangle,t} \middle| \mathbf{Y_t}, \mathbf{X_t}, \boldsymbol{\theta}\right)$, $i = 1, 2$.

Step2. Obtain the value of the covariate, $\mathbf{x_h} = \left(\mathbf{w_h}, \mathbf{c_h}, \mathbf{z_h}\right)$; $\mathbf{h} = \{t+1, \cdots, t+\tau\}$.[39]

Step3. With the covariate $\left(\mathbf{w_h}; h = \{t+1, \cdots, t+\tau\}\right)$, and the parameter $\boldsymbol{\varphi}' = \left(\boldsymbol{\varphi}'_{\langle 1\rangle}, \boldsymbol{\varphi}'_{\langle 2\rangle}\right)$, forecast the state transition probabilities $\left(p_{\langle ii\rangle,\mathbf{h}};\ i = 1, 2;\ \mathbf{h} = \{t+1, \cdots, t+\tau\}\right)$, and arrange them in matrix format to obtain $\left(\mathbf{P_h};\ \mathbf{h} = \{t+1, \cdots, t+\tau\}\right)$.

Step4. Forecast the state probability vector. Let,

$$\mathbf{S}'_t = \left[s_{\langle 1\rangle,t}, s_{\langle 2\rangle,t}\right] \quad \text{and} \quad \widehat{\mathbf{S}}'_t = \left[\widehat{s_{\langle 1\rangle,t}}, \widehat{s_{\langle 2\rangle,t}}\right], \tag{4.60}$$

[39] Note that $\mathbf{x_h}$ does not contain any lagged terms of y_h.

Then,

$$\left(\widehat{\mathbf{S}_{t+\tau}}\middle|\psi_t\right)=\left(\widehat{\mathbf{S}_t}\middle|\psi_t\right)\bullet\prod_{h=t+1}^{t+\tau}\mathbf{P}_h\,. \tag{4.61}$$

Step5. With the covariate $\mathbf{c}_{t+\tau}$ and the parameters $\boldsymbol{\alpha}=\left(\boldsymbol{\alpha}_{\langle 1\rangle},\boldsymbol{\alpha}_{\langle 2\rangle}\right)$, forecast the expected return at each state, $\widehat{\mu_{\langle i\rangle,t+\tau}}$.

Step6. With the covariate $\mathbf{z}_{t+\tau}$ and the parameters $\boldsymbol{\beta}=\left(\boldsymbol{\beta}_{\langle 1\rangle},\boldsymbol{\beta}_{\langle 2\rangle}\right)$, forecast the volatility at each state, $\widehat{\sigma_{\langle i\rangle,t+\tau}}$.

$$\begin{aligned}\widehat{\sigma^2_{t+\tau}}=&\Pr\left(\Delta_{t+\tau}=1\middle|\psi_{t+\tau-1}\right)\bullet\left[E\left(\sigma^2_{\langle 1\rangle,t+\tau}\middle|\psi_{t+\tau-1}\right)+E^2\left(\mu^2_{\langle 1\rangle,t+\tau}\middle|\psi_{t+\tau-1}\right)\right]\\&+\Pr\left(\Delta_{t+\tau}=2\middle|\psi_{t+\tau-1}\right)\bullet\left[E\left(\sigma^2_{\langle 2\rangle,t+\tau}\middle|\psi_{t+\tau-1}\right)+E^2\left(\mu^2_{\langle 2\rangle,t+\tau}\middle|\psi_{t+\tau-1}\right)\right]\\&-E^2\left(\mu^2_{t+\tau}\middle|\psi_{t+\tau-1}\right).\end{aligned} \tag{4.62}$$

The above steps are trivial except for step 2, where the value of the covariate $\left(\mathbf{x}_{\mathbf{h}};\ \mathbf{h}=\{t+1,\cdots,t+\tau\}\right)$ should be obtained according to different model specifications.

5. APPLICATION & EMPIRIAL STUDY

Chapter 1 to Chapter 4 proposed a parallel hidden Markov-switching volatility model. This chapter applies the parallel hidden Markov-switching volatility model to analytically study the S&P 500 stock index. The results of this study suggest two additional observations:-

(Obs. 7) ***Asymmetric cycle and lasting tranquility:*** It is observed that the tranquil state is more lasting than the turbulent state. The duration of tranquility (boom) and turbulence (bust) is thus asymmetric. The economics implications of this are outlined in section 6.1.

(Obs. 8) ***Time varying sensitivity to recent shocks:*** According to the empirical study carried out, at the turbulent state, volatility is more sensitive to recent shocks, while the opposite is true for the tranquil state.

In this chapter the empirical study is compared with the single regime GARCH and the MS-GARCH Gray/Klaassen models. The chapter closes by summarizing the study and comparing how different models describe the findings and address the oversights respectively. The empirical study follows the model inference, estimation, and forecast methodology given previously in chapter 2 to chapter 4. In order to provide a benchmark comparison for the weekly volatility estimation and forecast, the daily return data is used to obtain the "integrated weekly volatility".[40]. The mean squared error (MSE) and mean absolute deviation (MAD) suggest that the HMS-V model has greater precision in volatility forecasting. It is observed that the HMS-V model can capture some interesting phenomena of the stock index volatility. For example, it shows that the volatility at the

[40] Please refer to section 5.2.3

tranquil state is best modeled as a constant, while the volatility at the turbulent period is more significantly correlated with the lagged ranges.

5.1. Choice of covariate

Ideally, the chosen covariate should be efficient. The following discusses the rationale behind choosing the lagged observed "range" to construct the covariate of the proposed HMS-V model.[41]

5.1.1. Range vs. observable range

To model the volatility of S&P 500 weekly return, the observed intraweek range is used to construct the covariate of the Logit link functions and the volatility link functions. The "range" is the difference between the log of the highest and lowest price within a certain period of time[42]. According to the literature reviewed in Appendix XIII, the "range" has been proven as an efficient variable to estimate volatility.

Let the time interval be 1 and suppose that the price from $t-1$ to t is *continuously* observable, then the "true" high price, low price, and percentage range from time $t-1$ to t are:-

$$price_t^{(high)} := price_{[t-1,t]}^{(high)} = \max_{\tau \in [t-1,t]} \left(price_\tau \right), \tag{5.1}$$

$$price_t^{(low)} := price_{[t-1,t]}^{(low)} = \min_{\tau \in [t-1,t]} \left(price_\tau \right), \text{ and} \tag{5.2}$$

$$\delta_t := \delta_{[t-1,t]} = 100 \bullet \left[\log\left(price_{[t-1,t]}^{(high)} \right) - \log\left(price_{[t-1,t]}^{(low)} \right) \right], \tag{5.3}$$

[41] In Appendix X, a simple method is also provided to forecast the observable "range", so that the covariate of the HMS-V model can be estimated, which enables the multiple-step volatility forecasting.
[42] Note that only the range is used as the covariate of the regression, rather than regarding the range as the volatility of the time series.

respectively.

If the price of the security from time $t-1$ to t is not continuously observable, then let R_t be defined as the percentage observed range:

$$price_t^{(high)}(\text{obs}) := price_{[t-1,t]}^{(high)}(\text{obs}) = \max_{\tau\in[t-1,t]}\left(price_\tau(\text{obs})\right), \tag{5.4}$$

$$price_t^{(low)}(\text{obs}) := price_{[t-1,t]}^{(low)}(\text{obs}) = \min_{\tau\in[t-1,t]}\left(price_\tau(\text{obs})\right), \text{ and} \tag{5.5}$$

$$R_t := R_{[t-1,t]} = 100\bullet\left[\log\left(price_{[t-1,t]}^{(high)}(\text{obs})\right)-\log\left(price_{[t-1,t]}^{(low)}(\text{obs})\right)\right], \tag{5.6}$$

The close to close percentage return is defined as,

$$y_t := 100\bullet\left[\log\left(price_t^{(close)}\right)-\log\left(price_{t-1}^{(close)}\right)\right]. \tag{5.7}$$

Note that according to the definition of range given in equation (5.1) to equation (5.3), the "true" weekly range δ_t differs from the observed intraweek range R_t defined by equation (5.4) to equation (5.6). A detailed discussion and case study of the stock price of Citigroup Inc. are outlined in Appendix IX.

Parkinson (1980) suggested that if the price of a security follows a normal diffusion process, then a closed form relationship between range and volatility exists. Further details are provided in Appendix VIII. Unfortunately, for an equity index, the price is not continuously observable, which is attributed to the missing price information when the stock exchange is closed. As a result, the observed daily high (low) price is a downward (upward) biased estimate of the "true" high (low) price. Thus, the assumption exploited

by Parkinson (1980) should be modified (or rescaled) for[43] stock indices. In this thesis, the closed form relationship will not be applied. Despite the imperfections, due to the high value information provided by the observed range, the lagged observable range is proposed as covariate for the HMS-V model.

5.1.2. Forecasting range

Based on the relationship between the current volatility and the lagged observable range, given that the range can be forecasted then the multiple-step-ahead volatility may also be forecasted. Further details are available in Appendix X.

5.2. Data

The raw data for the empirical study is the S&P 500 daily close, high and low prices from January 05, 1983.[44] to November 21, 2007. The data source is finance.yahoo.com.

Let Thursday morning[45] be the beginning of the week, and the following Wednesday afternoon[46] be the end of the week. The daily closing price of all trading days of a given week is then used to construct a more reliable benchmark of the "true" volatility (see Appendix X). A total number of 1297 observations for y_t and R_t are thus obtained. *Table 6* gives a statistical summary of the data. The available sample has been divided into two parts. The first 989 observations, corresponding to the period from January 05, 1983 to December 26, 2001, will be used for estimation whereas the remaining 308

[43] If Parkinson's result were to be applied, some rescaling would be required to reflect the amount of missing information.
[44] Data prior to 1983 is not used because the S&P 500 modified the compilation method of the high and low prices around April, 1982.
[45] 09:30 ET
[46] 16:00 ET

observations will be employed for out-of-sample evaluation purposes[47].

Statistical summary for the percentage returns and observed ranges of weekly S&P500 index from January 05, 1983 to November 21, 2007				
	Return	Absolute return	Observed range	Log observed range
Mean	0.17	1.58	2.97	0.95
Median	0.30	1.27	2.52	0.92
Maximum	10.18	16.66	34.37	3.54
Minimum	-16.66	0.00	0.69	-0.37
Standard deviation	2.13	1.43	1.89	0.50
Skewness	-0.6	2.5	5.0	0.3
Kurtosis	7.6	16.3	66.0	3.4
Sample auto-correlation functions (lags)				
ACF (1)	-0.06	0.23	0.51	0.54
ACF (2)	0.01	0.17	0.43	0.50
ACF (3)	0.02	0.11	0.36	0.45
ACF (4)	-0.04	0.16	0.33	0.42
ACF (5)	-0.02	0.13	0.32	0.40
ACF (6)	0.07	0.11	0.34	0.41

Table 6

Descriptive statistics of the weekly percentage returns and the percentage observed intraweek ranges

Figure 39 to *Figure 43* provide visualizations of the data, including volatility clustering, fat tails and the long memory of the time series consists of absolute return.

5.3. Model selection criteria

This section discusses some statistical model selection criteria in order to:-

- study the goodness of fit of the in-sample estimate,
- study the accuracy of out-of-sample forecasting,
- choose the functional form, and
- compare different model specifications.

[47] The percentage log return and percentage log range because of the rescaling convenience mentioned in Appendix III.

5.3.1. Likelihood ratio test

Likelihood-based methods play a central role in parametric testing problems, among which the likelihood ratio (LR) test is perhaps the most popular. Under standard regularity conditions, a LR test has a simple and elegant asymptotic theoretical outline. However, the asymptotic null distribution of the ordinary LR test is violated in a mixture regression model. Therefore, the conventional LR test should not be used to test models with different number of states or non-nested link functional forms. However, the LR test may be applicable in the case of two models which have exactly the same number of states and link functional forms. This is equivalent to testing if one of the models is nested in the other (e.g. regression 02 and regression 03, *Table 8*, section 5.4).

5.3.2. AIC and BIC

The Akaike information criterion (AIC) and Bayesian information criterion (BIC) are among the most popular model selection methods (see Box, Jenkins, and Reinsel, 1994, pp 200-201). AIC was developed by Akaike (1973) as a decision-making strategy based on the Kullback-Leibler (1951) information measure. It provides a natural criterion for ordering alternate statistical models,

$$AIC = -2LLF_{obs}\left(\hat{\theta}\right) + 2K, \tag{5.8}$$

where $\hat{\theta}$ is the MLE and K is the number of parameters. BIC (Schwartz, 1978) finds its roots in the context of Bayesian model choice,

$$BIC = -2LLF_{obs}\left(\hat{\theta}\right) + K \cdot \log\left(T\right) \tag{5.9}$$

It is well known that BIC favors more parsimonious models than AIC due to its harsher

penalization. They may serve as additional (informal) assessments of (relative) model fitness.

5.3.3. MSE, MAD and integrated volatility

From examination of relevant literature, the practice of minimizing a statistical loss function, such as the mean square error (*MSE*) or the mean absolute deviation (*MAD*) is widespread. For the in-sample goodness-of-fit, the loss functions assume the following form[48]:

$$MSE_1\left(\text{in-sample}\right) = T^{-1}\sum_{t=1}^{T}\left(V_t^{1/2} - E\left(\widehat{\sigma_t}\middle|\psi_T\right)\right)^2, \tag{5.10}$$

$$MAD_1\left(\text{in-sample}\right) = T^{-1}\sum_{t=1}^{T}\left|V_t^{1/2} - E\left(\widehat{\sigma_t}\middle|\psi_T\right)\right|. \tag{5.11}$$

where,

$$V_t = (y_t - \overline{y})^2;\ \ \overline{y} = T^{-1}\sum_{t=1}^{T} y_t;\ \ t = 1,\cdots,T; \tag{5.12}$$

If the out-of-sample size is T^*, for a one-step-ahead forecast, the loss functions are,

$$MSE_1\left(\text{out-of-sample}\right) = T^{*-1}\bullet\sum_{t=T+1}^{T+T^*}\left(V_t^{*1/2} - E\left(\widehat{\sigma_t}\middle|\psi_{t-1}\right)\right)^2, \tag{5.13}$$

$$MAD_1\left(\text{out-of-sample}\right) = T^{*-1}\sum_{t=T+1}^{T+T^*}\left|V_t^{*1/2} - E\left(\widehat{\sigma_t}\middle|\psi_{t-1}\right)\right|. \tag{5.14}$$

where,

$$V_t^* = (y_t - \overline{y}^*)^2;\ \text{and}\ \overline{y} = T^{*-1}\sum_{t=T+1}^{T+T^*} y_t;\ \ t = T+1,\cdots,T+T^* \tag{5.15}$$

---

[48] $MSE = T^{*-1}\bullet\sum_{t=T+1}^{T+T^*}\left(V_t^* - E\left(\widehat{\sigma_t^2}\middle|\psi_{t-1}\right)\right)^2$ is not used as some outliners of the square term $\widehat{\sigma_t^2}$ (e.g. the Black Monday, 1987) may greatly skew the result.

The MAD criteria are generally more robust to the possible presence of outliers (e.g. the Black Monday in Oct 1987) than the *MSE* criteria.

However, as pointed out by Anderson and Bollerslev (1998) as well as others, for time series reflecting the return of a closing price to the following closing price, its square provides an unbiased but very noisy[49] measurement of the true volatility. Hence for the more reliable use MSE and MAD, a more efficient measurement of the true volatility is required. As pointed out by Anderson and Bollerslev (1998), to obtain a more efficient measurement of volatility, the square return of higher frequency time series can be pooled. Such a pooled volatility using higher frequency data is defined as "integrated volatility", which will be used as the "true" volatility for benchmarking purposes.

To obtain the "true" intraday volatility, Andersen, Bollerslev and Diebold (2003) discretely sampled a number of prices at regular intervals throughout the day, which lasted from time t to $t+1$. They then obtained the intraday returns and determined the variance of the discrete-time returns over the one-day interval on the sample path $\{\sigma_{t+\tau}\}_{\tau=0}^{1}$:

$$\bar{\sigma}_{t+1}^{2} = \int_{0}^{1} \sigma_{t+\tau}^{2} d\tau \,. \tag{5.16}$$

The integrated volatility, $\bar{\sigma}_{t+1}^{2}$, thus provides a canonical and natural measure of volatility. However, the *daily* integrated volatility is inherently difficult to measure, especially for the stock price since:-

- the weights that are assigned to the opening and closing hours are hard to determine. Qualitatively, it is probable that the prices at the open have richer

[49] For example, if the intraday price jumps up by 10% and jumps down by 10%, the squared return is 0, which is not an efficient benchmark of the true volatility.

information than other trading hours, thus greater weights are allocated to them. However it is very hard to quantify; also

- the integrated intraday price data is of very high frequency and sensitive to variations in the bid-ask bounce.

In order to obtain a relatively robust "true" weekly volatility as a benchmark, the *weekly* volatility is "modeled" by the pooling the square of daily close to close returns,

$$\bar{\sigma}_{t+1} = \sqrt{\sum y_{t+\tau}^2}, \qquad \tau = \left\{\frac{1}{5}, \frac{2}{5}, \cdots, 1\right\}. \tag{5.17}$$

where $y_{t+\tau}^2$ is the squared daily return in the week as defined by equation (5.7). Since the expected daily return is usually much smaller than the standard deviation, $E\left(y_{t+\tau}^2\right)$ injects only a small amount of bias to $\sigma_{t+\tau}^2$. The integrated volatility $\bar{\sigma}_{t+1}^2$ is thus a more efficient representation. In addition, it is insensitive to the rich information of the opening and closing hours as well as the bid-ask bounce. Using the same definition of a week defined in section 5.1.3, the daily closing price of all trading days is adopted to construct a more reliable benchmark of the "true" weekly volatility.

The estimated in-sample weekly volatility can then be compared with the realized integrated volatility:-

$$MSE_2\left(\text{in-sample}\right) = T^{-1}\sum_{t=1}^{T}\left(\bar{\sigma}_t - E\left(\widehat{\sigma_t}\middle|\psi_T\right)\right)^2, \tag{5.18}$$

$$MAD_2\left(\text{in-sample}\right) = T^{-1}\sum_{t=1}^{T}\left|\bar{\sigma}_t - E\left(\widehat{\sigma_t}\middle|\psi_T\right)\right|. \tag{5.19}$$

If the out-of-sample size is T^*, for a one-step-ahead forecast, the loss functions are,

$$MSE_2\left(\text{out-of-sample}\right) = T^{*-1} \bullet \sum_{t=T+1}^{T+T^*} \left(\overline{\sigma}_t - E\left(\widehat{\sigma}_t \middle| \psi_{t-1} \right) \right)^2, \tag{5.20}$$

$$MAD_2\left(\text{out-of-sample}\right) = T^{*-1} \sum_{t=T+1}^{T+T^*} \left| \overline{\sigma}_t - E\left(\widehat{\sigma}_t \middle| \psi_{t-1} \right) \right|. \tag{5.21}$$

Since "integrated volatility" is more efficient, only the MSE_2 and MAD_2 defined by equation (5.18) to equation (5.21) are reported instead of MSE_1 and MAD_1 defined by equation (5.10) to equation (5.14).

5.3.4. The auto-correlation of the forecasting error

If the forecasted volatility, $E\left[\sigma_t \middle| \psi_{t-1}\right]$, were unbiased then the difference between the "true" volatility and the forecasted volatility, $\sigma_t\left(error\right) = E\left[\sigma_t \middle| \psi_{t-1}\right] - \sigma_t$,[50] should have zero auto-correlation. If the forecasted volatility tends to continuously over-forecast or under-forecast volatility, then $\sigma_t\left(error\right)$ would display significant autocorrelations.

As documented in chapter 3, forcefully fitting a Markov-switching model to a linear single regime model like GARCH, the GARCH parameter would likely suggest a long memory process. As suggested by Hamilton and Susmel (1994), amongst others, the high persistency of long memory could be spurious. As a result of the *linear* learning and dis-learning of information, the forecasted volatility can be rendered too "smooth". Therefore, apart from MSE and MAD studies of the forecasting error, the auto-correlation of $\sigma_t\left(error\right)$ is also investigated as a more significant auto-correlation suggests greater complication.

[50] In this test the integrated volatility $\overline{\sigma}_t$ is used as a proxy for the "true" volatility σ_t.

5.3.5. The c.d.f. of out-of-sample forecast

If the model is appropriately specified then the *c.d.f.* of the one-step-ahead forecast should be uniformly *i.i.d.* This is a universal rule that applies to any model with any distribution assumption. Let the information up to time t be ψ_t, the probability density function of y_{t+1} be $f(y_{t+1}|\psi_t)$ and the associated cumulative distributional function be $F(y_{t+1}|\psi_t)=\int_{-\infty}^{y_{t+1}} f(v|\psi_t)dv$. Assuming an appropriate model specification, then $F(y_{t+1}|\psi_t)$ should follow an independent and identical uniform distribution. Following the work of Berkowitz (2001), a second transformation can be performed as,

$$\widehat{\eta_{t+1}}=\Phi^{-1}\left(CDF\ \left(y_{t+1}|\psi_t\right)\middle|0,1\right),\quad t=T,\cdots,T+T^*-1 \tag{5.22}$$

and test if $\widehat{\eta_{t+1}}$ is *i.i.d.* normal, $N(0,1)$. According to Berkowitz (2001), the series $\left\{\widehat{\eta_{t+1}}\right\}$ inherits all the bias of $\left\{CDF\ \left(y_{t+1}|\psi_t\right)\right\}$. A similar and but more intuitive method of the Berkowitz test is to study the skew and kurtosis of $\left\{\widehat{\eta_{t+1}}\right\}$.

To implement the test, $\left\{\widehat{\eta_{t+1}}\right\}$ needs to be initially determined. Note that according to the HMS-V model and estimated parameters $\hat{\theta}$, given ψ_t and y_{t+1}, the conditional one-step-ahead *c.d.f.* is,

$$\begin{aligned}\widehat{h_{t+1}} &= CDF\ \left(y_{t+1}\middle|\psi_t,\hat{\theta}\right)\\ &= \widehat{s_{\langle 1\rangle,t+1}}\int_{v=-\infty}^{y_{t+1}}\phi\left(v\middle|\widehat{\mu_{\langle 1\rangle,t+1}},\widehat{\sigma^2_{\langle 1\rangle,t+1}}\right)dv\\ &\quad+\left(1-\widehat{s_{\langle 1\rangle,t+1}}\right)\int_{v=-\infty}^{y_{t+1}}\phi\left(v\middle|\widehat{\mu_{\langle 2\rangle,t+1}},\widehat{\sigma^2_{\langle 2\rangle,t+1}}\right)dv\end{aligned}, \tag{5.23}$$

where,

$$\widehat{s_{\langle 1\rangle,t+1}} = E\left\{s_{\langle 1\rangle,t+1} \middle| \psi_t\right\}; t = T,\cdots,T+T^*-1. \tag{5.24}$$

Note tha an E-step must be performed based on the information ψ_t to obtain $\widehat{s_{\langle 1\rangle,t}} = E\left\{s_{\langle 1\rangle,t} \middle| \psi_t, \hat{\theta}\right\}$ before forecasting $\widehat{s_{\langle 1\rangle,t+1}}$. More thorough details were previously outlined in section 4.6.

5.4. HMS-V without covariate: Bull vs. bear, a qualitative assessment

The HMS-V model without covariate is the simplest and also most robust hidden Markov-switching model. Simple as it is, this model merits investigation as it is the most transparent, tractable and thus intuitively understandable. Although from a statistical perspective, the performance of HMS-V without covariate may not exceed that of GARCH models, economically, it does provide some basic but differing economic implications. For example, as reviewed in chapter 3, for stock indices, some literature suggests that the bull is associated with tranquility and the bear with turbulence.

Without covariate, every link function at each state refers to a constant. *Table 7* gives the parameter estimate and model performance metrics listed in section 5.3.

HMS-V model without covariate				
Model index		**HMS Reg. 01**		
First state	**Estimate**[51]	***Standard error*** **SEM**	***Standard error*** **Bootstrap**	***Standard error*** **Numerical difference**
$\widehat{\mu_{\langle 1\rangle}}$	0.301	(0.0648)	(0.0673)	(0.0648)
$\widehat{\sigma_{\langle 1\rangle}}$	1.564	(0.0597)	(0.0548)	(0.0595)
$\widehat{p_{\langle 11\rangle}}$	0.988	(0.0074)	(0.0060)	(0.0073)
Second state				
$\widehat{\mu_{\langle 2\rangle}}$	0.035	(0.1661)	(0.1738)	(0.1663)
$\widehat{\sigma_{\langle 2\rangle}}$	2.937	(0.1440)	(0.1391)	(0.1444)
$\widehat{p_{\langle 22\rangle}}$	0.980	(0.0145)	(0.0127)	(0.0146)
In-sample fit				
Log likelihood	-2091.4			
AIC	4194.7			
BIC	4224.1			
MSE2	1.36			
MAD2	0.69			
out-of-sample forecast (one-step-ahead)				
MSE2	0.82			
MAD2	0.65			
Skewness of . $\widehat{\eta_t}$	0.02			
Kurtosis of $\widehat{\eta_t}$	3.6			
ACF of the first 4 lags of $E[\sigma_t \mid \psi_{t-1}] - \overline{\sigma}$	0.39 0.40 0.28 0.21			

Table 7

The regression results of the HMS-V model without covariate. The standard errors are calculated through the SEM algorithm, the parametric bootstrap and one-sided numerical difference respectively. For any model, the c.d.f. of the out-of-sample forecast should be i.i.d. uniform; therefore, the transformed $\hat{\eta}_t$ *should follow i.i.d. normal with zero mean, unit variance, zero skewness and kurtosis=3.*

$$y_t \mid \psi_{t-1} \sim \begin{cases} N\left(\mu_{\langle 1\rangle}, \sigma_{\langle 1\rangle}\right) & w.p. \quad \Pr\left(\Delta_t = 1\right) \\ N\left(\mu_{\langle 2\rangle}, \sigma_{\langle 2\rangle}\right) & w.p. \quad \Pr\left(\Delta_t = 2\right), \end{cases}$$

$$p_{\langle ij\rangle,t} = \Pr\left(\Delta_t = j \mid \Delta_{t-1} = i\right).$$

[51] The standard errors obtained by the numerical difference and SEM algorithm are closely matched but the standard errors obtained by the parametric bootstrap are dissimilar. This is likely to be caused by two reasons:

- this particular time series under study has only 989 observations, which is less than the 4213 observations used previously for the Monte Carlo study in section 4.5; and
- the staying probabilities for both states is very high (0.988 for State 1 and 0.98 for State 2).

As a result, the parametric bootstrap may not give an accurate estimation of the standard errors, since for the simulated time series, the hidden state may only transit for a very limited number of times.

The standard errors of the transition probabilities and the standard deviation at each state are relatively small, indicating the existence of two states with distinctly different volatilities. State 1 describes a tranquil market coupled with higher expected weekly return, *0.30*, and lower volatility, *1.57*; on the contrary, state 2 describes a turbulent market characterized with lower expected return *0.036*, and higher volatility, *2.94*.

If the risk free interest rate were 0.0%[52], the shape ratios holding the stock indices return at the two states are:-

$$\text{Shape ratio at State 1 (tranquil state)} = \frac{0.30 - Rate_{Free}}{1.56} \approx 0.19 \quad ,$$

$$\text{Shape ratio at State 2 (turbulent state)} = \frac{0.036 - Rate_{Free}}{2.94} = 0.012 \quad .$$

Therefore, during a turbulent period, to hold this stock index, an investor suffers not only from greater uncertainty but also lower return. Chapter 6 provides some mutually reciprocal mechanisms in volatility and return to determine the economic implication.

The high staying probabilities are determined as $\widehat{p_{\langle 11\rangle}} = 0.988$, $\widehat{p_{\langle 22\rangle}} = 0.980$. It can be seen that $\widehat{p_{\langle 11\rangle}} > \widehat{p_{\langle 22\rangle}}$, thus a tranquil state is on average more lasting than a turbulent state. The economic implications of a longer lasting tranquility will be discussed later in chapter 6.

Based on Regression 01 (see *Figure 16* to *Figure 17*),

- the following periods are tranquil: Year 1983 to 1986, 1988 to early 1990, 1992 to

[52] For ease of presentation, this is an over-simplification.

early 1998, and 2004 to early 2007.

- The other periods are turbulent, often associated with stock market crashes and/or economic recessions. During a turbulent regime, volatility traders and short sellers often thrive[53].

As indicated by *Figure 17*, since August 2007, the market has most likely switched to a turbulent regime. The U.S. housing slump is affecting growth, the credit problems in the U.S. have spread to other parts of the world, the global economy faces downside risks from the rout of capital markets and at the same time, inflation is above the comfort level of most major central banks. Hence simple as the model is, it does give a qualitative assessment of the question: "is the volatility considered high or low"?

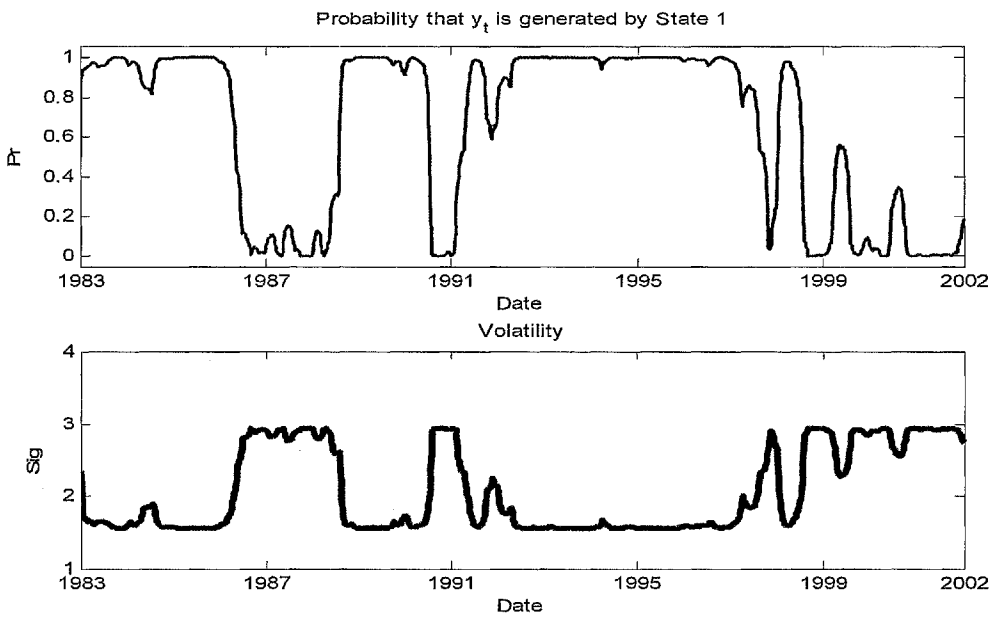


Figure 16

Regression 01, in-sample (HMS-V without covariate). The upper panel plots $\left(\widehat{s_{(1),T}}|\psi_T\right)$, $\mathbf{T}=\{1,\cdots,T\}$, *that is, the in-sample state 1 probability, the lower panel plots the corresponding* $\left(\widehat{\sigma_T}|\psi_T\right)$.

[53] During such turbulent times, one simple strategy is to trade the VIX index by CBOE. Since the volatility of the stock market is mean reverting (see Dueker, 1997), during a turbulent age, one can long the VIX index when it is below the long-term equilibrium and short the VIX index when it is above the long-term equilibrium.

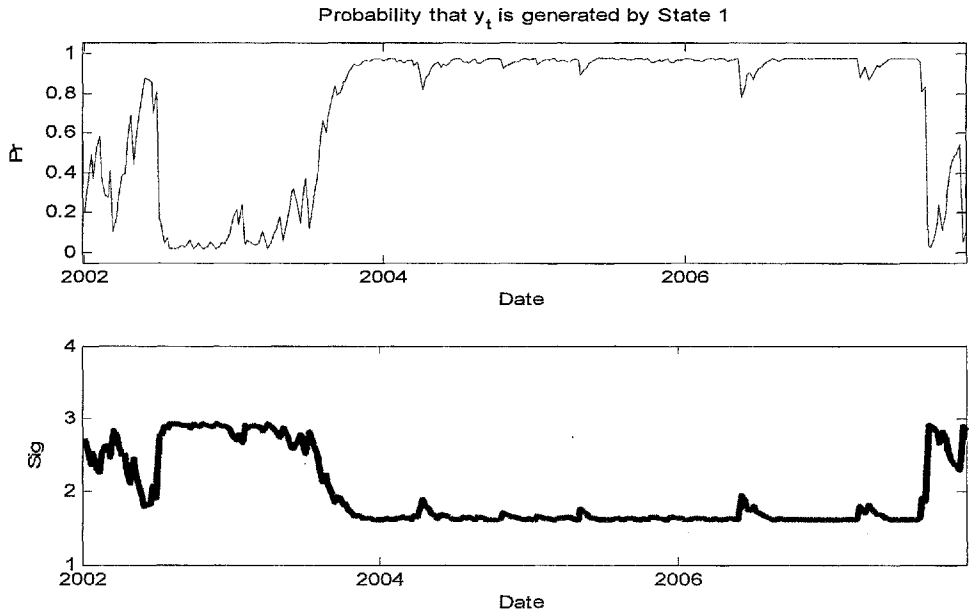


Figure 17

Regression 01, out-of-sample (HMS-V without covariate). The upper panel plots $\left\{\widehat{s_{(1),t}}|\psi_{t-1}\right\}$; $t = T, T+1, \cdots, T+T^*$, *that is, the one-step-ahead out-of-sample state 1 probability. The lower panel plots the corresponding* $\left\{\widehat{\sigma}_t|\psi_{t-1}\right\}$.

5.5. HMS-V with covariates: Quantitative fine-tuning

Regression 01 qualitatively accesses volatility and Regression 02 to Regression 05 quantitatively fine-tune it, where the volatility link functions and logic link function are used[54]. It is further observed that compared to the tranquil state, the turbulent state is far more reactive to recent shocks.

5.5.1. Link functions and covariates

The link functions are:-

1) The transition probability is determined by the Logit link function,

[54] Actually, six link functions can be used to replace each of the six parameters of Regression 01, but of the six parameters, the means of two states are still assumed constant.

$$p_{\langle ii\rangle,t} = LOGIT\left(\mathbf{w}_t \boldsymbol{\varphi}_{\langle i\rangle}\right); \qquad \left(i = 1,2;\, t = 2,\cdots T\right). \tag{5.25}$$

If the current intraweek range has a homogeneous impact on the transition probabilities within the next n weeks, then the row vector $\mathbf{w}_t$ has only two elements, $\mathbf{w}_t = \left(\mathbf{w}_{t,0}, \mathbf{w}_{t,1}\right) = \left(1, \mathbf{w}_{t,1}\right)$, where,

$$\mathbf{w}_{t,1} = n^{-1} \sum_{k=1}^{n} R_{t-k} \tag{5.26}$$

is the average of the last n intraweek ranges[55].

Alternatively, if a more recent intraweek range has greater impact on the transition probability, an exponential weighted form can be specified as,

$$\mathbf{w}_{t,1} = \sum_{k=1}^{\infty} \lambda^i \left(1-\lambda\right) R_{t-k}\,. \tag{5.27}$$

One can let $\lambda = 0.94$ according to RiskMetrics™ [56] (although $\lambda = 0.94$ is probably not the optimal parameter, it is used only to illustrate the potential improvement of such an exponentially-weighted method).

2) The expected return of y_t, given that $\Delta_t = i$, is assumed constant,

$$\mu_{\langle i\rangle,t} = \mu_{\langle i\rangle}; \qquad \left(i = 1,2;\, t = 1,\cdots T\right).$$

This assumption based on the following two reasons:-

- the autocorrelations (or clustering) of the weekly rerun are addressed by the clustering feature of the Markov-switching model; and

[55] Previous versions proposed by the author imposed the condition, $w_t = \frac{1}{n}\sum_{k=1}^{n} \log R_{t-k}$, however recent investigations show that, $w_t = \frac{1}{n}\sum_{k=1}^{n} R_{t-k}$ is a slightly better specification.

[56] To be more faithful to RiskMetrics™, one can let $\mathbf{w}_{t,1} = \sqrt{\sum_{k=1}^{\infty} \lambda^i \left(1-\lambda\right) R^2_{t-k}}$.

- the clustering of bullish or bearish market sentiments are of more interest.

3) Volatility is determined by an exponential function: given that $\Delta_t = i$, the standard deviation of y_t is,

$$\sigma_{\langle i\rangle,t} = \exp\left(\mathbf{z}_t \boldsymbol{\beta}_{\langle i\rangle}\right); \quad (i = 1,2;\ t = 2,\cdots T). \tag{5.28}$$

Let $\mathbf{z}_t = \left(\mathbf{z}_{t,0}, \mathbf{z}_{t,1}, \cdots, \mathbf{z}_{t,m}\right) = \left(1, \log R_{t-1}, \cdots, \log R_{t-m}\right)$ be the covariate of the volatility link function[57] and assume that the volatility of y_t is correlated with the m lagged observable ranges. For the parameter $\boldsymbol{\beta}_{\langle i\rangle} = \left(\boldsymbol{\beta}_{\langle i\rangle,0}, \boldsymbol{\beta}_{\langle i\rangle,1}, \cdots, \boldsymbol{\beta}_{\langle i\rangle,m}\right)$, $\boldsymbol{\beta}_{\langle i\rangle,0}$ is the intersection term and $\left\{\boldsymbol{\beta}_{\langle i\rangle,k};\ k = 1,\cdots,m\right\}$ describes the impacts of the lagged observable ranges on $\sigma_{\langle i\rangle,t}$.

5.5.2. Time varying sensitivity to recent shocks

The regression results of Regression 02 to Regression 05 are presented in this section. Regression 02 and Regression 03 both assume $w_t = \frac{1}{26}\sum_{k=1}^{26} R_{t-k}$. According to the findings of Regression 02, compared to the turbulent state, at the tranquil state, volatility is far less sensitive to recent shocks. Therefore, Regression 03 imposes the condition, $\beta_{\langle 1\rangle,1} = \beta_{\langle 1\rangle,2} = 0$ to test the hypothesis that volatility at the tranquil state is constant. Regression 04 and Regression 05 both assume $\mathbf{w}_{t,1} = \sum_{k=1}^{\infty} \lambda^t (1-\lambda) R_{t-k}$, so that a more recent observation has greater impact on the transition probability. Regression 05

[57] Previous work adopted, $\mathbf{z}_t = \left(1,\ R_{t-1}, \cdots, R_{t-m}\right)$, however the obtained results were slightly inferior to those obtained when using log range.

also imposes that $\beta_{\{1\},1} = \beta_{\{1\},2} = 0$, $\beta_{\{2\},2} = 0$, that is, the volatility at the tranquil state is constant and volatility at the turbulent state is only correlated with one lag of the intraweek range. Regressions 02 with 03 are then compared (see *Table 8*), as are Regressions 04 with 05 (see *Table 9*).

Model index	**Reg. 02**		**Reg. 03**	
m and n value	***m=2; n=26***		***m=2; n=26***	
	Estimate	***Standard error***	**Estimate**	***Standard error***
First state				
$\widehat{\mu}_{\langle 1\rangle}$	0.44	(0.10)	0.48	(0.11)
$\widehat{\beta}_{\langle 1\rangle,0}$	0.02	(0.13)	0.20	(0.09)
$\widehat{\beta}_{\langle 1\rangle,1}$	0.15	(0.13)	-	-
$\widehat{\beta}_{\langle 1\rangle,2}$	0.13	(0.14)	-	-
Logit link				
$\widehat{\varphi}_{\langle 1\rangle,0}$	4.18	(1.83)	4.93	(2.58)
$\widehat{\varphi}_{\langle 1\rangle,1}$	-1.39	(0.67)	-1.92	(1.15)
Second state				
$\widehat{\mu}_{\langle 2\rangle}$	-0.24	(0.23)	-0.15	(0.15)
$\widehat{\beta}_{\langle 2\rangle,0}$	0.51	(0.13)	0.45	(0.12)
$\widehat{\beta}_{\langle 2\rangle,1}$	0.31	(0.10)	0.32	(0.09)
$\widehat{\beta}_{\langle 2\rangle,2}$	0.14	(0.09)	0.14	(0.08)
Logit link				
$\widehat{\varphi}_{\langle 2\rangle,0}$	-2.92	(2.91)	-4.04	(2.06)
$\widehat{\varphi}_{\langle 2\rangle,1}$	0.84	(0.97)	1.26	(0.60)
	In-sample fit			
Log Likelihood	-2059.3		-2061.1	
AIC	4142.7		4142.2	
BIC	4201.4		4191.2	
MSE2	1.12		1.15	
MAD2	0.59		0.60	
	out-of-sample forecast (one-step-ahead)			
MSE2	0.60		0.61	
MAD2	0.57		0.58	
ACF of the first 4 lags of $E[\sigma_t \mid \psi_{t-1}] - \bar{\sigma}$	0.11 0.19 0.12 0.06		0.13 0.22 0.13 0.07	
Skewness of $\widehat{\eta}_t$	0.02		0.03	
Kurtosis of $\widehat{\eta}_t$	2.77		2.85	

Table 8

The regression results of the HMS-V model with covariate.

$$y_t \mid \psi_{t-1} \sim \begin{cases} N\left(\mu_{\langle 1\rangle,t}, \sigma_{\langle 1\rangle,t}\right) & w.p. \quad \Pr\left(\Delta_t = 1\right) \\ N\left(\mu_{\langle 2\rangle,t}, \sigma_{\langle 2\rangle,t}\right) & w.p. \quad \Pr\left(\Delta_t = 2\right) \end{cases},$$

$$\sigma_{\langle i\rangle,t} = \exp\left(\beta_{\langle i\rangle,0} + \beta_{\langle i\rangle,1} \cdot \log R_{t-1} + \beta_{\langle i\rangle,2} \cdot \log R_{t-2}\right),$$

$$p_{\langle ii\rangle,t} = \Pr\left(\Delta_t = j \mid \Delta_{t-1} = i\right).$$

$$p_{\langle ii\rangle,t} = LOGIT\left(\varphi_{\langle i\rangle,0} + \varphi_{\langle i\rangle,1} \cdot w_t\right), \text{ where } w_t = \frac{1}{26}\sum_{k=1}^{26} R_{t-k}$$

Model index	Reg. 04		Reg. 05	
	$\mathbf{w}_{t,1}=\sum_{k=1}^{\infty}\lambda^{i}(1-\lambda)R_{t-k}$, λ=0.94		$\mathbf{w}_{t,1}=\sum_{k=1}^{\infty}\lambda^{i}(1-\lambda)R_{t-k}$, λ=0.94	
	Estimate	***Standard error***	**Estimate**	***Standard error***
First state				
$\widehat{\mu_{\langle 1\rangle}}$	0.37	(0.08)	0.33	(0.08)
$\widehat{\beta_{\langle 1\rangle,0}}$	0.13	(0.12)	0.31	(0.06)
$\widehat{\beta_{\langle 1\rangle,1}}$	0.16	(0.12)	-	-
$\widehat{\beta_{\langle 1\rangle,2}}$	0.01	(0.11)	-	-
Logit link				
$\widehat{\varphi_{\langle 1\rangle,0}}$	9.07	(3.08)	19.03	(15.52)
$\widehat{\varphi_{\langle 1\rangle,1}}$	-3.31	(1.40)	-7.05	(5.80)
Second state				
$\widehat{\mu_{\langle 2\rangle}}$	-0.15	(0.21)	-0.03	(0.17)
$\widehat{\beta_{\langle 2\rangle,0}}$	0.65	(0.13)	0.64	(0.11)
$\widehat{\beta_{\langle 2\rangle,1}}$	0.29	(0.09)	0.35	(0.09)
$\widehat{\beta_{\langle 2\rangle,2}}$	0.07	(0.09)	-	-
Logit link				
$\widehat{\varphi_{\langle 2\rangle,0}}$	-8.62	(3.99)	-9.02	(3.23)
$\widehat{\varphi_{\langle 2\rangle,1}}$	2.28	(1.02)	2.45	(0.91)
	In-sample fit			
Log Likelihood	-2044.7		-2045.9	
AIC	4113.4		4109.8	
BIC	4172.2		4153.9	
MSE2	1.12		1.14	
MAD2	0.58		0.60	
	out-of-sample forecast (one-step-ahead)			
MSE2	0.60		0.61	
MAD2	0.57		0.57	
ACF of the first 4 lags of $E[\sigma_t\|\psi_{t-1}]-\bar{\sigma}$	0.15 0.25 0.13 0.07		0.18 0.29 0.16 0.08	
Skewness of $\widehat{\eta_t}$	-0.04		-0.12	
Kurtosis of $\widehat{\eta_t}$	2.90		3.09	

Table 9

The regression result of the HMS-V model with covariate.

$$y_t|\psi_{t-1}\sim\begin{cases}N(\mu_{\langle 1\rangle,t},\sigma_{\langle 1\rangle,t}) & w.p. \quad \Pr(\Delta_t=1)\\ N(\mu_{\langle 2\rangle,t},\sigma_{\langle 2\rangle,t}) & w.p. \quad \Pr(\Delta_t=2)\end{cases},$$

$$\sigma_{\langle i\rangle,t}=\exp\left(\beta_{\langle i\rangle,0}+\beta_{\langle i\rangle,1}\cdot\log R_{t-1}+\beta_{\langle i\rangle,2}\cdot\log R_{t-2}\right),$$

$$p_{\langle ij\rangle,t}=\Pr(\Delta_t=j\mid\Delta_{t-1}=i).$$

$$p_{\langle ii\rangle,t}=LOGIT\left(\varphi_{\langle i\rangle,0}+\varphi_{\langle i\rangle,1}\cdot w_t\right),\ \text{where}\ \sum_{k=1}^{\infty}\lambda^{i}(1-\lambda)R_{t-k},\ \lambda=0.94.$$

The Regression 02, which does not impose any constraint on the parameters, is analyzed here. $\widehat{\varphi_{\langle 1\rangle,1}} = -1.39$ is negative and statistically significant, which indicates that given that the previous observation is at the first state (i.e. tranquil state), there is an increased likelihood that the next observation will switch to the turbulent state with increasing lagged intraweek ranges. The converse is true for $\widehat{\varphi_{\langle 2\rangle,1}} = 0.84$. Furthermore, the volatility at the turbulent state is significantly positively correlated with the lagged intraweek ranges[58]. At the turbulent state, $\widehat{\boldsymbol{\beta}_{\langle 2\rangle,1}} = 0.31$, $\widehat{\boldsymbol{\beta}_{\langle 2\rangle,2}} = 0.14$, with the t-statistics also being significant. On the other hand, the *t*-statistics of $\widehat{\boldsymbol{\beta}_{\langle 1\rangle,1}}$ and $\widehat{\boldsymbol{\beta}_{\langle 1\rangle,2}}$ are far less significant[59] at the tranquil state. Therefore, compared to the tranquil state, volatility at the turbulent state is much more sensitive to the lagged observed range. Further insights are evident from the raw data.

According to f*Figure 16* and *Figure 17* most of the observations, y_t, from the period May/13/1992 to Dec/06/1995, i.e., from the 487th observation to the 673rd observation, are likely to have been generated by the tranquil state. Table *10* describes the sample ACFs of y_t^2, and the sample cross-correlation function (XCFs) between y_t^2 and the lagged terms of $\log(R_t)$. The sample ACFs are clearly not significant, indicating very weak auto-regressive [G]ARCH effect, if any at all. The sample XCFs of the first 5 lagged terms are all less than 0.10, indicating relatively weak correlations of the lagged intraweek ranges and volatility at the tranquil state.

---

[58] For $\widehat{\boldsymbol{\beta}_{\langle 2\rangle,1}}$, the t-statistic is 0.31/0.10=3.1 ; and for $\widehat{\boldsymbol{\beta}_{\langle 2\rangle,2}}$, the t-statistics is 0.14/0.09=1.5 .

[59] For $\widehat{\boldsymbol{\beta}_{\langle 1\rangle,1}}$, the t-statistic is 0.15/0.13=1.1; and for $\widehat{\boldsymbol{\beta}_{\langle 1\rangle,2}}$, the t-statistics is 0.13/0.14=0.9.

Weekly observations from May/13/1992 to Dec/06/1995 Sample Auto-Correlation function of squared return (left) Sample Cross-Correlation function between squared return and log range (right)			
Auto-Correlation function (lag)		*Cross-Correlation function (lag)*	
ACF (1)	-0.03	XCF (1)	0.09
ACF (2)	-0.03	XCF (2)	0.02
ACF (3)	-0.04	XCF (3)	0.08
ACF (4)	0.05	XCF (4)	0.06
ACF (5)	-0.11	XCF (5)	-0.13
ACF (6)	0.06	XCF (6)	0.07
ACF (7)	0.19	XCF (7)	0.14
ACF (8)	-0.01	XCF (8)	-0.06

Table 10

This table gives the sample ACFs of y_t^2 *and the sample XCF between* y_t^2 *and the lagged term of* $\log(R_t)$ *from May/13/ 1992 to Dec/06/1995.*

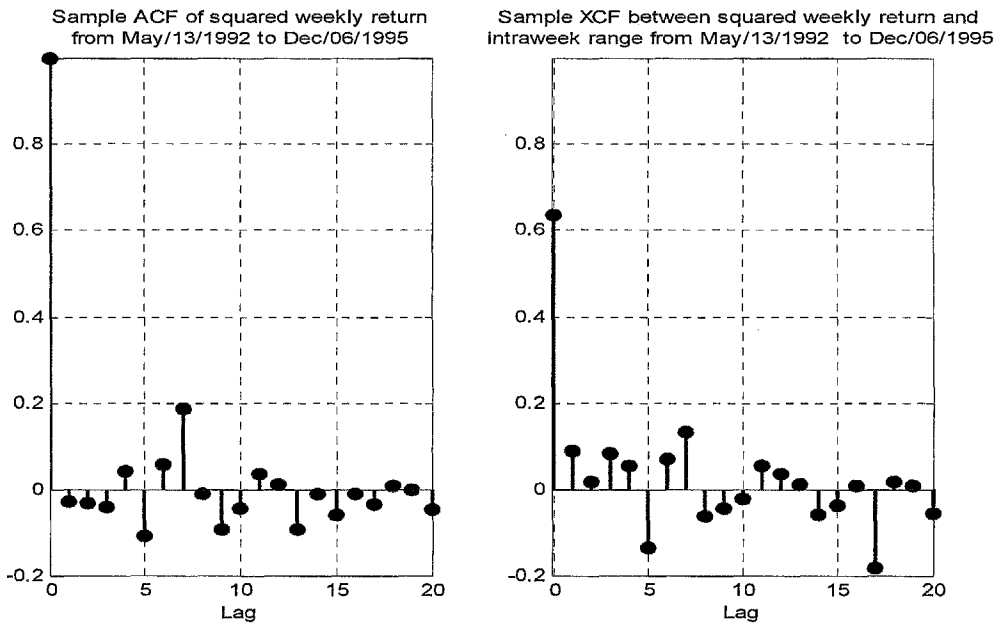


Figure 18

The left panel plots the sample autocorrelation function (sample ACFs) of y_t^2 *from 13th May 1992 to 6th December 1995. The right panel plots the sample cross correlation function (XCFs) between* y_t^2 *and the lagged term of* $\log(R_t)$ *in the same period.*

A similar finding for the period Jan/07/2004 to Jun/27/2007 is shown in *Table 11* and *Figure 19*.

Weekly observations from Jan/07/2004 to Jun/27/2007 Sample Auto-Correlation function of squared return (left) Sample Cross-Correlation function between squared return and log range (right)			
Auto-Correlation function (lag)		*Cross-Correlation function (lag)*	
ACF (1)	-0.02	XCF (1)	0.03
ACF (2)	-0.06	XCF (2)	0.11
ACF (3)	0.06	XCF (3)	0.04
ACF (4)	0.05	XCF (4)	0.18
ACF (5)	-0.05	XCF (5)	0.01
ACF (6)	-0.03	XCF (6)	-0.04
ACF (7)	0.06	XCF (7)	0.04
ACF (8)	-0.01	XCF (8)	0.01

Table 11

The table gives the sample ACFs of the y_t^2 *and the sample XCF between* y_t^2 *and the lagged term of* $\log(R_t)$ *from Jan/07/2004 to Jun/27/2007.*

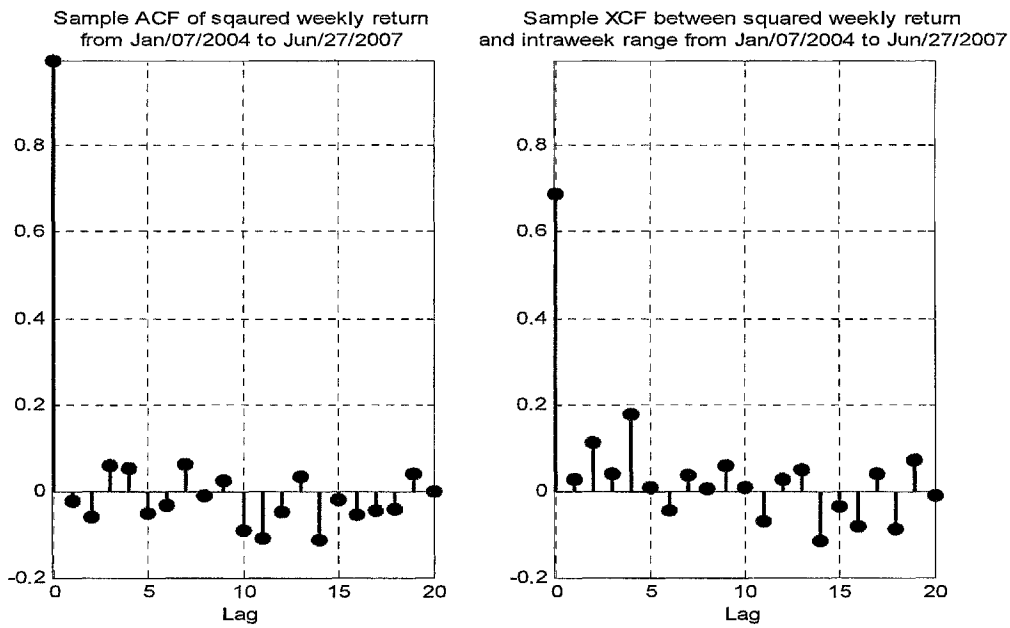


Figure 19

The left panel plots the sample autocorrelation function (sample ACFs) of y_t^2 *from Jan/07/2004 to Jun/27/2007. The right panel plots the sample cross correlation function (XCFs) between* y_t^2 *and the lagged term of* $\log(R_t)$ *during the same period.*

Conversely, the estimated value and *t*-statistics of $\widehat{\beta_{\langle 2\rangle,1}}$ and $\widehat{\beta_{\langle 2\rangle,2}}$ in Regression 02 indicate that at the turbulent period, the lagged intraweek range has significant

explanatory power of volatility. According to the top panel of *Figure 16* and the top panel of *Figure 17*, the vast majority of observations from Jan/08/1997 (the 730th observation) to Jun/02/2003 (the 1068th observation) are likely to reside at the turbulent state.

Table 12 and *Figure 20* describe the sample ACFs of y_t^2 as well as the sample XCFs between y_t^2 and the lagged terms of $\log(R_t)$. Both the sample ACFs and XCFs are much more statistically significant than those at the tranquil periods (e.g. May/13/1992 to Dec/06/1995), indicating obvious structure breaks.

Weekly observations from Jan/08/1997 to Jun/02/2003 Sample Auto-Correlation function of squared return (left) Sample Cross-Correlation function between squared return and log range (right)			
Auto-Correlation function (lag)		*Cross-Correlation function (lag)*	
ACF (1)	0.13	XCF (1)	0.22
ACF (2)	0.04	XCF (2)	0.24
ACF (3)	0.05	XCF (3)	0.12
ACF (4)	0.15	XCF (4)	0.12
ACF (5)	0.10	XCF (5)	0.09
ACF (6)	-0.04	XCF (6)	0.05
ACF (7)	0.06	XCF (7)	0.05
ACF (8)	0.01	XCF (8)	0.00

Table 12

The table gives the sample ACFs of y_t^2 *and the sample XCF between* y_t^2 *and the lagged term of* $\log(R_t)$ *from January 08, 1997 to June 02, 2003.*

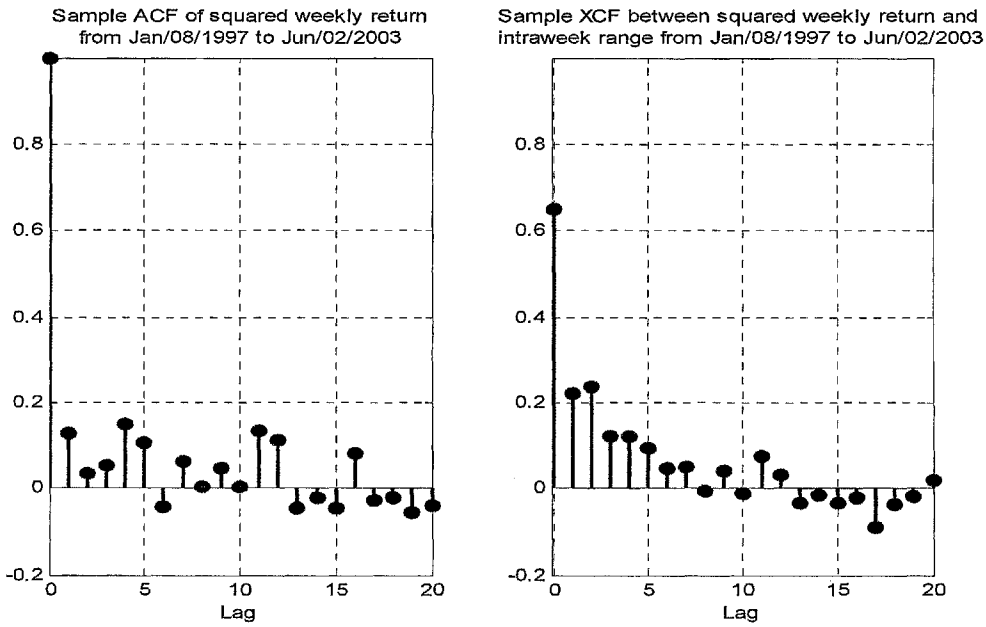


Figure 20

The left panel plots the sample autocorrelation function (sample ACFs) of y_t^2 *from January 08, 1997 to June 02, 2003. The right panel plots the sample cross correlation function (XCFs) between* y_t^2 *and the lagged term of* $\log(R_t)$ *during the same period.*

The analysis of the raw data is inline with the parameter estimate carried out for Regression 02. Hence, as opposed to using only a single set of one-size-fit-all parameters, different sets of parameters are adopted to describe the differing volatility patterns. Further study is merited if volatility at the tranquil state can be specified as a constant. Imposing restrictions on the two parameters ($\beta_{\langle 1\rangle,1}, \beta_{\langle 1\rangle,2}$), Regressions 03 and 02 are compared to examine the hypothesis:-

$$\mathbf{H}_0: \ \beta_{\langle 1\rangle,1} = 0 \ \text{ and } \ \beta_{\langle 1\rangle,2} = 0 .$$

Defining the parameter space that satisfies $\beta_{\langle 1\rangle,1} = 0$ and $\beta_{\langle 1\rangle,2} = 0$ as Θ_0, the LR test suggests:

$$-2\log\left(\frac{\sup\{LF_{obs}(\theta):\theta\in\Theta_0\}}{\sup\{LF_{obs}(\theta):\theta\in\Theta\}}\right) = -2(-2061.1+2059.3) \approx 3.6, \qquad (5.29)$$

where the P value is $1-\chi_2^2(3.6)=16.5\%$ and the null hypothesis, $\boldsymbol{\theta}\in\boldsymbol{\Theta}_0$, is not rejected at a 90% confidence interval. In addition, according to *Table 8*, the BIC values for Regression 02 and Regression 03 are 4201.4 and 4191.2 respectively, which favor Regression 03. Similarly, the LR test prefers Regression 05 to Regression 04, that is, the volatility at the tranquil state cannot be specified as a constant. As shown in *Table 8* and *Table 9*, if the BIC were to be used as the yardstick, Regression 05 would demonstrate the best performance among the 5 regressions. The in-sample estimation of Regression 05 is depicted in *Figure 21* to *Figure 23*, with its out-of-sample forecast shown in *Figure 24* to *Figure 26*.

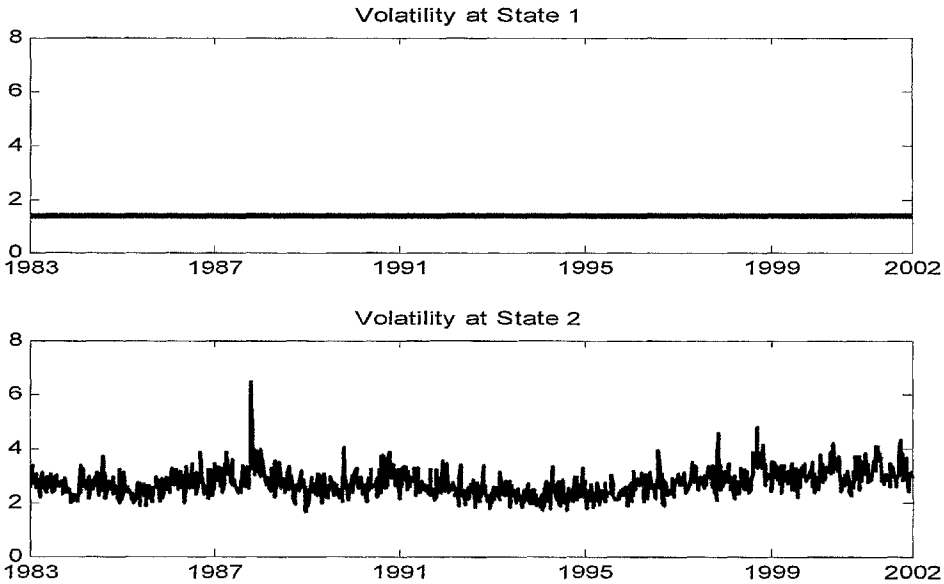


Figure 21

Regression 05, in-sample ($\beta_{\langle 1\rangle,1} = \beta_{\langle 1\rangle,2} = 0$, $\beta_{\langle 2\rangle,2} = 0$).

This figure plots $\left(\widehat{\sigma_{\langle 1\rangle,\mathbf{T}}}\middle|\psi_T\right)$ and $\left(\widehat{\sigma_{\langle 2\rangle,\mathbf{T}}}\middle|\psi_T\right)$; $\mathbf{T} = \{1,\cdots,T\}.$, $T = 989$.

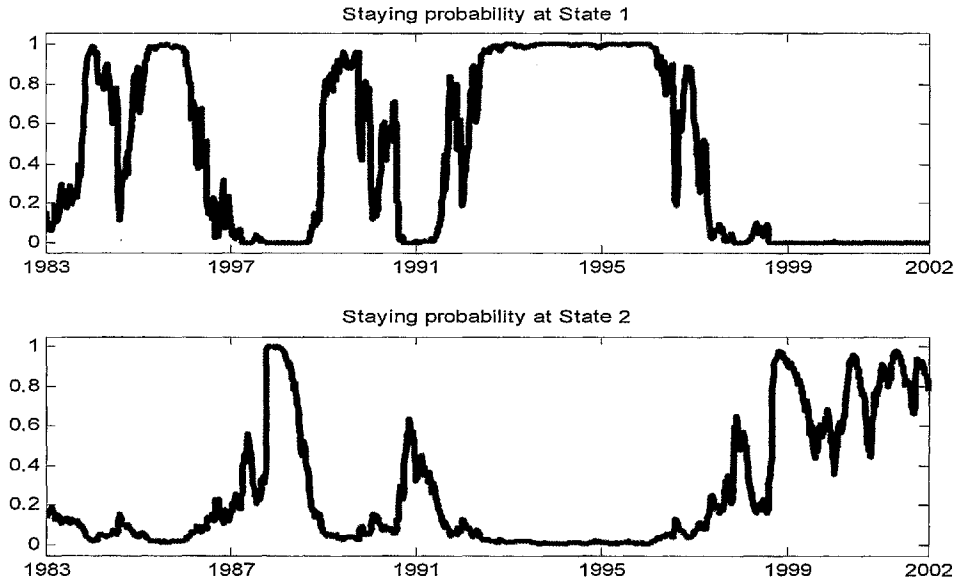


Figure 22

Regression 05, in-sample ($\beta_{\langle 1\rangle,1} = \beta_{\langle 1\rangle,2} = 0$, $\beta_{\langle 2\rangle,2} = 0$)

This figure plots $\left(\widehat{p_{\langle 11\rangle,\mathbf{T}}}\middle|\psi_T\right)$ *and* $\left(\widehat{p_{\langle 22\rangle,\mathbf{T}}}\middle|\psi_T\right)$; $\mathbf{T} = \{1,\cdots,T\}.$, $T = 989$.

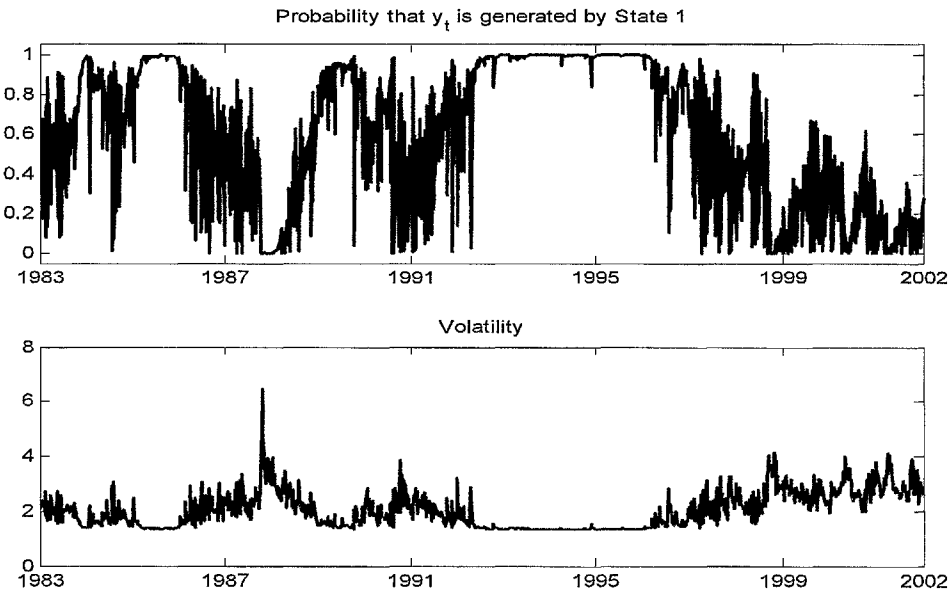


Figure 23

Regression 05, in-sample ($\beta_{\{1\},1} = \beta_{\{1\},2} = 0$, $\beta_{\{2\},2} = 0$)

This figure plots $\left(\widehat{\boldsymbol{s}_{\{1\},\mathbf{T}}} \middle| \psi_T\right)$ and $\left(\widehat{\sigma_{\mathbf{T}}} \middle| \psi_T\right)$; $\mathbf{T} = \{1, \cdots, T\}$, $T = 989$.

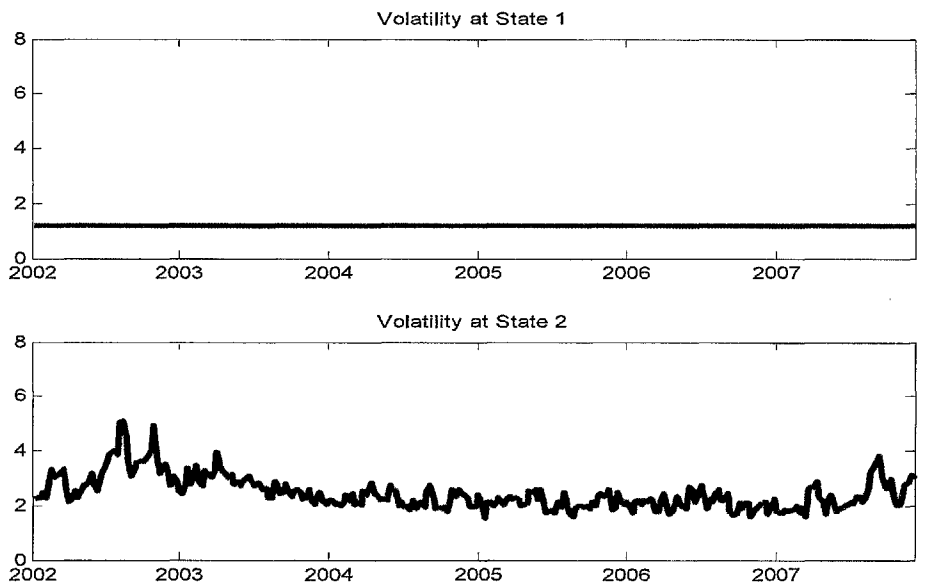


Figure 24

Regression 05, out-of-sample ($\beta_{\{1\},1} = \beta_{\{1\},2} = 0$, $\beta_{\{2\},2} = 0$)

This figure plots $\left\{\widehat{\sigma_{\{1\},t}} \middle| \psi_{t-1}\right\}$ and $\left\{\widehat{\sigma_{\{2\},t}} \middle| \psi_{t-1}\right\}$, $t = T, T+1, \cdots, T+T^*$, $T = 989$, $T^* = 1297$.

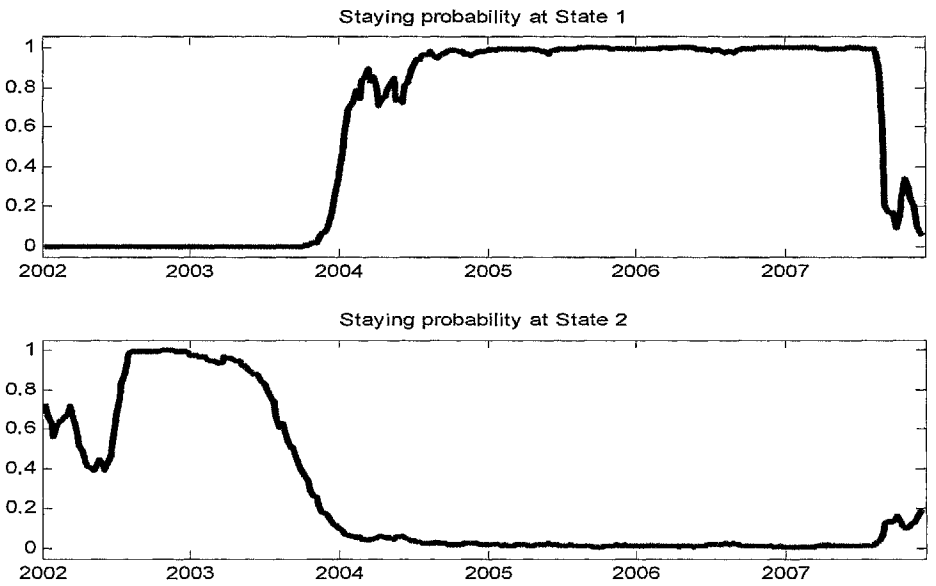


Figure 25

Regression 05, out-of-sample ($\beta_{\langle 1\rangle,1} = \beta_{\langle 1\rangle,2} = 0$, $\beta_{\langle 2\rangle,2} = 0$)

This figure plots $\left\{\widehat{p_{\langle 11\rangle,t}}\middle|\psi_{t-1}\right\}$ and $\left\{\widehat{p_{\langle 22\rangle,t}}\middle|\psi_{t-1}\right\}$, $t = T, T+1, \cdots, T+T^*$.,

$T = 989$, $T^* = 1297$.

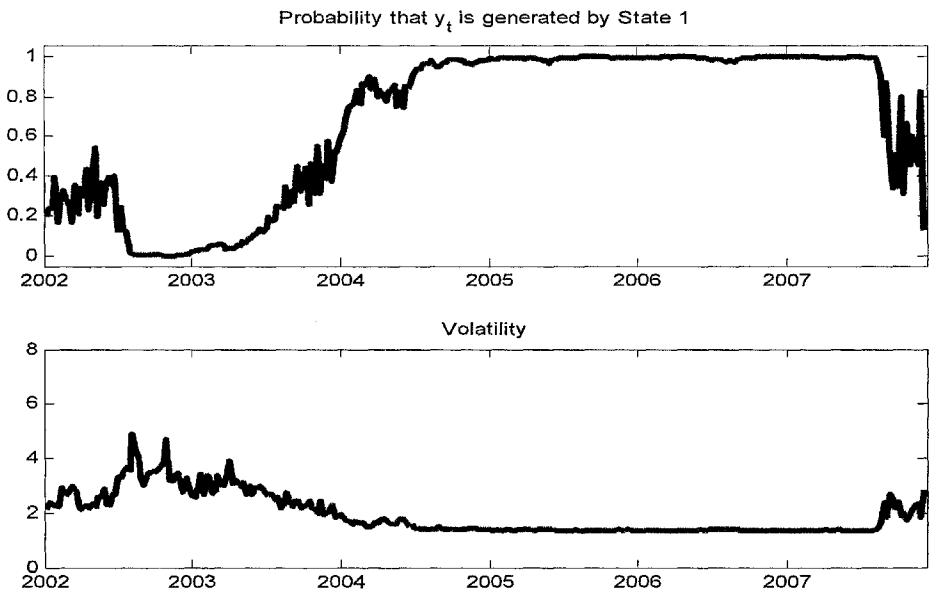


Figure 26

Regression 05, out-of-sample ($\beta_{\langle 1\rangle,1} = \beta_{\langle 1\rangle,2} = 0$, $\beta_{\langle 2\rangle,2} = 0$)

This figure plots $\left\{\widehat{s_{\langle 1\rangle,t}}\middle|\psi_{t-1}\right\}$ and $\left\{\widehat{\sigma_t}\middle|\psi_{t-1}\right\}$; $t = T, T+1, \cdots, T+T^*$., $T = 989$,

$T^* = 1297$.

5.6. Model Comparisons (empirical study)

5.6.1. Comparison with single regime models

5.6.1.1. Comparison using statistical loss functions

Using the same data described in section 5.2,

Table 13 reports the results of other popular models amongst the research community, namely the GARCH-Normal and GARCH-T. The results of Reg. 03 and Reg. 05 are also provided. It is evident that GARCH-T significantly out-performs GARCH-Normal in terms of likelihood based criteria such as AIC, BIC. This is expected as the student's *t* distribution is a far more potent conditional distribution for fat-tailed returns. Amongst all the regression samples, that of 05 has the most favorable performance in terms of AIC and BIC. However, when the "integrated volatility" is used as the proxy for the "true volatility", the GARCH-N and GARCH-T model have very similar performance in terms of in-sample and out-of-sample MSE and MAD. From these results the merits of the proposed model are compounded as it out-performs both single regime GARCH models.

Model index	GARCH(1,1) Normal		GARCH(1,1) Student's t		Model index	Reg. 03		Reg. 05	
						m=2; n=26		$w_{t,1} = \sum_{k=1}^{\infty} \lambda^{t}(1-\lambda) R_{t-k}$, λ=0.94	
	Est.	***Std. Error***	**Est.**	***Std. Error***		**Est.**	***Std. Error***	**Est.**	***Std. Error***
$\widehat{\mu}$	0.25	(0.0064)	0.32	(0.057)	$\widehat{\mu_{\langle 1\rangle}}$	0.48	(0.11)	0.33	(0.08)
$\widehat{\rho}$	-0.06	(0.036)	-0.08	(0.033)	$\widehat{\beta_{\langle 1\rangle,0}}$	0.20	(0.09)	0.31	(0.06)
$\widehat{\beta_0}$	0.075	(0.025)	0.067	(0.036)	$\widehat{\beta_{\langle 1\rangle,1}}$	-	-	-	-
$\widehat{\beta_1}$	0.08	(0.013)	0.06	(0.017)	$\widehat{\beta_{\langle 1\rangle,2}}$	-	-	-	-
$\widehat{\beta_2}$	0.91	(0.017)	0.92	(0.020)	$\widehat{\varphi_{\langle 1\rangle,0}}$	4.93	(2.58)	19.03	(15.52)
$\widehat{DoF}$	-	-	6.67	(1.143)	$\widehat{\varphi_{\langle 1\rangle,1}}$	-1.92	(1.15)	-7.05	(5.80)
					$\widehat{\mu_{\langle 2\rangle}}$	-0.15	(0.15)	-0.03	(0.17)
					$\widehat{\beta_{\langle 2\rangle,0}}$	0.45	(0.12)	0.64	(0.11)
					$\widehat{\beta_{\langle 2\rangle,1}}$	0.32	(0.09)	0.35	(0.09)
					$\widehat{\beta_{\langle 2\rangle,2}}$	0.14	(0.08)	-	-
					$\widehat{\varphi_{\langle 2\rangle,0}}$	-4.04	(2.06)	-9.02	(3.23)
					$\widehat{\varphi_{\langle 2\rangle,1}}$	1.26	(0.60)	2.45	(0.91)
In-sample fit									
Log Likelihood	-2097.3		-2067.0		*Log Likelihood*	-2061.1		-2045.9	
AIC	4204.7		4146.1		*AIC*	4142.2		4109.8	
BIC	4229.1		4175.5		*BIC*	4191.2		4153.9	
MSE2	1.41		1.41		*MSE2*	1.15		1.14	
MAD2	0.71		0.70		*MAD2*	0.60		0.60	
Out-of-sample forecast									
MSE2	0.79		0.79		*MSE2*	0.61		0.61	
MAD2	0.63		0.63		*MAD2*	0.58		0.57	
ACF of the first 4 lags of $E[\sigma_t \mid \psi_{t-1}] - \overline{\sigma}$	0.35		0.36		*ACF of the first 4 lags of* $E[\sigma_t \mid \psi_{t-1}] - \overline{\sigma}$	0.13		0.18	
	0.35		0.36			0.22		0.29	
	0.22		0.24			0.13		0.16	
	0.12		0.14			0.07		0.08	
Skewness of $\widehat{\eta_t}$	-0.45		-0.20		*Skewness of* $\widehat{\eta_t}$	0.03		-0.12	
Kurtosis of $\widehat{\eta_t}$	4.05		2.85		*Kurtosis of* $\widehat{\eta_t}$	2.85		3.09	

Table 13
Comparison between single regime GARCH and the proposed model

For GARCH-Normal, the persistence level is $0.076+0.911=0.988$, indicating that the half life time of a shock is, $\log(0.5)/\log(0.988)=57$ weeks. For GARCH-T, the persistence level is $0.064+0.923=0.987$, which is similar to that of GARCH-Normal. According to Hamilton and Susmel (1994), the persistence level suggested by GARCH-Normal (1,1) is suspiciously long. As discussed previously in section 3.3, if the persistence level were spurious, the model will then either continuously over-forecast and under-forecast volatility, and as a result, $\sigma_t(error)$ will exhibit positive auto-correlation. This is indeed the case for both GARCH models, as the positive auto-correlations of $\sigma_t(error)$ are far more significant than those of the proposed model. Finally, as the skewness and kurtosis of $\widehat{\eta_t}$ show, the proposed model is far more capable of addressing both the negative skewness and the excess kurtosis of S&P 500 return, when compared with both GARCH models.

5.6.1.2. Comparison by closer examinations of two episodes

Apart from the statistical superiority in volatility forecast demonstrated by the proposed HMS-V model, a superior economic implication is also provided. For comparison fairness[60], the realistic volatility forecasting process is mimicked for all models. For example, σ_t is forecasted based on the parameters and the information up to ψ_{t-1}. *Figure 27* and *Figure 28* plot the in-sample estimate and out-of-sample forecast using both the HMS-V and GARCH-T models. The weekly integrated volatility (equation

[60] Note that for the Markov models, the values of $E\left[\sigma_t|\psi_T,\hat{\theta}\right]$ and $E\left[\sigma_t|\psi_{t-1},\hat{\theta}\right]$ are different for the proposed model, while they are the same for GARCH models. Therefore we will use $E\left[\sigma_t|\psi_{t-1},\hat{\theta}\right]$ as the yardstick. This also applies to the expected return.

(5.16)) is also plotted as the benchmark for the "true" volatility. According the figures, HMS-V is more capable of describing the change in volatility when a tranquil state transits to a turbulent state and the forecasted volatility also decreases in a more timely manner when the real volatility (proxied by "integrated volatility") calms down. Hence, in this aspect, HMS-V out-performs GARCH-T.

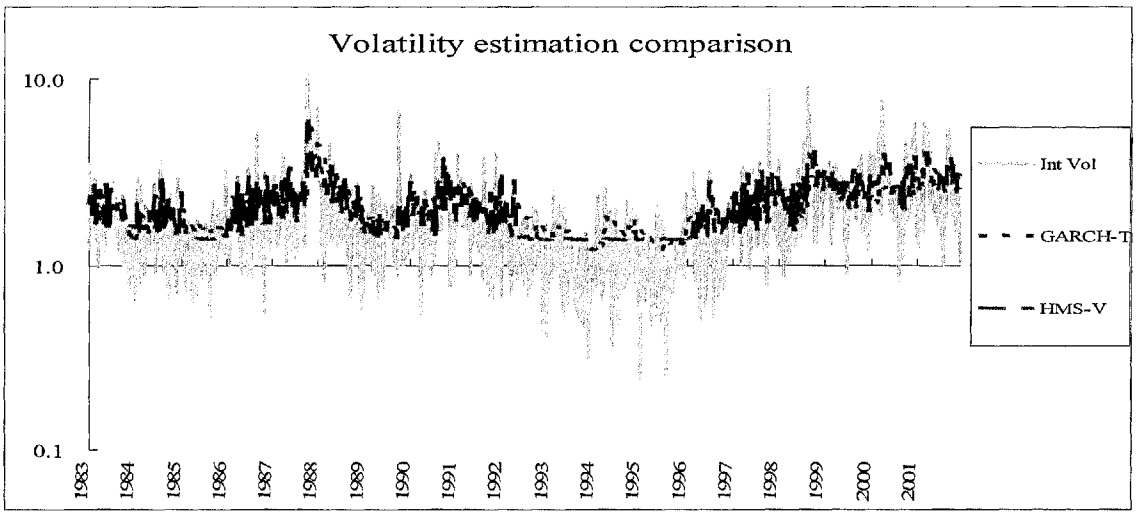


Figure 27

Volatility forecasting $E\left[\sigma_t \middle| \psi_{t-1}, \hat{\theta}\right]$ *using different methods from Jan/05/1983 to Dec/26/2001. The solid line plots the weekly "integrated volatility", the dashed line plots the volatility forecasted by the proposed HMS-V model, and the dotted line plots that of the GARCH-T model. Note: the y-axis of this figure is log scaled. This is due to the broad volatility range caused by Black Monday, 1987.*

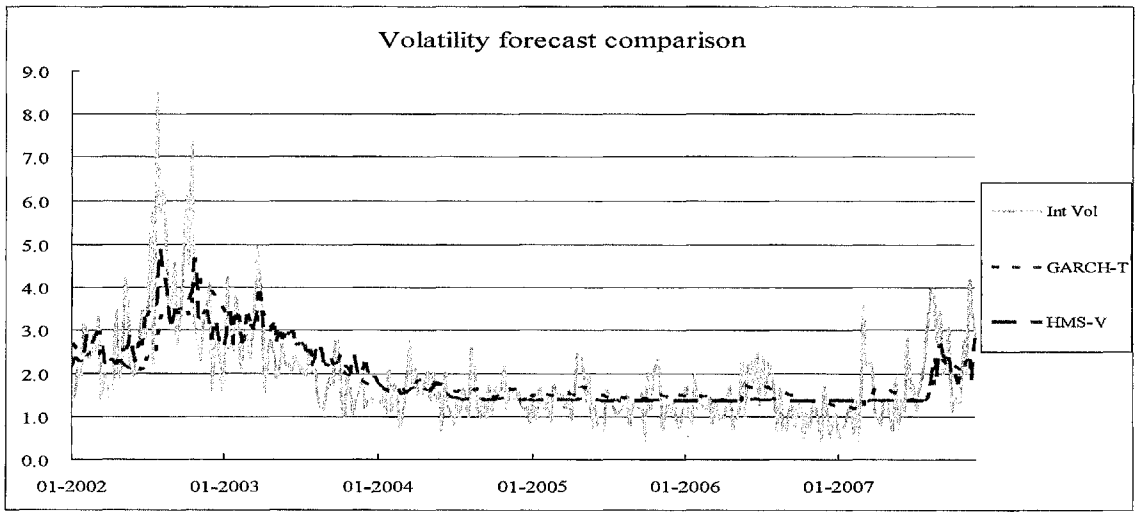


Figure 28

Volatility forecasting $E\left[\sigma_t \middle| \psi_{t-1}, \hat{\theta}\right]$ *using different methods from Dec/26/2001 to Nov/21/2007*
The solid line plots the weekly "integrated volatility", the dashed line plots the forecasted volatility by the proposed HMS-V mode, and the dotted line plots that of the GARCH-T model.

It can be seen that HMS-V suggests that the mean is lower at the turbulent state and higher at the tranquil state. On one hand, at the tranquil state, the expected return of y_t should be higher but not unrealistically high. However, on the other hand, GARCH models suggest a dramatically different situation: y_t is negatively correlated with y_{t-1} (see *Table 13*). To closely compare HMS-V with GARCH-T, two chosen periods are examined in details: Jan/ 07/1987 to Dec/28/1988 and Jan/03/2007 to Nov/21/2007.

The first chosen episode

Jan/07/1987 to Dec/28/1988 includes the infamous Black Monday, 1987. The S&P 500 was relatively stable before Oct/1987, dropped by 22.9% on Oct/19/1987, and remained very turbulent for the subsequent few months. In the wake of the Black Monday, the Fed and other central banks pumped liquidity into the system to prevent further downdrift. The integrated weekly volatility from Oct/14/1987 to Oct/21/1987 (refer to equation (5.17)) was 25.71%, which is much higher than the average of integrated volatility of

1.89% from Jan/05/1983 to Nov/21/2007. Partially due to the effort by the central banks, the market calmed down from Mar/1988 onwards and the average integrated weekly volatility from Mar/02/1988 to April/20/1988 decreased to 2.3%.

Table 14 and *Figure 29* compare the differences of the volatility forecasted by HMS-V and GARCH-T.[61]. The first impression is that the forecasted volatility by HMS-V peaks in Oct/28/1987 (the immediate week following Black Monday) whereas the forecasted volatility by GARCH-T peaks at Nov/04/1987 (two weeks after Black Monday). Therefore, it is clear that HMS-V is more capable of capturing sudden spikes in volatility.

According to HMS-V, at the shock of Black Monday, the state probability $\Pr(\Delta_t = 1|\psi_{t-1})$ immediately lowered to 0.55%, indicating a high likelihood of residing at the turbulent state. Consequently, the forecasted volatility (the week ended on Oct/28/1987) rose to 6.44% immediately after the shock, which is quite a dramatic increase from 2.32% on Oct/7/1987. As shown in *Figure 29*, a few weeks after Black Monday, the market calmed down and the forecasted volatility by HMS-V eased to 2.95% on Feb/17/1988. In this case, at the outset of the turbulent state, the HMS-V obtained an immediate higher volatility forecast through its two processes: (i) the underlying process switched to the turbulent state, (ii) the volatility link function at the turbulent state absorbed the shock with great speed. The joint effect of the two processes describe the demonstrated spike in volatility.

The results from the GARCH-T differ from that of HMS-V. It reacts linearly and too

[61] Given the estimated parameters, we proxy the forecasting process in reality[61], that is, forecasting σ_t based on information ψ_{t-1}.

sluggishly. According to GARCH-T, the forecasted volatility (the week ended on Oct/28/1987) only rose to 4.96% after Black Monday. Compared with the estimation from HMS-V, this is a relatively small increment from 2.28% on Oct/7/1987. The forecasted volatility by GARCH-T only peaks on Nov/04/1987, while the forecasted volatility by HMS-V peaks immediately on Oct/28/1987. In addition, after a few months following Black Monday, the forecasted volatility by GARCH-T eased only to 3.89% on Feb/17/1988, which is persistently and substantially greater than the contemporary integrated volatility. Therefore, when structure breaks take place, a one-size-fit-all linear model is inferior to the proposed model. Note that GARCH-T reacts to the shock through only a single process: the volatility link function absorbs the shock with high speed and is hence less capable than that dual processes of HMS-V.

Therefore, not only is HMS-V is more capable of forecasting volatility, the forecasted return is also more sensible. *Table 14* and *Figure 31* summarize the mean estimated by HMS-V and GARCH-T from Oct/07/1987 to Mar/02/1988. On the shock of Black Monday, according to HMS-V, the forecasted mean was low for a few months, which is not suggested by GARCH-T. On the shock of Black Monday, the expected return shot up to 1.67% and the expectation of the mean for the period Oct/07/1987 to Mar/02/1988 fluctuated widely.

Date	Return %	Int Vol %	HMS-V			GARCH-T	
			$\Pr(\Delta_t = 1 \mid \psi_{t-1})$ HMS-V	$E(\sigma_t \mid \psi_{t-1})$ HMS-V %	$E(\mu_t \mid \psi_{t-1})$ HMS-V %	$\widehat{\sigma_t}$ GARCH-T %	$\widehat{\mu_t}$ GARCH-T %
7-Oct-1987	-1.03	3.23	21.45%	2.32	0.05	2.21	0.301
14-Oct-1987	-4.27	3.86	46.94%	2.44	0.14	2.16	0.401
21-Oct-1987	-16.66	25.71	1.71%	3.22	-0.02	2.41	0.665
28-Oct-1987	-10.22	9.82	0.55%	6.44	-0.03	4.96	1.674
4-Nov-1987	6.51	6.16	0.33%	4.64	-0.03	5.64	1.150
11-Nov-1987	-2.88	4.52	0.25%	4.09	-0.03	5.59	-0.212
18-Nov-1987	1.50	3.50	0.35%	3.88	-0.03	5.42	0.552
25-Nov-1987	-0.59	2.96	0.41%	3.00	-0.03	5.22	0.196
2-Dec-1987	-4.46	4.64	0.31%	3.31	-0.03	5.03	0.366
9-Dec-1987	2.30	5.26	0.23%	3.87	-0.03	4.99	0.681
16-Dec-1987	3.77	4.31	0.23%	3.93	-0.03	4.82	0.130
23-Dec-1987	2.03	3.51	0.30%	3.56	-0.03	4.73	0.010
30-Dec-1987	-2.12	2.98	0.43%	3.11	-0.03	4.58	0.153
6-Jan-1988	4.35	3.69	0.40%	2.96	-0.03	4.44	0.490
13-Jan-1988	-5.18	7.30	0.28%	3.64	-0.03	4.39	-0.037
20-Jan-1988	-1.30	3.82	0.32%	3.95	-0.03	4.42	0.740
27-Jan-1988	2.74	2.86	0.36%	3.33	-0.03	4.29	0.424
3-Feb-1988	1.13	2.65	0.52%	3.36	-0.03	4.17	0.094
10-Feb-1988	1.75	2.38	0.72%	2.93	-0.02	4.02	0.226
17-Feb-1988	0.99	1.13	1.15%	2.95	-0.02	3.89	0.175
24-Feb-1988	1.99	2.17	1.50%	2.58	-0.02	3.75	0.237
2-Mar-1988	1.33	2.35	2.15%	2.93	-0.02	3.64	0.155

Table 14

Comparison between HMS-V and GARCH-T forecasted volatility $E\left[\sigma_t \middle| \psi_{t-1}\right]$ *and mean* $E\left[\mu_t \middle| \psi_{t-1}\right]$ *using S&P 500 weekly data from Oct/07/1987 to Mar/02/1988.*

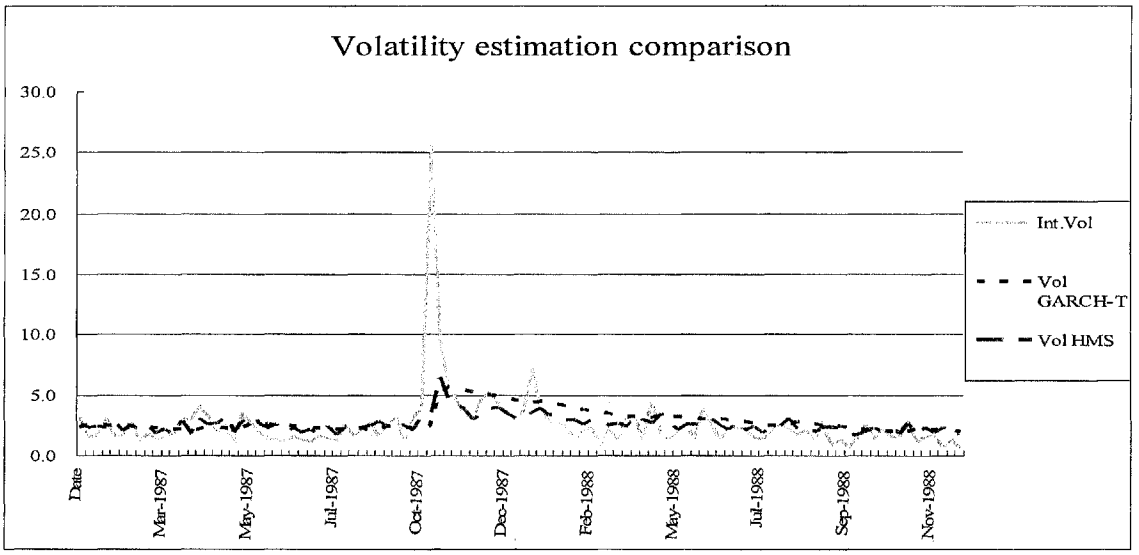


Figure 29

Forecasting $E\left[\sigma_t \middle| \psi_{t-1}\right]$ *by HMS-V and GARCH-T from Oct/07/1987 to Mar/02/1988.*

The solid line plots the weekly "integrated volatility", the dashed line plots the HMS-V and the dotted line plots the GARCH-T model.

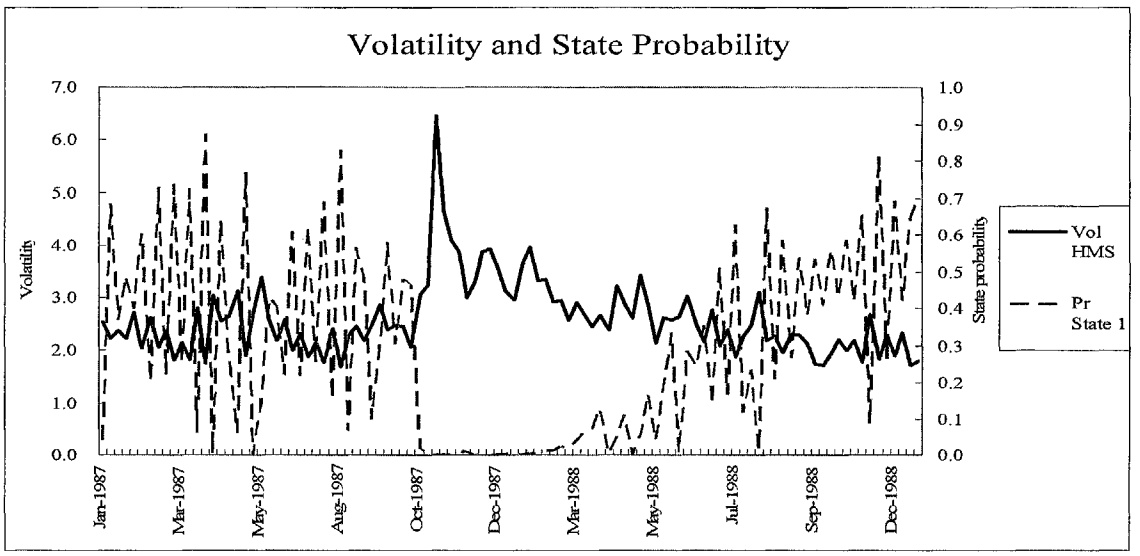


Figure 30

Forecasting $E\left[\sigma_t \middle| \psi_{t-1}\right]$ *by HMS-V model from Oct/07/1987 to Mar/02/1988.*

The solid line plots $E\left[\sigma_t \middle| \psi_{t-1}\right]$ *(left scale of the y-axis) and the dashed line plots* $\Pr\left(\Delta_t = 1 \middle| \psi_{t-1}\right)$ *(right scale of the y-axis).*

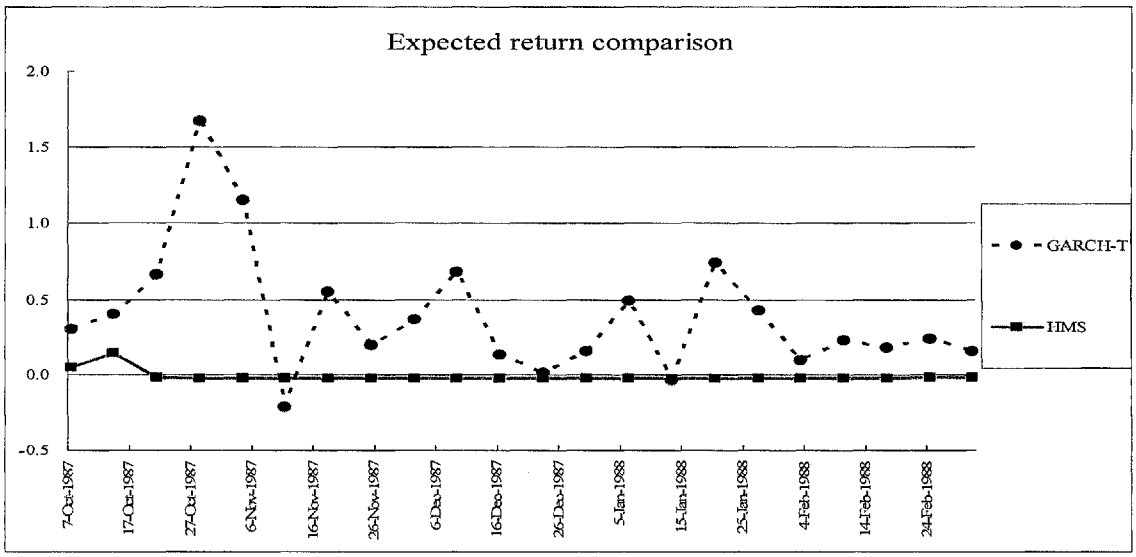


Figure 31

Forecasting mean $E\left[\mu_t \middle| \psi_{t-1}\right]$ *with the proposed HMS-V and GARCH-T from Oct/07/1987 to Mar/02/1988. The solid line plots HMS-V with the dotted line plotting GARCH-T.*

The second chosen episode

The second period of examination is from Jan/03/2007 to Nov/21/2007, which includes the period of the US subprime implosion which may lead the US into recession in 2008 or 2009. Before mid Jul/2007, the S&P 500 was relatively stable and showed little sign of stress but the price fluctuated and dropped from 1553 on Jul/19/2007 to 1440 on Nov/23/2007. The first wave of the subprime crunch hit in early Aug/2007 as the weekly integrated volatility shot up from 1.9% in mid Jul/2007 to 3.9% in early Aug/2007. The S&P 500 rebounded from Sep/2007 to Oct/2007 on the hope that the credit crunch was short-lived, but it plummeted again in Nov/2007 as the credit crunch spread to the rest of banking industry and the real economy. Some unconventional measures have been taken by the U.S. government agencies to avoid the systematic collapse of the financial system.

As shown in *Table 15* and *Figure 33*, according to HMS-V, prior to the subprime implosion, on Jul/25/2007, the forecasted tranquil state probability $\Pr\left(\Delta_t = 1 \middle| \psi_{t-1}\right)$ was

as high as 99.0%. Immediately after the shock of early August, $\Pr(\Delta_t = 1 | \psi_{t-1})$ drops to 60.3%. As a result, on Aug/15/2007, the forecasted volatility immediately shoots up to 2.38%.This compares favourably to the linear and slow reaction of GARCH-T, whose forecasted volatility at 15/Aug/2007 remained at 1.72%.

Table 15 and *Figure 34* also summarize the forecasted mean by HMS-V and GARCH-T from Jul/25/2007 to Nov/21/2007. HMS-V suggests that from 15/Aug/2007 onwards, the market was relatively bearish and the mean was below 0.30%. GARCH-T suggests otherwise, for example, the forecasted mean for $\widehat{y}_t$ shot up as high as 0.715% on Nov/14/2007. This could be due to the AR-GARCH-T model attempting to describe the nonlinear pattern in mean with a linear model specification, which is also the case for volatility.

Date	Return %	Int Vol %	HMS-V			GARCH-T	
			$\Pr(\Delta_t = 1 \mid \psi_{t-1})$ HMS-V	$E(\sigma_t \mid \psi_{t-1})$ HMS-V %	$E(\mu_t \mid \psi_{t-1})$ HMS-V %	$\widehat{\sigma_t}$ GARCH-T %	$\widehat{\mu_t}$ GARCH-T %
18-Jul-2007	1.79	1.93	99.5%	1.37	0.32	1.34	0.35
25-Jul-2007	-1.83	2.48	99.0%	1.38	0.32	1.36	0.17
1-Aug-2007	-3.50	3.37	96.3%	1.45	0.31	1.43	0.47
8-Aug-2007	2.14	3.93	90.4%	1.67	0.29	1.72	0.60
15-Aug-2007	-6.25	3.79	60.3%	2.38	0.19	1.72	0.14
22-Aug-2007	4.00	2.71	87.2%	1.82	0.28	2.33	0.83
29-Aug-2007	-0.02	3.52	54.3%	2.68	0.16	2.39	-0.01
5-Sep-2007	0.58	1.96	33.8%	2.46	0.09	2.31	0.32
12-Sep-2007	-0.05	2.22	50.4%	2.20	0.15	2.23	0.27
19-Sep-2007	3.83	3.10	34.2%	2.37	0.09	2.16	0.32
26-Sep-2007	-0.24	1.11	79.8%	1.89	0.25	2.28	0.01
3-Oct-2007	0.92	1.48	31.2%	2.00	0.08	2.20	0.34
10-Oct-2007	1.48	1.32	66.0%	1.75	0.21	2.14	0.24
17-Oct-2007	-1.37	1.29	45.2%	1.96	0.13	2.09	0.20
24-Oct-2007	-1.66	2.78	60.3%	2.09	0.19	2.07	0.43
31-Oct-2007	2.19	1.97	49.3%	2.31	0.15	2.07	0.45
7-Nov-2007	-4.88	4.21	45.6%	2.34	0.13	2.05	0.14
14-Nov-2007	-0.34	3.44	82.2%	1.85	0.26	2.36	0.71
21-Nov-2007	-3.73	2.81	12.6%	2.83	0.02	2.30	0.35

Table 15

Comparison between HMS-V and GARCH-T forecasting volatility $E\left[\sigma_t \middle| \psi_{t-1}\right]$ *and mean* $E\left[\mu_t \middle| \psi_{t-1}\right]$, *using S&P 500 weekly data from Jul/18/2007 to Nov/21/2007.*

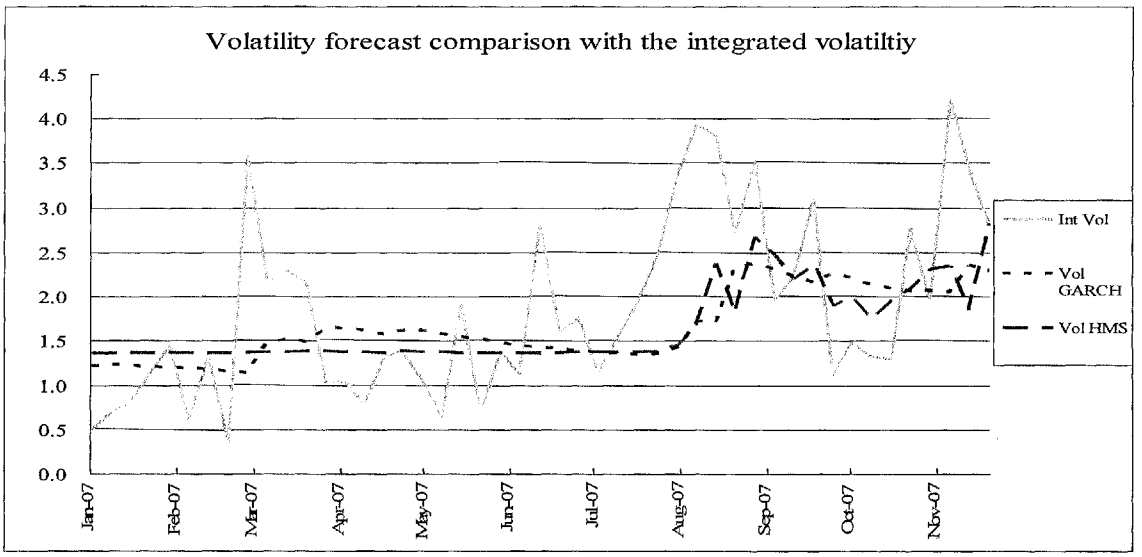


Figure 32

Forecasting volatility $E\left[\sigma_t \middle| \psi_{t-1}\right]$ *from Jul/25/2007 to Nov/21/2007. The solid line plots the "integrated volatility", the dashed line plots HMS-V and the dotted line plots GARCH-T.*

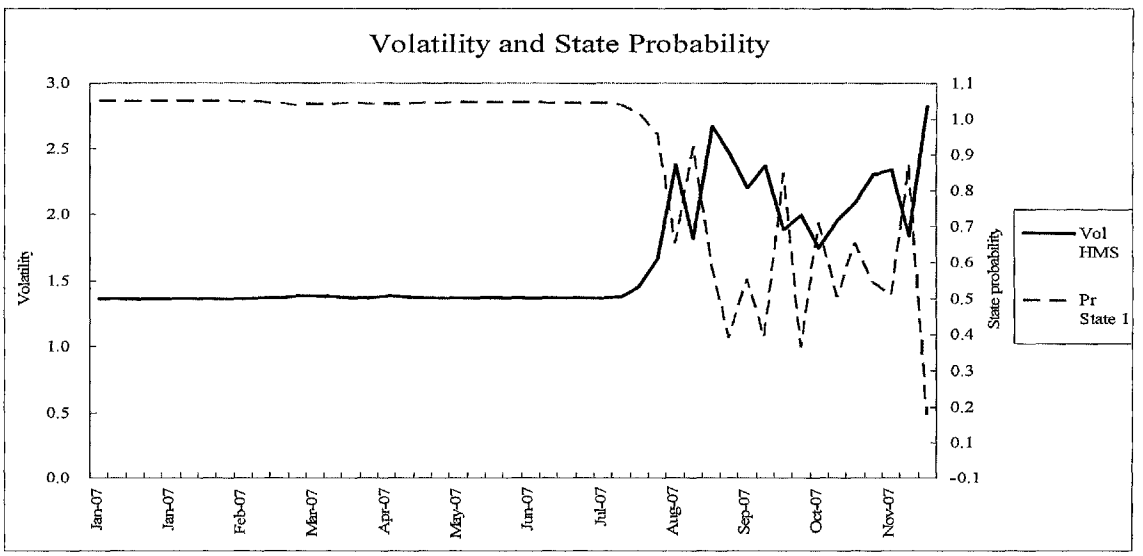


Figure 33

Forecasting $E\left[\sigma_t \middle| \psi_{t-1}\right]$ *using the HMS-V model from Jul/25/2007 to Nov/21/2007. The solid line plots* $E\left[\sigma_t \middle| \psi_{t-1}\right]$ *(left scale of the y-axis), with the dashed line plotting* $\Pr\left(\Delta_t = 1 \middle| \psi_{t-1}\right)$ *(right scale of the y-axis).*

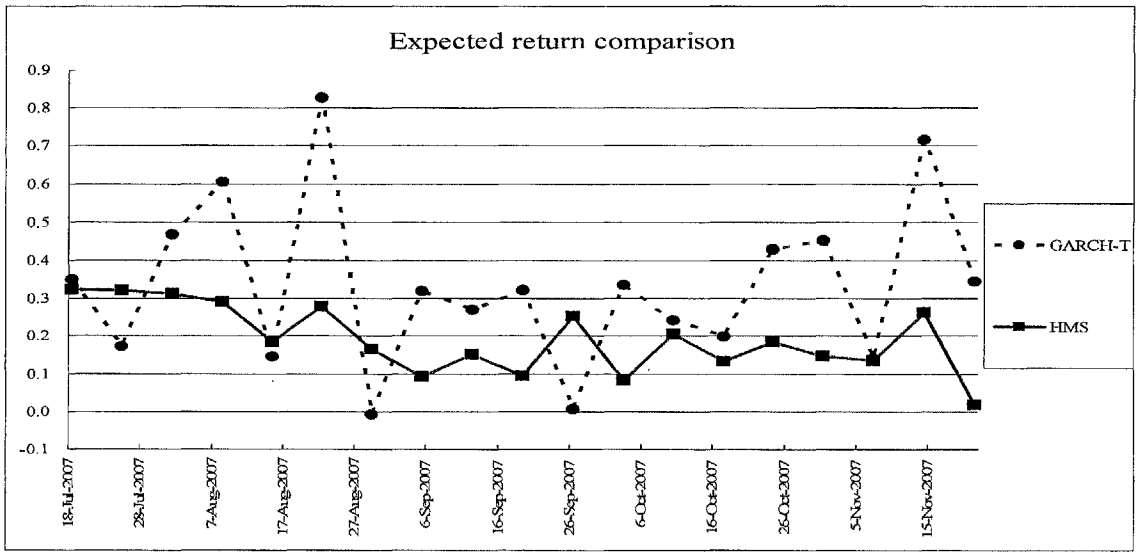


Figure 34

Forecasting mean $E\left[\mu_t \middle| \psi_{t-1}\right]$ *by HMS-V and GARCH-T from Jul/25/2007 to Nov/21/2007. The solid line plots HMS-V, whereas the dotted line plots GARCH-T.*

Therefore, the performance of the proposed HMS-V model during Black Monday in 1987 and the Credit Crunch in 2007 illustrates how regime-switching helps to explain the sudden jump of volatility.

5.6.2. Comparison with MS-GARCH

For a fair comparison, the same time series is used to fit the MS-GARCH Klaassen (2002)[62] model. The MLE suggests:

$$\underset{(0.054)}{\widehat{\mu_{\langle 1\rangle}} = 0.33}; \qquad \underset{(0.301)}{\widehat{\mu_{\langle 2\rangle}} = -3.69};$$

$$\underset{(0.052)}{\sigma^2_{\langle 1\rangle,t} = 0.14} + \underset{(0.020)}{0.034 \bullet \varepsilon^2_{t-1}} + \underset{(0.026)}{0.834 \bullet \sigma^2_{t-1}}$$

[62] Computational code and results for the thesis can be made available upon request. The computational code for MS-GARCH Klaassen is adapted from the original code written by Marcucci (2005). Marcucci's (2005) used the MS-GARCH Klaassen model to study the S&P 100 stock index. Marcucci's (2005) article and MatLab code can be found through the following url: http://www.bepress.com/snde/vol9/iss4/art6/
1,000 randomly generated starting values are used to obtain the result. This is because when compared to the proposed model, the MS-GARCH is more sensitive to the choice of starting values.

$$\sigma^2_{\langle 2\rangle,t} = \underset{(1.04)}{0.00} + \underset{(0.020)}{0.406 \bullet \varepsilon^2_{t-1}} + \underset{(0.411)}{0.594 \bullet \sigma^2_{t-1}} \qquad (5.30)$$

$$\widehat{p_{\langle 11\rangle}} = \underset{(0.008)}{0.960}; \qquad \widehat{p_{\langle 22\rangle}} = \underset{(0.008)}{0.054};$$

$$LLF = -2063.4$$

Note that according to equation (5.30), for the volatility link function at the second regime, two boundary conditions are hit and it is essentially an Exponentially Weighted Moving Average (EWMA) process[63]. Equation (5.30) suggests that:-

- State 2 has negative expected return, i.e., bearish; State 1 is bullish.
- State 1 (State 2) is far less (more) sensitive to the latest shock, ε^2_{t-1}. Thus State 1 can be referred to as the "jogging state", and State 2 as the "sprinting state"[64].
- In some cases, $\sigma_{\langle 2\rangle,t} \geq \sigma_{\langle 1\rangle,t}$, and in others, $\sigma_{\langle 2\rangle,t} < \sigma_{\langle 1\rangle,t}$.
- As a result, higher volatility is sometimes associated with bullish markets and sometimes with bearish markets. It is difficult to find convincing economic implications.

Again, for comparison fairness, $E\left[\sigma_t \middle| \psi_{t-1}\right]$ is examined for both models[65]. The two chosen periods of data are again: Jan/07/1987 to Dec/28/1988, and Jan/03/2007 to

[63] For the volatility link function of Gray/Klaassen MS-GARCH model, it is required that
$\sigma^2_{\langle 2\rangle,t} = \beta_{\langle 2\rangle,0} + \beta_{\langle 2\rangle,1} \bullet \varepsilon^2_{\langle 2\rangle,t-1} + \beta_{\langle 2\rangle,2} \bullet \sigma^2_{t-1}$
$\beta_{\langle 2\rangle,0} \geq 0; \beta_{\langle 2\rangle,1} + \beta_{\langle 2\rangle,2} \leq 1$

[64] On one hand, if the time series transits from a tranquil state to a turbulent state, then the volatility at State 2 will shoot up faster than State 1, and $\sigma_{\langle 2\rangle,t} > \sigma_{\langle 1\rangle,t}$. On the other hand, if the time series transits from a turbulent state to a tranquil state, then the volatility at State 2 will also calm down faster than State 1, that is, $\sigma_{\langle 2\rangle,t} < \sigma_{\langle 1\rangle,t}$.

[65] Note that to avoid confusion, the comparison using $E\left[\sigma_t \middle| \psi_t\right]$ or $E\left[\sigma_t \middle| \psi_T\right]$ is not presented.

Nov/21/2007.

The first chosen episode

Both models absorb the sudden shock induced by Black Monday with high speed, but through different mechanisms (see *Table 16*).

- For HMS-V, the forecasted turbulent state probability increases from 53.1% on Oct/14/1987 to 99.5% on Oct/28/1987. As before, at the outset of the turbulent period, the shock of Black Monday is absorbed by two engines: (i) greater turbulent state probability; (ii) the volatility link function at the turbulent state reacts to the turbulence. Note that $E\left[\widehat{\sigma_t}\middle|\psi_{t-1}\right]$ peaks on Oct/28/1987 (the week immediately following Black Monday).
- The MS-GARCH model of Gray/Klaassen (2002) suggests otherwise. The forecasted turbulent state probability only fluctuates within a small range (about from 4% to 6%). Although immediately after Black Monday, $E\left[\widehat{\sigma_{(2),t}}\middle|\psi_{t-1}\right]$ peaks at 11.05, $E\left[\widehat{\sigma_t}\middle|\psi_{t-1}\right]$ only peaks two weeks later. Therefore, at the outset of the turbulent period, the shock of Black Monday is absorbed by only one engine: the volatility link function of the second state, which describes the INTER-state dynamics from the tranquil state to turbulent state, and the INTRA-state dynamics from the turbulent state to the turbulent state.

Date	Return %	Int Vol %	HMS-V				MS-GARCH (1,1)			
			$\Pr(\Delta_t = 1 \mid \psi_{t-1})$	$\widehat{\sigma}_{\langle 1\rangle,t}$ %	$\widehat{\sigma}_{\langle 2\rangle,t}$ %	$E(\sigma_t \mid \psi_{t-1})$ %	$\Pr(\Delta_t = 1 \mid \psi_{t-1})$	$\widehat{\sigma}_{\langle 1\rangle,t}$ %	$\widehat{\sigma}_{\langle 2\rangle,t}$ %	$E(\sigma_t \mid \psi_{t-1})$ %
7-Oct-1987	-1.03	3.23	21.4%	1.36	2.52	2.32	95.9%	2.11	1.76	2.10
14-Oct-1987	-4.27	3.86	46.9%	1.36	3.09	2.44	95.9%	2.04	1.86	2.03
21-Oct-1987	-16.66	25.71	1.7%	1.36	3.24	3.22	95.5%	2.68	3.55	2.72
28-Oct-1987	-10.22	9.82	0.5%	1.36	6.46	6.44	94.6%	4.51	11.05	5.08
4-Nov-1987	6.51	6.16	0.3%	1.36	4.65	4.64	95.6%	6.53	8.64	6.63
11-Nov-1987	-2.88	4.52	0.3%	1.36	4.10	4.09	95.9%	6.17	6.53	6.19
18-Nov-1987	1.50	3.50	0.4%	1.36	3.89	3.88	95.9%	5.74	5.20	5.72
25-Nov-1987	-0.59	2.96	0.4%	1.36	3.01	3.00	95.9%	5.28	4.53	5.26
2-Dec-1987	-4.46	4.64	0.3%	1.36	3.31	3.31	95.9%	4.87	4.13	4.84
9-Dec-1987	2.30	5.26	0.2%	1.36	3.87	3.87	95.9%	4.60	4.82	4.61
16-Dec-1987	3.77	4.31	0.2%	1.36	3.93	3.93	95.9%	4.27	3.84	4.26
23-Dec-1987	2.03	3.51	0.3%	1.36	3.56	3.56	95.9%	3.99	4.03	3.99
30-Dec-1987	-2.12	2.98	0.4%	1.36	3.12	3.11	95.9%	3.71	3.33	3.70
6-Jan-1988	4.35	3.69	0.4%	1.36	2.96	2.96	95.9%	3.51	3.28	3.50
13-Jan-1988	-5.18	7.30	0.3%	1.36	3.64	3.64	95.9%	3.33	3.81	3.35
20-Jan-1988	-1.30	3.82	0.3%	1.36	3.96	3.95	95.8%	3.48	4.45	3.52
27-Jan-1988	2.74	2.86	0.4%	1.36	3.33	3.33	95.9%	3.30	2.94	3.29
3-Feb-1988	1.13	2.65	0.5%	1.36	3.37	3.36	95.9%	3.09	3.04	3.09
10-Feb-1988	1.75	2.38	0.7%	1.36	2.93	2.93	95.9%	2.88	2.49	2.87
17-Feb-1988	0.99	1.13	1.1%	1.36	2.96	2.95	95.9%	2.68	2.45	2.67
24-Feb-1988	1.99	2.17	1.5%	1.36	2.60	2.58	95.9%	2.50	2.15	2.49
2-Mar-1988	1.33	2.35	2.1%	1.36	2.95	2.93	96.0%	2.35	2.26	2.34

Table 16

Comparison between HMS-V and MS-GARCH, using S&P 500 weekly data from Oct/07/1987 to Mar/02/1988.

The second chosen episode

The second episode is from Jan/03/2007 to Nov/21/2007, which includes the beginning of the US subprime implosion (see *Table 17*). As of Jul/18/2007, the market had been

tranquil for a long duration. As HMS-V suggests, the volatility at the turbulence state is relatively independent of the recent tranquility, and it remained as high as $\widehat{\sigma_{\langle 2\rangle,t}} = 2.35$, hence the turbulence of distant history is always been respected. Gray/Klaassen's model suggests otherwise since both $\widehat{\sigma_{\langle 1\rangle,t}} = 1.03$ and $\widehat{\sigma_{\langle 2\rangle,t}} = 0.88$ are very low. The memory of turbulence has been under-represented.

Date	Return %	Int Vol %	HMS-V				MS-GARCH (1,1)			
			$\Pr(\Delta_t = 1 \mid \psi_{t-1})$	$\widehat{\sigma_{\langle 1\rangle,t}}$ %	$\widehat{\sigma_{\langle 2\rangle,t}}$ %	$E(\sigma_t \mid \psi_{t-1})$ %	$\Pr(\Delta_t = 1 \mid \psi_{t-1})$	$\widehat{\sigma_{\langle 1\rangle,t}}$ %	$\widehat{\sigma_{\langle 2\rangle,t}}$ %	$E(\sigma_t \mid \psi_{t-1})$ %
18-Jul-2007	1.79	1.93	99.5%	1.36	2.35	1.37	96.0%	1.03	0.88	1.02
25-Jul-2007	-1.83	2.48	99.0%	1.36	2.58	1.38	96.0%	1.06	1.30	1.07
1-Aug-2007	-3.50	3.37	96.3%	1.36	2.89	1.45	95.8%	1.54	1.83	1.55
8-Aug-2007	2.14	3.93	90.4%	1.36	3.39	1.67	95.4%	2.50	3.09	2.53
15-Aug-2007	-6.25	3.79	60.3%	1.36	3.37	2.38	95.9%	2.38	2.34	2.37
22-Aug-2007	4.00	2.71	87.2%	1.36	3.61	1.82	95.3%	3.09	4.73	3.19
29-Aug-2007	-0.02	3.52	54.3%	1.36	3.66	2.68	95.9%	3.00	3.48	3.02
5-Sep-2007	0.58	1.96	33.8%	1.36	2.86	2.46	95.9%	2.83	2.38	2.81
12-Sep-2007	-0.05	2.22	50.4%	1.36	2.80	2.20	95.9%	2.63	2.22	2.62
19-Sep-2007	3.83	3.10	34.2%	1.36	2.74	2.37	95.9%	2.46	2.07	2.45
26-Sep-2007	-0.24	1.11	79.8%	1.36	3.19	1.89	96.0%	2.38	3.01	2.41
3-Oct-2007	0.92	1.48	31.2%	1.36	2.22	2.00	95.9%	2.27	1.92	2.26
10-Oct-2007	1.48	1.32	66.0%	1.36	2.31	1.75	95.9%	2.12	1.82	2.11
17-Oct-2007	-1.37	1.29	45.2%	1.36	2.33	1.96	96.0%	1.99	1.84	1.98
24-Oct-2007	-1.66	2.78	60.3%	1.36	2.85	2.09	95.9%	1.97	1.91	1.97
31-Oct-2007	2.19	1.97	49.3%	1.36	2.94	2.31	95.9%	2.00	2.03	2.00
7-Nov-2007	-4.88	4.21	45.6%	1.36	2.91	2.34	96.0%	1.91	2.01	1.91
14-Nov-2007	-0.34	3.44	82.2%	1.36	3.24	1.85	95.2%	2.74	3.85	2.81
21-Nov-2007	-3.73	2.81	12.6%	1.36	2.98	2.83	95.9%	2.63	2.24	2.61

Table 17

Comparison between HMS-V and MS-GARCH, forecasting volatility based on information ψ_{t-1}, using S&P 500 weekly data from Jul/18/2007 to Nov/21/2007.

5.6.3. Summary

As discussed in chapter 1, the proposed model aims to address empirical observations and address the oversights found in the literature.

(Obs. 1) ***Clustering:***.

- HMS-V

  Volatility is clustered through two channels: (i) the first (also primary) channel is through the clustering of Markov states. For example, S&P 500 from early 2004 to mid 2007 had a low volatility and was explained by the clustering at the tranquil state; (ii) the second channel is the volatility link function, for example, the volatility at the turbulent state has higher correlation with the lagged intraweek range.

- GARCH

  The only source of volatility clustering is the ARCH (∞) volatility link function.

- MS-GARCH

  As equation (5.30) suggests, for the MS-GARCH model of Gray/Klaassen, $\widehat{p_{\langle 11\rangle}} + \widehat{p_{\langle 22\rangle}} = 1.014 \approx 1$, therefore, the model resembles a single regime GARCH-MixN. As a result, volatility clustering is primarily explained by the ARCH (∞) volatility link functions and the clustering of Markov regimes plays a much weaker role.

(Obs. 2) ***Structure breaks:***.

- HMS-V

  A structure break takes place when the Markov state switches from one state to the other. Using Regression 05 as the example, volatility was low from mid 1992 to mid 1996, and from early 2004 to mid 2007, hence it is likely that the

contemporary state is the tranquil Markov state. Volatility becomes high from 1998 to 2003, hence it is more likely that the contemporary Markov state is the turbulent state. Also note that when the time series remains in the turbulent state, volatility is more correlated with the lagged range, while the tranquil state suggests otherwise.

- GARCH

  GARCH uses a one-size-fits-all parameter to describe the dynamics and is hence unable to capture non-linear structure changes.

- MS-GARCH

  Volatility clustering is primarily explained by the ARCH (∞) volatility link functions and the clustering of regimes plays a much weaker role. Hence, it barely explains the structure breaks.

(Obs. 3) ***Reoccurrence of distant history:***

- HMS-V

  According to *Table 17* on Jul/18/2007, $\widehat{\sigma_{\langle 1\rangle}} = 1.36$ was low, which reflects the recent history; $\widehat{\sigma_{\langle 2\rangle}} = 2.35$ was still high, which keeps the memory of the distant history. The reoccurrence of distant history has therefore not been ruled out.

- GARCH

  On Jul/18/2007, the single regime GARCH-T model suggests that $\hat{\sigma} = 1.34$, therefore the memory of the distant turbulent history has faded away.

- MS-GARCH

  On Jul/18/2007, $\widehat{\sigma_{\langle 1\rangle}} = 1.03$ and $\widehat{\sigma_{\langle 2\rangle}} = 0.88$, which are both low. The reoccurrence of distant turbulences has clearly not been reflected.

(Obs. 4) ***Long memory:***.

- HMS-V

Volatility exhibits distinctively differing patterns at different states, and volatility clustering has primarily been explained by Markov-state clustering. The long memory is a consequence of structure breaks.

- GARCH

  GARCH addresses long memory by a near unity persistence, which is probably spuriously high.

- MS-GARCH

  Long memory is primarily explained by the ARCH (∞) volatility link function, and the clustering of regimes plays a much weaker role.

(Obs. 5) ***Interplay between volatility and expected return:***.

- HMS-V

  At the volatile state, the expected return is lower, while the opposite is true for the turbulent state. This finding is consistent with the reviewed literature and the economic implication given later in chapter 6.

- GARCH

  GARCH hardly suggests the clustering of bearish and bullish market.

- MS-GARCH

  As shown in *Table 16* and *Table 17*, for the Gray/Klaassen model, it may be difficult to relate the bearish / bullish markets to volatility

(Obs. 6) ***Leptokurtosis and asymmetry:***

All three models are able to model Leptokurtosis and the asymmetry of time series.

(Obs. 7) ***Asymmetric cycle and thus lasting tranquility:***.

- HMS-V

  For the forecasted state 1 probability $\left(\widehat{\Pr(\Delta_t = 1)} \middle| \psi_{t-1}\right)$, of 1297 observations,

only 499 observations were less than 0.5. Therefore, more observations are likely to reside at the tranquil state, which is consistent with the asymmetric business cycles discussed next in chapter 6.

- GARCH

  The same persistence level is imposed on all observations, hence it is not known if the tranquility / turbulence lasts longer.

- MS-GARCH

  As shown in equation (5.30), regardless of the current regime, the forecasted regime is similar.

(Obs. 8) ***Time varying sensitivity to recent shocks:***.

- HMS-V

  As suggested by equation (4.61), at the tranquil state, volatility is less sensitive to recent shocks[66]; the opposite is true for the turbulent state. Hence the sensitivity is also clustered (see *figure 18, figure 19 and figure 20* for validation using raw data)

- GARCH

  It is unable to describe the change of sensitivity.

- MS-GARCH

  As shown in *table 16* and *table 17*, if the volatilities at the two states are less distinctly different, the change of sensitivity is significant.

(Oversight. 1) ***Same volatility link function for both INTER-state and INTRA-state dynamics:***.

- HMS-V

  Using Regression 05 as an example, the volatility link function only caters for the

[66] This is further justified by the LR test between Regression 02 and Regression 03.

INTRA-state dynamics, with the INTER-state "jump" being explained by the switching of the regimes.

- GARCH

  This oversight is not addressed by the single-regime GARCH model.

- MS-GARCH

  According to equation (5.30), the volatility link functions manage both the INTER-state and INTRA-state dynamics. The regime switching only indicates the switching of coefficients, instead of the underlying process.

(Oversight. 2) ***Less tractable dynamics:***.

- HMS-V

  The two states are more independent, and the parallel structure makes the dynamics of the Markov-switching process more tractable.

- GARCH

  A single regime GARCH is most tractable.

- MS-GARCH

  The dynamics of the MS-GARCH suggest the intricately complex relationship between the two states. Tractability is reduced and there are more ambiguous economic implications.

(Oversight. 3) ***Possibly excessive adaptation to "less valuable" signals:***.

- HMS-V

  Even when one regime adapts excessively to the recent history, the other regime remains relatively independent. For example, the memory of distant turbulent history is maintained even though the S&P 500 had been tranquil from early 2004

to mid 2007.

- GARCH

  This oversight cannot be addressed by a single regime model.

- MS-GARCH

  The non-parallel structure suggests a less independent dynamics. As shown in *Table 16* and *Table 17*, it is difficult to tell which state is turbulent and which is tranquil. In this case, the economic implication can be confusing, in particular the interplay between the volatility and expected return.

(Oversight. 4) ***Passive adaptation and hence missing memory:***

- HMS-V

  As of Jul/18/2007, the market had been tranquil for a long duration. For the HMS-V model, the volatility at the tranquil state adapted aggressively to recent history, $\widehat{\sigma_{\langle 1\rangle,t}} = 1.36$; but the volatility at the turbulence state was relatively independent and remained as high as $\widehat{\sigma_{\langle 2\rangle,t}} = 2.35$.

- GARCH

  Using GARCH-T as an example, as of Jul/18/2007, $\widehat{\sigma_t} = 1.34$ is low, since it excessively adapted to the recent tranquility.

- MS-GARCH

  The estimated MS-GARCH does no better in memorizing the distant but still important history. For example, as shown in *Table 17*, as of Jul/18/2007, for the MS-GARCH Gray/Klaassen model, $\widehat{\sigma_{\langle 1\rangle,t}} = 1.03$, $\widehat{\sigma_{\langle 2\rangle,t}} = 0.88$. Hence, both states have been adapting aggressively to the recent tranquility.

(Oversight. 5) ***Possibly un-identifiable states:.***

- HMS-V

  For example, for Regression 02, despite the market being volatile or tranquil, volatility at the second state is always significantly greater than the first. It is easier to identify the two different states.

- GARCH

  This is not applicable to a single regime GARCH model.

- MS-GARCH

  Firstly, the model identifiability condition is lacking in the literature. Secondly, according to table 16 and table 17, it is difficult to tell which state is turbulent and which is tranquil.

(Oversight. 6) ***Long memory not addressed by regime-switching:***

- HMS-V

  As shown by *Figure 23* and *Figure 25*, the Markov state exhibits significant autocorrelation and the long memory can be explained by the periodic shift of regimes.

- GARCH

  The long memory is explained by a near unity persistence.

- MS-GARCH

  As indicated by equation (5.30), volatility clustering and hence long memory are largely explained by the ARCH(∞) volatility link function, but not by the clustering and switching of regimes.

(Oversight. 7) ***Possibly inconsistent economic implication of regime switching:***

- HMS-V

  For Regression 02 to Regression 05, there are distinct economic implications of the two Markov states: one is turbulent with bearish return and the other is tranquil with the bullish return. Such a distinct implication is consistent with the literature reviewed in chapter 3 and the following detailed discussions of chapter 6.

- GARCH

  This does not apply for a single-regime GARCH model.

- MS-GARCH

  It is not known which state is turbulent and which is tranquil. In this case, State 1 is a jogger, who reacts slowly to news; State 2 is a sprinter, who reacts rapidly to news. To the authors knowledge, such an explanation has not been previously given in the literature. In addition, it is difficult to understand the implication that a higher return is associated with the jogger while the opposite is true for the sprinter.

(Oversight. 8) ***Excessive computational burden:***.

This issue has been intensively discussed previously in chapter 4.

6. *Economic Implication of the empirical studies*

According to the empirical results using S&P 500 stock index, the HMS-V suggests at least two different regimes, namely, a boom regime (bullish and tranquil) and a bust regime (bearish and turbulent). The S&P 500 has a long period of large positive returns with low volatility that is cast into a sharp profile by a crash with large negative return and excessive volatility. In this chapter, the following economic implications of the statistical findings are discussed:-

- **Asymmetric duration of different states:** Section 6.1 discusses the asymmetric business cycle and its relation to the Markov-switching volatility model.
- **Reason for the clustering of Markov states:** Section 6.2 discusses the self-reinforcing process of a tranquil and bull market, which explains why the Markov states are clustered, in particular for the long lasting tranquility.
- **Reason for the interplay between volatility and mean:** Section 6.3 gives the economic implications why the tranquil (turbulent) state is associated with bullish (bearish) return.

6.1. Asymmetric business cycle, velocity of time and alternative time scales

Regarding the price of commodities, Clark (1973) suggested that the time scale for the evolution of commodity prices was one based on information flows rather than the calendar time. If a similar argument is applicable to a wider range of asset prices,

including the stock indices, then in the turbulence of 2008[67], a fast-forwarding time scale exists. For example, in one single day in Mar/2008, the information flow (and hence the velocity of economic time) was much greater within one single day in Mar/2004. Therefore, measured by the economic time scale, one single calendar day in Mar/2008 could be equivalent to a few calendar days in Mar/2004, and one day (calendar) volatility in Mar/2008 could be equivalent to a few days (calendar) volatility in Mar/2004[68].

It is noticed that the boom and bust of stock indices are usually closely related to the business cycle. The turbulent period is often associated with lower expected return and a sluggish economy with the reverse being true for the tranquil state. This could because there are more intense information flows during the turbulent period (when the business cycle is down)[69]. The asymmetric boom and bust business cycle suggests a more intensive information flow at the bust state. The asymmetric business cycle is well documented in the literature (e.g. Keynes 1936 and Hicks 1950). Sichel (1993) used the following graph to describe the asymmetry of the business cycles in terms of steepness and deepness:

[67] The information flows in 2007 and 2008 were most intensive such as the crash of excessive credit expansion, the failure of some major financial institutions, the housing bubble burst in the U.S. (probably also U.K.), a bank run in the U.K., the freeze of inter-bank lending, and some unconventional measures taken by the central banks.
[68] This is only a very rough illustration.
[69] In chapter 3, some Hidden Markov models were reviewed for the business cycle featured with structural breaks, which were documented by Hamilton (1989) and Hamilton and Lin (1996).

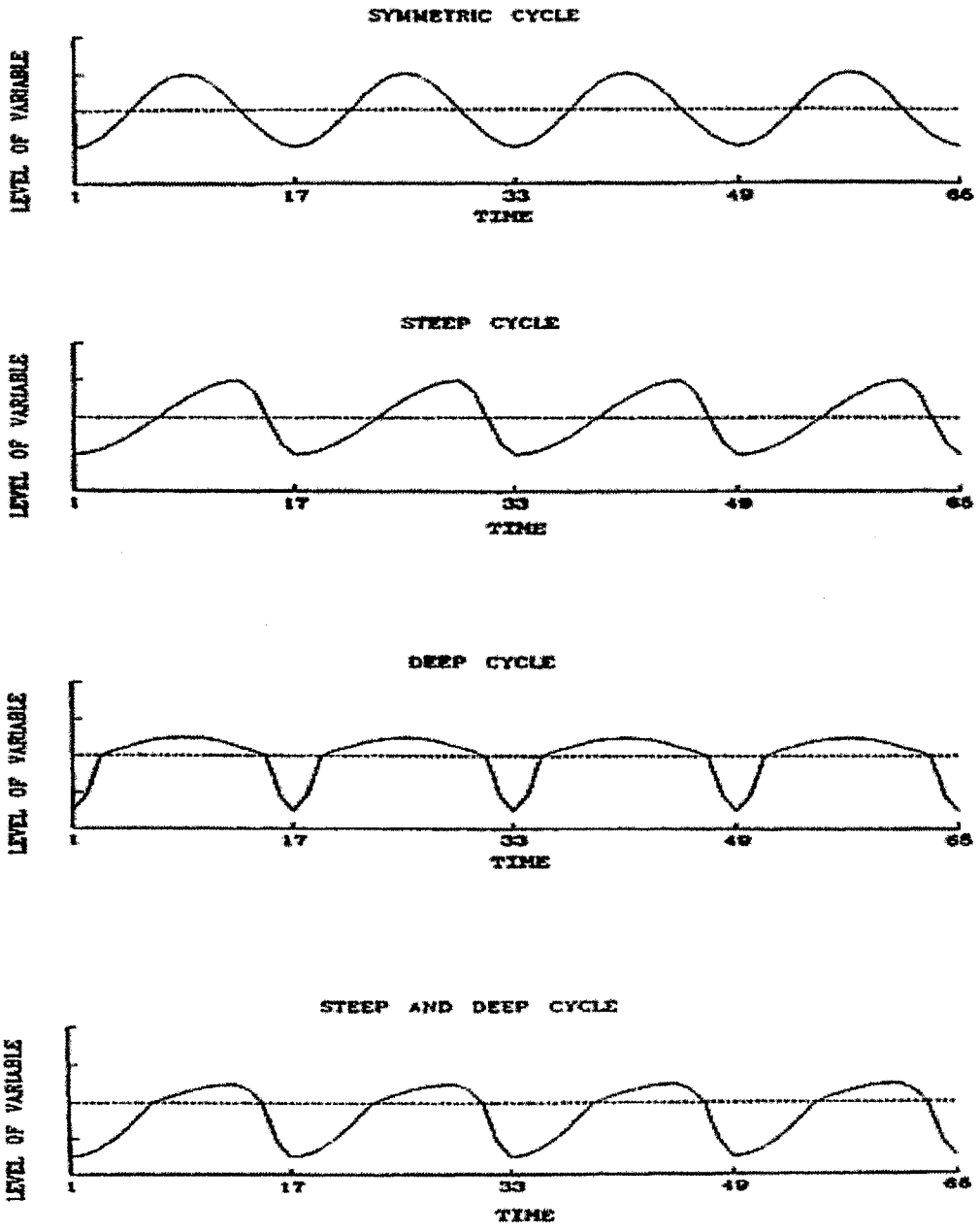


Figure 35

Asymmetric business cycle, velocity of time and alternative time scales. This graph is taken from Sichel (1993)'s paper.

As shown in *Figure 35*, compared with expansions, recessions tend to be more pronounced but shorter lived. Similar to Clark's (1973) idea, Stock (1987) studied cyclical time scale transformations. Suppose that:

- Stock's (1987) thesis in transforming the time scale is valid and the thesis is applicable also to the stock market.
- Using the economic time scale, one set of parameters is sufficient to model the volatility dynamics of the time series.

Therefore, for time series sampled according to the calendar time[70], two sets of parameters would be required. The hidden Markov model can also be understood as an alternative approach to rescale the calendar time scale into the economic time scale[71]: (i) when the time series resides at the tranquil state, the information flow carries fewer surprises (is less intensive or business as usual) and the economic time slows down; (ii) when the time series resides at the turbulent state, the information flow carries more surprises and the economic time speeds up. As a result, two sets of parameters are needed: one to describe the time series when time slows down (tranquil state) and another when time speeds up (turbulent state).

Interestingly, according to the plot of *Figure 35*, if the business cycle were asymmetrical, the expansionary state lasts longer than the contraction state. If the stock market is a reasonably good indicator of the business cycle, the tranquil state should also be more lasting, which has been validated by the empirical study carried out previously in chapter 5.

[70] That is, one set of parameters for the turbulent state while the volatility is evolving fast-forwardly, and another set of parameters for the tranquil state while the volatility is evolving slow-forwardly.
[71] Of course, economic modeling is always more or less subjective, and the proposed model out-performs previous algorithms in capturing and mimicking some real world economic observations.

6.2. Recursive loop of expectation and realization, self-reinforcing or self-correcting

The proposed model attempts to strike a balance between the influence of recent and distant history. However, in reality, economic agents often make decisions based only on the most recent history, which is one of the key reasons that recent history is most likely to repeat and that Markov states are clustered. For example, at the expansionary and tranquil state, an economic agent has a relatively exuberant expectation based on recent history. The agent will be forward testing his/her hypothesis by observing the realizations in the near future. This section discussed how exuberant expectations are repeatedly validated by the near future realizations:-

- Economic agents have relatively exuberant expectations based on recent history.
- Collectively, economic agents make economic decisions that influence the realization of the experiment, which is the key to a self-fulfilling process. For example, based on the passive exuberant expectation, agents positively promise greater investments[72] or releases loans[73]. Due to short-term pressure, some economic agents are forced to be more "upbeat"[74].
- The positive decisions by the agent boost the asset price.
- The exuberant expectation of the agent has been self-fulfilled and strengthened.

Therefore, during an expansionary state, even though the self-fulfilling process is further

[72] based on the expectation of higher earnings and/or the existence of more bullish investors
[73] based on the expectation of higher collateral value and/or lower default probability
[74] Quote from "The New York Time", Jul/10/2007
Citigroup's chief executive, Charles O. Prince, says his bank hasn't pulled back from making loans to provide funds for private equity deals, despite a skittish credit market and concerns that the recent run of big buyout deals could be losing steam. But Mr. Prince used an interesting metaphor to describe his company's situation as a major provider of financing for leveraged buyouts. "As long as the music is playing, you've got to get up and dance," he told The Financial Times on Monday, adding, "We're still dancing."

and further away from the long-run equilibrium (fundamental), it needs not revert to the equilibrium until a much later stage. As a result, the bullish sentiment and tranquil states are more likely to be clustered. In addition, some economic agents may be forced to take more optimistic[75] outlooks, since being less exuberant may results in a loss during the expansionary regime.

If the ex-post realization of the future contradicts the ex-ante forecast, a self-correcting process will begin. *Figure 36* describes the binomial outcome of the self-fulfilling and self-defeating process. Due to the agent's positive decision, the self-fulfilling scenario is far more likely to occur then the self-correcting scenario.

[75] A less "bullish" corporate executive or fund manager may be fired because of short-term peer pressure.

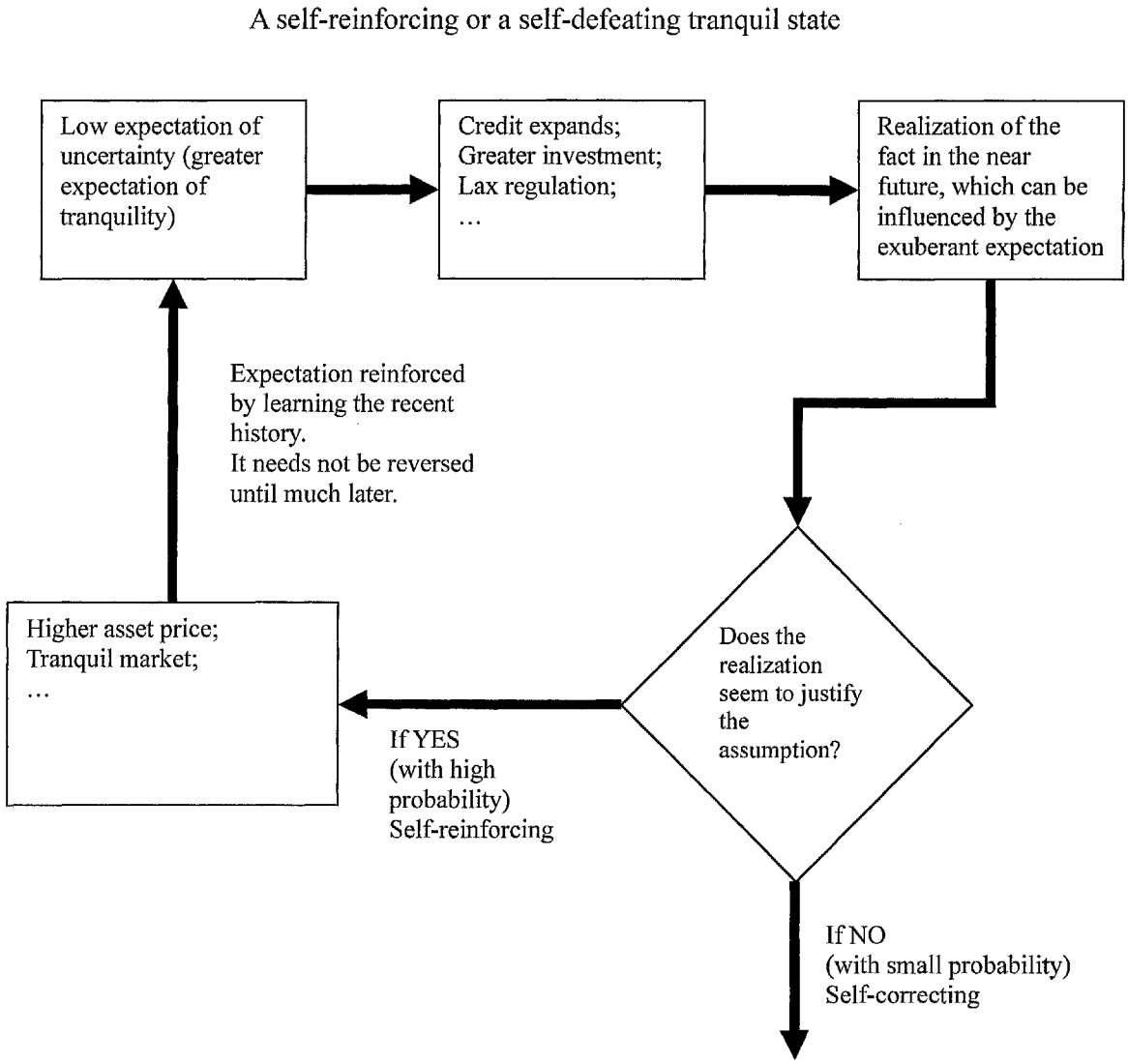


Figure 36

A self-reinforcing or a self-defeating tranquil state

6.3. Asymmetric surprises, sensitivities, correlations and risk premia

This section explains why the tranquil state is associated with higher return, while the

reverse is true for the turbulent state:-

- **Surprises, and the "uncertainty of assumptions"**: At the tranquil state, there are fewer surprises and less challenges on the assumptions of asset pricing, the reverse is true for the turbulent state.
- **Sensitivities, and the leverage ratios:** The leverage ratios of financial institutions and / or households are higher at the turbulent state and the price is more sensitive to shocks. The reverse is true for the tranquil state[76].
- **Correlations, and the diversification benefit:** The correlations of equity returns increase during the turbulent state and hence the diversification effect is reduced. The reverse is true for the tranquil state;
- **Time varying risk premia demanded by the investors:** The unexpected rise of volatility leads to an unexpected rise of risk premia in the future, which can only be compensated by an immediately lowered current price. This is how a disappointing return and turbulence are associated.

6.3.1. Fewer surprises at the tranquil state

During the boom regime, the economic outlook and government policy are considered more visible and stable. It is then easier to reach a "consensus" on the bullish asset pricing assumptions. Less volatile assumptions lead to less volatile asset prices:

1) **Greater consensus on revenue forecasting:** At the expansionary state, higher consumer confidence and more active investment activity boost the sale and revenue of a firm. Based only on recent history, there could be greater "consensus" on the extrapolating recent corporate revenue trend to the future.

[76] Please note that due to length considerations, sometimes this thesis only articulates the scenarios for a turbulent state as the scenario at the tranquil state is often simply the opposite.

2) **Greater consensus on forecasting central bank's policy, inflation and hence associated cost:** During a tranquil state, the monetary and regular policies might also be either unable to identify or unwilling to rein in the asset price boom. Note that it is a somewhat judgmental call to determine whether an asset price appreciation is "excessive", and a common objection against proactive monetary policies is that it requires the authorities to outperform market participants in assessing the fundamental values of asset prices (Bernanke and Gertler, 2001). The authorities may lack the intelligence to identify the "excessiveness", and it is not the mandate of central banks and thus politically inconvenient to intervene. Hence regulation has been lax[77]. It is easy to forecast the relevant policies (e.g. Fed target rate, regulation on non-deposit financial institutions).

3) **Greater consensus on forecasting the cost of fund:** Besides the monetary policy, the uncertainty in the cost of fund greatly depends on credit. Sometimes, a bullish stock market is also fueled by a credit expansion[78]. During such a bullish state, expectation of the future earnings and collateral values are more certain and credit could be abundant. Nobuhiro and Moore (1997) built an elegant model to explain the interaction between the credit limit and collateral price. The dynamic interaction turns out to be a powerful transmission mechanism by which the effects of shocks persist, amplify, and spill over into other sections, especially the stock market. Increasingly risky investments (often, over-investment) are thus financed. The cost of fund is low and can be forecasted with reasonable certainty, as it is easy to raise capital.

---

[77] The regulator may have an idealistic mentality on the effectiveness of the market and thus prefer lax regulation. A quote from "The age of turbulence" by Greenspan A (2007)
"Since markets have become too complex for effective human intervention, the most promising anti-crisis policies are those that maintain maximum market flexibility—freedom of action for key market participants such as hedge funds, private equity funds, and investment banks."

[78] It is more so for the financial crisis in the mid 2000s but less so for the tech bubble in the late 1990s.

4) **Lack of credible challenges to the above consensus:** To challenge and "surprise" bullish assumptions[79], a credible action is for investors to pull out their money and/or short sell the stock. Only by doing so, will the firmly believed bullish assumption be shaken. Market participants are defined as an investor who hold or short-sells stock[80]. During such a tranquil and bullish state, the market favors the optimists. One who believes a stock is too high can short it, borrowing shares and selling them in the hopes of replacing them later at a cheaper price. This can be costly however, both in the fees and in the risk of huge losses should the stock price keep rising[81]. According to Abreu (2003), if all the rational investors could agree to bet against an unrealistic assumption, they could make big profits. Without coordination, it's risky for any one of them to individually bet against the rising price. Many big investors rarely short stocks. Most investors like fund managers have tremendous short-term performance pressure, and thus have to follow the collectively mis-parameterized bullish assumptions[82]. When differences between bullish and bearish investors are extreme, many of the bears simply move to the sidelines, that is, they are no longer market players and lose the ability to change the bullish assumptions. The optimists are unchecked, and the bullish assumptions are left un-surprised. Catalysts to credibly challenge the

[79] Those who have positions on the shares are considered as the market players.
[80] Note that short-sell interests are only a few percentages of the total shares.

[81] **Short Sales Risk:** The short sale position will suffer if a security that it has sold short appreciates in value. The short seller position may also suffer if it is required to close out a short position earlier than it had intended. This would occur if the securities lender requires it to deliver the securities that short seller borrows at the commencement of the short sale and the short seller is unable to borrow the securities from other securities lenders.

[82] Quote from the "Bubbles and Crashes" by Abreu and Markus (2003):
"when Stanley Druckenmiller, who managed George Soros's $ 8.2 million Quantum Fund, was asked why he didn't get out of internet stocks earlier even though he knew that technology stocks were overvalued, he replied that he thought the party wasn't going to end so quickly. In his words "We thought it was the eighth inning, and it was the ninth."
Another example is Pequot Capital Management, Pequot Capital Management boarded the Internet bandwagon early, investing in America Online in 1994. It was heavily invested in tech stocks through the late 1990s. When they started falling in March 2000, Pequot got hurt. But it was agile enough to take bearish positions on the stocks, and its funds posted strong performances for the year. Mr. Brunnermeier saw the bubble, too. He thought people were crazy for buying tech stocks. But as both the hedge funds' gains and his theoretical work suggest, even if you know there's a bubble, it might be smart to go along.

bullish assumptions are lacking. Although articles or speeches by conscious economists express rational "challenge", they do not hold massive short positions to move the market and the "challenge" is not potent. Although truth might be with the minority, money and hence voting power is with the majority. Irrational money talks louder than articles.

As illustrated in *Figure 36*, an optimistic expectation and benign realization of the economic reality is likely to be self-reinforcing for a prolonged period of time, and the hidden danger of the economy need not reveal itself until a much later stage.

6.3.2. More surprises at the turbulent state

At the end of the tranquil state, problems are likely to "build up" (e.g. elevated earning expectation and hence elevated share price, elevated collateral value expectation and hence elevated debt[83]). There could be two possible outcomes: (i) the cumulative problems remain hidden, the bullish assumptions remain un-challenged and the market remains calm; (ii) the long-run level of productivity and a more realistic discounted value of the earning will be revealed. Note that the arrival of news on long-run equilibrium can come as a great surprise to both policy makers[84] and the buoyant market[85]. Greater

[83] Quote from "The age of turbulence" by Greenspan A. (2007)
"For the US alone, of the nearly $3 trillion of home mortgage originations in 2006, a fifth were subprime and another fifth were so-called Alt-A mortgages."

[84] For example, the Fed's stance on the U.S. housing bubble as of Oct/2005
Quote from "The economic outlook" by B. Bernanke at October 20, 2005:
"House prices have risen by nearly 25 percent over the past two years. Although speculative activity has increased in some areas, at a national level these price increases largely reflect strong economic fundamentals,..."
http://www.whitehouse.gov/cea/econ-outlook20051020.html

[85] For example, the collapse of LTCM in 1998 was beyond the imagination of most market players:
Quote from the LTCM Confidential Memorandum, Jan/1999
"The result was a downward spiral which fed upon itself driving market positions to unanticipated extremes well beyond the levels incorporated in risk management and stress loss discipline."

uncertainty in the asset pricing assumptions immediately translates into more volatile asset (e.g. equity) prices:-

1) **It is harder to forecast revenue:** The consumer confidence and hence company revenue may be low and volatile.
2) **It is harder to forecast the interest rate**: Depending on the central bank's policy, both inflation and deflation are possible, hence there is greater uncertainty on the future discount rate. The central bank may be forced to take some radical actions that are unthinkable to most market players[86], which disrupts the expectation of future interest rate, future inflation and hence future discount rate.
3) **It is harder to forecast the cost of fund**: This is because the banking system may also be stressed and banks have to hoard cash to reserve capital. Therefore, besides a risk free interest rate, it is also hard to forecast the credit spread chargeable. An overly leveraged firm may find that its funding unexpectedly dries up. If the money market dries up because of the toxic combination of credit and liquidity crunch, and it will impact those who rely too heavily on short-term funding. In addition, the bust in the collateral value restricts the firms' ability to self-finance their operations. Hence some promising projects have to be aborted[87].
4) **It is harder to forecast the future cost**: Inflation is a wildcard following the dilemma of central banks.
5) **Greater challenges on asset pricing assumptions**: As volatility attracts short sellers, both bullish and bearish assumptions will be constantly challenged, there is a greater diversity of market participants and the stock market is a classic battle-field.

[86] For example, in Mar/2008, the Fed took a move with no precedent since the Great Depression—that it would extend unlimited credit for six months not only to commercial banks, but also to investment banks and brokerage houses

[87] "A bank is a place where they lend you an umbrella in fair weather and ask for it back when it begins to rain." – Frost R.

The hidden turbulent state thus takes over the volatility dynamics within a short period of time. Note that it is hard to time when the asset bubble will burst and turbulence may remain hidden and undetected for long periods of time. Therefore, although the stock market only violently reacted to the burst of credit bubble since Aug/2007, the hidden cumulative danger could have burst earlier when the market was still tranquil. This hidden danger should be incorporated into models, as discussed previously in section 5.6.2.

Note that during the turbulent regime, surprises are often clustered. For example, the near collapse of Bear Stearns in 2008 raised the fear of the crash of the inter-connected banking system, which is a downside surprise for the stock market. The Fed's ensuing bail out and loosened monetary policy shocked the market again at the upside. However, the aggressive moves by the Fed erode the value of the USD, lifted commodity prices, and increased the fear of inflation, which was the downside surprise. As one surprise leads to another, the greater uncertainties of the asset pricing assumptions are also clustered. This corresponds to the empirical finding that volatility is more sensitive to recent shocks during a turbulent regime.

6.3.3. Greater sensitivity at the turbulent state

In economics, leverage (or gearing) magnifies the potential positive or negative outcome. Unfortunately, leveraging of financial institutions may have been accumulated during the tranquil state, which then peaks at the beginning of a turbulent state. Hence there is greater sensitivity at the turbulent state.

For a financial institution, leverage takes the form of aggressive lending / borrowing or using margin to increase the potential return in a trading position. A greater leverage ratio leads to magnified profit and loss. Thus leverage is considered an important driver of how much (i.e. volatility) reaction there is to news. As of early 2008, it is believed that the leverage ratios of major US brokerage firms had gone up to over 30 times[88]. Therefore, at the end of the tranquil state, although the volatility of stock indices is temporarily low, there has always been a hidden danger lurking, and once the crisis is triggered, volatility can shoot up in a very short period of time.

The possible credit and liquidity crisis during the turbulent period also contributes greatly to higher volatility. With a stressed and volatile collateral value, financial institutions may become reluctant to provide liquidity to each other. To the anguish of the market, creditors (or trade counterparty) would also demand a greater haircut[89] and thus more collateral as compensation. When those adverse conditions simultaneously motorize, financial institutions would have no other choice but to de-leverage. This usually means raising capital, selling assets (often in the form of fire-sale), and/or hoarding cash (reluctance to lend). Fire-sales often lead to volatile and bearish stock markets. *Figure 37* shows how the sharp decline in price and rising volatility feed on themselves during such a turbulent period,

[88] Source: "The Financial Times" April/2008

[89] Note that for most financial institutions, a haircut is a function of volatility.
Quote from Basel II Accord, 156:
"In calculating the haircuts, a 99th percentile, one-tailed confidence interval is to be used."
Therefore, the greater the value-at-risk (VaR), the greater the haircut is required.

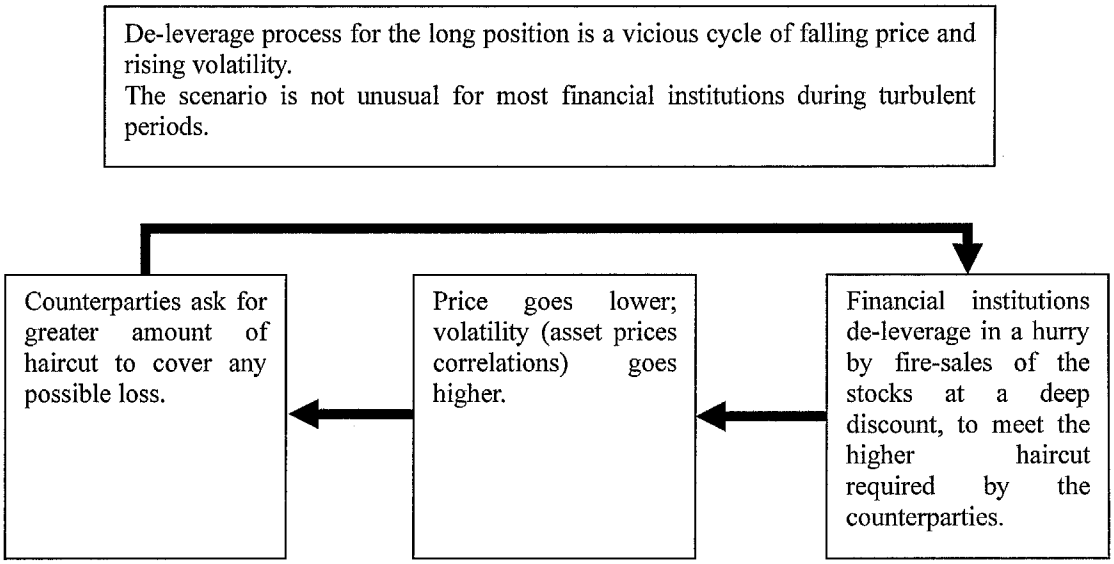


Figure 37
The sharp decline in price and rising volatility are self-reinforcing.

At the same time, the market is also more sensitive to upside shocks. This is because to some short-sellers and speculators, bear market and volatility represents opportunity[90]. Short-selling usually involves the usage of margin and hence leverage. Those short-sellers are not long-term passive investors and they tend to get in and out of the market frequently. Short-sellers help to push the price downward in a faster manner, however they are also prone to losing money when market moves against them. If the market moves against short-sellers, to meet the margin call, they tend to get out in a hurry, buying back the shorted shares at a higher price and thus pushing the price up further. In addition, short-selling involves stock-lending and the standard stock-lending practice is that the loan must be repaid on demand. This practice exposes short-seller to the risk of being "squeezed." A short squeeze occurs when the lender of the borrowed share refuses to lend the shares anymore. If the short-seller is unable to find an alternative lender, the

[90] The New York Stock Exchange (NYSE) says 16bn NYSE-listed shares were sold short as of Mar/31/2008. On Mar/2007, this number is only 10.5bn.
http://www.nyxdata.com/nysedata/asp/download.asp?s=xls&prod=shortint
Essentially, if the volatility is low, the price movements are insufficient to provide a level of expected profit in excess of transaction costs.

short-seller must repurchase the share in the open market to repay the loan and close the position, i.e., short-covering. Short-covering is likely to let the market enjoy a sharp rally in a short period of time and aggrandize the volatility. Note that "short-covering" can be understood as a de-leveraging process of the short position. The magnitude of the market rebound could be far beyond the imagination of the market player who uses leverage. As a result, during the turbulent state, short-selling is also more active, and it contributes to higher volatility.

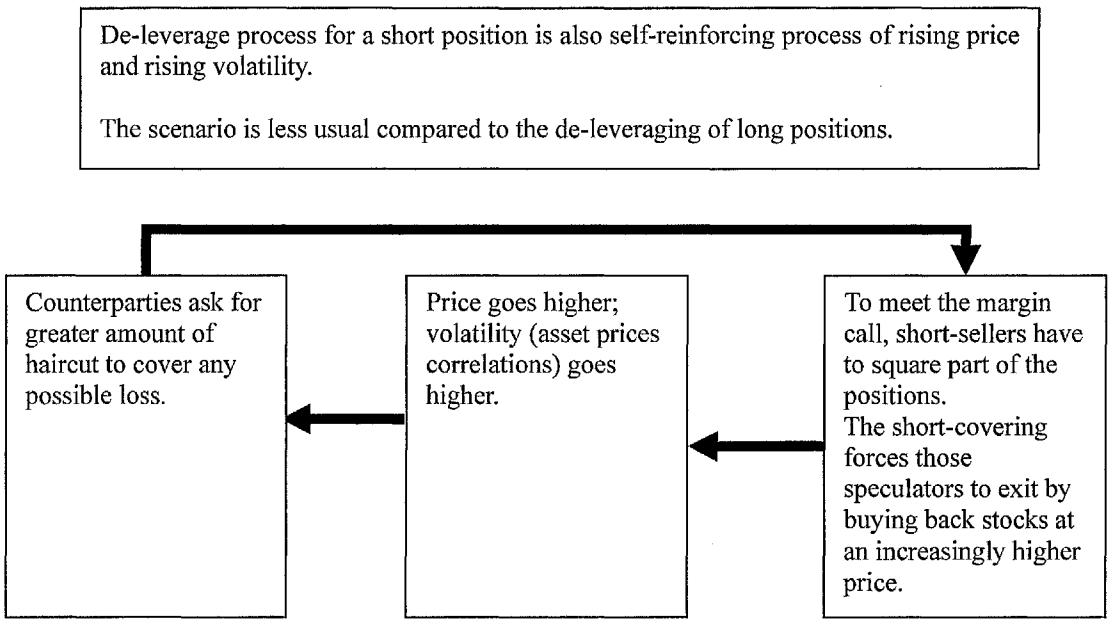


Figure 38

The de-leverage process for the short position is also a self-reinforcing process of sharply rising price and volatility.

The activities of short-selling, profit taking, portfolio insurance and buying on dips are often also clustered. As trend following speculators try to profit from the volatility of the market, the expectation of volatility becomes self-fulfilling. This also corresponds to the previous empirical finding that volatility is more sensitive to more recent shocks during a

turbulent regime.

6.3.4. Increased equity return correlation at the turbulent state

The increased volatility of the stock indices can be partially attributable to the increased correlation of equity returns during the turbulent state. The S&P 500 is a well balanced U.S. Large-Cap portfolio, and lower correlations between the individual stock returns reduce the portfolio risk. However, the benefit of diversification may be materiality damaged during the turbulent state, as the correlation of equity returns is likely to shoot up substantially.

These asymmetric correlations were tested by Hong, Tu and Zhou (2008). Their empirical study suggested stocks tend to have greater correlations with the market when the market goes down than when it goes up. Hong, Tu and Zhou (2008) also provided such tests for asymmetric betas and covariances. Andrew and Chen's (2002) empirical study also suggested that correlations between U.S. stocks and the aggregate U.S. market are much greater for downside moves, especially for extreme downside moves, than for upside moves. Longin and Solnik (2001) plotted downside and upside correlations of the U.S. equity market and showed that downside correlations were much larger than upside correlations. The correlation breakdowns weaken the benefit of diversification during the turbulent state and the volatility of the S&P 500 is more likely to go up.

6.3.5. Greater risk premia requested by the investor at the turbulent market

Turner, Startz and Nelson (1989) constructed a Markov switching volatility model in which agents are uncertain of the state and their findings indicated that agents are consistently surprised by high-variance period. Thus there is a negative correlation between movements in volatility and in excess returns (i.e., a turbulent state is often associated with a bear market). If the regime unexpectedly switches from the tranquil state to the turbulent state, the investors would immediately ask for higher risk premia. The expected higher risk premia immediately requires a lower current stock price. Therefore, a surprise often causes both higher volatility and a sharp decline in stock price.

7. Thesis Summary

Chapter 1 to chapter 4 proposed a hidden Markov-switching volatility model. The most unique feature of the proposed model is the parallel structure, which separates the INTER-state and INTRA-state dynamics, enhances greater transparency, balances the memory of both recent and distant history, provides more consistent economic implication and greatly stabilizes the subsequent EM algorithm.

Chapter 5 gave an empirical example of the HMS-V using the S&P 500 weekly return. The proposed model identified two regimes with distinctly different patterns both in volatility and mean. The expected return at the tranquil state was higher, while the expected return at the turbulent state was lower. It was also observed that the tranquil state lasted longer than the turbulent state.

Chapter 6 explored the economic implication of the empirical findings. The long lasting tranquility was related to the asymmetric business cycle and a self-reinforcing mechanism during an expansionary state. The reasoning behind an often bullish tranquil state was also discussed, with the reverse often being true for the turbulent state.

However, this thesis only discussed a two-state hidden Markov-switching model based on the parsimonious principal, while a model with more states could better describe the time series. In addition, while the time series discussed is only univariate, the proposed framework can be readily extended to a time series with higher dimensions. The identifiability condition for the mixture multinomial distribution is also well established,

and the hidden Markov-switching process can be used to model the dynamic correlation matrix. The semi-positivity condition of the variance-covariance matrices often requires a simpler covariance link function, especially within a hidden Markov-switching framework. The logit link function can then be utilized to realize some interesting dynamics. For example, for the foreign exchange rates of the EUR to the U.S. dollar (EUR/USD) and the Icelandic Krona to the U.S. dollar (ISK/USD), the dynamics of EUR/USD would be expected to have a greater impact on the ISK/USD whereas the inverse would not be expected due to the different sizes of their economies. In order to describe the dynamics, the sum of historical squared return of EUR/USD can be used to construct the covariate for the Logit link function. As a result, the dynamics of the EUR/USD would have a higher impact on the state transition probability. The covariate for the Logit link function with historical square return of both the EUR/USD and the ISK/USD can also be constructed but give the former a greater weight according to some economic factors.

ADDITIONAL TABLES

SEM algorithm of the HMS-V model without covariate for (to be continue)

Starting value		*Global optimum*	
$\mu_{\langle 1\rangle}^{(0)}$	0.000	$\mu_{\langle 1\rangle}^{(131)}$	0.301
$\sigma_{\langle 1\rangle}^{(0)}$	1.075	$\sigma_{\langle 1\rangle}^{(131)}$	1.564
$\mu_{\langle 2\rangle}^{(0)}$	0.000	$\mu_{\langle 2\rangle}^{(131)}$	0.035
$\sigma_{\langle 2\rangle}^{(0)}$	4.301	$\sigma_{\langle 2\rangle}^{(131)}$	2.937
$p_{\langle 11\rangle}^{(0)}$	0.800	$p_{\langle 11\rangle}^{(131)}$	0.988
$p_{\langle 22\rangle}^{(0)}$	0.800	$p_{\langle 22\rangle}^{(131)}$	0.980
LLF_{obs}	-1266.5		

DM matrix

0.0733	-0.0077	-0.1426	0.0192	-0.0022	-0.0048
-0.0153	0.3505	-0.1707	0.4311	0.0160	0.0066
-0.0218	-0.0130	0.0577	-0.0503	0.0003	0.0026
0.0062	0.0659	-0.1014	0.2835	-0.0001	-0.0106
-0.4706	1.7129	0.4835	-0.0882	0.3638	0.6344
-0.3260	0.2267	1.1782	-2.3948	0.2017	0.4806

Eigen-value of DM matrix

0.84	0.55	0.11	0.09	0.02	5.78E-08

Corresponding Eigen-vectors

0.007	-0.007	-0.235	-0.131	-0.002	0.873
-0.027	-0.283	0.197	-0.240	0.017	0.117
-0.002	0.020	0.071	0.028	-0.001	0.458
0.008	-0.098	-0.052	0.112	0.013	0.118
-0.835	-0.816	-0.843	0.944	-0.894	0.002
-0.549	0.495	-0.432	0.143	0.448	-0.001

Information matrix of the complete data $\mathbf{I}_{com}\left(\hat{\boldsymbol{\theta}}\right)$

263	-1.08E-13	0	0	0	0
-1.08E-13	526	0	0	0	0
0	0	40	-5.73E-15	0	0
0	0	-5.73E-15	80	0	0
0	0	0	0	58085	0
0	0	0	0	0	18520

Information matrix of the missing data $\mathbf{I}_{miss}\left(\hat{\boldsymbol{\theta}}\right)$

19	-4	-6	2	-125	-89
-4	184	-7	35	930	122
-6	-7	2	-4	19	49
2	35	-4	23	-7	-197
-124	901	19	-7	21133	11749
-86	119	47	-192	11714	8901

SEM algorithm of the HMS-V model without covariate (continued)

Information matrix of the observable data $\mathbf{I}_{obs}\left(\hat{\boldsymbol{\theta}}\right)$ *by* ***SEM***

244	4	6	-2	125	89
4	342	7	-35	-930	-122
6	7	38	4	-19	-49
-2	-35	4	57	7	197
124	-901	-19	7	36952	-11749
86	-119	-47	192	-11714	9619

Variance-covariance matrix by V ***parametric bootstrap***

4.53E-03	-2.49E-04	-1.00E-03	-2.04E-04	1.71E-05	-1.88E-05
-2.49E-04	3.00E-03	-3.18E-04	1.15E-03	4.92E-05	3.38E-05
-1.00E-03	-3.18E-04	3.02E-02	-4.40E-05	-1.60E-05	-1.11E-04
-2.04E-04	1.15E-03	-4.40E-05	1.93E-02	8.47E-07	-2.45E-04
1.71E-05	4.92E-05	-1.60E-05	8.47E-07	3.57E-05	1.68E-05
-1.88E-05	3.38E-05	-1.11E-04	-2.45E-04	1.68E-05	1.61E-04

The inverse of the Variance-covariance matrix by ***parametric bootstrap***

224	21	8	2	-154	46
21	352	4	-21	-468	-52
8	4	33	0	-5	24
2	-21	0	54	-14	88
-154	-468	-5	-14	30200	-3087
46	-52	24	88	-3087	6681

Information matrix of the observable data by ***numerical difference***

244	4	6	-2	124	88
4	342	7	-35	-902	-119
6	7	38	4	-20	-47
-2	-35	4	57	8	191
124	-902	-20	8	36411	-11367
88	-119	-47	191	-11367	9351

Standard errors of MLE by different methods

SEM	Bootstrap	Numerical difference
0.0648	0.0673	0.0648
0.0597	0.0548	0.0595
0.1661	0.1738	0.1663
0.1440	0.1391	0.1444
0.0074	0.0060	0.0073
0.0145	0.0127	0.0146

Table 18

The SEM algorithm and comparison with previous models for Regression 01

ADDITIONAL FIGURES

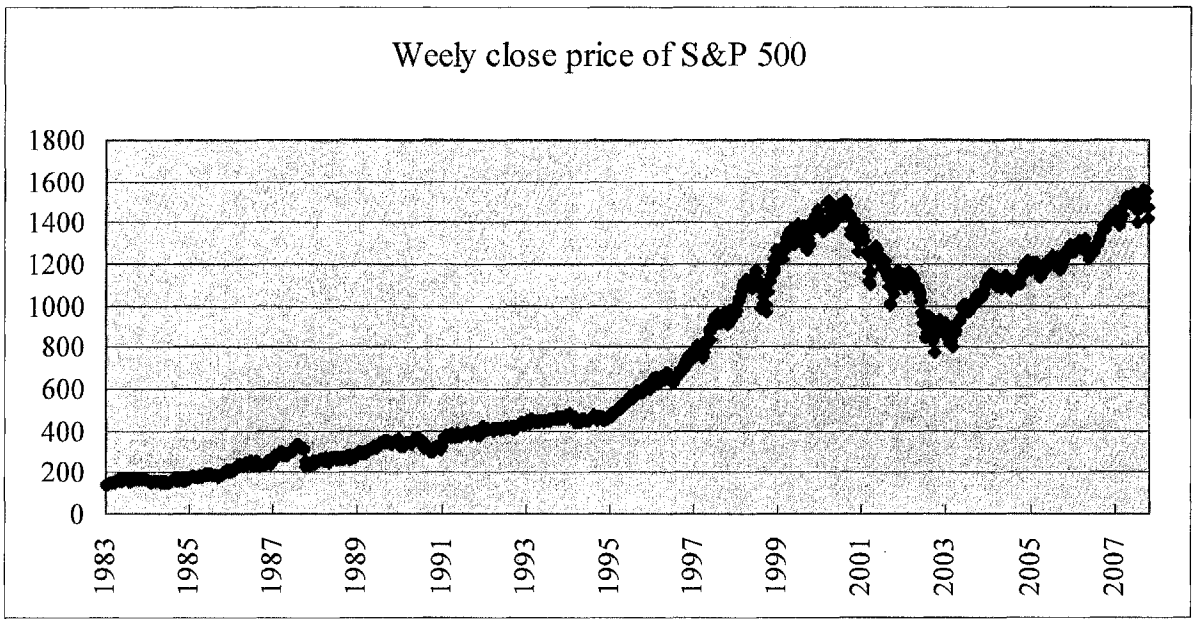


Figure 39
The historical price of the S&P500 stock index from Jan/05/1983 to Nov/21/2007.

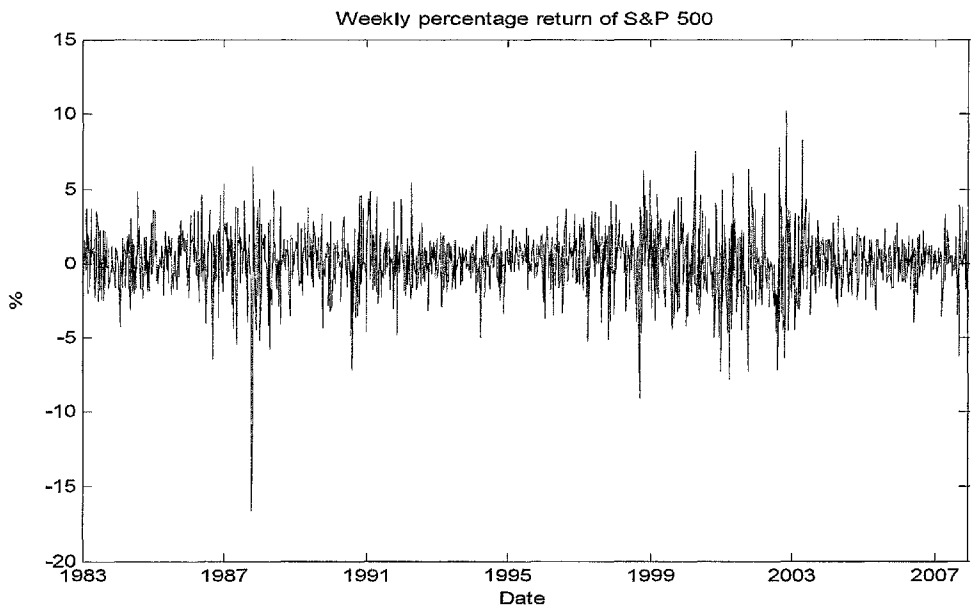


Figure 40
The historical weekly percentage return on S&P500 from Jan/05/1983 to Nov/21/2007.

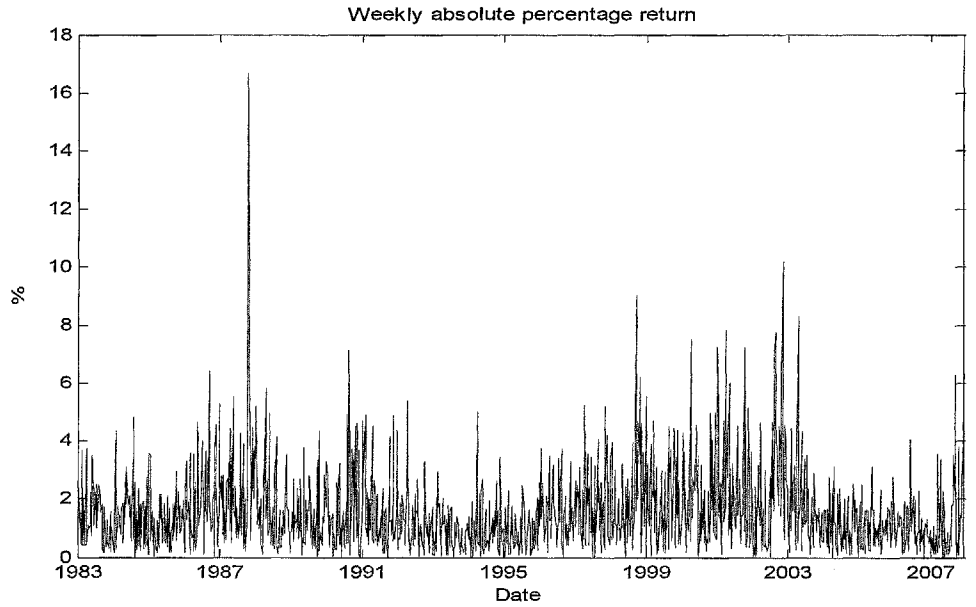


Figure 41

The historical absolute weekly percentage return of S&P 500 from Jan/05/1983 to Nov/21/2007.

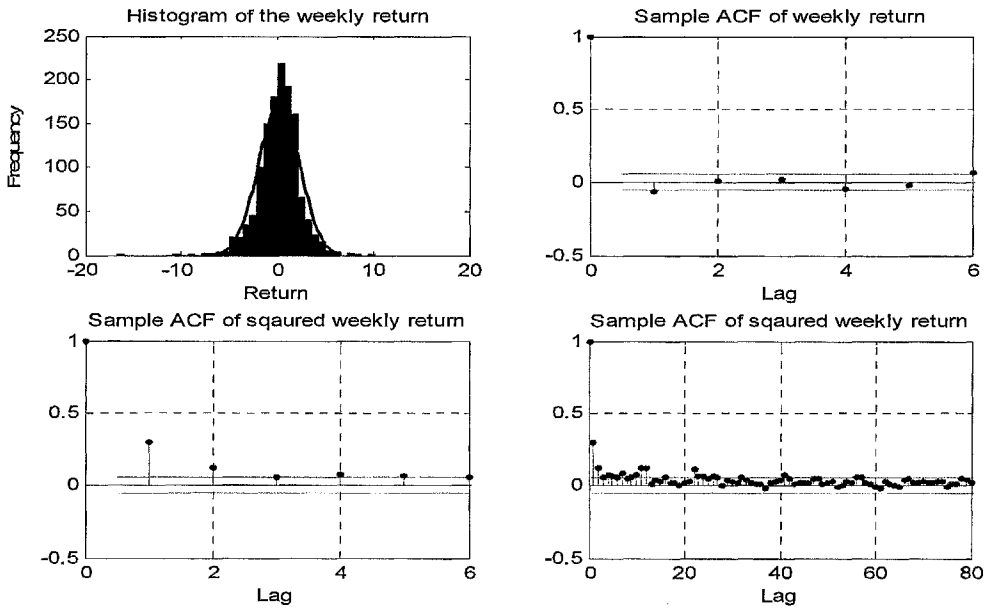


Figure 42

The historical weekly percentage return on the S&P500 index and a normal distribution.	*The ACF of weekly return*
The ACF of squared weekly return.	*The ACF of squared weekly return (800 lags).*

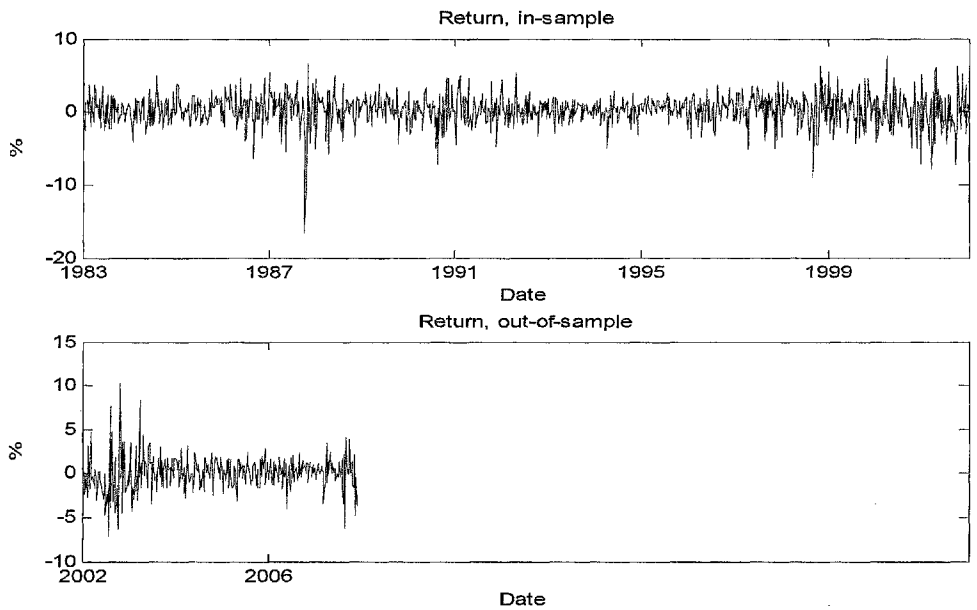


Figure 43

Spited return data: the first 989 observations are for the in-sample inference and the rest 308 observations are for the out-of-sample inference.

APPENDIX I

The appendix gives a sufficient identifiability condition of the proposed model.

App1.1. The property of the mixture regression

The *p.d.f.* of a mixture normal distribution as

$$MixN\left(\nu\middle|\lambda_1,\mu_1,\sigma_1^2,\mu_2,\sigma_2^2\right)=\lambda_1\phi\left(\nu\middle|\mu_1,\sigma_1^2\right)+\left(1-\lambda_1\right)\phi\left(\nu\middle|\mu_2,\sigma_2^2\right). \qquad \text{(AppI.1)}$$

For a mixture regression, the underlying state is hidden, and y_t follows a two-component mixture normal distribution. The weight on the component i at time t is the rate of the successful Bernoulli trail $\Delta_t=i$. Let f_t be the *p.d.f.* of such conditional mixture normal distribution with covariate,

$$\begin{aligned} f_t &= f\left(y_t \mid \mathbf{c}_t,\ \mathbf{z}_t, \boldsymbol{\theta}\right) \\ &= \Pr\left(y_t \mid \mathbf{X}_\mathbf{T}, \boldsymbol{\theta}\right) \\ &= \Pr\left(\Delta_t=1\right)\bullet f_{\langle 1\rangle,t}+\Pr\left(\Delta_t=2\right)\bullet f_{\langle 2\rangle,t} \\ &= \Pr\left(\Delta_t=1\right)\bullet f_{\langle 1\rangle}\left(y_t \mid \mathbf{c}_t,\mathbf{z}_t,\Delta_t=1,\boldsymbol{\alpha}_{\langle 1\rangle},\boldsymbol{\beta}_{\langle 1\rangle}\right) \\ &\qquad +\Pr\left(\Delta_t=2\right)\bullet f_{\langle 2\rangle}\left(y_t \mid \mathbf{c}_t,\mathbf{z}_t,\Delta_t=2,\boldsymbol{\alpha}_{\langle 2\rangle},\boldsymbol{\beta}_{\langle 2\rangle}\right) \\ &= \Pr\left(\Delta_t=1\right)\bullet\phi\left(y_t\middle|\mu_{\langle 1\rangle,t},\sigma^2_{\langle 1\rangle,t}\right)+\Pr\left(\Delta_t=2\right)\bullet\phi\left(y_t\middle|\mu_{\langle 2\rangle,t},\sigma^2_{\langle 2\rangle,t}\right) \end{aligned} \qquad \text{(AppI.2)}$$

According to Equation (2.17), the log likelihood function of the proposed HMS-V is

$$\begin{aligned} LLF_{obs}(\boldsymbol{\theta}) &= \log\left(f(\mathbf{Y}_\mathbf{T}\mid\mathbf{X}_\mathbf{T},\boldsymbol{\theta})\right) \\ &= \log\left\{\sum_{i=1}^{2}\sum_{\Delta_2=1}^{2}\cdots\sum_{\Delta_T=1}^{2}\left[\Pr(\Delta_1=i)\bullet f_{\langle i\rangle,1}\bullet\prod_{t=2}^{T}\left(p_{\langle\Delta_{t-1}\Delta_t\rangle,t}\bullet f_{\langle i\rangle,t}\right)\right]\right\}. \end{aligned}$$

Note that the sample likelihood function is NOT simply the products of f_t, because the two variables $\left(y_t,\ \Delta_t\right)$ are not pair-wise independent.

App1.2. A sufficient identifiability condition

Identifiability condition is a prerequisite for a mixture distribution or a regression involves with a mixture distribution. A sufficient identifiability condition for the proposed model will be given in this appendix.

The proposed model is only identifiable if there exists a unique global maximizer $\boldsymbol{\theta}$ for $LLF_{obs}(\boldsymbol{\theta})$. Without covariate, Teicher (1961, 1963) shows that all mixture normal distributions are identifiable. With covariate, Hennig (2000) proves that in theory the full rank of the covariate is still not strict enough to guarantee the model identifiability, although in practice the full rank of the covariate is usually good enough. He points out that the problems lies in a label switching scenario, and gives two counterexamples as illustrations. We will incorporate his discovery, and give a sufficient condition for the identifiability for the proposed HMS-V model. Note that the log likelihood function of the proposed HMS-V model is:-

$$\begin{aligned} LLF_{obs}(\boldsymbol{\theta}) &= \log\left(f\left(\mathbf{Y}_{\mathbf{T}} \mid \mathbf{X}_{\mathbf{T}}, \boldsymbol{\theta}\right)\right) \\ &= \log\left\{\sum_{i=1}^{2}\sum_{\Delta_2=1}^{2}\cdots\sum_{\Delta_T=1}^{2}\left[\Pr(\Delta_1 = i)\cdot f_{\langle i\rangle,1}\cdot\prod_{t=2}^{T}\left(p_{\langle\Delta_{t-1}\Delta_t\rangle,t}\cdot f_{\langle i\rangle,t}\right)\right]\right\} \end{aligned} \qquad \text{(AppI.3)}$$

Definition. Consider the class of probability models, $\Pr(\mathbf{Y}_{\mathbf{T}} \mid \mathbf{X}_{\mathbf{T}}, \boldsymbol{\theta})$, with $\Pr\left(y_t \mid \mathbf{X}_{\mathbf{T}}, \boldsymbol{\theta}\right)$ defined by Equation (AppI.3), parameter space $\boldsymbol{\Theta}$, sample space $\mathcal{Y}_1 \times \cdots \times \mathcal{Y}_T$, and strictly exogenous covariate $\mathbf{X}_{\mathbf{T}} = \left(\mathbf{W}_{\mathbf{T}}, \mathbf{C}_{\mathbf{T}}, \mathbf{Z}_{\mathbf{T}}\right)$. The number of states of $\boldsymbol{\theta}$ is $\left|\Delta(\boldsymbol{\theta})\right| = 2$. We also require the parameters set in the space $\boldsymbol{\Theta}$ be sorted according to the following labeling rules.

- For $\left(\left(\boldsymbol{\alpha}_{\langle 1\rangle}, \boldsymbol{\beta}_{\langle 1\rangle}, \boldsymbol{\varphi}_{\langle 1\rangle}\right), \cdots, \left(\boldsymbol{\alpha}_{\langle|\Delta(\boldsymbol{\theta})|\rangle}, \boldsymbol{\beta}_{\langle|\Delta(\boldsymbol{\theta})|\rangle}, \boldsymbol{\varphi}_{\langle|\Delta(\boldsymbol{\theta})|\rangle}\right)\right) \in \boldsymbol{\Theta}$, the 1^{st} labeling criterion is

$$\mathbf{z}_1\boldsymbol{\beta}_{\langle 1\rangle} \le \mathbf{z}_1\boldsymbol{\beta}_{\langle 2\rangle} \le \cdots \le \mathbf{z}_1\boldsymbol{\beta}_{\langle |\Delta(\theta)|\rangle}. \qquad \text{(AppI.4)}$$

- If the relationship in (AppI.4) are all "<", then we are already able to differentiate them. If there is any "=" in (AppI.4), sort the parameter according the 2^{nd} criterion:

$$\mathbf{z}_2\boldsymbol{\beta}_{\langle 1\rangle} \le \mathbf{z}_2\boldsymbol{\beta}_{\langle 2\rangle} \le \cdots \le \mathbf{z}_2\boldsymbol{\beta}_{\langle |\Delta(\theta)|\rangle}. \qquad \text{(AppI.5)}$$

- Similarly, if for all the previous t labeling criteria, $\exists\ \boldsymbol{\beta}_{\langle i\rangle}$ and $\boldsymbol{\beta}_{\langle j\rangle}$, $\left(i \ne j;\ i,j = 1,\cdots,\left|\Delta(\theta)\right|\right)$ such that $\boldsymbol{\beta}_{\langle i\rangle}\mathbf{z}_\tau = \boldsymbol{\beta}_{\langle j\rangle}\mathbf{z}_\tau$ for $\tau = 1,\cdots,t$, then sort them according to the $(t+1)^{th}$ labeling criterion,

$$\mathbf{z}_{t+1}\boldsymbol{\beta}_{\langle i\rangle} \le \mathbf{z}_{t+1}\boldsymbol{\beta}_{\langle j\rangle}, \qquad \text{(AppI.6)}$$

until $t = T$.

- If $\exists\ \boldsymbol{\beta}_{\langle i\rangle}$ and $\boldsymbol{\beta}_{\langle j\rangle}$, who satisfy, $\mathbf{z}_t\boldsymbol{\beta}_{\langle i\rangle} = \mathbf{z}_t\boldsymbol{\beta}_{\langle j\rangle}$, $\forall t = 1,\cdots,T$, then sort them according to the $(T+1)^{th}$ criterion:

$$\mathbf{c}_1\boldsymbol{\alpha}_{\langle i\rangle} \ge \mathbf{c}_1\boldsymbol{\alpha}_{\langle j\rangle} \qquad \text{(AppI.7)}$$

until the $(T+T)^{th}$ criterion,

$$\mathbf{c}_T\boldsymbol{\alpha}_{\langle i\rangle} \ge \mathbf{c}_T\boldsymbol{\alpha}_{\langle j\rangle} \qquad \text{(AppI.8)}$$

If the relationships between all the above $T+T$ criteria are all "=", then the two components are not identifiable.

The class of probability model is identifiable if for $\boldsymbol{\theta}, \boldsymbol{\theta}^* \in \boldsymbol{\Theta}$, we have that

$$f\left(\mathbf{Y}_\mathbf{T} \mid \mathbf{X}_\mathbf{T}, \boldsymbol{\theta}\right) = f\left(\mathbf{Y}_\mathbf{T} \mid \mathbf{X}_\mathbf{T}, \boldsymbol{\theta}^*\right) \qquad \text{(AppI.9)}$$

for all $\left(y_1,\cdots,y_T\right) \in \mathcal{Y}_1 \times \cdots \times \mathcal{Y}_T$, if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}^*$.

Note the above labeling rule indicates that the two models are identical should they agree

up to the permutations of parameters. In practice, the first criterion (AppI.4) usually provides enough information to label the parameters. We now provide a sufficient condition for the identifiability.

Sufficient condition. The HMS-V regression model is identifiable up to $|\Delta|$ states if it satisfies both of the following two conditions

1) $$\min\left\{g:\left\{\left(\mathbf{c}_t^-,\mathbf{z}_t^-\right):t\in\mathbf{T}\right\}\subseteq\bigcup_{t=1}^{g}\mathbf{h}_t:\mathbf{h}_t\in\mathcal{H}_{D_\mathbf{C}+D_\mathbf{Z}-3}\right\}>|\Delta|=2, \qquad \text{(AppI.10)}$$

where

- the dimension of $\mathbf{c}_t$ is $D_\mathbf{C}$, and the first element of $\mathbf{c}_t$ is an intercept term;
- the last $D_\mathbf{C}-1$ terms of $\mathbf{c}_t$ are denoted as $\mathbf{c}_t^-$;
- the last $D_\mathbf{Z}-1$ terms of $\mathbf{z}_t$ are denoted as $\mathbf{z}_t^-$;
- $\mathbf{T}$ is a vector of time series index, $\mathbf{T}=1,2,\cdots,T$; and
- $\mathcal{H}_{D_\mathbf{C}+D_\mathbf{Z}-3}$ denotes the space of $D_\mathbf{C}+D_\mathbf{Z}-3=\left(D_\mathbf{C}-1\right)+\left(D_\mathbf{Z}-1\right)-1$ dimensional hyperplanes of $\Re^{D_\mathbf{C}+D_\mathbf{Z}-2}$.[91]

2) All of the three covariates, $\mathbf{W}_\mathbf{T},\mathbf{C}_\mathbf{T}$ and $\mathbf{Z}_\mathbf{T}$, are full rank.

Proof. Suppose that $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ both satisfy Equation (AppI.9), where their number of states are $|\Delta(\boldsymbol{\theta})|$ and $|\Delta(\boldsymbol{\theta}^*)|$ respectively. Without loss of generality, assume,

$$|\Delta|\geq|\Delta(\boldsymbol{\theta})|\geq|\Delta(\boldsymbol{\theta}^*)| \qquad \text{(AppI.11)}$$

According to Hennig (2000, pp 290-291), we can preclude the existence of some $\boldsymbol{\theta}$, such that

[91] Interested reader please refer to Hennig (2000) for mathematical details.

$$\bigcup_{\theta=\theta^*}\left\{\left(\mathbf{c}^-,\mathbf{z}^-\right):\ \mathbf{c}'\boldsymbol{\beta}=\mathbf{c}'\boldsymbol{\beta}^*\ \text{ and }\ \mathbf{z}'\boldsymbol{\alpha}=\mathbf{z}'\boldsymbol{\alpha}^*\right\}\supset\left\{\left(\mathbf{c}_t^-,\mathbf{z}_t^-\right):t\in\mathbf{T}\right\} \qquad \text{(AppI.12)}$$

Because then $g\le\left|\Delta\left(\theta^*\right)\right|\le\left|\Delta(\theta)\right|$, which leads to contradiction of condition 1.

If so, integrating both sides of Equation (AppI.9) for $y_2,\cdots,y_T$, respectively,

$$\begin{aligned}&\int_{y_2}\int_{y_3}\cdots\int_{y_T} f\left(\mathbf{Y_T}\,|\,\mathbf{X_T},\boldsymbol{\theta}\right)dy_T\cdots dy_3dy_2\\&=\int_{y_2}\int_{y_3}\cdots\int_{y_T} f\left(\mathbf{Y_T}\,|\,\mathbf{X_T},\boldsymbol{\theta}^*\right)dy_T\cdots dy_3dy_2\end{aligned} \qquad \text{(AppI.13)}$$

which yields

$$\sum_{i=1}^{2}\Delta_{ini}\langle i\rangle\phi\left(y_1\left|\mu_{\langle i\rangle,1},\sigma^2_{\langle i\rangle,1}\right.\right)=\sum_{i=1}^{2}\Delta^*_{ini}\langle i\rangle\phi\left(y_1\left|\mu^*_{\langle i\rangle,1},\sigma^{*2}_{\langle i\rangle,1}\right.\right) \qquad \text{(AppI.14)}$$

for all $y_1\in\mathcal{Y}_1$. Since each side of Equation (AppI.14) can be regarded as a finite normal mixture without covariates, Teicher (1961) classifies the mixture normal as the an "additively closed family" and his theorem (pp 245) indicates that

$$\begin{aligned}&\left|\Delta(\theta)\right|=\left|\Delta\left(\theta^*\right)\right|,\qquad \Pr(\Delta_1=i)=\Pr(\Delta_1^*=i)>0,\\&\mu_{\langle i\rangle,1}=\mu^*_{\langle i\rangle,1},\qquad \sigma_{\langle i\rangle,1}=\sigma^*_{\langle i\rangle,1}\end{aligned} \qquad \text{(AppI.15)}$$

With covariates, according to the non-existence of (AppI.12), (AppI.15) is also true.

Similarly to (AppI.13), integrating both side of Equation (AppI.9) for $y_3,\cdots,y_T$, respectively, yields,

$$\begin{aligned}&\sum_{i=1}^{2}\sum_{j=1}^{2}\Pr(\Delta_1=i)\bullet p_{\langle ij\rangle,2}\bullet\phi\left(y_1\left|\mu_{\langle i\rangle,1},\sigma^2_{\langle i\rangle,1}\right.\right)\bullet\phi\left(y_2\left|\mu_{\langle j\rangle,2},\sigma^2_{\langle j\rangle,2}\right.\right)\\&=\sum_{i=1}^{2}\sum_{j=1}^{2}\Pr(\Delta_1^*=i)\bullet p^*_{\langle ij\rangle,2}\bullet\phi\left(y_1\left|\mu^*_{\langle i\rangle,1},\sigma^{*2}_{\langle i\rangle,1}\right.\right)\bullet\phi\left(y_2\left|\mu^*_{\langle j\rangle,2},\sigma^{*2}_{\langle j\rangle,2}\right.\right),\end{aligned} \qquad \text{(AppI.16)}$$

for all $\left(y_1,y_2\right)\in\mathcal{Y}_1\times\mathcal{Y}_2$. Without covariates, so that the parameter permutation rule is imposed directly on the mean and/or standard deviation of the mixture normal distribution,

Teicher's (1963) result indicates that

$$\left|\Delta(\boldsymbol{\theta})\right| = \left|\Delta\left(\boldsymbol{\theta}^{*}\right)\right|, \qquad \Pr(\Delta_1 = i)\bullet p_{\langle ij\rangle,2} = \Pr(\Delta_1^{*} = i)\bullet p^{*}_{\langle ij\rangle,2},$$
$$\mu_{\langle i\rangle,2} = \mu^{*}_{\langle i\rangle,2}, \qquad \sigma_{\langle i\rangle,2} = \sigma^{*}_{\langle i\rangle,2}, \tag{AppI.17}$$

With covariates, according to the non-existence of (AppI.12), (AppI.17) is also true.

For each $t > 2$, integrating both size of Equation (AppI.9) for $y_1, \cdots, y_{t-1}$, $y_{t+1}, \cdots, y_T$ yields

$$\sum_{i=1}^{2}\left(\sum_{j=1}^{2}\cdots\sum_{\Delta t-1=1}^{2}\Pr(\Delta_1 = j)\bullet p_{\langle \Delta t-1\, i\rangle,t}\right)\bullet\phi\left(y_t\left|\mu_{\langle i\rangle,t},\sigma^{2}_{\langle i\rangle,t}\right.\right)$$
$$= \sum_{i=1}^{2}\left(\sum_{j=1}^{2}\cdots\sum_{\Delta t-1=1}^{2}\Pr(\Delta_1^{*} = j)\bullet p^{*}_{\langle \Delta t-1\, i\rangle,t}\right)\bullet\phi\left(y_t\left|\mu^{*}_{\langle i\rangle,t},\sigma^{*2}_{\langle i\rangle,t}\right.\right) \tag{AppI.18}$$

for all $y_t \in \mathcal{Y}_t$. According to the non-existence of (AppI.12), Equation (AppI.18) also indicates that

$$\sum_{j=1}^{2}\cdots\sum_{\Delta t-1=1}^{2}\Delta_{imi}\left\langle j\right\rangle p_{\langle \Delta t-1\, i\rangle,t} = \sum_{j=1}^{2}\cdots\sum_{\Delta t-1=1}^{2}\Delta^{*}_{imi}\left\langle j\right\rangle p^{*}_{\langle \Delta t-1\, i\rangle,t} \tag{AppI.19}$$

for $i, j = 1, \cdots, \left|\Delta\right|;\ t = 2, \cdots, T$

$$\left|\Delta(\boldsymbol{\theta})\right| = \left|\Delta\left(\boldsymbol{\theta}^{*}\right)\right|, \qquad \mu_{\langle i\rangle,t} = \mu^{*}_{\langle i\rangle,t}, \qquad \sigma_{\langle i\rangle,t} = \sigma^{*}_{\langle i\rangle,t}, \qquad \text{for } t = 2, \cdots, T \tag{AppI.20}$$

According to Equation (AppI.15) , (AppI.17) and (AppI.19), we can recursively conclude that

$$p_{\langle ij\rangle,t} = p^{*}_{\langle ij\rangle,t}; \quad \text{for } i, j = 1,2;\ t = 2, \cdots, T \tag{AppI.21}$$

Therefore,

$$LOGIT\left(\mathbf{w}_t\boldsymbol{\varphi}_{\langle i\rangle}\right) = LOGIT\left(\mathbf{w}_t\boldsymbol{\varphi}^{*}_{\langle i\rangle}\right); \quad i = 1,2; t = 2, \cdots, T \tag{AppI.22}$$

and because of the monotonic property of the logit function,

$$\mathbf{w}_t \boldsymbol{\varphi}_{\langle i \rangle} = \mathbf{w}_t \boldsymbol{\varphi}^*_{\langle i \rangle}; i = 1,2;\ t = 2,\cdots,T. \qquad \text{(AppI.23)}$$

This is equivalent to

$$\left(\boldsymbol{\varphi}_{\langle i \rangle} - \boldsymbol{\varphi}^*_{\langle i \rangle}\right)' \mathbf{w}_t = 0, \text{ for } i = 1,2;\ t = 2,\cdots,T,$$

or to

$$\left(\boldsymbol{\varphi}_{\langle i \rangle} - \boldsymbol{\varphi}^*_{\langle i \rangle}\right)' \mathbf{W}_T = 0; \quad \text{for } i = 1,2. \qquad \text{(AppI.24)}$$

Thus, a sufficient condition for $\boldsymbol{\varphi} = \boldsymbol{\varphi}^*$ is $\mathbf{W}_T$ is full rank. Similarly, the full rank of $\mathbf{C}_T$ and $\mathbf{Z}_T$ guarantee that $\boldsymbol{\alpha} = \boldsymbol{\alpha}^*$ and $\boldsymbol{\beta} = \boldsymbol{\beta}^*$ respectively.

Because $\Pr(\Delta_1 = i) = \Pr(\Delta_1^* = i)$, the sufficient identifiability condition is proved. In case the given sufficient condition is violated, we need to reparameterize the model accordingly. In practice, if the value of the covariate is continuous, the two-state Markov-switching model is usually identifiable.

APPENDIX II

This appendix discusses the mathematical details of the forward and backward recursive program at the E step of the EM algorithm.

The E-step of the EM algorithm is an important step of the mapping from $\theta^l \to \theta^{l+1}$. The following forward and backward recursive program will be adopted. Note that for convenience of notation, we will omit most of the iteration subscripts.

Define

$$p_{\langle ii\rangle,t} = p_{\langle ii\rangle}\left(\mathbf{w}_t, \widehat{\boldsymbol{\varphi}_{\langle i\rangle}^{(l)}}\right) \quad \text{for} \quad i=1,2;\ t=2,\cdots,T,$$

$$1-p_{\langle ij\rangle,t} = 1-p_{\langle ij\rangle}\left(\mathbf{w}_t, \widehat{\boldsymbol{\varphi}_{\langle i\rangle}^{(l)}}\right) \quad \text{for} \quad i=1,2 \ \text{ and } \ i\neq j,$$

$$f_{\langle i\rangle,t} = f_{\langle i\rangle}\left(y_t \,|\, \mathbf{c}_t,\ \mathbf{z}_t,\ \widehat{\boldsymbol{\vartheta}_{\langle i\rangle}^{(l)}}\right) \quad \text{for} \quad i=1,2 \ \text{ and } \ t=1,\cdots,T,$$

$$\begin{aligned}\widehat{s_{\langle i\rangle,t}} &= f\left(\Delta_t = i \middle| \mathbf{X}_\mathbf{T}, \mathbf{Y}_\mathbf{T}\right) \\ &= \frac{f\left(\mathbf{Y}_\mathbf{T}, \Delta_t = i \middle| \mathbf{X}_\mathbf{T}\right)}{\Pr f\left(\mathbf{Y}_\mathbf{T} \middle| \mathbf{X}_\mathbf{T}\right)} \\ &= \frac{f\left(\mathbf{Y}_\mathbf{T}, \Delta_t = i \middle| \mathbf{X}_\mathbf{T}\right)}{\sum_{i=1}^{2} f\left(\mathbf{Y}_\mathbf{T}, \Delta_t = i \middle| \mathbf{X}_\mathbf{T}\right)},\end{aligned}$$

$$\widehat{\xi}_{\langle ij\rangle,t} = \Pr\left(\Delta_{t-1}=i,\Delta_t=j\middle|\mathbf{X_T},\mathbf{Y_T}\right)$$
$$= \frac{f\left(\mathbf{Y_T},\Delta_{t-1}=i,\Delta_t=j\middle|\mathbf{X_T}\right)}{f\left(\mathbf{Y_T}\middle|\mathbf{X_T}\right)},$$

where

$$f\left(\mathbf{Y_T},\Delta_{t-1}=i,\Delta_t=j\middle|\mathbf{X_T}\right)$$
$$= f\left(y_1,\ldots,y_{t-1},\Delta_{t-1}=i\middle|\mathbf{X_T}\right)\cdot f\left(y_t,\ldots,y_T,\Delta_t=j\,|\,y_1,\ldots,y_{t-1},\Delta_{t-1}=i\middle|\mathbf{X_T}\right)$$
$$= \Pr\left(y_1,\ldots,y_{t-1},\Delta_{t-1}=i\middle|\mathbf{X_T}\right)\cdot f\left(y_t,\ldots,y_T,\Delta_t=j\,|\,\Delta_{t-1}=i,\mathbf{X_T}\right)$$
$$= f\left(y_1,\ldots,y_{t-1},\Delta_{t-1}=i\middle|\mathbf{X_T}\right)\cdot\Pr\left(\Delta_t=j\middle|\Delta_{t-1}=i\right)\cdot f\left(y_t,\ldots,y_T\middle|\Delta_{t-1}=i,\Delta_t=j,\mathbf{X_T}\right)$$
$$= f\left(y_1,\ldots,y_{t-1},\Delta_{t-1}=i\middle|\mathbf{X_T}\right)\cdot p_{\langle ij\rangle,t}\cdot f\left(y_t,\ldots,y_T\middle|\Delta_t=j,\mathbf{X_T}\right)$$
$$= f\left(y_1,\ldots,y_{t-1},\Delta_{t-1}=i\middle|\mathbf{X_T}\right)\cdot p_{\langle ij\rangle,t}\cdot f_{\langle j\rangle,t}\cdot f\left(y_{t+1},\ldots,y_T\middle|\Delta_t=j,\mathbf{X_T}\right),$$

We can therefore express the estimated state transition probability as:

$$\widehat{\xi}_{\langle ij\rangle,t} = \frac{f\left(y_1,\ldots,y_{t-1},\Delta_{t-1}=i\middle|\mathbf{X}_T\right)\cdot p_{\langle ij\rangle,t}\cdot f_{\langle j\rangle,t}\cdot f\left(y_{t+1},\ldots,y_T\,|\,\Delta_t=j,\mathbf{X}_T\right)}{f\left(\mathbf{Y}_T\middle|\mathbf{X}_T\right)}.$$

Using the following forward and backward recursive formula, we define:

$$A_{\langle i\rangle,t} = f\left(y_1,\ldots,y_t,\Delta_t=i\middle|\mathbf{X_T}\right), \qquad \text{for } t=2,\cdots,T, i=1,2,$$
$$A_{\langle i\rangle,1} = f\left(\Delta_1=i\right)\cdot f_{\langle i\rangle,t}, \qquad \text{for } i=1,2,$$
$$B_{\langle i\rangle,t} = f\left(y_{t+1},\ldots,y_T\middle|\Delta_t=i,\mathbf{X_T}\right), \qquad \text{for } t=1,2,\cdots,T-1,\ i=1,2,$$
$$B_{\langle i\rangle,T} = 1 \qquad \text{for } i=1,2.$$

We can therefore express the estimated state transition probability as:

$$\widehat{\xi}_{\langle ij\rangle,t} = \frac{f\left(y_1,\ldots,y_{t-1},\Delta_{t-1}=i\middle|\mathbf{X_T}\right)\cdot p_{\langle ij\rangle,t}\cdot f_{\langle j\rangle,t}\cdot f\left(y_{t+1},\ldots,y_T\,|\,\Delta_t=j\middle|\mathbf{X_T}\right)}{f\left(y_1,\ldots,y_T\middle|\mathbf{X_T}\right)}$$
$$= \frac{A_{\langle i\rangle,t-1}\cdot p_{\langle ij\rangle,t}\cdot f_{\langle j\rangle,t}\cdot B_{\langle j\rangle,t}}{A_{\langle 1\rangle,T}+A_{\langle 2\rangle,T}},$$

and the estimated probability that y_t is generated by state i is:

$$\widehat{s}_{\langle i\rangle,t} = \sum_{j=1}^{2}\widehat{\xi}_{\langle ij\rangle,t}.$$

With above setup, the following is the detailed recursive formula to compute $a_i(t)$ and $b_i(t)$.

For the forward recursive program, we have:

$$\begin{aligned}
A_{\langle j\rangle,t} &= f\left(y_1,...,y_t,\Delta_t = j \middle| \mathbf{X_T}\right) \\
&= \sum_{i=1}^{2} f\left(y_1,...,y_t,\Delta_t = j,\Delta_{t-1} = i \middle| \mathbf{X_T}\right) \\
&= \sum_{i=1}^{2} f\left(y_1,...,y_t,\Delta_t = j,\Delta_{t-1} = i \middle| \mathbf{X_T}\right) \bullet f\left(y_t,\Delta_t = j \mid y_1,...,y_{t-1},\Delta_{t-1} = i,\mathbf{X_T}\right) \\
&= \sum_{i=1}^{2} f\left(y_1,...,y_t,\Delta_t = j,\Delta_{t-1} = i \middle| \mathbf{X_T}\right) \bullet f\left(y_t,\Delta_t = j \mid \Delta_{t-1} = i,\mathbf{X_T}\right) \\
&= \sum_{i=1}^{2} f\left(y_1,...,y_t,\Delta_t = j,\Delta_{t-1} = i \middle| \mathbf{X}_T\right) \bullet f\left(\Delta_t = j \mid \Delta_{t-1} = i\right) \bullet \Pr\left(y_t \mid \Delta_{t-1} = i,\Delta_t = j,\mathbf{X}_T\right) \\
&= \sum_{i=1}^{2} f\left(y_1,...,y_{t-1},\Delta_{t-1} = i \middle| \mathbf{X_T}\right) \bullet p_{\langle ij\rangle,t} \bullet f\left(y_t \mid \Delta_t = j,\mathbf{X_T}\right) \\
&= \sum_{i=1}^{2} A_{\langle i\rangle,t-1} \bullet p_{\langle ij\rangle,t} \bullet f_{\langle j\rangle,t},
\end{aligned}$$

and

$$A_{\langle i\rangle,t} = \left[\sum_{j=1}^{2} A_{\langle j\rangle,t-1} \bullet p_{\langle ji\rangle,t}\right] \bullet f_{\langle i\rangle,t}.$$

For the backward recursive program, we have:

$$\begin{aligned}
B_{\langle i\rangle,t} &= f\left(y_{t+1},...,y_T \mid \Delta_t = i,\mathbf{X_T}\right) \\
&= \sum_{j=1}^{2} \Pr\left(y_{t+1},y_{t+2},...,y_T,\Delta_{t+1} = j \mid \Delta_t = i,\mathbf{X_T}\right) \\
&= \sum_{j=1}^{2} \Pr\left(y_{t+1} \mid \Delta_t = i,\mathbf{X_T}\right) \bullet f\left(y_{t+2},...,y_T,\Delta_{t+1} = j \mid y_{t+1},\Delta_t = i,\mathbf{X_T}\right) \\
&= \sum_{j=1}^{2} \Pr\left(y_{t+1} \mid \Delta_t = i,\mathbf{X_T}\right) \bullet f\left(\Delta_{t+1} = j \mid y_{t+1},\Delta_t = i,\mathbf{X_T}\right) \\
&\qquad \bullet f\left(y_{t+2},...,y_T \mid y_{t+1},\Delta_t = i,\Delta_{t+1} = j,\mathbf{X_T}\right),
\end{aligned}$$

where

$$
\begin{aligned}
& f\left(y_{t+1} \mid \Delta_t = i, \mathbf{X_T}\right) \bullet \Pr\left(\Delta_{t+1} = j \mid y_{t+1}, \Delta_t = i, \mathbf{X_T}\right) \\
&= f\left(y_{t+1} \mid \Delta_t = i, \mathbf{X_T}\right) \frac{f\left(\Delta_{t+1} = j, y_{t+1}, \Delta_t = i \middle| \mathbf{X_T}\right)}{f\left(y_{t+1}, \Delta_t = i \middle| \mathbf{X_T}\right)} \\
&= \frac{f\left(\Delta_t = j, y_{t+1}, \Delta_t = i \middle| \mathbf{X_T}\right)}{\Pr\left(\Delta_t = i \middle| \mathbf{Y_T}, \mathbf{X_T}\right)} \\
&= \frac{\Pr\left(\Delta_t = i \middle| \mathbf{Y_T}, \mathbf{X_T}\right) \bullet f\left(\Delta_t = j, y_{t+1} \mid \Delta_t = i, \mathbf{X_T}\right)}{\Pr\left(\Delta_t = i \middle| \mathbf{Y_T}, \mathbf{X_T}\right)} \\
&= \Pr\left(\Delta_{t+1} = j \mid \Delta_t = i\right) \bullet f\left(y_{t+1} \mid \Delta_t = i, \Delta_{t+1} = j, \mathbf{X_T}\right) \\
&= p_{\langle ij \rangle, t} \bullet f_{\langle j \rangle, t+1},
\end{aligned}
$$

therefore

$$
\begin{aligned}
& \sum_{j=1}^{2} f\left(y_{t+1} \middle| \Delta_t = i, \mathbf{X_T}\right) \bullet \Pr\left(\Delta_{t+1} = j \middle| y_{t+1}, \Delta_t = i, \mathbf{X_T}\right) \bullet f\left(y_{t+2}, \ldots, y_T \middle| y_{t+1}, \Delta_t = i, \Delta_{t+1} = j, \mathbf{X_T}\right) \\
&= \sum_{j=1}^{2} p_{\langle ij \rangle, t} \bullet f_{\langle j \rangle, t+1} \bullet f\left(y_{t+2}, \ldots, y_T \middle| \Delta_{t+1} = j, \mathbf{X_T}\right) \\
&= \sum_{j=1}^{2} p_{\langle ij \rangle, t} \bullet f_{\langle j \rangle, t+1} \bullet B_{\langle j \rangle, t+1},
\end{aligned}
$$

and

$$
B_{\langle i \rangle, t} = \sum_{j=1}^{2} p_{\langle ij \rangle, t} \bullet f_{\langle j \rangle, t+1} \bullet B_{\langle j \rangle, t+1}.
$$

In addition, the likelihood function for the observable data is the by-product of the forward program:

$$
\begin{aligned}
LLF_{obs}(\boldsymbol{\theta}) &= \Pr\left(\mathbf{Y_T} \middle| \mathbf{X_T}, \boldsymbol{\theta}\right) \\
&= \Pr\left(\mathbf{Y_T}, \Delta_T = 1 \middle| \mathbf{X_T}, \boldsymbol{\theta}\right) + \Pr\left(\mathbf{Y_T}, \Delta_T = 2 \middle| \mathbf{X_T}, \boldsymbol{\theta}\right) \\
&= A_{\langle 1 \rangle, T} + A_{\langle 2 \rangle, T}.
\end{aligned}
$$

APPENDIX III

The appendix discusses the possibility of both numerical overflow and underflow of the forward and backward programming at the E state of the EM algorithm.

Consider a time series, $\mathbf{Y}_{\mathbf{T}}^{\diamond}$, whose data generating processes is defined by the HMS-V model. Define another time series $\mathbf{Y}_{\mathbf{T}}^{\spadesuit} \equiv G\mathbf{Y}_{\mathbf{T}}^{\diamond}$, where G is a positive scaler. Consequently, Let their parameters be

$$\boldsymbol{\theta}^{\diamond} = \left(\boldsymbol{\varphi}_{\langle i\rangle}^{\diamond}, \boldsymbol{\alpha}_{\langle i\rangle}^{\diamond}, \boldsymbol{\beta}_{\langle i\rangle}^{\diamond}, \Delta_{ini}^{\diamond}\langle 1\rangle\right)', \quad \boldsymbol{\theta}^{\spadesuit} = \left(\boldsymbol{\varphi}_{\langle i\rangle}^{\spadesuit}, \boldsymbol{\alpha}_{\langle i\rangle}^{\spadesuit}, \boldsymbol{\beta}_{\langle i\rangle}^{\spadesuit}, \Delta_{ini}^{\spadesuit}\langle 1\rangle\right)'$$

$$\boldsymbol{\beta}_{\langle i\rangle}^{\diamond} = \left(\boldsymbol{\beta}_{\langle i\rangle,0}^{\diamond}, \boldsymbol{\beta}_{\langle i\rangle,1}^{\diamond}, \cdots, \boldsymbol{\beta}_{\langle i\rangle,k}^{\diamond}\right)', \quad \boldsymbol{\beta}_{\langle i\rangle}^{\spadesuit} = \left(\boldsymbol{\beta}_{\langle i\rangle,0}^{\spadesuit}, \boldsymbol{\beta}_{\langle i\rangle,1}^{\spadesuit}, \cdots, \boldsymbol{\beta}_{\langle i\rangle,k}^{\spadesuit}\right) \qquad (i = 1,2).$$

Then

$$\Pr(\Delta_1^{\diamond} = 1) = \Pr(\Delta_1^{\spadesuit} = 1), \quad \boldsymbol{\varphi}_{\langle i\rangle}^{\diamond} = \boldsymbol{\varphi}_{\langle i\rangle}^{\spadesuit}, \quad \left(\boldsymbol{\beta}_{\langle i\rangle,1}^{\diamond}, \cdots, \boldsymbol{\beta}_{\langle i\rangle,k}^{\diamond}\right)' = \left(\boldsymbol{\beta}_{\langle i\rangle,1}^{\spadesuit}, \cdots, \boldsymbol{\beta}_{\langle i\rangle,k}^{\spadesuit}\right)'$$

and

$$\boldsymbol{\alpha}_{\langle i\rangle}^{\diamond} = G\boldsymbol{\alpha}_{\langle i\rangle}^{\spadesuit}, \quad \boldsymbol{\beta}_{\langle i\rangle,0}^{\diamond} = \boldsymbol{\beta}_{\langle i\rangle,0}^{\diamond} + \log G.$$

Given their parameter estimates at the l^{th} step, $\boldsymbol{\theta}^{\diamond(l)}$ and $\boldsymbol{\theta}^{\spadesuit(l)}$, of which $\boldsymbol{\alpha}_{\langle i\rangle}^{\diamond(l)} = G\boldsymbol{\alpha}_{\langle i\rangle}^{\spadesuit(l)}$, and $\boldsymbol{\beta}_{\langle i\rangle,0}^{\diamond(l)} = \boldsymbol{\beta}_{\langle i\rangle,0}^{\spadesuit(l)} + \log G$, if we implement the E-step of both time series $\mathbf{Y}_T^{\diamond}$ and $\mathbf{Y}_T^{\spadesuit}$ in exactly the same way, then

$$p_{\langle ij\rangle,t}^{\diamond(l)} = p_{\langle ij\rangle,t}^{\spadesuit(l)}, \qquad i,j = 1,2;\ t = 2,\cdots,T$$

$$f_{\langle i\rangle}^{\diamond}\left(y_t \mid \mathbf{c}_t, \mathbf{z}_t, \boldsymbol{\vartheta}_{\langle i\rangle}^{\diamond(l)}\right) = f_{\langle i\rangle}^{\spadesuit}\left(y_t \mid \mathbf{c}_t, \mathbf{z}_t, \boldsymbol{\vartheta}_{\langle i\rangle}^{\spadesuit(l)}\right) \bullet \log G, \qquad i,j = 1,2;\ t = 1,\cdots,T$$

$$A_{\langle i\rangle,t}^{\diamond} = A_{\langle i\rangle,t}^{\spadesuit} \bullet \left(\log G\right)^t,$$

$$B^{\diamond}_{\langle i\rangle,t} = B^{\spadesuit}_{\langle i\rangle,t} \bullet \left(\log G\right)^{T-t},$$

$\hat{s}^{\diamond(l+1)}_{T} = \hat{s}^{\spadesuit(l+1)}_{T}$, and $\hat{\mathbf{s}}^{\diamond(l+1)}_{T} = \hat{\mathbf{s}}^{\spadesuit(l+1)}_{T}$.

Therefore, for a *continuous* and *scalable* distribution, given a value G very close to zero, the forward-backward probabilities of $\mathbf{Y}^{\spadesuit}_{T}$ will overflow, which corresponds to a positive log likelihood function.

The difficulty of the above recursive algorithm is that as t grows, a and b may converge either to 0 or diverge to infinity very quickly (depends on the value of $f_{\langle i\rangle,t}$). Therefore we need to rescale the forward and backward probabilities; otherwise the system will either underflow or overflow. We express the value of N by the following vector,

$N \rightarrow \{num, power\}$

so that

$N = num * \exp(-power), \qquad 1 \le num < e,$

The transformation converts a lot of the unnecessary power calculation (including the density function of normal distribution) into addition and subtraction. We suggest rescaling all the values of $f_{\langle i\rangle,t}$, $A_{\langle i\rangle,t}$, $B_{\langle i\rangle,t}$ and some other relevant functions if necessary.

APPENDIX IV

This appendix gives the closed form Gradient and Hessian of the ***FIRST*** sub-M step of the EM algorithm for the HMS-V model. The second sub-M step follows accordingly.

Given the estimate of the state probabilities, $\hat{s}^{(l+1)}_{\langle i\rangle,t}$, $i=1$, the sub-M step will obtain the $\widehat{\varphi^{(l+1)}_{\langle 1\rangle}}$. For visual clarity of notation, we drop most of the state and iteration subscripts.

The log likelihood function of the Logit link is

$$LLF_1^{(l+1)}\left(\varphi_{\langle 1\rangle}\right)=\sum_{t=2}^{T}\Big[\hat{s}^{(l+1)}_{\langle 11\rangle,t}\log p_{\langle 11\rangle}\left(\mathbf{w}_t,\varphi_{\langle 1\rangle}\right) + \hat{s}^{(l+1)}_{\langle 12\rangle,t}\log\left(1-p_{\langle 11\rangle}(\mathbf{w}_t,\varphi_{\langle 1\rangle})\right)\Big],$$

drop some subscripts, and it simplifies to

$$LLF\left(\varphi\right)=\sum_{t=2}^{T}\Big[\widehat{s_{\langle 11\rangle,t}}\log LOGIT\left(\mathbf{w}_t\varphi\right) + \widehat{s_{\langle 12\rangle,t}}\log\left(1-LOGIT\left(\mathbf{w}_t\varphi\right)\right)\Big].$$

Let $\sout{EXP}_t=\exp\left(\mathbf{w}_t\varphi\right)$; $t=2,3,\cdots,T$

then $p_{\langle 11\rangle,t}=\dfrac{\sout{EXP}_t}{1+\sout{EXP}_t}$.

The first order derivative of LLF_t w.r.t. φ can be obtained by chain rule,

$$\frac{\partial LLF_t}{\partial p_{\langle 11\rangle,t}}=\frac{\widehat{s_{\langle 11\rangle,t}}}{p_{\langle 11\rangle,t}}-\frac{\widehat{s_{\langle 12\rangle,t}}}{1-p_{\langle 11\rangle,t}};\qquad t=2,3,\cdots,T$$

$$\frac{\partial p_{\langle 11\rangle,t}}{\partial \sout{EXP}_t}=\frac{1}{\left(1+\sout{EXP}_t\right)^2};\qquad t=2,3,\cdots,T$$

$$\frac{\partial \cancel{EXP}_t}{\partial \boldsymbol{\varphi}} = \cancel{EXP}_t \mathbf{w}_t'; \qquad t = 2,3,\cdots,T,$$

and the Gradient is

$$\frac{\partial LLF}{\partial \boldsymbol{\varphi}} = \sum_{t=2}^{T} \frac{\partial LLF_t}{\partial p_{\langle 11\rangle,t}} \frac{\partial \partial p_{\langle 11\rangle,t}}{\partial \cancel{EXP}_t} \frac{\partial \cancel{EXP}_t}{\partial \boldsymbol{\varphi}}.$$

With the closed form Gradient, we apply the chain rule to obtain the second order derivative of LLF_t *w.r.t.* $\boldsymbol{\varphi}$,

$$\frac{\partial^2 LLF_t}{\partial \left(p_{\langle 11\rangle,t}\right)^2} = -\left(\frac{\widehat{\mathbb{S}_{\langle 11\rangle,t}}}{\left(p_{\langle 11\rangle,t}\right)^2} + \frac{\widehat{\mathbb{S}_{\langle 12\rangle,t}}}{\left(1-p_{\langle 11\rangle,t}\right)^2} \right)$$

thus the Hessian is

$$\frac{\partial^2 LLF}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}'} = \sum_{t=2}^{T} \left(\frac{\partial^2 LLF_t}{\partial \left(p_{\langle 11\rangle,t}\right)^2} \right) \left(\frac{\partial p_{\langle 11\rangle,t}}{\partial \cancel{EXP}_t} \right)^2 \left(\frac{\partial \cancel{EXP}_t}{\partial \boldsymbol{\varphi}} \right) \left(\frac{\partial \cancel{EXP}_t}{\partial \boldsymbol{\varphi}} \right)'.$$

APPENDIX V

This appendix gives the closed form Gradient and Hessian of the ***THIRD*** sub-M steps of the EM algorithm for the HMS-V model. The forth sub-M step follows accordingly.

Given the estimate of the state probabilities, $\hat{s}_{\langle i\rangle,t}^{(l+1)}$, $i=1$, the sub-M step will obtain $\widehat{\vartheta^{(l+1)}_{\langle 1\rangle}} = \left\{\widehat{\alpha^{(l+1)}_{\langle 1\rangle}}, \widehat{\beta^{(l+1)}_{\langle 1\rangle}}\right\}$. For visual clarity of notation, we drop most of the state and iteration subscripts.

At time t, given that y_t resides at State i=1, then

$y_t = \mu_{\langle i\rangle,t} + \varepsilon_t$,

$\varepsilon_t = \sigma_{\langle i\rangle,t} \bullet \eta_t$ $\qquad$ $\eta_t \overset{i.i.d}{\sim} N(0,1)$,

$\mu_{\langle i\rangle,t} = \mathbf{c}_t \boldsymbol{\alpha}_{\langle i\rangle}$,

$\sigma_{\langle i\rangle,t} = \exp\left(\mathbf{z}_t \boldsymbol{\beta}_{\langle i\rangle}\right)$, and

$$LLF_3\left(\vartheta_{\langle i\rangle}\right) = -\frac{\sum_{t=1}^{T} \widehat{s_{\langle i\rangle,t}}}{2}\log(2\pi) - \sum_{t=1}^{T} \widehat{s_{\langle i\rangle,t}}\left(\log\left(\sigma_{\langle i\rangle,t}\right) + \frac{(y_t - \mu_{\langle i\rangle,t})^2}{2\sigma^2_{\langle i\rangle,t}}\right).$$

Drop the state subscript i, we have

$$\begin{aligned} LLF(\vartheta) &= \sum_{t=1}^{T} LLF_t(\vartheta) \\ &= \sum_{t=1}^{T} -\frac{1}{2}\widehat{s}_t \log(2\pi) - \widehat{s}_t\left(\log(\sigma_t) + \frac{(y_t - \mu_t)^2}{2\sigma_t^2}\right). \end{aligned}$$

The first order derivative of LLF_t w.r.t. $\vartheta = \{\alpha, \beta\}$ can be obtained by chain rule,

$$\frac{\partial LLF_t}{\partial \mu_t} = \hat{s}_t \frac{y_t - \mu_t}{\sigma_t^2}$$

$$\frac{\partial LLF_t}{\partial \sigma_t} = \hat{s}_t \frac{-\sigma_t^2 + (y_t - \mu_t)^2}{\sigma_t^3}$$

$\frac{\partial \mu_t}{\partial \alpha} = \mathbf{c}_t'$, and

$\frac{\partial \mu_t}{\partial \beta} = \sigma_t \mathbf{z}_t'$.

The Gradient is

$$\frac{\partial LLF}{\partial \alpha} = \sum_{t=1}^{t=T} \hat{s}_t \frac{y_t - \mu_t}{\sigma_t^2} \mathbf{c}_t',$$

$$\frac{\partial LLF}{\partial \beta} = \sum_{t=1}^{t=T} \hat{s}_t \frac{-\sigma_t^2 + (y_t - \mu_t)^2}{\sigma_t^2} \mathbf{z}_t',$$

$$\frac{\partial LLF}{\partial \vartheta} = \left[\frac{\partial LLF}{\partial \alpha}, \frac{\partial LLF}{\partial \beta} \right]'.$$

The Hessian equals to

$$\begin{aligned} \frac{\partial^2 LLF_t}{\partial \alpha \partial \alpha'} &= \frac{\partial}{\partial \alpha'} \frac{\partial LLF_t}{\partial \alpha} \\ &= \frac{\partial}{\partial \alpha'} \left(\hat{s}_t \frac{y_t - \mu_t}{\sigma_t^2} \mathbf{c}_t' \right) \\ &= -\frac{\hat{s}_t}{\sigma_t^2} \mathbf{c}_t' \mathbf{c}_t, \end{aligned}$$

$$\frac{\partial^2 LLF_t}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}=\frac{\partial}{\partial\boldsymbol{\beta}'}\frac{\partial LLF_t}{\partial\boldsymbol{\beta}}=\frac{\partial}{\partial\boldsymbol{\beta}'}\left(\widehat{s}_t\frac{-\sigma_t^2+(y_t-\mu_t)^2}{\sigma_t^2}\mathbf{z}_t'\right)$$

$$=\frac{\partial}{\partial\boldsymbol{\beta}'}\left(\widehat{s}_t(y_t-\mu_t)^2\mathbf{z}_t'\sigma_t^{-2}\right)$$

$$=-2\widehat{s}_t(y_t-\mu_t)^2\mathbf{z}_t'\sigma_t^{-3}\bullet\sigma_t\mathbf{z}_t$$

$$=-2\widehat{s}_t(y_t-\mu_t)^2\sigma_t^{-2}\mathbf{z}_t'\mathbf{z}_t,$$

$$\frac{\partial^2 LLF_t}{\partial\boldsymbol{\beta}\partial\boldsymbol{\alpha}'}=\frac{\partial}{\partial\boldsymbol{\alpha}'}\frac{\partial LLF_t}{\partial\boldsymbol{\beta}}=\frac{\partial}{\partial\boldsymbol{\alpha}'}\left(\widehat{s}_t\frac{-\sigma_t^2+(y_t-\mu_t)^2}{\sigma_t^2}\mathbf{z}_t'\right)$$

$$=\frac{\partial}{\partial\boldsymbol{\alpha}'}\left(\widehat{s}_t\sigma_t^{-2}\mathbf{z}_t'(y_t-\mu_t)^2\right)$$

$$=-2\widehat{s}_t\sigma_t^{-2}(y_t-\mu_t)\mathbf{z}_t'\mathbf{c}_t,$$

$$\frac{\partial^2 LLF_t}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\beta}'}=-2\widehat{s}_t\sigma_t^{-2}(y_t-\mu_t)\mathbf{c}_t'\mathbf{z}_t,$$

and $$\frac{\partial^2 LLF_{com}}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}'}=\begin{bmatrix}\sum_{t=1}^{T}\frac{\partial^2 LLF_t}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\alpha}'} & \sum_{t=1}^{T}\frac{\partial^2 LLF_t}{\partial\boldsymbol{\alpha}\partial\boldsymbol{\beta}'}\\ \sum_{t=1}^{T}\frac{\partial^2 LLF_t}{\partial\boldsymbol{\beta}\partial\boldsymbol{\alpha}'} & \sum_{t=1}^{T}\frac{\partial^2 LLF_t}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\end{bmatrix}.$$

The information matrix follows accordingly.

APPENDIX VI

This appendix briefly discusses the choice of the starting values for the EM algorithm.

Despite of its relative stability and simplicity, a set of poorly chosen starting values may still spoil the EM algorithm, which increases the number of iterations and reduce the chance of converging to the global optimum.

With fixed covariate, Wang and Puterman (1999a) discuss the choice of starting values for a Markov-switching Poisson regression in details. Due to the property of financial time series, the thesis primary adopts the "Just Added" strategy. The principle idea is to start with a simpler model with no covariate, using the optimized parameter values, parsimoniously add new columns of the covariate, and optimize again. We also skip the technical details since it is also remotely related to the core idea of the thesis.

APPENDIX VII

This appendix briefly discusses if the Gradient and Hessian of the log likelihood function of the observed data are both continuous and bounded. It helps to explain the stability of the numerically obtained Hessian.

Section 4 suggests using the EM algorithm to provide comfortable starting values, and employ the quasi-Newton algorithm to obtain the MLE. Most software packages routinely calculate the "true" Hessian by numerical difference once the quasi-Newton converges. For example, after optimization, the MatLab 7.0 Optimization Toolbox automatically perturbs each element of the optimized parameter vector by a tiny amount in turns, and obtains the corresponding increment on the log likelihood. Using only the one-sided approximation of the Taylor series expansion, it calculates the second order derivatives. We can then obtain the numerical approximation of the information matrix and the standard errors of the MLE accordingly. Note that the default perturbation value for any coefficient by the optimization toolbox of MatLab 7.0 is fixed at 10^{-10} irrespective to its value. Therefore, when the relative magnitudes of the parameters differ greatly, the estimated information matrix is less precise. We solve such problem by rescaling the data to enhance stability. For example, if a return series has the true parameters, $\alpha_{\langle 1\rangle,0} = 0.002$ and $\alpha_{\langle 1\rangle,1} = 0.8$, we can simply multiply all the element of $\mathbf{Y}_\mathbf{T}$ by 100, and estimated the new time series again, whose true parameters are now more balanced, $\alpha^*_{\langle 1\rangle,0} = 0.2$ and $\alpha^*_{\langle 1\rangle,1} = 0.8$. According to Section 4.2.1, the property of the original data should be well kept. Of course, one can only use the rescaling method for a continuous distribution. Partly because of the above reason, we always use the percentage log return instead of the

log return in practice. According to our empirical study, the dominating majority of the estimated information matrices by the numerical difference method are positive definite and invertible.

APPENDIX XIII

This appendix reviews some literature on the relationship between the range and volatility based on a few key assumptions. One of the key assumptions is that the price is continuously observable within the specified time interval.

Parkinson (1980) makes the following assumptions, and he also concludes in terms of measuring volatility, range carries far more efficient information then the close to close return.

The problem to be solved may be stated as follows: Suppose a point particle undergoes a one-dimensional, continuous random walk with a diffusion constant D. Then, the probability of finding the particle in the interval $(\varepsilon, \varepsilon + d\varepsilon)$ at time t, if it started at point ε_0 at time $t = 0$, is $\left(d\varepsilon / \sqrt{2\pi Dt}\right) \exp\left[-(\varepsilon - \varepsilon_0)^2 / 2\pi Dt\right]$. By comparison with the normal distribution, Parkinson sees that D is the variance of the displacement $\varepsilon - \varepsilon_0$ after a unit time interval. This suggests the traditional way to estimate D: Parkinson measures $\varepsilon(t)$ for $t = 0,1,2,\cdots,n$. Then, defining $d_i =$ displacement during the ith interval, $d_i = \varepsilon(i) - \varepsilon(i-1)$, $i = 1,2,\cdots,n$, then

$$D_\varepsilon = \frac{1}{n-1} \sum_{i=1}^{n} \left(d_i - \bar{d}\right)^2$$

is an estimate for D; where

$$\bar{d} = \frac{1}{n} \sum_{m=1}^{n} d_m = \text{mean displacement}.$$

However, instead of measuring $\varepsilon(n)$, for $n=0,1,2,\cdots$, suppose we have measured only the difference l between the maximum and minimum position during each time interval. These differences should be capable of giving a good estimate for D, for it is intuitively clear that the average difference will get larger or smaller as D gets larger or smaller.

Defining $P(l,t)$ to be the probability that $(\varepsilon_{\max}-\varepsilon_{\min})\le l$ during time interval t, we have

$$P(l,t)=\sum_{n=1}^{\infty}(-1)^{n+1}n\left\{erfc\left[(n+1)l/\sqrt{2Dt}\right]\right.$$
$$\left.-2erfc\left(nl/\sqrt{2Dt}\right)+erfc\left[(n-1)l/\sqrt{2Dt}\right]\right\}$$

where $erfc(\varepsilon)=1-erf(\varepsilon)$ and $erf(\varepsilon)$ is the error function.

Parkinson then proves that

$$E\left[l^2\right]=\left[4\ln(2)\right]Dt \qquad \text{(AppIX.1)}$$

Computing the variance of D_ε and D_l, Parkinson finds that

$$E\left[(D_\varepsilon-D)^2\right]=\left[\frac{E(\varepsilon^4)}{E(\varepsilon^2)}-1\right]\frac{D^2}{N_\varepsilon}=\frac{2D^2}{N_\varepsilon} \qquad \text{(AppIX.2)}$$

and

$$E\left[(D_l-D)^2\right]=\left[\frac{E(l^4)}{E(l^2)}-1\right]\frac{D^2}{N_l}=\frac{0.41D^2}{N_l}, \qquad \text{(AppIX.3)}$$

where N is the number of observations. Thus, to obtain the same amount of variance using the two methods, we need $N_\varsigma\approx 5N_l$. Clearly, the range is far more informative than the close to close return in terms of estimating volatility.

In addition, data on the ranges are widely available for most currencies and futures contracts, and data on the observable ranges are also abundant for individual stocks and a lot of exchange-traded securities, not only at present but over a long historical spans. However, it is only until recently that researchers further develop the insight of Parkinson (1980) and incorporate it with other volatility models, such as the GARCH model. Andersen and Bollerslev (1998), Christoffersen (2003), Alizadeh, Brandt and Diebold (2002), Brandt and Jones (2002), Chou (2005), Brandt and Diebold (2006) and Martens and Dijk (2006), among the others, also conclude that the range is an efficient proxy of the volatility.

APPENDIX IX

This appendix discusses the reason that for some securities, there is missing price information when exchange is closed. We give an example using the stock price of Citigroup Inc (NYSE:C) during the trading day Dec 26, 2007 to Dec 28, 2007.

Note that according to the definition of range by Equation (5.3), the "true" weekly range δ_t is different from the observed intraweek range R_t described by Equation (5.6).

The statement "from time t to time $t+1$" strictly imposes the sampling period. For example, in order to get the daily range according to the definition, we can collect the prices from a precise time point to exactly the same time point 24 hours later. For currencies or future contracts traded around the clock, the fluctuation of price is usually available during the defined time interval, and the observed ranges are basically inline with the fundamental definition. But for securities traded only during specific hours in a day, it is complicated.

Some articles ignore the difference of the "true" next day's range and observable range, and assume that the underlying price is unchanged until today's opening hour.

The proxy, however, comes with a cost. A lot of relevant information arrives between the last closing hour to the next opening hour. For example, (1) some very important news is released when the market is closed; (2) while the domestic exchange is closed, the price fluctuation of related securities traded in a foreign exchange located at a different time zone is likely to let the domestic players drastically change their expectations; (3) some

trading activity of a relevant security continues during the closing hours of the stock exchange, for example, the OTC (over-the-counter) traded S&P 500 index futures contracts. Therefore, before the market opens, everyone updates the expectation and there is often a significant jump of the price at the opening hour. The expected size of the jump implies missing information when the market is closed. Thus, for stock indices, conceptually, the observed intraday range is NOT the daily range. The missing prices should have been the candidates for the high (low) price and the observed intraweek price is only a subset of the complete price information. As a result, the observable high (low) price is downward (upward) biased and the observed range is also downward biased.

To illustrate how the observed range is less than the "true" range for stock indices, we study the stock price of Citigroup Inc (NYSE:C) during the trading day Dec 26, 2007 to Dec 28, 2007. *Figure 44*[92] plots the price fluctuation of NYSE:C during the official trading hour, which is from 9:30 to 4:00 ET. We noticed that the price jumps significantly at the opening hours. To further investigate the reason of those jumps, *Figure 45* includes the price of NYSE:C during the extended trading hours. The price of the extended trading hours gives a more complete picture. For example, during the official trading hours at Dec/27/2007, the highest price of NYSE:C is USD30.06, while the extended hour reveals that the highest price should be at least USD30.60.

[92] Source: Finance.google.com

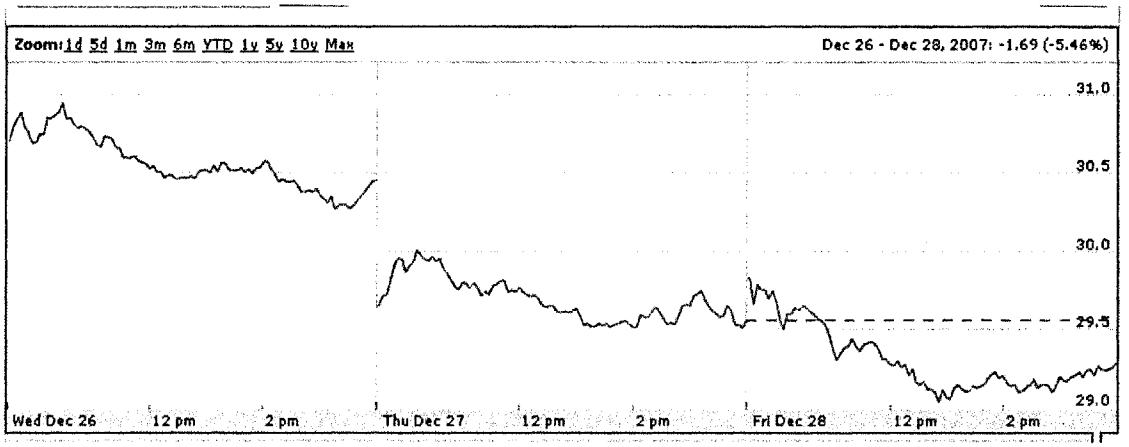


Figure 44

The price of Citigroup Inc at the official trading hours from Dec/26/2007 to Dec/28/2007

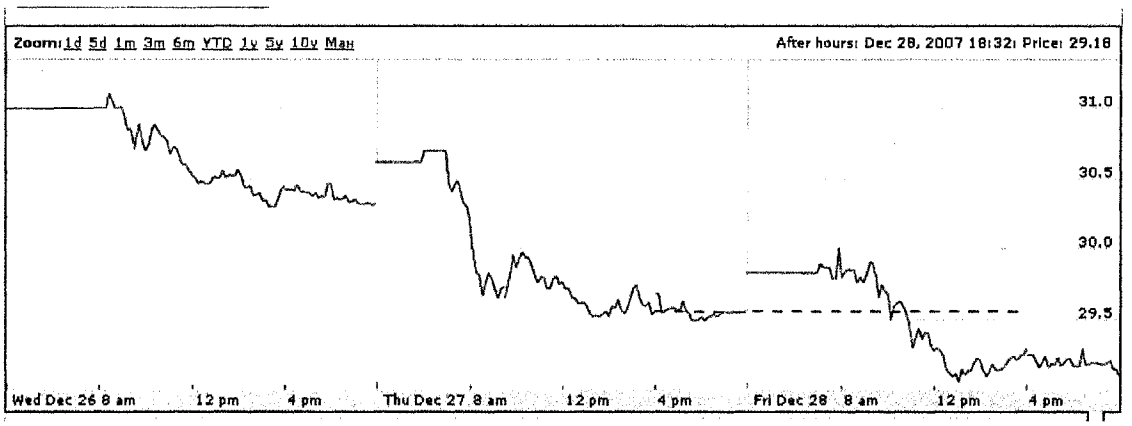


Figure 45

The price of Citigroup Inc at the official trading hours and extended hours[93] *Dec/26/2007 to Dec/28/2007*

In conclusion, because of the discrepancies in the definitions of sampling period of "range", the close-form relation between the range and volatility explored by Parkinson (1980) is more appropriate for currencies and 24 hours traded futures, but not for stock indices.

[93] U.S. exchange after-hour markets
The NYSE and ASE provide crossing sessions in which matching buy and sell orders can be executed at 5:00 p.m. based on the exchanges' 4:00 p.m. closing prices. The BSE and PSE have post-primary sessions that operate from 4:00 to 4:15. CHX and PCX operate their post-primary sessions until 4:30 p.m. Additionally CHX has an "E-Session" to handle limit orders from 4:30 to 6:30p.m.

APPENDIX X

This appendix proposes a simple model to forecast the range based on the lagged observed ranges.

Chou proposes the Conditional AutoRegressive Range (CARR) model in 2005; CARR models directly on the observable range instead of the volatility and Chou views the range as an alternative measure for volatility.

In the spirit of CARR, we find that we can also fit the *log* observable range with a very simple ARMA(1,1) model,

$$(1-\gamma_1 \bullet L)(\log(R_t)-\gamma_0)=(1-\gamma_2 \bullet L)\zeta_t, \qquad \text{(AppX.1)}$$

where L is the lag operator. The long-term (unconditional) value of log observable range is

$$\log R_\infty = \frac{\gamma_0}{1-\gamma_1}. \qquad \text{(AppX.2)}$$

The implied autocorrelation function is given by

$$\rho_1 = \frac{(1-\gamma_1 \bullet \gamma_2)(\gamma_1-\gamma_2)}{1+\gamma_1^2-2\gamma_1 \bullet \gamma_2},$$

and for displacement k greater than 1,

$$\rho_k = \gamma_1 \bullet \rho_{k-1}, \qquad k \geq 2.$$

The data used to empirically test the ARMA (1,1) model is the same dataset used for the hidden Markov-switching volatility model.

Let $R_t := 100 \cdot \left[\log\left(price_t^{(high)} \right) - \log\left(price_t^{(low)} \right) \right]$ be the observed weekly percentage range.

Table 19 gives the summary statistics of the 1297 observations on R_t

Summary statistics for the percentage observed ranges of weekly S&P500 index, from January 05, 1983 to November 21, 2007		
	Observed range	Log observed range
Mean	2.97	0.95
Median	2.52	0.92
Maximum	34.37	3.54
Minimum	0.69	-0.37
Standard deviation	1.89	0.50
Skewness	5.0	0.3
Kurtosis	66.0	3.4
J-B test	219790	36
P value of *J-B* test	0.000	0.000
Sample auto-correlation functions (lags)		
ACF (1)	0.51	0.54
ACF (2)	0.43	0.50
ACF (3)	0.36	0.45
ACF (4)	0.33	0.42
ACF (5)	0.32	0.40
ACF (6)	0.34	0.41

Table 19

Descriptive statistics of the percentage observed intraweek ranges

Notice that in *Table 19*, the *J-B* test indicates that the unconditional distribution of the log observed intraweek range is close to a normal distribution and the sample ACFs are more significant than the autocorrelations of the squared return. *Figure 46* plots the sample ACFs, PACFs and histogram of the log observed intraweek range. The sample ACFs and PACFs are both decaying, which resemblance an ARMA (1,1) process.

Notice that in *Table 6*, the *J-B* test[94] indicates that the unconditional distribution of the

[94] The test statistic *J-B* is defined as:-

$$J-B = \frac{n}{6}\left(Skewness^2 + \frac{(Kurtosis-3)^2}{4} \right)$$

log observed intraweek range is relatively closer to a normal distribution and the sample ACFs are statistically significant.[95]

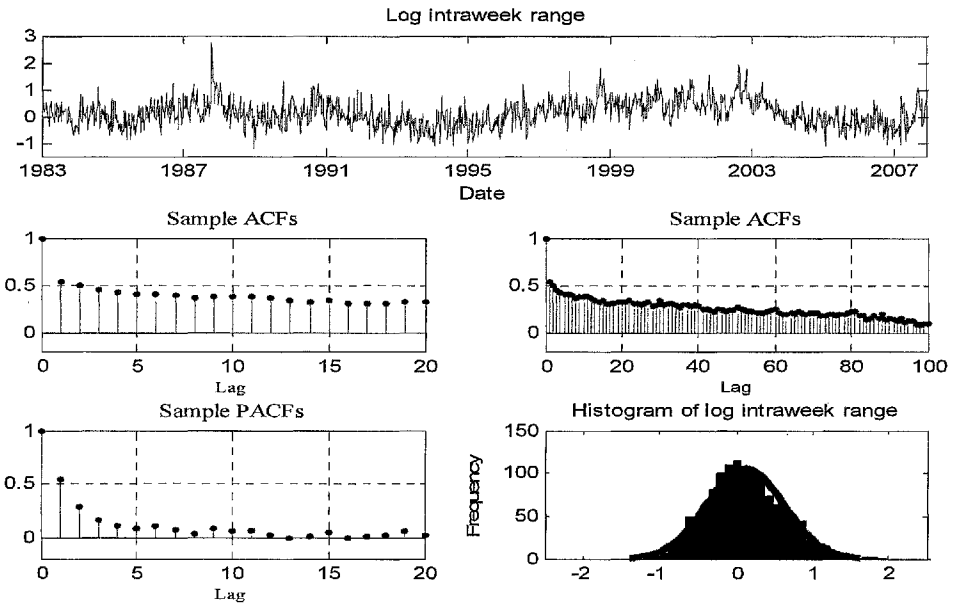


Figure 46
The log observed intraweek range and some preliminary data analysis (Sample ACFs and PACFs)

The available sample is divided into two parts. The first 989 observations, corresponding to the period from January 05, 1983 to December 26, 2001, will be used for estimation. The remaining 308 observations will be employed for out-of-sample evaluation purpose.

Assuming that the residuals of the regression (AppX.1) follow *i.i.d.* normal distribution, fit the in-sample log observed intraweek range to the ARMA(1,1) process,

where n is the degrees of freedom.
The statistic *J-B* has an asymptotic chi-square distribution with two degrees of freedom and can be used to test the null hypothesis that data are from a normal distribution.

[95] Figure 46 plots the sample ACFs, PACFs and histogram of the log observed intraweek range. The sample ACFs and PACFs are both decaying, which resemblance an ARMA (1,1) process.

$$\begin{array}{c}(1-0.9577L)\left(\log(R_t)-0.41\right)=(1+0.7460L)\zeta_t; \\ \quad (0.0132) \qquad\qquad (0.013) \qquad (0.0300) \\ \widehat{Var}(\zeta_t)=0.1606. \\ \qquad\quad (0.00595)\end{array} \qquad \text{(AppX.3)}$$

$$LLF_{ARMA}\left(\hat{\boldsymbol{\gamma}}\right)=-499.08,$$

and

$$\log \widehat{R_\infty}=\frac{\widehat{\gamma_0}}{1-\widehat{\gamma_1}}=0.97. \qquad \text{(AppX.4)}$$

The bottom panel of *Figure 48* plots the residual of the above regression, which does not exhibit any trend or clustering. *Figure 48* also diagnoses the residuals, whose sample ACFs and PACFs are not significantly different from zero. The QQ plot suggests that the normal distribution is reasonable. The LM test also indicates that there is no significant correlation of the squared residuals (not shown). We also try higher AR or MA orders, but the likelihood ratio (LR) test prefers ARMA (1,1) process at the 90% confidence interval[96].

[96] results not shown

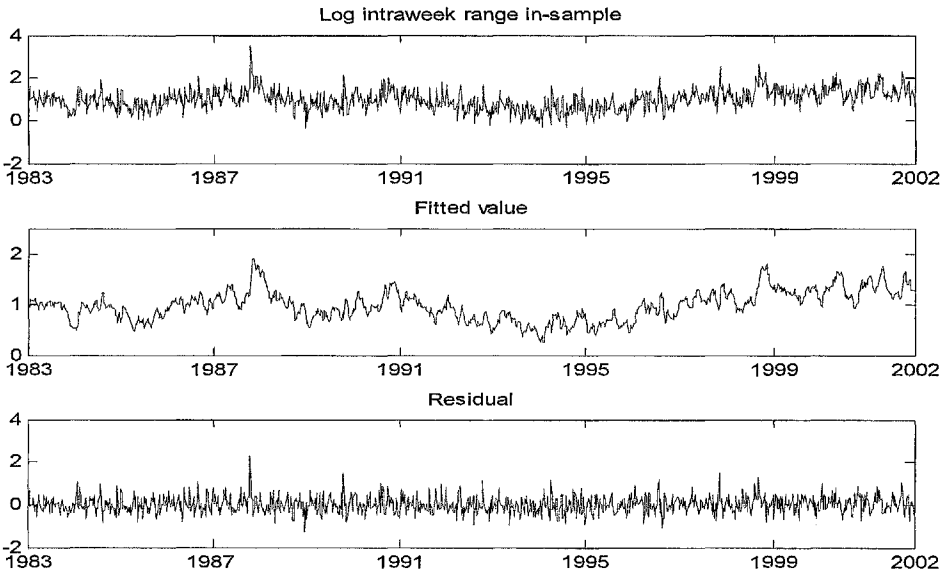


Figure 47

The top panel plots the in-sample Log intraweek range R_t;

the mid panel plots the in-sample fitted Log intraweek range $\widehat{R}_t$; and

the bottom panel plots the in-sample residuals $R_t - \widehat{R}_t$;

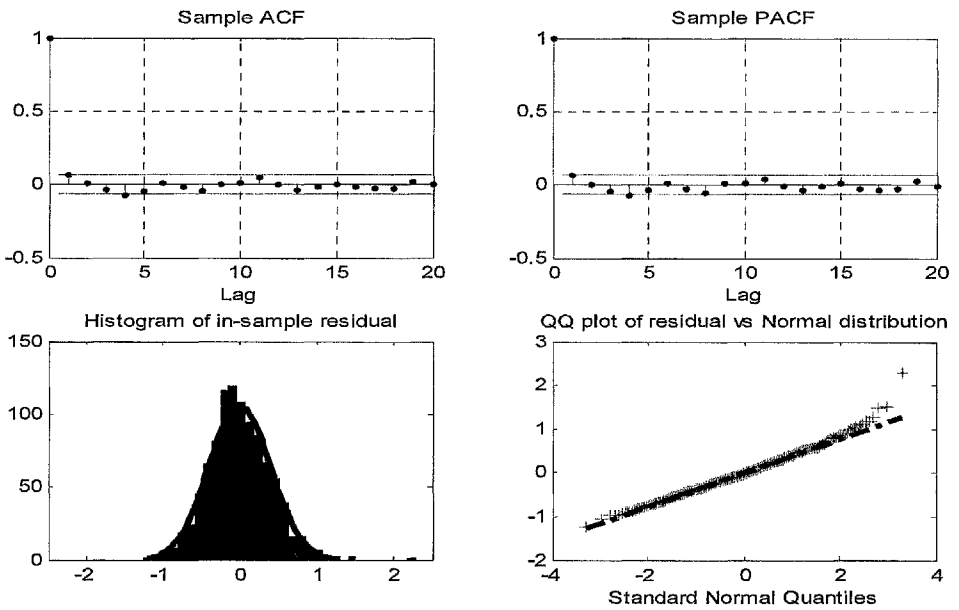


Figure 48

***upper-left panel**: the sample ACF of in-sample residuals*

***upper-right panel**: the sample PACF of in-sample residuals*

***lower-left panel**: the histogram of in-sample residuals against the normal distribution*

***lower-right panel**: the QQ plot of in-sample residuals against the normal distribution*

Based on the estimated parameters, we obtain $E\left\{R_t \middle| \psi_{t-1}, \hat{\gamma}\right\}$, for $t = 990, \cdots, 1297$ and the out-of-sample forecast errors.

Figure 49 plots the out-of-sample observations and the forecast errors. *Figure 50* diagnoses the forecast errors (sample ACFs, PACFs, QQ-plot), and there is no significant evidence that the residuals are not *i.i.d.* normal.

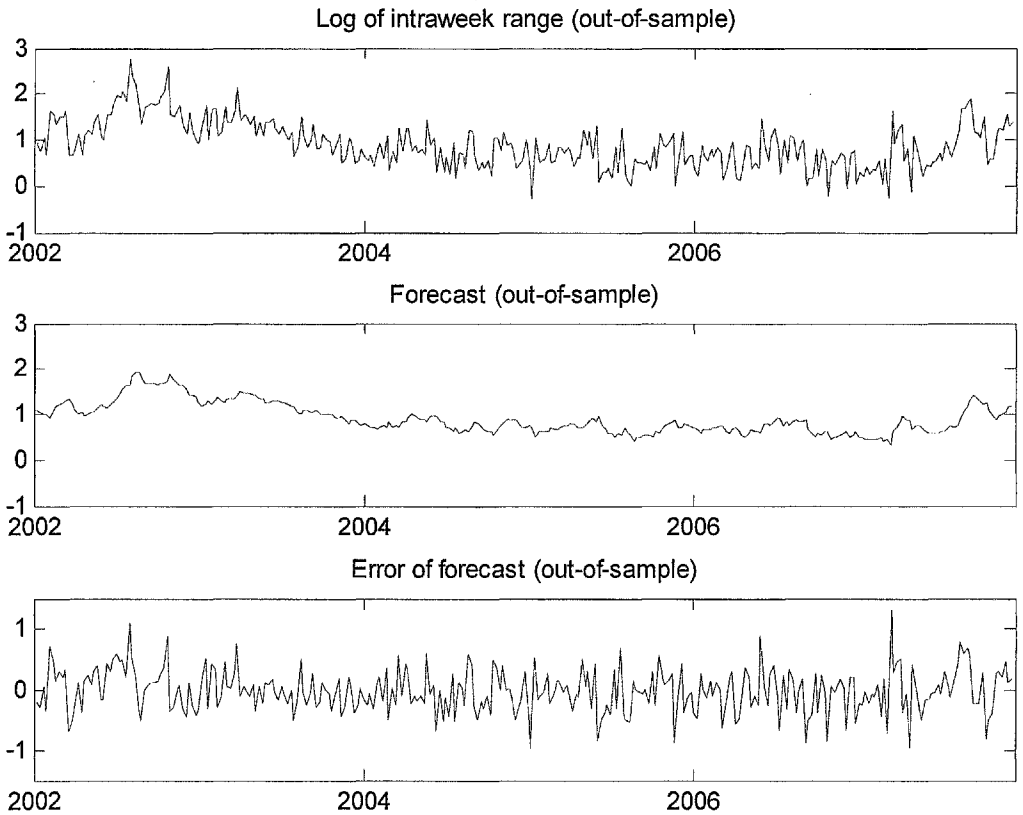


Figure 49

The top panel plots the out-of-sample Log intraweek range R_t;

the mid panel plots the out-of-sample fitted Log intraweek range $\widehat{R}_t$; and

the bottom panel plots the out-of-sample residuals $R_t - \widehat{R}_t$;

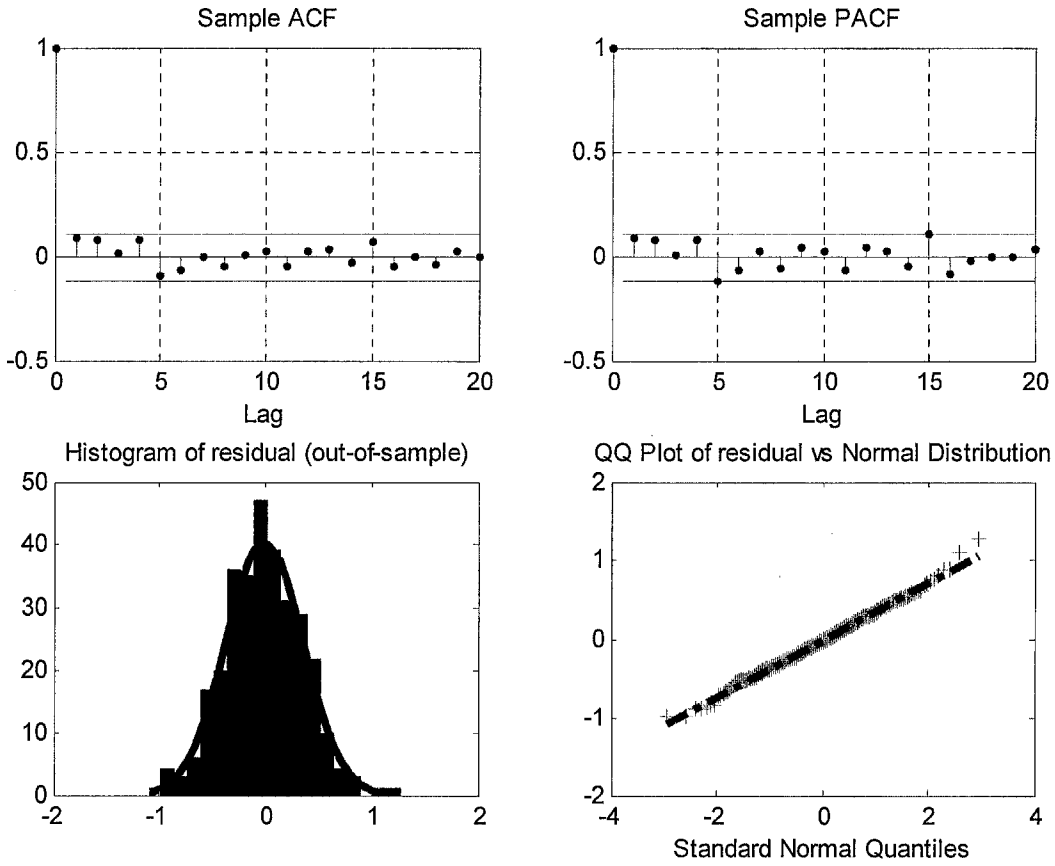


Figure 50

***upper-left panel**: the sample ACF of out-of-sample residuals*

***upper-right panel**: the sample PACF of out-of-sample residuals*

***lower-left panel**: the histogram of out-of-sample residuals against the normal distribution*

***lower-right panel**: the QQ plot of out-of-sample residuals against the normal distribution*

In summary, the ARMA(1,1) with normal residuals provide both adequately reasonable goodness of fit and out-of-sample forecast for the log intraweek range. Either the least square or the MLE can be used to estimate the parameters.

APPENDIX XI

This appendix studies a simple causality test between the intraweek range and weekly return using the S&P 500 stock index historical data from January 05, 1983 to November 21, 2007.

We define Thursday as the first day of the week, and the next Wednesday as the last day of the week. Let $price_t^{(close)}$ be the close price at the end of week t, $price_t^{(high)}$ be the observable highest price during week t; let $y_t := 100 \cdot \left[\log\left(price_t^{(close)} \right) - \log\left(price_{t-1}^{(close)} \right) \right]$ be the weekly percentage log return, and $R_t := 100 \cdot \left[\log\left(price_t^{(high)} \right) - \log\left(price_t^{(low)} \right) \right]$ be the observed weekly percentage range.

Considering the following two regression[97]:

$$\log R_t = a_0 + a_1 \log R_{t-1} + a_2 \log R_{t-2} + a_3 \log R_{t-3} + a_4 \log R_{t-4} + \varepsilon_t^{null} \qquad \text{(AppXI.1)}$$

$$\log R_t = b_0 + b_1 \log R_{t-1} + b_2 \log R_{t-2} + b_3 \log R_{t-3} + b_4 \log R_{t-4} + b_5 y_{t-1}^2 + b_6 y_{t-2}^2 + b_7 y_{t-3}^2 + b_8 y_{t-4}^2 + \varepsilon_t^{alt}. \qquad \text{(AppXI.2)}$$

In Equation (AppXI.1), we first attempt to forecast R_t using past terms of R_t (R_{t-1} to R_{t-4}). Then in Equation (AppXI.2), we try to forecast R_t using past terms of R_t (R_{t-1} to R_{t-4}) and y_t^2 (y_{t-1}^2 to y_{t-4}^2). If the second forecast is found to be much more successful, then the past of y_t^2 appears to contain information helping in forecasting R_t that is not in past R_t. If this were the case, y_{t-1}^2 to y_{t-4}^2 "Granger cause" R_{t+1}.

[97] Note that we do not use R_t as the independent variable, because there is strong evidence that the residuals do not follow normal distribution and we can not use simple OLS and F-test to study the causality.

Fit the above two equations using the intraweek range and weekly return data from Jan/05/1983 to Nov/21/2007, using OLS estimator, we find that neither the residuals ε^{null} nor ε_t^{alt} deviates significantly from the normal distribution[98]. Therefore, we use the simple OLS and F-test to study if the past of y_t^2 contain information helping in forecasting R_t that is not in past R_t.

$H_{cause}: b_5 = b_6 = b_7 = b_8 = 0$

According to the regression result exhibited in *Table 20*, none of the *t*-statistic for the parameters estimate $\widehat{b_5}$, $\widehat{b_6}$, $\widehat{b_7}$ and $\widehat{b_8}$ is significantly from zero. In addition, the F test fails to reject the hypothesis $b_5 = b_6 = b_7 = b_8 = 0$. Therefore, the past of y_t^2 appears to contain information helping in forecasting R_t that is not in past R_t, and y_t^2 does not "Granger cause" R_{t+1}.

[98] Please refer to the QQ plot at *Figure 51*.

using only the lagged terms of R_t			using only the lagged terms of both R_t and y_t		
	$\hat{a}$	$\sigma(\hat{a})$		$\hat{b}$	$\sigma(\hat{b})$
$\widehat{a_0}$	0.023	(0.0285)	$\widehat{b_0}$	0.021	(0.0314)
$\widehat{a_1}$	0.319	(0.0276)	$\widehat{b_1}$	0.302	(0.0322)
$\widehat{a_2}$	0.204	(0.0288)	$\widehat{b_2}$	0.210	(0.0332)
$\widehat{a_3}$	0.121	(0.0288)	$\widehat{b_3}$	0.135	(0.0332)
$\widehat{a_4}$	0.115	(0.0277)	$\widehat{b_4}$	0.144	(0.0319)
			$\widehat{b_5}$	0.005	(0.0063)
			$\widehat{b_6}$	-0.002	(0.0065)
			$\widehat{b_7}$	-0.005	(0.0065)
			$\widehat{b_8}$	-0.012	(0.0063)
$\sigma(\widehat{\varepsilon^{null}})$	0.1584		$\sigma(\widehat{\varepsilon^{alt}})$	0.1581	
SSR^{null}	204.6		SSR^{alt}	203.6	
F-test	1.5742				
P value of F-test					

Table 20

This table shows that adding lagged terms of y^2 *hardly improves the goodness-of-fit of* $\log(R_t)$

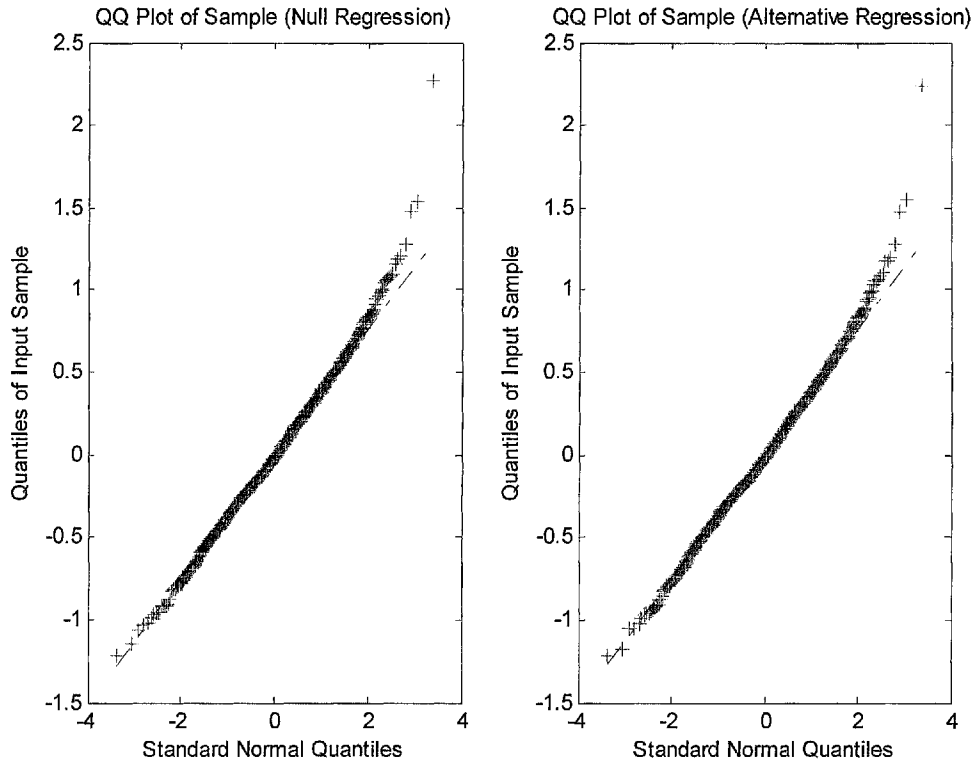


Figure 51

The left and right figure are the QQ plot of the residuals (ε_t^{null} and ε_t^{alt}) respectively. There is no strong evidence that the residuals ε_t^{null} or ε_t^{alt} does not follow normal distribution, therefore, it is probably reasonable to use an F-test to test the hypothesis.

Reference:

Abreu D. and Markus B., "Bubbles and Crashes", *Econometrica*, 2003, *71*, pp. 2013-2040

Akaike H., "Information Theory and an Extension of the Maximum Likelihood Principle", F. C. B.N. Petrov, *In 2nd International Symposium on Information Theory.* Budapest: Akademia Kiado, 1973, 267-281.

Alexander C.; Kaeck A. "Regimes in CDS Spreads: A Markov Switching Model of iTraxx Europ Indices", *Journal of Banking & Finance,* forthcoming.

Alexander C., "Short and Long Term Smile Effects: The Binomial Normal Mixture Diffusion Model", *Journal of Banking & Finance*, 2004, *28*, pp. 2957-2980.

Alizadeh A.; Nomikos N. and Pouliasis P., "A Markov Regime Switching Approach for Hedging Energy Commodities ", *Journal of Banking and Finance*, 2008, *forthcoming.*

Alizadeh S.; Brandt M. and Diebold F., "Range-Based Estimation of Stochastic Volatility Model", *Journal of Finance*, 2002, *57*, pp. 1047-1092.

Andersen T. and Bollerslev T. "Answering the Sceptics: Yes, Standard Volatility Models Do Provide Accurate Forecast", *International Economic Review*, 1998, *39*, pp. 885-905.

Ang A. and Bekaert G., "International Asset Allocation with Regime Shifts", *Review of Financial Studies*, 2002a, *15*, pp. 1137-1187.

____. "Regime Switching in Interest Rate", *Journal of Business & Economic Statistics*, 2002b, *20*, pp. 163-182.

Ang A. and Chen J., "Asymmetric Correlations of Equity Portfolios", *Journal of Financial Economics*, 2002, *63*, pp. 443-494

Baillie R.; Bollerslev T. and Mikkelsen H., "Fractionally Integrated Generalized Autoregressive Conditional Heteroscedasticity", *Journal of Econometrics*, 1996, *74*, pp. 3-30.

Baillie R.; Chung C. and Tieslau M., "Analysing Inflation by the Fractionally Integrated ARFIMA-GARCH Model", *Journal of Applied Econometrics*, 1996, *11*, pp. 23-40.

Barkoulas J. and Baum C., "Long Memory in the Greek Stock Markets", *Applied Financial Economics*, 2000, *10*, pp. 177-184.

Basel Committee on Banking Supervision, "*International Convergence of Capital Measurement and Capital Standards*", 2006

Banachewicz K.; Lucas A. and Vaart D., "Modeling Portfolio Defaults Using Hidden Markov Models with Covariates", *Econometrics Journal*, 2007, *10*, pp. 1-18.

Baum L. and Petire T., "Statistical Inference for Probabilistic Functions of Finite State Markov Chains", *Annals of Mathematical Statistics*, 1966, *37*, pp. 1554-1563.

Baum L.; Petrie T.; Soules G. and Weiss N., "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains", *The Annals of Mathematical Statistics*, 1970, *41*, pp. 164-171.

Beine M.; Laurent S. and Lecourt C., "Official Central Bank Interventions and Exchange Rate Volatility: Evidence from a Regime-Switching Analysis", *working paper*, 2003, *47*, pp. 891-911.

Bera A. and Higgins M., "ARCH Models: Properties, Estimation and Testing", *Journal of Economic survey*, 1993, *7*, pp. 305-366.

Bera A. and Jarque C., "Efficient Tests for Normality, Heteroscedaticity, and Serial Dependence of Regression Residuals", Economic Letters, 1980, 6, pp. 255-259.

Berkowitz J., "Testing Density Forecast, with Application to Risk Management", *Journal of Business & Economic Statistics*, 2001, *19*, pp. 465-474.

Bernake B., "The Macroeconomics of the Great Depression: A Comparative Approach", *Journal of Money, Credit and Banking,* 1995, *27,* pp.1-28

____."The Economic Outlook", *Testimony Before the Joint Economic Committee,* Oct/20/2005

____."The Recent Financial Turmoil and its Economic and Policy Consequences", Oct/2007, http://www.frbsf.org/news/speeches/2007/1009.html

____."Financial Markets, the Economic Outlook, and Monetary Policy", Jan/2008, http://www.federalreserve.gov/newsevents/speech/bernanke20080110a.htm

Bernake B. and Alan B., "Credit, Money and Aggregate Demand", *American Economic Review, 78,* pp. 435-439

Bernake B. and Gertler M., "Should Central Banks Respond to Movements in Asset Prices?", *The American Economic Review*, 2001, *91*, pp. 253-257

Bernake B. and Mihov I., "Measuring Monetary Policy", *Quarterly Journal of Economics,* 1998, *113*, pp. 869-902

Bickel P.; Ritoc Y. and Ryden T., "Asymptotic Normality of the Maximum Likelihood Estimator for General Hidden Markov Models", *Annals of Statistics*, 1998, *26*, pp. 1614-1635.

Black F., "Studies of Stock Market Volatility Changes", I. A. S. Association, *Proceedings of the American Statistical Association, Business and Economic Statistics Section.* Alexandria,VA: American Statistical Association, 1976.

Black F. and Cox J., "Valuing Corporate Securities: Some Effect of Bond Indenture Provisions", *Journal of Finance*, 1976, *31*, pp. 351-367.

Bollen N.; Gray S. and Whaley R., "Regime Switching in Foreign Exchange Rates: Evidence from Currency Option Prices", *Journal of Econometrics*, 2000, *94*, pp. 239-276.

Bollerslev T., "Generalized Autoregressive Conditional Herteroskedasticity", *Journal of Econometrics*, 1986, *31*, pp. 307-327.

Bollerslev T.; Engle R., and Wooldridge J., "A Capital Asset Pricing Model with Time-varying Covariances", *Journal of Political Economy*, 1988, *96*, pp. 116-131.

Boswijk H. P., Hommes C. H., and Manzan S., "Behavioral Heterogeneity in Stock Prices". *Journal of Economic Dynamics & Control*, 2007, *31*, pp. 1938-1970.

Box G., Jenkins G. M., and Reinsel G., "Time Series Analysis: Forecasting & Control ". 3th edition, *Prentice Hall*, 1994.

Brandt M. and Diebold F., "A No-Arbitrage Approach to Range-Based Estimation of Return Covariances and Correlations", *Journal of Business*, 2006, *79*, pp. 61-74.

Brandt M. and Jones C., "Volatility Forecasting with Range-Based EGARCH Models", *working paper*, 2002.

Brent R., Algorithms for Minimization without Derivatives. *Prentice-Hall,* 1973.

Brunnermeier M., "Deciphering the 2007/8 Liquidity and Credit Crunch", *Journal of Economic Perspective*, 2008 (forthcoming).

Cai J., "A Markov Model of Switching-Regime ARCH", *Journal of Business & Economic Statistics*, 1994, *12*, pp. 309-316.

Calvet L., "How to Forecast Long-Run Volatility: Regime Switching and the Estimation of Multifractal Process", *Journal of financial Econometrics*, 2004, *2*, pp. 49-83.

Charlotte C., "Regime Switching in the Yield Curve", *Journal of Future markets*, 2004, *24*, pp. 315-336.

Chen J.; Hong H. and Stein J., "Forecasting Crashes: Trading Volume, Past Returns and Conditional Skewness in Stock Prices", *Journal of Financial Economics*, 2001, *61*, pp. 345-381.

Chen S., "Does Monetary Policy Have Asymmetric Effects on Stock Returns?", *Journal of Money, Credit and Banking*, 2007, *39*, pp. 667-688

Cheung Y. and Erlandsson U., "Exchange Rates and Markov Switching Dynamics", *Journal of Business & Economic Statistics*, 2005, *23*, pp. 314-320.

Chou R., "Forecasting Financial Volatilities with Extreme Values: The Conditional Autoregressive Range (CARR) Model", *Journal of money, credit, and banking*, 2005, *37*, pp. 561-582.

Christensen C., "Regime Switching in the Yield Curve", *Journal of Futures Markets*, 2004, *24*, pp. 315-336.

Christie A., "The Stochastic Behavior of Common Stock Variance: Value, Leverage and interest rate effect", *Journal of Financial Economics*, 1982, *10*, pp. 407-432.

Clark P., "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices", *Econometrica,* 1986, *33*, pp.237-262

Clements M. and Krolzig H., "Business cycle asymmetries: Characterisation and testing based on markov-switching autoregressions", *Journal of Business and Economic Statistics,* 2003, *21*, pp. 196 - 211.

Cont R., "Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues", *Quantitative Finance*, 2001, *1*, pp. 223-236.

Cox C.; Ingersoll J. and Ross S., "A Theory of the Term Structure of Interest Rates", *Econometrica* 1985, *53*. pp. 385-407.

Dacco R. and Satchell S., "Why Do Regime-Switching Models Forecast So Badly?", *Journal of forecasting*, 1999, *18*, pp. 1-16.

Davig T., "Regime-Switching Debt and Taxation", *Journal of Monetary economics*, 2004, *51*, pp. 837-859.

Dempster A.; Laird N. and Rubin D., "Maximum Likelihood from Incomplete Data Via the EM

Algorithm", *The Journal of the Royal Statistical Society (Series B)*, 1977, *39*, pp. 1-38.
Dijk D.; Franses P. H. and Boswijk H. P., "Asymmetric and Common absorption of Shocks in Nonlinear Autoregressive Models" working paper.
Diebold F., "Modelling the Persistence of Conditional Variance: A Comments", *Econometric Reviews*, 1986, *5*, pp. 51-56.
Diebold F. and Inoue A., "Long Memory and Structural Change", *Journal of Econometrics, Forthcoming.*
Douc R. and Matias C., "Asymptotics of the Maximum Likelihood Estimator for General Hidden Markov Models", *Bernoulli*, 2001, *7*, pp. 381-420.
Dueker M., "Markov Switching in GARCH Processes and Mean-Reverting Stock-Market Volatility", *Journal of Business & Economic Statistics*, 1997, *15*, pp. 26-35.
Dueker M. and Neely C., "Can Markov Switching Models Predict Excess Foreign Exchange Returns?", *Journal of Banking & Finance*, 2007, *31*, pp. 279-296.
Engle C. and Hamilton J., "Long Swings in the Dollar: Are They in the Data and Do the Markets Know It?", *American Economic Review*, 1990, *80*, pp. 689-713.
Engle C., "Can the Markov Switching Model Forecast Exchange Rate?" *Journal of International Economics*, 1994, *36*, pp. 151-165.
Engle R., "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of Untied Kingdom Inflation", *Econometrica*, 1982, *50*, pp. 987-1007.
Engle R. and Lee G., "A Long-Run and Short-Run Component Model of Stock Return Volatility," *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W.J. Granger.* Oxford University Press, 1999, 475-497.
Engle R. and Patton A., "What Good Is a Volatility Model?" *Quantitative Finance*, 2001, *1*, pp. 237-245.
Everitt B. and Gand D., Finite Mixture Distributions. *Chapman & Hall*, 1981.
Falk B., "Further Evidence on the Asymmetric Behavior of Economic Time Series over the business Cycle", *Journal of Political Economy,* 1986, *94,* pp. 1096-1109
Fama E., "The Behaviour of Stock Market Prices", *Journal of Business*, 1965, *38*, pp. 34-105.
Friedman M., "Monetary Variability: United States and Japan", *Journal of Money, Credit and Banking,* 1983, *15,* pp. 339-343
Garcia R. and Perron P., "An Analysis of the Real Interest Rate under Regime Shifts", *Review of Economics and Statistics*, 1996, *78*, pp. 111-125.
Genon-Catalot V.; Heantheau T. and Laredo C. "Conditional Likelihood Estimators for Hidden Markov Models and Stochastic Volatility Models", *Board of the Foundation of the Scandinavian Journal of Statistics*, 2003, *30*, pp. 297-316.
Giampieri G.; Davis M. and Crowder M., "A Hidden Markov Model of Default Interaction", *Quantitative Finance,* 2005, *5*, pp. 27-34.
Goodfriend M.; Robert G. and King M., "The Incredible Volcker disinflation", *Journal of Monetary Economics,* 2005, *52,* pp. 981-1015
Granger C. and Ding Z., "Some Properties of Absolute Returns: An Alternative Measure of Risk", *working paper*, 1993.
Granger C. and Ding Z., "Varieties of Long Memory Models", *Journal of Econometrics*, 1996, *73*, pp. 61-77.
Gray S. "Modelling the Conditional Distribution of Interest Rate as a Regime-Switching Process", *Journal of Financial Economics*, 1996, *42*, pp. 27-62.
Greenlaw D.; Hatzius J., and Hyun S., "Leveraged Losses: Lessons From the Mortgage Market Meltdown", *US Monetary Policy Forum Conference Draft.* Feb/2008
http://www.chicagogsb.edu/usmpf/docs/usmpf2008confdraft.pdf
Greenspan A. "The age of turbulence: adventures in a new world", September 17, 2007, *Penguin Press HC, The.*
Guidolin M. and Timmermann A., "International Asset Allocation Under Regime Switching", *The Review of Financial Studies*, 2008, *21*, pp. 889-935.
____. "Size and Value Anomalies Under Regime Shifts", *Journal of Financial Economics*, 2008, *6, pp.* 1-48
Haas M.; Mittnik S. and Paolella M., "Mixed Normal Conditional Heteroscedasticity." *Journal of Financial Econometrics*, 2004a, *2*, pp. 211-250.
____. "A New Approach to Markov Switching GARCH Models." *Journal of Financial Econometrics*, 2004b, *4*, pp. 493-530.

Hamilton J., "Regime-Switching Models", *working paper*, 2005.
____. "Specification Testing in Markov-Switching Time-Series Models", *Journal of Econometrics*, 1996, *70*, pp. 127-157.
____. "Analysis of Time Series Subject to Changes in Regime", *Journal of Econometrics*, 1990, *45*, pp. 39-70.
____. "A New Approach to the Economics Analysis of Nonstationary Time Series and the Business Cycle", *Econometrica*, 1989a, *57*, pp. 357-384.
____. "Rational-Expectations Econometric Analysis of Changes in Regime: An Investigation of the Term Structure of Interest Rate", *Journal of Economic Dynamic and Control*, 1989b, *57*, pp. 357-384.
Hamilton J. and Gabriel P., "What Do the Leading Indicators Lead?", *Journal of Business*, 1996, *69*, pp. 27-49.
Hamilton J. and Lin G. "Stock Market Volatility and the Business Cycle", *Journal of Applied Econometrics*, 1996, *11*, pp. 573-593.
Hamilton J. and Susmel, R., "Autoregressive Conditional Heteroscedasticity and Changes in Regime", *Journal of Econometrics*, 1994, *64*, pp. 307-333.
Hansen B., "Autoregressive Conditional Density Estimation", *International Economic Review*, 1994, *35*, pp. 705-730.
Hennig C., "Identifiability of Models for Clusterwise Linear Regression", *Journal of Classification*, 2000, *17*, pp. 273-296.
Hicks J., A Contribution to the Theory of the Trade Cycle. *Oxford: Clarendon Press,* 1950.
Holzman H.; Munk A. and Gneiting T. "Identifiability of Finite Mixture of Elliptical Distributions", *Board of the Foundation of the Scandinavian Journal of Statistics*, 2006.
Hong Y.; Tu J. and Zhou G. "Asymmetries in Stock Returns ", *Review of Financial Studies,* Forthcoming.
Hyung N.; Poon S. and Granger C., "The Source of Long Memory in Financial Market Volatility", *unpublished manuscript*, 2005.
Jamshidian M. and Jennrich R., "Acceleration of the EM Algorithm by Using Quasi-Newton Methods", *Journal of Royal Statistical Society, Series B*, 1997, *59*, pp. 569-587.
____. "Standard Errors for EM Estimation", *Journal of Royal Statistical Society*, 2000, *62*, pp. 257-70.
Jeanne O. and Masson P. "Currency Crises, Sunspots, and Markov-Switching Regimes", *Journal of Business & Economic Statistics*, 2000, *50*, pp. 327-350.
Keynes J., The General Theory of Employment, Interest, and Money. *New York: Hardcourt, Brace,* 1936.
Kim C., "Dynamic Linear Models with Markov Switching", *Journal of Econometrics*, 1994, *60*, pp. 1-22.
____. "Markov-Switching Models with Endogenous Explanatory Variables", *Journal of Econometrics*, 2004, *122*, pp. 127-136.
Klaassen F., "Improving GARCH Volatility Forecasts with Regime-Switching GARCH", *Empirical Economics*, 2002, *27*, pp. 363-394.
Krolzig H., "Markov-Switching Procedures for Dating the Euro-Zone Business Cycle", *Quarterly Journal of Economic Research,* 2001, *3*, pp. 339-351
Lamoureux C. and Lastrapes W., "Persistence in Variance, Structural Change and the GARCH Model", *Journal of Business & Economic Statistics,* 1990, *8*, pp. 225-234.
Lange K., "A Gradient Algorithm Locally Equivalent to the EM Algorithm", *Journal of the Royal Statistical Society B,* 1995a, *57*: pp. 425-437.
____. "A Quasi-Newton Acceleration of the EM Algorithm." *Statistica Sinica* 1995b, *5*: 1-18.
Lanne M. and Saikkonen P., "Modelling the U.S. Short-Term Interest Rate by Mixture Autoregressive Processes", *Journal of Financial Econometrics*, 2003, *1*, pp. 96-125.
LeGland F. and Mevel L., "Exponential forgetting and geometric ergodicity in hidden markov models", *Mathematics of Control, Signals and Systems*, 2000. *13,* pp. 63-93.
Leroux B. and Puterman M., "Maximum-Penalized-Likelihood Estimation for Independent and Markov-Dependent Mixture Models", *Biometrics*, 1992, *48*, pp. 545-558.
Leroux B., "Consistent Estimation of a Mixing Distribution", *Annals of Statistics*, 1992, *20*, pp. 1350-1360.
Lobato I. and Savin N., "Real and Spurious Long-Memory Properties of Stock-Market Data", *Journal of Business & Economic Statistics*, 1998, *16*, pp. 261-268.
Longin F. and Solnik B., "Extreme Correlation of International Equity markets", *Journal of Finance,* 2001, *56,* pp. 649-674

Lowenstein R., When Genius Failed: the Rise and Fall of Long-Term Capital Management, 2001, *Random House Trade*

Lux T., "Long-Term Stochastic Dependence in Financial Prices: Evidence from the German Stock Market", *Aplied Economics Letters*, 1996, *3*, pp. 261-283.

Madhusoodanan T., "Long-Term Dependence in the Indian Stock Market", *Journal of Financial studies*, 1998, *3*, pp. 33-53.

Maheu J. and Mccurdy T., "Identifying Bull and Bear Markets in Stock Returns." *Journal of Business & Economic Statistics*, 2000, *18*(1), pp. 100.

Mandelbrot B., "The Valuation of Some Other Speculative Prices", *Journal of Business*, 1967, *40*, pp. 393-413.

____. "New Methods in Statistical Economics", *Journal of Political Economy*, 1963a, *71*, pp. 421-440.

Marcucci J., "Forecasting Stock Market Volatility with Regime-Switching GARCH Models", *Studies in Nonlinear Dynamics & Econometrics,* 2005, Vol. 9: No. 4, Article 6.

Markowitz H., "Portfolio Selection", *Journal of Finance*, 1952, *7,* pp. 77-91

Martens M., "Estimating Unbiased and Precise Realized Covariances", *working paper*, 2004.

Martens M. and Dijk D., "Measuring Volatility with the Realized Range", *Journal of Econometrics, forthcoming.*

Martens M.; Pooter M. and Dijk D. "Predicting the Daily Covariance Matrix of S&P100 Stocks Using Intraday Data - but Which Frequency to Use", *Econometric Review, forthcoming.*

McLachlan G. and Peer D., Finite Mixture Models. *New York: Wiley,* 2000.

McLachlan G. and Krishnan T., The EM Algorithm and Extensions. *New York: John Wiley & Sons.*, 1997.

Meng X. and Rubin D., "On the Global and Componentwise Rates of Convergence of the EM Algorithm", *Linear algebra and its applications*, 1994, *199*, pp. 413-425.

____. "Using EM to Obtain Asymptotic Variance-Covariance Matrices: The SEM Algorithm", *Journal of the American Statistical Association*, 1991, *86*, pp. 899-909.

Merton R., "On the pricing of corporate debt: The risk structure of interest rates", *Journal of Finance,* 1974, *29*, 449-470.

Naik P. A. and Raman K., "Understanding the Impact of Synergy in Multimedia Communications", *Journal of Marketing Research*, 2003, *40*, pp. 375-388.

Neftci S., "Are Economic Time Series Asymetric Over the Business Cycle?", *Journal of Political Economy*, 1984, *92*, pp. 307-328

Nelson D., "Conditional Heteroskedasticity in Asset Returns: A New Approach", *Econometrica*, 1991, *59*, pp. 347-370.

____. "Stationarity and Persistence in the GARCH(1,1) Models", *Econometric Theory*, 1990, *6*, pp. 318-334.

Nobuhiro K. and Moore J., "Credit Cycles", *Journal of Political Economy,* 1997, *105*, pp. 211-248

Parkinson M., "The Extreme Value Method for Estimating the Variance of the Rate of Return", *Journal of Business*, 1980, *53*, pp. 61-65.

Pelletier D., "Regime Switching for Dynamic Correlations", *Journal of Econometrics*, forthcoming.

Pinduck R., "Risk, Inflation, and the Stock Market", *American Economic Review,* 1984, *74,* pp.335-351

Poterba J. and Summers L., "The persistence of Volatility and Stock Market Flucturations", *American Economic Review,* 1986, *76,* pp.1142-1151

Puterman M., Markov Decision Processes: Discrete Stochastic Dynamic Programming, *John Wiley and Sons, New York, NY,* 1994.

Redner R. and Walker H., "Mixture densities, maximum likelihood and the EM algorithm", *SIAM review* 1984, *26*: pp. 195-239.

Reinhart C. and Rogoff K., "This Time is Different: A Panoramic View of Eight Centuries of Financial Crises", *Working Paper*, Mar 2008

Ritchey R., "Call Option Valuation for Discrete Normal Mixtures", *Journal of Financial Research*, 1990 *13*, pp. 771-818.

Schafer J., Analysis of Imcomplete Multavariate Data. *Chapman & Hall,* 1997.

Schwarz G., "Estimating the Dimension of a Model", *The Annals of Mathematical Statistics*, 1978, *6*, pp. 461-464.

Sharpe W., "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk", *Journal of Finance*, 1964, *19,* pp. 425-442

Shiller R., "Conversation, Information, and Herd Behavior", *American Economic Review,* 1995, *85,* pp.

181-185.
____., "Exuberant Reporting", *Harvard International Review,* 2001, *23,* pp. 60-65.
Sheppard K., UCSD GARCH Toolbox, 2005. Computational code
Sichel D., "Business Cycle Asymmetry: A Deeper Look", *Economic Inquiry*, 1993, 31, pp. 224-236
Siddhartha C., "Estimation and Comparison of Multiple Change-Point Models", *Journal of Econometrics*, 1998, *86*, pp. 221-241.
Sims C. and Tao Z., "Were There Switches in the U.S. Monetary Policy?", *American Economic Review,* 2006, *96, pp. 54-81.*
Smith A.; Naik P. and Tsai C. "Markov-Switching Model Selection Using Kullback-Leibler Divergence", *Journal of Econometrics*, (forthcoming).
Smith C. discussion on "Maximum Likelihood Estimation From Incomplete Data Via the EM Algorithm", by Dempster A., and Rubin, D. *Journal of the Royal Statistical Society. Series B,* 1997, *39,* pp 24-25.
Markov-Switching and Stochastic Volatility Diffusion Models of Short-Term Interest Rate", *Journal of Business & Economic Statistics*, 2002, *20*, pp. 183-197.
Smith D. "Markov-Switching and Stochastic Volatility Diffusion Models of Short-Term Interest Rate", *Journal of Business & Economic Statistics*, 2002, *20*, pp. 183-197.
Stock J., "Measuring Business Cycle Time", *Journal of Political Economy*, 1987, *95*, pp. 1240-1261
____. "Dynamic Economic Models Subject to Time Deformation", *Ph.D. Thesis, U. Cal, Berkeley 1983.*
Teicher H., "Identifiability of Finite Mixtures", *Annals of Mathematical Statistics*, 1963, *34*, pp. 1265-1269.
____. "Identifiability of Mixtures", *The Annals of Mathematical Statistics*, 1961, *32*, pp. 244-248.
Timmermann A., "Moments of Markov Switching Models", *Journal of Econometrics*, 2000, *96*, pp. 75-111.
____. "Structure Breaks, Incomplete Information and Stock Prices", *Journal of Business & Economic Statistics*, 2001, *19*, pp. 299-315.
Turner C.; Startz R. and Nelson C., "A Markov Model of Heteroskedasticity, Risk, and Learning In the Stock Market", *Journal of Financial Economics*, 1989, *25,* pp. 3-22.
Tolvi J., "Long-Term Memory in a Small Stock Market", *Economics Bulletin*, 2003, *7*, pp. 1-13.
Vlaar P. and Palm, F., "The Message in Weekly Exchange Rates in the European Monetary System: Mean Reversion, Conditional Heteroskedasticity and Jumps", *Journal of Business & Economic Statistics*, 1993, *11*, pp. 351-360.
Wang P. and Puterman M., "Markov Poisson Regression Models for Discrete Time Series. Part 2: Applications", *Journal of Applied Statistics*, 1999, *26*, pp. 871-882.
Wang P.; Puterman M.; Cockburn I. and Le N., "Mixed Poisson Regression Models with Covariate Dependent Rates", *Biometrics*, 1996, *52*, pp. 381-400.
Wang P., "A Bivariate Zero-Inflated Negative Binomial Regression Model for Count Data with Excess Zeros", *Economics Letters*, 2003, *78*, pp. 373-378.
Wang P. and Alba J., "A Zero-Inflated Negative Binomial Regression Model with Hidden Markov Chain", *Economic letters*, 2006, *92*, pp. 209-213.
Wang P. and Puterman M., "Markov Poisson Regression Models for Discrete Time Series. Part I: Methodology", *Journal of Applied Statistics*, 1999, *26*, pp. 855-69.
____. "Analysis of Longitudinal Data of Epileptic Seizure Counts-a Two-State Hidden Markov Regression Approach", *Biometrical Journal*, 2001, *43*, pp. 941-62.
Weigend A. and Shi S., "Predicting Daily Probability Distribution of S&P500 Returns", *Journal of Forecasting*, 2000, *19*, pp. 375-92.
Williams J., The Teory of Investment Value, *Havard University Press*, 1938
Wong C. and Li W., "On a Mixture Autoregressive Conditional Heteroskedastic Model", *Journal of the American Statistical Association*, 2001, *96*, pp. 982-995.
____. "On a Mixture Autoregressive Model", *Journal of the Royal Statistical Society*, 2000, *62*, pp. 95-115.
Wu C.F., "On the Convergence of the Em Algorithm", *Annals of Statistics*, 1983, *11*, pp. 95-103.
Yu W. and Zivit E., "Long Memory Versus Strcture Breaks in Modeling and Forecasting Realized Volatility", *working paper*, 2006.
Zhang X., "Information uncertainty and Stock Returns", *Journal of Finance*, 2006, *61*, pp. 105-137.