Forecasting Stock Volatility with Neural Network on Time Varying Transition Probability

Mr. Wasit Norakarntiansin

An Independent Study Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Financial Engineering
Department of Banking and Finance
FACULTY OF COMMERCE AND ACCOUNTANCY
Chulalongkorn University
Academic Year 2022

การพยากรณ์ความผันผวนของราคาหลักทรัพย์ด้วยโครงข่ายประสาทเทียมบนความน่าจะเป็นในการเปลี่ยนผ่านที่แปรผันตามเวลา

นายวสิษฐ นรการเทียนสิน

สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาวิศวกรรมการเงิน ภาควิชาการธนาคารและการเงิน
คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2565

Independent Study Title	Forecasting Stock Volatility with Neural Network on Time Varying Transition Probability
By	Mr. Wasit Norakarntiansin
Field of Study	Financial Engineering
Thesis Advisor	Associate Professor THAISIRI WATEWAI, Ph.D.

Accepted by the FACULTY OF COMMERCE AND ACCOUNTANCY, Chulalongkorn University in Partial Fulfillment of the Requirement for the Master of Science

INDEPENDENT STUDY COMMITTEE

.... Chairman
(Associate Professor SIRA SUCHINTABANDID, Ph.D.)

.... Advisor
(Associate Professor THAISIRI WATEWAI, Ph.D.)

.... Examiner
(Tanawit Sae-Sue, Ph.D.)

วสิษฐ นรการเทียนสิน : การพยากรณ์ความผันผวนของราคาหลักทรัพย์ด้วยโครงข่ายประสาทเทียมบนความน่าจะเป็นในการเปลี่ยนผ่านที่แปรผันตามเวลา. (Forecasting Stock Volatility with Neural Network on Time Varying Transition Probability) อ.ที่ปรึกษาหลัก : รศ. ดร.ไทยศิริ เวทไว

การพยากรณ์ความผันผวนของผลตอบแทนหลักทรัพย์มีความสำคัญทางด้านการเงินในหลากหลายด้าน เช่น การจัดกลุ่มหลักทรัพย์เพื่อการลงทุน การบริหารความเสี่ยง และการกำหนดกลยุทธ์การซื้อขายหลักทรัพย์ แบบจำลอง GARCH ถูกพัฒนามาอย่างยาวนานเพื่อใช้ในการศึกษาพลวัตที่แท้จริงของความผันผวนของผลตอบแทนหลักทรัพย์ การประยุกต์แนวความคิดการสลับเปลี่ยนแบบมาร์คอฟกับแบบจำลอง GARCH ช่วยให้สามารถระบุสาเหตุของภาวะความผันผวนสูงแบบชั่วคราวและการคงอยู่ของ shock ที่ส่งผลต่อเนื่องต่อความผันผวนที่เกิดขึ้นแบบชั่วคราวได้ ในสารนิพนธ์นี้ แบบจำลอง Markov switching GARCH ได้ถูกปรับปรุงต่อโดยประยุกต์ใช้แนวคิดของโครงข่ายประสาทเทียมในการช่วยประมาณความน่าจะเป็นในการเปลี่ยนผ่านที่แปรผันตามเวลาของแบบจำลอง การปรับปรุงแบบจำลองนี้มีจุดประสงค์เพื่อที่จะได้แบบจำลองที่สามารถพยากรณ์ความผันผวนของผลตอบแทนหลักทรัพย์ให้มีความแม่นยำมากขึ้น สารนิพนธ์นี้ศึกษาแบบจำลองดังกล่าวกับความผันผวนที่เกิดขึ้นจริง (Realized volatility) ของดัชนีราคาหุ้นตลาดหลักทรัพย์แห่งประเทศไทยและใช้ข้อมูลทางด้านการเงินและด้านเศรษฐกิจมหภาคเป็นส่วนหนึ่งในปัจจัยนำเข้าสำหรับโครงข่ายประสาทเทียม โดยใช้ข้อมูลตั้งแต่มกราคม พ.ศ. 2555 จนถึงธันวาคม พ.ศ. 2565 จากผลการทดสอบเชิงประจักษ์พบว่าแบบจำลองที่ใช้ Hidden layer 1 ชั้น ร่วมกับ Node จำนวน 3 ถึง 6 นั้นมีความเหมาะสม เนื่องจากสามารถปรับปรุงการพยากรณ์ได้แม้เพียงเล็กน้อย อย่างไรก็ตามผลลัพธ์ที่ดีขึ้นอย่างมีนัยสำคัญยังไม่ถูกพบอย่างชัดเจนรวมถึงแบบจำลองยังมีความเสี่ยงต่อปัญหา Overfitting

สาขาวิชา	วิศวกรรมการเงิน	ลายมือชื่อนิสิต	
ปีการศึกษา	2565	ลายมือชื่อ อ.ที่ปรึกษาหลัก	

6284064826 : MAJOR FINANCIAL ENGINEERING

KEYWORD: Neural Network, SET index, Realized Volatility, Volatility Forecasting

Wasit Norakarntiansin : Forecasting Stock Volatility with Neural Network on Time Varying Transition Probability. Advisor: Assoc. Prof. THAISIRI WATEWAI, Ph.D.

Forecasting volatilities of financial security returns are important for many financial applications e.g., portfolio investment construction, risk management and trading strategy. The GARCH model has long been refined to capture the true dynamic of volatility on a security return. By applying the Markov switching to the GARCH model, the source of the temporary high volatility and high persistence of a shock to the volatility can be captured. In this study, we refine the Markov switching GARCH model further by applying the notion of the neural network to approximate the time varying transition probabilities. We aim to achieve a model that provides better return volatility prediction. The realized volatility of the SET index is investigated while many financial-macro data are used as input factors for the neural network from January 2012 to December 2022. From the empirical results, the models of one hidden layer with 3 to 6 nodes are suggested since a minor improvement can be observed. However, the significant improvement is still unclear and the models are exposed to the overfitting problem.

Field of Study:	Financial Engineering	Student's Signature ....
Academic Year:	2022	Advisor's Signature ....

ACKNOWLEDGEMENTS

I am deeply grateful to my advisor, Dr. Thaisiri Watewai, for his support and guidance throughout my master's program. His expertise and patience have been invaluable to me and have played a crucial role in the success of this research project. I am also grateful to Dr. Tanawit Sae-Sue and Dr. Sira Suchintabandid for providing me invaluable suggestions and lessons to conduct my research.

Wasit Norakarntiansin

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

Chapter 1
Introduction

This study is about detecting the financial market regime, the state of market which is indirectly observed but conveys critical information for anyone who longs for opportunities. Among other ways, detecting market regimes can be achieved through a key measure of uncertainty in the financial market, volatility. Therefore, we can rely on the stylized facts of financial asset returns.

In the midst of working economic machines, the financial market has long been behaving like a complex system since there have been many participating parties that have been directly and indirectly involved with the market. With plenty of applications, the participants have observed prices of financial assets and have utilized them in many ways such as pricing financial derivatives, constructing portfolios, managing risks, etc. Since they all have to face financial market uncertainty, there have been great attempts to try to understand asset price dynamics in order to utilize observed data correctly. Nowadays, one of the challenging tasks is still forecasting and volatility forecasting turns out to be relatively more plausible than forecasting the direction of price movement, based on Danielsson (2011). Volatility has an impact on a wide range of applications and volatility forecasting task demands serious care on how a forecaster make assumptions on asset price dynamic. As pointed out by Tsay (2010), market volatility is unobservable directly, a forecaster or a modeler needs to rely on some of the stylized facts of financial asset returns. For example, there are periods when volatility stays high for a while and there are periods when it stays low and claim, referred to as volatility clustering. Volatility keeps changing over time but eventually, it reverts to some level or it seems stationary. Volatility reacts differently according to the direction of asset price movement, referred to as leverage effect. These characteristics can be widely observed in financial asset returns in many markets. The other famous stylized facts can be found in Cont (2001) which provides more stylized facts in detail.

Up until now, the common tools for modeling volatility have been GARCH model. It was designed to capture volatility clustering, mentioned above, with an aim to provide volatility measure (Engle (2001)). As surveyed by Bollerslev (2009), there was a long list for ARCH extension reflecting a variety of aspects in improving the ARCH process. However, one interesting and critical comment can be found in Hamilton and Susmel (1994) and Cai (1994) on the predictability problem which is due to the high persistence of shock to conditional volatility that the estimated GARCH model implies. Their main argument was that if the estimated model showed high persistence of change in stock price volatility, due to uncertainty, then why they still had an inaccurate forecast using the GARCH model? They pointed out to structural change in the financial market as a potential suspect and then proposed regime switching ARCH model, which implied that the high persistence was just temporary. There exist times when shock persistence is high, or high volatile regime, and times when shock persistence is low, or claim regime. The back and forth of staying in different market regimes, in the financial market, can be related to the notion of business cycles where there have been times for economic expansion, getting peak, decline, and recovery or it can be related to some structural change due to changes in regulatory laws, political risks, and wars. Many attempts have been made to improve the regime switching ARCH model since the model still suffered from path dependence problem in estimation process. The more appealing variants can be seen in Gray (1996), Dueker (1997), and Klaassen (2002) who modified the model and proposed methods to add Markov switching feature on GARCH processes, MS-GARCH model. The model allows us to have separate sets of GARCH parameters and allows for nonstationary volatility process for a highly volatile regime. The authors claimed that the model provided a better fit to data. We provide more details in the literature review part.

As for volatility, market regimes have been presumed to be unobserved. Hamilton (1989) proposed a method to infer regimes from time series data. Although he aims to analyze the business cycle of the postwar U.S. economic system, the method can be applied to infer financial market regimes, including regime inference via MS-GARCH model framework. The regime dynamic is typically based on a Markov chain process where the transitions, changing from one regime to the next regime, are governed by transition probabilities. Simple assumptions have been made for transition probabilities. They were assumed to be a set of constants that preserved nice properties of the Markov chain, and have commonly been treated as parameters to estimate. Filardo (1994) referred to Hamilton's model framework above as fixed transition probability Markov switching model, FTP-MS model, and argued that allowing the transition probabilities to vary across time did refine regime inference and prediction. He proposed time varying transition probability Markov switching model, TVTP-MS model, where the transition probabilities could depend on the dynamic of carefully preselected explanatory variables. The second reason for supporting TVTP-MS model is that the times, on average, that we stay in one regime can vary and depend on explanatory variables. This feature makes TVTP-MS model align with observed data and theoretical intuition. For the transition probability's part, a linear combination of explanatory variables, weighted by a set of parameters, is passed into a nonlinear function which has a range that aligns with the notion of probability e.g., logistic function or Normal CDF. And this is the point where this study is all about. In this study, we will change the way of exploring the relationship of explanatory variables to the time varying transition probability, and we will try to explore it under time varying transition probability Markov switching GARCH model or TVTP-MS-GARCH model.

In the field of artificial neural network, passing a sum of weighted explanatory variables into a nonlinear function can be viewed as a single hidden layer network or a shallow network. It can also be viewed as related to the generalized linear model in statistics. Referring to Zhang et al. (2021), although we can use a single hidden network to learn many functions and to explore relationship between variables, but it does not mean we should solve all problems using it. We can also use a deeper network to learn or approximate many functions. Utilizing the deeper network to approximate the transition probability part in TVTP-MS-GARCH model allows explanatory variables to have an impact on transition probability in a more sophisticated way by sacrificing interpretability. This extension may help refine a regime inference and may help provide a more realistic transition probability. The regime inference and regime prediction will have an impact on many applications e.g., refining trading strategy, forecasting volatility, improving hedging strategy, and constructing a portfolio. This extension may also help refine a volatility forecasting which is the main application in this study. Due to very scarce literature on the TVTP-MS-GARCH model, and not much literature paying attention to time varying transition probability part of the model, this study may help fill the gap.

1.1 Research Objective

Our aim is to refine the TVTP-MS-GARCH model to get a more accurate forecasted volatility and also to get a more reliable regime inference. The main research question is if we apply the notion of neural network[1] to approximate the transition probabilities, then do we achieve a model that provide better prediction out-of-sample? Note that we will separate data into three sets, the first 50% portion of a whole data set will be used for model estimation, referred to as the training set. The next 25% will be used for model selection, referred to as the validation set. The rest of the data will be used for accuracy evaluation, referred to as the test set or out-of-sample data.

For the training set, we will use it to estimate the model parameters. The set of explanatory variables will be selected. Many forms of neural network architecture will be estimated. Given the estimated models, we will use the data in the validation set to select the best model that provides the best accurate forecasted volatility. In this study, the notion of realized volatility[2] will be applied as suggested by Tsay (2010). We will use this measure for accuracy evaluation on the validation set and we will also use Mean Squared Error (MSE) and Mean Absolute Error (MAE) as the loss functions. The estimated model that provides forecasted volatility with the lowest MSE and MAE will be selected and will be used in the next step.

For the out-of-sample data period, the best model from above will be used for accuracy evaluation compared to our benchmark model. The FTP-MS-GARCH from the literature will be used as the benchmark model. We will use realized volatility for accuracy evaluation and also use MSE and MAE as the loss functions in the comparison. According to more flexibility of the neural network, the true relationship of explanatory variables to the transition probability may be partly captured or be approximated. We expect that the proposed model can reveal a reliable dynamic of transition probability and regime inference. This would result in higher accurate forecasts or lower MSE and MAE compared to the benchmark model.

The rest of this proposal is organized as follows: Chapter 2 provides background reasoning on why regime detection in the financial market is related to how volatility behavior changes. Some details about MS-GARCH model and the related problem issue are included. Background on time varying transition probability assumption is reviewed and examples on using TVTP-MS-GARCH model are

[1] The idea of neural network will be described later in Chapter 3.
[2] For example, to calculated daily realized volatility, the sum of the squared 30-minute log returns will be calculated on each trading day. For one week ahead volatility, we sum the calculated daily realized volatility over 5 days ahead.

provided. Chapter 3 provides detail of the proposed model and all methods and also describes the data that will be used and the scope of this study.

Chapter 2
Literature Reviews

In order to predict asset return volatility, we can rely on stylized facts and non-stationarity in return data series to back up our choice of tools, Markov switching GARCH (MS-GARCH) model, which has long been used and studied with interesting history.

2.1 From persistence of shock on conditional variance to regime switching

The beginning of MS-GARCH model came surprisingly from arguments against the founder of ARCH and GARCH model. In 1986, not too long after GARCH model had been introduced, Engle and Bollerslev (1986) proposed a new model of conditional variance which captured the persistence of shock on conditional variance, known as integrated GARCH (IGARCH) model. They spotted this feature from fitting GARCH model to long period weekly asset return series. The IGARCH model implied that multi-step forecasts did not revert back to unconditional variance due to contribution of current information on variance process persisted. Since stock volatility has an impact on stock price through risk premium, the persistence is critical to stock return modeling. Although both authors did provide sound evidence and aligned theories, there were studies that went against their idea.

The famous one was from Lamoureux and Lastrapes (1990), who investigated whether GARCH parameters were time varying or not. They estimated GARCH with daily stock returns and stock market index by allowing shift in constant term of GARCH process. Their results confirmed that there were structural changes in unconditional variance that caused the seemingly high persistence in conditional variance, which implied that there was a misspecification bias in using one regime GARCH model. They referred to Hamilton's regime switching model as a potential solution. By this motivation, Hamilton and Susmel (1994) investigated further on the context of forecasting performance, they showed that constant-variance assumption beat one regime GARCH model in forecasting. Even in a longer forecast period, the one regime GARCH provided no better performance to constant-variance assumption. So, this questions directly the evidence of high persistence of shock to conditional variance. Mikosch and Stărică (2004) also showed that it was non-stationarity in data, a change in unconditional variance, that caused one to observe IGARCH effect in the estimated GARCH model. Non-stationarity in data made some statistical tools provide spurious results especially when one assumed stationarity.

Hamilton and Susmel (1994) and Cai (1994) proposed a better model, Markov switching ARCH model, where the switching process was governed exogenously by the Markov chain. The model implied that, for two market regimes and time varying in ARCH parameters, we can seize the momentum in conditional variance, highly volatile market regime, and we can also seize the reversion to unconditional variance, claim market regime. At that time, they stated that using the GARCH form with the Markov switching were restricted by an estimation problem.

2.2 Markov switching GARCH model

Because the MS-ARCH model still did not satisfy researchers enough, they explored the methods to have a simple GARCH form instead of many lag terms on the information period for each regime. In this section, we will state common model setup, emphasize the estimation problem, and provide details on how researchers tackle the problem. Note that the model setup and all notations in this section are mainly adapted from Dueker (1997) and Marcucci (2005).

In the case of simple two market regimes, discrete state variable S_t at time t follows the Markov process with $S_t \in \{0,1\}$ for all of the time. The transition probabilities are stated as follows,

$$\begin{aligned} Prob(S_t = 0 | S_{t-1} = 0) &= e_0 \\ Prob(S_t = 1 | S_{t-1} = 0) &= 1 - e_0 \\ Prob(S_t = 1 | S_{t-1} = 1) &= e_1 \\ Prob(S_t = 0 | S_{t-1} = 1) &= 1 - e_1 \end{aligned} \tag{2.1}$$

The dependent variable r_t represents an asset return at time t and is assumed to have the following form,

$$\begin{aligned} r_t &= \mu_t + \epsilon_t \\ \epsilon_t &\sim iid(0, \sigma_t^2) \end{aligned} \tag{2.2}$$

Conditional mean μ_t is assumed to depend on state S_t. While ϵ_t is assumed to have zero mean and conditional variance σ_t^2. Dueker (1997) assumed $\sigma_t^2 = f(h_t)$ where the function f could take many forms. The variable h_t was assumed to follows GARCH (1,1) process with Markov switching parameters. The general form of h_t is stated as follows,

$$h_t(S_t, S_{t-1}, \dots, S_0) = \gamma(S_t) + \alpha(S_t)\epsilon_{t-1}^2 + \beta(S_t) h_{t-1}(S_{t-1}, S_{t-2}, \dots, S_0) \tag{2.3}$$

We will refer to this equation as the GARCH equation. Note that h_t in the general form depends on the entire history of state variable S_t. In model estimation, the calculation of h_t places too high burden on evaluation of the likelihood function with information up to time t and so this is called the path dependent problem for estimating MS-GARCH model. However, there are solutions.

For this case, Dueker (1997) collapsed the entire paths by setting h_t to depend only on S_t and S_{t-1}. We use $h_t^{m,n}$ for the value of h_t when $S_t = m$ and $S_{t-1} = n$. Let $\mathcal{F}_t$ denote the set of all information up to time t. To approximate equation 2.3, the author set up h_t^m to be the proxy for value of

h_t when $S_t = m$. The proxy h_t^m can be calculated by integrating out S_{t-1} from $h_t^{m,n}$, using probability that state variable $S_{t-1} = n$ given $\mathcal{F}_t$ and the value of $S_t = m$. The description can be stated as follow,

$$h_t^m(S_t) = \sum_{n=0}^{1} Prob(S_{t-1} = n | S_t = m, \mathcal{F}_t) h_t^{m,n}(S_t, S_{t-1}) \qquad (2.4)$$

The author used equation 2.4 to simplify h_{t-1} on the right-hand-side of equation 2.3. So, the GARCH equation 2.3 can be approximated by using the following equation,

$$h_t^{m,n}(S_t, S_{t-1}) = \gamma(S_t) + \alpha(S_t)\epsilon_{t-1}^2 + \beta(S_t) h_{t-1}^n(S_{t-1}) \qquad (2.5)$$

The variable h_t now depends on S_t and S_{t-1} instead of the entire path history of S_t. With this set up, estimation can be done using the idea from Hamilton (1990). Dueker (1997) applied the method with many conditional variance specifications $\sigma_t^2 = f(h_t)$. Recall that function f could take many forms. The author used the models to study daily percentage changes of the S&P 500 Index. His models showed improvement on fitting with the data by comparing to one regime GARCH and MS-ARCH of Hamilton and Susmel (1994). However, the results showed that his MS-GARCH framework still did not pass a goodness-of-fit test for the out-of-sample period.

Another resemble method can be found in Gray (1996) which showed another way to tackle the path dependent problem by integrating out the path of state variable S up to time $t-1$. Let ρ_t^m denote the probability that $S_t = m$, given information up to time $t-1$, or $\rho_t^m = Prob(S_t = m | \mathcal{F}_{t-1})$. Recall that $S_t \in \{0,1\}$ then we have $\rho_t^0 = Prob(S_t = 0 | \mathcal{F}_{t-1})$ and $\rho_t^1 = 1 - \rho_t^0$. We use $\mu_t^2(S_t = m)$ to denote the square of conditional mean at time t when $S_t = m$. From equations 2.2 and 2.3 above, Gray (1996) simply assumed $\sigma_t^2 = f(h_t) = h_t$, but we state the difference by using specific notations as following. We use $\tilde{h}_t^m(S_t)$ to denote conditional variance at time t when $S_t = m$ and we also use $\tilde{h}_t(S_t)$ for a proxy of conditional variance under Gray (1996) method. The value of $\tilde{h}_t(S_t)$ can be calculated as follows,

$$\begin{aligned} \tilde{h}_t(S_t) &= E[r_t^2 | \mathcal{F}_{t-1}] - (E[r_t | \mathcal{F}_{t-1}])^2 \\ \tilde{h}_t(S_t) &= \rho_t^0\left(\mu_t^2(S_t = 0) + \tilde{h}_t^0\right) + \rho_t^1\left(\mu_t^2(S_t = 1) + \tilde{h}_t^1\right) \\ &\quad - [\rho_t^0 \mu_t(S_t = 0) + \rho_t^1 \mu_t(S_t = 1)]^2 \end{aligned} \qquad (2.6)$$

Using equation 2.6 to approximate h_{t-1} on the right-hand-side of equation 2.3, we have the following approximation:

$$\tilde{h}_t^m(S_t) = \gamma(S_t) + \alpha(S_t)\epsilon_{t-1}^2 + \beta(S_t)\tilde{h}_{t-1}(S_{t-1}) \qquad (2.7)$$

Now the $\tilde{h}_t$ in equation 2.7 does not depend on the entire history of state variable.

Like in Dueker (1997), the estimation method follows the idea from Hamilton (1990). We can observe that both methods above are not too different. Gray (1996) also showed good performance of the model in fitting with data. The author did investigate the model on short term U.S. interest rate data and claimed that his MS-GARCH framework beat one regime GARCH for an out-of-sample forecasting test.

The main difference for these two methods is how they integrate out the history path of state variable S and how they calculate the involved probabilities. However, the key feature is that, as stressed by Marcucci (2005), the modified method allows switching in regime to cause the persistence of shock to volatility. So, with these modifications, there are two sources for shock on volatility persistence, one come from within regime and one from switching to high volatile regime.

The advantage on using Gray's (1996) model specification is that we can extend the model to investigate the case where the transition probabilities are time varying. The details on this point will be explained in the next section below. However, there are limitation on flexibility of the model to estimate the transition probabilities and our proposed model will refine it. We provide more details in Chapter 3.

There are other forms of MS-GARCH model. But in this study, it is suitable to start from Gray (1996) version since the author clearly provided a method to allow for time varying transition probability assumption which is the main feature for our study.

2.3 Time varying transition probability assumption

The model assumption in equation 2.1 can be traced back to Hamilton (1989) in which the notion of the Markov chain was used to model turning points of the underlying states of the U.S. economy. The assumption states that the probability of switching to the next regime depends only on the current regime. In model estimation, the value of the transition probabilities, in equation 2.1 above, were treated as model parameters to be estimated. We can call the model in the previous part as fixed transition probability MS-GARCH or FTP-MS-GARCH in short. Filardo (1994) and Diebold et al. (1994) generalized the probability of switching to depend on the current regime and a set of explanatory variables. Note that they did not apply the notion on MS-GARCH model in their studies. They investigated on a fixed transition probability Markov switching (FTP-MS) model and a time varying transition probability Markov switching (TVTP-MS) model but their ideas could be applied directly to MS-GARCH model in our study. To illustrate, let X_t denote the vector of the value of the explanatory variables at time t. Let $\psi_{S_{t-1}}$ denote a function that can map a set of real numbers to range [0,1], to align with the properties of probability. The modified equation 2.1 can be stated as follows,

$$\begin{aligned}
&Prob(S_t = 0|S_{t-1} = 0, X_{t-1}) = e_{0t} = \psi_0(a_0 + b_0' X_{t-1}) \\
&Prob(S_t = 1|S_{t-1} = 0, X_{t-1}) = 1 - e_{0t} \\
&Prob(S_t = 1|S_{t-1} = 1, X_{t-1}) = e_{1t} = \psi_1(a_1 + b_1' X_{t-1}) \\
&Prob(S_t = 0|S_{t-1} = 1, X_{t-1}) = 1 - e_{1t}
\end{aligned} \quad (2.8)$$

where a_s and b_s, $s = 0, 1$ are the model parameters. Notice that p_t and q_t now depend on X_{t-1}. According to Tsay (2010), the return tends to stay in the state $S_{t-1} = 0$ with expected duration $1/(1 - e_{0t})$ and it tends to stay in the state $S_{t-1} = 1$ with expected duration $1/(1 - e_{1t})$. If the strength point of using Markov switching is creating meaningful forecasts with flexibility as claimed by Hamilton (1994), then this generalization can provide further flexibility.

Filardo (1994) also provided supporting reasons which can be summarized as follow. First, the model allowed transition probability to rise just before and after a change of an economic regime and this provided model flexibility adding on to the Hamilton's Markov switching model. For the second reason, the model opened to more complex temporal persistence of the investigated dependent variable. Since both models, FTP-MS model, and TVTP-MS model, can capture two sources of the persistence, one from autoregressive form of the dependent variable within each regime and another one from regime persistence. But the regime persistence of the TVTP-MS model depends on explanatory variables and then the source of regime persistence comes partly from the dynamic of the explanatory variables. The final reason is that the expected duration of staying at each regime under the FTP-MS model depends on the estimated values of the transition probability. Allowing them to be time varying and to depend on explanatory variables also provides more intuitive interpretation of the expected duration. These advantages of the TVTP-MS model are directly linked to the inference of the market regime. Diebold et al. (1994) conducted a simulation, under the TVTP-MS model, and showed that regime detection was improved in the case that we know the right explanatory variables. This implies that explanatory variables selection is critical to the success of the model.

All these points above will follow when we apply the extension to MS-GARCH. Gray (1996) generalized FTP-MS-GARCH to achieve time varying transition probability Markov switching GARCH (TVTP-MS-GARCH) model, and showed that the extension improved performance in fitting with data. The author also pointed out that the source of persistence of shock to conditional variance can depend on expanding phase of high volatile regime, indicated by time vary transition probability. So, the empirical result that was observed in Engle and Bollerslev (1986), and other studies, can be captured by TVTP-MS-GARCH model, which raises the important role of market regime inference. Although good results were seen, the promising result was still in question due to limited studies on this model. Some of studies that utilized and supported the model are described below.

Brunetti et al. (2008) studied market regime of exchange rate market using TVTP-MS-GARCH model. The authors found that exchange rate, in southeast Asia, exhibited calm phase and turbulent phase by distinguishing conditional mean and conditional variance processes. They also showed that macro and financial variables such as real effective exchange rate, money supply, stock index return, and volatility contained information to detect market regime.

Tan et al. (2021) utilized daily trading volume and the number of daily Google searches, “Bitcoin”, as explanatory variables for varying transition probability of Bitcoin volatility regimes. The authors showed that TVTP-MS-GARCH with only trading volume and Student-t error distribution could beat one regime GARCH, GJR-GARCH, and FTP-MS-GARCH for out-of-sample forecast.

Wang et al. (2021) studied the impact of geopolitical risk on volatility shift in crude oil dynamic using an asymmetric time-varying transition probability Markov regime switching GARCH (AS-TVTP-MS-GARCH) model. The authors used geopolitical risk index as explanatory variable and modified TVTP-MS-GARCH to capture the different impacts of positive and negative changes in geopolitical risk on crude oil volatility regimes. They compared the smoothed probabilities of different regimes for MS-GARCH, TVTP-MS-GARCH, and AS-TVTP-MSGARCH and claimed that the proposed model gave a better in-sample fit and the geopolitical risk index conveyed information to regime detection. Later, Hong et al. (2022) extended the investigation on international crude oil market using financial stress index but the authors focused on out-of-sample forecasting. They claimed that financial stress has affected the supply-demand structure of crude oil, and so the model with an asymmetric feature provided better performance in forecasting.

Chapter 3
Methodology and Data

3.1 Methodology: The proposed model

For the proposed model specification, let r_t denote log-return of asset price at time index t. Here, we use weekly return since daily return may contain more noise. Assume that we have a discrete state variable S_t which represents the market regime at time t and it follows a first order Markov process. Note that the value of state variable $S_t \in \{0, 1\}$ for all time t. When $S_t = 0$, the concerning asset return is assumed to stay in low a volatile market regime at time period t. When $S_t = 1$, the return is assumed to stay in a high volatile market regime. We use $\mu(S_t)$ to denote the conditional mean of asset return at time t which depends on state variable S_t. We will add a subscript i to represent the value of the variables when state variable $S_t = i$. So, we use μ_i for the value of $\mu(S_t)$ when the value of state variable $S_t = i$ and we will apply this logic to all other notations.

Let σ_t^2 denote the conditional variance at time t. We use ε_t to represent the innovation term and we assume it to be identically distributed with standard normal distribution for simplicity. We use ϕ_t to represent the diffusion part of asset return. The first element of the proposed model, when $S_t = i$, is specified as follows,

$$r_t = \mu_i + \phi_t \tag{3.1}$$

$$\phi_t = \sigma_t \varepsilon_t \tag{3.2}$$

$$\varepsilon_t \sim iid\ N(0,1)$$

$$\sigma_t^2 = \omega_i + \alpha_i \phi_{t-1}^2 + \beta_i \sigma_{t-1}^2 \quad (3.3)$$

Equation 3.3 is referred to as the GARCH equation which is GARCH (1,1) with Markov switching parameters, denoted as $(\omega(S_t), \alpha(S_t), \beta(S_t))$. The GARCH equation 3.3 is assumed to follow Gray's (1996) specification and σ_{t-1}^2 on the right-hand side will be approximated to get a state-independent average of past conditional variance later in this section. Note that σ_t and hence ϕ_t depend on the current state S_t.

For the second element of the proposed model, we will define notations for variables, and the probabilities that will be needed later. Let $\mathcal{F}_t$ denote the set of asset return information up to time t. We use $\mathcal{M}_t$ to denote the set of all explanatory variables information up to time t. Assume that we consider m explanatory variables and we use x_{mt} for value of the m^{th} explanatory variable at time t. That is,

$$\mathcal{F}_t = \{r_0, r_1, r_2, \dots, r_t\} \quad (3.4)$$

$$\mathcal{M}_t = \{x_{10}, x_{20}, \dots, x_{m0}, \dots, x_{1t}, x_{2t}, \dots, x_{mt}\} \quad (3.5)$$

In this study, we follow Gray's (1996) specification to have time varying transition probabilities and we propose an extension on this part. We use $\boldsymbol{X}_t$ for the vector of the values of the explanatory variables at time t. To simplify the notations, we augmented the first element of vector $\boldsymbol{X}_t$ to be value 1, illustrated below.

$$\boldsymbol{X}_t = \begin{bmatrix} 1 \\ x_{1t} \\ \vdots \\ x_{mt} \end{bmatrix} \quad (3.6)$$

According to Zhang et al. (2021), given a set of explanatory variables we can apply a neural network method. The neural network architecture composes of input layer, hidden layer, and output layer. The input layer is where the information of the values of the explanatory variables are passed into the neural network. If we pick a set of information that contains m values then each value can be called a node. In this case, our input layer has m nodes. The input layer only passes information to the next layer which is the hidden layer. A hidden layer takes inputs from the input layer or from the other hidden layers. It processes the information and sends output to the next layer which can be another hidden layer or an output layer. One neural network architecture can possess many hidden layers. One hidden layer can take many nodes, each of which is where the input data are processed and creates output data per node. The output layer takes information, which is the output data from the hidden layer, and creates the desired final output. The output layer also can possess many nodes.

In our context, vector $\boldsymbol{X}_{t-1}$ can represent an input layer that contains m nodes. Each input node represents data at time $t-1$, so we can actually utilize a set of data within $\mathcal{M}_{t-1}$ for each time period. For example, assume we consider 1 hidden layer with 3 nodes. We use $\boldsymbol{\theta}_{jk}$ to represent a vector of weights and a bias term of the j^{th} hidden layer at the k^{th} node. We use θ_{jkl} for the value of the weight of the j^{th} hidden layer, k^{th} node, and the corresponding l^{th} input. Note that the bias terms are denoted by θ_{jk0} which corresponds with value 1 that we have augmented in $\boldsymbol{X}_t$ above. In our example, the vector of weights and the bias term of node k can be stated as follows,

$$\boldsymbol{\theta}_{1k}= \begin{bmatrix} \theta_{1k0} \\ \theta_{1k1} \\ \vdots \\ \theta_{1km} \end{bmatrix} \tag{3.7}$$

$$bias\ term = \ \theta_{1k0} \tag{3.8}$$

Each node in the hidden layer processes input by setting up a linear combination of vector $\boldsymbol{X}_{t-1}$ and $\boldsymbol{\theta}_{1k}$. Then the linear combinations are passed into the activation function. In the literature, the activation function can take many forms and depends on the context of the study or researcher's desire. In this study, we follow Gray (1996) and use the logistic function as an activation function for simplicity. Let φ_{jk} denote the activation function that we apply at the j^{th} layer and k^{th} node. We use z_{jk} for the output value that we get from applying φ_{jk} to the input. In our example, the one hidden layer with 3 nodes processes input and sends out a set of z_{jk} as output to the next layer. This can be illustrated as follows,

$$\begin{aligned} z_{11}&= \varphi_{11}(\boldsymbol{\theta}_{11}{}'\boldsymbol{X}_{t-1}) \\ z_{12}&= \varphi_{12}(\boldsymbol{\theta}_{12}{}'\boldsymbol{X}_{t-1}) \\ z_{13}&= \varphi_{13}(\boldsymbol{\theta}_{13}{}'\boldsymbol{X}_{t-1}) \end{aligned} \tag{3.9}$$

The output layer takes the output from the hidden layer and processes them to create the final result. In our context, we use the explanatory variables and neural network to approximate transition probabilities so the activation function at the output layer is required to have a range from 0 to 1 to align with the notion of probability. Let $\boldsymbol{Z}_n$ denote a vector containing the output from each node of the n^{th} hidden layer and $\boldsymbol{Z}_n$, likes vector $\boldsymbol{X}_{t-1}$, is always augmented by value 1 at the first element. Let h denote the number of nodes in the output layer. Then we use $\widetilde{\boldsymbol{\theta}}_h$ for a vector of weights and a bias term of the output layer at the h^{th} node. Let $\tilde{\varphi}_h$ denote the activation function that we apply at the output layer and at the h^{th}node. Since we assume 2 market regimes, we can set 2 nodes for the output layer. For our example, the vector $\boldsymbol{Z}_n$ is stated below as equation 3.10. The transition probabilities, denoted by p_t^{ij} for the probability that $S_t = i$ given $S_{t-1} = j$, and the information up to time $t-1$, and the neural network architecture in our example can be stated as follows,

$$\boldsymbol{Z}_1= \begin{bmatrix} 1 \\ z_{11} \\ z_{12} \\ z_{13} \end{bmatrix} \tag{3.10}$$

$$p_t^{00}= \tilde{\varphi}_1\left(\widetilde{\boldsymbol{\theta}}_1'\boldsymbol{Z}_1\right) \tag{3.11}$$

$$p_t^{11}= \tilde{\varphi}_2\left(\widetilde{\boldsymbol{\theta}}_2'\boldsymbol{Z}_1\right) \tag{3.12}$$

Note that in this example, $[3(m+1)+2(3+1)-2]$ more parameters are needed to be estimated compared to the two regimes FTP-MS-GARCH model.

To state the model extension clearly, let $\boldsymbol{\varphi}$ denote a whole neural network architecture and $\boldsymbol{\theta}$ denote all weights and bias terms from $\boldsymbol{\varphi}$**.** We use $\boldsymbol{\varphi}(\boldsymbol{\theta}, \boldsymbol{X}_t)|_{node=i}$ for output value from the i^{th} node of the output layer given all weights, bias terms, and a set of input data at time period t. The transition probability for our proposed model can be stated as follows,

$$
\begin{aligned}
&Prob(S_t = 0|S_{t-1} = 0, \mathcal{M}_{t-1}) = p_t^{00} = \boldsymbol{\varphi}(\boldsymbol{\theta}, \boldsymbol{X}_{t-1})|_{node=1} \\
&Prob(S_t = 1|S_{t-1} = 0, \mathcal{M}_{t-1}) = 1 - p_t^{00} \\
&Prob(S_t = 1|S_{t-1} = 1, \mathcal{M}_{t-1}) = p_t^{11} = \boldsymbol{\varphi}(\boldsymbol{\theta}, \boldsymbol{X}_{t-1})|_{node=2} \\
&Prob(S_t = 0|S_{t-1} = 1, \mathcal{M}_{t-1}) = 1 - p_t^{11}
\end{aligned}
\tag{3.13}
$$

The conditional mean $\mu(S_t)$ from equation 3.1, the GARCH equation's parameters $(\omega(S_t), \alpha(S_t), \beta(S_t))$ from equation 3.3, and all weights and bias terms within a neural network architecture $\boldsymbol{\theta}$ will be referred to as model parameters. The number of the hidden layers and the number of nodes of each hidden layer are treated as hyperparameters. Note that in this study, many combinations of the hyperparameter will be set up. The example above is just one model set up.

In this study, we use FTP-MS-GARCH model as the benchmark model, so the transition probability for our benchmark model can be stated as follows,

$$
\begin{aligned}
&Prob(S_t = 0|S_{t-1} = 0) = e_0 \\
&Prob(S_t = 1|S_{t-1} = 0) = 1 - e_0 \\
&Prob(S_t = 1|S_{t-1} = 1) = e_1 \\
&Prob(S_t = 0|S_{t-1} = 1) = 1 - e_1
\end{aligned}
\tag{3.14}
$$

The fixed transition probabilities e_0 and e_1 are part of the model parameters of the benchmark model.

3.2 Methodology: Estimation, model selection and accuracy evaluation

We will apply the filtering method from Hamilton (1990). First, we set the probability of being at market regime i at time t, denoted by p_{it}, as follows,

$$
\begin{aligned}
&p_{0t} = Prob(S_t = 0|\mathcal{F}_{t-1}, \mathcal{M}_{t-1}) \\
&p_{1t} = Prob(S_t = 1|\mathcal{F}_{t-1}, \mathcal{M}_{t-1}) = 1 - p_{0t}
\end{aligned}
\tag{3.15}
$$

To continue from equation 3.3 above, the path dependent problem is solved following Gray's (1996) method. For equation 3.16 below, we use $\sigma_{i,t}^2$ to denote the value of the conditional variance of the return at time t given $S_t = i$. The conditional variance σ_{t-1}^2 from the right-hand side of equation 3.3 can be approximated as follows,

$$Variance(r_{t-1}|\mathcal{F}_{t-2},\mathcal{M}_{t-2}) = E[r_{t-1}^2|\mathcal{F}_{t-2},\mathcal{M}_{t-2}] - E[r_{t-1}|\mathcal{F}_{t-2},\mathcal{M}_{t-2}]^2$$

$$\sigma_{t-1}^2 \approx p_{0t-1}(\mu_0^2 + \sigma_{0,t-1}^2) + (1 - p_{0t-1})(\mu_1^2 + \sigma_{1,t-1}^2) - [p_{0t-1}\mu_0 + (1 - p_{0t-1})\mu_1]^2 \quad (3.16)$$

Gray (1996) also approximated the variable ϕ_{t-1} as follows,

$$\phi_{t-1} \approx r_{t-1} - E[r_{t-1}|\mathcal{F}_{t-2},\mathcal{M}_{t-2}]$$

$$\phi_{t-1} \approx r_{t-1} - [p_{0t-1}\mu_0 + (1 - p_{0t-1})\mu_1] \quad (3.17)$$

Note that we only use equations 3.16 and 3.17 to approximate the terms in the right-hand side of equation 3.3. Next, we evaluate the term p_{it}, from equation 3.15, only for the case when $S_t = 0$ as follows:

$$\begin{aligned} p_{0t} &= Prob(S_t = 0|\mathcal{F}_{t-1},\mathcal{M}_{t-1}) \\ &= \sum_{i=0}^{1} Prob(S_t = 0|S_{t-1} = i,\mathcal{F}_{t-1},\mathcal{M}_{t-1}) Prob(S_{t-1} = i|\mathcal{F}_{t-1},\mathcal{M}_{t-1}) \\ &= p_t^{00} Prob(S_{t-1} = 0|\mathcal{F}_{t-1},\mathcal{M}_{t-1}) + (1 - p_t^{11}) Prob(S_{t-1} = 1|\mathcal{F}_{t-1},\mathcal{M}_{t-1}) \\ &= p_t^{00} Prob(S_{t-1} = 0|\mathcal{F}_{t-1},\mathcal{M}_{t-1}) + (1 - p_t^{11})[1 - Prob(S_{t-1} = 0|\mathcal{F}_{t-1},\mathcal{M}_{t-1})] \end{aligned} \quad (3.18)$$

To evaluate the term $Prob(S_{t-1} = 0|\mathcal{F}_{t-1},\mathcal{M}_{t-1})$, we assume that given the set of information $\mathcal{F}_{t-1}$ and $\mathcal{M}_{t-1}$, it is sufficient to evaluate the probability that $S_{t-1} = 0$ as follows:

$$\begin{aligned} Prob(S_{t-1} = 0|\mathcal{F}_{t-1},\mathcal{M}_{t-1}) &= Prob(S_{t-1} = 0|r_{t-1},\mathcal{F}_{t-2},\mathcal{M}_{t-2}) \\ &= \frac{Prob(r_{t-1},S_{t-1} = 0|\mathcal{F}_{t-2},\mathcal{M}_{t-2})}{Prob(r_{t-1}|\mathcal{F}_{t-2},\mathcal{M}_{t-2})} \\ &= \frac{Prob(r_{t-1}|S_{t-1} = 0,\mathcal{F}_{t-2},\mathcal{M}_{t-2}) Prob(S_{t-1} = 0|\mathcal{F}_{t-2},\mathcal{M}_{t-2})}{\sum_{i=0}^{1} Prob(r_{t-1},S_{t-1} = i|\mathcal{F}_{t-2},\mathcal{M}_{t-2})} \\ &= \frac{Prob(r_{t-1}|S_{t-1} = 0,\mathcal{F}_{t-2},\mathcal{M}_{t-2}) Prob(S_{t-1} = 0|\mathcal{F}_{t-2},\mathcal{M}_{t-2})}{\sum_{i=0}^{1} Prob(r_{t-1}|S_{t-1} = i,\mathcal{F}_{t-2},\mathcal{M}_{t-2}) Prob(S_{t-1} = i|\mathcal{F}_{t-2},\mathcal{M}_{t-2})} \end{aligned}$$

(3.19)

The term $Prob(r_{t-1}|S_{t-1} = i,\mathcal{F}_{t-2},\mathcal{M}_{t-2})$ can be evaluated as follows,

$$Prob(r_{t-1}|S_{t-1} = i,\mathcal{F}_{t-2},\mathcal{M}_{t-2}) = \frac{1}{\sqrt{2\pi}\sigma_{i,t-1}} exp\left\{\frac{-(r_{t-1} - \mu_i)^2}{2\sigma_{i,t-1}^2}\right\} \quad (3.20)$$

In this study, we use realized volatility as the target output. To calculate daily realized volatility, the sum of the squared 30-minute log returns will be calculated on each trading day. For one week ahead volatility, we sum the calculated daily realized volatility over 5 days ahead. This quantity will be included in a whole data set which will be described in Section 3.3 below. Given the realized volatility as the target output and the estimated conditional variance, which will be calculated by using equation 3.16 above, the difference between these two quantities will be used to set the objective function. To estimate the model parameters, we use Mean Squared Error (MSE) as the objective function. We will minimize the objective function by altering all model parameters. For our proposed model, however, the analytical gradient and hessian are hard to formulate but we can rely on numerical approximation. So, in this study, the minimization problem will be done numerically. Note that we will initialize all weights and bias terms within a neural network architecture $\boldsymbol{\theta}$ by using standard normal distribution. Re-initialize the model parameters are required for each model.

The whole data set will be separated into a training set, a validation set, and a test set. The training set will be used in the model estimation. We will select a set of combinations of hyperparameters. One model has one combination of hyperparameters. For example, if we use 4 explanatories data as inputs to a neural network with 3 nodes in the first hidden layer and 5 nodes in the second hidden layer, so we can express the combination as (4-3-5-2) to represent the model. Recall that we have 2 nodes in the output layer of our proposed model. The hyperparameters of each model are fixed along the accuracy evaluation processes. All variants of our proposed model will be trained using all data from the training set.

For model prediction, we update equation 3.13 by using $\mathcal{M}_t$ and utilizing the estimated model. Next, we use the updated transition probability, $\mathcal{F}_t$, $\mathcal{M}_t$, and p_{it} to calculate the predicted probability $\hat{p}_{it+1}$ by using equations 3.18 and 3.20. For given $\mathcal{F}_t$ and the estimated model, we also can update equations 3.16 and 3.17, and then we can update GARCH equation 3.3. Let $\hat{\sigma}^2_{it+1|t}$ denote the one-step ahead forecast variance at time $t+1$ from updating GARCH equation 3.3 given $S_t = i$, $\mathcal{F}_t$, and $\mathcal{M}_t$. Let $\hat{\sigma}^2_{t+1|t}$ denote the one step ahead forecast variance given $\mathcal{F}_t$ and $\mathcal{M}_t$. We can use $\hat{p}_{it+1}$, $\hat{\sigma}^2_{it+1|t}$, and equation 3.16 to calculate $\hat{\sigma}^2_{t+1|t}$ as follow,

$$\begin{aligned}\hat{\sigma}^2_{t+1|t} = \hat{p}_{0t+1}(\hat{\mu}_0^2 + \hat{\sigma}^2_{0t+1|t}) + (1-\hat{p}_{0t+1})(\hat{\mu}_1^2 + \hat{\sigma}^2_{1t+1|t}) \\ - [\hat{p}_{0t+1}\hat{\mu}_0 + (1-\hat{p}_{0t+1})\hat{\mu}_1]^2\end{aligned} \tag{3.21}$$

After we estimate all models, we will use the estimated models and equation 3.21 above to calculate one step ahead forecast. Next, we update the training set by including one time step data from the validation set. We will re-estimate all models and calculate one step ahead forecast. Note that all of the model parameters will be re-estimated while all of the hyperparameters will be fixed. These steps will be repeated until we utilize all data in the validation set. MSE and MAE will be used to measure accuracy of the forecasted volatility on realized volatility. The model that provides the lowest MSE and MAE value will be selected and will be used for the next accuracy evaluation step. Note that we can have one model with the lowest MSE and another one with the lowest MAE.

The test set data or out-of-sample data will be used for the last accuracy evaluation process. In this process, we use all data from the training set and validation set to estimate the benchmark model.

Note that we use only asset return data and realized volatility for the benchmark model estimation which is the same method to our proposed model. Given the best model from the previous model selection process and the estimated benchmark model, we will use them to calculate one step ahead forecasts. Next, we update the data set, that we have already used in estimation, to include one-time step data from the test set. We will re-estimate both models and calculate one step ahead forecasts. These steps will be repeated until we utilize all data in the test set.

Note that one limitation of Gray's (1996) method is that the multi-step-ahead volatility forecasts turn out to be too complicated. So, we focus on one-step-ahead forecast, or one week ahead forecast in this study. To answer our research question, we use $\hat{\sigma}^2_{w+1|w}$ for the one week ahead forecasted volatility, given that we utilize all data up until week w, from the benchmark model and the selected model. We denote the calculated realized volatility, one week ahead from week w, by $\hat{h}_{w+1|w}$. MSE and MAE[3] will be used in accuracy evaluation for the given n periods out-of-sample data. The MSE and MAE will be calculated as follows,

$$MSE_1 = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\sigma}^2_{w+1|w} - \hat{h}_{w+1|w}\right)^2 \quad (3.22)$$

$$MSE_2 = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{\sigma}_{w+1|w} - \hat{h}^{1/2}_{w+1|w}\right)^2 \quad (3.23)$$

$$MAE_1 = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{\sigma}^2_{w+1|w} - \hat{h}_{w+1|w}\right| \quad (3.24)$$

$$MAE_2 = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{\sigma}_{w+1|w} - \hat{h}^{1/2}_{w+1|w}\right| \quad (3.25)$$

3.3 Data and scope of study

In this study, we will investigate our proposed model described in the previous section on Thailand stock market index. The main dependent variable r_t is the weekly log return on the SET index. Our target data set starts from January 2012 to the end of December 2022. The realized volatility will be calculated on the whole data set.

We will set the first 50% of the data set as the training set, from January 2012 to the end of June 2017, for the model estimation purpose. The next 25%, from July 2017 to the end of March 2020, will be used for model selection and validation. The data is referred to as the validation set. The rest of the data set, referred to as the test set, will be used as the out-of-sample period for the accuracy evaluation purpose. For explanatory variables X_t, some candidates are proposed as follows,

---

[3] Note that the main difference between MSE and MAE is that MSE will place more serious weight on the outlier or large forecast error. So, they provide different interpretations which depend on how the outlier affect the use of the forecasted value in an application.

- Trading volume of the SET index. We can refer to the study of Tan et al. (2021) which was described in Section 2.3.
- The volatility of other leading markets. Aloy et al. (2014) studied the shift-volatility in East Asian stock market that came from the regional market or came from larger global market. They used volatility of some leading markets such as S&P, Nikkei, Hang Seng, and Singapore indices as explanatory variables in TVTP-MS of market return. They also found that Hongkong, Singapore, and Japan markets were influenced mainly by volatility of the U.S. market while Malaysia, Philippines, and Thai markets were influenced mainly by volatility of Hongkong and Singapore markets. In this study, we will use the realized volatility of both markets.
- Term spread from short-term government bond yield and the long-term yield, 10Y-2Y and 5Y-2Y. Estrella (2005) claimed that these variables could be used as a leading indicator for economic downturn. So, this might also help in detecting regime for the stock market.
- Active Thai CDS prices. These prices may reflect foreign-investor confidence on the Thai macroeconomic and political risk.
- There are macro-financial variables such as spread of Thai interest rate and the U.S. interest rate, oil price, or exchange rate. These variables may have an impact on the expectation of investors to Thailand macroeconomic and the stock market. In this study, we will use Effective federal funds rate (EFFR), exchange rate (USB/THB), oil prices, gold price and spread between EFFR and the BOT monetary policy rate.

All the data in this study will come from data providers such as Bloomberg and Tradingview.com.

Chapter 4
Results

4.1 Observed data

In this study, we use realized volatility of the SET index as the target output which is calculated by using 30-minute historical close data. We follow the calculation from Tsay (2010) and include all variations that occurred from the market close and re-open session. The realized volatility of the HSI index and the STI index are calculated the same way but they are used as a part of input factors. All of the 30-minute historical close data come from Tradingview.com, while the rest of the historical weekly data come from Bloomberg. Table 1 below summarizes all of the observed data from January 2012 to the end of December 2022.

Table 1: Observed data

Roles	Data name	Calculation	Number of observations
Dependent variable	SET index	Log-return	34,525*

Roles	Data name	Calculation	Number of observations
Target output	SET index	Realized volatility	
Input factor (1)	SET index trading volume	Standardization	
Input factor (2)	Thai government bond spread 10Y-2Y	Standardization	574
Input factor (3)	Thai government bond spread 5Y-2Y	Standardization	
Input factor (4)	HSI index	Realized volatility, Standardization	34,936*
Input factor (5)	STI index	Realized volatility, Standardization	43,910*
Input factor (6)	HSI index	Log-return, Standardization	34,936*
Input factor (7)	STI index	Log-return, Standardization	43,910*
Input factor (8)	Thai 5Y CDS rate	First difference, Standardization	
Input factor (9)	Effective federal funds rate (EFFR)	First difference, Standardization	
Input factor (10)	Exchange rate (USD/THB)	Log-return, Standardization	
Input factor (11)	Oil price (WTI)	Log-return, Standardization	574
Input factor (12)	Oil price (BRENT)	Log-return, Standardization	
Input factor (13)	Gold price	Log-return, Standardization	
Input factor (14)	Spread between EFFR and BOT rate	Standardization	

* The data are calculated to get a weekly basis before they are used in the model estimation.

For the Thai 5Y CDS rate and EFFR, the directions of change are more matter than the level itself in our context of this study. So, we apply first difference method for them. The same reason is applied to all of the prices but we use log-return method to get the direction of change. The main reason that we standardize all of the input factors is to prevent the gradient vanishing problem when using backpropagation to train our model. In the model estimation, we load all required data from the training set, standardize all the input factors, and then start the estimation process. For the model validation test, we reload all required data from the training set and include data from the validation set for one-time step. We standardize the input factors from the reloaded data, and then restart the estimation process. These steps are repeated until we utilize all data in the validation set. For the out-of-sample data set, we use the same method.

4.2 Neural network configuration and model training

Before we train our model, we need to decide on the number of the hidden layers and the number of nodes in each hidden layer. Since the optimal structure of feed forward network architecture is still an open research question, we follow the trial-and-error scheme. However, there is the suggestion that problems which require more than one hidden layer are rare. So, It is suggested to start with one hidden layer (Heaton, 2008).

Heaton (2008) also suggested that the nodes in the hidden layer should be between the number of input factors and the number of outputs. The author stated that too few nodes might create the underfitting problem while too many nodes would cause the overfitting problem. There were no widely accepted rules on the issue. Although we have 14 input factors, we will try up to 20 nodes in this study. The decisions are made due to the limited resources for training our model.

In training our model, we need to randomly initialize the values of the weight and the bias parameters of the neural network. The random values are generated from the standard normal distribution. Each of the models are trained with many initializations and then we select the best one by considering the results.

4.3 Validation evaluation and model selection

In this study, we separate our data set into 3 parts, the first part contains 50%. The second part contains 25% and the third part contains the rest. The first part is used in model training while the second part is used for validation evaluation. The last part or out-of-sample data set, will be used for accuracy evaluation in the next section. Table 2 below summarizes the measured errors from the 1-week ahead forecasts of the realized volatility and the calculated realized volatility of the SET index. The calculations in the table follow the equation 3.22 to 3.25. Note that we use the notation (14-n-2) to represent the model that has 14 inputs, 1 hidden layer with n nodes and 2 outputs. Also, note that we scale all the error values by 10^6.

Table 2: Measured error from each of the models in the validation evaluation

Model	MSE_1	Rank	MSE_2	Rank	MAE_1	Rank	MAE_2	Rank
(14-1-2)	4.63	6	123.13	4	400.83	13	5,109.49	8
(14-2-2)	5.26	11	156.33	14	397.00	11	5,251.06	11
(14-3-2)	4.50	4	123.28	5	342.07	1	4,773.15	3
(14-4-2)	5.16	8	150.99	9	394.53	10	4,994.11	5
(14-5-2)	4.25	2	114.60	2	360.75	3	4,749.43	2
(14-6-2)	5.27	12	154.32	12	384.40	5	5,134.21	10
(14-7-2)	4.34	3	120.62	3	383.68	4	4,973.62	4
(14-8-2)	5.52	18	166.14	17	387.12	6	5,102.65	7
(14-9-2)	5.56	19	182.68	20	416.05	18	5,573.49	20
(14-10-2)	5.44	16	163.70	16	391.98	9	5,132.71	9
(14-11-2)	5.43	15	162.69	15	411.06	16	5,390.67	19
(14-12-2)	3.74	1	100.55	1	343.21	2	4,578.61	1
(14-13-2)	5.52	17	169.09	18	388.63	7	5,333.38	15
(14-14-2)	5.33	13	152.64	11	404.38	15	5,282.48	12
(14-15-2)	5.24	10	149.19	8	389.55	8	5,009.59	6
(14-16-2)	5.18	9	147.14	7	403.41	14	5,306.15	14
(14-17-2)	5.07	7	151.52	10	397.62	12	5,284.55	13
(14-18-2)	5.40	14	155.55	13	411.83	17	5,358.29	18
(14-19-2)	4.51	5	128.39	6	417.45	20	5,348.17	17
(14-20-2)	5.58	20	170.89	19	417.07	19	5,333.54	16

At this point, our result above suggests that we may choose the model with 12 nodes (14-12-2) as the first-best model. The second-best model seems to be the model with 3 nodes (14-3-2). The third-best model is the model with 5 nodes (14-5-2). Now, we have 3 models for the next step which is the accuracy evaluation with our benchmark model.

4.4 Accuracy evaluation against the benchmark

In this step, we compare model performance with the benchmark model which is the FTP-MS-GARCH model. We retrain each model and utilize out-of-sample data set in each time step. Table 3 below states the model parameters from the benchmark model and the selected models from the previous section. Note that we only present the GARCH parameters here while the neural network parameters will be discussed in the following section.

Table 3: Parameters of the benchmark model and the top three models

Parameters	Descriptions	Benchmark	(14-12-2)	(14-3-2)	(14-5-2)
μ_0	Conditional mean (regime 0)	0.726%	2.566%	1.793%	1.323%
ω_0	GARCH(1,1) parameters (regime 0)	3.11E-05	8.33E-05	7.66E-23	6.45E-05
α_0		0.204	0.025	0.001	0.027
β_0		0.358	0.309	0.432	0.496
μ_1	Conditional mean (regime 1)	4.651%	10.619%	-0.204%	4.273%
ω_1	GARCH(1,1) parameters (regime 1)	1.97E-05	1.39E-07	7.33E-06	7.11E-17
α_1		0.421	1.264	0.879	1.220
β_1		0.340	0.510	0.669	0.514
e_0	Fixed transition probability (from regime 0 to regime 0)	0.990			
e_1	Fixed transition probability (from regime 1 to regime 1)	0.799			

Note that we put only non-negative constraints on the GARCH parameters in the model estimation process. So, we allow $\alpha_i + \beta_i$ from equation 3.3 to be greater than 1 and then we could have a model with no unconditional variance. Notice that the top three models have no unconditional variance in regime 1 since $\alpha_1 + \beta_1$ from all three models are greater than 1. From Table 3 above, we can define regime 0 as a low persistence volatility state since $\alpha_0 + \beta_0$ from all models are relatively lower than $\alpha_1 + \beta_1$. The contribution of shock to the conditional variance is relatively low in this regime and the persistence of shock from the one-time lag is relatively low too. So, regime 1 is referred to as the high persistence volatility state.

The GARCH parameters of regime 1 from the top three models imply that the high volatility and high persistence are temporary and highly depend on the shock to the market. In the high persistence volatility state (regime 1), volatilities are allowed to diverge since $\alpha_1 + \beta_1$ of the models are high and greater than 1. The convergence of the volatilities comes from the transition back to the low persistence volatility state (regime 0) where a shock has a relatively low impact. Notice that the conditional mean of the benchmark model, the first-best model and the third-best model are aligned for both regimes while the conditional mean of the second-best model is not. This impacts the interpretation of the market regimes. A positive conditional mean for the high persistence volatility state can be related to a market rebound after a market downturn or a crash while a negative value implies a market downturn itself. If this is the case then we might need more than two regimes to capture the market behavior of the SET index. In my opinion, the cause may be a few experiences of long and serious market downturns in the training period. The second possible cause is the overfitting problem.

Table 4 below summarizes the measured errors from the 1-week ahead forecasts of the realized volatility and the calculated realized volatility of the SET index. Note that we scale all the error values by 10^6.

Table 4: Measured error from each of the models in accuracy evaluation

Model	MSE_1	Rank	MSE_2	Rank	MAE_1	Rank	MAE_2	Rank
(14-1-2)	0.125	3	48.01	8	190.38	6	4713.36	11
(14-2-2)	0.143	12	49.53	14	203.58	17	4847.86	19
(14-3-2)	0.132	5	49.48	12	194.59	10	4821.88	16
(14-4-2)	0.123	2	44.25	3	176.36	1	4304.51	1
(14-5-2)	0.140	10	50.12	15	189.74	5	4595.70	7
(14-6-2)	0.116	1	41.54	1	179.32	2	4348.31	2
(14-7-2)	0.166	18	55.88	21	207.99	20	4991.49	21
(14-8-2)	0.148	14	50.72	16	198.89	13	4751.14	13
(14-9-2)	0.137	7	44.08	2	188.12	4	4458.49	3
(14-10-2)	0.156	17	53.81	20	201.16	15	4827.58	17
(14-11-2)	0.194	21	51.68	17	216.47	21	4762.84	15
(14-12-2)	0.150	15	53.55	19	199.59	14	4835.25	18
(14-13-2)	0.137	8	49.27	11	197.68	11	4746.96	12
(14-14-2)	0.177	20	47.89	7	205.43	19	4550.90	6
(14-15-2)	0.141	11	52.36	18	194.34	9	4849.61	20
(14-16-2)	0.172	19	47.15	6	202.29	16	4656.24	9
(14-17-2)	0.147	13	48.78	10	203.58	18	4753.49	14
(14-18-2)	0.136	6	44.98	4	186.61	3	4468.70	4
(14-19-2)	0.151	16	49.50	13	198.67	12	4687.78	10
(14-20-2)	0.128	4	46.02	5	191.80	7	4625.83	8
Benchmark	0.137	9	48.06	9	191.80	7	4493.69	5

It turns out that our first-best model fails to outperform the benchmark model and delivers relatively poor performance. This is a sign of the overfitting problem. However, the second-best model can outperform the benchmark model for MSE_1 measure and the third-best model can outperform the benchmark model for MAE_1 measure. At this point, we cannot state confidently that the notion of a neural network does improve the forecast ability of the TVTP-MS-GARCH model. We can only observe some minor improvements. However, we can observe that the model of 1 hidden layer with 3-6 nodes can be the optimal choice.

The following figures will show the comparison between the calculated realized volatility of the SET index on a weekly basis and one week ahead of forecasted volatility from the benchmark model and the top three models from the validation test. The out-of-sample data start from April 2020 to December 2022.

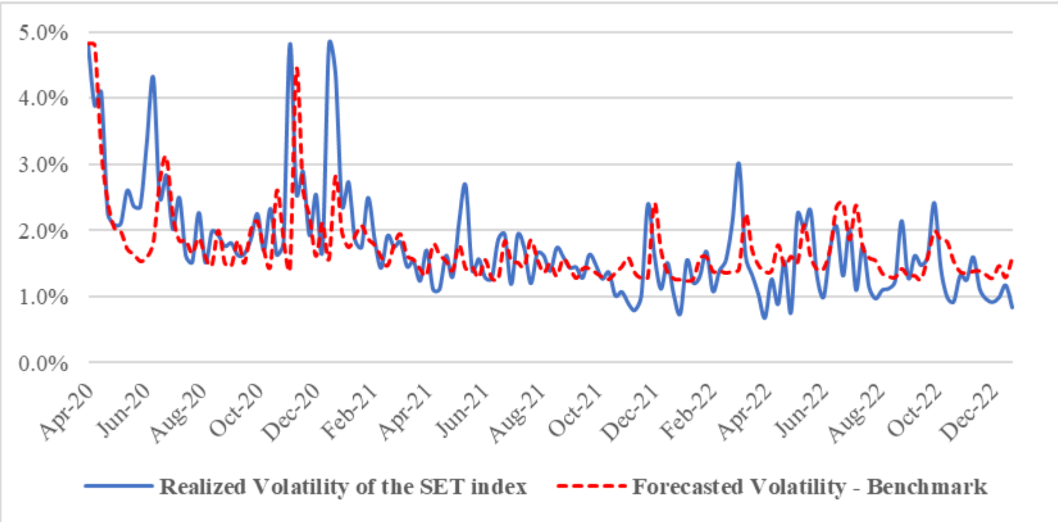

Figure 1: Compare the realized volatility of the SET index with the forecasted volatility from the benchmark model

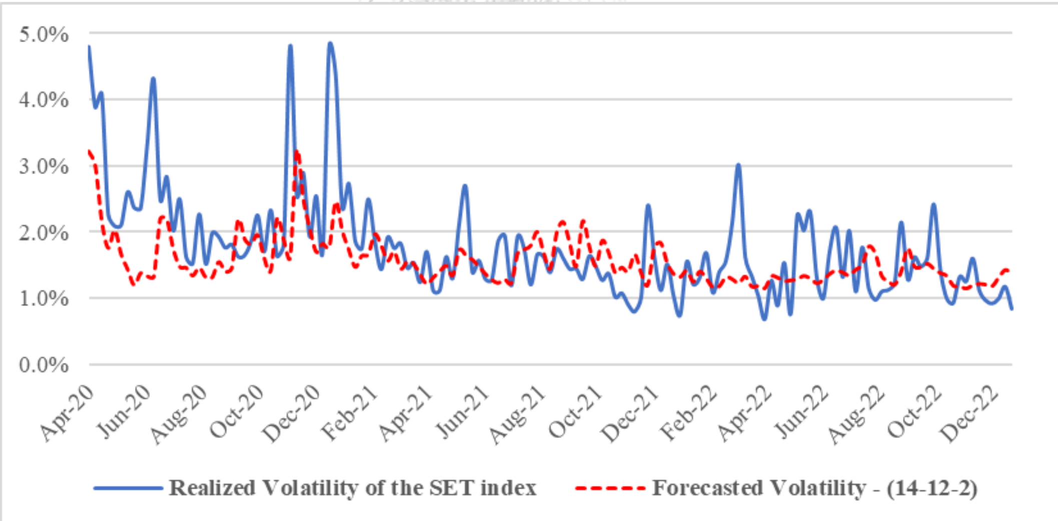

Figure 2: Compare the realized volatility of the SET index with the forecasted volatility from the first-best model (14-12-2)

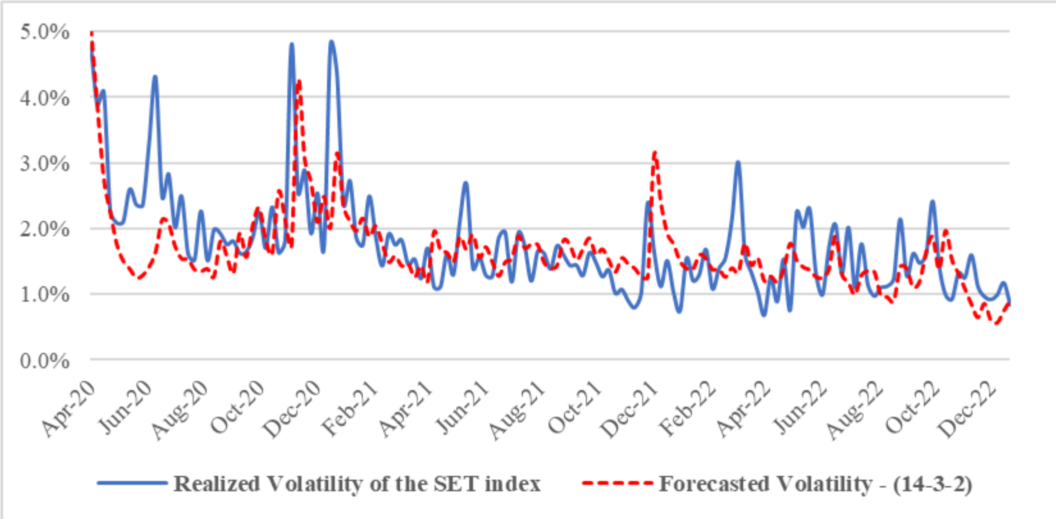

Figure 3: Compare the realized volatility of the SET index with the forecasted volatility from the second-best model (14-3-2)

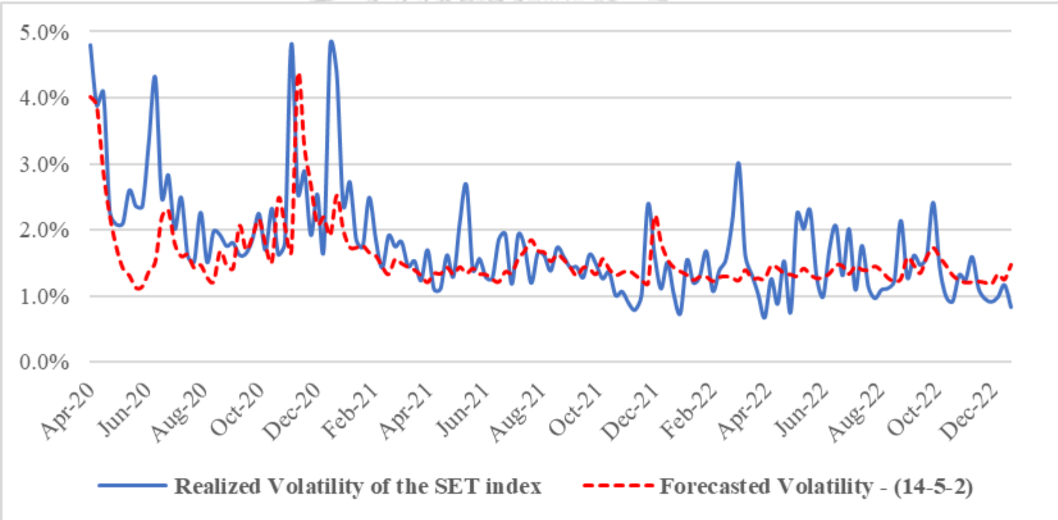

Figure 4: Compare the realized volatility of the SET index with the forecasted volatility from the third-best model (14-5-2)

From all of the figures above, it seems that the second-best model (14-3-2) can deliver more sensible forecasts after February 2022 to December 2022. While the first-best model and the third-best model deliver relatively flatter and the forecasted values are centered near 1.5% from February 2022 to December 2022. In my opinion, it is hard to state that the second-best model is better than the benchmark model if we judge them from the figures.

4.5 Usefulness of the input factors

The success of the model comes partly from the preselected input factors. So, we investigate further for the necessary of them. We consider the trained weights, at the input layer, from model (14-3-2) and (14-5-2). If the trained weight of a factor gets close to zero then the factor contributes less impact on our model performance. We take absolute on each of the weights and average over each node on the same factor. Table 5 below summarizes the averaged weight per one factor and we rank them from the highest impact to the lowest.

Table 5: Averaged weight for each factor at the input layer

Data name	Calculation	Averaged weight (14-3-2)	Rank	Averaged weight (14-5-2)	Rank
SET trading volume	Standardization	0.294	9	0.566	1
Thai government bond spread 10Y-2Y	Standardization	0.465	3	0.371	5
Thai government bond spread 5Y-2Y	Standardization	0.345	8	0.293	9
HSI index	Realized volatility, Standardization	0.347	7	0.392	2
STI index	Realized volatility, Standardization	0.183	13	0.331	8
HSI index	Log-return, Standardization	0.356	6	0.377	4
STI index	Log-return, Standardization	0.169	14	0.237	12
Thai 5Y CDS rate	First difference, Standardization	0.504	2	0.353	7
Effective federal funds rate (EFFR)	First difference, Standardization	0.414	5	0.225	13
Exchange rate USD/THB	Log-return, Standardization	0.231	12	0.267	11
Oil price (WTI)	Log-return, Standardization	0.256	10	0.357	6
Oil price (BRENT)	Log-return, Standardization	0.448	4	0.275	10
Gold price	Log-return, Standardization	0.556	1	0.380	3
Spread between EFFR and BOT rate	Standardization	0.246	11	0.180	14

The above table indicates the input factors that have a high and low impact on the model performance. The log-return of the STI index and spread between EFFR and BOT rate are seemed to have a low impact on model performance. However, most of the values do not get close to zero clearly. So, we cannot confidently drop them from the set of input factors.

Chapter 5
Conclusion

In this study, we aim to refine the TVTP-MS-GARCH model in forecasting realized volatility. By applying the notion of the neural network, we can extend the transition probability part to depend on external factors in a more complex way. We expect that the neural network can approximate the true relationship of the factors with the change of the market regimes. If it is so, then we expect to achieve a model that provides better prediction out-of-sample.

We do study the realized volatility of the SET index which is calculated from 30-minute close data from January 2012 to December 2022. We select the input factors, based on related literatures, from the leading stock market (Hong Kong and Singapore market) and financial-macro data e.g., exchange rate, oil price, gold price, spread of interest rate and CDS price. We separate the whole data set into 3 parts. The first part is used for the model estimation. The second part is used in the model validation and selection. We select the models that deliver relatively low error on the 1-week ahead forecasted volatility and the calculated realized volatility. The last part of the data set is used in accuracy evaluation or out-of-sample tests. We keep training the selected models and compare the forecasting performance of the models with the benchmark model, the FTP-MS-GARCH model.

Our result shows that the best model from the validation test underperforms the benchmark model on the out-of-sample test. Our best model signals the overfitting problem. However, the second-best model and the third-best model unclearly outperform the benchmark model out-of-sample. So, we cannot state confidently that our model can deliver better forecasting performance than the FTP-MS-GARCH model.

For the next study, we can improve or extend the model specification in many ways. For example, we can extend the GARCH equation by adding a one-time lag of the realized volatility. Another example is that we can re-select and make a new combination of input factors or just focus on some interesting data. In my opinion, adding more hidden layers may induce the model to the overfitting problem but this way might be worth a try.

REFERENCES

Aloy, M., De Truchis, G., Dufrénot, G., & Keddad, B. (2014). Shift-Volatility Transmission in East Asian Equity Markets: New Indicators. In G. Dufrénot, F. Jawadi, & W. Louhichi (Eds.), *Market Microstructure and Nonlinear Dynamics: Keeping Financial Crisis in Context* (pp. 273-291). Springer.

Bollerslev, T. (2009). Glossary to ARCH (GARCH). *CREATES Research Papers*, *49*, 1-41.

Brunetti, C., Scotti, C., Mariano, R. S., & Tan, A. H. (2008). Markov switching GARCH models of currency turmoil in Southeast Asia. *Emerging Markets Review*, *9*(2), 104-128.

Cai, J. (1994). A Markov model of unconditional variance in ARCH. *Journal of Business and Economic Statistics*, *12*(3), 309-316.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative finance*, *1*(2), 223-236.

Danielsson, J. (2011). *Financial risk forecasting: the theory and practice of forecasting market risk with implementation in R and Matlab*. John Wiley & Sons, Inc.

Diebold, F. X., Lee, J.-H., & Weinbach, G. C. (1994). Regime switching with time-varying transition probabilities. In C. P. Hargreaves (Ed.), *Nonstationary Time Series Analysis and Cointegration* (pp. 283-302). Oxford University Press.

Dueker, M. J. (1997). Markov switching in GARCH processes and mean-reverting stock-market volatility. *Journal of Business & Economic Statistics*, *15*(1), 26-34.

Engle, R. (2001). GARCH 101: The use of ARCH/GARCH models in applied econometrics. *Journal of economic perspectives*, *15*(4), 157-168.

Engle, R. F., & Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric reviews*, *5*(1), 1-50.

Estrella, A. (2005). The yield curve as a leading indicator: frequently asked questions. 1-23. https://www.newyorkfed.org/medialibrary/media/research/capital_markets/ycfaq.pdf

Filardo, A. J. (1994). Business-cycle phases and their transitional dynamics. *Journal of Business & Economic Statistics*, *12*(3), 299-308.

Gray, S. F. (1996). Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, *42*(1), 27-62.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the econometric society*, 357-384.

Hamilton, J. D. (1990). Analysis of time series subject to changes in regime. *Journal of econometrics*, *45*(1-2), 39-70.

Hamilton, J. D. (1994). *Time series analysis*. Princeton University Press.

Hamilton, J. D., & Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of econometrics*, *64*(1-2), 307-333.

Heaton, J. (2008). *Introduction to Neural Networks with Java* (2nd ed.). Heaton Research, Inc.

Hong, Y., Wang, L., Liang, C., & Umar, M. (2022). Impact of financial instability on international crude oil volatility: New sight from a regime-switching framework. *Resources Policy*, *77*(102667), 1-11.

Klaassen, F. (2002). Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Economics*, *27*, 363–394. https://doi.org/https://doi.org/10.1007/s001810100100

Lamoureux, C. G., & Lastrapes, W. D. (1990). Persistence in variance, structural change, and the GARCH model. *Journal of Business & Economic Statistics*, *8*(2), 225-234.

Marcucci, J. (2005). Forecasting stock market volatility with regime-switching GARCH models. *Studies in Nonlinear Dynamics & Econometrics*, *9*(4), 1-42.

Mikosch, T., & Stărică, C. (2004). Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *Review of Economics and Statistics*, *86*(1), 378-390.

Tan, C.-Y., Koh, Y.-B., Ng, K.-H., & Ng, K.-H. (2021). Dynamic volatility modelling of Bitcoin using time-varying transition probability Markov-switching GARCH model. *The North American Journal of Economics and Finance*, *56*(101377), 1-17.

Tsay, R. S. (2010). *Analysis of financial time series* (3rd ed.). John Wiley & Sons, Inc.

Wang, L., Ma, F., Hao, J., & Gao, X. (2021). Forecasting crude oil volatility with geopolitical risk: Do time-varying switching probabilities play a role? *International Review of Financial Analysis*, *76*(101756), 1-12.

Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2021). *Dive into deep learning* (Vol. arXiv:2106.11342). arXiv preprint https://arxiv.org/abs/2106.11342

VITA

NAME	Wasit Norakarntiansin
DATE OF BIRTH	23 June 1988
PLACE OF BIRTH	Bangkok, Thailand
HOME ADDRESS	Bangkok, Thailand