FORECASTING ASSET RETURN VOLATILITY FROM PAST PRICE DATA

by

Nobuaki Kato

A dissertation submitted in partial fulfillment
Of the requirements for the degree of

Doctor of Philosophy
[EDHEC Business School]

April 2022

Dissertation Committee

Emmanuel Jurczenko, PhD
Francis X. Diebold, PhD
Laurent Calvet, PhD
Enrique Schroth, PhD

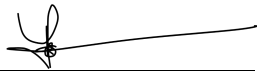
EDHEC BUSINESS SCHOOL

APPROVAL FORM

CANDIDATE'S NAME: Nobuaki KATO

TITLE OF DISSERTATION: "Forecasting Asset Return Volatility from Past Price Data"

APPROVAL:

Emmanuel Jurczenko		*Chair*
Francis X. Diebold		*External Examiner*
Laurent Calvet		*Member*
Enrique Schroth		*Member*

☐ with minor revisions

☐ with major revisions

DATE: 5 April 2022

Abstract

This thesis develops a specification of return dynamics that incorporates price thresholds for volatility forecasting. Because the asset price volatility exhibits a sharp change in the volatility and long memory, it is challenging to incorporate these characteristics into the model at the same time. For example, GARCH has a tradeoff between the reproduction of these two characteristics. On the other hand, the Markov Switching (MS) model can achieve these features simultaneously. However, traditionally, the MS model assumes a constant transition probability of regime-switching. Moreover, the MS model tends to lack enough states to match the variety of volatility distribution since it is difficult to extend beyond a few economic states.

The new forecasting approach is to specify a time-varying transition matrix of the MS model in which the relative position of an asset price to its threshold determines future volatility. This price threshold is defined by multiplying an empirically determined ratio by the observed moving average. Then, the approach relates the probabilities of the current price crossing over the price thresholds to the transition probabilities of the states. The model has a closed-form likelihood, and its parameters can be estimated by the maximum likelihood estimation. The main benefit of this specification is two folds. First, the model can endogenously determine the time-varying transition probability of regime-switching. As far as we can get the time series of the asset price, the model can specify the time-varying transition probability of the MS model. Second, extending it to a multiple-state model is easy, which has been a significant challenge to the traditional MS model, where

the number of parameters must increase exponentially. In the proposed model, only adding one more parameter to the three-state model accepts to change the number of states from three to infinity.

The first chapter establishes the foundation of the model with three states. Then, the second chapter extends the three-states model to the multi-state model. The thesis evaluates the point forecast, interval forecast, and density forecast in each chapter by the proposed model. The first chapter examines the performance of the out-of-sample forecast from the proposed model with the competing models, such as the traditional constant transition probability MS model and GARCH (1,1) model. As a result, the proposed three-state model outperforms competing models for forecasting the return distribution. The second chapter conducts the out-of-sample analysis by increasing the number of states of the proposed model. As a result, the second chapter shows that extending the number of states improves the model's forecasting ability.

The proposed model's practical application would be to use it as an indicator of dynamic asset allocation, conditional on the filtered economic states. Also, the model is helpful for risk management monitoring. For future research, it will be interesting to use the model's characteristic of endogenous time-varying transition probability to study long memory's origin and appearance in volatility.

Contents

Acknowledgments

First, I would like to express my most profound appreciation to my supervisor, Laurent Calvet, for his guidance and various helpful advice throughout my research. Although I had several times when I struggled to proceed in the middle of my research, Professor Calvet gave me many hints to overcome the obstacles and unblock the situation. Without his wealth of knowledge and expertise, I could not have been able to complete my thesis.

I am grateful to Abraham Lioui, Christophe Croux, Emmanuel Jurczenko, Enrique Schroth, Ichiro Tange, Mirco Rubin, Nikolaos Tessaromatis, Raman Uppal, Riccardo Rebonato, and Vladislav Gounas for thoughtful comments and suggestions throughout my research. I'm glad to have so many great people surrounding me during my academic life. I am also thankful to Brigitte Bogaerts-Chevillott, Marine Castel, and Mathilde Legrand for their various logistic support and valuable administrative assistance.

I would like to express my gratitude to Francis X. Diebold for accepting the external examination of my thesis. Professor Diebold provided me with valuable feedback and suggestions for future research.

Last but not least, I am deeply indebted to my wife Michi for her understanding, patience, and encouragement during the last years of my doctorate. Without her extensive support, I could not have been able to allocate my countless weekdays and weekends to pursue the long and winding road. I want to dedicate this thesis to my family.

List of Tables

List of Figures

1 Paper 1, Forecasting Volatility With Price Thresholds: Three States Model

Forecasting Volatility With Price Thresholds

Nobuaki Kato[1]

31st May 2021

Abstract

This paper develops a specification of return dynamics that incorporates price thresholds for volatility forecasting. The new forecasting approach is to specify a time-varying transition matrix of the Markov Switching (MS) model in which the relative position of an asset price to its threshold determines future volatility. This price threshold is defined by multiplying an empirically determined ratio to the observed moving average. Then, the approach relates the probabilities of the current price crossing over the price thresholds to the transition probabilities of the states. The model has a closed-form likelihood, and its parameters can be estimated by the maximum likelihood estimation. The paper evaluates the point forecast, interval forecast, and density forecast by the proposed model and the competing models, such as the constant transition probability MS model and GARCH (1,1) model, from the out-of-sample of CRSP S&P 500 return data. As a result, the proposed model outperforms competing models for forecasting the return distribution.

JEL classification: C13; C32; G17

Keywords: Time-varying transition probabilities, Markov switching, maximum likelihood, price threshold, moving average, out-of-sample forecasts

[1]EDHEC Business School (nobuaki.kato@edhec.com). I am indebted to Laurent Calvet for his guidance and various helpful advice through my research. I am grateful to Abraham Lioui, Christophe Croux, Ichiro Tange, Mirco Rubin, Nikolaos Tessaromatis, Raman Uppal, Riccardo Rebonato, and Vladislav Gounas for their thoughtful comments and suggestions. I am also thankful to Brigitte Bogaerts-Chevillotte and Mathilde Legrand for various logistic supports.

1.1 Introduction

Modeling and forecasting accurate asset return distribution is an important problem for investors and risk managers. Investors/asset managers allocate assets dynamically based on the estimate of the economic conditions, expecting they can maximize wealth and minimize loss better than a naïve static allocation. Risk managers need to quantify the appropriate risks associated with these portfolio decisions and correctly assign risk budgets. Return distributions are driven by the underlying state variables. However, because the underlying states are not observable in general, asset managers and risk managers may use several indicators to infer the latent regime. The Markov Switching (MS) model can derive one such indicator.

Goldfeld and Quandt (1973) and Cosslett and Lee (1985) pioneered the MS model, and Hamilton (1989) applied it to macroeconomic analysis. The MS model is a widely used econometric model that can filter the latent states from observed time series and describe a corresponding volatility process conditional on such states. The model is practical and easy to employ, as a few parameters govern its switching dynamics via a transition probability matrix dependent only on the particular periods of previous states. Since then, the MS model is popularly used in various applications, such as business cycle analysis, monetary policy studies, and volatility forecasts in financial time series (Hamilton and Raj, 2002).

However, in econometrics, most of these applications use constant transition probabilities (CTP). As Diebold et al. (1994) and Filardo (1994) argue, the CTP is not realistic because the underlying economic condition changes over time. MacRae (1977) considers that the transition probabilities vary in response to the previous period's aggregated exogenous variables. Diebold et al. (1994) and Filardo (1994) introduce time-varying transition probabilities (TVTP) as a logistic function of exogenous variables and past returns. By specifying the TVTP, we may improve the state's predictability and a better knowledge of the return distribution.

Because the exogenous variables' economic specification to TVTP was initially left for

further research, many researchers searched for suitable candidates for these variables. Filardo (1998) states that an exogenous variable must be conditionally uncorrelated with the Markov process state. Gray (1996) and Fong and See (2002) assume the exogenous variable as a level of the short rate and derive that the TVTP is a weighted average of switching probabilities from low regime to high regime and probabilities from high regime to low regime. Van Norden and Schaller (1997) specify the exogenous variable as a price-dividend ratio. Peria (2002) uses five macroeconomic variables[2], and Yuan (2011) also examines four macroeconomic comparators for the exogenous variables in exchange rate application[3].

One problem with these types of the specification is that the model becomes vulnerable to misspecification risk. If we use macroeconomic variables, it is not handy to use such a model for asset allocation quickly as the observations are generally infrequent. Moreover, it is not easy to use the same economic model for different asset classes or different types of analysis. For example, there is little economic reason to use purchasing power parity for detecting a regime in the stock market. Thus, depending on the subject of the analysis, researchers must change the exogenous variables every time. Rather than finding economically meaningful exogenous variables, I posit that the empirical prices already contain ample information, allowing econometricians to extract the transition probabilities directly from the data.

This paper defines the TVTP of the MS model endogenously through the market price. As Hayek (1945) states and the market microstructure literature shows through transactions by the informed traders, the market price contains rich information, and it reasonably reflects the current economic condition[4]. Further, as Morck et al. (1990) and Bond et al.

[2]Peria (2002) models TVTP by the growth of domestic credit, the ratio of imports to exports, the real exchange rate, the unemployment rate, and the fiscal deficit.

[3]Yuan (2011) examines the purchasing power parity, the linear combination of money supply and economic output, the output from real interest rate differential model, the output from portfolio balance model

[4]See Wilson (1974), Milgrom (1979), Glosten and Milgrom (1985), Kyle (1985), Palfrey (1985), and Easley and O'Hara (1987) (1991) (1992) for the examples.

(2012) show, the asset price movement has a feedback effect on the real economy. Hence, the price is not only a helpful guide to know the current underlying state but also helps to infer the future state. From this point of view, I assume that the sudden deviation from the current price level in unusual magnitude is large enough to impact the real economy, and therefore, implies the change in the underlying state.

Volatility exhibits two well-documented features. One is volatility asymmetry. There is a negative correlation between stock returns and return volatility, where when the stock price increases, the volatility tends to decrease, and when the price decreases, the volatility tends to increase[5]. Therefore, we can specify different volatility regimes according to the price direction (up or down). Another is a long memory. Once a shock in the volatility is induced, it persists over multiple periods. If we consider these two features, it is reasonable to consider that the significant price movement triggers a change in volatility magnitude and persists for some time. This phenomenon is effectively regime-switching. For example, if the asset price crashes dramatically in one day (e.g., near 10% by dot-com bubble crash and Lehman Brothers bankruptcy, or 20% on Black Monday), the volatility increases, implying that the market has entered in a high volatility regime. Contrary, if the asset price jumps in one day (e.g., 5% by "Whatever it takes" speech given by then ECB governor Mario Draghi), it may imply that the market has entered a low volatility regime.

How much price change enough to switch the regime (e.g., 5% or 10% in the above examples) can be measured by a distance between a latent "trigger" price level and the current market price. If the market price crosses this latent price level, it triggers a change in the underlying economic condition. This intuition is similar to the technical analysis popularly used by the practitioners, which see the trend change when the price breaks support or resistance level. To estimate this latent price threshold empirically, we can apply

[5]There are two possible reasons which cause this volatility asymmetry. One reason could be through a firm's leverage, which increases when the stock price decreases (leverage effect). Another reason is through volatility feedback where the future rise in volatility must be compensated by the rise in return, therefore the current price must drop (risk premium). See Black (1976), Christie (1982), French et al. (1987) and a good review by Bekaert and Wu (2000).

the likelihood estimation or the generalized method of moment.

When we posit the existence of such a price threshold, we can relate the distance of the current asset price from this threshold into a probability of the current asset price crossing over this threshold level. This paper defines this probability as the transition probability of the regime-switching model. The above specification of the probability is analogous to computing an exercise probability of out-of-the-money European option security (Black and Scholes, 1973) or a firm's default probability (Merton, 1973). In the TVTP setting, the strike price is equal to the threshold (or "border") of the two different economic regimes, and the maturity is one day. This specification brings the benefit of only requiring the price dynamics. It is free from specifying which exogenous variable matters to the specification of the economic model of transition probabilities, a central discussion of the literature for a long time. Also, because the model is independent of the exogenous variables specific to any particular asset class, we can apply the same principle to several asset classes (e.g., stocks, fixed incomes, commodities). Moreover, it is not necessary to consider the infrequency of the data, which is inevitable for macroeconomic data.

There are a few observation-driven models proposed in the literature. Engel and Hakkio examine EMS exchange rates (1996), which also uses a distance of the price level from its band as an input to the logistic function. Engel and Hakkio's approach found that the volatility increases when the exchange rates rapidly go close to the edge of ERM bands[6]. However, their analysis is specific to the ERM band, which does not exist anymore. Another example is Bazzi et al. (2016). They take an approach that specifies the transition probabilities as a specific transformation of lagged observations. The authors generate the innovation of the time-varying probability by the score of the predictive likelihood function[7] This approach determines the parameter which drives the transition probabilities by

[6]Before the introduction of euros, European Monertary System (EMS) controlled the exchange rates of currencies in the Exchange Rate Mechanism (ERM)

[7]Bazzi et al. (2016) uses econometric approach called Generalized Autoregressive Score (GAS), developed by Creal et al. (2013).

the steepest ascent (Newton) method on the conditional density of the previous point of time $t-1$ for the model's local fit at time t. While this method seems powerful, it is based on a purely econometric technique, and it is challenging to grasp fundamental economics intuitively. Because asset managers and risk managers often encounter an occasion that requires an explanation of the economics behind it, it is not easy to translate the economics of the model's result. The last example is Wang et al. (2019). They specify the transition probability as a linear combination of change in realized volatility[8]. However, the authors do not provide much economic rationale for this specification.

This paper's main contribution is to derive TVTP from observing the time series without requiring any assumption on the exogenous variable, but with some intuitive economics. The paper relates some concepts on market microstructure and technical analysis to the econometric approach. The specification is straightforward, parsimonious, and has a closed-form, so the Maximum Likelihood Estimation can easily estimate the parameters.

The remainder of the paper is organized as follows: Section 1.2 develops the TVTP model. Section 1.3 presents an empirical analysis of TVTP and provides an in-sample comparison of simulated return distributions. Section 1.4 conducts out-of-sample analysis; point forecast, interval forecast, and density forecast. Section 1.5 concludes.

1.2 TVTP Model

1.2.1 Motivation

There are two fundamental premises in this paper. First, the security price contains rich information, and thus, we can filter the underlying economic state by observing the current price. Second, the security price movement has a feedback effect on the economic state. Therefore, we can infer the future states by observing the price evolution.

For the first premise, if the market is efficient (Fama, 1970), the price reflects all the

[8] I will explain the realized volatility measure in out-of-sample section of this paper.

relevant information. Even if the price does not always contain the full relevant information contemporaneously, it will reveal the underlying information over time through insider transactions[9]. On the other hand, each market participant may trade for different reasons. Even in this case, the price aggregates the various pieces of information and collectively converges to the underlying value[10]

The second premise is more subtle than the first one. Traditionally, the literature assumes that the price evolution in financial markets has little or no effect on the real economy. However, Morck et al. (1990) and Bond et al. (2012) argue a feedback effect where the prices affect the decision makers' actions of the firm. This feedback arises because the decision-makers, even though they are supposed to be most informed, also use the information inherent in the price to guide corporate decisions (Chen et al., 2007, Bakke and Whited, 2014). The decision-makers do not necessarily have perfect information, and the market as a whole can be more informed than the insiders (Grossman, 1976). Additionally, the firm managers' compensations are generally tied to the firm's share price. Moreover, credit rating agencies and regulators of banking who influence the firm's cash flows monitor the market prices very closely. Thus, the managers' actions will be influenced by the expected outcome from the stock price movement.

The price movement influences financial-market traders as well. Dow et al. (2017) show how a slight decline in fundamentals may reduce price efficiency and amplify the original adverse economic shocks. In their model, a firm manager decides whether to invest or not. However, for the decision-making, the manager relies on some information reflected in the company's stock price, which implies the investment is good or bad. This information is produced by speculative traders who are informed and motivated by trading profit. However, speculators collect information only when their marginal benefit exceeds the cost of acquiring information (Grossman and Stiglitz, 1980). If the fewer the speculators are will-

[9]Glosten and Milgrom (1985), Kyle (1985), Easley and O'Hara (1987), (1991), (1992), for example.

[10]See Wilson (1974), Milgrom (1979), and Palfrey (1985).

ing to acquire the information and play the game, the less information is reflected in the price, and the less sensitive the price becomes to the firm's manager's decision. However, on the other hand, if a risk-averse manager thinks the stock price to be irrelevant to her decision, she does not make any investment. Then, if the speculators see that the firm's manager does not invest in increasing the firm's value, they may not expect a benefit that exceeds the cost of information acquisition. Thus, if speculators foresee the manager's decision not to invest, they find that the information acquisition is less profitable than the firm's manager would have invested. This reluctance to trade erodes the price efficiency further. Therefore, even if a decline in fundamentals is small initially, it reduces the probability of a firm's investment and discourages speculators from producing information. As a result, the price becomes less efficient, and the firm's investment declines further.

From the above discussion, it seems that the observed market price can influence the real economy. Analyzing the price dynamics serves as a clue to infer the underlying economic state and influences the future latent state. When an economic regime switches from one to another, there must be a trigger of economic state change, reflected in the price. Once the trigger is exercised, the price movement direction could indicate which economic regime we are in. What is essential to know is how far the current asset price level hovers from this trigger. If the price is close to the trigger, it is more likely to cross this level. If the price is far from the trigger, it is less probable to cross the trigger. For this reason, it is reasonable to assume that the price movement directly affects the Markov switching model's probability transition matrix.

1.2.2 Specification

This paper examines the basic case where there are one asset and three states of the economy in which the asset has low volatility (stable state: S_s), normal volatility (middle state: S_m), and high volatility (volatile state: S_v). The volatility of each of the three states is characterized such that

$$\sigma(S_t) = \{\sigma_s, \sigma_m, \sigma_v\} \in \mathbb{R}^3_+ \tag{1.1}$$

where $\sigma_s \equiv \sigma(S_s)$, $\sigma_m \equiv \sigma(S_m)$, $\sigma_v \equiv \sigma(S_v)$. Then, assume that the one-period logarithmic total return of the asset $r_t \equiv \ln(P_t/P_{t-1})$ conditional on S_t follows a process, such that

$$r_{t+1}(S_t) = \mu - \frac{\sigma^2(S_t)}{2} + \sigma(S_t)\varepsilon_{t+1} \tag{1.2}$$

where μ is a positive constant, ε_{t+1} is a random variable that follows IID standard Gaussians $\mathcal{N}(0,1)$. Here, μ is a long-run mean of return because $\mathbb{E}[e^{r_{t+1}}] = e^{\mu}$.

At time t with the state $s_t = S_j$, $j \in \{s,m,v\}$, (given the state $s_{t-1} = S_i$ at time $t-1$, $i \in \{s,m,v\}$, and the state $s_{t-2} = S_k$ at time $t-2$, $k \in \{s,m,v\}$, and so on,) the transition probability of first-order Markov chain is specified as

$$\begin{aligned} \gamma_{ij,t} &\equiv \mathbb{P}(s_t = S_j | s_{t-1} = S_i, s_{t-2} = S_k, ...; \mathscr{I}_{1:t-1}) \\ &= \mathbb{P}(s_t = S_j | s_{t-1} = S_i; \mathscr{I}_{1:t-1}) \end{aligned} \tag{1.3}$$

where $\mathscr{I}_{1:t-1}$ is cumulative information observed by econometrician from the past price sequence from time $t = 1$ until $t = t-1$. Then, the state dynamics of the MS model are defined by the time-varying transition matrix such that

$$A_t \equiv \begin{bmatrix} \gamma_{ss,t} & \gamma_{sm,t} & \gamma_{sv,t} \\ \gamma_{ms,t} & \gamma_{mm,t} & \gamma_{mv,t} \\ \gamma_{vs,t} & \gamma_{vm,t} & \gamma_{vv,t} \end{bmatrix} \tag{1.4}$$

This time-varying transition matrix is this paper's contribution to the literature. Once the time-varying transition matrix is specified, it can be easily applied to the conventional Markov process described in Section 1.3.

1.2.3 Price Dynamics

To specify the time-varying transition matrix as in equation (1.4), suppose that there is a feedback effect on states from the price dynamics, as discussed previously. I characterize this feedback effect by defining two thresholds for the asset price level (upper threshold and lower threshold), which the representative investor sets. When the asset price crosses these thresholds instantaneously, it triggers the transition of one state to another. It is intuitive to measure this threshold as a distance from the fundamental asset price level. Because the fundamental asset price is not observable, I use the long-term mean as an estimator. For a long time, practitioners have been using moving averages as a key indicator to see the changes in the trend (the technical analysis). The literature also reports the effectiveness of the technical analysis, which is based on a moving average[11]. The logic of technical analysis can be extended to the regime-switching analysis. I consider the price dynamics problem with the following settings: if the asset price hovers around its mean, the economy is in a middle state S_m. On the other hand, if the price is far above its mean, the economy enters a stable state S_s. Similarly, if the price is far below its mean, the economy enters a volatile state S_v. The upper threshold is the mean of asset price plus some positive distance, and the lower threshold is the mean minus some positive distance.

Switching Thresholds of Economic States

First, I begin by defining the switching thresholds for the economy in the middle state. Given the asset price series $\{P_t\}_{t=1}^{T}$, an observed exponential weighted moving average

[11] See Brock et al. (1992), Lo et al. (2000), Menkhoff and Taylor (2007), Neely et al. (2013) for the discussion of the profitability and forecasting power of the moving average indicator, and Zhou and Zhu (2013) for the equilibrium model incorporating moving average.

(EWMA) is defined with a specific parameter $\delta \in (0,1)$ such that

$$\begin{aligned} EWMA_t &\equiv \delta P_t + (1-\delta) EWMA_{t-1} \\ EWMA_1 &\equiv P_1 \end{aligned} \tag{1.5}$$

then, the two threshold price levels conditional on the middle state are specified with the previous period's EWMA as[12]

$$\begin{aligned} K^m_{u,t} &\equiv (1+\psi_u) EWMA_{t-1} \\ K^m_{l,t} &\equiv (1-\psi_l) EWMA_{t-1} \end{aligned} \tag{1.6}$$

where $K^m_{u,t}$ is the upper threshold price level during the middle state, $K^m_{l,t}$ is the lower threshold price level during the middle state, $\psi_u \in (0, 1/\delta - 1)$ and $\psi_l \in (0, 1/\delta - 1)$ are constant. The three economic states are defined as

$$S_t \equiv \begin{cases} S_s & \text{if } K^m_{u,t} < P_t \\ S_m & \text{if } K^m_{l,t} < P_t < K^m_{u,t} \\ S_v & \text{if } P_t < K^m_{l,t} \end{cases} \tag{1.7}$$

Therefore, from equation (1.3), the transition probabilities conditional on the middle state can be restated as

[12] See Appendix A for the proof of that the price threshold at time t can be characterized by the EWMA at time $t-1$, not at time t

$$
\begin{aligned}
\gamma_{ms,t} &= \mathbb{P}[S_s|S_m] = \mathbb{P}[P_t > K^m_{u,t}|S_{t-1} = S_m; EWMA_{t-1}] \\
\gamma_{mv,t} &= \mathbb{P}[S_v|S_m] = \mathbb{P}[P_t < K^m_{l,t}|S_{t-1} = S_m; EWMA_{t-1}] \\
\gamma_{mm,t} &= \mathbb{P}[S_m|S_m] = 1 - \gamma_{ms,t} - \gamma_{mv,t}
\end{aligned}
\tag{1.8}
$$

Similar to equation (1.6), suppose there are two thresholds of the asset price level, which trigger an economic transition from the stable state to the middle state and the volatile state. These thresholds conditional on the stable state are defined as

$$
\begin{aligned}
K^s_{m,t} &\equiv (1 - \psi_l \cdot \lambda^s) EWMA_{t-1} \\
K^s_{l,t} &\equiv \frac{(1 - \psi_l \cdot \lambda^s)(1 - \psi_l)}{1 + \psi_u} EWMA_{t-1}
\end{aligned}
\tag{1.9}
$$

where $\lambda^s = \sigma_s/\sigma_m$ is a ratio of volatility in the stable state and the middle state.

Parameter λ serves as an adjustment factor, which controls the bandwidth of price thresholds from the EWMA depending on the volatility level. Imagine if the threshold distances from EWMA are equal across different states. It would be less probable to cross the threshold from the stable states because the price fluctuates less, and it would be more probable to cross from the volatile states because the price fluctuates more. Therefore, if the bandwidth is unconditionally constant regardless of the state, an asset price becomes very stable for a long time (no switch from the stable state), and the volatility memory dies out very quickly (immediate exit from the volatile state). However, as an asset price empirically shows long volatility memory (i,e, volatile state lasts long once the economy enters the volatile state), the constant thresholds distance is not coherent with the data. Hence, an adjustment factor must be introduced for the consistency. In the stable state, λ becomes smaller than 1, which shrinks the bandwidth, and in the volatile states, λ becomes larger than 1, which expands the bandwidth. This adjustment increases switching probability from the stable state and

reduces switching probability from the volatile state.

Another highlight here is that the middle price threshold in the stable state, $K^s_{m,t}$, is now lower than the EWMA. I assume the transition from the stable state to the middle state follows the same mechanism of switching from the middle state to the volatile state. From the stable state point of view, the middle state is more volatile. In the proposed model's framework, the price must go down to switch from a less volatile state to a more volatile state. This is the same relationship between the middle state and the volatile state in equation (1.6). If I extend this view to the switch from a stable state to a volatile state, the price must go below the lower threshold of the stable state once, to enter the middle state, and even go further down the lower threshold of the middle state to enter the volatile state. This is why $K^s_{l,t}$ is defined as equation (1.9). Imagine there is another price level, $EMWA^{'}_t$, underneath the $K^s_{m,t}$, which satisfies $EWMA^{'}_t(1+\psi_u) = EWMA_t(1-\psi_l\lambda^s)$. Then this $EMWA^{'}_t$ can be considered as an imaginary EWMA if the economy were in the middle state from the stable state's perspective. This imaginary EWMA of middle state can be written by $EMWA^{'}_t = EWMA_t \times (1-\psi_l\lambda^s)/(1+\psi_u)$. Then, by using this imaginary EWMA, the lower threshold to switch to the volatile state can be expressed as $EMWA^{'}_t(1-\psi_l)$. Hence, $K^s_{l,t}$ is defined as equation (1.9).

The same discussion to equation (1.9) is also applied to the price thresholds in the volatile state. Suppose there are two thresholds for the asset price level, which trigger a transition from the volatile state to the middle state and the stable state. Conditional on the volatile state, these thresholds are defined as

$$\begin{aligned} K^v_{m,t} &\equiv (1+\psi_u\cdot\lambda^v)EWMA_{t-1} \\ K^v_{u,t} &\equiv \frac{(1+\psi_u\cdot\lambda^v)(1+\psi_u)}{1-\psi_l}EWMA_{t-1} \end{aligned} \tag{1.10}$$

where $\lambda^v = \sigma_v/\sigma_m$ is a volatility ratio in the volatile and middle states, which is the adjustment factor for the price threshold distance. In the volatile state, the price threshold to

switch back to the middle state lies above the EWMA and the threshold to switch to the stable state lies further above.

Finally, using equation (1.3), (1.9), and (1.10), the transition probabilities conditional on the stable state and the volatile state can be restated as,

$$\begin{aligned}
\gamma_{sm,t} &= \mathbb{P}[S_m|S_s] = \mathbb{P}[K^s_{l,t} < P_t < K^s_{m,t}|S_{t-1} = S_s; EWMA_{t-1}] \\
\gamma_{sv,t} &= \mathbb{P}[S_v|S_s] = \mathbb{P}[P_t < K^s_{l,t}|S_{t-1} = S_s; EWMA_{t-1}] \\
\gamma_{ss,t} &= \mathbb{P}[S_s|S_s] = 1 - \gamma_{sm,t} - \gamma_{sv,t}
\end{aligned} \tag{1.11}$$

$$\begin{aligned}
\gamma_{vm,t} &= \mathbb{P}[S_m|S_v] = \mathbb{P}[K^v_{m,t} < P_t < K^v_{u,t}|S_{t-1} = S_v; EWMA_{t-1}] \\
\gamma_{vs,t} &= \mathbb{P}[S_s|S_v] = \mathbb{P}[P_t > K^v_{u,t}|S_{t-1} = S_v; EWMA_{t-1}] \\
\gamma_{vv,t} &= \mathbb{P}[S_v|S_v] = 1 - \gamma_{vm,t} - \gamma_{vs,t}
\end{aligned} \tag{1.12}$$

As described in the following subsection, the above specification is analogous to an exercise probability of out-of-the-money European option security (Black and Scholes, 1973) conditional on the latent states.

Transition Probabilities

Now, I quantify the next period's state probability at time $t-1$ in equations (1.8), (1.11), and (1.12). First, I describe how to quantify the probability of price going beyond the upper thresholds. From equation (1.2), at time $t-1$, the next period's price P_t being higher than the upper threshold level $K_{u,t}$, conditional on the state $S_{t-1} = S_i$, where $i \in \{m, v\}$, is written as

$$
\begin{aligned}
P_t > K^i_{u,t} &\Leftrightarrow \ln P_t > \ln K^i_{u,t} \\
&\Leftrightarrow \ln P_{t-1} + \mu - \frac{\sigma^2(S_i)}{2} + \sigma(S_i)\varepsilon_{t+1} > \ln K^i_{u,t} \\
&\Leftrightarrow -\varepsilon_{t+1} < \frac{\ln\left(\frac{P_{t-1}}{K^i_{u,t}}\right) + \mu - \frac{\sigma^2(S_i)}{2}}{\sigma(S_i)}
\end{aligned}
\tag{1.13}
$$

Therefore, alike Black and Scholes (1973), if I define $d^i_{u,t}$ such that

$$
d^i_{u,t} \equiv \frac{\ln\left(\frac{P_{t-1}}{K^i_{u,t}}\right) + \mu - \frac{\sigma^2(S_i)}{2}}{\sigma(S_i)} \tag{1.14}
$$

then $d^i_{u,t}$ follows IID standard Gaussian $\mathcal{N}(0,1)$. Thus, from equation (1.8) and (1.12),

$$
\gamma_{is,t} = \mathbb{P}[P_t > K^i_{u,t} | S_{t-1} = S_i] = \Phi(d^i_{u,t}) \tag{1.15}
$$

where Φ is Gaussian cumulative distribution function. Note that, the transition probabilities of crossing the upper threshold of time t, conditional on the state i at time $t = t-1$, are fully quantified with the information available at time $t = t-1$.

Similarly, to quantify the probability of price being lower than the lower threshold, by defining $d^j_{l,t}$, where $j \in \{m, s\}$, such that

$$
d^j_{l,t} \equiv \frac{\ln\left(\frac{P_{t-1}}{K^j_{l,t}}\right) + \mu - \frac{\sigma^2(S_j)}{2}}{\sigma(S_j)} \tag{1.16}
$$

leads to the transition probabilities of crossing the lower threshold at time t, conditional on the state j at time $t = t-1$, which can be written from equation (1.8) and (1.11) as

$$\gamma_{jv,t} = \mathbb{P}[P_t < K^j_{l,t} | S_{t-1} = S_j] = 1 - \Phi(d^j_{l,t}) \tag{1.17}$$

Finally, defining the threshold for transitioning to the middle state either from the stable state or from the volatile state, $d^k_{m,t}$, where $k \in \{s, v\}$, such that

$$d^k_{m,t} \equiv \frac{\ln\left(\frac{P_{t-1}}{K^k_{m,t}}\right) + \mu - \frac{\sigma^2(S_k)}{2}}{\sigma(S_k)} \tag{1.18}$$

then, from equation (1.11) and (1.12), the transition probabilities for the above conditions are

$$\begin{aligned} \gamma_{sm,t} &= \mathbb{P}[K^s_{l,t} < P_t < K^s_{m,t} | S_{t-1} = S_s] = \Phi(d^s_{l,t}) - \Phi(d^s_{m,t}) \\ \gamma_{vm,t} &= \mathbb{P}[K^v_{l,t} < P_t < K^v_{m,t} | S_{t-1} = S_v] = \Phi(d^v_{m,t}) - \Phi(d^v_{u,t}) \end{aligned} \tag{1.19}$$

From equation (1.15), (1.17), and (1.19), the transition probabilities in the matrix, equation (1.4) is restated as

$$
\begin{aligned}
\gamma_{ss,t} &= 1 - \gamma_{sm,t} - \gamma_{sv,t} \\
\gamma_{sm,t} &= \Phi(d^s_{l,t}) - \Phi(d^s_{m,t}) \\
\gamma_{sv,t} &= 1 - \Phi(d^s_{l,t})
\end{aligned}
$$

$$
\begin{aligned}
\gamma_{ms,t} &= \Phi(d^m_{u,t}) \\
\gamma_{mm,t} &= 1 - \gamma_{ms,t} - \gamma_{mv,t} \\
\gamma_{mv,t} &= 1 - \Phi(d^m_{l,t})
\end{aligned}
\tag{1.20}
$$

$$
\begin{aligned}
\gamma_{vs,t} &= \Phi(d^v_{u,t}) \\
\gamma_{vm,t} &= \Phi(d^v_{m,t}) - \Phi(d^v_{u,t}) \\
\gamma_{vv,t} &= 1 - \gamma_{vs,t} - \gamma_{vm,t}
\end{aligned}
$$

Each transition probability at time t, $\gamma_{ij,t}$, is fully characterized by the observations at time $t-1$ with the Gaussian cumulative distribution function. When the price gets close to a threshold, the transition probability becomes high. When the price gets far from the threshold, the transition probability becomes low. The price threshold TVTP MS model has a closed-form likelihood, and the maximum likelihood estimation can easily estimate its parameters. For the recapitulation, the TVTP requires only seven parameters as follows.

$$
\theta = (\sigma_s \ \sigma_m \ \sigma_v \ \psi_u \ \psi_l \ \delta \ \mu) \in \mathbb{R}^7_+ \tag{1.21}
$$

where σ_s, σ_m, σ_v are volatilities conditional on each state $i \in \{s, m, v\}$, ψ_u and ψ_l are price threshold parameters, δ is an exponential moving average parameter, μ is the long-term mean of the asset returns.

1.3 Empirical Analysis

1.3.1 Maximum Likelihood Estimation

Now I estimate the model parameters by a historical dataset. The conditional probability density function with three states is

$$\omega_t = \left(f(r_t|s_t = S_s, r_{1:t-1}; \theta)\ f(r_t|s_t = S_m, r_{1:t-1}; \theta)\ f(r_t|s_t = S_v, r_{1:t-1}; \theta)\right)' \tag{1.22}$$

where $f(r_t|s_t = S_j, r_{1:t-1}; \theta)$ is

$$f(r_t|s_t = S_j; r_{1:t-1}; \theta) = \frac{1}{\sqrt{2\pi\sigma^2(S_j)}} \exp\left[-\frac{\{r_t - \mu + \sigma^2(S_j)/2\}^2}{2\sigma^2(S_j)}\right] \tag{1.23}$$

The filtered probabilities of the latent states conditional on the observed returns satisfy[13]

$$\hat{\Pi}_t = \frac{\omega_t * (\hat{\Pi}_{t-1} A_t)}{[\omega_t * (\hat{\Pi}_{t-1} A_t)]\iota} \tag{1.24}$$

where $*$ denotes element-by-element multiplication and $\iota = (1\ 1\ 1)'$. Then the likelihood function is written as[14]

$$\ln L(\theta; r_{1:T}) = \sum_{t=1}^{T} \ln[\omega_t \cdot (\hat{\Pi}_{t-1} A_t)] \tag{1.25}$$

Then, the maximum likelihood estimator is defined by

[13] See Appendix B for the proof.

[14] See Appendix B for the proof.

$$\hat{\theta}_{ML} = \arg\max_{\theta} \{\ln L(\theta; r_{1:T})\} \tag{1.26}$$

Since the objective function of equation (1.26) is non-linear in general, it is difficult to solve analytically. Therefore, the optimization is conducted numerically in practice.

1.3.2 Sample Data

I retrieve return data from Prof. Kenneth French's website[15]. This data contains the daily excess return on the value-weighted US equity index calculated by the Center for Research in Security Prices (CRSP) over July 1st, 1926 to December 31st, 2020, which has $T = 24,896$ observations. Over this period, the excess return series has a historical mean of 0.030302% per day (about 7% per year). The Augmented Dickey-Fuller (ADF) test (1996) brings no evidence of non-stationarity in the return data[16]. Using excess return has merit compared to using raw stock return for the long time horizon analysis because it is not affected by the interest regime change in the US in 1951 (Treasury-Federal Reserve Accord). I construct the price level of the US equity at time t from the excess return by $P_t \equiv \prod_{k=0}^{t}(1+R_k)$, where R_t is observed excess return at time t.

1.3.3 Model Parameter Estimation

For optimizing equation (1.26), I derive the historical mean from the sample data as $\hat{\mu} = 0.030302\%$ and use it as a calibrated value for μ[17]. Since the optimization process is based

[15] https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/

[16] The p-value of ADF test is less than 0.01 for 0 to 20 lags.

[17] The reason for the calibration is because the conditional expectation of the return is μ ($\because \mathbb{E}[e^{r_t}] = e^{\mu}$), and I want the estimated model to be consistent with reasonable values for the long-run mean. Also, I set optimization constraints to parameters such that $\sigma_s \in (0.001, 0.100)$, $\sigma_m \in (0.001, 0.100)$, $\sigma_v \in (0.001, 0.100)$, $\psi_u \in (0.001, 0.100)$, $\psi_l \in (0.001, 0.100)$, and $\delta \in (0, 1)$. These constraints are required to make sure that the Hessian matrix is invertible during the optimization. Therefore, I use the optimization method in Byrd et al. (1995). This optimization method incorporates box constraints (an upper bound

on a numerical search, initial values for the optimization influences the result. Sometimes the optimization result get stacked at local minima. To mitigate this risk, I performed a robustness check described at Appendix E.

Table 1 summarizes the maximum likelihood estimation results of the TVTP MS model along with the conventional, three-state CTP MS model and GARCH(1,1) model over the whole sample[18]. I use these models as benchmark models for the in-sample and out-of-sample comparison in later sections. To see whether the proposed model is statistically different from a no-regime model, I also conduct a Wald test for the null hypothesis where $\theta_0 = \{\sigma_s, \sigma_m\ \sigma_v\ \psi_u\ \psi_l\ \delta\} = \{\bar{\sigma}\ \bar{\sigma}\ \bar{\sigma}\ 0\ 0\ 1\}$, where $\bar{\sigma}$ is long-run volatility ($\bar{\sigma}$ = 0.010781). The p-values for all the estimators are very close to zero ($p < 0.00001$). This rejects the null hypothesis confidently.

Table 1 shows that the likelihood of the proposed model, $\ln L$, is higher than its peers. In general, when the number of parameters increases, the likelihood becomes higher. Therefore, I also compared the Bayesian Information Criterion (BIC), which adjusts the likelihood by penalizing the high number of parameters. The BIC criterion is given by $BIC = T^{-1}(-2\ln L + NP \cdot \ln T)$ where NP is the number of free parameters in the specification of the model. Even though the BIC penalizes CTP MS and helps GARCH(1,1) to be better off, TVTP MS still has the lowest BIC among the peers. Note that TVTP and CTP MS are not nested each other, but the estimated conditional volatilities ($\hat{\sigma}_s$, $\hat{\sigma}_m$, and $\hat{\sigma}_v$) are almost at the same level. Thus the difference of the likelihood between TVTP MS and CTP MS seems to purely come from the specification of the transition matrix (time-varying or constant).

---

and a lower bound) to a Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton method. See Broyden (1970), Fletcher (1970), Goldfarb (1970), and Shanno (1970)

[18]The three-state CTP MS model has the same return process as the proposed model, $\tilde{r}_{t+1}(S_t) = \mu - \sigma^2(S_t)/2 + \sigma(S_t)\varepsilon_{t+1}$, but has a constant transition probability matrix. This matrix is defined as 3x3 matrix such that, $A = \{(1-\gamma_{sm}-\gamma_{sv}\ \gamma_{sm}\ \gamma_{sv}); (\gamma_{ms}\ 1-\gamma_{ms}-\gamma_{mv}\ \gamma_{mv}); (\gamma_{vs}\ \gamma_{vm}\ 1-\gamma_{vs}-\gamma_{vs})\}$. GARCH(1,1) model has the return process which follows, $\tilde{r}_{t+1} = \mu - \sigma^2_{g,t+1}/2 + \sigma_{g,t+1}\varepsilon_t$ where $\sigma^2_{g,t+1} = \omega^2 + \alpha(r_t - \mu + \sigma^2_{g,t}/2)^2 + \beta\sigma^2_{g,t}$. I do the same calibration as TVTP model for maximum likelihood optimization, such that $\mu = 0.030302\%$ for the two benchmark models.

Figure 1 shows the filtered probability of latent states by the model using the estimated parameters. The blue area corresponds to the probability of being in the stable state S_s, yellow is the probability of being in the middle state S_m, and red is the probability of being in the volatile state S_v. For example, the model filters almost 100% probability of being in a volatile state during the Great Depression (1929), the Black Monday (1987), the Great Recession (2008), and the Covid-19 Pandemic (2020). The time average of filtered state probabilities over the sample period was 60.47% in the stable state ($\bar{\pi}_s \equiv \sum_{t=1}^{T} \hat{\pi}_{s,t}$ = 60.47%), 30.36% in the middle state ($\bar{\pi}_m \equiv \sum_{t=1}^{T} \hat{\pi}_{m,t}$ = 30.36%), and 9.16% in the volatile state ($\bar{\pi}_v \equiv \sum_{t=1}^{T} \hat{\pi}_{v,t}$ = 9.16%). Together with the estimates in Table 1, the expectation of the volatility are roughly equal to the long-run volatility ($\mathbb{E}[\sigma^2(S_t)] \equiv \sum_{i=\{s,m,v\}} \bar{\pi}_i \hat{\sigma}_i^2 = 0.010939^2 \approx \bar{\sigma}^2$). Therefore, the estimated parameters are consistent with the observations.

1.3.4 Return Simulation

Next, I see whether the proposed model generates a return distribution closer to the empirical data than the CTP MS model and the GARCH(1,1) model. I use the estimators in Table 1 to generate simulated return series from each model. Figure 2 shows the observed logarithmic return series from the data, $\{r_t\}_{t=1}^{T}$ (Figure 2 (a)), and the simulated logarithmic return series, $\{\tilde{r}_t\}_{t=1}^{T}$, by each of the three models (Figure 2 (b), (c), (d)). All the graphs have exactly the same scale for the x-axis and y-axis. From Figure 2 (a), the observed return series exhibits apparent volatility clustering, especially around the Great Depression (1929, extremely volatile period on the left part of the chart), the Black Monday (1987, middle right where the significant negative return was observed), and Lehman Brothers bankruptcy (2008, volatile period on the right part of the graph). Also, the observed return distribution shows negative skewness (the magnitude of negative returns is larger than positive returns) with infrequent long tails. Figure (d) shows that GARCH(1,1) can generate nice characteristics of infrequent but clear volatility clusterings like the real returns, but the return pattern shows wild swing and is somewhat symmetric from its center

(zero percent return). Figures (b) and (c) show that the two MS models produce a similar magnitude of volatility as the observed return series in general conditions. The difference between Figure (b) and (c) is return asymmetry. While CTP MS yields a nice pattern of volatility clustering, the return distribution does not show enough skewness (or asymmetry of returns). Though TVTP MS produces slightly more frequent spikes than its peers, the return distribution shows clear skewness on the downside.

1.4 Out-of-sample Comparison with Alternative Models

1.4.1 Point Forecasts

I now compare the out-of-sample performance of the three models. First, I do the maximum likelihood estimation of each model's parameters using only first half of observations. Table 2 shows the maximum likelihood estimators of the parameters over the first half of the sample data[19]. Again, the likelihood of the proposed model is higher than the benchmark models, so is BIC.

To begin with, I compare how well each model performs a point forecast of volatility over a different forecasting horizon. Here, I perform Mincer-Zarnowitz regressions (Mincer and Zarnowitz, 1969) of the volatility forecasts from competing models over one business day, five business day (one week), ten business day (two weeks), twenty business day (one month), forty business day (two months), and sixty business day (one quarter) horizons. Mincer-Zarnowitz regression is a commonly used method in financial econometric literature to evaluate the accuracy of the forecast, which can be written as

[19] Again, I calibrated for the mean for the optimization with the mean of returns over the fist half sample ($\hat{\mu} = 0.031391\%$) for all the models. The fist half of the sample contains 12,448 observations. The same robustness check as for the full sample is applied as shown in the Appendix E.

$$\sigma^2_{t,n} = \gamma_0 + \gamma_1 \mathbb{E}_{t-n}\sigma^2_{t,n} + \varepsilon_t \tag{1.27}$$

where $\sigma^2_{t,n}$ is a variance over n days from time $t-n$ and ε_t is white noise. When the model is correctly specified, the model produces perfect variance forecasts. Thus, the null hypothesis of the unbiased forecast implies, $H_0 : \gamma_0 = 0 \cap \gamma_1 = 1$.

Though the method concept is straightforward, we need to be careful that the true variance $\sigma^2_{t,n}$ is not observable. Hence, I use realized volatility as a proxy for the latent variance as Andersen and Bollerslev (1998). If we replace the true variance $\sigma^2_{t,n}$ in equation (1.27) with a estimate $\hat{\sigma}^2_{t,n}$, which is

$$\hat{\sigma}^2_{t,n} = RV_{t,n} = \sigma^2_{t,n} + \eta_t \tag{1.28}$$

where $RV_{t,n} = \sum_{s=t-n+1}^{t} r_s^2$. $RV_{t,n}$ is realized volatility over n day period from $t-n$, and η_t is white noise and $\mathbb{E}[\eta_t] = 0$. Then equation (1.27) becomes

$$RV_{t,n} = \gamma_0 + \gamma_1 \mathbb{E}_{t-n} RV_{t,n} + u_t \tag{1.29}$$

where $u_t = \varepsilon_t + \eta_t$, and $\mathbb{E}_{t-n}RV_{t,n} = B^{-1}\sum_{k=1}^{B} \hat{RV}_{t,n,k}$, and $\hat{RV}_{t,n,k}$ is the forecast of realized volatility over n day period at the $t-n$ point of time from the model. The multiple-period forecast is path-dependent and requires keeping the error term in the return specification (equation (1.2)) for the forecast. Hence, the single-path variance forecast is noisy. Therefore, I generate $B = 10,000$ paths of forecast at each point of time $(t-n)$ for each forecast horizon (n). Then, I take the average of the B paths of the forecast at each point of time to get the less noisy version of the realized volatility forecast $\mathbb{E}_{t-n}RV_{t,n}$.

If the model concerned is correctly specified, the coefficients of Mincer-Zarnowitz will be $\gamma_0 = 0$ and $\gamma_1 = 1$. Table 3 shows the result of Mincer-Zarnowitz evaluation of realized volatility point forecast. I use the latter half of the subsample to evaluate the performance of forecasting the realized volatility $RV_{t,n}$ for the dependent variable in equation (1.29). In this subsample, there are $T' = 12,448$ observations. Table 3 also shows the p-value of single parameter Wald test for individual parameters against the null hypothesis, such that $H_0 : \gamma_0 = 0$ and $H_0 : \gamma_1 = 1$. Therefore, higher p-values are better because the higher the p-values are, the less the null hypothesis is rejected. The figures in the parentheses are the standard error with Heteroskedasticity and Autocorrelation Consistent (HAC) adjustment using the method of Newey and West (1987), with automatic lag parameter selection (West and Newey, 1994) as described in Appendix C.

The two MS model (TVTP and CTP) yields similar forecast pattern from the out-of-sample observations. The result shows that MS models forecast with higher accuracy than GARCH(1,1)'s prediction, as the intercepts are closer to zero ($\gamma_0 \approx 0$) and the slopes are closer to one ($\gamma_1 \approx 1$). Both MS models tend to forecast the correct level of variance magnitude in the short horizon but tend to forecast less variance in the long horizon since γ_1 monotonically increases and becomes larger than one. This tendency is opposite from GARCH(1,1)'s pattern, where the model forecasts larger variance when the time horizon becomes longer as γ_1 monotonically decreases. Comparing the result from the two MS models, we see there is a slight difference. The TVTP MS model is relatively better at forecasting a longer horizon (more than twenty days) than the CTP MS model, while CTP MS is slightly better at a shorter horizon (less than ten days).

1.4.2 Interval Forecasts

To confirm the magnitude of the swings and negative skewness of the observations from the simulated returns (Figure 2), I evaluate the return distribution with a specific measure that summarizes the distribution pattern. One such metric can be the Value-at-Risk metric

(VaR). While VaR is commonly applied to a value of a portfolio, I define the estimator of VaR by a model at time t to be the pth quantile ($p \in [0,1]$) of the conditional return distribution over the period from time $t+1$ to $t+n$ as

$$\hat{VaR}_{t,n}(p) \equiv \hat{F}_{t,n}(p) \tag{1.30}$$

where $\hat{F}_{t,n}$ is conditional return forecast distribution generated by the model. Therefore, the model expects to experience an asset return, which goes further down the tail of the distribution less than $\hat{VaR}_{t,n}(p)$ over the next n days, only p percent of the time. The accuracy of the VaR, or unconditional coverage property, is verified by recording the number of "hit" rates (Kupiec, 1995, Christoffersen, 1998). The "hit" function flags one when the observed return r_t exceeds the predicted $\hat{VaR}_{t,n}(p)$. The failure rate is the number of times where the "hit" function flags over the given sample period. Denoting the observed cumulative return over n days from time t as $r_{t,t+n} = \sum_{k=t}^{n} r_k$, the indicator ("hit") function which flags the breach of the VaR is defined as

$$I_{t+n}(p) \equiv \begin{cases} 1 & \text{if } r_{t,t+n} < \hat{VaR}_{t,n}(p) \\ 0 & \text{if } r_{t,t+n} \geq \hat{VaR}_{t,n}(p) \end{cases} \tag{1.31}$$

Then the "failure" rate is written as

$$\hat{U}_n(p) \equiv \frac{1}{T'} \sum_{t=1}^{T'-n} I_{t+n}(p) \tag{1.32}$$

If the VaR is accurately estimated, the failure rate converges to p; $\hat{U}_n(p) \to p$. To obtain each model's conditional forecast distribution $\hat{F}_{t,n}$, at each time t, I simulate $\tilde{r}_{t,t+n}$ for

$B = 10,000$ paths $\{\tilde{r}_{t,t+n}\}_{b=1}^{B}$. Then, I simply take the p^{th} quantile of $\{\tilde{r}_{t,t+n}\}_{b=1}^{B}$ at each time t, to assign it to $\hat{VaR}_{t,n}(p)$. Then obtaining $\{I_{t+n}(p)\}_{t=1}^{T'}$ is straightforward. From the construction of the hit function, the time series $\{I_{t+n}(p)\}_{t=1}^{T'}$ is highly likely to have autocorrelation. Therefore, once the $\{I_{t+n}(p)\}_{t=1}^{T'}$ series is generated, I conduct the Ljung-Box test (1978) to evaluate whether the hit indicators exhibit autocorrelation. Since the test statistic of the Ljung-Box test rejects the null hypothesis (H_0: no autocorrelation) for all of the generated hit series $\{I_{t+n}(p)\}_{t=1}^{T'}$ ($n = 1,5,10,20,40,60$) at 3% confidence level, the hit indicator (equation (1.31)) reveals to have a strong autocorrelation. Thus, I use Newey-West Heteroskedasticity and Autocorrelation Consistent standard errors (1987) for the failure rate forecast as described in Appendix C.

Table 4 and Table 5 reports the failure rate of the TVTP MS model, CTP MS model and GARCH(1,1) for $n = \{1,5,10,20,40,60\}$ day forecasts and confidence level at lower tail and upper tail, p = { 1%, 5%, 10%, 90%, 95%, 99% }. The p-value is evaluated against a null hypothesis $H_0 : \hat{U}_n(p) = p$. If the p-value is larger, it is less probable to reject that the forecasted VaR level is indeed at the realized quantile from the observed returns. Therefore, larger p-values imply better forecasts. The boldface numbers are statistically indifferent from the target value-at-risk level p at the 3% confidence level. The result in Table 4 and Table 5 shows that the price threshold TVTP MS model forecasts a more accurate distribution for both the lower tail and the upper tail than competing models. In one day horizon, GARCH(1,1) slightly outperforms its peer at some levels (VaR at 5%, 10%, and 99%). However, when forecasting longer than the five-day horizon, GARCH(1,1) does not forecast VaR well on the lower tail at all (i.e., VaR = 1%, 5%, and 10%). The TVTP and CTP MS models compete with each other. However, at almost all the time horizons and each percentile, the TVTP MS model shows better forecasts than CTP MS by looking at the p-values. The only exceptions where CTP MS outperforms TVTP MS are $p = 1\%$ for twenty, forty, and sixty days and $p = 90\%$ for ten, twenty, and forty days. The proposed model shows a better ability to predict the VaR metric closer to the real return distribution's

percentile from this out-of-sample test. Therefore, we expect to do better risk management by introducing the TVTP to the MS model than other models.

1.4.3 Density Forecasts

Lastly, I evaluate density forecasts by each model with the Probability Integral Transform (Diebold, et al., 1998). The probability integral transform of a sequence of generated forecasts by a model $\{f_t(r_{t,n}|\Omega_t)\}_{t=1}^{T'}$ is a cumulative density function corresponding to the density $f_t(r_{t,n})$ evaluated at r such that

$$\hat{F}_{t,n}(r) \equiv \mathbb{P}_t(r_{t,n} \leq r|\Omega_t) = \int_{-\infty}^{r} f_t(y)dy \tag{1.33}$$

If the forecast generating model is specified correctly, the random variables $U_{t,n} = F_{t,n}(y_{t,n})$ follow an uniform distribution on the interval of $[0,1]$. Thus, if the probability integral transform is closer to a uniform distribution, it can be said that the forecast is more accurate. Therefore, I measure the distance of these two distributions and compare them among the three models. This is an extension of the evaluation of interval forecasting, which checks every interval of the distribution. Following Diebold et al. (1998), I look at the graphical plot of $U_{t,n}$ to see how close the probability integral transform of the forecasts is to the uniform distribution.

Figure 3 and Figure 4 show the probability integral transform of forecasts generated by corresponding models. Figure 3 shows the result for the short horizon, and Figure 4 illustrates the long horizon. The horizontal axis corresponds to percentile from zero to one, and the vertical axis corresponds to the frequency for each percentile bucket. Figure 3 and Figure 4 have a hundred buckets. In each figure, the top three charts are the probability integral transform of forecasts by the TVTP MS model. Similarly, the three charts in the middle row are the probability integral transform of forecasts by the CTP MS model. The

bottom three charts are the probability integral transform of GARCH(1,1). In Figure 3, the three charts in the left column are one-day forecasts by the corresponding models, the center column summarizes the five-day forecasts, and the right column shows the ten-day forecasts. Similarly, in Figure 4, the left column is twenty-day forecasts, the center column summarizes the forty-day forecasts, and the right column shows the sixty-day forecasts. The red horizontal lines in each chart indicate the uniform distribution, and the blue lines are 5% distant from the uniform distribution line.

Overall, the probability integral transform of forecasts by the proposed model has the closest distributions to the uniform distribution among the three models. Especially, the TVTP MS model performs well at forecasting the tail probabilities. If the tail forecast is too loose, the frequency at the tail percentile shoots up, which is seen in the distributions of the CTP MS and GARCH (1,1) models. The graphical result here is consistent with the result of interval forecasts seen in Table 4 and Table 5. The Cramer von-Mise (CVM) statistics in Table 6 also confirm this graphical interpretation. As in Appendix D, the CVM criterion measures the distances between two given distributions. Hence the smaller the CVM is, the closer the distance between the probability integral transform and the uniform distribution, and the better the forecast is. Table 6 shows that the price threshold TVTP mode produces strictly smaller CVM statistics for all the forecast time horizons. The graphical result in Figure 3, Figure 4, and the smaller CVM values in Table 6 indicate that the distribution of the out-of-sample density forecast by the proposed model outperforms other models.

1.5 Conclusion

This paper proposes a specification of the TVTP of an MS model by measuring the distance between the current asset price level and the hypothetical threshold price levels and computing the likelihood of crossing these levels. This paper relates the crossing probability directly to the transition probabilities without specifying the TVTP through a logistic func-

tion and exogenous variables. The model has closed form, is parsimonious, and requires only seven parameters to estimate. From the empirical analysis, the model is able to better predict the return distribution of the US equity index than benchmark models such as the CTP MS model and GARCH(1,1). I compare the out-of-sample point forecast, the out-of-sample interval forecast, and the out-of-sample density forecast of the three models and find that the proposed model predicts more accurate predictions. Even though this paper uses CRSP US equity returns for the analysis, the proposed approach is not limited to equities. The model is also very handy for other asset classes since it only requires asset price dynamics. This model's practical application would be to use it as an indicator of dynamic asset allocation based on the filtered regime of the capital market and as a more accurate quantile-based risk management monitoring. One interesting future research would be to extend the proposed model from three states to a higher number of states to describe better the volatility clustering observed in the data. Also, the lognormal distribution assumption in the error term of the return specification would be too strong since the return distribution exhibits a fat tail. It is interesting to explore further the appropriate distribution.

2 Paper 2, Forecasting Volatility With Price Thresholds: Multiple States Model

Forecasting Volatility With Multi-State Price Threshold

Nobuaki Kato[1]

31st Dec 2021

Abstract

This paper extends the three-state Price Threshold Volatility forecasting model (PTV-3) developed by Kato (2021). The original PTV model assumes only a limited number of states in the economy, which does not have enough volatility components to match the variety of magnitude of volatility in the real world. The new model extends the number of states to $2k+1$ by expanding the time-varying transition probability matrix with one single parameter added to the PTV model. This multi-state PTV model (PTV-M) is parsimonious and has a closed-form likelihood. Therefore, the maximum likelihood estimation can apply to its parameters. The paper evaluates its performance of the point forecast, the interval forecast, and the density forecast for various state numbers k based on the out-of-sample of CRSP S&P 500 return data. As a result, the increase in the number of states improves the return distribution and the performance of out-of-sample forecasting.

JEL classification: C13; C32; G17

Keywords: Time-varying transition probabilities, Markov switching, maximum likelihood, price threshold, moving average, out-of-sample forecasts, multi-states

[1]EDHEC Business School (nobuaki.kato@edhec.com). I am indebted to Laurent Calvet for his guidance and various helpful advice through my research. I am grateful to Abraham Lioui, Christophe Croux, Ichiro Tange, Mirco Rubin, Nikolaos Tessaromatis, Raman Uppal, Riccardo Rebonato, and Vladislav Gounas for their thoughtful comments and suggestions. I am also thankful to Brigitte Bogaerts-Chevillotte, Marine Castel, and Mathilde Legrand for various logistic supports.

2.1 Introduction

Accurate volatility forecast plays a central role in the risk management of financial assets. However, assessing volatility is not an easy task because it varies significantly over time. Even when the volatility seemed to be at a certain level at a certain point in time, it could change dramatically at a different point in time. Also, financial asset typically shows infrequent but largely negative returns, which has a significant economic consequence but is hard to capture by a single estimate because these rare events get averaged out. For this reason, a simple standard deviation from the sample time series is not appropriate for risk management. Therefore, there are extensive researches on volatility modeling in the financial literature.

Several streams of the volatility models try to capture these volatility dynamics and the negative skewness in the return distribution of the financial asset. One such stream is the ARCH/GARCH family, beginning with Engle (1982) and Bollerslev (1986). It addresses the sharp change in the volatility and volatility clustering exhibited in the time series observation. However, by construction, the model cannot produce a sharp change in volatility and long memory of volatility at the same time because there is a trade-off between these features[2].

Another popular stream of volatility modeling is the Markov Switching model (MS). Since Hamilton (1989) applied to the macroeconomic analysis, the MS model has been widely used for financial applications[3]. It models the volatility conditional on the latent states (or "regimes"). This regime switches according to transition probability, and the switch in regime leads to a change in magnitude of volatility. Also, because this switching

[2]In the GARCH model, the stationary condition imposes a trade-off between the reactivity to the shocks and the persistence. For example, the GARCH (1,1) model is written as $\sigma_t^2 = \omega + (\alpha+\beta)\sigma_{t-1}^2 + \alpha\sigma_{t-1}^2(\varepsilon_{t-1}^2 - 1)$, where σ is asset volatility, ε_t is white noise, ω is an intercept, α is a sensitivity to a volatility shock, and $\alpha+\beta$ is decay speed of volatility shock. The GARCH (1,1) requires $\alpha+\beta < 1$ to be stationary. Therefore, if the α becomes relatively large for a stationary process, β becomes relatively small (e.g., spikier and less persistent), and vice versa.

[3]See Hamilton and Raj, (2002) for the review and Calvet and Fisher (2001), (2004), (2007) for the extension.

probability is typically low, one volatility regime tends to persist for a certain period. This persistence leads to a nice feature of volatility clustering. Contrary to the ARCH/GARCH family, the MS model can combine the sensitivity to the shocks and the persistence.

However, the MS model has two drawbacks. First, the typical MS models assume the transition probability constant over time. This constant transition probability (CTP) is a restrictive assumption (Diebold et al., 1994, Filardo, 1994) because the underlying economy changes over time, and the economic condition should impact how easy to switch from one regime to another regime. Second, the MS model requires a large number of states to reproduce the various magnitudes of the volatility. For example, the two-state MS model can produce only two return distributions (for example, "stable-state" distribution and "volatile-state" distribution). However, from the actual return observations, it is clear that the return dynamics show more variety and have more components than the two return distributions. Thus, while the MS model is suitable to capture the volatility dynamics "on average," it cannot capture granular dynamics unless the model increases the number of states.

If the number of the state is small (e.g., two-state MS model), the MS model does not produce enough magnitude in the spike of volatility which can be occasionally seen in the actual historical return distribution of the asset price. This lack of magnitude in volatility spike is due to the very low frequency of such volatility in the real world. Suppose a forecasting model assumes only a small number of states. In that case, these rare volatility spikes are ignored or averaged out because statistical methods, such as Maximum Likelihood Estimation (MLE) or Generalized Method of Moment (GMM), tend to match with more frequent observation when the number of parameter sets is limited. Hence, to reproduce the infrequent but largely negative return, it is necessary to increase the number of states to enough levels where the statistical estimation can capture those rare observations. However, the increase in the number of states quickly increases the number of parameters.

Typically, each transition probability requires different parameters[4].

Many researchers propose solutions to the CTP problem by introducing time-varying transition probabilities (TVTP)[5] but fail to address the lack of states simultaneously. Some use exogenous macroeconomic variables, and some use observations for the input of the TVTP function. However, these TVTP solutions typically focus on a small number of states (e.g., two states) and still have difficulty extending to multiple states. Typical TVTP solutions use a logistic function in each transition probability, which again requires an increase in the parameter number if we extend to multiple states. Therefore, these TVTP models may be able to produce a reasonable return distribution "on average," but they still face a challenge to capture more granular volatility dynamics.

In order to solve the two drawbacks of the MS model, this paper extends the earlier work of the Price Threshold Volatility forecasting model (PTV) developed by Kato (2021). The earlier PTV model assumes three states in the economy (hence, this paper calls it PTV-3), where each state has a different volatility level (the stable state, the middle state, and the volatile state). Here, the initial point in the economy is the middle state. The model further assumes that the economy enters a stable state if the asset price rises and crosses above a certain price threshold. On the other hand, the economy enters a volatile state if the price declines further below a certain threshold. Based on this premise, the PTV model quantifies the probability of the price crossing these thresholds by measuring the distance between the current price level and each threshold. Hence, the model transforms the threshold crossing probability to the transition probability. The PTV-3 model shows improvement of volatility forecast when the proposed TVTP model is compared with conventional models such as the three-state CTP MS and GARCH(1,1) models. This paper aims to extend this PTV-3 model to a multi-state PTV model (PTV-M).

---

[4]For example, typical k-state CTP MS models requires $k(k-1)$ parameters to specify the transition probability matrix. Calvet and Fisher (2001), (2004), (2007) propose a multi-fractal approach to CTP MS model to extend the number of states.

[5]See Engel and hakkio (1996), Gray (1996), Van Norden and Schaller (1997), Filardo (1998), Fong and See (2002), Peria (2002), Yuan (2011), Bazzi et al. (2014), and Wang et al. (2019), for example.

A nice feature of the PTV model is that it does not assume a logistic function to specify each transition probability. Moreover, the specification of the price threshold to quantify the TVTP in the PTV model allows a recursive structure. These characteristics of the PTV model allow the extension to a multiple-state model quite easily. Unlike other MS models, the proposed PTV-M model requires only one additional parameter to control the number of states. This single additional parameter gives an econometrician the flexibility to change from three states to an infinite number of states. The specification of the model is also parsimonious and has a closed-form.

Also, there are economic rationales to use the price dynamics as the TVTP function in the PTV model as follows:

1. The market microstructure literature shows that asset price contains rich information about the economic condition[6]. Since the asset price reflects underlying and anticipated economic conditions, it is reasonable to infer the latent states from the asset price.

2. The asset price shows the so-called leverage effect, which shows a negative correlation between the asset returns and the return volatility[7]. Therefore, the positive asset return is often associated with the decrease in volatility, and the negative asset return is often associated with the increase in volatility.

3. The price movement of the financial asset has a feedback effect on the real economy[8]. The current price movement influences the future action of decision-makers of the firms and financial market participants.

The remainder of the paper is organized as follows: Section 2.2 develops the extension of the PTV-3 model. Section 2.3 presents an empirical analysis of the extended PTV-M

[6]See Wilson (1974), Milgrom (1979), Glosten and Milgrom (1985), Kyle (1985), Palfrey (1985), and Easley and O'Hara (1987) (1991) (1992) for the examples.

[7]See Black (1976), Christie (1982), French et al. (1987) and a good review by Bekaert and Wu (2000).

[8]See Morck et al. (1990) and Bond et al. (2012).

model and provides an in-sample comparison of simulated return distributions. Section 2.4 conducts out-of-sample analysis; point forecast, interval forecast, and density forecast. Section 2.5 concludes.

2.2 Multi-States TVTP Model

2.2.1 Specification

This paper extends the PTV-3 model to the PTV-M model, which examines a case where there is one asset and latent $2k+1$ states of the economy at one point in time. This state of economy varies over time and is described such that

$$S_t \equiv S_i = \{S_{-k}, S_{-k+1}, ..., S_{-1}, S_0, S_1, ..., S_{k-1}, S_k\} \in \mathbb{R}_+^{2k+1} \tag{2.1}$$

$i \in \{-k, -k+1, ..., -1, 0, 1, ...k-1, k\}$ is an indicator variable, S_0 is the median state of the economy, S_k is the most stable state, and S_{-k} is the most volatile state. When S_t goes toward S_k, the state becomes stabilized monotonically (the volatility in S_1 is lower than volatility in S_0, S_2 is lower than S_1, and so on). On the other hand, when S_t goes toward S_{-k} the state becomes volatile monotonically (the volatility in S_{-1} is higher than volatility in S_0, S_{-2} is higher than S_{-1}, and so on). The volatility of the asset price conditional on a state S_t is characterized as

$$\sigma(S_t) \equiv \sigma_i = \begin{cases} \bar{\sigma} \times a^i & \text{if } i \geq 0 \\ \bar{\sigma} \times b^i & \text{if } i < 0 \end{cases} \tag{2.2}$$

where $\bar{\sigma} \in \mathbb{R}_+$ is a constant variable, and $a \in (0,1)$ and $b \in (0,1)$ are spacing parameters. Since a and b are positive constant but less than one, $\sigma_{-k} > \sigma_{-k+1} > ... > \sigma_{-1} > \bar{\sigma} > \sigma_1 >$

$... > \sigma_{k-1} > \sigma_k$. If the spacing parameters a and b become close to one, the change from σ_i to σ_{i+1} becomes small, and if a and b become close to zero, the change from σ_i to σ_{i+1} becomes large.

The one-period logarithmic total return of the asset $r_t \equiv \ln(P_t/P_{t-1})$ follows the process, such that

$$r_{t+1}(S_t) = \mu - \frac{\sigma^2(S_t)}{2} + \sigma(S_t)\varepsilon_{t+1} \tag{2.3}$$

where μ is a positive constant, ε_{t+1} is a random variable that follows IID standard Gaussians $\mathcal{N}(0,1)$. Here, μ is a long-run mean of return because $\mathbb{E}[e^{r_{t+1}}] = e^{\mu}$.

At time t with the state $s_{t-1} = S_i$, the transition probability of the first-order Markov chain is specified as

$$\begin{aligned} \gamma_{i,j,t} &\equiv \mathbb{P}(s_t = S_j | s_{t-1} = S_i, s_{t-2} = S_m, ...; \mathscr{I}_{1:t-1}) \\ &= \mathbb{P}(s_t = S_j | s_{t-1} = S_i; \mathscr{I}_{1:t-1}) \end{aligned} \tag{2.4}$$

where $j \in \{-k, -k+1, ..., -1, 0, 1, ..., k-1, k\}$ is an indicator of next state S_j, $\mathscr{I}_{1:t-1}$ is cumulative information observed by econometrician from the past price sequence from time $t = 1$ until $t = t-1$. Then, the state dynamics of the MS model are defined by the time-varying transition matrix such that

$$A_t \equiv \begin{bmatrix} \gamma_{k,k,t} & \gamma_{k,k-1,t} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \gamma_{k,-k+1,t} & \gamma_{k,-k,t} \\ \gamma_{k-1,k,t} & \gamma_{k-1,k-1,t} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \gamma_{k-1,-k+1,t} & \gamma_{k-1,-k,t} \\ \vdots & \vdots & \ddots & & & & & & & \vdots & \vdots \\ \vdots & \vdots & & \ddots & & & & & & \vdots & \vdots \\ \vdots & \vdots & & & \ddots & & & & & \vdots & \vdots \\ \gamma_{0,k,t} & \gamma_{0,k-1,t} & \cdots & \cdots & \gamma_{0,-1,t} & \gamma_{0,0,t} & \gamma_{0,1,t} & \cdots & \cdots & \gamma_{0,-k+1,t} & \gamma_{0,-k,t} \\ \vdots & \vdots & & & & & \ddots & & & \vdots & \vdots \\ \vdots & \vdots & & & & & & \ddots & & \vdots & \vdots \\ \vdots & \vdots & & & & & & & \ddots & \vdots & \vdots \\ \gamma_{-k+1,k,t} & \gamma_{-k+1,k-1,t} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \gamma_{-k+1,-k+1,t} & \gamma_{-k+1,-k,t} \\ \gamma_{-k,k,t} & \gamma_{-k,k-1,t} & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \gamma_{-k,-k+1,t} & \gamma_{-k,-k,t} \end{bmatrix} \tag{2.5}$$

The next sub-section describes the specification of this $(2k+1) \times (2k+1)$ matrix.

2.2.2 Price Dynamics

In the PTV model, the TVTP to a certain state is quantified as a probability of an asset price crossing thresholds at a certain price level corresponding to the states. This probability is measured by the distance between the current price and the price thresholds and the price volatility with the assumption of normal distribution. Thinking about a building structure is helpful to show this switching mechanism as follows.

Imagine we are in a building with k floors above the ground and k basements floors under the ground. To clarify, this paper uses a European notation of counting the floor, such as the ground floor (the level zero) for the first floor in the US. The ground floor is the median floor of this building, and we start from here. If we would like to go up to the next floor, we use stairs to cross the ceiling of the ground floor. After climbing up to the next floor, we may stop there or may go up again to the second floor. On the other hand, we can go down to the first basement or the second basement. If our feeling to decide to go upstairs or downstairs is governed by a normal distribution, we are more likely to go up or down to the adjacent floor than to go up or down a floor two floors away from the current floor at one time. Also, if we tend to move longer distance in a unit of time, we are more likely to go up to the next floor or go down to the basement.

These ascending and descending stairs are analogous to the model setting. Consider an exponentially weighted moving average (EWMA) as a median state of the economy, and we start from here, which is the ground floor in the building analogy. An upper threshold lies above this EWMA, and if we cross this upper threshold, we go to a stable state S_1 (the first floor). We may stop there or may cross another upper threshold again to go to a more stable state S_2 (the second floor) at once. On the other hand, if we cross the lower threshold below the EWMA, we go to a volatile state S_{-1} (the first basement). We may go back again to the median state S_0 (the ground floor), or may go further to a more volatile state. How often we may cross these thresholds is determined by the price volatility. The price return has a normal distribution in the error term, and the crossing probability is quantified in the same way to quantify an exercise probability of out-of-the-money European option security (Black and Scholes, 1973). The higher the volatility is, the more likely the price is to cross the threshold. Also, the price is more likely to cross the adjacent threshold than the thresholds two states away from the current state.

Because the price volatility is vital to quantify the probability in this setting, a careful treatment is required. As explained later, when we quantify the probability of crossing multiple thresholds at one unit of time (for example, going to the second floor from the ground floor without staying on the first floor), we need to adjust the volatility accordingly. This is because the magnitude of volatility is different across the states by definition. For example, the volatility of the stable state (the first floor) is lower than the volatility of the middle state (the ground floor). Thus, if we measure the probability of going to a more stable state (for example, the second state), the proposed model uses a weighted average of volatility in these two states to quantify the probability, as explained later in this paper.

Switching Thresholds of Economic States

Given the asset price series $\{P_t\}_{t=1}^{T}$, an observed exponentially weighted moving average (EWMA) is defined with a specific parameter $\delta \in (0,1)$ as below.

$$
\begin{aligned}
EWMA_t &\equiv \delta P_t + (1-\delta) EWMA_{t-1} \\
EWMA_1 &\equiv P_1
\end{aligned}
\tag{2.6}
$$

I introduce a price threshold $K^i_{j,t}$ such that

$$
K^i_{j,t} \equiv EWMA_{t-1} \times \kappa^i_j \tag{2.7}
$$

where κ^i_j is a (constant) multiplier, which gives a distance of each threshold from EWMA (ceilings and floors of each floor in the building analogy). Here again, i is an indicator of the current state, and j is an indicator of the next state. (For example, if we are on the second floor now, $i = 2$, and if we quantify the distance to the fourth floor, $j = 4$.) Note that the $K^i_{j,t}$ is determined at time $t-1$[9]. For "climbing" the thresholds (going to upstairs, which is $j \geq i$), define this multiplier $\kappa^i_j = \kappa^i_{u,j}$ as a function of spacing parameters a, and b, with additional parameters ψ_u and ψ_l such that

$$
\begin{aligned}
\kappa^i_{u,j} &\equiv \frac{1+\psi_u a^{j-1}}{1-\psi_l b^{j-1}} \kappa^i_{u,j-1} && \text{if } j > i+1 \\
\kappa^i_{u,j} &\equiv 1+\psi_u a^i && \text{if } j = i+1 \\
\kappa^i_{u,j} &\equiv 1 && \text{if } j = i
\end{aligned}
\tag{2.8}
$$

where $\psi_u \in (0, 1/\delta - 1)$ and $\psi_l \in (0, 1/\delta - 1)$ are constant. Note that this threshold multiplier is not time-varying. Similarly, define a multiplier $\kappa^i_j = \kappa^i_{l,j}$ for "descending" the thresholds (going to downstairs, which is $j \leq i$), such that

[9] See Appendix A for why the price threshold of time t can be fixed at $t-1$.

$$
\begin{aligned}
\kappa^i_{l,j} &\equiv \frac{1-\psi_l b^{j+1}}{1+\psi_u a^{j+1}}\kappa^i_{l,j+1} && \text{if } j < i-1 \\
\kappa^i_{l,j} &\equiv 1-\psi_l b^i && \text{if } j = i-1 \\
\kappa^i_{l,j} &\equiv 1 && \text{if } j = i
\end{aligned}
\tag{2.9}
$$

The threshold multiplier $\kappa^i_j = \{\kappa^i_{u,j}, \kappa^i_{l,j}\}$ has a recursive structure. Since a, b, ψ_u, and ψ_l are all positive constant, $\kappa^i_{u,j+1}$ is always larger than $\kappa^i_{u,j}$, and $\kappa^i_{l,j+1}$ is always larger than $\kappa^i_{l,j}$ for the same state i. When $j-i$ increases, the upward distances from a state i becomes larger (or downward distance becomes shorter), and when $j-i$ decreases, the downward distances from the state i become larger (or upward distance becomes shorter).

The spacing parameters a and b serve as an adjustment factor, which controls the bandwidth of price thresholds. Due to this adjustment factor, the marginal distance between each "upper" state (overground floors) diminishes when the price crosses a state which lies above EWMA ($j > 0$). For example, the distance between the second threshold and the third threshold above the EWMA is shorter than the distance between the first threshold and the second threshold. On the other hand, the marginal distance between each "lower" state (underground floors) increases when the price crosses a state which lies below from EWMA ($j < 0$). For example, the distance between the second threshold and the third threshold below from the EWMA is longer than the distance between the first threshold and the second threshold. The reason is as follows.

Imagine if the marginal threshold distances were equal across different states. It would be less probable to cross the threshold from the stable states because the price fluctuates less, and the distance between the two thresholds becomes relatively longer. Also, it would be more probable to cross from the volatile states because the price fluctuates more, and the relative distance between the two thresholds becomes shorter. Therefore, if the bandwidth is unconditionally constant regardless of the state, an asset price becomes very stable for a long time once it enters a stable state because there is less switch from the stable state.

Also, the volatility memory dies out very quickly because even though the asset price enters a volatile state, it would immediately exit from the state thanks to heightened asset fluctuation. However, as an asset price empirically shows long volatility memory (i,e, volatile state lasts long once the economy enters the volatile state), the constant thresholds distance is not coherent with the data. Therefore, there must be some adjustment on the marginal distance conditional on the state. This adjustment increases switching probability from a stable state and reduces switching probability from the volatile state.

Another characteristics of κ_j^i is that the form is different for $j > i$ and $j < i$. Imagine that the price is now in the median state S_0 and there is another price level, $EMWA_t^{'}$, above the first upper threshold K_1^0, which satisfies $EWMA_t^{'} \times \kappa_0^1 = EWMA_t \times \kappa_1^0$. Then this $EMWA_t^{'}$ can be viewed as an imaginary EWMA if the economy were in the one-step upper state S_1, looking from S_0. This condition means that there are two different states S_0 and S_1, which share the same borderline as a lower threshold in perspective of $EWMA_t^{'}$ and an upper threshold in the perspective of $EWMA_t$. This imaginary EWMA can be written by $EMWA_t^{'} = EWMA_t \times \kappa_1^0/\kappa_0^1 = EWMA_t(1+\psi_u)/(1-\psi_l b)$. Then, by using this imaginary EWMA, the upper threshold to switch to the the more stable state (S_2) from the median state (S_0) in one unit of time can be expressed as $K_2^0 = EMWA_t^{'} \times \kappa_2^1 = EWMA_t\{(1+\psi_u)/(1-\psi_l b)\} \times (1+\psi_u a) = EWMA_t\{(1+\psi_u a)/(1-\psi_l b)\} \times \kappa_1^0$. This discussion is applied to higher j recursively. Hence, $\kappa_{u,j}^i$ is defined as equation (2.8). The same argument is applied for $\kappa_{l,j}^i$, and therefore it is defined as equation (2.9).

In summary, the price thresholds are described as a matrix as below.

$$K_t \equiv \begin{bmatrix} EWMA & K^{k}_{k-1,t} & \dots & \dots & \dots & \dots & \dots & \dots & \dots & K^{k}_{-k+1,t} & K^{k}_{-k,t} \\ K^{k-1}_{k,t} & EWMA & \dots & \dots & \dots & \dots & \dots & \dots & \dots & K^{k-1}_{-k+1,t} & K^{k-1}_{-k,t} \\ \vdots & \vdots & \ddots & & & & & & & \vdots & \vdots \\ \vdots & \vdots & & \ddots & & & & & & \vdots & \vdots \\ \vdots & \vdots & & & \ddots & & & & & \vdots & \vdots \\ K^{0}_{k,t} & K^{0}_{k-1,t} & \dots & \dots & K^{0}_{1,t} & EWMA & K^{0}_{-1,t} & \dots & \dots & K^{0}_{-k+1,t} & K^{0}_{-k,t} \\ \vdots & \vdots & & & & & \ddots & & & \vdots & \vdots \\ \vdots & \vdots & & & & & & \ddots & & \vdots & \vdots \\ \vdots & \vdots & & & & & & & \ddots & \vdots & \vdots \\ K^{-k+1}_{k,t} & K^{-k+1}_{k-1,t} & \dots & \dots & \dots & \dots & \dots & \dots & \dots & EWMA & K^{-k+1}_{-k,t} \\ K^{-k}_{k,t} & K^{-k}_{k-1,t} & \dots & \dots & \dots & \dots & \dots & \dots & \dots & K^{-k}_{-k+1,t} & EWMA \end{bmatrix} \tag{2.10}$$

Using this price threshold, each state in the economy is defined as

$$S_t \equiv \begin{cases} S_k & \text{if } K_{k,t} < P_t \\ S_{k-1} & \text{if } K_{k-1,t} < P_t < K_{k,t} \\ \vdots & \\ S_1 & \text{if } K_{1,t} < P_t < K_{2,t} \\ S_0 & \text{if } K_{-1,t} < P_t < K_{1,t} \\ S_{-1} & \text{if } K_{-2,t} < P_t < K_{-1,t} \\ \vdots & \\ S_{-k+1} & \text{if } K_{-k,t} < P_t < K_{-k+1,t} \\ S_{-k} & \text{if } P_t < K_{-k,t} \end{cases} \tag{2.11}$$

Therefore, from equation (2.4), the transition probabilities can be restated as

$$
\begin{aligned}
\gamma_{i,k,t} &= \mathbb{P}[S_k|S_i] = \mathbb{P}[K^i_{k,t} < P_t | S_{t-1} = S_i; EWMA_{t-1}] && \text{if } i \neq j,\ j = k \\
\gamma_{i,j,t} &= \mathbb{P}[S_j|S_i] = \mathbb{P}[K^i_{j,t} < P_t < K^i_{j+1,t} | S_{t-1} = S_i; EWMA_{t-1}] && \text{if } i \neq j,\ 0 < j < k \\
\gamma_{i,0,t} &= \mathbb{P}[S_0|S_i] = \mathbb{P}[K^i_{-1,t} < P_t < K^i_{1,t} | S_{t-1} = S_i; EWMA_{t-1}] && \text{if } i \neq j,\ j = 0 \\
\gamma_{i,j,t} &= \mathbb{P}[S_j|S_i] = \mathbb{P}[K^i_{j-1,t} < P_t < K^i_{j,t} | S_{t-1} = S_i; EWMA_{t-1}] && \text{if } i \neq j,\ -k < j < 0 \\
\gamma_{i,-k,t} &= \mathbb{P}[S_{-k}|S_i] = \mathbb{P}[P_t < K^i_{-k,t} | S_{t-1} = S_i; EWMA_{t-1}] && \text{if } i \neq j,\ j = -k \\
\gamma_{i,i,t} &= \mathbb{P}[S_i|S_i] = 1 - \sum_{m=-k, m\neq i}^{k} \gamma_{i,m,t} && \text{if } i = j
\end{aligned}
\tag{2.12}
$$

As described in the following subsection, quantification of the probability in the specification above is analogous to an exercise probability of out-of-the-money European option security (Black and Scholes, 1973) conditional on the latent states.

Transition Probabilities

Now, let's quantify the state probability of time t at time $t-1$ in equation (2.12). From equation (2.3), at time $t-1$, the next period's price P_t being higher than the upper threshold level $K^i_{j,t}$, conditional on the state $S_{t-1} = S_i$, where $i \in \{-k, ..., -1, 0, 1, ..., k\}$, is written as

$$
\begin{aligned}
P_t > K^i_{j,t} &\Leftrightarrow \ln P_t > \ln K^i_{j,t} \\
&\Leftrightarrow \ln P_{t-1} + \mu - \frac{\sigma^2(S_i)}{2} + \sigma(S_i)\varepsilon_{t+1} > \ln K^i_{j,t} \\
&\Leftrightarrow -\varepsilon_{t+1} < \frac{\ln\left(\frac{P_{t-1}}{K^i_{j,t}}\right) + \mu - \frac{\sigma^2(S_i)}{2}}{\sigma(S_i)}
\end{aligned}
\tag{2.13}
$$

Since ε follows IID standard Gaussian, the above quantity follows the same. Thus, if we transform the price threshold $K^i_{j,t}$ by the volatility $\sigma(S_i)$, it follows IID Gaussian, and we can quantify the probability.

However, this volatility must be adjusted when the asset price goes across multiple thresholds in one unit of time, such that

$$\begin{aligned} h^i_j &\equiv \sigma_i && \text{if } j = i+1 \text{ or } j = i-1 \\ h^i_j &\equiv \frac{\kappa^i_{j-1}-1}{\kappa^i_j-1} h^i_{j-1} + \frac{\kappa^i_j-\kappa^i_{j-1}}{\kappa^i_j-1}\sigma_{j-1} && \text{if } j > i+1 \\ h^i_j &\equiv \frac{1-\kappa^i_{j+1}}{1-\kappa^i_j} h^i_{j+1} + \frac{\kappa^i_{j+1}-\kappa^i_j}{1-\kappa^i_j}\sigma_{j-1} && \text{if } j < i-1 \end{aligned} \tag{2.14}$$

h^i_j is a weighted average of the volatility in the state j, and the weighted average of volatility from median state to the state $j-1$ (when ascending, $j > i+1$), or until the state $j+1$ (when descending, $j < i+1$).

Imagine that we are at the median state (the ground floor). If we go up to the second upper state (the second floor) in one unit of time, we first go up to the first upper state and then climb up to the second upper state. We know that, as soon as we reach the first upper state, we calm down a little bit and become a little bit less likely to move around than we were at the ground floor (volatility becomes more stable in the first upper state). Therefore, when we aim to go to the second floor from the ground floor at once, we need to take into account this expected change in the behavior at the first upper state. If we want to measure the probability of crossing the second upper threshold from the median state, we cannot apply the single volatility of median state (σ_0) all the way up to the second upper threshold. σ_0 is only applicable up to the first upper threshold, and from there, the new volatility σ_1 must be applied. Thus, the adjusted volatility is a weighted average of volatilities with distances between each threshold.

The marginal distance between the EWMA and the first upper threshold, and between the first upper threshold and the second upper threshold are governed by threshold multiplier κ^i_j, as shown in equation (2.8) and (2.9). For example, the adjusted volatility from EWMA to the second upper threshold is determined by $(K^0_{2,t} - EWMA_t)h^0_2 = (K^0_{1,t} - EWMA_t)h^0_1 +$

$(K^0_{2,t} - K^0_{1,t})\sigma_1 \Leftrightarrow h^0_2 = \{(K^0_{1,t} - EWMA_t)/(K^0_{2,t} - EWMA_t)\}h^0_1 + \{(K^0_{2,t} - K^0_{1,t})/(K^0_{2,t} - EWMA_t)\}\sigma_1$. Each term of the right hand side of this equation contains $EWMA$ and cancels out between numerator and denominator, we get equation (2.14).

Alike Black and Scholes (1973), define $d^i_{j,t}$ using this adjusted volatility h^i_j such that

$$d^i_{j,t} \equiv \frac{\ln\left(\frac{P_{t-1}}{K^i_{j,t}}\right) + \mu - \frac{1}{2}(h^i_j)^2}{h^i_j} \tag{2.15}$$

then $d^i_{j,t}$ follows IID standard Gaussian $\mathcal{N}(0,1)$. Thus, from equation (2.12) and (2.15), the probability of the asset price crossing above the uppermost threshold $j = k$, conditional on the state i is

$$\gamma_{i,k,t} = \mathbb{P}[P_t > K^i_{k,t} | S_{t-1} = S_i] = \Phi(d^i_{k,t}) \tag{2.16}$$

where Φ is Gaussian cumulative distribution function. Note that the transition probabilities of crossing the upper threshold of time t, at time $t = t-1$, are fully quantified with the information available at time $t = t-1$ (because the threshold $K^i_{j,t}$ is determined at the time $t-1$).

Similarly, to quantify the probability of price being lower than the lowermost threshold $j = -k$, $d^i_{j,t}$ leads to the transition probabilities of crossing the lowermost threshold at time t, conditional on the state i at time $t = t-1$, which can be written from equation (2.12) and (2.15) as

$$\gamma_{i,-k,t} = \mathbb{P}[P_t < K^i_{-k,t} | S_{t-1} = S_i] = 1 - \Phi(d^i_{-k,t}) \tag{2.17}$$

Finally, from equations (2.12) and (2.15), the transition probabilities for the intermediate

state conditions ($i \neq j$, $j \neq k$ and $j \neq -k$) are

$$\begin{aligned}
\gamma_{ij,t} &= \mathbb{P}[K^i_{j,t} < P_t < K^i_{j+1,t} | S_{t-1} = S_i] = \Phi(d^i_{j,t}) - \Phi(d^i_{j+1,t}) \text{ if } j > 0 \\
\gamma_{i0,t} &= \mathbb{P}[K^i_{-1,t} < P_t < K^i_{1,t} | S_{t-1} = S_i] = \Phi(d^i_{-1,t}) - \Phi(d^i_{1,t}) \text{ if } j = 0 \\
\gamma_{ij,t} &= \mathbb{P}[K^i_{j-1,t} < P_t < K^i_{j,t} | S_{t-1} = S_i] = \Phi(d^i_{j-1,t}) - \Phi(d^i_{j,t}) \text{ if } j < 0
\end{aligned} \tag{2.18}$$

Each transition probability at time t, $\gamma_{ij,t}$, is fully characterized by the observations at time $t-1$ with the Gaussian cumulative distribution function. When the price gets close to a threshold, the transition probability becomes high. When the price gets far from the threshold, the transition probability becomes low. This price threshold model has a closed-form likelihood, and the maximum likelihood estimation can apply to the parameters.

To summarize the above discussion, the PTV-M model requires eight parameters as follows.

$$\theta = (\bar{\sigma}\ a\ b\ \psi_u\ \psi_l\ \delta\ \mu\ k) \in \mathbb{R}^8_+ \tag{2.19}$$

where $\bar{\sigma}$ is volatility in the median state, a, b are spacing parameters for price thresholds, ψ_u and ψ_l are price threshold parameters, δ is an exponential moving average parameter, μ is the long-term mean of the asset returns, and k is the adjustment parameter for the number of states. Note that the extension of PTV-3 model to multiple states (PTV-M) only requires one additional parameter k, and using two spacing parameters, a and b, instead of volatility in each state. This model is fully flexible as it can accept any natural number of k.

2.3 Empirical Analysis

2.3.1 Maximum Likelihood Estimation

Now I estimate the model parameters by a historical dataset. The conditional probability density function with $2k+1$ states is

$$\omega_t = (f(r_t|s_t = S_k, r_{1:t-1}; \theta) \ldots f(r_t|s_t = S_0, r_{1:t-1}; \theta) \ldots f(r_t|s_t = S_{-k}, r_{1:t-1}; \theta))' \quad (2.20)$$

where $f(r_t|s_t = S_j, r_{1:t-1}; \theta)$ is

$$f(r_t|s_t = S_j; r_{1:t-1}; \theta) = \frac{1}{\sqrt{2\pi\sigma^2(S_j)}} \exp\left[- \frac{\{r_t - \mu + \sigma^2(S_j)/2\}^2}{2\sigma^2(S_j)} \right] \quad (2.21)$$

The filtered probabilities of the latent states conditional on the observed returns satisfy[10]

$$\hat{\Pi}_t = \frac{\omega_t * (\hat{\Pi}_{t-1} A_t)}{[\omega_t * (\hat{\Pi}_{t-1} A_t)]\iota} \quad (2.22)$$

where $*$ denotes element-by-element multiplication and $\iota = (1 \ldots 1)' \in \mathbb{R}^{2k+1}$. Then the likelihood function is written as[11]

$$\ln L(\theta; r_{1:T}) = \sum_{t=1}^{T} \ln[\omega_t \cdot (\hat{\Pi}_{t-1} A_t)] \quad (2.23)$$

Then, the maximum likelihood estimator is defined by

[10] See Appendix B for the proof.

[11] See Appendix B for the proof.

$$\hat{\theta}_{ML} = \arg\max_{\theta} \{\ln L(\theta; r_{1:T})\} \tag{2.24}$$

Since the objective function of equation (2.24) is non-linear in general, it is difficult to solve analytically. Therefore, the optimization is conducted numerically in practice.

2.3.2 Sample Data

This paper retrieves the return data from Prof. Kenneth French's website for the empirical analysis[12]. This data contains the daily excess return on the value-weighted US equity index calculated by the Center for Research in Security Prices (CRSP) over July 1st, 1926 to December 31st, 2020, which has $T = 24,896$ observations. Over this period, the excess return series has a historical mean of 0.030302% per day (about 7% per year). The Augmented Dickey-Fuller (ADF) test (1996) brings no evidence of non-stationarity in the return data[13]. Using excess return has merit compared to using raw stock return for the long time horizon analysis because it is not affected by the interest regime change in the US in 1951 (Treasury-Federal Reserve Accord). I construct the price level of the US equity at time t from the excess return by $P_t \equiv \prod_{k=0}^{t}(1+R_k)$, where R_t is observed excess return at time t.

2.3.3 Model Parameter Estimation

For optimizing equation (2.24), the historical mean is derived from the sample data as $\hat{\mu} = 0.030302\%$ and is used as a calibrated value for μ[14].

[12] https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/

[13] The p-value of ADF test is less than 0.01 for 0 to 20 lags.

[14] The reason for the calibration is because the conditional expectation of the return is μ ($\because \mathbb{E}[e^{\tilde{r}_t}] = e^{\mu}$), and it is preferable that the estimated model to be consistent with reasonable values for the long-run mean. Also, optimization constraints are set to parameters such that $\bar{\sigma} \in (0.001, 0.100)$, $a \in (0.001, 0.999)$, $b \in (0.001, 0.999)$, $\psi_u \in (0.001, 0.100)$, $\psi_l \in (0.001, 0.100)$, and $\delta \in (0,1)$. These constraints are required

For the empirical analysis, I conduct MLE with various $k \in \{1,2,...,10\}$. To proceed with the optimization process, first, I estimated parameters for the $k = 1$ model with some robustness check[15]. Then, I used the MLE estimators of $k = 1$ as an initial value for the estimation of $k = 2$. The same principle applies to all the subsequent ks (for example, the MLE estimators of $k = 2$ for the initial value to estimate $k = 3$, the MLE estimators of $k = 3$ for the initial value to estimate $k = 4$, and so on).

Table 7 summarizes the maximum likelihood estimation results of the PTV-M model by increasing k from 1 to 10 over the whole sample. I use the $k = 1$ model as a benchmark model for the in-sample and out-of-sample comparison in later sections. From the table, the likelihood, $\ln L$, rises as k increases monotonically up to $k = 8$ and then stabilizes at that point. $\hat{\bar{\sigma}}$ is almost stable at about one percent, but gradually and monotonically decreases when k increases. This is because the asset price exhibits negative skewness in its return distribution. When k increases, the model matches a larger magnitude on the downside volatility than the upside volatility. But, the model needs to counterweight both sides of volatility to match with historical volatility because the model assumes the symmetric number of states on the upper side and the lower side (k states on the upper side and k states on the lower side). Therefore, it needs to reposition the median state volatility $\bar{\sigma}$ on the more stable side of the economy (upper side).

The estimates, $\hat{a}$ and $\hat{b}$ increase monotonically and approach to one when k increases. From equations (2.2), (2.8) and (2.9), a and b are spacing parameters, and if a and b approach zero, the bandwidth between one threshold to the next threshold widens. Contrary, when a and b approach one, the bandwidth of thresholds narrows. Hence, it does make sense that these spacing parameters approach one, because, as it is like cutting a cake to

---

to make sure that the Hessian matrix is invertible during the optimization. Because the optimization is done in a numerical fashion, actual optimization was done by scaling a, b, and δ by $1/10$ to make all the parameters' size in the same order to ease numerical search. Therefore, the optimization method in Byrd et al. (1995) is preferable. This optimization method incorporates box constraints (an upper bound and a lower bound) to a Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-Newton method. See Broyden (1970), Fletcher (1970), Goldfarb (1970), and Shanno (1970).

[15] See Appendix F for the robustness check.

more slices and each slice becomes thinner, increasing state number k narrows distances of each threshold. Also, $\hat{b}$ is smaller than $\hat{a}$ in general. From equations (2.2), this $a > b$ implies an asymmetry in the volatility states that the downside volatility (mainly governed by b) is disproportionally higher than the upside volatility (mainly governed by a).

The cake analogy can be applied to parameters $\hat{\psi}_u$ and $\hat{\psi}_l$, which also controls bandwidths of thresholds. These parameters decrease almost monotonically as k increases. This decrease is also because cutting the economy by multiple states makes each bandwidth thinner. Lastly, $\hat{\delta}$ hovers between 0.60 and 0.64, but no clear relationship with k is observed.

2.3.4 Model Selection

I calculate Vuong statistics (1989) to see whether there is a statistically significant difference among the likelihood produced by each k model. If we consider two densities f and g, the Vuong statistics is written such that,

$$T^{-1/2}LR_T = \frac{1}{\sqrt{T}}(\ln L_T^f - \ln L_T^g) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \ln \frac{f(r_t|r_1,...,r_{t-1})}{g(r_t|r_1,...,r_{t-1})} \tag{2.25}$$

where LR_T is the likelihood ratio of $\ln L_T^f$ and $\ln L_T^g$, L_T^f is the likelihood function of density f with T samples, and L_T^g is the likelihood function of density g with T samples. I compare the density of the $k \neq 10$ models with the $k = 10$ model, hence the null hypothesis is $H_0 : L_T^k = L_T^{10}$ where $k \neq 10$. When the return observation $\{r_t\}$ is IID, Vuong (1989) shows that the likelihood ratio LR_T is asymptotically normal under the null, and the variance of this likelihood ratio is consistently estimated by the sample variance of the addends $\ln[f(r_t|r_1,...,r_{t-1})/g(r_t|r_1,...,r_{t-1})]$. However, since the financial return series is typically not IID (for example, the return is affected by the latent states), I construct a heteroskedasticity and autocorrelation consistent (HAC) version of the Vuong test by following Calvet

and Fisher (2004)[16].

Table 9 shows the result of the Vuong test. Panel A shows the t-statistics and p-values for the full sample, and Panel B shows HAC adjusted version. Panel C and Panel D corresponds to the same test as Panel A and Panel B rspectively, but for the half sample. Table 9 shows that the p-value increases when the k approaches 10. For the full sample, with the negative Vuong statistics, the $k = 10$ model significantly outperforms the model with $k \in \{1,2,3\}$ and the model with $k = 4$ at a 2% significance level. From $k = 5$, the significance level decreases monotonically as k approaches 10 and hits bottom (largest p-value) at $k = 8$. This tendency is the same for the half sample, where the model with $k = 10$ significantly outperforms the models up to $k = 2$ and $k = 3$ with 3% significance. From $k = 4$, the significance level reduces monotonically until $k = 8$. From these observations, I conclude that the increase of k up to $k = 8$ improves the model, but from $k = 9$, the model improvement is limited even if we increase k.

2.3.5 Return Simulation

Next, I see whether increasing the number of state k improves the generation of a return distribution closer to the empirical data. I use the estimators in Table 7 to generate simulated return series from each k. Figure 5 and 6 show the observed logarithmic return series from the data, $\{r_t\}_{t=1}^{T}$ ("(a) Observed Time Series"), and the simulated logarithmic return series, $\{\tilde{r}_t\}_{t=1}^{T}$, by each k (figures from (b) to (f)). All the graphs have exactly the same scale for the x-axis and y-axis. From Figure 5 (a) and Figure 6 (a), the observed return series exhibits volatility clustering. Also, the observed return distribution shows negative skewness with infrequent long tails. Figure 5 and Figure 6 show that an increase of k can help generate nice characteristics of infrequent spike and clear volatility clusterings like the real returns. When the number of the state is low (for example, $k = 1$), the model can not produce a large spike as the real-world experience. However, as k increases, the model

[16]See Appendix C for Newey-West heteroskedasticity and autocorrelation consistent standard errors.

becomes able to produce large but infrequent spikes as the observed return series while maintaining a relatively stable period. Also, as k increases, the model becomes capable of producing more visible volatility clustering compared to the $k = 1$ model. Observing simulated return distributions, increasing k improves the generation of return distributions.

2.4 Out-of-sample Comparison with Various k Models

2.4.1 Parameter Estimation

I now compare the out-of-sample performance by varying k. First, I do the maximum likelihood estimation of each model's parameters using only the first half of observations. The way MLE estimation was performed was exactly the same as the full sample estimation. I did robustness checks for the $k = 1$ parameters[17] and use the result for subsequent model $k = 2$. I perform this until $k = 10$. Table 8 shows the maximum likelihood estimators of the parameters over the first half of the sample data[18]. Again, when the number of state k increases, the log-likelihood increases monotonically up to $k = 8$, and it plateaus from there. The parameter variations by k are also consistent with the result in the full sample, where $\hat{\bar{\sigma}}$, $\hat{\psi}_u$, $\hat{\psi}_u$ decreases monotonically, and $\hat{a}$ and $\hat{b}$ increase monotonically by increasing k. $\hat{\bar{\delta}}$ is stable over k.

2.4.2 Point Forecasts

In this subsection, I compare how well each model performs a point forecast of volatility over a different forecasting horizon, by Mincer-Zarnowitz regressions (Mincer and Zarnowitz, 1969) over one business day, five business day (one week), ten business day (two weeks), twenty business day (one month), forty business day (two months), and sixty

[17] See Appendix F.

[18] Again, I calibrated the mean for the optimization with the mean of returns over the first half sample ($\hat{\mu} = 0.031391\%$) for all the models. The first half of the sample contains 12,448 observations.

business day (one quarter) horizons. Mincer-Zarnowitz regression can be written as

$$\sigma^2_{t,n} = \gamma_0 + \gamma_1 \mathbb{E}_{t-n}\sigma^2_{t,n} + \varepsilon_t \tag{2.26}$$

where $\sigma^2_{t,n}$ is a variance over n days from time $t-n$ and ε_t is white noise. When the model is correctly specified, the model produces perfect variance forecasts. Thus, the null hypothesis of the unbiased forecast implies $H_0 : \gamma_0 = 0 \cap \gamma_1 = 1$. I translate equation (2.26) to my out-of-sample analysis, such that

$$RV_{t,n} = \gamma_0 + \gamma_1 \mathbb{E}_{t-n} RV_{t,n} + u_t \tag{2.27}$$

where $RV_{t,n} = \sum_{s=t-n+1}^{t} r_s^2$. $RV_{t,n}$ is realized volatility over n day period from $t-n$, with $\mathbb{E}_{t-n}RV_{t,n} = B^{-1}\sum_{k=1}^{B} \hat{RV}_{t,n,k}$. $\hat{RV}_{t,n,k}$ is the forecast of realized volatility over n day period at the $t-n$ point of time from the model, and u_t is white noise ($\mathbb{E}[u_t] = 0$). The multiple-period forecast is path-dependent and requires keeping the error term in the return specification (equation (2.3)) for the forecast. Hence, the single-path variance forecast is noisy. Therefore, I generate $B = 10,000$ paths of forecast at each point of time $(t-n)$ for each forecast horizon (n = {1, 5, 10, 20, 40, 60 } days). Then, I take the average of the B paths of the forecast at each point of time to get the less noisy version of the realized volatility forecast $\mathbb{E}_{t-n}RV_{t,n}$.

If the model concerned is correctly specified, the coefficients of Mincer-Zarnowitz will be $\gamma_0 = 0$ and $\gamma_1 = 1$. Table 10 shows the result of Mincer-Zarnowitz evaluation of realized volatility point forecast. I use the latter half of the subsample to evaluate the performance of forecasting the realized volatility $RV_{t,n}$ for the dependent variable in equation (2.27). In this subsample, there are $T' = 12,448$ observations. The table shows the coefficients of Mincer-Zarnowitz regression and R^2. The figures in the parentheses are the standard error with

Heteroskedasticity and Autocorrelation Consistent (HAC) adjustment using the method of Newey and West (1987), with automatic lag parameter selection (West and Newey, 1994) as described in Appendix C.

Since only looking at the coefficient of Mincer-Zarnowitz is not fully informative and can be misleading, we must check its graphical representation. Figures 7 to Figure 12 show the result of Mincer-Zarnowitz regression. Figure 7 shows Mincer-Zarnowitz regression of one-day out-of-sample forecast, Figure 8 shows five-day forecast, Figure 9 shows ten-day forecast, Figure 10 shows twenty-day forecast, Figure 11 shows forty-day forecast, and Figure 12 shows sixty-day forecast for each $k = \{1, 2, ..., 10\}$. Each panel starts with $k = 1$ from the top left, $k = 2$ on the top right, to $k = 10$ to the bottom right. Each graph has realized volatility forecast from the model on the x-axis, and the real realized volatility observation corresponding to that forecast is on the y-axis. The scale of the x-axis and y-axis are the same within the panel. The blue diagonal line is the regression line.

As shown in the graphs, the model with lower k is only capable of producing up to a certain level of volatility (e.g., $k = 1$ on the top left, and $k = 2$ on the right next, which creates a vertical "wall" of dot points), but gradually become able to produce higher volatility as k increases (as we go toward the right one by one in the panel). This ability to produce higher volatility is clearly an improvement. Even with a low number of states with limited volatility, the slope of the regression line can be close to one, but not in granularity (like one-day forecast by $k = 1$ model in Figure 7). In such a case, the standard error is significant, and the explanatory power (R^2) is low (e.g., 0.084). If the model can produce higher volatility, it improves the R^2 of the regression.

Table 10 shows that the increase of k states improves out-of-sample point forecasts. The one-day forecast tends to be good for any k in terms of slope coefficient γ_1. However, when k is small, the model does not perform well on longer horizons. Increasing k contributes to a significant improvement on forecasting long horizons while forecasting equally well on the short horizons. From $k = 6$ to $k = 8$, the model performs equally well on short and long

horizons. The model with $k = 9$ and $k = 10$ look good on the long horizon forecast, but start to deviate on short horizon. However, the table shows that the HAC-adjusted standard errors decrease almost monotonically when k increases. This monotonic improvement is similar for the intercept of the regression coefficient γ_0. Though the intercept is almost zero for the $k = 1$ model for all the horizons, if we look carefully, the increase in k still improves. In terms of R^2, it improves gradually up to $k = \{6, 7, 8\}$ and stabilizes. This tendency of stabilizing at $k = 8$ is consistent with the MLE result and the Vuong test, where $k = 8$ yields the highest log-likelihood. The R^2 of the $k = 8$ model is almost double of R^2 of the $k = 1$ model for the shorter horizons and about 1.5 times better for longer horizons. Overall, an increase of k states improves forecasting performance.

2.4.3 Interval Forecasts

I evaluate the return distribution with the Value-at-Risk metric (VaR). The VaR at time t is defined to be the pth quantile ($p \in [0, 1]$) of the conditional return distribution over the period from time $t + 1$ to $t + n$ as

$$\hat{VaR}_{t,n}(p) \equiv \hat{F}_{t,n}(p) \tag{2.28}$$

where $\hat{F}_{t,n}$ is conditional return forecast distribution generated by the model. The accuracy of the VaR, or unconditional coverage property, is verified by recording the number of "hit" rates[19]. The "hit" function flags one when the observed return r_t exceeds the predicted $\hat{VaR}_{t,n}(p)$. The failure rate is the number of times where the "hit" function flags over the given sample period. Denoting the observed cumulative return over n days from time t as $r_{t,t+n} = \sum_{k=t}^{n} r_k$, the indicator ("hit") function which flags the breach of the VaR is defined as

[19] Kupiec (1995), Christoffersen (1998)

$$I_{t+n}(p) \equiv \begin{cases} 1 & \text{if } r_{t,t+n} < \hat{VaR}_{t,n}(p) \\ 0 & \text{if } r_{t,t+n} \geq \hat{VaR}_{t,n}(p) \end{cases} \tag{2.29}$$

Then the "failure" rate is written as

$$\hat{U}_n(p) \equiv \frac{1}{T'} \sum_{t=1}^{T'-n} I_{t+n}(p) \tag{2.30}$$

If the VaR is accurately estimated, the failure rate converges to p; $\hat{U}_n(p) \to p$. To obtain each model's conditional forecast distribution $\hat{F}_{t,n}$, at each time t, I simulate $\tilde{r}_{t,t+n}$ for $B = 10,000$ paths $\{\tilde{r}_{t,t+n}\}_{b=1}^{B}$. Then, I simply take the p^{th} quantile of $\{\tilde{r}_{t,t+n}\}_{b=1}^{B}$ at each time t, to assign it to $\hat{VaR}_{t,n}(p)$ and obtain $\{I_{t+n}(p)\}_{t=1}^{T'}$. From the construction of the hit function, the time series $\{I_{t+n}(p)\}_{t=1}^{T'}$ is highly likely to have autocorrelation. Thus, I use Newey-West Heteroskedasticity and Autocorrelation Consistent standard errors by Newey and West (1987) for the failure rate forecast[20].

Table 11, to Table 16 report the failure rate of the model with each k for n = {1, 5, 10, 20, 40, 60 } day forecasts and confidence level at lower tail and upper tail of p = { 1%, 5%, 10%, 90%, 95%, 99% }. The p-value is evaluated against a null hypothesis $H_0 : \hat{U}_n(p) = p$. If the p-value is larger, it is less probable to reject that the forecasted VaR level is indeed at the realized quantile from the observed returns. Therefore, larger p-values imply better forecasts. In these tables, the boldface numbers are statistically indifferent from the target value-at-risk level p at the 3% confidence level.

Table 11 and Table 12 show the improvement by the increase of k is mild for the short horizon. However, looking at it carefully, the failure rate on the middle interval (e.g., ten percentile or ninety percentile) becomes closer to the target percentile when k increases.

[20]Described in Appendix C

For forecasting five-day horizon, an increase of k improves forecasting the right tail (ninety-nine percentile) because the p-value is increasing. This improvement is particularly true for forecasts on more than a twenty-day horizon from Table 14 to Table 16. The p-value gets higher both on the left tail and the right tail for the longer horizons when k increases (except for the ninety-nine percentile of the sixty-day horizon forecast). This means, especially for longer horizons such as more than twenty-day forecast, the increasing k improves the model's ability to produce return distribution similar to real distribution, while maintaining a similar, or slightly higher accuracy on the shorter horizons.

2.4.4 Density Forecasts

I evaluate density forecasts by each model with the Probability Integral Transform (Diebold, et al. 1998). The probability integral transform of a sequence of generated forecasts by a model $\{f_t(r_{t,n}|\Omega_t)\}_{t=1}^{T'}$ is a cumulative density function corresponding to the density $f_t(r_{t,n})$ evaluated at r such that

$$\hat{F}_{t,n}(r) \equiv \mathbb{P}_t(r_{t,n} \leq r|\Omega_t) = \int_{-\infty}^{r} f_t(y)dy \tag{2.31}$$

If the forecast generating model is specified correctly, the random variables $U_{t,n} = F_{t,n}(y_{t,n})$ follow an uniform distribution on the interval of $[0,1]$. Thus, if the probability integral transform is closer to a uniform distribution, it can be said that the forecast is more accurate. This is an extension of the evaluation of interval forecasting, which checks every interval of the distribution. Following Diebold et al. (1998), I look at the graphical plot of $U_{t,n}$ to see how close the probability integral transform of the forecasts is to the uniform distribution.

Figure 13, 14 and 15 show the probability integral transform of forecasts generated by corresponding models. Figure 13 shows the result for the short horizon (Panel A: one day forecast and Panel B: five-day forecast), Figure 14 illustrates the middle horizon (Panel

A: ten-day forecast and Panel B: twenty-day forecast) and Figure 15 illustrates the long horizon (Panel A: forty-day forecast and Panel B: sixty-day forecast). Each panel contains ten forecasts beginning with $k = 1$ at the top left toward $k = 10$ on the bottom right. The horizontal axis corresponds to percentile from zero to one, and the vertical axis corresponds to the frequency for each percentile bucket. Figure 13, 14, and 15 have a hundred buckets for each graph. The red horizontal lines in each chart indicate the uniform distribution, and the blue lines are 5% distant from the uniform distribution line.

Overall, the probability integral transform of forecasts improves by an increase in k. The probability integral transform tends to have a butterfly shape, such that there are peaks of frequency at the left tail and at ninety percentile, and valleys of frequency at the middle range and right tail (ninety-nine percentile). These peaks gradually become lower, and valleys gradually become higher toward the uniform distribution line (red line) when k increases. The convergence to the uniform distribution is especially seen on the right tail (ninety-nine percentile). The graphical result here is consistent with the result of interval forecasts seen in Table 11, to 16, where VaR forecasts tend to be good at the left and right tails, and the forecasting ability improves by k.

The Cramer von-Mise (CVM) statistics in Table 17 also confirm this graphical interpretation. As in Appendix D, the CVM criterion measures the distances between two given distributions. Hence the smaller the CVM is, the closer the distance between the probability integral transform and the uniform distribution, and the better the forecast is. Table 17 shows that when k increases, the CVM statistics (the distance) get smaller overall. The CVM criterion improves for the forecast horizons from one day to twenty days when k becomes larger. Each model performs almost equally well for the forty-day horizon. On the other hand, the CVM criterion gradually increases for the sixty-day horizon. However, the value for the sixty-day horizon is still low compared to other forecasting horizons except for the one-day horizon, even for the higher k. Considering the short and medium improvements, the slight underperformance on the sixty-day horizon is acceptable. Together with

the graphical result in Figure 13, Figure 14, and Figure 15, the smaller CVM values in Table 17 indicate that the distribution of the out-of-sample density forecast by the proposed model outperforms other models.

2.5 Conclusion

This paper proposes an extension of the PTV-3 model to consider a higher number of different states. In this extended PTV-M model, the price threshold has an excellent recursive structure, easy to expand to multiple states. This extension in the number of states in the PTV model overcomes the weakness of the MS model, which cannot easily produce various return distribution groups contemporaneously. The model is parsimonious and flexible in accepting any number of states by changing parameter k. This feature is convenient because typical MS requires an exponentially expanding number of parameters when increasing the number of states. From the empirical analysis, the model with higher k can better predict the return distribution of the US equity index. I compare the out-of-sample point forecast, the out-of-sample interval forecast, and the out-of-sample density forecast of the models with $k = \{1, 2, ..., 10\}$ and find that the model with higher k predicts more accurate predictions than lower k.

This model's practical application would be to use it as an indicator of dynamic asset allocation based on the filtered regime of the capital market and as a more accurate quantile-based risk management monitoring. One interesting work could be applying this PTV-M model to different asset classes, such as fixed incomes or commodities. Another direction is to study the implications of endogenous transition probabilities for the appearance of long memory in volatility. The realized asset return volatility is well known to have long memory (Andersen et al. 2003), but the origin of this long memory is not yet well-understood. On the other hand, the long memory is well known to be closely related to Markov Switching (Diebold and Inoue, 2001), in the classic CTP case. It will be interesting to explore the

further implication of the relation between the long memory and Markov Switching in the TVTP setting.

Bibliography

[1] Andersen, T.G., Bollerslev, T., 1998 *Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts*. International Economic Review, Vol. 39, No. 4, Pages 885-905.

[2] Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003 *Modeling and Forecasting Realized Volatility*. Econometrica, Vol. 71, No. 4, No. 2, Pages 579-625.

[3] Andersen, T.G., Bollerslev, T., Meddahi, N. 2005 *Correcting The Errors: Volatility Forecast Evaluation Using High-Frequency Data and Realized Volatility*. Econometrica, Vol. 73, No. 1, Pages 279-296.

[4] Bakke, T., Whited, T.M., 2010 *Which Firms Follow the Market? An Analysis of Corporate Investment Decisions*. Review of Financial Studies, Vol. 23, No. 5, Pages 1941-1980.

[5] Bazzi, M., Blasques, F., Koopman, S.J., Lucas, A., 2016 *Time Varying Transition Probabilities for Markov Regime Switching Models*. Journal of Time Series Analysis, Vol 38, No. 3, Pages 458-478.

[6] Bekaert, G., Wu, G., 2000 *Asymmetric Volatility and Risk in Equity Markets*. Review of Financial Studies, Vol. 13, No. 1, Pages 1-42.

[7] Black, F., 1976. *Studies of Stock Price Volatility Changes*. Proceeding of the American Statistical Association, Business and Economical Statistics Section, Pages 177-181.

[8] Black, F., Scholes, M., 1973. *The Pricing of Options and Corporate Liabilities*. Journal of Political Economy, Vol. 81, No. 3, Pages 637-654.

[9] Bollerslev, T., 1986. *Generalized Autoregressive Conditional Heteroskedasticity*. Journal of Econometrics, Vol. 31, No. 3, Pages 307-327.

[10] Bond, P., Edmans, A., Goldstein, I.,2012 *The Real Effects of Financial Markets*. Annual Review of Financial Economics, Vol. 4, No. 1, 339-360.

[11] Brock, W., Lakonishok, J., LeBaron, B., 1992. *Simple Technical Trading Rules and the Stochastic Properties of Stock Returns*. Journal of Finance, Vol. 47, No. 5, Pages 1731-1764.

[12] Broyden, C.G., 1970 *The convergence of a class of double-rank minimization algorithms*. IMA Journal of Applied Mathematics, Vol. 6, 76-90.

[13] Byrd, R.H., Lu, P., Nocedal, J., Zhu, C.,1995 *A Limited Memory Algorithm for Bound Constrained Optimization*. SIAM Journal on Scientific Computing, Vol. 16, 1190-1208.

[14] Calvet, L.E., Fisher, A.J., 2001. *Forecasting Multifractal Volatility*. Journal of Econometrics, Vol. 105, No. 1, Pages 27-58.

[15] Calvet, L.E., Fisher, A.J., 2004. *How To Forecast Long-run Volatility: Regime-switching And The Estimation of Multifractal Processes*. Journal of Financial Econometrics, Vol. 2, No. 1, Pages 49-83.

[16] Calvet, L.E., Fisher, A.J., 2007. *Multifrequency News and Stock Returns*. Journal of Financial Economics, Vol. 86, No. 1, Pages 178-212.

[17] Chen, Q., Goldstein, I., Jiang, W.,2007. *Price Informativeness and Investment Sensitivity to The Stock Price*. Review of Financial Studies, Vol. 20, No. 3, Pages 619-650.

[18] Christie, A.A., 1982. *The Stochastic Behavior of Common Stock Variances - Value, Leverage and Interest Rate Effects*. Journal of Financial Economics, Vol. 10, No. 4, Pages 407-432.

[19] Christoffersen, P., 1998. *Evaluating Interval Forecasts*. International Economic Review, Vol. 39, No. 4, Pages 841-862.

[20] Cosslett, S.R., Lee, L.F., 1985. *Serial Correlation in Latent Discrete Variable Models*. Journal of Econometrics, Vol. 27, No. 1, Pages 79-97.

[21] Creal, D., Koopman, S.J., Lucas, A., 2013. *Generalized Autoregressive Score Models With Applications*. Journal of Applied Econometrics, Vol. 28, No. 5, Pages 777-795.

[22] Diebold, F.X., Gunther, T.A., Tay, A.S., 1998. *Evaluating Density Forecasts with Applications to Financial Risk Management*. International Economic Review, Vol. 39, No. 4, Pages 863-883.

[23] Diebold, F.X., Inoue, A., 2001. *Long Memory and Regime Switching*. Journal of Econometrics, Vol. 105, No. 1, Pages 131-159.

[24] Diebold, F.X., Lee, J., Weinbach, G.C., 1994. *Regime Switching with Time-Varying Transition Probabilities*. in C. Hargreaves (ed.), Nonstationary Time Series Analysis and Cointegration, Pages 283-302. Oxford University Press.

[25] Dow, J., Goldstein, I., Guembel, A., 2017. *Incentives for Information Production in Markets Where Prices Affect Real Investment*. Journal of the European Economic Association, Vol. 15, No. 4, Pages 877–909.

[26] Easley, D., O'Hara, M., 1987. *Price, Trade Size, and Information in Securities Markets*. Journal of Financial Economics, Vol. 19, No. 1, Pages 69-90.

[27] Easley, D., O'Hara, M., 1991. *Order Form and Information in Securities Markets*. Journal of Finance, Vol. 46, No. 3, Pages 905-927.

[28] Easley, D., O'Hara, M., 1992. *Time and The Process of Security Price Adjustment*. Journal of Finance, Vol. 47, No. 2, Pages 577-605.

[29] Engel, C., Hakkio, C.S., 1996. *Distribution of Exchange Rates in the EMS*. International Journal of Finance & Economics, Vol. 1, No. 1, Pages 55-67.

[30] Engle, R.F., 1982. *Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation*. Econometrica, Vol. 50, No. 4, Pages 987-1007.

[31] Fama, E.F., 1970. *Efficient Capital Markets: A Review of Theory and Empirical Work*. Journal of Finance, Vol. 25, No. 2, Pages 383-417.

[32] Filardo, A.J., 1994. *Business-cycle Phases and Their Transitional Dynamics*. Journal of Business & Economic Statistics, Vol. 12, No. 3, Pages 299-308.

[33] Filardo, A.J., 1998. *Choosing Information Variables For Transition Probabilities in A Time-Varying Transition Probability Markov Switching Model*. Research Working Paper 98-09, Federal Reserve Bank of Kansas City.

[34] Fletcher, R., 1970. *A New Approach to Variable Metric Algorithms*. Computer Journal, Vol. 13, No. 3, Pages 317-322.

[35] Fong, W.M., See, K.H., 2002. *A Markov Switching Model of The Conditional Volatility of Crude Oil Futures Prices*. Energy Economics, Vol. 24, No. 1, Pages 71-95.

[36] French, K.R., Schwert, G.W., Stambaugh, R., 1987. *Expected Stock Returns and Volatility*. Journal of Financial Economics, Vol. 19, No. 1, Pages 3-29.

[37] Fuller, W.A.,1996 *Introduction to Statistical Times Series*. John Wiley and Sons

[38] Glosten, L.R., Milgrom, P.R., 1985. *Bid, Ask, and Transaction Prices In A Specialist Market With Heterogeneously Informed Traders*. Journal of Financial Economics, Vol. 14, No. 1, Pages 71-100.

[39] Goldfarb, D., 1970. *A Family of Variable Metric Updates Derived by Variational Means*. Mathematics of Computation, Vol. 24, No. 109, Pages 23-26.

[40] Goldfeld, S.M., Quandt, R.E., 1973. *A Markov Switching Model For Switching Regressions*. Journal of Econometrics, Vol. 1, No. 1, Pages 3-15.

[41] Gray, S.F., 1996. *Modeling The Conditional Distribution of Interest Rates as A Regime-switching Process*. Journal of Financial Economics, Vol. 42, No. 1, Pages 27-62.

[42] Grossman, S.J., 1976. *On The Efficiency of Competitive Stock Markets Where Traders Have Diverse Information*. Journal of Financial, Vol. 31, No. 2, Pages 573-585.

[43] Grossman, S.J., Stiglitz J.E. 1980. *On The Impossibility of Informationally Efficient Markets*. American Economic Review, Vol. 70, No. 3, Pages 393-408.

[44] Hamilton, J., 1989. *A New Approach to The Economic Analysis of Nonstationary Time Series and The Business Cycle*. Econometrica, Vol. 57, No. 2, Pages 357-384.

[45] Hamilton, J., Raj, B., 2002. *New Directions in Business Cycle Research and Financial Analysis*. Empirical Economics, Vol. 27, No. 2, Pages 149-162.

[46] Hayek, F.A., 1945. *The Use of Knowledge in Society*. The American Economic Review, Vol. 35, No. 4, Pages 519-530.

[47] Kato, N., 2021. *Forecasting Volatility with Price Threshold*. Working Paper.

[48] Kupiec, P., 1995. *Techniques for Verifying The Accuracy of Risk Management Models*. Journal of Derivatives, Vol. 3, Pages 73-84.

[49] Kyle, A.S., 1985. *Continuous Auctions and Insider Trading*. Econometrica, Vol. 53, No. 6, Pages 1315-1335.

[50] Ljung, G.M., Box, G.E.P., 1978. *On a Measure of Lack of Fit in Time Series Models*. Biometrika, Vol. 65, No. 2, Pages 297-303.

[51] Lo, A., Mamaysky, H., Wang, J., 2000. *Fundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation*. Journal of Finance, Vol. 55, No. 4, Pages 1705-1765.

[52] MacRae, E.C., 1977. *Estimation of Time-Varying Markov Processes with Aggregate Data*. Econometrica, Vol. 45, No. 1, Pages 183-198.

[53] Menkhoff, L., Taylor, M., 2007. *The obstinate Passion of Foreign Exchange Professionals: Technical Analysis*. Journal of Economic Literature, Vol. 45, Pages 936-972.

[54] Merton, R.C., 1973. *On the Pricing of Corporate Debt: The Risk Structure of Interest Rates*. Journal of Finance, Vol. 29, No. 2, Pages 449-470.

[55] Milgrom, P., 1979. *A Convergence Theorem for Competitive Bidding and Differential Information*. Econometrica, Vol. 47, No. 3, Pages 921-943.

[56] Mincer, J., Zarnowitz, V., 1969. *The Evaluation of Economic Forecasts*. In Economic Forecasts and Expectations, National Bureau of Economic Research.

[57] Morck, R., Shleifer, A., Vishny, R.W., Shapiro, M., Poterba, J.M., 1990. *The Stock Market and Investment: Is the Market a Sideshow?*. Brookings Papers on Economic Activity, Vol. 1990, No. 2, Pages 157-215.

[58] Neely, C.J., Rapach, D.E., Tu, J., Zhou, G., 2013. *Forecasting The Equity Risk Premium: The Role of Technical Indicators*. Management Science, Vol. 60, No. 7, Pages 1772-1791

[59] Newey, W., West, K., 1987. *A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix*. Econometrica, Vol. 55, No. 3, Pages 703-708.

[60] Palfrey, T., 1985. *Uncertainty Resolution Private Information Aggregation, and The Cournot Competitive Limit*. Review of Economic Studies, Vol. 52, No. 1, Pages 69-83.

[61] Peria, M.S.M., 2002. *A Regime-Switching Approach to the Study of Speculative Attacks: A Focus on EMS Crises* Empirical Economics, Vol. 27, No. 2, Pages 299-334.

[62] Shanno, D.F., 1970. *Conditioning of quasi-Newton methods for function minimization*. Mathematics of Computation, Vol. 24, No. 111, Pages 647-656.

[63] Van Norden, S., Schaller, H., 1997. *Regime Switching Stock Market Returns*. Applied Financial Economics, Vol. 7, No. 2, Pages 177-191.

[64] Vuong, Q., 1989. *Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses*. Econometrica, Vol. 57, No. 2, Pages 307-333.

[65] Wang, X., Shrestha, K., Sun, Q. 2019. *Forecasting Realised Volatility: A Markov Switching Approach with Time-varying Transition Probabilities*. Accounting & Finance, Vol. 59, No. 2, Pages 1947-1975.

[66] West, K., Newey, W., 1994. *Automatic Lag Selection in Covariance Matrix Estimation*. Review of Economic Studies, Vol. 61, No. 4, Pages 631-653.

[67] Wilson, R., 1974. *Informational Economies of Scale*. The Bell Journal of Economics, Vol. 6, No. 1, Pages 184-195.

[68] Yuan, C., 2011. *The Exchange Rate and Macroeconomic Determinants: Time-varying transitional Dynamics*. The North American Journal of Economics and Finance, Vol. 22, No. 2, Pages 197-220.

[69] Zhou, G., Zhu, Y., 2013. *An Equilibrium Model of Moving-average Predictability and Time-series Momentum*. SSRN.

Appendices

A. Proofs of equations in Section 1.2.3 and 2.2.2

Assumption 1. *The price threshold K at time t is characterized by the exponential weighted moving average, $EWMA$ at time t, with a positive constant $c \in \mathbb{R}_+$, such that*

$$K_t = c \times EWMA_t$$

Proposition 1. *If Assumption 1 holds, the price threshold K at time t is characterized by the exponential weighted moving average, $EWMA$ at time $t-1$, with a constant $c' \in \mathbb{R}$, such that*

$$K_t = c' \times EWMA_{t-1}$$

Proof. An asset price at time t goes beyond this price threshold is written as

$$\begin{aligned} P_t > K_t &= c \times EWMA_t \\ &= c\{\delta P_t + (1-\delta) EWMA_{t-1}\} \\ &= c\delta P_t + c(1-\delta) EWMA_{t-1} \\ \Leftrightarrow P_t &> \frac{c(1-\delta)}{1-c\delta} EWMA_{t-1} \end{aligned}$$

Therefore, defining as followings, K_t is characterized by the exponential weighted moving average at time $t-1$.

$$c' \equiv \frac{c(1-\delta)}{1-c\delta}$$

When $c \in (0, 1/\delta)$, c' becomes also a positive constant. The same discussion is also applied when an asset price at time t goes below the price threshold ($P_t < K_t$). □

B. Proofs of equations in Section 1.3.1 and 2.3.1

Proposition 2. *The filtered probabilities of the N latent states at time t, conditional on the observed returns satisfies*

$$\hat{\Pi}_t = \frac{\omega_t * (\hat{\Pi}_{t-1} A_t)}{[\omega_t * (\hat{\Pi}_{t-1} A_t)]\iota} \in \mathbb{R}^N$$

where $*$ *denotes element-by-element multiplication and* $\iota = (1 \ldots 1)' \in \mathbb{R}^N$

Proof. From Bayes' formula

$$\begin{aligned}
\hat{\Pi}_t^j &= \mathbb{P}(s_t = S_j | r_t, r_{1:t-1}; \mathscr{I}_{1:t-1}) \\
&= \frac{f(r_t | s_t = S_j, r_{1:t-1}; \mathscr{I}_{1:t-1}) \mathbb{P}(s_t = S_j | r_{1:t-1}; \mathscr{I}_{1:t-1})}{f(r_t | r_{1:t-1}; \mathscr{I}_{1:t-1})} \\
&= \frac{\omega_{j,t} \mathbb{P}(s_t = S_j | r_{1:t-1}; \mathscr{I}_{1:t-1})}{f(r_t | r_{1:t-1}; \mathscr{I}_{1:t-1})}
\end{aligned}$$

Because, the transition probability is a Markov chain, such that

$$\mathbb{P}(s_t = S_j | s_{t-1} = S_i, r_{1:t-1}; \mathscr{I}_{1:t-1}) = \mathbb{P}(s_t = S_j | s_{t-1} = S_i; \mathscr{I}_{1:t-1})$$

Hence,

$$\begin{aligned}
&\mathbb{P}(s_t = S_j | r_{1:t-1}; \mathscr{I}_{1:t-1}) \\
&= \sum_{i=1}^{N} \mathbb{P}(s_t = S_j | s_{t-1} = S_i, r_{1:t-1}; \mathscr{I}_{1:t-1}) \mathbb{P}(s_{t-1} = S_i | r_{1:t-1}; \mathscr{I}_{1:t-1}) \\
&= \sum_{i=1}^{N} \mathbb{P}(s_t = S_j | s_{t-1} = S_i; \mathscr{I}_{1:t-1}) \mathbb{P}(s_{t-1} = S_i | r_{1:t-1}; \mathscr{I}_{1:t-1}) \\
&= \sum_{i=1}^{N} \gamma_{ij,t} \hat{\Pi}_{t-1}^i \\
&= (\hat{\Pi}_{t-1} A_t)_j
\end{aligned}$$

Also, since $\hat{\Pi}_t \iota = \iota$

$$f(r_t | r_{1:t-1}; \mathscr{I}_{1:t-1}) = [\omega_t * (\hat{\Pi}_{t-1} A_t)]\iota$$

Therefore

$$\hat{\Pi}_t^j = \frac{\omega_{j,t}(\hat{\Pi}_{t-1}A_t)_j}{[\omega_t * (\hat{\Pi}_{t-1}A_t)]\iota}$$

□

Proposition 3. *The analytical expression for the likelihood function of the filtered probabilities of the latent states is*

$$\ln L(\theta; r_{1:T}) = \sum_{t=1}^{T} \ln[\omega_t \cdot (\hat{\Pi}_{t-1}A_t)]$$

Proof.

$$\begin{aligned}
\ln L(\theta; r_{1:T}) &= \sum_{t=1}^{T} \ln f(r_t | r_{1:t-1}) \\
&= \sum_{t=1}^{T} \ln[\omega_t * (\hat{\Pi}_{t-1}A_t)]\iota \\
&= \sum_{t=1}^{T} \ln \sum_{j=1}^{N} [\omega_{j,t} * (\hat{\Pi}_{t-1}A_t)_j] \\
&= \sum_{t=1}^{T} \ln[\omega_t \cdot (\hat{\Pi}_{t-1}A_t)]
\end{aligned}$$

□

C. Newey-West Heteroskedasticity and Autocorrelation Consistent Standard Errors

The Newey-West Heteroskedasticity and Autocorrelation Consistent covariance matrix $\hat{\Omega}_n$ (Newey and West, 1987) is described as

$$\hat{\Omega}_n = \hat{\Gamma}_{0,n} + \sum_{v=1}^{q} w(v,q)\{\hat{\Gamma}_{v,n} + \hat{\Gamma}'_{v,n}\}$$

where $\hat{\Gamma}_v$ is a sample autocovariance and $w(v,q) = 1 - \{v/(q+1)\}$ is a Bartlett kernel, which makes $\hat{\Omega}_n$ positive semi-definite. q is a truncation parameter, which can be automatically chosen by West and Newey's method (1994) such that

$$q = 1.447 \cdot \left[\frac{w' \hat{S} w}{w' \hat{S}^0 w} T \right]^{\frac{1}{3}}$$

where w is $g \times 1$ weighting vector $w = (0\ 1\ 1\ ...\ 1)'$, g is number of estimators to test, and $\hat{S} = \sum_{j=-l}^{l} |j|^k \hat{\Gamma}_j$ where $k = \{0,1\}$, and $l = 4(T/100)^{2/9}$.

For the Mincer-Zarnowitz regression, $\hat{\Gamma}_v$ is written as such that

$$\hat{\Gamma}_{v,n} = \frac{1}{T} \sum_{t=v+1}^{T} \left\{ RV_{t,n} - \frac{1}{T'} \sum_{t=T-T'}^{T} RV_{t,n} \right\} \hat{u}_t \hat{u}'_{t-v} \left\{ RV_{t-v,n} - \frac{1}{T'} \sum_{t=T-T'}^{T} RV_{t,n} \right\}'$$

Then the Wald test statistic, which follows $\mathscr{X}^2$ distribution with g degrees of freedom, is calculated by

$$W_n = \{\theta - \theta_0\} \hat{\Omega}_n^{-1} \{\theta - \theta_0\}' \sim \mathscr{X}^2(g)$$

where $\theta = \{\gamma_0, \gamma_1\}$ and $\theta_0 = \{0,1\}$.

For the failure rate forecast, $\hat{\Gamma}_v$ is written as such that

$$\hat{\Gamma}_{v,n}(p) = \frac{1}{T} \sum_{t=v+1}^{T} \{I_{t,n}(p) - \bar{I}_{T,n}(p)\} \{I_{t-v,n}(p) - \bar{I}_{T,n}(p)\}'$$

Then the Wald test statistic is

$$W_n(p) = \{\hat{U}_n(p) - p\}\hat{\Omega}_n^{-1}\{\hat{U}_n(p) - p\}' \sim \mathscr{X}^2(g)$$

D. Cramer von-Mises Criterion

Cramer von-Mises (CVM) statistics measures the distance between two distributions such that

$$CVM_n = T' \int_0^1 [y - \hat{F}_{U,n}(y)]^2 dy$$

where $\hat{F}_{U,n}$ is the empirical distribution of the transforms $U_{t,n}$. Hence CVM statistics measures the norm-2 distance between the observation y (in this paper, the out-of-sample forecast by the model) and the empirical distribution (realized return distribution). Under the null hypothesis, the distribution of y is $\hat{F}_{U,n}$. The smaller the CVM statistic becomes, the less likely the null hypothesis is rejected.

E. Robustness check of MLE estimators for the three-state model

MLE estimators can be different by the initial value for the optimization of maximizing likelihood. The optimization process is numerical search, and the result could be trapped by a local maxima. Thus, I performed a robustness check on the MLE estimators by changing initial value of each parameters. Table 18, 19, 20, 21, and 22 show the MLE estimators by changing initial value of each parameter. The default initial parameters are set such that: $\sigma_s = 0.005, \sigma_m = 0.010, \sigma_v = 0.020, \psi_u = 0.02, \psi_l = 0.02, \delta = 0.60, \mu = 0.000303$.

Each table show the MLE estimation result by changing one parameter. Table 18, 19, 20 show the MLE estimators when I change σ_s, σ_m, and σ_v from 0.0025 to 0.0250 with step by 0.0025 accordingly. Table 21 shows the MLE estimators by changing initial value of ψ_u and ψ_l from 0.005 to 0.050 with step by 0.005. And lastly, Table 22 shows the MLE estimators by changing initial value of δ from 0.10 to 1.00 with step by 0.1. These tables show that the initial value affects the result significantly. However, in any of these cases, I get the maximum likelihood (around 84,510) when the estimated parameters are at the level of: $\hat{\sigma}_s = 0.0054, \hat{\sigma}_m = 0.0108, \hat{\sigma}_v = 0.0270, \hat{\psi}_u = 0.0210, \hat{\psi}_l = 0.0240, \hat{\delta} = 0.650, \hat{\mu} = 0.000303$. Hence, I take these value for the final MLE estimators.

Similarly, Table 23, 24, 25, 26, and 27 shows the same analysis for the first half sample of the observation. The MLE estimators are robust for this period.

F. Robustness check of MLE estimators for the multi-state model

MLE estimators can be different by the initial value for the optimization of maximizing likelihood. The optimization process is numerical search, and the result could be trapped by a local maxima. Thus, I performed a robustness check on the MLE estimators with $k = 1$ by changing initial value of each parameters. Table 28, 29, 30, and 31 show the MLE estimators by changing initial value of each parameter. The default initial parameters are set such that: $\bar{\sigma} = 0.010, a = 0.50, b = 0.40, \psi_u = 0.02, \psi_l = 0.02, \delta = 0.60, \hat{\mu} = 0.000303$. Each table shows the MLE estimation result by changing one parameter. Table 28 shows the MLE estimators when I change $\bar{\sigma}$ from 0.0025 to 0.0250, step by 0.0025 accordingly. Table 29 shows the MLE estimators by changing the initial value of a and b from 0.10 to 1.00 with step by 0.1. Table 30 shows the MLE estimators by changing the initial value of ψ_u and ψ_l from 0.005 to 0.050 with step by 0.005. And lastly, Table 31 shows the MLE estimators by changing the initial value of δ from 0.10 to 1.00 with step by 0.1.

These tables show that the final MLE estimators converge to very similar values with a very similar log-likelihood level in any of these cases. I get stable estimated parameters, which are at the level of: $\hat{\tilde{\sigma}} = 0.0106, \hat{a} = 0.48, \hat{b} = 0.40, \hat{\psi}_u = 0.0210, \hat{\psi}_l = 0.0240, \hat{\delta} = 0.600, \hat{\mu} = 0.000303$. Hence the MLE estimators are consistent.

Similarly, Table 32, 33, 34, and 35 shows the same analysis for the first half sample of the observation. The MLE estimators are robust for this period.

Tables and Figures

Table 1: Maximum Likelihood Estimation Over Full Sample (Three-state Model)

This table displays the model parameters from the maximum likelihood estimation over $T = 24,896$ in-sample data. For the optimization, $\mu = 0.000303$ (long-run mean of return) is calibrated for all the models. To make the Hessian invertible during the optimization, I also put optimization constraints of the form as following: $\sigma_s \in (0.000, 0.100)$, $\sigma_m \in (0.000, 0.100)$, $\sigma_v \in (0.000, 0.100)$, $\psi_u \in (0.000, 0.100)$, $\psi_l \in (0.000, 0.100)$, $\delta \in (0.000, 1.000)$, $\gamma_{ij} \in (0.000, 0.500)(i, j \in \{s, m, v\})$, $\omega \in (0.000, 1.000)$, $\alpha \in (0.000, 1.000)$, $\beta \in (0.000, 1.000)$. Therefore, I use the method from Byrd et al.(1995) for the optimization method, which can perform a constrained optimization. BIC is the Bayesian Information Criterion, which adjusts the likelihood by the number of parameters. Small BIC means the adjusted likelihood is higher. BIC is given by $BIC = T^{-1}(-2\ln L + NP \cdot \ln T)$, where NP is the number of free parameters in the specification.

A. TVTP MS

$\hat{\sigma}_s$	$\hat{\sigma}_m$	$\hat{\sigma}_v$	$\hat{\psi}_u$	$\hat{\psi}_l$	$\hat{\delta}$
0.005291	0.010577	0.026725	0.020899	0.023271	0.648252
(0.000054)	(0.000156)	(0.000518)	(0.000487)	(0.000542)	(0.019568)
$\hat{\mu}$				$\ln L$	BIC
0.000303				84,513.3	-13.58

B. CTP MS

$\hat{\sigma}_s$	$\hat{\sigma}_m$	$\hat{\sigma}_v$	$\hat{\gamma}_{sm}$	$\hat{\gamma}_{sv}$	$\hat{\gamma}_{ms}$
0.005332	0.010813	0.026800	0.014534	0.001149	0.030144
(0.000059)	(0.000178)	(0.000523)	(0.001530)	(0.000561)	(0.003082)
$\hat{\gamma}_{mv}$	$\hat{\gamma}_{vs}$	$\hat{\gamma}_{vm}$	$\hat{\mu}$	$\ln L$	BIC
0.005223	0.004995	0.030467	0.000303	84,419.4	-13.56
(0.000959)	(0.006970)	(0.004870)			

C. GARCH(1,1)

$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\mu}$	$\ln L$	BIC
0.001164	0.105696	0.883252	0.000303	84,062.6	-13.50
(0.000044)	(0.004198)	(0.004454)			

Table 2: Maximum Likelihood Estimation Over The First Half Sample (Three-state Model)

This table displays the model parameters from the maximum likelihood estimation over the first half of the sample, $T = 12,448$. I use these parameters for the out-of-sample analysis. For the optimization, $\mu = 0.000314$ (long-run mean of return) is calibrated for all the models. To make the Hessian invertible during the optimization, I also put optimization constraints of the form as following: $\sigma_s \in (0.000, 0.100)$, $\sigma_m \in (0.000, 0.100)$, $\sigma_v \in (0.000, 0.100)$, $\psi_u \in (0.000, 0.100)$, $\psi_l \in (0.000, 0.100)$, $\delta \in (0.000, 1.000)$, $\gamma_{ij} \in (0.000, 0.500)(i, j \in \{s, m, v\})$, $\omega \in (0.000, 1.000)$, $\alpha \in (0.000, 1.000)$, $\beta \in (0.000, 1.000)$. Therefore, I use the method of Byrd et al. (1995) for the optimization method, which can perform a constrained optimization. BIC is the Bayesian Information Criterion, which adjusts the likelihood by the number of parameters. Small BIC means the adjusted likelihood is higher. BIC is given by $BIC = T^{-1}(-2\ln L + NP \cdot \ln T)$, where NP is the number of free parameters in the specification.

A. TVTP MS

$\hat{\sigma}_s$	$\hat{\sigma}_m$	$\hat{\sigma}_v$	$\hat{\psi}_u$	$\hat{\psi}_l$	$\hat{\delta}$
0.004700	0.009672	0.025814	0.017603	0.019639	0.648060
(0.000069)	(0.000218)	(0.000579)	(0.000650)	(0.000700)	(0.024896)
$\hat{\mu}$				$\ln L$	BIC
0.000314				43,042.5	-13.83

B. CTP MS

$\hat{\sigma}_s$	$\hat{\sigma}_m$	$\hat{\sigma}_v$	$\hat{\gamma}_{sm}$	$\hat{\gamma}_{sv}$	$\hat{\gamma}_{ms}$
0.004595	0.009520	0.025986	0.026153	0.000120	0.042408
(0.000084)	(0.000246)	(0.000636)	(0.003689)	(0.001077)	(0.006093)
$\hat{\gamma}_{mv}$	$\hat{\gamma}_{vs}$	$\hat{\gamma}_{vm}$	$\hat{\mu}$	$\ln L$	BIC
0.012589	0.000001	0.037275	0.000314	42,997.6	-13.81
(0.002294)	(0.001473)	(0.006484)			

C. GARCH(1,1)

$\hat{\omega}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\mu}$	$\ln L$	BIC
0.001113	0.117550	0.874175	0.000314	42,654.1	-13.70
(0.000056)	(0.006482)	(0.006485)			

Table 3: Mincer-Zarnowitz Regression of Point Forecast (Three-state Model)

This table displays the coefficient of Mincer-Zarnowitz evaluation of the realized volatility forecast $RV_{t,n} = \gamma_0 + \gamma_1 \mathbb{E}_{t-n} RV_{t,n} + u_t$. The regression is performed over $T' = 12,448$ out-of-sample observations on forecasts produced by each model. If the realized volatility is accurately predicted, the intercept γ_0 should be close to zero, and the slope γ_1 should be close to one. The standard errors in the table are HAC adjusted (Newey and West, 1987). The table also shows the p-value of single-parameter Wald statistic, which is performed for a null hypothesis $H_0 : \gamma_0 = 0$, and $H_0 : \gamma_1 = 1$ correspondingly. The closer the p-value becomes to one, the less probable the null hypothesis is rejected. Therefore, the closer p-value to one implies better forecast ability.

	One Day		Five Days		Ten Days	
	γ_0	γ_1	γ_0	γ_1	γ_0	γ_1
TVTP MS	0.0000	0.9237	0.0000	0.9432	-0.0001	0.9599
	(0.000)	(0.186)	(0.000)	(0.195)	(0.000)	(0.238)
p-value	85.34%	68.09%	61.41%	77.12%	53.81%	86.65%
CTP MS	0.0000	0.9645	0.0000	0.9752	-0.0001	0.9940
	(0.000)	(0.198)	(0.000)	(0.200)	(0.000)	(0.251)
p-value	96.72%	85.79%	83.84%	90.10%	75.43%	98.09%
GARCH(1,1)	0.0000	0.8224	0.0001	0.7556	0.0004	0.6704
	(0.000)	(0.107)	(0.000)	(0.118)	(0.000)	(0.142)
p-value	9.16%	9.65%	3.11%	3.86%	1.46%	2.01%

	Twenty Days		Forty Days		Sixty Days	
	γ_0	γ_1	γ_0	γ_1	γ_0	γ_1
TVTP MS	-0.0004	0.9832	-0.0013	1.0752	-0.0027	1.1729
	(0.001)	(0.288)	(0.001)	(0.267)	(0.002)	(0.273)
p-value	46.73%	95.35%	20.18%	77.85%	9.62%	52.58%
CTP MS	-0.0002	1.0280	-0.0009	1.1119	-0.0021	1.2092
	(0.000)	(0.301)	(0.001)	(0.292)	(0.002)	(0.282)
p-value	63.45%	92.51%	36.57%	70.18%	18.13%	45.87%
GARCH(1,1)	0.0009	0.5664	0.0023	0.4726	0.0038	0.4189
	(0.000)	(0.167)	(0.001)	(0.178)	(0.002)	(0.153)
p-value	0.62%	0.95%	1.67%	0.31%	3.72%	0.01%

Table 4: Failure Rate of Value-at-Risk Forecast (Three-state Model: Short Day Horizons)

This table displays the frequency of returns that go further down than the VaR forecasted by the corresponding model (failure rate). The failure rate is calculated using $T' = 12,448$ out-of-sample observations. If the VaR is accurately predicted, the failure rate should be close to the target percentile, p. Boldface numbers are the failure rate, which are statistically indifferent from the target p at the 3% confidence level. The standard errors in the table are HAC adjusted (Newey and West, 1987).

Quantile	1%	5%	10%	90%	95%	99%
			A. One Day Horizon			
TVTP MS	1.49%	6.27%	11.11%	89.33%	**94.87%**	99.26%
	(0.11%)	(0.21%)	(0.28%)	(0.26%)	(0.19%)	(0.08%)
p-value	0.00%	0.00%	0.00%	0.55%	25.83%	0.04%
CTP MS	1.61%	6.33%	11.28%	89.25%	**94.81%**	99.35%
	(0.11%)	(0.23%)	(0.30%)	(0.25%)	(0.18%)	(0.07%)
p-value	0.00%	0.00%	0.00%	0.21%	13.14%	0.00%
GARCH(1,1)	1.94%	5.52%	**9.94%**	91.27%	95.59%	**99.11%**
	(0.13%)	(0.22%)	(0.29%)	(0.21%)	(0.16%)	(0.08%)
p-value	0.00%	0.84%	42.21%	0.00%	0.01%	8.87%
			B. Five Day Horizon			
TVTP MS	1.53%	7.22%	13.26%	**89.40%**	**95.51%**	99.34%
	(0.16%)	(0.38%)	(0.52%)	(0.47%)	(0.32%)	(0.11%)
p-value	0.11%	0.00%	0.00%	10.40%	5.29%	0.10%
CTP MS	1.76%	7.58%	13.37%	89.03%	**95.52%**	99.44%
	(0.20%)	(0.43%)	(0.57%)	(0.44%)	(0.30%)	(0.10%)
p-value	0.02%	0.00%	0.00%	1.53%	5.51%	0.00%
GARCH(1,1)	2.15%	7.30%	12.28%	**90.62%**	95.88%	99.35%
	(0.22%)	(0.42%)	(0.56%)	(0.40%)	(0.29%)	(0.12%)
p-value	0.00%	0.00%	0.00%	6.17%	0.10%	0.08%
			C. Ten Day Horizon			
TVTP MS	**1.04%**	6.26%	12.24%	**90.69%**	96.02%	99.44%
	(0.18%)	(0.50%)	(0.70%)	(0.61%)	(0.43%)	(0.14%)
p-value	41.16%	0.57%	0.07%	13.22%	0.86%	0.09%
CTP MS	1.53%	7.04%	12.68%	**90.50%**	96.13%	99.55%
	(0.24%)	(0.56%)	(0.77%)	(0.59%)	(0.39%)	(0.13%)
p-value	0.84%	0.02%	0.03%	19.40%	0.19%	0.00%
GARCH(1,1)	2.44%	7.02%	11.96%	91.25%	96.45%	99.36%
	(0.32%)	(0.59%)	(0.76%)	(0.56%)	(0.38%)	(0.15%)
p-value	0.00%	0.03%	0.49%	1.22%	0.01%	0.88%

Table 5: Failure Rate of Value-at-Risk Forecast (Three-state Model: Long Day Horizons)

This table displays the frequency of returns that go further down than the VaR forecasted by the corresponding model (failure rate). The failure rate is calculated using $T' = 12,448$ out-of-sample observations. If the VaR is accurately predicted, the failure rate should be close to the target percentile, p. Boldface numbers are the failure rate, which are statistically indifferent from the target p at the 3% confidence level. The standard errors in the table are HAC adjusted (Newey and West, 1987).

Quantile	1%	5%	10%	90%	95%	99%
			A. Twenty Day Horizon			
TVTP MS	0.60%	**5.60%**	12.09%	**90.90%**	96.30%	99.64%
	(0.20%)	(0.61%)	(0.95%)	(0.85%)	(0.56%)	(0.11%)
p-value	2.01%	16.10%	1.37%	14.47%	1.08%	0.00%
CTP MS	**1.15%**	6.95%	12.63%	**90.80%**	96.67%	99.69%
	(0.27%)	(0.75%)	(1.02%)	(0.80%)	(0.50%)	(0.11%)
p-value	29.13%	0.48%	0.60%	17.71%	0.08%	0.00%
GARCH(1,1)	2.33%	7.28%	12.32%	91.86%	96.60%	99.41%
	(0.41%)	(0.77%)	(1.02%)	(0.76%)	(0.51%)	(0.18%)
p-value	0.06%	0.16%	1.13%	0.71%	0.08%	0.97%
			B. Forty Day Horizon			
TVTP MS	**0.70%**	**4.60%**	**10.82%**	**91.71%**	96.65%	99.95%
	(0.31%)	(0.75%)	(1.20%)	(1.09%)	(0.65%)	(0.04%)
p-value	16.62%	29.75%	24.89%	5.78%	0.55%	0.00%
CTP MS	**1.20%**	**6.25%**	**11.92%**	**91.54%**	96.99%	99.95%
	(0.41%)	(0.95%)	(1.34%)	(1.07%)	(0.59%)	(0.03%)
p-value	31.19%	8.71%	7.07%	7.29%	0.03%	0.00%
GARCH(1,1)	2.36%	7.29%	**12.10%**	92.33%	97.01%	99.78%
	(0.53%)	(1.05%)	(1.36%)	(1.02%)	(0.62%)	(0.10%)
p-value	0.53%	1.50%	6.20%	1.11%	0.07%	0.00%
			C. Sixty Day Horizon			
TVTP MS	**0.68%**	**4.29%**	**10.95%**	**91.84%**	97.25%	99.85%
	(0.32%)	(0.83%)	(1.40%)	(1.21%)	(0.59%)	(0.10%)
p-value	15.65%	19.36%	24.77%	6.48%	0.01%	0.00%
CTP MS	**1.19%**	**6.41%**	**12.36%**	**91.86%**	97.62%	99.92%
	(0.45%)	(1.08%)	(1.54%)	(1.17%)	(0.52%)	(0.08%)
p-value	32.99%	8.84%	6.09%	5.97%	0.00%	0.00%
GARCH(1,1)	2.36%	7.50%	**12.23%**	92.51%	97.46%	99.73%
	(0.61%)	(1.20%)	(1.51%)	(1.13%)	(0.55%)	(0.18%)
p-value	1.33%	1.87%	6.98%	1.29%	0.00%	0.00%

Table 6: Goodness-of-fit: Cramer-von Mises Criterion of Probability Integral Transform (Three-state Model)

The table shows the Cramer-von Mise (CVM) distance between a uniform distribution and the empirical distribution of the probability integral transform of one day, five days, ten days, twenty days, forty days, and sixty days forecast. If the distribution is correctly predicted by the model, the probability integral transform become close to the uniform distribution (null hypothesis). The smaller the CVM statistic is, the less likely the null hypothesis is rejected. Therefore, the smaller CVM criterion implies better forecast ability.

	One Day	Five Days	Ten Days	Twenty Days	Forty Days	Sixty Days
TVTP MS	1.4809	7.7905	4.8739	4.3723	2.2906	1.1791
CTP MS	1.5849	10.6190	10.0120	11.5930	9.2199	7.8532
GARCH(1,1)	5.3180	6.2412	7.0131	8.8511	7.9298	7.7959

Table 7: Maximum Likelihood Estimation Over Full Sample (Multi-state Model)

This table displays the model parameters from the maximum likelihood estimation over $T = 24,896$ in-sample data. For the optimization, $\hat{\mu} = 0.000303$ (long-run mean of return over full sample) is calibrated for all the models. To make the Hessian invertible during the optimization, I also put optimization constraints of the form as following: $\bar{\sigma} \in (0.000, 0.100)$, $a \in (0.000, 1.000)$, $b \in (0.000, 1.000)$, $\psi_u \in (0.000, 0.100)$, $\psi_l \in (0.000, 0.100)$, $\delta \in (0.000, 1.000)$. Therefore, I use the method of Byrd et al. (1995) for the optimization method, which can perform a constrained optimization.

k	$\hat{\bar{\sigma}}$	$\hat{a}$	$\hat{b}$	$\hat{\psi}_u$	$\hat{\psi}_l$	$\hat{\delta}$	$\ln L$
1	0.010576	0.485822	0.401171	0.021359	0.025781	0.637379	84,539.85
	(0.00017)	(0.00488)	(0.00616)	(0.00051)	(0.00063)	(0.01890)	
2	0.010529	0.614678	0.513070	0.021857	0.024664	0.611423	84,974.25
	(0.00024)	(0.00580)	(0.00757)	(0.00058)	(0.00079)	(0.01601)	
3	0.010085	0.698387	0.592994	0.019859	0.021732	0.604298	85,045.90
	(0.00029)	(0.00612)	(0.00891)	(0.00061)	(0.00082)	(0.01465)	
4	0.009995	0.755199	0.657604	0.017763	0.019261	0.618038	85,068.45
	(0.00036)	(0.00631)	(0.00922)	(0.00071)	(0.00096)	(0.01418)	
5	0.009946	0.792360	0.697489	0.015906	0.01722	0.627195	85,078.85
	(0.00046)	(0.00695)	(0.00988)	(0.00082)	(0.00116)	(0.01363)	
6	0.009838	0.820308	0.728760	0.014245	0.01536	0.638627	85,084.65
	(0.00058)	(0.00756)	(0.01029)	(0.00092)	(0.00134)	(0.01343)	
7	0.009460	0.842569	0.750696	0.012692	0.013452	0.643592	85,088.60
	(0.00062)	(0.00767)	(0.01049)	(0.00088)	(0.00129)	(0.01329)	
8	0.008966	0.858307	0.765904	0.011593	0.011939	0.642466	85,091.25
	(0.00059)	(0.00791)	(0.01094)	(0.00081)	(0.00113)	(0.01331)	
9	0.008878	0.859290	0.764605	0.011693	0.011996	0.643012	85,089.90
	(0.00061)	(0.00908)	(0.01217)	(0.00082)	(0.00111)	(0.01365)	
10	0.008439	0.868329	0.773204	0.010985	0.010951	0.64128	85,091.80
	(0.00049)	(0.00937)	(0.01220)	(0.00075)	(0.00092)	(0.01355)	

Table 8: Maximum Likelihood Estimation Over The First Half Sample (Multi-state Model)

This table displays the model parameters from the maximum likelihood estimation over the first half of the sample, $T^{'} = 12,448$. I use these parameters for the out-of-sample analysis. For the optimization, $\hat{\mu} = 0.000314$ (long-run mean of return over half sample) is calibrated for all the models. To make the Hessian invertible during the optimization, I also put optimization constraints of the form as following: $\bar{\sigma} \in (0.000, 0.100)$, $a \in (0.000, 1.000)$, $b \in (0.000, 1.000)$, $\psi_u \in (0.000, 0.100)$, $\psi_l \in (0.000, 0.100)$, $\delta \in (0.000, 1.000)$. Therefore, I use the method of Byrd et al. (1995) for the optimization method, which can perform a constrained optimization.

k	$\hat{\bar{\sigma}}$	$\hat{a}$	$\hat{b}$	$\hat{\psi}_u$	$\hat{\psi}_l$	$\hat{\delta}$	$\ln L$
1	0.009780	0.461054	0.379203	0.018419	0.022074	0.647169	43,050.88
	(0.00021)	(0.00631)	(0.00786)	(0.00060)	(0.00070)	(0.02317)	
2	0.009626	0.600924	0.497141	0.018523	0.020601	0.639573	43,274.25
	(0.00028)	(0.00740)	(0.01053)	(0.00066)	(0.00083)	(0.01970)	
3	0.009442	0.694834	0.59364	0.016629	0.018017	0.641181	43,306.56
	(0.00038)	(0.00786)	(0.01171)	(0.00079)	(0.00101)	(0.01867)	
4	0.008942	0.756763	0.653619	0.013975	0.014791	0.652767	43,317.93
	(0.00054)	(0.00896)	(0.01206)	(0.00095)	(0.00127)	(0.01787)	
5	0.008496	0.798691	0.692578	0.011989	0.012399	0.660539	43,324.11
	(0.00049)	(0.00837)	(0.01204)	(0.00082)	(0.00105)	(0.01748)	
6	0.008140	0.826986	0.719912	0.010650	0.010783	0.66565	43,327.08
	(0.00050)	(0.00928)	(0.01303)	(0.00084)	(0.00106)	(0.01739)	
7	0.007864	0.845562	0.737222	0.009777	0.009716	0.667451	43,328.29
	(0.00052)	(0.01007)	(0.01425)	(0.00082)	(0.00103)	(0.01738)	
8	0.007680	0.858374	0.749416	0.00928	0.009068	0.667495	43,328.39
	(0.00052)	(0.00995)	(0.01444)	(0.00076)	(0.00095)	(0.01746)	
9	0.007494	0.869120	0.759318	0.008834	0.008503	0.665583	43,328.11
	(0.00051)	(0.00977)	(0.01441)	(0.00070)	(0.00086)	(0.01737)	
10	0.007185	0.873673	0.761149	0.008561	0.008028	0.666526	43,327.71
	(0.00047)	(0.01342)	(0.01776)	(0.00080)	(0.00087)	(0.01741)	

Table 9: Vuong Test for The Model Selection

This table displays the Vuong (1989) test statistic on the log-likelihood of each k against that of $k = 10$. Higher t-statistics means the likelihood difference is more statistically significant. Vuong test statistics is given by $V = (\ln L^k_{T'} - \ln L^{10}_{T'})/\sqrt{T}$, where $\ln L^k_{T'}$ is the likelihood function of $k \neq 10$ model, and $\ln L^{10}_{T'}$ is the likelihood function of the $k = 10$ model. Panel A shows the t-statistics and p-values of the Vuong test for the full sample. Panel B shows the HAC adjusted version (1987) of the Vuong test for the full sample. Panel C and Panel D shows the same for the half sample.

Full Sample									
k	1	2	3	4	5	6	7	8	9
Vuong	-3.498	-0.745	-0.291	-0.148	-0.082	-0.045	-0.020	-0.003	-0.012
A. Vuong Test									
t-stats	-12.24	-5.546	-3.319	-2.121	-1.368	-0.874	-0.493	-0.123	-0.959
p-value	0.000	0.000	0.001	0.017	0.086	0.191	0.311	0.451	0.169
B. HAC-adjusted Vuong Test									
t-stats	-8.209	-5.106	-3.178	-2.038	-1.343	-0.864	-0.475	-0.124	-1.008
p-value	0.000	0.000	0.001	0.021	0.090	0.194	0.318	0.451	0.157
Half Sample									
k	1	2	3	4	5	6	7	8	9
Vuong	-2.482	-0.479	-0.190	-0.088	-0.032	-0.006	0.005	0.006	0.004
C. Vuong Test									
t-stats	-6.044	-2.433	-1.350	-0.765	-0.347	-0.079	0.099	0.291	0.291
p-value	0.000	0.000	0.028	0.140	0.312	0.456	0.556	0.597	0.596
D. HAC-adjusted Vuong Test									
t-stats	-4.222	-2.298	-1.309	-0.747	-0.344	-0.080	0.105	0.191	0.196
p-value	0.000	0.001	0.032	0.145	0.313	0.455	0.559	0.606	0.609

Table 10: Mincer-Zarnowitz Regression of Point Forecast (Multi-state Model)

This table displays the coefficient of Mincer-Zarnowitz evaluation of the realized volatility forecast $RV_{t,n} = \gamma_0 + \gamma_1 \mathbb{E}_{t-n} RV_{t,n} + u_t$. The regression is performed over $T' = 12,448$ out-of-sample observations on forecasts produced by each model. If the realized volatility is accurately predicted, the intercept γ_0 should be close to zero, and the slope γ_1 should be close to one. The standard errors in the table are HAC adjusted (Newey and West, 1987).

	One Day		Five Days		Ten Days		Twenty Days		Forty Days		Sixty Days	
k	γ_0	γ_1	γ_0	γ_1	γ_0	γ_1	γ_0	γ_1	γ_0	γ_1	γ_0	γ_1
1	0.000	1.127	-0.00	1.218	-0.00	1.322	-0.00	1.505	-0.00	1.887	-0.01	2.218
	(0.00)	(0.03)	(0.00)	(0.02)	(0.00)	(0.02)	(0.00)	(0.03)	(0.00)	(0.04)	(0.00)	(0.05)
R^2	0.084		0.195		0.219		0.213		0.186		0.158	
2	0.000	1.267	-0.00	1.309	-0.00	1.325	-0.00	1.353	-0.00	1.495	-0.00	1.645
	(0.00)	(0.03)	(0.00)	(0.02)	(0.00)	(0.02)	(0.00)	(0.02)	(0.00)	(0.02)	(0.00)	(0.03)
R^2	0.148		0.331		0.346		0.308		0.253		0.212	
3	0.000	1.258	-0.00	1.291	-0.00	1.294	-0.00	1.302	-0.00	1.403	-0.00	1.508
	(0.00)	(0.03)	(0.00)	(0.02)	(0.00)	(0.02)	(0.00)	(0.02)	(0.00)	(0.02)	(0.00)	(0.03)
R^2	0.147		0.346		0.349		0.310		0.257		0.221	
4	0.000	1.241	-0.00	1.308	-0.00	1.344	-0.00	1.403	-0.00	1.586	-0.00	1.753
	(0.00)	(0.03)	(0.00)	(0.02)	(0.00)	(0.02)	(0.00)	(0.02)	(0.00)	(0.02)	(0.00)	(0.03)
R^2	0.170		0.371		0.389		0.330		0.267		0.221	
5	0.000	0.982	0.000	1.029	-0.00	1.046	-0.00	1.085	-0.00	1.218	-0.00	1.324
	(0.00)	(0.02)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.02)	(0.00)	(0.02)
R^2	0.177		0.387		0.389		0.339		0.265		0.227	
6	0.000	1.035	-0.00	1.071	-0.00	1.073	-0.00	1.086	-0.00	1.178	-0.00	1.258
	(0.00)	(0.02)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.02)	(0.00)	(0.02)
R^2	0.179		0.389		0.393		0.344		0.281		0.230	
7	0.000	1.040	-0.00	1.077	-0.00	1.074	-0.00	1.075	-0.00	1.151	-0.00	1.219
	(0.00)	(0.02)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.02)	(0.00)	(0.02)
R^2	0.177		0.391		0.395		0.345		0.281		0.229	
8	0.000	0.943	0.000	0.969	0.000	0.966	-0.00	0.971	-0.00	1.048	-0.00	1.111
	(0.00)	(0.02)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.02)	(0.00)	(0.02)
R^2	0.177		0.388		0.392		0.343		0.282		0.230	
9	0.000	0.870	0.000	0.888	0.000	0.880	0.000	0.883	-0.00	0.955	-0.00	1.013
	(0.00)	(0.02)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.02)
R^2	0.173		0.380		0.383		0.337		0.279		0.228	
10	0.000	0.868	0.000	0.887	0.000	0.880	0.000	0.883	-0.00	0.956	-0.00	1.014
	(0.00)	(0.02)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.02)
R^2	0.173		0.380		0.383		0.337		0.279		0.228	

Table 11: Failure Rate of Value-at-Risk Forecast (Multi-state Model: One Day Horizon)

This table displays the frequency of returns that go further down than the VaR forecasted by the corresponding model (failure rate). The failure rate is calculated using $T' = 12,448$ out-of-sample observations. If the VaR is accurately predicted, the failure rate should be close to the target percentile, p. Boldface numbers are the failure rate, which is statistically indifferent from the target p at the 3% confidence level. The standard errors in the table are HAC adjusted (Newey and West, 1987).

Quantile	1%	5%	10%	90%	95%	99%
k=1	1.49%	6.27%	11.11%	89.33%	**94.87%**	99.26%
	(0.11%)	(0.21%)	(0.28%)	(0.26%)	(0.19%)	(0.08%)
p-value	0.00%	0.00%	0.00%	0.84%	16.47%	0.00%
k=2	1.44%	6.11%	11.11%	**89.73%**	**95.08%**	99.48%
	(0.11%)	(0.21%)	(0.27%)	(0.25%)	(0.18%)	(0.06%)
p-value	0.00%	0.00%	0.00%	14.43%	31.94%	0.00%
k=3	1.38%	5.83%	11.13%	**89.81%**	**95.10%**	99.51%
	(0.10%)	(0.21%)	(0.28%)	(0.25%)	(0.18%)	(0.06%)
p-value	0.01%	0.00%	0.00%	22.05%	28.87%	0.00%
k=4	1.42%	6.05%	11.17%	**89.52%**	**95.13%**	99.47%
	(0.10%)	(0.21%)	(0.27%)	(0.25%)	(0.18%)	(0.07%)
p-value	0.00%	0.00%	0.00%	3.06%	24.75%	0.00%
k=5	1.50%	6.03%	11.14%	**89.75%**	**95.16%**	99.46%
	(0.11%)	(0.21%)	(0.27%)	(0.26%)	(0.19%)	(0.07%)
p-value	0.00%	0.00%	0.00%	16.41%	19.89%	0.00%
k=6	1.43%	5.95%	11.04%	**89.77%**	**95.19%**	99.48%
	(0.11%)	(0.21%)	(0.27%)	(0.26%)	(0.18%)	(0.06%)
p-value	0.00%	0.00%	0.00%	18.80%	15.01%	0.00%
k=7	1.49%	5.94%	11.07%	**89.67%**	**95.10%**	99.41%
	(0.11%)	(0.21%)	(0.26%)	(0.26%)	(0.19%)	(0.07%)
p-value	0.00%	0.00%	0.00%	9.83%	29.47%	0.00%
k=8	1.46%	5.92%	11.04%	**89.77%**	**95.16%**	99.42%
	(0.11%)	(0.21%)	(0.26%)	(0.26%)	(0.18%)	(0.07%)
p-value	0.00%	0.00%	0.00%	18.78%	19.70%	0.00%
k=9	1.48%	5.98%	11.01%	**89.73%**	**95.13%**	99.43%
	(0.11%)	(0.21%)	(0.26%)	(0.26%)	(0.18%)	(0.07%)
p-value	0.00%	0.00%	0.01%	14.37%	23.52%	0.00%
k=10	1.45%	5.93%	11.08%	**89.68%**	**95.17%**	99.42%
	(0.11%)	(0.21%)	(0.26%)	(0.26%)	(0.18%)	(0.07%)
p-value	0.00%	0.00%	0.00%	10.57%	18.30%	0.00%

Table 12: Failure Rate of Value-at-Risk Forecast (Multi-state Model: Five Day Horizon)

This table displays the frequency of returns that go further down than the VaR forecasted by the corresponding model (failure rate). The failure rate is calculated using $T' = 12,448$ out-of-sample observations. If the VaR is accurately predicted, the failure rate should be close to the target percentile, p. Boldface numbers are the failure rate, which is statistically indifferent from the target p at the 3% confidence level. The standard errors in the table are HAC adjusted (Newey and West, 1987).

Quantile	1%	5%	10%	90%	95%	99%
k=1	1.53%	7.22%	13.26%	**89.40%**	**95.51%**	99.34%
	(0.17%)	(0.38%)	(0.52%)	(0.47%)	(0.32%)	(0.11%)
p-value	0.21%	0.00%	0.00%	6.65%	7.14%	0.02%
k=2	1.63%	7.03%	13.06%	**89.45%**	**95.46%**	99.30%
	(0.18%)	(0.39%)	(0.53%)	(0.47%)	(0.31%)	(0.12%)
p-value	0.03%	0.00%	0.00%	11.99%	7.13%	0.77%
k=3	1.60%	6.88%	12.76%	**89.60%**	95.66%	99.39%
	(0.18%)	(0.39%)	(0.53%)	(0.46%)	(0.30%)	(0.11%)
p-value	0.04%	0.00%	0.00%	19.63%	1.50%	0.01%
k=4	1.54%	7.06%	12.97%	**89.21%**	**95.41%**	99.34%
	(0.18%)	(0.38%)	(0.53%)	(0.48%)	(0.32%)	(0.11%)
p-value	0.12%	0.00%	0.00%	4.99%	9.85%	0.15%
k=5	1.62%	6.92%	12.82%	**89.33%**	**95.44%**	**99.23%**
	(0.18%)	(0.37%)	(0.52%)	(0.48%)	(0.32%)	(0.13%)
p-value	0.03%	0.00%	0.00%	8.30%	8.43%	3.83%
k=6	1.57%	6.87%	12.73%	**89.40%**	**95.49%**	99.27%
	(0.18%)	(0.37%)	(0.52%)	(0.48%)	(0.31%)	(0.13%)
p-value	0.08%	0.00%	0.00%	10.51%	6.02%	1.87%
k=7	1.66%	6.86%	12.74%	**89.21%**	**95.37%**	**99.23%**
	(0.18%)	(0.37%)	(0.52%)	(0.49%)	(0.32%)	(0.13%)
p-value	0.01%	0.00%	0.00%	5.23%	12.61%	3.59%
k=8	1.72%	6.92%	12.76%	**89.31%**	**95.33%**	**99.23%**
	(0.19%)	(0.37%)	(0.51%)	(0.48%)	(0.32%)	(0.12%)
p-value	0.01%	0.00%	0.00%	7.78%	14.94%	3.02%
k=9	1.68%	6.91%	12.79%	**89.21%**	**95.37%**	**99.21%**
	(0.19%)	(0.37%)	(0.52%)	(0.49%)	(0.32%)	(0.13%)
p-value	0.01%	0.00%	0.00%	5.17%	12.76%	5.62%
k=10	1.67%	6.96%	12.78%	**89.32%**	**95.38%**	**99.23%**
	(0.19%)	(0.37%)	(0.52%)	(0.48%)	(0.32%)	(0.13%)
p-value	0.02%	0.00%	0.00%	7.95%	12.33%	3.14%

Table 13: Failure Rate of Value-at-Risk Forecast (Multi-state Model: Ten Day Horizon)

This table displays the frequency of returns that go further down than the VaR forecasted by the corresponding model (failure rate). The failure rate is calculated using $T' = 12,448$ out-of-sample observations. If the VaR is accurately predicted, the failure rate should be close to the target percentile, p. Boldface numbers are the failure rate, which is statistically indifferent from the target p at the 3% confidence level. The standard errors in the table are HAC adjusted (Newey and West, 1987).

Quantile	1%	5%	10%	90%	95%	99%
k=1	**1.07%**	6.34%	12.20%	**90.30%**	95.93%	99.46%
	(0.18%)	(0.50%)	(0.70%)	(0.61%)	(0.43%)	(0.14%)
p-value	36.99%	0.46%	0.12%	31.49%	1.35%	0.06%
k=2	**1.25%**	6.46%	12.33%	**90.27%**	95.95%	99.35%
	(0.21%)	(0.52%)	(0.72%)	(0.62%)	(0.43%)	(0.15%)
p-value	11.22%	0.26%	0.06%	32.85%	1.36%	1.08%
k=3	**1.21%**	6.34%	12.02%	**90.53%**	96.11%	99.39%
	(0.20%)	(0.52%)	(0.73%)	(0.61%)	(0.42%)	(0.15%)
p-value	15.06%	0.52%	0.29%	19.02%	0.38%	0.49%
k=4	**1.23%**	6.46%	12.28%	**90.02%**	**95.84%**	**99.26%**
	(0.21%)	(0.52%)	(0.72%)	(0.63%)	(0.43%)	(0.16%)
p-value	13.24%	0.24%	0.08%	48.47%	2.62%	5.93%
k=5	**1.32%**	6.30%	12.03%	**90.08%**	**95.79%**	**99.22%**
	(0.21%)	(0.50%)	(0.71%)	(0.64%)	(0.44%)	(0.17%)
p-value	6.95%	0.49%	0.21%	45.00%	3.55%	10.59%
k=6	**1.32%**	6.26%	12.03%	**90.25%**	95.90%	**99.25%**
	(0.22%)	(0.50%)	(0.71%)	(0.63%)	(0.44%)	(0.17%)
p-value	6.73%	0.60%	0.22%	34.51%	1.97%	7.01%
k=7	**1.35%**	6.26%	12.11%	**90.21%**	**95.89%**	**99.21%**
	(0.22%)	(0.50%)	(0.71%)	(0.63%)	(0.44%)	(0.17%)
p-value	6.00%	0.56%	0.15%	37.03%	2.14%	11.58%
k=8	**1.33%**	6.24%	12.08%	**90.06%**	**95.79%**	**99.18%**
	(0.22%)	(0.50%)	(0.71%)	(0.64%)	(0.44%)	(0.18%)
p-value	6.37%	0.65%	0.17%	46.46%	3.83%	15.63%
k=9	**1.39%**	6.29%	12.11%	**90.11%**	**95.82%**	**99.18%**
	(0.23%)	(0.50%)	(0.71%)	(0.64%)	(0.44%)	(0.17%)
p-value	4.29%	0.51%	0.15%	42.98%	3.23%	14.37%
k=10	**1.35%**	6.25%	12.17%	**90.12%**	**95.83%**	**99.19%**
	(0.22%)	(0.50%)	(0.71%)	(0.64%)	(0.44%)	(0.18%)
p-value	5.83%	0.63%	0.11%	42.49%	3.09%	13.53%

Table 14: Failure Rate of Value-at-Risk Forecast (Multi-state Model: Twenty Day Horizon)

This table displays the frequency of returns that go further down than the VaR forecasted by the corresponding model (failure rate). The failure rate is calculated using $T' = 12,448$ out-of-sample observations. If the VaR is accurately predicted, the failure rate should be close to the target percentile, p. Boldface numbers are the failure rate, which is statistically indifferent from the target p at the 3% confidence level. The standard errors in the table are HAC adjusted (Newey and West, 1987).

Quantile	1%	5%	10%	90%	95%	99%
k=1	**0.82%**	**6.05%**	12.08%	**90.37%**	**96.04%**	99.47%
	(0.20%)	(0.61%)	(0.95%)	(0.85%)	(0.56%)	(0.11%)
p-value	20.35%	6.28%	1.66%	33.38%	3.46%	0.20%
k=2	**1.03%**	**6.24%**	12.19%	**90.48%**	**96.06%**	99.46%
	(0.24%)	(0.68%)	(0.98%)	(0.85%)	(0.57%)	(0.17%)
p-value	45.82%	3.51%	1.25%	28.78%	3.20%	0.30%
k=3	**1.19%**	**6.30%**	11.93%	**90.72%**	96.21%	99.47%
	(0.26%)	(0.70%)	(0.98%)	(0.84%)	(0.55%)	(0.16%)
p-value	22.71%	3.29%	2.46%	19.56%	1.47%	0.21%
k=4	**1.07%**	6.32%	12.20%	**90.27%**	**95.83%**	99.36%
	(0.24%)	(0.68%)	(0.97%)	(0.87%)	(0.59%)	(0.18%)
p-value	39.25%	2.67%	1.18%	37.64%	8.12%	2.16%
k=5	**1.10%**	6.33%	12.10%	**90.09%**	**95.67%**	**99.22%**
	(0.25%)	(0.67%)	(0.96%)	(0.89%)	(0.61%)	(0.21%)
p-value	34.57%	2.38%	1.46%	46.04%	13.64%	15.36%
k=6	**1.07%**	6.28%	12.21%	**90.44%**	**95.73%**	**99.30%**
	(0.24%)	(0.67%)	(0.97%)	(0.87%)	(0.61%)	(0.19%)
p-value	39.36%	2.78%	1.11%	30.43%	11.37%	5.82%
k=7	**1.01%**	6.26%	12.20%	**90.21%**	**95.67%**	**99.25%**
	(0.24%)	(0.66%)	(0.96%)	(0.89%)	(0.61%)	(0.20%)
p-value	48.50%	2.75%	1.11%	40.64%	13.37%	10.71%
k=8	**1.07%**	6.38%	12.24%	**90.19%**	**95.61%**	**99.23%**
	(0.24%)	(0.67%)	(0.96%)	(0.89%)	(0.61%)	(0.20%)
p-value	37.78%	1.89%	1.02%	41.70%	15.99%	12.68%
k=9	**1.07%**	6.29%	12.29%	**90.18%**	**95.62%**	**99.23%**
	(0.24%)	(0.66%)	(0.97%)	(0.88%)	(0.62%)	(0.20%)
p-value	39.20%	2.56%	0.89%	42.05%	15.95%	13.54%
k=10	**1.10%**	6.30%	12.28%	**90.24%**	**95.60%**	**99.25%**
	(0.25%)	(0.66%)	(0.97%)	(0.89%)	(0.62%)	(0.20%)
p-value	34.64%	2.37%	0.92%	39.25%	16.42%	10.43%

Table 15: Failure Rate of Value-at-Risk Forecast (Multi-state Model: Forty Day Horizon)

This table displays the frequency of returns that go further down than the VaR forecasted by the corresponding model (failure rate). The failure rate is calculated using $T' = 12,448$ out-of-sample observations. If the VaR is accurately predicted, the failure rate should be close to the target percentile, p. Boldface numbers are the failure rate, which is statistically indifferent from the target p at the 3% confidence level. The standard errors in the table are HAC adjusted (Newey and West, 1987).

Quantile	1%	5%	10%	90%	95%	99%
k=1	**1.18%**	**5.42%**	**11.34%**	**90.64%**	**95.89%**	99.63%
	(0.42%)	(0.88%)	(1.28%)	(1.16%)	(0.75%)	(0.16%)
p-value	33.69%	31.78%	14.82%	29.19%	11.62%	0.01%
k=2	**1.10%**	**5.74%**	**11.64%**	**90.66%**	**96.05%**	99.71%
	(0.39%)	(0.90%)	(1.27%)	(1.15%)	(0.74%)	(0.12%)
p-value	40.10%	20.63%	9.91%	28.32%	7.62%	0.00%
k=3	**1.28%**	**5.84%**	**11.51%**	**90.77%**	**96.05%**	99.66%
	(0.41%)	(0.92%)	(1.27%)	(1.14%)	(0.74%)	(0.15%)
p-value	25.26%	17.92%	11.72%	24.87%	7.70%	0.00%
k=4	**1.18%**	**5.90%**	**11.97%**	**90.36%**	**95.75%**	99.60%
	(0.40%)	(0.91%)	(1.31%)	(1.17%)	(0.78%)	(0.16%)
p-value	32.65%	16.24%	6.58%	38.08%	16.60%	0.01%
k=5	**1.23%**	**5.85%**	**11.75%**	**90.21%**	**95.50%**	99.48%
	(0.40%)	(0.90%)	(1.28%)	(1.18%)	(0.80%)	(0.19%)
p-value	28.29%	17.29%	8.54%	42.92%	26.82%	0.53%
k=6	**1.13%**	**5.75%**	**11.75%**	**90.64%**	**95.73%**	99.58%
	(0.39%)	(0.89%)	(1.29%)	(1.15%)	(0.78%)	(0.16%)
p-value	36.93%	19.80%	8.71%	28.97%	17.55%	0.02%
k=7	**1.08%**	**5.71%**	**11.82%**	**90.48%**	**95.62%**	99.55%
	(0.36%)	(0.88%)	(1.28%)	(1.16%)	(0.80%)	(0.16%)
p-value	41.08%	20.92%	7.75%	33.84%	21.89%	0.04%
k=8	**1.07%**	**5.87%**	**11.88%**	**90.49%**	**95.62%**	99.53%
	(0.37%)	(0.89%)	(1.29%)	(1.17%)	(0.80%)	(0.17%)
p-value	42.95%	16.53%	7.18%	33.64%	21.97%	0.08%
k=9	**1.07%**	**5.84%**	**11.82%**	**90.54%**	**95.56%**	99.53%
	(0.37%)	(0.90%)	(1.28%)	(1.16%)	(0.80%)	(0.17%)
p-value	43.03%	17.37%	7.78%	32.10%	24.10%	0.10%
k=10	**1.03%**	**5.81%**	**11.83%**	**90.47%**	**95.62%**	99.47%
	(0.36%)	(0.89%)	(1.28%)	(1.17%)	(0.80%)	(0.19%)
p-value	47.26%	18.13%	7.58%	34.46%	21.89%	0.64%

Table 16: Failure Rate of Value-at-Risk Forecast (Multi-state Model: Sixty Day Horizon)

This table displays the frequency of returns that go further down than the VaR forecasted by the corresponding model (failure rate). The failure rate is calculated using $T' = 12,448$ out-of-sample observations. If the VaR is accurately predicted, the failure rate should be close to the target percentile, p. Boldface numbers are the failure rate, which is statistically indifferent from the target p at the 3% confidence level. The standard errors in the table are HAC adjusted (Newey and West, 1987).

Quantile	1%	5%	10%	90%	95%	99%
k=1	**1.15%**	**5.79%**	**11.86%**	**90.82%**	**95.88%**	99.50%
	(0.48%)	(1.00%)	(1.51%)	(1.32%)	(0.79%)	(0.23%)
p-value	38.00%	21.59%	10.98%	26.81%	13.25%	1.51%
k=2	**1.06%**	**5.96%**	**12.10%**	**90.70%**	**96.27%**	99.64%
	(0.43%)	(1.00%)	(1.51%)	(1.28%)	(0.74%)	(0.20%)
p-value	44.68%	17.04%	8.19%	29.22%	4.34%	0.05%
k=3	**1.19%**	**6.11%**	**11.99%**	**90.77%**	**96.21%**	99.62%
	(0.47%)	(1.03%)	(1.51%)	(1.31%)	(0.75%)	(0.22%)
p-value	33.81%	13.99%	9.27%	27.64%	5.16%	0.21%
k=4	**1.15%**	**6.26%**	**12.42%**	**90.19%**	**95.57%**	99.53%
	(0.43%)	(1.06%)	(1.50%)	(1.34%)	(0.83%)	(0.23%)
p-value	36.77%	11.74%	5.33%	44.50%	24.71%	0.96%
k=5	**1.09%**	**6.21%**	**12.30%**	**90.03%**	**95.43%**	99.51%
	(0.43%)	(1.04%)	(1.53%)	(1.35%)	(0.85%)	(0.24%)
p-value	41.83%	12.17%	6.66%	49.04%	30.56%	1.77%
k=6	**0.98%**	**5.92%**	**12.19%**	**90.35%**	**95.82%**	99.62%
	(0.39%)	(1.02%)	(1.51%)	(1.32%)	(0.80%)	(0.21%)
p-value	47.62%	18.37%	7.35%	39.66%	15.18%	0.16%
k=7	**0.94%**	**6.02%**	**12.26%**	**90.06%**	**95.67%**	99.60%
	(0.39%)	(1.00%)	(1.49%)	(1.34%)	(0.82%)	(0.21%)
p-value	44.35%	15.30%	6.44%	48.31%	20.86%	0.22%
k=8	**0.97%**	**6.01%**	**12.35%**	**90.14%**	**95.62%**	99.59%
	(0.39%)	(1.02%)	(1.51%)	(1.34%)	(0.82%)	(0.22%)
p-value	46.79%	16.02%	5.98%	45.91%	22.45%	0.40%
k=9	**0.94%**	**6.02%**	**12.36%**	**90.11%**	**95.64%**	99.61%
	(0.39%)	(1.00%)	(1.51%)	(1.33%)	(0.81%)	(0.21%)
p-value	43.49%	15.35%	5.93%	46.61%	21.55%	0.22%
k=10	**0.95%**	**6.00%**	**12.32%**	**90.20%**	**95.58%**	99.59%
	(0.38%)	(1.01%)	(1.50%)	(1.31%)	(0.83%)	(0.22%)
p-value	45.02%	16.09%	6.09%	43.90%	24.22%	0.35%

Table 17: Goodness-of-fit: Cramer-von Mises Criterion of Probability Integral Transform (Multi-state Model)

The table shows the Cramer-von Mise (CVM) distance between a uniform distribution and the empirical distribution of the probability integral transform of one day, five days, ten days, twenty days, forty days, and sixty days forecast. If the model correctly predicts the distribution, the probability integral transform becomes close to the uniform distribution (null hypothesis). The smaller the CVM statistic is, the less likely the null hypothesis is rejected. Therefore, the smaller CVM criterion implies better forecast ability.

	One Day	Five Days	Ten Days	Twenty Days	Forty Days	Sixty Days
k=1	1.4915	8.1909	4.9579	4.4430	2.1289	0.8829
k=2	1.4486	8.3581	5.3296	5.0028	2.8062	1.2970
k=3	1.3320	7.7052	4.8234	4.4128	2.1912	1.0131
k=4	1.3997	7.8948	4.8613	4.6875	2.8630	1.5022
k=5	1.3715	7.0235	4.0024	3.8058	2.7040	1.7290
k=6	1.3807	6.9587	4.1211	3.9707	2.5621	1.4340
k=7	1.3977	6.8821	3.9742	3.8931	2.7642	1.7848
k=8	1.3881	6.7435	3.9785	3.8851	2.7755	1.8132
k=9	1.3913	6.6860	3.9207	3.8039	2.8145	1.8763
k=10	1.4052	6.7182	3.8732	3.8034	2.7868	1.8821

Table 18: Robustness Analysis (Three-state model: MLE Estimators by Changing Initial σ_s)

This table displays the model parameters from the maximum likelihood estimation over $T = 24,896$ in-sample data. For the optimization, $\hat{\mu} = 0.000303$ (long-run mean of return) is calibrated for all the models. To make the Hessian invertible during the optimization, I also put optimization constraints of the form as following: $\sigma_s \in (0.000, 0.100)$, $\sigma_m \in (0.000, 0.100)$, $\sigma_v \in (0.000, 0.100)$, $\psi_u \in (0.000, 0.100)$, $\psi_l \in (0.000, 0.100)$, $\delta \in (0.000, 1.000)$. Therefore, I use the method of Byrd et al. (1995) for the optimization method, which can perform a constrained optimization. In this table, I changed the initial value of σ_s with $0.0025 \times n$ where $n = \{1, 2, 3, ..., 10\}$. For example, the first result in this table: $\hat{\sigma_s} = 0.02612$, $\hat{\sigma_m} = 0.00545$, $\hat{\sigma_v} = 0.01069$, $\hat{\psi_u} = 0.01392$, $\hat{\psi_l} = 0.01225$, and $\hat{\delta} = 0.67280$ is the MLE estimators when the initial values are set such that: $\sigma_s = 0.0025 \times n = 0.0025 \times 1 = 0.0025$, $\sigma_m = 0.01$, $\sigma_v = 0.02$, $\psi_u = 0.02$, $\psi_l = 0.02$, and $\delta = 0.60$.

n	$\hat{\sigma_s}$	$\hat{\sigma_m}$	$\hat{\sigma_v}$	$\hat{\psi_u}$	$\hat{\psi_l}$	$\hat{\delta}$	$\ln L$
Initial Value	$0.0025 \times n$	0.010	0.020	0.02	0.02	0.60	
1	0.02612	0.00545	0.01069	0.01392	0.01225	0.67280	84,171.42
	(0.00050)	(0.00005)	(0.00015)	(0.00028)	(0.00027)	(0.02322)	
2	0.00529	0.01058	0.02673	0.02090	0.02327	0.64824	84,513.31
	(0.00005)	(0.00016)	(0.00052)	(0.00049)	(0.00054)	(0.01957)	
3	0.00534	0.01085	0.02806	0.02110	0.02387	0.64882	84,509.03
	(0.00005)	(0.00016)	(0.00060)	(0.00049)	(0.00055)	(0.01950)	
4	0.00537	0.01080	0.02712	0.02118	0.02401	0.65041	84,511.87
	(0.00005)	(0.00016)	(0.00053)	(0.00049)	(0.00056)	(0.02027)	
5	0.01048	0.00552	0.02475	0.01439	0.01365	0.62062	84,103.68
	(0.00016)	(0.00006)	(0.00044)	(0.00040)	(0.00036)	(0.03101)	
6	0.01048	0.00552	0.02569	0.01412	0.01359	0.62246	84,107.61
	(0.00016)	(0.00006)	(0.00048)	(0.00037)	(0.00034)	(0.02837)	
7	0.01061	0.00555	0.02676	0.01442	0.01362	0.61782	84,100.08
	(0.00016)	(0.00006)	(0.00056)	(0.00038)	(0.00034)	(0.02821)	
8	0.01025	0.00531	0.02691	0.01217	0.01189	0.99758	84,195.78
	(0.00015)	(0.00006)	(0.00058)	(0.00025)	(0.00022)	(0.04767)	
9	0.00537	0.01075	0.02685	0.02103	0.02354	0.66459	84,512.17
	(0.00005)	(0.00016)	(0.00052)	(0.00049)	(0.00054)	(0.02050)	
10	0.01049	0.00551	0.02561	0.01410	0.01347	0.62822	84,110.16
	(0.00016)	(0.00006)	(0.00048)	(0.00037)	(0.00033)	(0.02823)	

Table 19: Robustness Analysis (Three-state model: MLE Estimators by Changing Initial σ_m)

This table displays the model parameters from the maximum likelihood estimation over $T = 24,896$ in-sample data. For the optimization, $\hat{\mu} = 0.000303$ (long-run mean of return) is calibrated for all the models. To make the Hessian invertible during the optimization, I also put optimization constraints of the form as following: $\sigma_s \in (0.000, 0.100)$, $\sigma_m \in (0.000, 0.100)$, $\sigma_v \in (0.000, 0.100)$, $\psi_u \in (0.000, 0.100)$, $\psi_l \in (0.000, 0.100)$, $\delta \in (0.000, 1.000)$. Therefore, I use the method of Byrd et al. (1995) for the optimization method, which can perform a constrained optimization. In this table, I changed the initial value of σ_m with $0.0025 \times n$ where $n = \{1, 2, 3, ..., 10\}$. For example, the first result in this table: $\hat{\sigma}_s = 0.01031$, $\hat{\sigma_m} = 0.00546$, $\hat{\sigma}_v = 0.02509$, $\hat{\psi}_u = 0.01429$, $\hat{\psi}_l = 0.01374$, and $\hat{\delta} = 0.58545$ is the MLE estimators when the initial values are set such that: $\sigma_s = 0.005$, $\sigma_m = 0.0025 \times n = 0.0025 \times 1 = 0.0025$, $\sigma_v = 0.02$, $\psi_u = 0.02$, $\psi_l = 0.02$, and $\delta = 0.60$.

n	$\hat{\sigma}_s$	$\hat{\sigma_m}$	$\hat{\sigma}_v$	$\hat{\psi}_u$	$\hat{\psi}_l$	$\hat{\delta}$	$\ln L$
Initial Value	0.005	$0.0025 \times n$	0.020	0.02	0.02	0.60	
1	0.01031	0.00546	0.02509	0.01429	0.01374	0.58545	84,094.24
	(0.00016)	(0.00006)	(0.00046)	(0.00039)	(0.00036)	(0.02659)	
2	0.00539	0.01087	0.02689	0.02181	0.02450	0.61934	84,509.76
	(0.00005)	(0.00016)	(0.00051)	(0.00052)	(0.00058)	(0.01908)	
3	0.00537	0.01090	0.02746	0.02165	0.02406	0.64140	84,510.23
	(0.00005)	(0.00016)	(0.00055)	(0.00051)	(0.00055)	(0.01944)	
4	0.00529	0.01058	0.02673	0.02090	0.02327	0.64824	84,513.31
	(0.00005)	(0.00016)	(0.00052)	(0.00049)	(0.00054)	(0.01957)	
5	0.00537	0.01074	0.02603	0.02166	0.02409	0.64049	84,510.51
	(0.00005)	(0.00016)	(0.00047)	(0.00052)	(0.00058)	(0.02040)	
6	0.00536	0.01088	0.02772	0.02136	0.02391	0.65164	84,510.26
	(0.00005)	(0.00016)	(0.00057)	(0.00050)	(0.00055)	(0.01978)	
7	0.00535	0.01086	0.02737	0.02159	0.02370	0.64988	84,510.25
	(0.00005)	(0.00016)	(0.00055)	(0.00051)	(0.00054)	(0.01948)	
8	0.00539	0.01086	0.02696	0.02145	0.02408	0.65245	84,511.40
	(0.00005)	(0.00016)	(0.00052)	(0.00050)	(0.00056)	(0.02043)	
9	0.00534	0.01074	0.02636	0.02107	0.02326	0.69352	84,510.47
	(0.00005)	(0.00016)	(0.00049)	(0.00049)	(0.00053)	(0.02219)	
10	0.00554	0.02365	0.00955	0.05171	0.05315	0.78650	84,067.64
	(0.00007)	(0.00043)	(0.00018)	(0.00133)	(0.00132)	(0.02527)	

Table 20: Robustness Analysis (Three-state model: MLE Estimators by Changing Initial σ_v)

This table displays the model parameters from the maximum likelihood estimation over $T = 24,896$ in-sample data. For the optimization, $\hat{\mu} = 0.000303$ (long-run mean of return) is calibrated for all the models. To make the Hessian invertible during the optimization, I also put optimization constraints of the form as following: $\sigma_s \in (0.000, 0.100)$, $\sigma_m \in (0.000, 0.100)$, $\sigma_v \in (0.000, 0.100)$, $\psi_u \in (0.000, 0.100)$, $\psi_l \in (0.000, 0.100)$, $\delta \in (0.000, 1.000)$. Therefore, I use the method of Byrd et al. (1995) for the optimization method, which can perform a constrained optimization. In this table, I changed the initial value of σ_v with $0.0025 \times n$ where $n = \{1, 2, 3, ..., 10\}$. For example, the first result in this table: $\hat{\sigma_s} = 0.00560$, $\hat{\sigma_m} = 0.02391$, $\hat{\sigma_v} = 0.00979$, $\hat{\psi_u} = 0.05248$, $\hat{\psi_l} = 0.05535$, and $\hat{\delta} = 0.69793$ is the MLE estimators when the initial values are set such that: $\sigma_s = 0.005$, $\sigma_m = 0.01$, $\sigma_v = 0.0025 \times n = 0.0025 \times 1 = 0.0025$, $\psi_u = 0.02$, $\psi_l = 0.02$, and $\delta = 0.60$.

n	$\hat{\sigma_s}$	$\hat{\sigma_m}$	$\hat{\sigma_v}$	$\hat{\psi_u}$	$\hat{\psi_l}$	$\hat{\delta}$	$\ln L$
Initial Value	0.005	0.010	$0.0025 \times n$	0.02	0.02	0.60	
1	0.00560	0.02391	0.00979	0.05248	0.05535	0.69793	84,041.13
	(0.00007)	(0.00043)	(0.00020)	(0.00137)	(0.00142)	(0.02133)	
2	0.00541	0.01113	0.02823	0.02260	0.02490	0.65679	84,503.56
	(0.00005)	(0.00017)	(0.00061)	(0.00054)	(0.00058)	(0.02150)	
3	0.00536	0.01081	0.02644	0.02192	0.02407	0.65224	84,510.59
	(0.00005)	(0.00016)	(0.00049)	(0.00053)	(0.00057)	(0.02089)	
4	0.00539	0.01087	0.02695	0.02144	0.02418	0.64838	84,511.26
	(0.00005)	(0.00016)	(0.00052)	(0.00050)	(0.00057)	(0.02029)	
5	0.00532	0.01074	0.02725	0.02069	0.02353	0.65907	84,511.88
	(0.00005)	(0.00016)	(0.00055)	(0.00047)	(0.00054)	(0.01996)	
6	0.00538	0.01087	0.02697	0.02136	0.02409	0.65760	84,511.53
	(0.00005)	(0.00016)	(0.00052)	(0.00050)	(0.00056)	(0.02073)	
7	0.00540	0.01099	0.02730	0.02143	0.02439	0.65417	84,509.85
	(0.00005)	(0.00017)	(0.00054)	(0.00050)	(0.00057)	(0.02052)	
8	0.00529	0.01058	0.02673	0.02090	0.02327	0.64824	84,513.31
	(0.00005)	(0.00016)	(0.00052)	(0.00049)	(0.00054)	(0.01957)	
9	0.00537	0.01075	0.02656	0.02112	0.02374	0.65242	84,512.36
	(0.00005)	(0.00016)	(0.00050)	(0.00049)	(0.00055)	(0.02018)	
10	0.00536	0.01084	0.02695	0.02152	0.02389	0.64870	84,511.71
	(0.00005)	(0.00016)	(0.00052)	(0.00051)	(0.00055)	(0.01988)	

Table 21: Robustness Analysis (Three-state model: MLE Estimators by Changing Initial ψ_u and ψ_l)

This table displays the model parameters from the maximum likelihood estimation over $T = 24,896$ in-sample data. For the optimization, $\hat{\mu} = 0.000303$ (long-run mean of return) is calibrated for all the models. To make the Hessian invertible during the optimization, I also put optimization constraints of the form as following: $\sigma_s \in (0.000, 0.100)$, $\sigma_m \in (0.000, 0.100)$, $\sigma_v \in (0.000, 0.100)$, $\psi_u \in (0.000, 0.100)$, $\psi_l \in (0.000, 0.100)$, $\delta \in (0.000, 1.000)$. Therefore, I use the method of Byrd et al. (1995) for the optimization method, which can perform a constrained optimization. In this table, I changed the initial value of ψ_u and ψ_l with $0.005 \times n$ where $n = \{1, 2, 3, ..., 10\}$. For example, the first result in this table: $\hat{\sigma_s} = 0.00537$, $\hat{\sigma_m} = 0.01083$, $\hat{\sigma_v} = 0.02682$, $\hat{\psi_u} = 0.02155$, $\hat{\psi_l} = 0.02402$, and $\hat{\delta} = 0.66263$ is the MLE estimators when the initial values are set such that: $\sigma_s = 0.005$, $\sigma_m = 0.01$, $\sigma_v = 0.02$, $\psi_u = 0.005 \times n = 0.005 \times 1 = 0.005$, $\psi_l = 0.005 \times n = 0.005 \times 1 = 0.005$, $\delta = 0.60$.

n	$\hat{\sigma_s}$	$\hat{\sigma_m}$	$\hat{\sigma_v}$	$\hat{\psi_u}$	$\hat{\psi_l}$	$\hat{\delta}$	$\ln L$
Initial Value	0.005	0.010	$0.0025 \times n$	0.02	0.02	0.60	
1	0.00537	0.01083	0.02682	0.02155	0.02402	0.66263	84,511.63
	(0.00005)	(0.00016)	(0.00051)	(0.00051)	(0.00056)	(0.02126)	
2	0.00529	0.01058	0.02673	0.02090	0.02327	0.64824	84,513.31
	(0.00005)	(0.00016)	(0.00052)	(0.00049)	(0.00054)	(0.01957)	
3	0.00535	0.01073	0.02695	0.02132	0.02385	0.62427	84,511.37
	(0.00005)	(0.00016)	(0.00052)	(0.00050)	(0.00056)	(0.01858)	
4	0.00533	0.01064	0.02661	0.02035	0.02337	0.61500	84,507.79
	(0.00005)	(0.00016)	(0.00051)	(0.00047)	(0.00055)	(0.01732)	
5	0.00541	0.01090	0.02681	0.02231	0.02511	0.58247	84,503.92
	(0.00005)	(0.00016)	(0.00051)	(0.00054)	(0.00061)	(0.01778)	
6	0.00542	0.01095	0.02681	0.02227	0.02532	0.57432	84,501.72
	(0.00005)	(0.00016)	(0.00051)	(0.00053)	(0.00062)	(0.01742)	
7	0.00543	0.01093	0.02682	0.02251	0.02554	0.56271	84,498.87
	(0.00005)	(0.00016)	(0.00051)	(0.00054)	(0.00063)	(0.01718)	
8	0.00541	0.01091	0.02680	0.02289	0.02565	0.54685	84,494.54
	(0.00005)	(0.00016)	(0.00050)	(0.00056)	(0.00064)	(0.01656)	
9	0.00538	0.01075	0.02608	0.02195	0.02480	0.56138	84,499.22
	(0.00005)	(0.00016)	(0.00047)	(0.00053)	(0.00061)	(0.01654)	
10	0.00539	0.01085	0.02688	0.02135	0.02403	0.65907	84,511.45
	(0.00005)	(0.00016)	(0.00052)	(0.00050)	(0.00056)	(0.02078)	

Table 22: Robustness Analysis (Three-state model: MLE Estimators by Changing Initial δ)

This table displays the model parameters from the maximum likelihood estimation over $T = 24,896$ in-sample data. For the optimization, $\hat{\mu} = 0.000303$ (long-run mean of return) is calibrated for all the models. To make the Hessian invertible during the optimization, I also put optimization constraints of the form as following: $\sigma_s \in (0.000, 0.100)$, $\sigma_m \in (0.000, 0.100)$, $\sigma_v \in (0.000, 0.100)$, $\psi_u \in (0.000, 0.100)$, $\psi_l \in (0.000, 0.100)$, $\delta \in (0.000, 1.000)$. Therefore, I use the method of Byrd et al. (1995) for the optimization method, which can perform a constrained optimization. In this table, I changed the initial value of δ with $0.1 \times n$ where $n = \{1, 2, 3, ..., 10\}$. For example, the first result in this table: $\hat{\sigma}_s = 0.00537$, $\hat{\sigma_m} = 0.01083$, $\hat{\sigma}_v = 0.02682$, $\hat{\psi}_u = 0.02155$, $\hat{\psi}_l = 0.02402$, and $\hat{\delta} = 0.66263$ is the MLE estimators when the initial values are set such that: $\sigma_s = 0.005$, $\sigma_m = 0.01$, $\sigma_v = 0.02$, $\psi_u = 0.02$, $\psi_l = 0.02$, and $\delta = 0.10 \times n = 0.10 \times 1 = 0.10$.

n	$\hat{\sigma}_s$	$\hat{\sigma_m}$	$\hat{\sigma}_v$	$\hat{\psi}_u$	$\hat{\psi}_l$	$\hat{\delta}$	$\ln L$
Initial Value	0.005	0.010	$0.0025 \times n$	0.02	0.02	0.60	
1	0.00947	0.00514	0.02355	0.01330	0.01305	0.48916	84,025.59
	(0.00019)	(0.00008)	(0.00053)	(0.00035)	(0.00034)	(0.01556)	
2	0.01065	0.00539	0.02584	0.01225	0.01218	1.00000	84,195.46
	(0.00016)	(0.00006)	(0.00051)	(0.00024)	(0.00021)	(0.04516)	
3	0.01051	0.00538	0.02548	0.01231	0.01204	1.00000	84,197.00
	(0.00015)	(0.00006)	(0.00048)	(0.00024)	(0.00021)	(0.04462)	
4	0.00549	0.01112	0.02716	0.02432	0.02791	0.47054	84,455.64
	(0.00005)	(0.00016)	(0.00052)	(0.00061)	(0.00074)	(0.01473)	
5	0.00541	0.01080	0.02685	0.02196	0.02503	0.56819	84,501.09
	(0.00005)	(0.00016)	(0.00051)	(0.00052)	(0.00061)	(0.01704)	
6	0.00529	0.01058	0.02673	0.02090	0.02327	0.64825	84,513.31
	(0.00005)	(0.00016)	(0.00052)	(0.00049)	(0.00054)	(0.01957)	
7	0.00535	0.01089	0.02787	0.02065	0.02357	0.73502	84,502.62
	(0.00005)	(0.00016)	(0.00059)	(0.00047)	(0.00053)	(0.02512)	
8	0.00539	0.01087	0.02702	0.02115	0.02381	0.70629	84,507.85
	(0.00005)	(0.00016)	(0.00053)	(0.00049)	(0.00054)	(0.02384)	
9	0.00537	0.01090	0.02843	0.02079	0.02363	0.78673	84,489.68
	(0.00005)	(0.00017)	(0.00063)	(0.00047)	(0.00053)	(0.03092)	
10	0.00542	0.01111	0.02785	0.02108	0.02416	0.86564	84,466.84
	(0.00005)	(0.00018)	(0.00059)	(0.00048)	(0.00054)	(0.04688)	

Table 23: Robustness Analysis (Three-state model: Half Sample. MLE Estimators by Changing Initial σ_s)

This table displays the model parameters from the maximum likelihood estimation over $T = 12,448$ in-sample data. For the optimization, $\hat{\mu} = 0.000314$ (long-run mean of return over the half sample period) is calibrated for all the models. To make the Hessian invertible during the optimization, I also put optimization constraints of the form as following: $\sigma_s \in (0.000, 0.100)$, $\sigma_m \in (0.000, 0.100)$, $\sigma_v \in (0.000, 0.100)$, $\psi_u \in (0.000, 0.100)$, $\psi_l \in (0.000, 0.100)$, $\delta \in (0.000, 1.000)$. Therefore, I use the method of Byrd et al. (1995) for the optimization method, which can perform a constrained optimization. In this table, I changed the initial value of σ_s with $0.0025 \times n$ where $n = \{1, 2, 3, ..., 10\}$. For example, the first result in this table: $\hat{\sigma_s} = 0.00472$, $\hat{\sigma_m} = 0.00971$, $\hat{\sigma_v} = 0.02646$, $\hat{\psi_u} = 0.01759$, $\hat{\psi_l} = 0.01964$, and $\hat{\delta} = 0.62878$ is the MLE estimators when the initial values are set such that: $\sigma_s = 0.0025 \times n = 0.0025 \times 1 = 0.0025$, $\sigma_m = 0.01$, $\sigma_v = 0.02$, $\psi_u = 0.02$, $\psi_l = 0.02$, and $\delta = 0.60$.

n	$\hat{\sigma_s}$	$\hat{\sigma_m}$	$\hat{\sigma_v}$	$\hat{\psi_u}$	$\hat{\psi_l}$	$\hat{\delta}$	$\ln L$
Initial Value	$0.0025 \times n$	0.010	0.020	0.02	0.02	0.60	
1	0.00472	0.00971	0.02646	0.01759	0.01964	0.62878	43,040.42
	(0.00007)	(0.00021)	(0.00062)	(0.00057)	(0.00062)	(0.02143)	
2	0.00478	0.00995	0.02634	0.01833	0.02056	0.64066	43,040.88
	(0.00007)	(0.00022)	(0.00061)	(0.00060)	(0.00067)	(0.02336)	
3	0.00477	0.01002	0.02681	0.01875	0.02069	0.64071	43,039.84
	(0.00007)	(0.00022)	(0.00064)	(0.00062)	(0.00066)	(0.02348)	
4	0.00478	0.00997	0.02683	0.01840	0.02072	0.63271	43,040.06
	(0.00007)	(0.00022)	(0.00064)	(0.00060)	(0.00067)	(0.02300)	
5	0.00931	0.00497	0.02395	0.01269	0.01184	0.61475	42,771.76
	(0.00024)	(0.00010)	(0.00057)	(0.00056)	(0.00052)	(0.03456)	
6	0.00960	0.00494	0.02465	0.01253	0.01163	0.61556	42,774.11
	(0.00028)	(0.00009)	(0.00063)	(0.00053)	(0.00048)	(0.03229)	
7	0.00994	0.00497	0.02730	0.01265	0.01217	0.61167	42,761.89
	(0.00028)	(0.00009)	(0.00080)	(0.00052)	(0.00050)	(0.03666)	
8	0.01033	0.00513	0.02856	0.01358	0.01179	0.61377	42,739.90
	((0.00033)	(0.00008)	(0.00115)	(0.00049)	(0.00038)	(0.03210)	
9	0.00478	0.00995	0.02639	0.01823	0.02055	0.65416	43,041.06
	(0.00007)	(0.00022)	(0.00061)	(0.00060)	(0.00067)	(0.02429)	
10	0.00939	0.00494	0.02409	0.01249	0.01172	0.62257	42,775.65
	(0.00026)	(0.00010)	(0.00059)	(0.00056)	(0.00052)	(0.03487)	

Table 24: Robustness Analysis (Three-state model: Half Sample. MLE Estimators by Changing Initial σ_m)

This table displays the model parameters from the maximum likelihood estimation over $T = 12,448$ in-sample data. For the optimization, $\hat{\mu} = 0.000314$ (long-run mean of return over the half sample period) is calibrated for all the models. To make the Hessian invertible during the optimization, I also put optimization constraints of the form as following: $\sigma_s \in (0.000, 0.100)$, $\sigma_m \in (0.000, 0.100)$, $\sigma_v \in (0.000, 0.100)$, $\psi_u \in (0.000, 0.100)$, $\psi_l \in (0.000, 0.100)$, $\delta \in (0.000, 1.000)$. Therefore, I use the method of Byrd et al. (1995) for the optimization method, which can perform a constrained optimization. In this table, I changed the initial value of σ_m with $0.0025 \times n$ where $n = \{1, 2, 3, ..., 10\}$. For example, the first result in this table: $\hat{\sigma}_s = 0.00921$, $\hat{\sigma_m} = 0.00485$, $\hat{\sigma}_v = 0.02449$, $\hat{\psi}_u = 0.01256$, $\hat{\psi}_l = 0.01152$, and $\hat{\hat{\delta}} = 0.58798$ is the MLE estimators when the initial values are set such that: $\sigma_s = 0.005$, $\sigma_m = 0.0025 \times n = 0.0025 \times 1 = 0.0025$, $\sigma_v = 0.02$, $\psi_u = 0.02$, $\psi_l = 0.02$, and $\delta = 0.60$.

n	$\hat{\sigma}_s$	$\hat{\sigma_m}$	$\hat{\sigma}_v$	$\hat{\psi}_u$	$\hat{\psi}_l$	$\hat{\hat{\delta}}$	$\ln L$
Initial Value	0.005	$0.0025 \times n$	0.020	0.02	0.02	0.60	
1	0.00921	0.00485	0.02449	0.01256	0.01152	0.58798	42,768.87
	(0.00025)	(0.00011)	(0.00063)	(0.00060)	(0.00054)	(0.03144)	
2	0.00475	0.00987	0.02685	0.01835	0.02094	0.61611	43,039.33
	(0.00007)	(0.00021)	(0.00064)	(0.00060)	(0.00070)	(0.02256)	
3	0.00479	0.00997	0.02641	0.01836	0.02071	0.63936	43,040.56
	(0.00007)	(0.00022)	(0.00061)	(0.00060)	(0.00067)	(0.02349)	
4	0.00478	0.00995	0.02634	0.01833	0.02056	0.64066	43,040.88
	(0.00007)	(0.00022)	(0.00061)	(0.00060)	(0.00067)	(0.02336)	
5	0.00471	0.00982	0.02740	0.01765	0.02010	0.64026	43,039.74
	(0.00007)	(0.00021)	(0.00069)	(0.00057)	(0.00065)	(0.02257)	
6	0.00477	0.01000	0.02715	0.01833	0.02072	0.64248	43,040.00
	(0.00007)	(0.00022)	(0.00066)	(0.00060)	(0.00067)	(0.02353)	
7	0.00478	0.00996	0.02668	0.01848	0.02092	0.64292	43,040.34
	(0.00007)	(0.00022)	(0.00063)	(0.00061)	(0.00069)	(0.02426)	
8	0.00478	0.00994	0.02638	0.01825	0.02054	0.64518	43,041.03
	(0.00007)	(0.00022)	(0.00061)	(0.00060)	(0.00067)	(0.02367)	
9	0.00476	0.01000	0.02642	0.01835	0.02069	0.65631	43,041.24
	(0.00007)	(0.00022)	(0.00061)	(0.00060)	(0.00068)	(0.02459)	
10	0.00510	0.02397	0.00890	0.05065	0.05279	0.80127	42,766.43
	(0.00008)	(0.00053)	(0.00023)	(0.00182)	(0.00170)	(0.03444)	

Table 25: Robustness Analysis (Three-state model: Half Sample. MLE Estimators by Changing Initial σ_v)

This table displays the model parameters from the maximum likelihood estimation over $T = 12,448$ in-sample data. For the optimization, $\hat{\mu} = 0.000314$ (long-run mean of return over the half sample period) is calibrated for all the models. To make the Hessian invertible during the optimization, I also put optimization constraints of the form as following: $\sigma_s \in (0.000, 0.100)$, $\sigma_m \in (0.000, 0.100)$, $\sigma_v \in (0.000, 0.100)$, $\psi_u \in (0.000, 0.100)$, $\psi_l \in (0.000, 0.100)$, $\delta \in (0.000, 1.000)$. Therefore, I use the method of Byrd et al. (1995) for the optimization method, which can perform a constrained optimization. In this table, I changed the initial value of σ_v with $0.0025 \times n$ where $n = \{1, 2, 3, ..., 10\}$. For example, the first result in this table: $\hat{\sigma_s} = 0.00509$, $\hat{\sigma_m} = 0.02355$, $\hat{\sigma_v} = 0.00904$, $\hat{\psi}_u = 0.04943$, $\hat{\psi}_l = 0.05212$, and $\hat{\delta} = 0.71048$ is the MLE estimators when the initial values are set such that: $\sigma_s = 0.005$, $\sigma_m = 0.01$, $\sigma_v = 0.0025 \times n = 0.0025 \times 1 = 0.0025$, $\psi_u = 0.02$, $\psi_l = 0.02$, and $\delta = 0.60$.

n	$\hat{\sigma_s}$	$\hat{\sigma_m}$	$\hat{\sigma_v}$	$\hat{\psi}_u$	$\hat{\psi}_l$	$\hat{\delta}$	$\ln L$
Initial Value	0.005	0.010	$0.0025 \times n$	0.02	0.02	0.60	
1	0.00509	0.02355	0.00904	0.04943	0.05212	0.71048	42,752.81
	(0.00008)	(0.00049)	(0.00024)	(0.00174)	(0.00165)	(0.02807)	
2	0.00477	0.00989	0.02575	0.01813	0.02059	0.64324	43,040.57
	(0.00007)	(0.00022)	(0.00057)	(0.00060)	(0.00068)	(0.02377)	
3	0.00478	0.00992	0.02634	0.01839	0.02067	0.64712	43,041.01
	(0.00007)	(0.00022)	(0.00061)	(0.00060)	(0.00067)	(0.02420)	
4	0.00474	0.00986	0.02587	0.01921	0.02092	0.63941	43,039.40
	(0.00007)	(0.00022)	(0.00058)	(0.00065)	(0.00070)	(0.02490)	
5	0.00470	0.00967	0.02581	0.01760	0.01964	0.64806	43,042.48
	(0.00007)	(0.00021)	(0.00058)	(0.00057)	(0.00063)	(0.02290)	
6	0.00478	0.00999	0.02648	0.01828	0.02064	0.64956	43,040.94
	(0.00007)	(0.00022)	(0.00062)	(0.00060)	(0.00067)	(0.02398)	
7	0.00476	0.00997	0.02830	0.01774	0.02049	0.64946	43,035.75
	(0.00007)	(0.00022)	(0.00076)	(0.00057)	(0.00065)	(0.02340)	
8	0.00478	0.00995	0.02634	0.01833	0.02056	0.64066	43,040.88
	(0.00007)	(0.00022)	(0.00061)	(0.00060)	(0.00067)	(0.02336)	
9	0.00476	0.01002	0.02747	0.01886	0.02093	0.63823	43,038.59
	(0.00007)	(0.00022)	(0.00068)	(0.00062)	(0.00067)	(0.02377)	
10	0.00476	0.00990	0.02564	0.01845	0.02059	0.64152	43,040.50
	(0.00007)	(0.00022)	(0.00056)	(0.00061)	(0.00068)	(0.02380)	

Table 26: Robustness Analysis (Three-state model: Half Sample. MLE Estimators by Changing Initial ψ_u and ψ_l)

This table displays the model parameters from the maximum likelihood estimation over $T = 12,448$ in-sample data. For the optimization, $\hat{\mu} = 0.000314$ (long-run mean of return over the half sample period) is calibrated for all the models. To make the Hessian invertible during the optimization, I also put optimization constraints of the form as following: $\sigma_s \in (0.000, 0.100)$, $\sigma_m \in (0.000, 0.100)$, $\sigma_v \in (0.000, 0.100)$, $\psi_u \in (0.000, 0.100)$, $\psi_l \in (0.000, 0.100)$, $\delta \in (0.000, 1.000)$. Therefore, I use the method of Byrd et al. (1995) for the optimization method, which can perform a constrained optimization. In this table, I changed the initial value of ψ_u and ψ_l with $0.005 \times n$ where $n = \{1, 2, 3, ..., 10\}$. For example, the first result in this table: $\hat{\sigma_s} = 0.00478$, $\hat{\sigma_m} = 0.00997$, $\hat{\sigma_v} = 0.02640$, $\hat{\psi_u} = 0.01823$, $\hat{\psi_l} = 0.02060$, and $\hat{\delta} = 0.65115$ is the MLE estimators when the initial values are set such that: $\sigma_s = 0.005$, $\sigma_m = 0.01$, $\sigma_v = 0.02$, $\psi_u = 0.005 \times n = 0.005 \times 1 = 0.005$, $\psi_l = 0.005 \times n = 0.005 \times 1 = 0.005$, $\delta = 0.60$.

n	$\hat{\sigma_s}$	$\hat{\sigma_m}$	$\hat{\sigma_v}$	$\hat{\psi_u}$	$\hat{\psi_l}$	$\hat{\delta}$	$\ln L$
Initial Value	0.005	0.010	$0.0025 \times n$	0.02	0.02	0.60	
1	0.00478	0.00997	0.02640	0.01823	0.02060	0.65115	43,041.06
	(0.00007)	(0.00022)	(0.00061)	(0.00060)	(0.00067)	(0.02408)	
2	0.00478	0.00995	0.02634	0.01833	0.02056	0.64066	43,040.88
	(0.00007)	(0.00022)	(0.00061)	(0.00060)	(0.00067)	(0.02336)	
3	0.00473	0.00990	0.02734	0.01791	0.02031	0.62910	43,039.30
	(0.00007)	(0.00021)	(0.00068)	(0.00058)	(0.00065)	(0.02200)	
4	0.00483	0.01022	0.02673	0.01946	0.02185	0.62098	43,035.97
	(0.00007)	(0.00023)	(0.00062)	(0.00065)	(0.00073)	(0.02378)	
5	0.00492	0.01074	0.02781	0.02029	0.02369	0.56911	43,021.40
	(0.00007)	(0.00026)	(0.00070)	(0.00069)	(0.00083)	(0.02136)	
6	0.00482	0.01004	0.02645	0.01918	0.02174	0.57076	43,032.43
	(0.00007)	(0.00022)	(0.00060)	(0.00063)	(0.00073)	(0.02053)	
7	0.00482	0.01005	0.02640	0.01942	0.02196	0.55746	43,029.57
	(0.00007)	(0.00022)	(0.00060)	(0.00064)	(0.00074)	(0.02004)	
8	0.00473	0.00980	0.02612	0.01936	0.02157	0.55666	43,031.00
	(0.00007)	(0.00021)	(0.00059)	(0.00065)	(0.00074)	(0.02030)	
9	0.00483	0.01003	0.02639	0.01961	0.02221	0.53707	43,024.54
	(0.00007)	(0.00022)	(0.00060)	(0.00065)	(0.00075)	(0.01922)	
10	0.00483	0.01010	0.02631	0.02000	0.02260	0.52130	43,019.65
	(0.00007)	(0.00022)	(0.00059)	(0.00066)	(0.00077)	(0.01866)	

Table 27: Robustness Analysis (Three-state model: Half Sample. MLE Estimators by Changing Initial δ)

This table displays the model parameters from the maximum likelihood estimation over $T = 12,448$ in-sample data. For the optimization, $\hat{\mu} = 0.000314$ (long-run mean of return over the half sample period) is calibrated for all the models. To make the Hessian invertible during the optimization, I also put optimization constraints of the form as following: $\sigma_s \in (0.000, 0.100)$, $\sigma_m \in (0.000, 0.100)$, $\sigma_v \in (0.000, 0.100)$, $\psi_u \in (0.000, 0.100)$, $\psi_l \in (0.000, 0.100)$, $\delta \in (0.000, 1.000)$. Therefore, I use the method of Byrd et al. (1995) for the optimization method, which can perform a constrained optimization. In this table, I changed the initial value of δ with $0.1 \times n$ where $n = \{1, 2, 3, ..., 10\}$. For example, the first result in this table: $\hat{\sigma_s} = 0.01000$, $\hat{\sigma_m} = 0.00562$, $\hat{\sigma_v} = 0.02462$, $\hat{\psi_u} = 0.02383$, $\hat{\psi_l} = 0.02418$, and $\hat{\delta} = 0.23488$ is the MLE estimators when the initial values are set such that: $\sigma_s = 0.005$, $\sigma_m = 0.01$, $\sigma_v = 0.02$, $\psi_u = 0.02$, $\psi_l = 0.02$, and $\delta = 0.10 \times n = 0.10 \times 1 = 0.10$.

n	$\hat{\sigma_s}$	$\hat{\sigma_m}$	$\hat{\sigma_v}$	$\hat{\psi_u}$	$\hat{\psi_l}$	$\hat{\delta}$	$\ln L$
Initial Value	0.005	0.010	$0.0025 \times n$	0.02	0.02	0.60	
1	0.01000	0.00562	0.02462	0.02383	0.02418	0.23488	42,503.59
	(0.00022)	(0.00010)	(0.00054)	(0.00160)	(0.00180)	(0.01917)	
2	0.00893	0.00466	0.02475	0.01023	0.00997	0.78117	42,818.51
	(0.00021)	(0.00009)	(0.00062)	(0.00036)	(0.00033)	(0.03635)	
3	0.00478	0.01001	0.02688	0.01822	0.02060	0.66678	43,040.43
	(0.00007)	(0.00022)	(0.00065)	(0.00059)	(0.00066)	(0.02513)	
4	0.00486	0.01019	0.02715	0.02022	0.02325	0.48197	43,003.27
	(0.00007)	(0.00022)	(0.00064)	(0.00066)	(0.00077)	(0.01684)	
5	0.00483	0.01008	0.02663	0.01938	0.02202	0.56013	43,029.75
	(0.00007)	(0.00022)	(0.00061)	(0.00064)	(0.00074)	(0.02016)	
6	0.00478	0.00995	0.02634	0.01833	0.02056	0.64066	43,040.88
	(0.00007)	(0.00022)	(0.00061)	(0.00060)	(0.00067)	(0.02336)	
7	0.00476	0.00990	0.02642	0.01785	0.02004	0.72747	43,038.23
	(0.00007)	(0.00022)	(0.00062)	(0.00058)	(0.00064)	(0.02955)	
8	0.00476	0.00989	0.02632	0.01786	0.02007	0.72164	43,038.76
	(0.00007)	(0.00022)	(0.00061)	(0.00058)	(0.00064)	(0.02917)	
9	0.00478	0.01001	0.02657	0.01792	0.02016	0.80456	43,026.29
	(0.00007)	(0.00023)	(0.00064)	(0.00058)	(0.00064)	(0.03933)	
10	0.00472	0.00979	0.02805	0.01717	0.01947	0.89584	43,006.60
	(0.00007)	(0.00022)	(0.00077)	(0.00054)	(0.00060)	(0.05553)	

Table 28: Robustness Analysis (Multi-state model: $k = 1$ MLE Estimators by Changing Initial $\bar{\sigma}$)

This table displays the model parameters from the maximum likelihood estimation over $T = 24,896$ in-sample data. For the optimization, $\hat{\mu} = 0.000303$ (long-run mean of return) is calibrated for all the models. To make the Hessian invertible during the optimization, I also put optimization constraints of the form as following: $\bar{\sigma} \in (0.000, 0.100)$, $a \in (0.000, 1.000)$, $b \in (0.000, 1.000)$, $\psi_u \in (0.000, 0.100)$, $\psi_l \in (0.000, 0.100)$, $\delta \in (0.000, 1.000)$. Therefore, I use the method of Byrd et al. (1995) for the optimization method, which can perform a constrained optimization. In this table, I changed the initial value of $\bar{\sigma}$ with $0.0025 \times n$ where $n = \{1, 2, 3, ..., 10\}$. For example, the first result in the table: $\bar{\sigma} = 0.01061$, $\hat{a} = 0.48532$, $\hat{b} = 0.40105$, $\hat{\psi}_u = 0.02165$, $\hat{\psi}_l = 0.02618$, $\hat{\delta} = 0.60949$ is the MLE estimator when the initial value is set such that: $\bar{\sigma} = 0.0025 \times n = 0.0025 \times 1 = 0.0025$, $a = 0.50$, $b = 0.40$, $\psi_u = 0.02$, $\psi_l = 0.02$, $\delta = 0.60$.

Initial Value	$0.0025 \times n$	0.50	0.40	0.02	0.02	0.60	
n	$\hat{\bar{\sigma}}$	$\hat{a}$	$\hat{b}$	$\hat{\psi}_u$	$\hat{\psi}_l$	$\hat{\delta}$	$\ln L$
1	0.01061	0.48532	0.40105	0.02165	0.02618	0.60949	84,538.55
	(0.00017)	(0.00488)	(0.00619)	(0.00052)	(0.00065)	(0.01777)	
2	0.01061	0.48525	0.40105	0.02166	0.02618	0.60992	84,538.55
	(0.00017)	(0.00488)	(0.00619)	(0.00052)	(0.00066)	(0.01781)	
3	0.01058	0.48656	0.40070	0.02164	0.02618	0.61056	84,538.65
	(0.00017)	(0.00490)	(0.00616)	(0.00052)	(0.00066)	(0.01786)	
4	0.01056	0.49092	0.40891	0.0217	0.02607	0.60728	84,537.25
	(0.00016)	(0.00493)	(0.00622)	(0.00052)	(0.00065)	(0.01788)	
5	0.01057	0.48741	0.40478	0.02139	0.02576	0.62934	84,539.5
	(0.00017)	(0.00488)	(0.00620)	(0.00051)	(0.00063)	(0.01857)	
6	0.01059	0.48539	0.40039	0.02156	0.02605	0.61737	84,539.15
	(0.00017)	(0.00488)	(0.00617)	(0.00051)	(0.00065)	(0.01806)	
7	0.01057	0.48589	0.40086	0.02136	0.02576	0.63723	84,539.85
	(0.00017)	(0.00488)	(0.00615)	(0.00051)	(0.00063)	(0.01889)	
8	0.01057	0.48604	0.40089	0.02135	0.02578	0.63601	84,539.8
	(0.00017)	(0.00489)	(0.00615)	(0.00051)	(0.00063)	(0.01884)	
9	0.01057	0.48515	0.39589	0.02165	0.02628	0.60433	84,537.75
	(0.00017)	(0.00489)	(0.00613)	(0.00052)	(0.00067)	(0.01752)	
10	0.01056	0.48503	0.39566	0.02164	0.02626	0.60297	84,537.55
	(0.00017)	(0.00489)	(0.00613)	(0.00052)	(0.00067)	(0.01742)	

Table 29: Robustness Analysis (Multi-state model: $k = 1$ MLE Estimators by Changing Initial a and b)

This table displays the model parameters from the maximum likelihood estimation over $T = 24,896$ in-sample data. For the optimization, $\hat{\mu} = 0.000303$ (long-run mean of return) is calibrated for all the models. To make the Hessian invertible during the optimization, I also put optimization constraints of the form as following: $\bar{\sigma} \in (0.000, 0.100)$, $a \in (0.000, 1.000)$, $b \in (0.000, 1.000)$, $\psi_u \in (0.000, 0.100)$, $\psi_l \in (0.000, 0.100)$, $\delta \in (0.000, 1.000)$. Therefore, I use the method of Byrd et al. (1995) for the optimization method, which can perform a constrained optimization. In this table, I changed the initial value of a and b with $0.10 \times n$ where $n = \{1, 2, 3, ..., 10\}$. For example, the first result in this table: $\bar{\sigma} = 0.01041$, $\hat{a} = 0.48919$, $\hat{b} = 0.39595$, $\hat{\psi}_u = 0.02077$, $\hat{\psi}_l = 0.02510$, and $\hat{\delta} = 0.68394$ is the MLE estimators when the initial values are set such that: $\bar{\sigma} = 0.010$, $a = 0.10 \times n = 0.10 \times 1 = 0.10$, $b = 0.10 \times n = 0.10 \times 1 = 0.10$, $\psi_u = 0.02$, $\psi_l = 0.02$, $\delta = 0.60$.

Initial Value	0.01	$0.10 \times n$	$0.10 \times n$	0.02	0.02	0.6	
n	$\hat{\bar{\sigma}}$	$\hat{a}$	$\hat{b}$	$\hat{\psi}_u$	$\hat{\psi}_l$	$\hat{\delta}$	$\ln L$
1	0.01041	0.48919	0.39595	0.02077	0.02510	0.68394	84,536.15
	(0.00016)	(0.00492)	(0.00601)	(0.00048)	(0.00061)	(0.02109)	
2	0.01061	0.48373	0.40199	0.02105	0.02534	0.68173	84,537.4
	(0.00017)	(0.00486)	(0.00616)	(0.00049)	(0.00061)	(0.02105)	
3	0.01060	0.48384	0.40144	0.02108	0.02542	0.66650	84,538.8
	(0.00017)	(0.00486)	(0.00617)	(0.00049)	(0.00061)	(0.02016)	
4	0.01059	0.48534	0.40124	0.02137	0.02577	0.63906	84,539.85
	(0.00017)	(0.00488)	(0.00616)	(0.00051)	(0.00063)	(0.01900)	
5	0.01056	0.48617	0.40101	0.02132	0.02573	0.63555	84,539.85
	(0.00017)	(0.00488)	(0.00616)	(0.00050)	(0.00063)	(0.01878)	
6	0.01056	0.485552	0.40088	0.02129	0.02567	0.63786	84,539.85
	(0.00017)	(0.00487)	(0.00616)	(0.00050)	(0.00063)	(0.01884)	
7	0.01068	0.47623	0.39487	0.02082	0.02518	0.71170	84,531.45
	(0.00017)	(0.00480)	(0.00617)	(0.00048)	(0.00060)	(0.02241)	
8	0.01057	0.48602	0.40129	0.02133	0.02575	0.63789	84,539.85
	(0.00017)	(0.00488)	(0.00616)	(0.00050)	(0.00063)	(0.01892)	
9	0.01057	0.485427	0.40473	0.02051	0.02467	0.78973	84,514.7
	(0.00017)	(0.00489)	(0.00614)	(0.00046)	(0.00056)	(0.02949)	
10	0.01058	0.485451	0.40106	0.02135	0.02577	0.63734	84,539.85
	(0.00017)	(0.00488)	(0.00616)	(0.00051)	(0.00063)	(0.01888)	

Table 30: Robustness Analysis (Multi-state model: $k = 1$ MLE Estimators by Changing Initial ψ_u and ψ_l)

This table displays the model parameters from the maximum likelihood estimation over $T = 24,896$ in-sample data. For the optimization, $\hat{\mu} = 0.000303$ (long-run mean of return) is calibrated for all the models. To make the Hessian invertible during the optimization, I also put optimization constraints of the form as following: $\bar{\sigma} \in (0.000, 0.100)$, $a \in (0.000, 1.000)$, $b \in (0.000, 1.000)$, $\psi_u \in (0.000, 0.100)$, $\psi_l \in (0.000, 0.100)$, $\delta \in (0.000, 1.000)$. Therefore, I use the method of Byrd et al. (1995) for the optimization method, which can perform a constrained optimization. In this table, I changed the initial value of ψ_u and ψ_l with $0.005 \times n$ where $n = \{1, 2, 3, ..., 10\}$. For example, the first result in this table: $\hat{\bar{\sigma}} = 0.01064$, $\hat{a} = 0.48507$, $\hat{b} = 0.40355$, $\hat{\psi}_u = 0.02169$, $\hat{\psi}_l = 0.02616$, and $\hat{\delta} = 0.61363$, is the MLE estimators when the initial values are set such that: $\bar{\sigma} = 0.010$, $a = 0.50$, $b = 0.40$, $\psi_u = 0.005 \times n = 0.005 \times 1 = 0.005$, $\psi_l = 0.005 \times n = 0.005 \times 1 = 0.005$, $\delta = 0.60$.

Initial Value	0.01	0.50	0.40	$0.005 \times n$	$0.005 \times n$	0.6	
n	$\hat{\bar{\sigma}}$	$\hat{a}$	$\hat{b}$	$\hat{\psi}_u$	$\hat{\psi}_l$	$\hat{\delta}$	$\ln L$
1	0.01064	0.48507	0.40355	0.02169	0.02616	0.61363	84,538.75
	(0.00017)	(0.00488)	(0.00622)	(0.00052)	(0.00065)	(0.01802)	
2	0.01060	0.48522	0.40091	0.02163	0.02613	0.61128	84,538.7
	(0.00017)	(0.00488)	(0.00619)	(0.00052)	(0.00065)	(0.01781)	
3	0.01059	0.48533	0.40077	0.02161	0.02612	0.61183	84,538.75
	(0.00017)	(0.00487)	(0.00618)	(0.00052)	(0.00065)	(0.01785)	
4	0.01061	0.48519	0.40102	0.02165	0.02617	0.60959	84,538.55
	(0.00017)	(0.00488)	(0.00619)	(0.00052)	(0.00065)	(0.01777)	
5	0.01049	0.49332	0.40768	0.02170	0.02610	0.59225	84,535.25
	(0.00016)	(0.00495)	(0.00620)	(0.00052)	(0.00065)	(0.01718)	
6	0.01062	0.48765	0.40802	0.02162	0.02605	0.61286	84,538.2
	(0.00017)	(0.00489)	(0.00625)	(0.00051)	(0.00064)	(0.01800)	
7	0.01059	0.48611	0.40215	0.02153	0.02599	0.62085	84,539.35
	(0.00017)	(0.0488)	(0.00618)	(0.00051)	(0.00064)	(0.01824)	
8	0.01061	0.48793	0.40757	0.02169	0.02608	0.61006	84,538.05
	(0.00017)	(0.00490)	(0.00624)	(0.00052)	(0.00064)	(0.01791)	
9	0.01059	0.48621	0.40297	0.02140	0.02581	0.63305	84,539.75
	(0.00017)	(0.00488)	(0.00618)	(0.00051)	(0.00063)	(0.01873)	
10	0.01061	0.48742	0.40653	0.02170	0.02613	0.60751	84,538
	(0.00017)	(0.00489)	(0.00624)	(0.00052)	(0.00065)	(0.01779)	

Table 31: Robustness Analysis (Multi-state model: $k = 1$ MLE Estimators by Changing Initial δ)

This table displays the model parameters from the maximum likelihood estimation over $T = 24,896$ in-sample data. For the optimization, $\hat{\mu} = 0.000303$ (long-run mean of return) is calibrated for all the models. To make the Hessian invertible during the optimization, I also put optimization constraints of the form as following: $\bar{\sigma} \in (0.000, 0.100)$, $a \in (0.000, 1.000)$, $b \in (0.000, 1.000)$, $\psi_u \in (0.000, 0.100)$, $\psi_l \in (0.000, 0.100)$, $\delta \in (0.000, 1.000)$. Therefore, I use the method of Byrd et al. (1995) for the optimization method, which can perform a constrained optimization. In this table, I changed the initial value of δ with $0.10 \times n$ where $n = \{1, 2, 3, ..., 10\}$. For example, the first result in this table: $\hat{\bar{\sigma}} = 0.01058$, $\hat{a} = 0.48526$, $\hat{b} = 0.40075$, $\hat{\psi}_u = 0.02131$, $\hat{\psi}_l = 0.02572$, and $\hat{\delta} = 0.64322$, is the MLE estimators when the initial values are set such that: $\bar{\sigma} = 0.010$, $a = 0.50$, $b = 0.40$, $\psi_u = 0.02$, $\psi_l = 0.02$, $\delta = 0.10 \times n = 0.10 \times 1 = 0.10$.

Initial Value	0.01	0.50	0.40	0.02	0.02	$0.10 \times n$	
n	$\hat{\bar{\sigma}}$	$\hat{a}$	$\hat{b}$	$\hat{\psi}_u$	$\hat{\psi}_l$	$\hat{\delta}$	$\ln L$
1	0.01058	0.48526	0.40075	0.02131	0.02572	0.64322	84,539.8
	(0.00017)	(0.00488)	(0.00616)	(0.00050)	(0.00063)	(0.01916)	
2	0.01064	0.48259	0.40097	0.02144	0.02586	0.63537	84,539.6
	(0.00017)	(0.00485)	(0.00620)	(0.00051)	(0.00064)	(0.01878)	
3	0.01058	0.48317	0.39649	0.02136	0.02584	0.63481	84,539.45
	(0.00017)	(0.00486)	(0.00613)	(0.00051)	(0.00064)	(0.01867)	
4	0.01057	0.48609	0.40154	0.02137	0.02579	0.63575	84,539.8
	(0.00017)	(0.00488)	(0.00616)	(0.00051)	(0.00063)	(0.01885)	
5	0.01072	0.47676	0.39084	0.02180	0.02658	0.60651	84,535.3
	(0.00017)	(0.00482)	(0.00617)	(0.00053)	(0.00068)	(0.01753)	
6	0.01061	0.48532	0.40105	0.02165	0.02618	0.60949	84,538.55
	(0.00017)	(0.00488)	(0.00619)	(0.00052)	(0.00065)	(0.01777)	
7	0.01062	0.48286	0.40199	0.02105	0.02535	0.67970	84,537.6
	(0.00017)	(0.00485)	(0.00618)	(0.00049)	(0.00061)	(0.02089)	
8	0.01061	0.48538	0.40430	0.02139	0.02575	0.63895	84,539.7
	(0.00017)	(0.00487)	(0.00621)	(0.00051)	(0.00063)	(0.01901)	
9	0.01058	0.48537	0.40118	0.02136	0.02577	0.63808	84,539.85
	(0.00017)	(0.00488)	(0.00616)	(0.00051)	(0.00063)	(0.01893)	
10	0.01057	0.48573	0.40097	0.02134	0.02576	0.63734	84,539.85
	(0.00017)	(0.00488)	(0.00616)	(0.00051)	(0.00063)	(0.01888)	

Table 32: Robustness Analysis, Half Sample (Multi-state model: $k = 1$ MLE Estimators by Changing Initial $\bar{\sigma}$)

This table displays the model parameters from the maximum likelihood estimation over $T = 12,448$ in-sample data. For the optimization, $\hat{\mu} = 0.000314$ (long-run mean of return over the half sample period) is calibrated for all the models. To make the Hessian invertible during the optimization, I also put optimization constraints of the form as following: $\bar{\sigma} \in (0.000, 0.100)$, $a \in (0.000, 1.000)$, $b \in (0.000, 1.000)$, $\psi_u \in (0.000, 0.100)$, $\psi_l \in (0.000, 0.100)$, $\delta \in (0.000, 1.000)$. Therefore, I use the method of Byrd et al. (1995) for the optimization method, which can perform a constrained optimization. In this table, I changed the initial value of $\bar{\sigma}$ with $0.0025 \times n$ where $n = \{1, 2, 3, ..., 10\}$. For example, the first result in the table: $\bar{\sigma} = 0.00982$, $\hat{a} = 0.46016$, $\hat{b} = 0.37968$, $\hat{\psi}_u = 0.01849$, $\hat{\psi}_l = 0.02217$, $\hat{\delta} = 0.64615$ is the MLE estimator when the initial value is set such that: $\bar{\sigma} = 0.0025 \times n = 0.0025 \times 1 = 0.0025$, $a = 0.50$, $b = 0.40$, $\psi_u = 0.02$, $\psi_l = 0.02$, $\delta = 0.60$.

Initial Value	$0.0025 \times n$	0.50	0.40	0.02	0.02	0.60	
n	$\hat{\bar{\sigma}}$	$\hat{a}$	$\hat{b}$	$\hat{\psi}_u$	$\hat{\psi}_l$	$\hat{\delta}$	$\ln L$
1	0.00982	0.46016	0.37968	0.01849	0.02217	0.64615	43,050.85
	(0.00021)	(0.00632)	(0.00787)	(0.00060)	(0.00070)	(0.02318)	
2	0.00982	0.46006	0.37976	0.01849	0.02217	0.64628	43,050.84
	(0.00021)	(0.00633)	(0.00787)	(0.00060)	(0.00070)	(0.02319)	
3	0.00983	0.45963	0.38050	0.01847	0.02213	0.64452	43,050.83
	(0.00021)	(0.00631)	(0.00791)	(0.00060)	(0.00070)	(0.02301)	
4	0.00982	0.46016	0.37974	0.01849	0.02216	0.64647	43,050.85
	(0.00023)	(0.00961)	(0.00945)	(0.00060)	(0.00068)	(0.02435)	
5	0.00981	0.46033	0.37977	0.01852	0.02220	0.64414	43,050.83
	(0.00021)	(0.00632)	(0.00787)	(0.00061)	(0.00070)	(0.02312)	
6	0.00982	0.46006	0.37977	0.01849	0.02217	0.64624	43,050.84
	(0.00021)	(0.00633)	(0.00787)	(0.00060)	(0.00070)	(0.02319)	
7	0.00982	0.46010	0.37977	0.01849	0.02217	0.64627	43,050.84
	(0.00021)	(0.00633)	(0.00787)	(0.00060)	(0.00070)	(0.02319)	
8	0.00980	0.46037	0.37912	0.01848	0.02218	0.64407	43,050.85
	(0.00021)	(0.00631)	(0.00786)	(0.00060)	(0.00070)	(0.02309)	
9	0.00980	0.46033	0.37938	0.01845	0.02211	0.64637	43,050.87
	(0.00021)	(0.00631)	(0.00787)	(0.00060)	(0.00070)	(0.02314)	
10	0.00982	0.46006	0.37984	0.01849	0.02217	0.64651	43,050.84
	(0.00021)	(0.00633)	(0.00787)	(0.00060)	(0.00070)	(0.02321)	

Table 33: Robustness Analysis, Half Sample (Multi-state model: $k = 1$ MLE Estimators by Changing Initial a and b)

This table displays the model parameters from the maximum likelihood estimation over $T = 12,448$ in-sample data. For the optimization, $\hat{\mu} = 0.000314$ (long-run mean of return over the half sample period) is calibrated for all the models. To make the Hessian invertible during the optimization, I also put optimization constraints of the form as following: $\bar{\sigma} \in (0.000, 0.100)$, $a \in (0.000, 1.000)$, $b \in (0.000, 1.000)$, $\psi_u \in (0.000, 0.100)$, $\psi_l \in (0.000, 0.100)$, $\delta \in (0.000, 1.000)$. Therefore, I use the method of Byrd et al. (1995) for the optimization method, which can perform a constrained optimization. In this table, I changed the initial value of a and b with $0.10 \times n$ where $n = \{1, 2, 3, ..., 10\}$. For example, the first result in this table: $\bar{\sigma} = 0.00982$, $\hat{a} = 0.46041$, $\hat{b} = 0.37959$, $\hat{\psi}_u = 0.01853$, $\hat{\psi}_l = 0.02221$, and $\hat{\delta} = 0.64377$ is the MLE estimators when the initial values are set such that: $\bar{\sigma} = 0.010$, $a = 0.10 \times n = 0.10 \times 1 = 0.10$, $b = 0.10 \times n = 0.10 \times 1 = 0.10$, $\psi_u = 0.02$, $\psi_l = 0.02$, $\delta = 0.60$.

Initial Value	0.01	$0.10 \times n$	$0.10 \times n$	0.02	0.02	0.6	
n	$\hat{\bar{\sigma}}$	$\hat{a}$	$\hat{b}$	$\hat{\psi}_u$	$\hat{\psi}_l$	$\hat{\delta}$	$\ln L$
1	0.00982	0.46041	0.37959	0.01853	0.02221	0.64377	43,050.83
	(0.00021)	(0.00633)	(0.00786)	(0.00061)	(0.00070)	(0.02310)	
2	0.00979	0.46082	0.37900	0.01843	0.02211	0.64538	43,050.87
	(0.00021)	(0.00632)	(0.00786)	(0.00060)	(0.00070)	(0.02308)	
3	0.00982	0.46063	0.38046	0.01859	0.02228	0.63630	43,050.71
	(0.00021)	(0.00632)	(0.00788)	(0.00061)	(0.00071)	(0.02273)	
4	0.00980	0.46052	0.37940	0.01847	0.02214	0.64563	43,050.86
	(0.00021)	(0.00632)	(0.00786)	(0.00060)	(0.00070)	(0.02312)	
5	0.00981	0.46007	0.37877	0.01845	0.02213	0.64545	43,050.86
	(0.00021)	(0.00632)	(0.00786)	(0.00060)	(0.00070)	(0.02307)	
6	0.00981	0.46032	0.37960	0.01844	0.02211	0.64853	43,050.86
	(0.00021)	(0.00633)	(0.00787)	(0.00060)	(0.00070)	(0.02325)	
7	0.00980	0.45978	0.37956	0.01827	0.02195	0.65980	43,050.75
	(0.00021)	(0.00629)	(0.00789)	(0.00059)	(0.00069)	(0.02376)	
8	0.00981	0.46018	0.37922	0.01842	0.02212	0.64549	43,050.86
	(0.00021)	(0.00632)	(0.00787)	(0.00060)	(0.00070)	(0.02304)	
9	0.00982	0.46007	0.37975	0.01849	0.02217	0.64630	43,050.84
	(0.00021)	(0.00633)	(0.00787)	(0.00060)	(0.00070)	(0.02319)	
10	0.00982	0.46020	0.37971	0.01848	0.02217	0.64629	43,050.85
	(0.00021)	(0.00633)	(0.00787)	(0.00060)	(0.00070)	(0.02319)	

Table 34: Robustness Analysis, Half Sample (Multi-state model: $k = 1$ MLE Estimators by Changing Initial ψ_u and ψ_l)

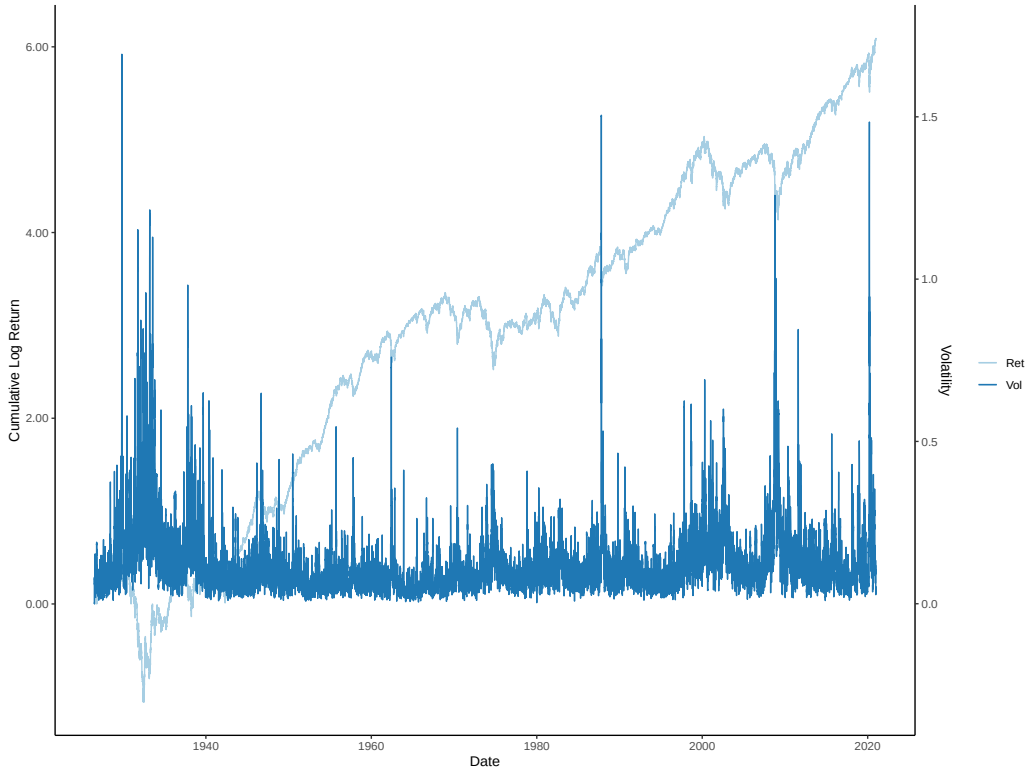
This table displays the model parameters from the maximum likelihood estimation over $T = 12,448$ in-sample data. For the optimization, $\hat{\mu} = 0.000314$ (long-run mean of return over the half sample period) is calibrated for all the models. To make the Hessian invertible during the optimization, I also put optimization constraints of the form as following: $\bar{\sigma} \in (0.000, 0.100)$, $a \in (0.000, 1.000)$, $b \in (0.000, 1.000)$, $\psi_u \in (0.000, 0.100)$, $\psi_l \in (0.000, 0.100)$, $\delta \in (0.000, 1.000)$. Therefore, I use the method of Byrd et al. (1995) for the optimization method, which can perform a constrained optimization. In this table, I changed the initial value of ψ_u and ψ_l with $0.005 \times n$ where $n = \{1, 2, 3, ..., 10\}$. For example, the first result in this table: $\bar{\sigma} = 0.00964$, $\hat{a} = 0.48334$, $\hat{b} = 0.36951$, $\hat{\psi}_u = 0.01768$, $\hat{\psi}_l = 0.01965$, and $\hat{\delta} = 0.66561$, is the MLE estimators when the initial values are set such that: $\bar{\sigma} = 0.010$, $a = 0.50$, $b = 0.40$, $\psi_u = 0.005 \times n = 0.005 \times 1 = 0.005$, $\psi_l = 0.005 \times n = 0.005 \times 1 = 0.005$, $\delta = 0.60$.

Initial Value	0.01	0.50	0.40	$0.005 \times n$	$0.005 \times n$	0.6	
n	$\hat{\bar{\sigma}}$	$\hat{a}$	$\hat{b}$	$\hat{\psi}_u$	$\hat{\psi}_l$	$\hat{\delta}$	$\ln L$
1	0.00980	0.45994	0.38010	0.01839	0.02200	0.65308	43,050.85
	(0.00021)	(0.00630)	(0.00790)	(0.00060)	(0.00069)	(0.02343)	
2	0.00982	0.46006	0.37954	0.01853	0.02221	0.64489	43,050.83
	(0.00021)	(0.00632)	(0.00786)	(0.00061)	(0.00070)	(0.02317)	
3	0.00982	0.46041	0.37907	0.01860	0.02231	0.63457	43,050.68
	(0.00021)	(0.00634)	(0.00786)	(0.00061)	(0.00071)	(0.02261)	
4	0.00982	0.46004	0.37976	0.01850	0.02217	0.64681	43,050.84
	(0.00021)	(0.00632)	(0.00787)	(0.00060)	(0.00070)	(0.02324)	
5	0.00982	0.46006	0.37975	0.01849	0.02217	0.64622	43,050.84
	(0.00021)	(0.00633)	(0.00787)	(0.00060)	(0.00070)	(0.02319)	
6	0.00982	0.46013	0.37989	0.01849	0.02218	0.64591	43,050.84
	(0.00021)	(0.00633)	(0.00787)	(0.00060)	(0.00070)	(0.02318)	
7	0.00982	0.46008	0.37984	0.01850	0.02218	0.64628	43,050.84
	(0.00021)	(0.00632)	(0.00787)	(0.00060)	(0.00070)	(0.02321)	
8	0.00978	0.46096	0.37998	0.01844	0.02210	0.64609	43,050.87
	(0.00021)	(0.00630)	(0.00788)	(0.00060)	(0.00070)	(0.02316)	
9	0.00982	0.46030	0.37990	0.01850	0.02219	0.64621	43,050.84
	(0.00021)	(0.00633)	(0.00787)	(0.00060)	(0.00070)	(0.02321)	
10	0.00982	0.46009	0.37977	0.01849	0.02217	0.64638	43,050.84
	(0.00021)	(0.00633)	(0.00787)	(0.00060)	(0.00070)	(0.02320)	

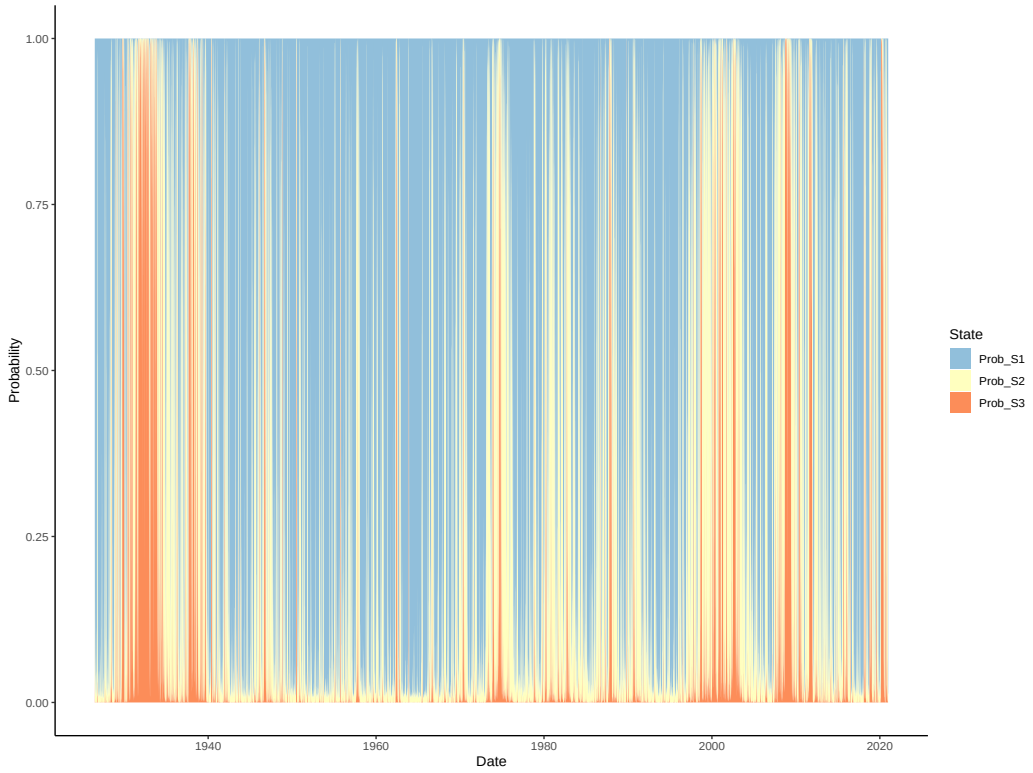
Table 35: Robustness Analysis, Half Sample (Multi-state model: $k = 1$ MLE Estimators by Changing Initial δ)

This table displays the model parameters from the maximum likelihood estimation over $T = 12,448$ in-sample data. For the optimization, $\hat{\mu} = 0.000314$ (long-run mean of return over the half sample period) is calibrated for all the models. To make the Hessian invertible during the optimization, I also put optimization constraints of the form as following: $\bar{\sigma} \in (0.000, 0.100)$, $a \in (0.000, 1.000)$, $b \in (0.000, 1.000)$, $\psi_u \in (0.000, 0.100)$, $\psi_l \in (0.000, 0.100)$, $\delta \in (0.000, 1.000)$. Therefore, I use the method of Byrd et al. (1995) for the optimization method, which can perform a constrained optimization. In this table, I changed the initial value of δ with $0.10 \times n$ where $n = \{1, 2, 3, ..., 10\}$. For example, the first result in this table: $\hat{\bar{\sigma}} = 0.00981$, $\hat{a} = 0.46027$, $\hat{b} = 0.37977$, $\hat{\psi}_u = 0.01848$, $\hat{\psi}_l = 0.02216$, and $\hat{\delta} = 0.64763$, is the MLE estimators when the initial values are set such that: $\bar{\sigma} = 0.010$, $a = 0.50$, $b = 0.40$, $\psi_u = 0.02$, $\psi_l = 0.02$, $\delta = 0.10 \times n = 0.10 \times 1 = 0.10$.

Initial Value	0.01	0.50	0.40	0.02	0.02	$0.10 \times n$	
n	$\hat{\bar{\sigma}}$	$\hat{a}$	$\hat{b}$	$\hat{\psi}_u$	$\hat{\psi}_l$	$\hat{\delta}$	$\ln L$
1	0.00981	0.46027	0.37977	0.01848	0.02216	0.64763	43,050.85
	(0.00021)	(0.00631)	(0.00787)	(0.00060)	(0.00070)	(0.02330)	
2	0.00980	0.46070	0.38047	0.01846	0.02211	0.64478	43,050.85
	(0.00021)	(0.00632)	(0.00789)	(0.00060)	(0.00070)	(0.02305)	
3	0.00982	0.46009	0.37982	0.01849	0.02217	0.64618	43,050.84
	(0.00021)	(0.00633)	(0.00787)	(0.00060)	(0.00070)	(0.02318)	
4	0.00978	0.46105	0.37920	0.01842	0.02207	0.64717	43,050.88
	(0.00021)	(0.00631)	(0.00786)	(0.00060)	(0.00070)	(0.02317)	
5	0.00981	0.46029	0.38000	0.01845	0.02212	0.64583	43,050.86
	(0.00021)	(0.00632)	(0.00788)	(0.00060)	(0.00070)	(0.02310)	
6	0.00981	0.46006	0.37930	0.01849	0.02218	0.64636	43,050.85
	(0.00021)	(0.00632)	(0.00786)	(0.00060)	(0.00070)	(0.02322)	
7	0.00982	0.46004	0.37938	0.01848	0.02215	0.64717	43,050.85
	(0.00021)	(0.00632)	(0.00787)	(0.00060)	(0.00070)	(0.02323)	
8	0.00981	0.46014	0.38003	0.01838	0.02205	0.64731	43,050.87
	(0.00021)	(0.00631)	(0.00790)	(0.00060)	(0.00070)	(0.02307)	
9	0.00982	0.45994	0.37972	0.01846	0.02215	0.64641	43,050.85
	(0.00021)	(0.00632)	(0.00788)	(0.00060)	(0.00070)	(0.02315)	
10	0.00980	0.45982	0.37967	0.01839	0.02202	0.64849	43,050.87
	(0.00021)	(0.00630)	(0.00790)	(0.00060)	(0.00069)	(0.02313)	

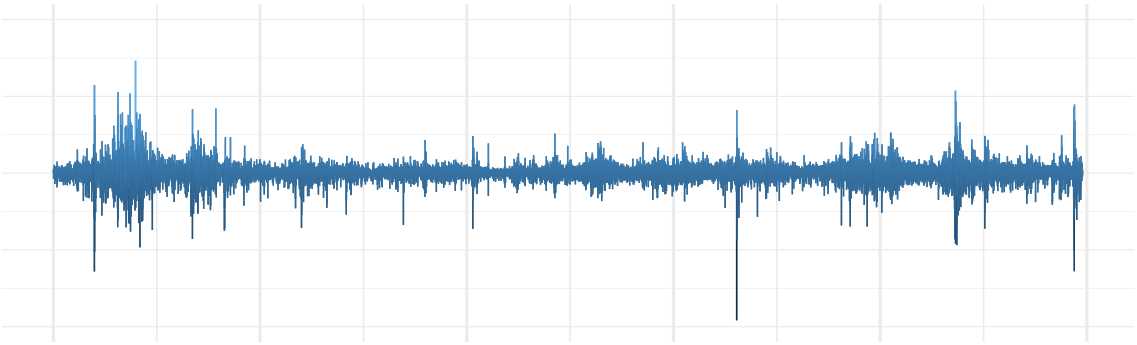


(a) Cumulative Log Return (Light Blue) and Five Days Volatility (Dark Blue)

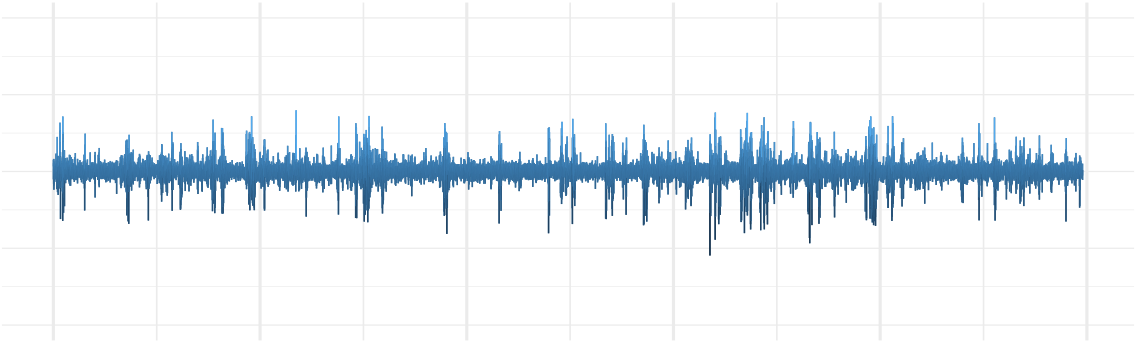


(b) Filtered States (Blue: Stable, Yellow: Middle, Red: Volatile)

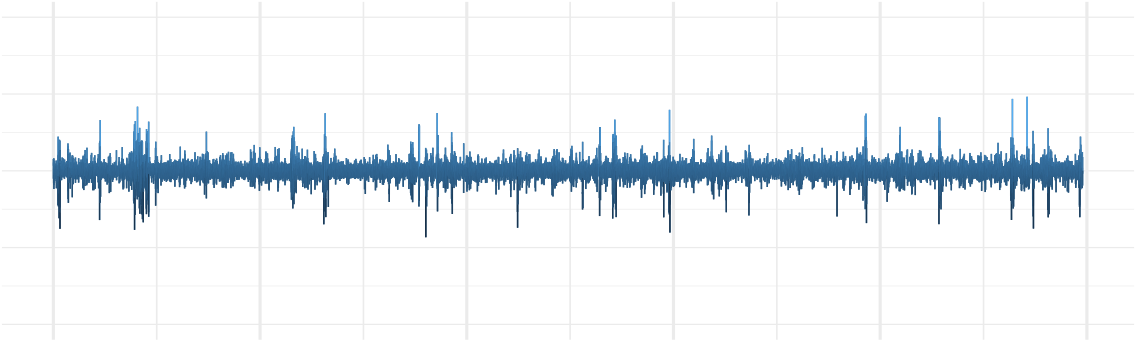
Figure 1: Filtered States by Time-varying Transition Probability Markov Switching Model (Three-state Model)



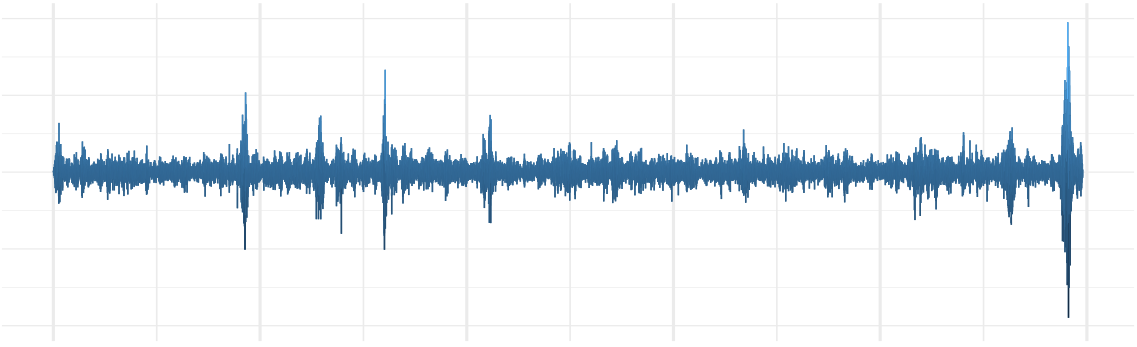
(a) Observed Time Series



(b) Simulation from TVTP MS Model



(c) Simulation from CTP MS Model



(d) Simulation from GARCH(1,1) Model

Figure 2: Simulated Log Return Series (Three-state Model)

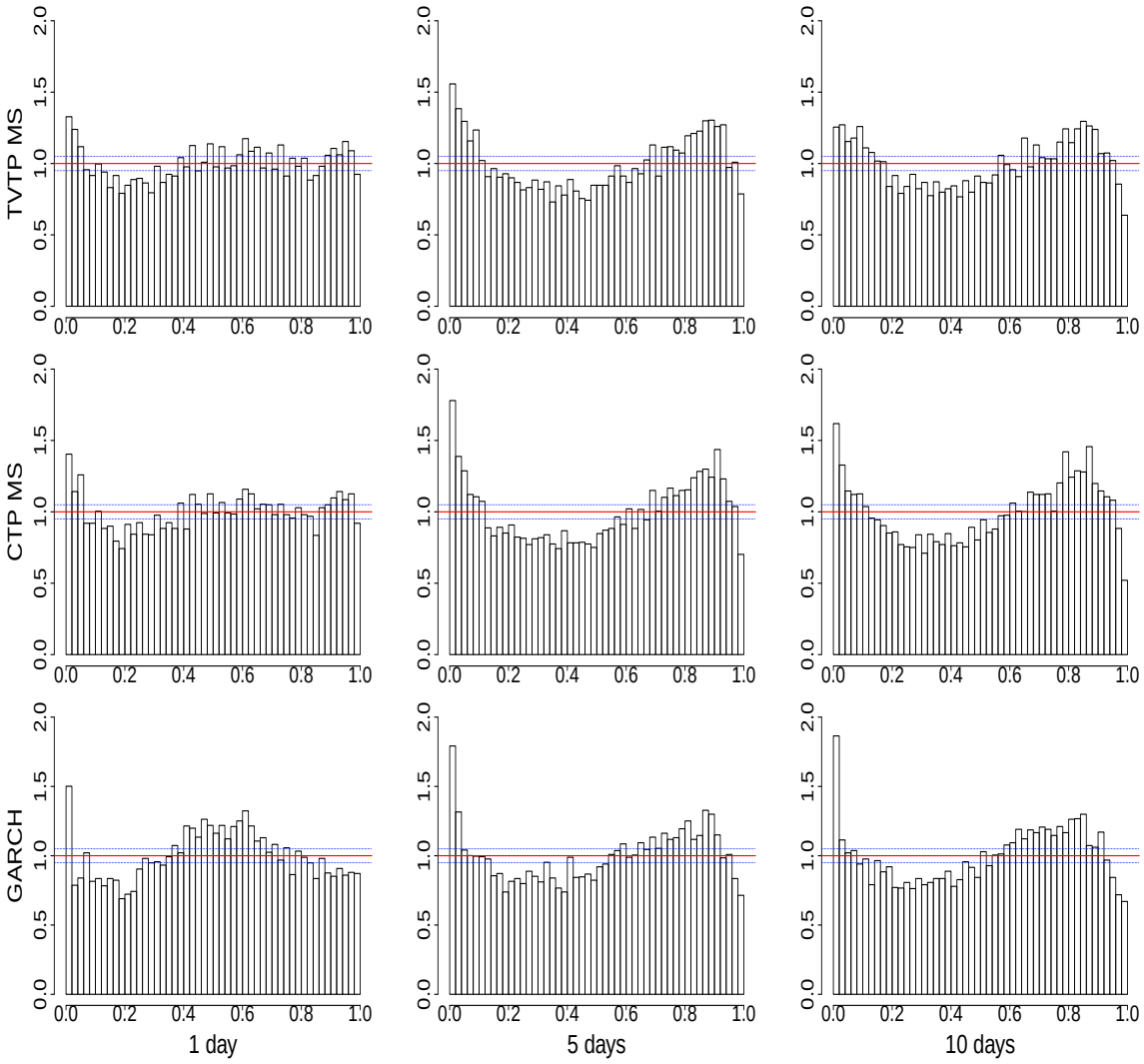


Figure 3: Probability Integral Transform (Three-state Model: Short Day Horizons)

The above graphs are the probability integral transform of the forecast made by three models, the Time-varying Transition Probability Markov Switching (TVTP MS) Model, the Constant Transition Probability Markov Switching (CTP MS) Model, and GARCH(1,1), from top to bottom. For each model, the three graphs from left to right correspond to the forecast for one day, five days, and ten days. In each graph, the red line indicates the uniform distribution, and the blue line indicates the +/- 5% level from the red line. If the model is correctly specified, a produced probability integral transform becomes close to uniform distribution (red line).

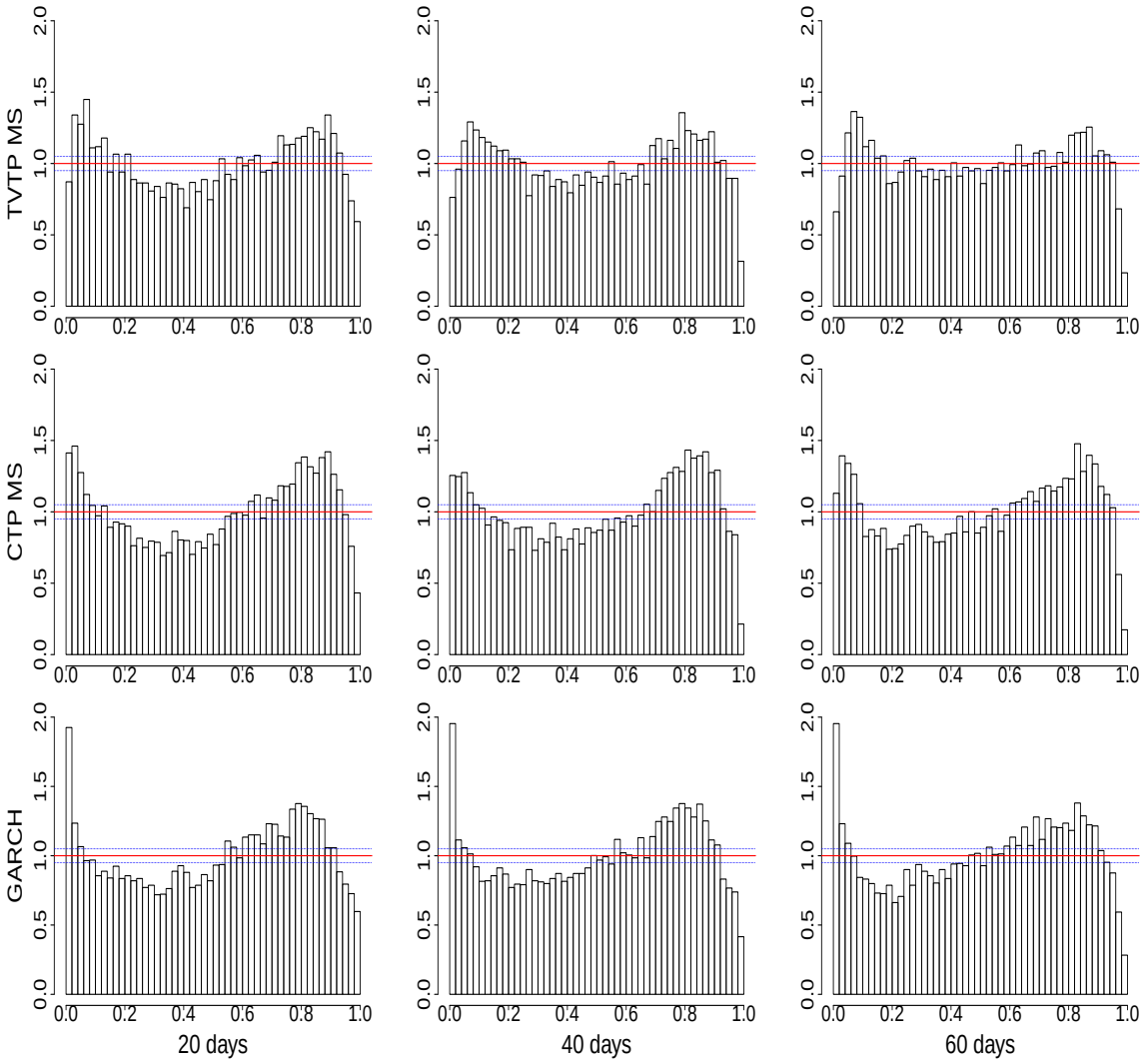
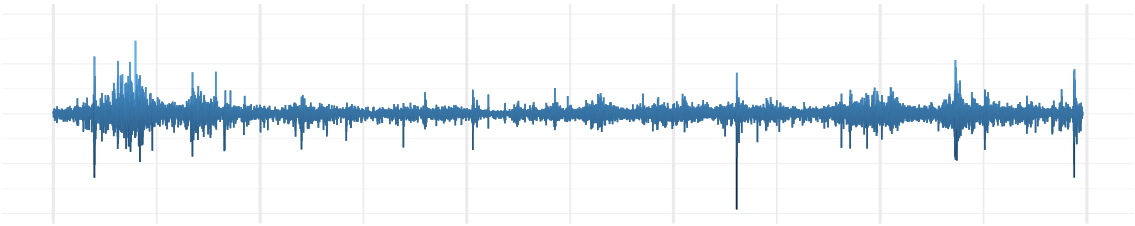
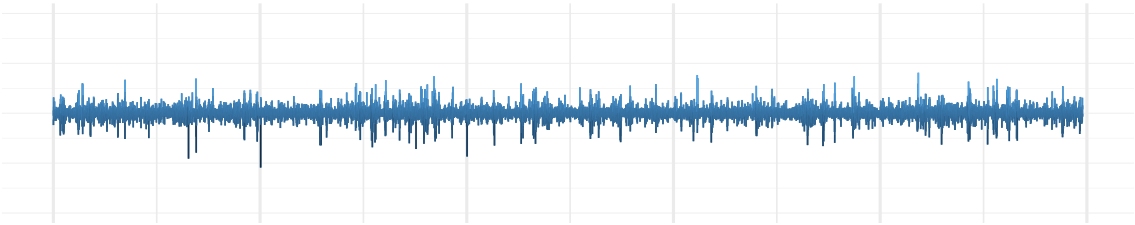


Figure 4: Probability Integral Transform (Three-state Model: Long Day Horizons)

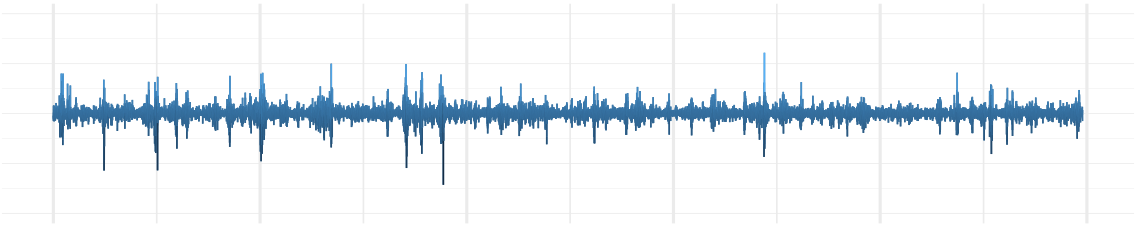
The above graphs are the probability integral transform of the forecast made by three models, the Time-varying Transition Probability Markov Switching (TVTP MS) Model, the Constant Transition Probability Markov Switching (CTP MS) Model, and GARCH(1,1), from top to bottom. For each model, the three graphs from left to right correspond to the forecast for twenty day, forty days, and sixty days. In each graph, the red line indicates the uniform distribution, and the blue line indicates the +/- 5% level from the red line. If the model is correctly specified, a produced probability integral transform becomes close to uniform distribution (red line).



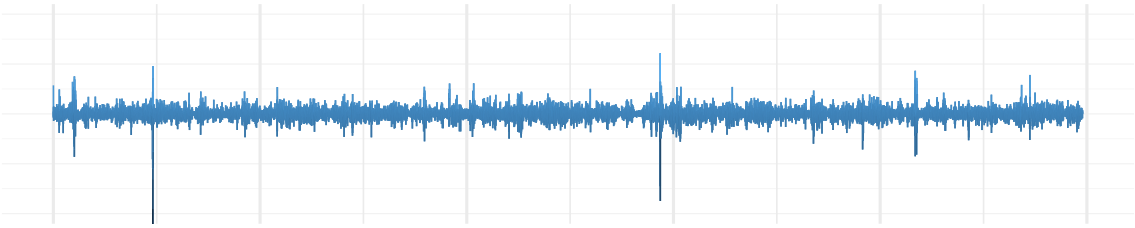
(a) Observed Time Series



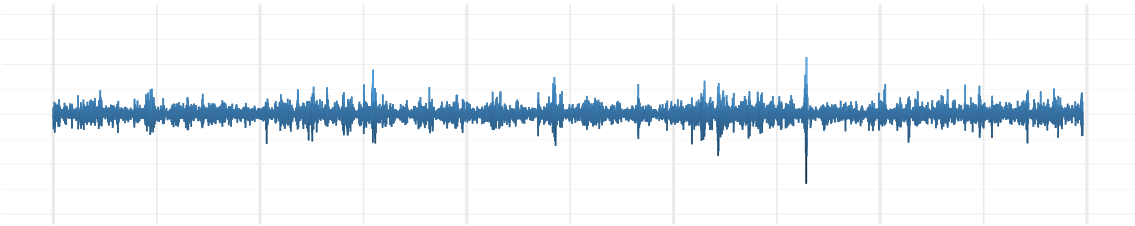
(b) Simulation (k=1)



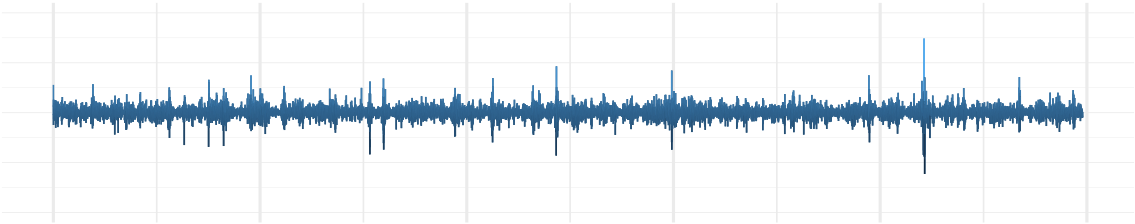
(c) Simulation (k=2)



(d) Simulation (k=3)

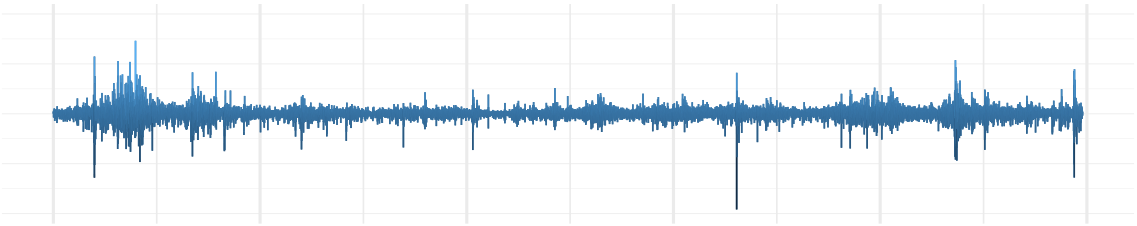


(e) Simulation (k=4)

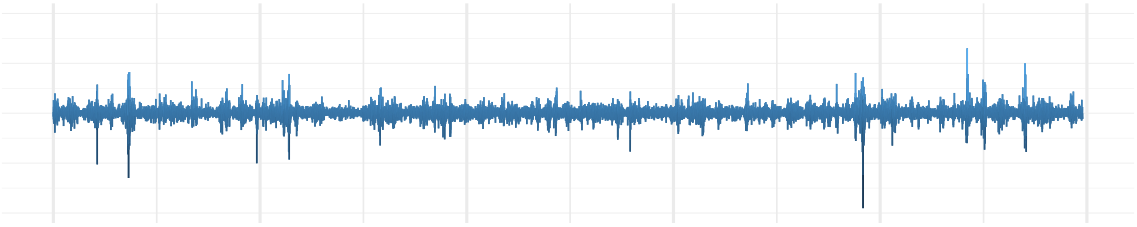


(f) Simulation (k=5)

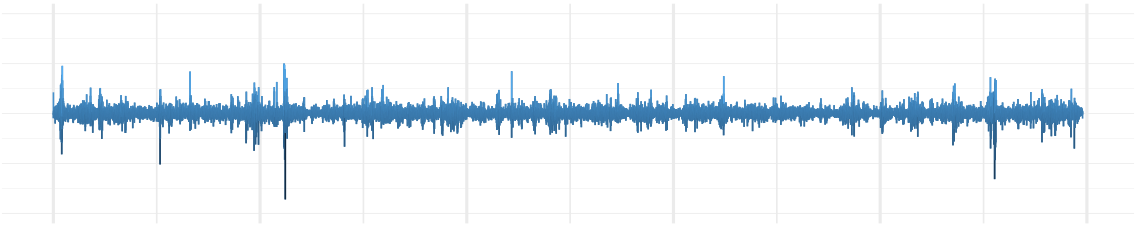
Figure 5: Simulated Log Return Series (Multi-state Model)



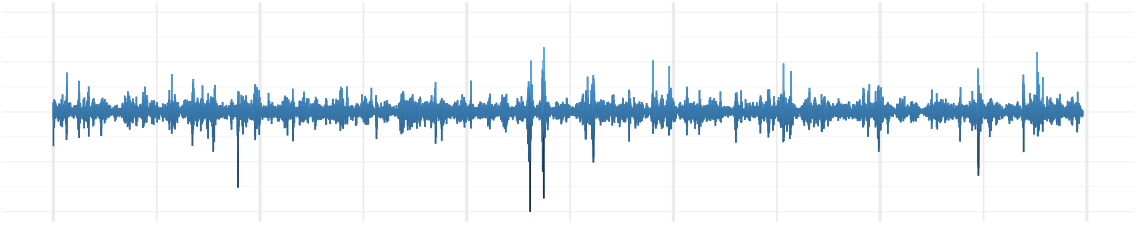
(a) Observed Time Series



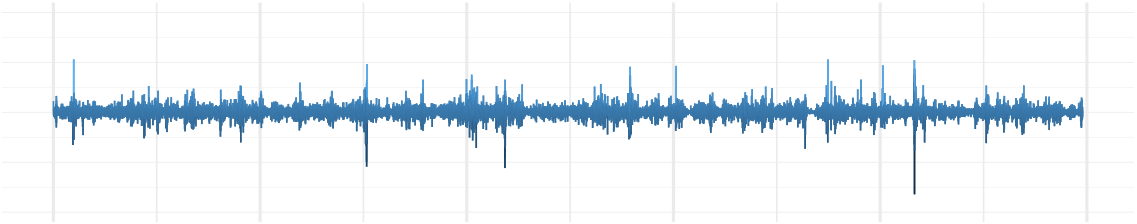
(b) Simulation (k=6)



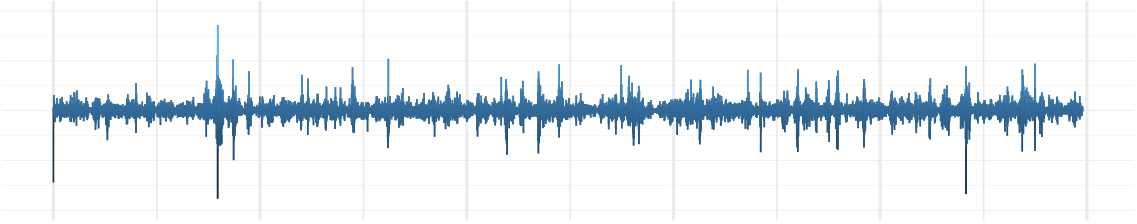
(c) Simulation (k=7)



(d) Simulation (k=8)



(e) Simulation (k=9)



(f) Simulation (k=10)

Figure 6: Simulated Log Return Series (Multi-state Model)

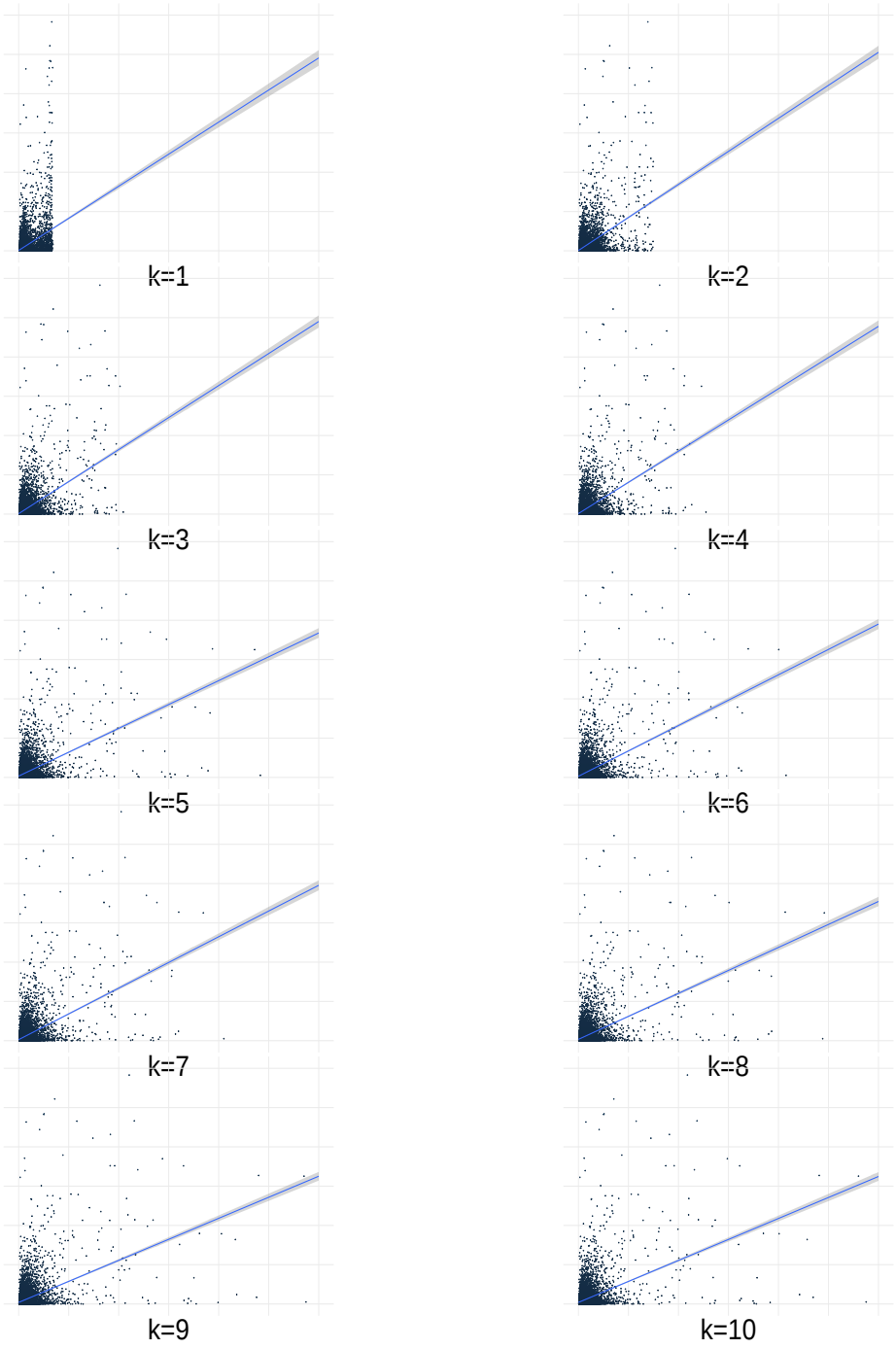


Figure 7: Mincer-Zarnowitz Regression (Multi-state Model: One Day Horizon)

The above graphs are the graphical result of Mincer-Zarnowitz regression on the forecast made by changing k from 1 to 10. In each graph, the x-axis is the realized volatility forecast by the model, and the y-axis is the contemporaneous realized volatility from the historical return. Every graph has the same scale of x-axis and y-axis. The blue diagonal line is the regression line. The graph starts with $k = 1, 2, ...$ from the top left to $k = 10$ to the bottom right.

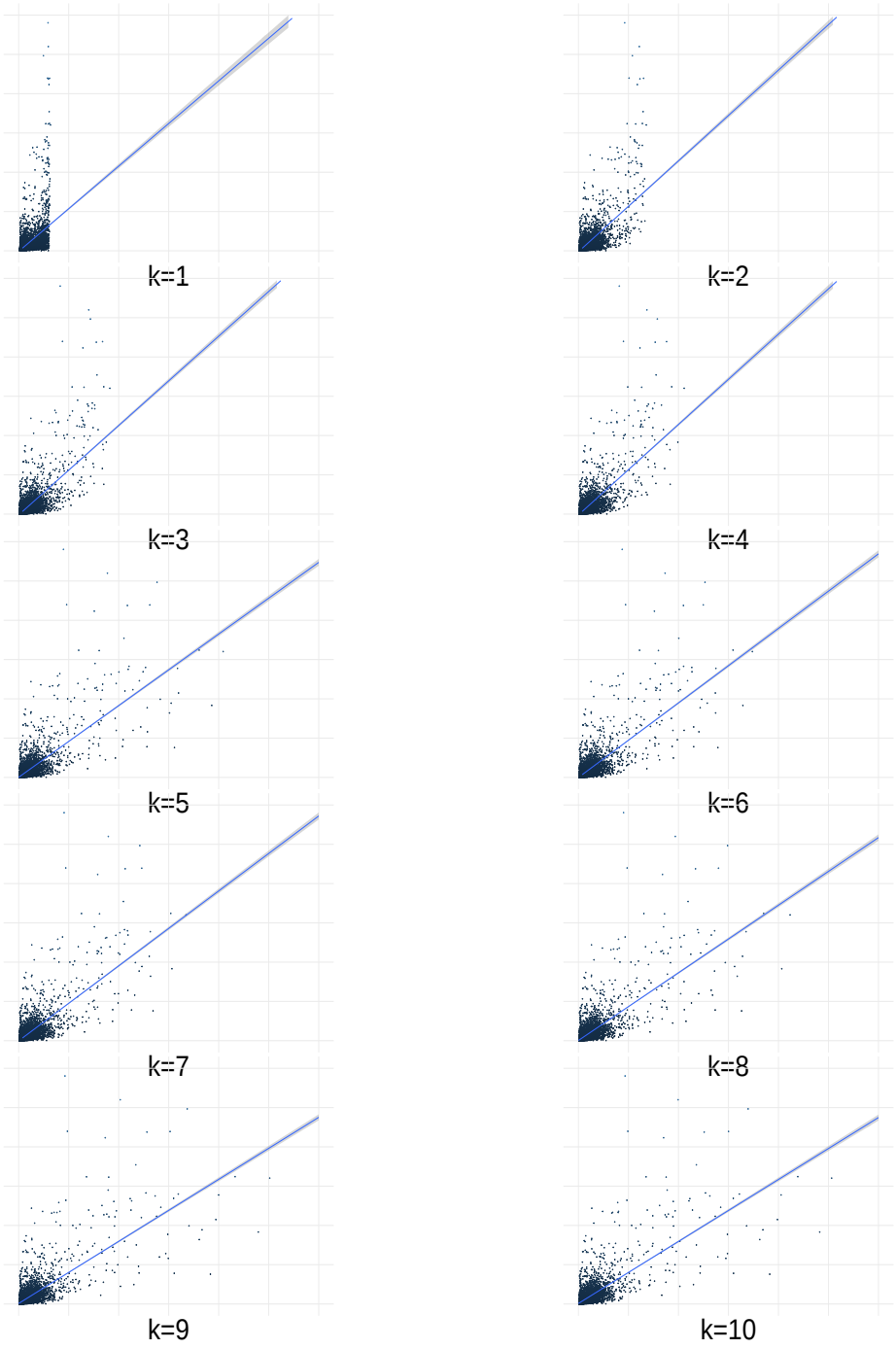


Figure 8: Mincer-Zarnowitz Regression (Multi-state Model: Five Day Horizon)

The above graphs are the graphical result of Mincer-Zarnowitz regression on the forecast made by changing k from 1 to 10. In each graph, the x-axis is the realized volatility forecast by the model, and the y-axis is the contemporaneous realized volatility from the historical return. Every graph has the same scale of x-axis and y-axis. The blue diagonal line is the regression line. The graph starts with $k = 1, 2, ...$ from the top left to $k = 10$ to the bottom right.

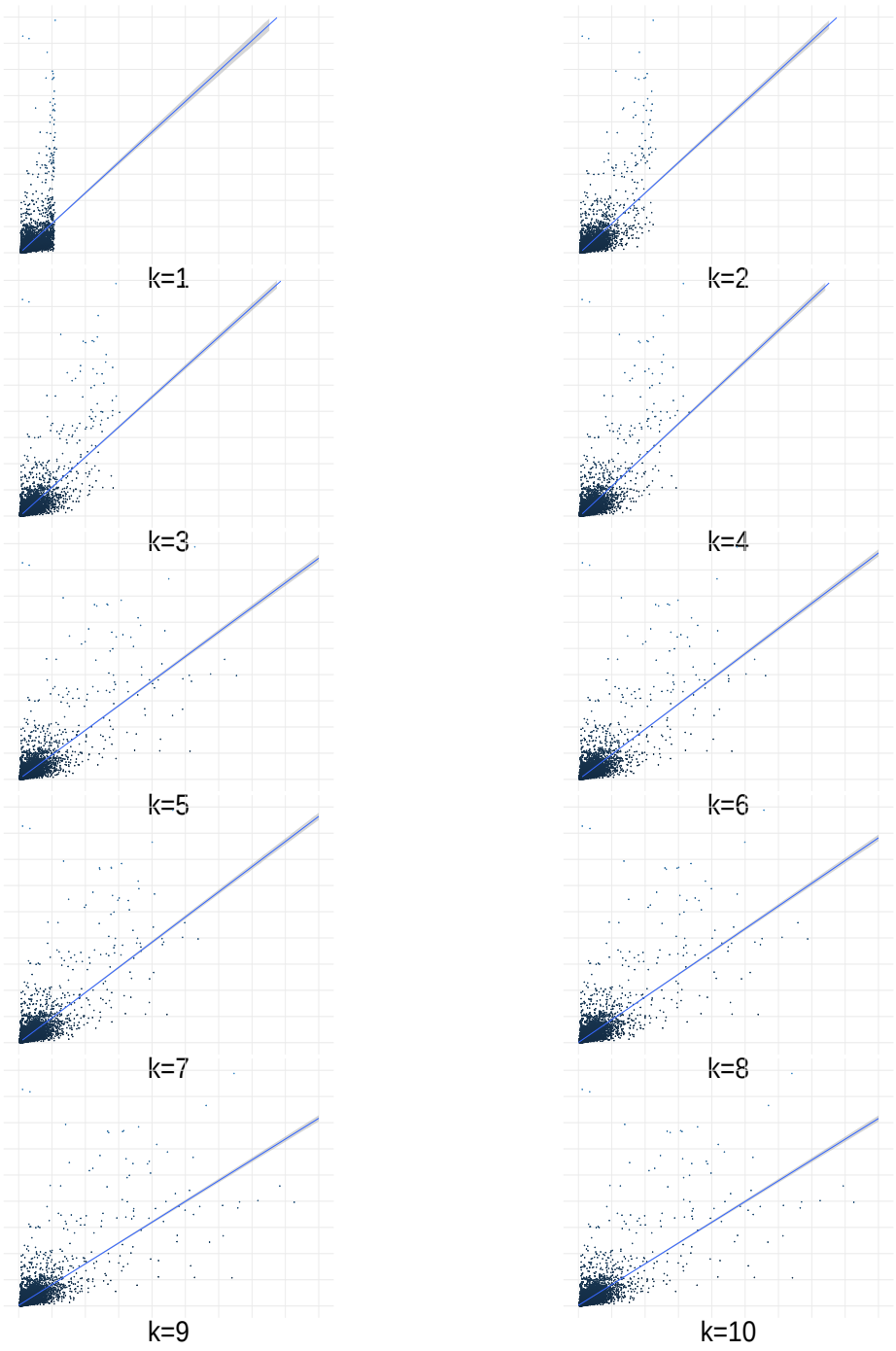


Figure 9: Mincer-Zarnowitz Regression (Multi-state Model: Ten Day Horizon)

The above graphs are the graphical result of Mincer-Zarnowitz regression on the forecast made by changing k from 1 to 10. In each graph, the x-axis is the realized volatility forecast by the model, and the y-axis is the contemporaneous realized volatility from the historical return. Every graph has the same scale of x-axis and y-axis. The blue diagonal line is the regression line. The graph starts with $k = 1, 2, ...$ from the top left to $k = 10$ to the bottom right.

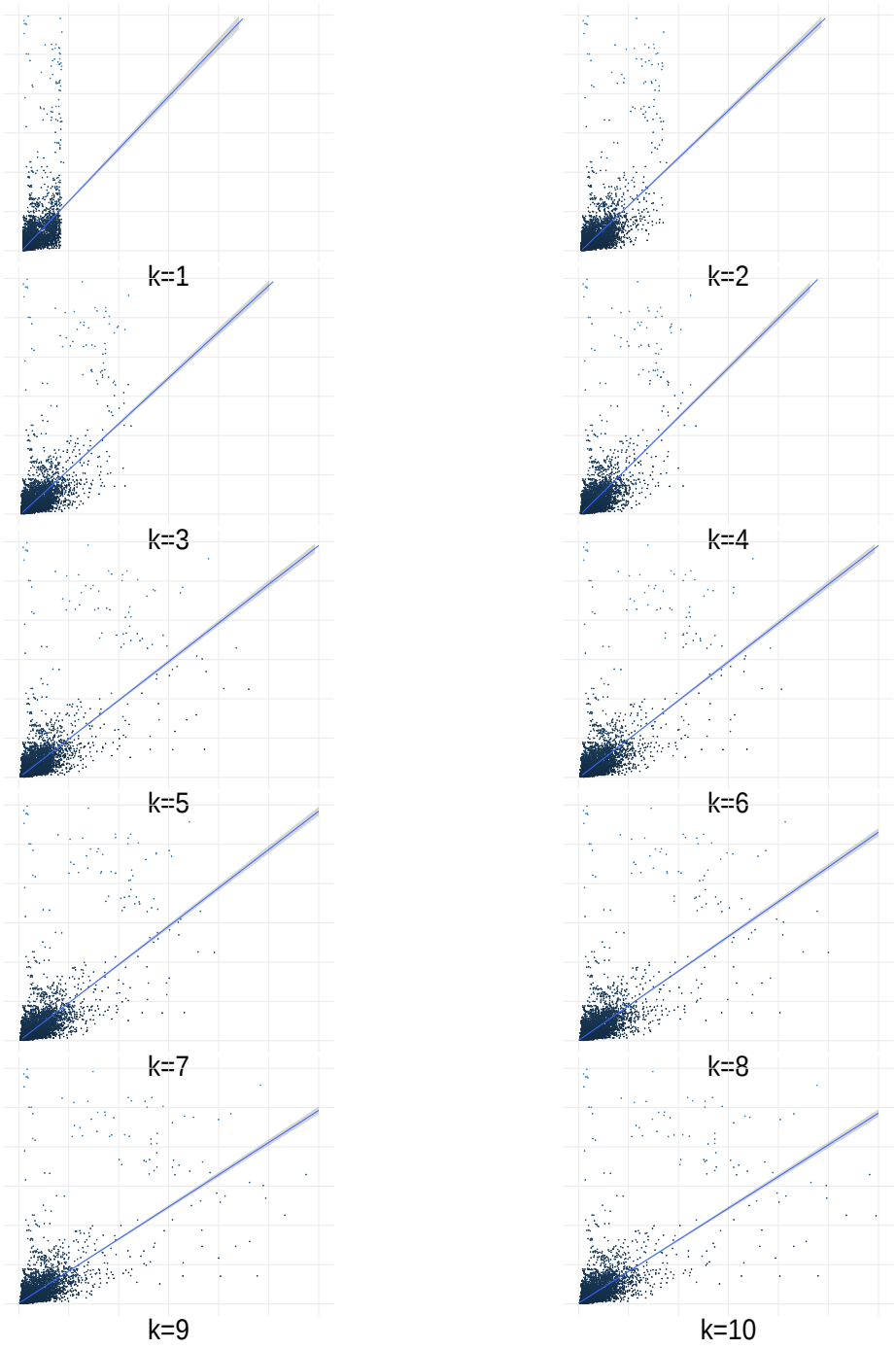


Figure 10: Mincer-Zarnowitz Regression (Multi-state Model: Twenty Day Horizon)

The above graphs are the graphical result of Mincer-Zarnowitz regression on the forecast made by changing k from 1 to 10. In each graph, the x-axis is the realized volatility forecast by the model, and the y-axis is the contemporaneous realized volatility from the historical return. Every graph has the same scale of x-axis and y-axis. The blue diagonal line is the regression line. The graph starts with $k = 1, 2, ...$ from the top left to $k = 10$ to the bottom right.

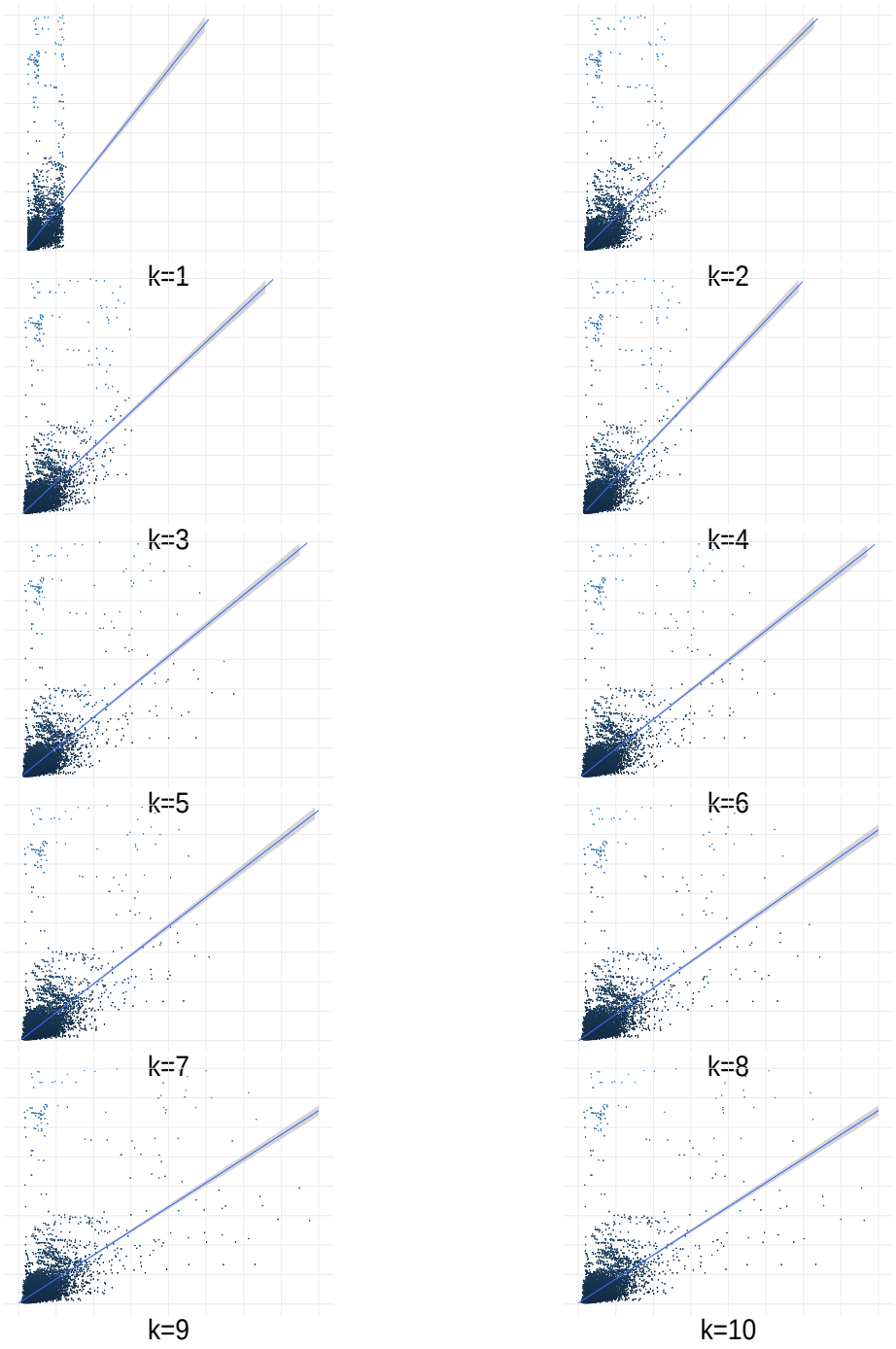


Figure 11: Mincer-Zarnowitz Regression (Multi-state Model: Forty Day Horizon)

The above graphs are the graphical result of Mincer-Zarnowitz regression on the forecast made by changing k from 1 to 10. In each graph, the x-axis is the realized volatility forecast by the model, and the the y-axis is the contemporaneous realized volatility from the historical return. Every graph has the same scale of x-axis and y-axis. The blue diagonal line is the regression line. The graph starts with $k = 1, 2, ...$ from the top left to $k = 10$ to the bottom right.

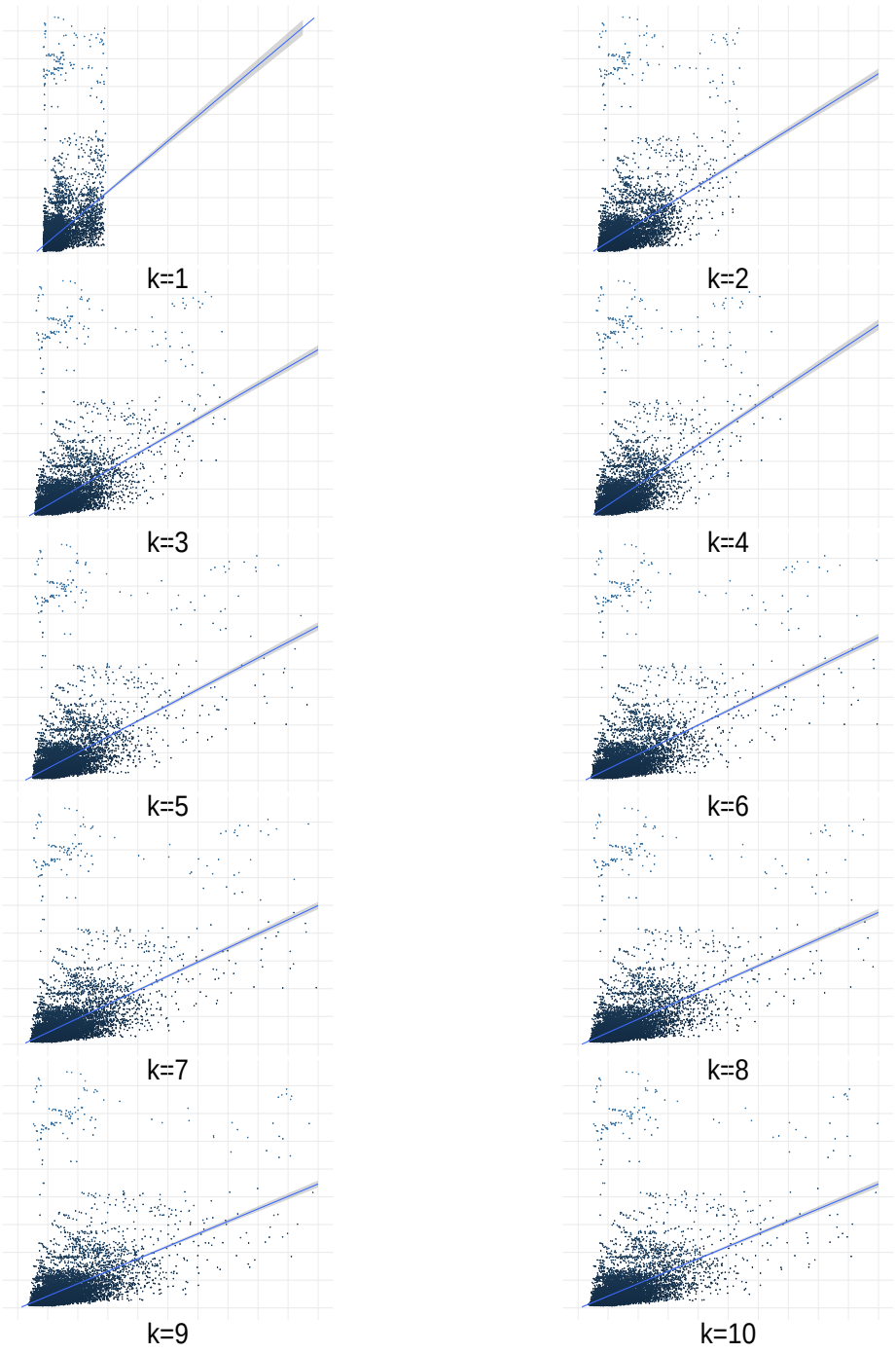


Figure 12: Mincer-Zarnowitz Regression (Multi-state Model: Sixty Day Horizon)

The above graphs are the graphical result of Mincer-Zarnowitz regression on the forecast made by changing k from 1 to 10. In each graph, the x-axis is the realized volatility forecast by the model, and the the y-axis is the contemporaneous realized volatility from the historical return. Every graph has the same scale of x-axis and y-axis. The blue diagonal line is the regression line. The graph starts with $k = 1, 2, ...$ from the top left to $k = 10$ to the bottom right.

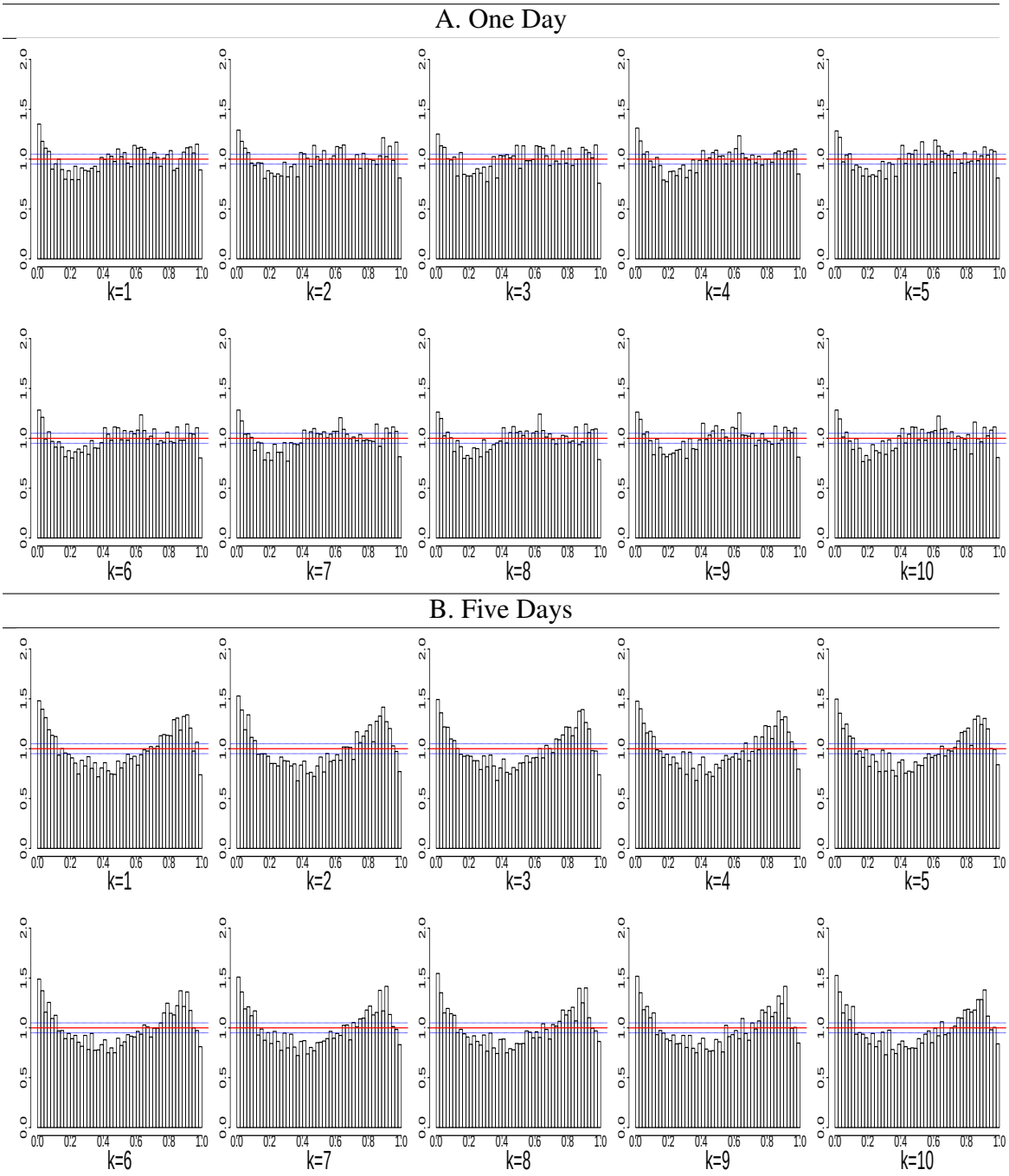


Figure 13: Probability Integral Transform (Multi-state Model: Short Horizons)

The above graphs are the probability integral transform of the forecast made by changing k from 1 to 10. Panel A is a one-day forecast, and panel B is a five-day forecast. The red line indicates the uniform distribution in each graph, and the blue line indicates the +/- 5% level from the red line. If the model is correctly specified, a produced probability integral transform becomes close to uniform distribution (red line).

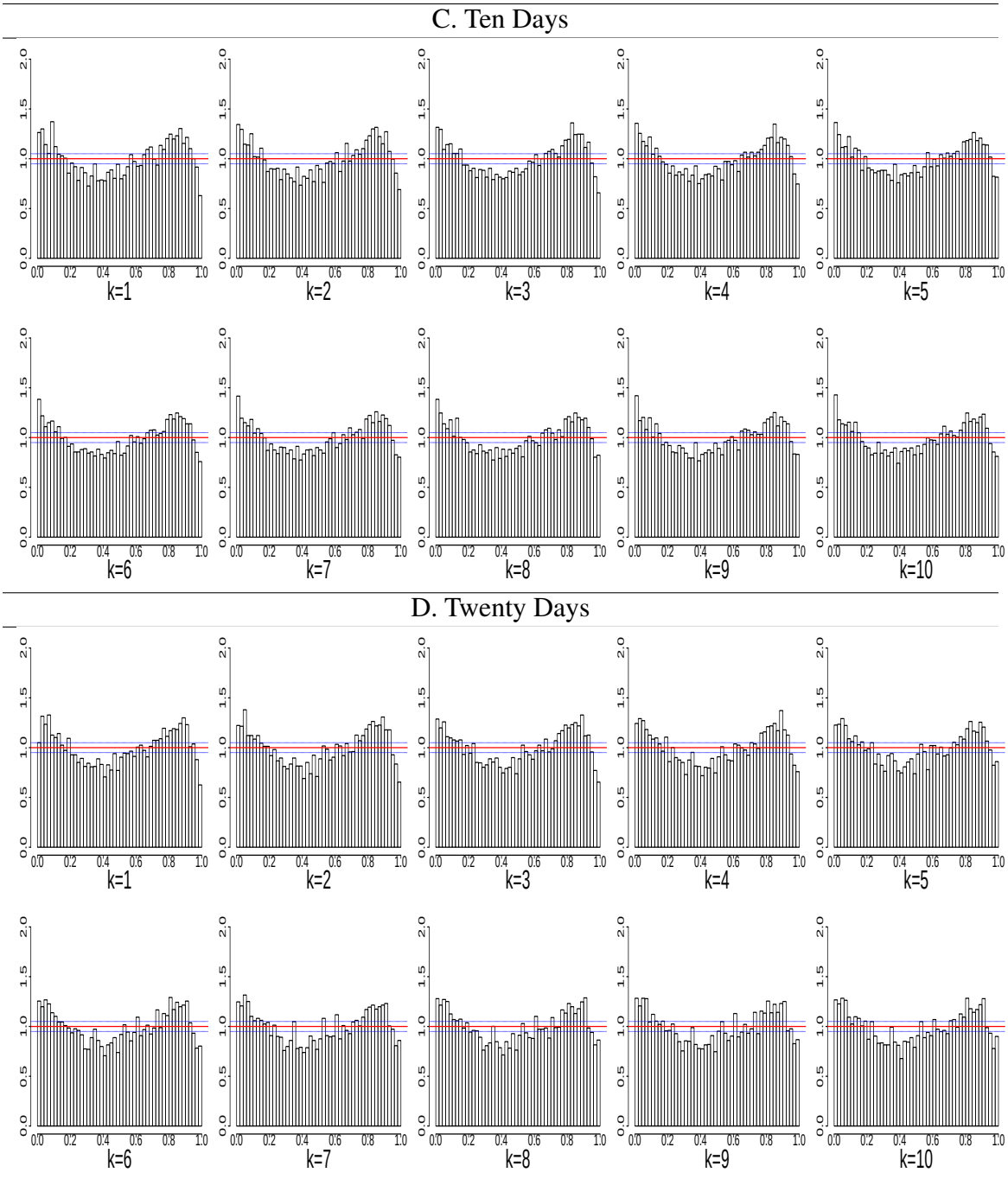


Figure 14: Probability Integral Transform (Multi-state Model: Middle Horizons)

The above graphs are the probability integral transform of the forecast made by changing k from 1 to 10. Panel C is a ten-day forecast and panel D is a twenty-day forecast. For each model, the three graphs from left to right correspond to the forecast for one day, five days, and ten days. The red line indicates the uniform distribution in each graph, and the blue line indicates the +/- 5% level from the red line. If the model is correctly specified, a produced probability integral transform becomes close to uniform distribution (red line).

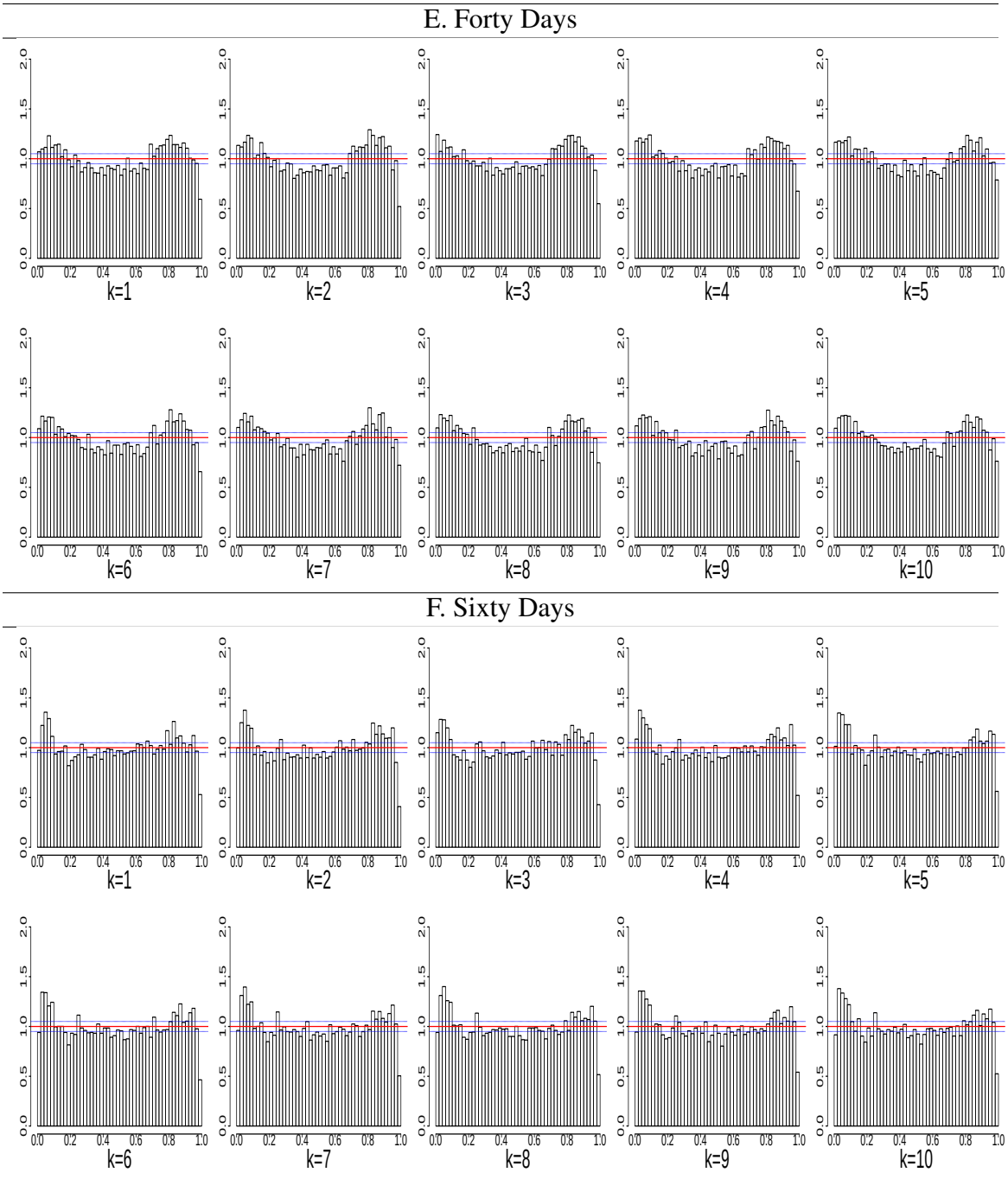


Figure 15: Probability Integral Transform (Multi-state Model: Long Horizons)

The above graphs are the probability integral transform of the forecast made by changing k from 1 to 10. Panel E is a forty-day forecast and panel F is a sixty-day forecast. For each model, the three graphs from left to right correspond to the forecast for one day, five days, and ten days. The red line indicates the uniform distribution in each graph, and the blue line indicates the +/- 5% level from the red line. If the model is correctly specified, a produced probability integral transform becomes close to uniform distribution (red line).