Two-State Markov Switching Volatility Model for Ultra-High-Frequency Data of JGB Futures*

Soo Nam Park** · Young-Jae Kim***

This paper specifies two-state Markov-switching volatility models and investigates the volatility behavior of the ultra-high-frequently observed returns on Japanese government bond futures transaction. In addition, we test the duration and volume effects on transition probabilities with a time-varying probability model. Our main findings are as follows: First, MS-GARCH models are very effective to reduce the autocorrelation of volatility, since the Ljung-Box statistics for squared standardized residuals of the models are dramatically reduced and present significantly smaller values in contrast to the single-regime GARCH model. Second, the volatilities of MS-GARCH models respond to new information more sensitively than those of the single-regime model. Third, the duration decreases volatility mainly by reducing the transition probability from high-variance regime to high-variance regime in the time-varying transition probability model, while the trading volume decreases both transition probabilities so that the transactions lead to a shift from one regime to another.

JEL Classification: C58, G13
Keywords: Markov-Switching model, UHF data, Japanese Government Bond Futures, duration

---

* Received August 8, 2001. Accepted November 3, 2001. This work was supported by the Korea Research Foundation Grant funded by the Korean Government (KRF-2009-32A-B00032).
** Department of Economics, Pusan National University, E-mail: snpark@pusan.ac.kr
*** Author for correspondence, Department of Economics, Pusan National University, San 30, Jangjeon-dong, Guemjeong-gu, Busan, South Korea, 609-735, Tel: +82-51-510-2539, Fax: +82-51-581-3143, E-mail: kimyj@pusan.ac.kr
1. INTRODUCTION

Modeling the price volatility of an asset is very important in finance since the volatility represents the risk of an asset and a determinant of its price. Recently, since intraday transaction data in financial markets are increasingly available, studies examining these data have appeared quite often. In addition, many studies have investigated nonlinearity on high-frequency returns.

This paper examines the volatility behavior of ultra-high-frequently observed returns on Japanese government bond (JGB) futures transactions, using two-state Markov-switching models that allow heterogeneous variances, and demonstrates a time-varying transition probability of the volatility by incorporating the effects of duration and volume.

It is well known that high-frequency asset returns are nonlinear and persistent, and suddenly change in variance. Regarding the futures market, Hsieh (1995) suggests that most returns from futures are nonlinear since high-frequency futures returns are not autocorrelated but their absolute values are strongly autocorrelated. McMillan and Speight (2002) show that futures returns are nonlinear and have conditional heteroskedasticity, applying STAR-EGARCH and STAR-CGARCH models to the high-frequency data of Long Gilt futures transactions. Shi and Lee (2008), using time domain and frequency domain methods, report that high-frequency data of JGB futures transaction have persistent volatility and support the mixed-distribution-hypothesis (MDH).

This paper models futures returns rather than spot returns, as Bansal et al. (2010) did, for two reasons. First, bond futures contracts are more openly traded than spot bonds traded over the counter. Thus, as Bansal et al. (2010) points out, the futures returns mitigate potential microstructure-related measurement concerns. Second, the futures contracts are derivatives in zero net supply with a maturity of a few months. Thus, Bansal et al. (2010) argue that the trading volume in contracts can better be characterized as having a direct relation to hedging demand over a specific and modest horizon.
This paper can be featured as analyzing ultra-high-frequently observed data, which are defined as the irregularly spaced records of transactions in a financial market. Using ultra-high-frequency (UHF) data, intraday volatility can be exquisitely analyzed, in contrast to high-frequency (HF) data. For example, suppose the futures price rises 1 tick during 1 minute 30 seconds. In this case, the HF data, which are recorded in 1-minute units, distort the price volatility, indicating that no price change occurs during the first 1 minute and the price rise of 1 tick occurs during the second 1 minute. However, UHF records do not cause this distortion.

Sometimes, futures price volatility may exhibit nonlinear behavior, so that it suddenly increases or decreases. One possible explanation for this nonlinearity of futures volatility is the existence of heterogeneous traders. For example, if high-risk traders are concentrated, then futures volatility will suddenly increase, while an increase in low-risk traders will reduce the volatility. We consider the heterogeneous futures volatility as a two-state Markov-switching process, which is classified into high-variance and low-variance regimes. Thus, we characterize the volatility in the UHF data of JGB futures transactions by this model.

Engle and Russell (1998) developed an autoregressive conditional duration (ACD) model to analyze UHF data of an asset market. Engle (2000) incorporated the conditionally expected duration, estimated by the ACD model, into the GARCH model, in order to build a model on the conditional variance of UHF data. McCulloch and Tsay (2001), using the threshold ACD(1, 1) model, showed that UHF financial data are characterized by nonlinearity. Although ACD models well characterize UHF data, we adopt GARCH-type models to analyze JGB futures transactions in order to directly estimate regime-switching variances. Cai (1994) and Hamilton and Susmel (1994) capture the abrupt change in the volatility of asset returns by modeling regime-switching in variance, applying the Markov chain. Dueker (1997) classified Markov-switching volatility models into four specifications. Gray (1996) generalized the Markov-switching volatility model and further estimated the conditional variance series of US Treasury bill rates. Klaassen
460

Soo Nam Park - Young-Jae Kim


Our work prepares the data through a two-stage procedure, and adopts Gray’s (1996) methodology to analyze this adjusted data to model the volatility. This paper differs from previous works in the following points. First, we analyze UHF data on JGB futures, not observed by regular time intervals, measuring the variance of UHF returns per unit time as Engle (2000) did. Second, we capture the sudden changes in the volatility of UHF returns, specifying Markov-switching models to characterize the heterogeneous volatility. Third, we suggest a model that reflects time-varying transition probability by duration and trading volume. Especially, our TVP model has the merit that the relation between conditional variance and duration could be nonlinearized, while Engle (2000)’s UHF-GARCH implicitly supposes linearized relation.

The next section specifies the models and introduces the estimation methodology. The third section describes the data used for the empirical analysis and explains the data preparation procedure. The fourth section presents the estimation results, followed by the conclusions in the final section.

2. MODEL AND METHOD

2.1. Distributional Assumptions

We start with the assumption that the futures price \( P_t \) moves as a geometric Brownian motion,

\[
P_t \sim \text{Geometric Brownian Motion}.
\]
Let $i = 0$ be the first futures transaction, the subscript indicating the transaction that causes the $i$th price change in a futures market. In addition, let the duration $d_i$ be the $i$th time interval between the $(i-1)$th and $i$th price changes; we assume it is an exogenous variable. Then, the log-return $R_i$ between the $(i-1)$th and $i$th price changes is normally distributed as follows:

$$R_i = \log(P_i / P_{i-1}) \sim N(d_i m_i, d_i h_i),$$

where $m_i$ and $h_i$ represent the conditional expectation of the return and the conditional variance per unit time, respectively, given the information set $\Omega_{i-1}$, which includes the history of all the transactions until the $(i-1)$th transaction.

$$m_i \equiv E\left(\frac{R_i}{d_i} \bigg| \Omega_{i-1}\right), \quad h_i \equiv \text{Var}\left(\frac{R_i}{d_i} \bigg| \Omega_{i-1}\right).$$

Hence, the log-return between the two price changes is expressed by the sum of its expectation $d_i m_i$ and the error term $\sqrt{d_i h_i} \varepsilon_i$.

$$R_i = d_i m_i + \sqrt{d_i h_i} \varepsilon_i, \quad \varepsilon_i \sim N(0, 1).$$

In the above, $\varepsilon_i$ denotes the standardized error, which is distributed as standard normal.

Now, let the futures return be measured again per square root of time.

$$r_i \equiv \frac{R_i}{\sqrt{d_i}}.$$

Then, the re-defined futures return $r_i$ can be expressed as follows:
where $\mu_i \equiv E(r_i | \Omega_{i-1})$ is the conditional expectation of the just measured return, given the previous one-step information set; the new disturbance $u_i \equiv \sqrt{h_i} \varepsilon_i$ should be normally distributed as $N(0, h_i)$, whose conditional variance is $h_i$, since the standardized error $\varepsilon_i$ is a random variable assumed as a standard normal distribution. Thus, the futures return $r_i$ should be a normal distribution whose conditional expectation is $\mu_i$ and conditional variance $h_i$, given the information set $\Omega_{i-1}$.

\[
r_i | \Omega_{i-1} \sim N(\mu_i, h_i).
\]

In the above discussion, we could identify the volatility per unit time as $h_i$, since $h_i$ is the variance of $r_i$ per unit time, following Engle (2000).

### 2.2. Assumptions on the Regimes, Mean, and Variance

A natural approach to modeling economic time series with nonlinear models seems to be to define different states of the world, or regimes, and to allow for the possibility that the dynamic behavior of economic variables depends on the regime that occurs at any given point in time (Priestley, 1980, 1988). Some regime-switching models assume that the regime cannot actually be observed but is determined by an underlying unobservable stochastic process. This implies that one can never be certain that a particular regime has occurred at a particular point in time, but can only assign probabilities to the occurrence of the different regimes (Franses and Dijk, 2000). These models assume that the regime that occurs at a point of time $i$, with our notation, cannot be observed, as it is determined by an
Two-State Markov Switching Volatility Model for Ultra-High-Frequency Data of JGB Futures

unobservable process, which we denote as $S_i$.

In the case of only two regimes, $S_i$ can be simply assumed to take the values of 1 and 2. The most popular model in this type, which was proposed by Hamilton (1989), is the Markov-switching model, in which the process $S_i$ is assumed to be a first-order Markov-process. This implies that the current regime $S_i$ depends only on the regime of the previous step, $S_{i-1}$. Hence, the model is completed by defining the transition probabilities of moving from one state to the other:

$$P(S_i = 1 | S_{i-1} = 1) = p^{(11)},$$

$$P(S_i = 2 | S_{i-1} = 1) = 1 - p^{(11)},$$

$$P(S_i = 2 | S_{i-1} = 2) = p^{(22)},$$

$$P(S_i = 1 | S_{i-1} = 2) = 1 - p^{(22)}.$$

Thus, $p^{(j)}$ is equal to the probability that regime $j$ at the point of time $i - 1$ is followed by the same regime at the point of time $i$. This is to show that for the two-state Markov-switching model the unconditional probability of the regime 1 is given by (Hamilton, 1994)

$$p^* \equiv P(S_i = 1) = \frac{1 - p^{(22)}}{2 - p^{(11)} - p^{(22)}}. \quad (2)$$

The conditional probability of the regime 1 can be defined as follows, given the information set $\Omega_{i-1}$:

$$p_i^* \equiv P(S_i = 1 | \Omega_{i-1}).$$

To be parsimonious, let us assume that the conditional expectations of the return are identical in the two regimes and are generated by the $q$th-order
autoregressive process.

$$\mu_i = E(r_i | \Omega_{i-1}) = \mu_i^{(1)} = \mu_i^{(2)} = \phi_0 + \phi_1 r_{i-1} + \cdots + \phi_q r_{i-q}.$$  \hfill (3)

However, suppose the conditional variance of the return would be state-dependent and time-varying, given the information set $\Omega_{i-1}$, as follows:

$$h_i^{(j)} = \text{Var}(r_i | S_i = j, \Omega_{i-1}), \ j = 1, 2.$$  \hfill (4)

According to Gray’s (1996) manipulation, the two-state Markov-switching conditional variance is represented by the following equation:

$$h_i = \text{Var}(r_i | \Omega_{i-1}) = E(r_i^2 | \Omega_{i-1}) - \left[ E(r_i | \Omega_{i-1}) \right]^2$$

$$= p_i^* [ (\mu_i^{(1)})^2 + h_i^{(1)} ] + (1 - p_i^*) [ (\mu_i^{(2)})^2 + h_i^{(2)} ] - [ p_i^* \mu_i^{(1)} + (1 - p_i^*) \mu_i^{(2)} ]^2$$  \hfill (4)

In both equations (3) and (4), the superscript $(j)$ indicates the $j$th regime. In equation (4), the conditional variance $h_i$ seems to be similarly formed as the two-component model suggested by Ding and Granger (1996), which would be determined by the weighted summation of the variance $h_i^{(1)}$ in the first regime and the variance $h_i^{(2)}$ in the second regime, with the weight $p_i^*$. However, our model can be distinguished from that of Ding and Granger (1996) by the fact that the weight $p_i^*$ is a conditional probability and state-dependent, and thus time-varying.

2.3. Markov-Switching GARCH Models

Under the assumptions in the previous subsections, the conditional distribution of futures returns could be represented as follows:
Two-State Markov Switching Volatility Model for Ultra-High-Frequency Data of JGB Futures

(5)

\[ r_t | \Omega_{t-1} \sim \begin{cases} N(\mu_i, h_i^{(1)}) & \text{with prob. } p_i^* \\ N(\mu_i, h_i^{(2)}) & \text{with prob. } 1 - p_i^* \end{cases} \]

We specify the state-dependent conditional variance \( h_i^{(j)} \) in each regime as the popular GARCH(1,1) model in the following equation:

(6)

\[ h_i^{(j)} = h_i^{(0)} + \omega_i (u_{i-1}^2 - h_i^{(j)}) + \alpha_i (h_{i-1} - h_i^{(j)}), \quad j = 1, 2. \]

The GARCH(1,1) model on \( h_i^{(j)} \), as specified in equation (6), has the advantage that we can directly estimate the unconditional variance \( h_i^{(j)} \) instead of the constant term in the variance equation. Let us denote the Markov-switching GARCH(1,1) model, specified as expressions (5) and (6), as a Markov-switching GARCH constant probability (MS-GARCH-CP) model — such that the transition probabilities are constant.

Let \( f_i^{(j)} \) be the state-dependent conditional density of a disturbance \( u_i \). Then, with the conditional distribution of \( r_t \) as in equation (5), the density \( f_i^{(j)} \) is given as follows:

(7)

\[ f_i^{(j)} = f(u_i | S_i = j, \Omega_{t-1}; \theta) = \frac{1}{\sqrt{2\pi h_i^{(j)}}} \exp \left( -\frac{u_i^2}{2h_i^{(j)}} \right), \]

where \( \theta \) denotes the parameter vector. In addition, the conditional density of \( u_i \) should be expressed by the following equation, using the conditional probability \( p_i^* \) of the regime 1.

(8)

\[ f_i = f(u_i | \Omega_{t-1}; \theta) = p_i^* f_i^{(1)} + (1 - p_i^*) f_i^{(2)}. \]

Thus, the log-likelihood of the disturbance \( u_i \) is

(9)

\[ \log L = \sum_{i=1}^{N} \log(f_i) = \sum_{i=1}^{N} \log[p_i^* f_i^{(1)} + (1 - p_i^*) f_i^{(2)}]. \]
and we can obtain the estimators of parameters $\theta$ by maximizing the above log-likelihood. According to Gray (1996), in equations (5), (8), and (9), the conditional probability $p^*_i$ of the regime 1 can be obtained by the following equation:

$$p^*_i = P(S_i = 1 | \Omega_{t-1}; \theta) = \frac{p^{(11)}f^{(1)}_{i-1} + (1 - p^{(22)})(1 - p^{(2)})f^{(2)}_{i-1}}{p^{(11)}f^{(1)}_{i-1} + (1 - p^{(2)})f^{(2)}_{i-1}}. \quad (10)$$

On the other hand, the transition probability $p^{(j)}$ should just have a value within $[0,1]$ because it is only a “probability,” but in the actual estimation it is difficult to obtain an estimate of $p^{(j)}$ that satisfies this condition. Hence, we specify the transition probability in the form of a logistic function with a parameter $\gamma^{(j)}_0$ so that we can easily obtain an estimate of $p^{(j)}$ within $[0,1]$.

$$p^{(j)} = \frac{1}{1 + \exp(-\gamma^{(j)}_0)}. \quad (11)$$

The assumption on each transition probability could be more relaxed so that we can specify the transition probabilities to be also time-varying. Then, the time-varying transition probability could be denoted as $p^{(j)}_i$ with subscript $i$. We can specify this transition probability $p^{(j)}_i$ as a form of logistic function similar to equation (11), however, introducing exogenous variables as follows:

$$p^{(j)}_i = \frac{1}{1 + \exp[-x^{(j)}_{i-1}\gamma^{(j)}]} \quad (12)$$

where $x^{(j)}_{i-1}$ is the exogenous variables vector which varies $p^{(j)}_i$, and $\gamma^{(j)}$ is its coefficients vector. Adopting equations (5), (6), and the transition probability function, equation (12), let us denote the Markov-switching GARCH(1,1) model simply as a Markov-switching GARCH time-varying probability (MS-GARCH-TVP) model — such that the transition
probabilities are time-varying. Our MS-GARCH-TVP model, which replaces equation (11) with equation (12), can be estimated with equations (9), (10), and (12) as well as conditional probability $p_i^*$ of the regime 1, following equation (13), which substitutes for equation (10).

$$p_i^* = P(S_i = 1 | \Omega_{i-1}; \theta) = \frac{p_i^{(1)} p_{i-1}^{(1)} f_{i-1}^{(1)} + (1-p_i^{(2)}) (1-p_{i-1}^{(2)}) f_{i-1}^{(2)}}{p_{i-1} f_{i-1}^{(2)} + (1-p_{i-1}) f_{i-1}^{(2)}}. \quad (13)$$

Engle (2000) reports that the duration between price changes decreases the volatility per second, and supports Easley and O’Hara’s (1992) formulation. Most of the literature, which presents the theoretical background of the financial market microstructure, suggests that trading volumes involve information flows. Admati and Pfleiderer (1988) include heterogeneous traders, called informed traders, liquidity traders, and market makers. In particular, liquidity traders are classified as discretionary or non-discretionary types. In addition, numerous empirical research studies such as Tauchen and Pitts (1983), Karpoff (1987), Lamoureux and Lastrapes (1990), Jones et al. (1994), Xu and Wu (1999), and Huang and Masulis (2003) demonstrate that the relationship between volatility and volume is significant. Girma and Mougoue (2002) and Hadsell (2006) show significant relationships between volatility and volume in futures markets. Thus, we introduce duration and volume into equation (12) as variables that vary transition probabilities $p_i^{(b)}$, so that we specify $x_i = (1, d_i, v_i)$ and $\gamma_i^{(b)} = (\gamma_i^{(b)}, \gamma_i^{(b)}, \gamma_i^{(b)})$, where $v_i$ is the trading volume per second at the $i$–1th price $P_{i-1}^{*}$ during duration $d_i$.

3. THE DATA AND ADJUSTMENT

The ultra-high-frequency data of 10-year JGB futures transactions in the Tokyo Stock Exchange are taken from the electronic trading system of a brokerage company in Korea. The data cover the period from January 8 to
November 28, 2008. To avoid a regular pattern around the last and first trading day, we eliminated all transactions from March 3 to 21, from June 2 to 20, and from September 1 to 19. The regular trading hours of JGB futures are divided into three sessions every day. The morning session is from 9:00 to 11:00, the afternoon session from 12:30 to 15:00, and the evening session from 15:30 to 18:00 on each trading day. The raw data contain the records of transaction price, date, point of time, and volume. The transaction time is recorded in seconds.

In order to estimate conditional variances using UHF data, the data should go through a preparatory procedure given by Engle (2000). The procedure is divided into two stages. The first stage is a procedure for thinning the data, and the second stage is a procedure for the removal of diurnal patterns in the thinned data.

The first stage, thinning the data, is a procedure to pick out the points at which significant price change occurred in a price process. From the raw data, following Engle (2000), we remove only the points at which no change of price occurred, so that all the points that recorded price change of 1 tick or more are collected as sample observations. To put it concretely, our first stage of preparing the data for empirical analysis is as follows: First, we dropped any transaction that occurred outside regular trading hours. Second, we eliminated all the transaction data with zero duration. We treated these transactions as one single transaction at the last price of these transactions, summing up all the volumes. Third, the returns between transactions were computed as the difference of the log of the prices. Then, we eliminated the data with zero return and computed the duration and volume between price changes. Fourth, we erased the data of the initial price and transactions at this price in each trading session to exclude breaking time effect on the return and volume. After this first-stage preparing procedure, our sample contains

1) Transaction data on January 24 and 29, February 13, April 8 and 25, July 24 and 30, August 6 and 15, September 22 and 24, October 1, 6, and 7, and November 10, 17, 20, and 21 are omitted by the communication blackout of our computer system. However, this is a matter of small importance since our sample is remarkably large which contains 251,511 observations.
251,511 observations. After the first stage, we re-define the return dividing it by the square root of the duration to measure the volatility per unit time. Similarly, we define the volume between price changes dividing it by duration.

The second stage of the preparing procedures is a procedure to remove the diurnal patterns, what is called the time-of-day effect, in intraday transactions. It is a well-known fact that transactions exhibit a typical daily pattern over the course of the trading day. This repetition, which occurs periodically every day, may act on variables as a deterministic component, as the seasonality of monthly time series. In order to remove this feature from the data, the time series of a defined return, duration, and volume were diurnally adjusted.

According to Engle and Russell (1998), a preliminary variable $Y_i$ is expressed by the following multiplicative model:

$$ Y_i = C(t_i) \cdot y_i, $$

where $C(t_i)$ represents the diurnal pattern as the function of time $t_i$ when the $i$th price change occurs. $y_i$ is the ratio of preliminary variable $Y_i$ to the diurnal pattern $C(t_i)$, which represents the stochastic component of $Y_i$ to exclude the deterministic component. There are two methods to estimate the model of stochastic component $y_i$. In the one-step method, the model of the stochastic component $y_i$ and $C(t_i)$ are estimated simultaneously. In the two-step method, the model of $y_i$ is estimated after the estimation of $C(t_i)$. Zhang et al. (2001) argue that the difference between joint and two-step estimation becomes negligible when the sample size is large. Hence, we choose the two-step approach in order to pay attention to the stochastic component $y_i$.

For the diurnal adjustment, we regressed the absolute values of returns, durations, and volumes on a piecewise cubic spline according to equation (15) with knots at 9:00, 10:00, and 11:00 for the morning session, at 12:30, 13:30, 14:30, and 15:00 for the afternoon session, and at 15:30, 16:30, 17:30, and
18:00 for the evening session.

\[ Y_i = C_1(t_i) + C_2(t_i) + C_3(t_i) + e_i, \]  

(15)

where

\[ C_1(t_i) = \sum_{n=1}^{2} \left[ c_{n0} + c_{n1} (t_i - k_n) + c_{n2} (t_i - k_n)^2 + c_{n3} (t_i - k_n)^3 \right] I_n, \]

\[ C_2(t_i) = \sum_{n=1}^{2} \left[ c_{n0} + c_{n1} (t_i - k_n) + c_{n2} (t_i - k_n)^2 + c_{n3} (t_i - k_n)^3 \right] I_n, \]

\[ C_3(t_i) = \sum_{n=6}^{8} \left[ c_{n0} + c_{n1} (t_i - k_n) + c_{n2} (t_i - k_n)^2 + c_{n3} (t_i - k_n)^3 \right] I_n, \]

\[ I_n = \begin{cases} 1, & \text{if } t_i \in [k_n, k_{n+1}) \text{,} \\ 0, & \text{otherwise} \end{cases}, \quad n = 1, \ldots, 8. \]

In equation (15), \( Y_i \) is the preliminary variable of the \( i \)th absolute value of return, duration, or volume, which is obtained from the first stage of the preparing procedures, \( C_1(t_i) + C_2(t_i) + C_3(t_i) = C(t_i) \) is the diurnal pattern according to equation (14), and \( C_1(t_i), \ C_2(t_i), \text{ and } C_3(t_i) \) are the cubic splines for each trading session. The coefficients of cubic splines are restricted by the continuousness conditions of the cubic function, and the first and second derivatives in each trading session, but disconnection is allowed in the breaking time between the trading sessions. We measure the time point \( t_i \) and each knot \( k_n \) in seconds after midnight daily; thus, each knot in the regression has a value of \( k_1 = 32400 \) at 9:00, \( k_2 = 36000 \) at 10:00, \( k_3 = 45000 \) at 12:30, \( k_4 = 48600 \) at 13:30, \( k_5 = 52200 \) at 14:30, \( k_6 = 55800 \) at 15:30, \( k_7 = 59400 \) at 16:30, and \( k_8 = 63000 \) at 17:30 every day.

Figure 1 shows fitting curves of return, duration, and volume, which are estimated by cubic spline in each trading session. As shown in figure 1, the
Figure 1  Diurnal Pattern Curves of Return, Duration, and Volume

Notes: This figure shows the curves of intraday seasonality in the return, duration, and volume. Each curve represents the fitted series on the cubic spline which is restricted by the continuousness conditions of the regression function and the first and second derivatives in each trading session everyday, but allowed disconnection in breaking time between the sessions. The top panel contains the diurnal pattern of absolute return ($|r_i|$), the middle panel, duration ($d_i$), and the bottom panel, trading volume ($v_i$). In each panel, the first curve represents the diurnal pattern in the morning session, the second, in the afternoon session, and the third, in the evening session.

return and volume are characterized by typical $U$ patterns, and the duration is characterized by an inverse $U$ pattern in each trading session. We constructed the adjusted series by dividing each preliminary series by these spline forecasts.
Table 1  Summary Statistics

<table>
<thead>
<tr>
<th></th>
<th>Preliminary</th>
<th></th>
<th>Adjusted</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r_i$</td>
<td>$d_i$</td>
<td>$v_i$</td>
<td>$r_i$</td>
</tr>
<tr>
<td>Mean</td>
<td>-8.40E-08</td>
<td>15.88072</td>
<td>3.710221</td>
<td>-0.002565</td>
</tr>
<tr>
<td>Median</td>
<td>-3.98E-06</td>
<td>5</td>
<td>1.6</td>
<td>-0.149069</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.000955</td>
<td>2381</td>
<td>283</td>
<td>35.63030</td>
</tr>
<tr>
<td>Minimum</td>
<td>-0.001250</td>
<td>1</td>
<td>0.000420</td>
<td>-30.55585</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>5.01E-05</td>
<td>39.24233</td>
<td>6.550856</td>
<td>1.254460</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.069246</td>
<td>10.58939</td>
<td>6.121331</td>
<td>-0.065201</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>10.54546</td>
<td>220.1405</td>
<td>78.48163</td>
<td>14.06697</td>
</tr>
<tr>
<td>Observations</td>
<td></td>
<td></td>
<td></td>
<td>251,511</td>
</tr>
</tbody>
</table>

Notes: Adjusted Series = Preliminary Series + Spline Forecast. All the series represent data prepared for the empirical analysis. That is, $r_i$ denotes the log return divided by square root of duration. $d_i$ is the duration between price changes, and $v_i$ denotes trading volume per second.

Table 1 reports summary statistics of the data. The left panel of the table presents the statistics of the preliminary series obtained through the first stage of preparing procedures. The mean return of JGB futures, measured by square root seconds, is close to zero. The duration between price changes averages 15.9 seconds, and its range and standard deviation show its wide fluctuation. Transaction volume has a mean of 3.7 contracts per second, and its range and standard deviation indicate their wide fluctuation, too.

The right panel of the table reports summary statistics of the adjusted series by cubic spline. It is desirable that the mean values of adjusted series are equal to one, except for the returns that can be negative, since they are the ratios of the original series to the diurnal pattern. The mean values of adjusted duration and volume are close to one, so that the adjustment by cubic spline seems to be adequate. The adjusted return shows a range of $-30.6 \sim 35.6$ and a standard deviation of 1.3. These facts suggest that the
volatility of the stochastic component of intraday returns is very large. However, the reduced ranges and standard deviations of duration and volume show that the fluctuations of the stochastic components decrease by the adjustment. Hereafter, we adopt the adjusted series obtained from the preparing procedures for empirical analysis.

4. ESTIMATION RESULTS

The preliminary linear AR($q$) model for the specification of the mean equation (3) is determined by reference to the Schwarz criterion. According to the statistics in table 2, this identification procedure indicates that an AR(6) process is appropriate.

Table 3 reports the parameter estimates and related statistics for the GARCH(1,1) models suggested in the previous section. The estimates for the MS-GARCH models are compared with those for the single-regime model. The unconditional variance in single regime is estimated at 1.489, which seems to be similar to square of the sample standard deviation of 1.254 in table 1. For the MS-GARCH-CP model, the unconditional variance is 3.886 in regime 1, and 0.788 in regime 2. Thus, these estimates identify that regime 1 is a high-variance regime and regime 2 is a low-variance regime. The unconditional variances for the MS-GARCH-TVP model are similar to those for the MS-GARCH-CP model.

<table>
<thead>
<tr>
<th>Table 2 Preliminary Selection of the AR Order</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q = 1$</td>
</tr>
<tr>
<td>Schwarz</td>
</tr>
</tbody>
</table>

Notes: In this table, $q$ denotes the order of an autoregressive model, $r = \phi_0 + \phi_1 r_{-1} + \cdots + \phi_q r_{-q} + u$. At $q = 6$, the Schwarz criterion indicates minimum value, so that the AR(6) model is the best selection for the mean equation.
**Table 3**  Parameter Estimates for GARCH(1,1) Models

<table>
<thead>
<tr>
<th></th>
<th>Single Regime</th>
<th>Markov Switching</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MS-GARCH-CP</td>
</tr>
<tr>
<td></td>
<td>Coeff.</td>
<td>z-Stat.</td>
</tr>
<tr>
<td>$\phi_0$</td>
<td>0.000</td>
<td>0.088</td>
</tr>
<tr>
<td>$\phi_1$</td>
<td>-0.285</td>
<td>-147.4</td>
</tr>
<tr>
<td>$\phi_2$</td>
<td>-0.044</td>
<td>-21.45</td>
</tr>
<tr>
<td>$\phi_3$</td>
<td>0.004</td>
<td>1.967</td>
</tr>
<tr>
<td>$\phi_4$</td>
<td>0.011</td>
<td>5.024</td>
</tr>
<tr>
<td>$\phi_5$</td>
<td>0.006</td>
<td>2.641</td>
</tr>
<tr>
<td>$\phi_6$</td>
<td>0.001</td>
<td>0.595</td>
</tr>
<tr>
<td>$\alpha^{(1)}$</td>
<td>1.489</td>
<td>63.02</td>
</tr>
<tr>
<td>$\alpha^{(2)}$</td>
<td>0.065</td>
<td>96.13</td>
</tr>
<tr>
<td>$\beta^{(1)}$</td>
<td>0.923</td>
<td>1409</td>
</tr>
<tr>
<td>$\beta^{(2)}$</td>
<td>0.788</td>
<td>51.22</td>
</tr>
<tr>
<td>$\gamma_{0}^{(1)}$</td>
<td>7.130</td>
<td>56.33</td>
</tr>
<tr>
<td>$\gamma_{d}^{(1)}$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\gamma_{s}^{(1)}$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\gamma_{0}^{(2)}$</td>
<td>7.298</td>
<td>72.32</td>
</tr>
<tr>
<td>$\gamma_{d}^{(2)}$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\gamma_{s}^{(2)}$</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Log-lik. $-384,609.4$ $-383,719.2$ $-383,692.9$

Notes: The abbreviations CP and TVP represent constant and time-varying transition probabilities, respectively. The $p$-values are calculated from a standard normal distribution; thus significance statistics of parameters are denoted as the $z$-Stat. The mean and variance equations are as follows:

$$r_t = \phi_0 + \phi_1 r_{t-1} + \phi_2 r_{t-2} + \phi_3 r_{t-3} + \phi_4 r_{t-4} + \phi_5 r_{t-5} + \phi_6 r_{t-6} + u_t,$$

$$h^{(1)}(t) = \alpha^{(1)} \tau_{t-1} + \beta^{(1)} (h^{(1)}(t-1) - h^{(1)}),$$

where the superscript $j$ indicates a specific regime, with $j = 1$ for a high-variance regime or $j = 2$ for a low-variance regime in Markov-switching models. The transition probabilities could be estimated as follows:

$$p^{(1)} = \frac{1}{1 + \exp(-\gamma_{c}^{(1)})}$$

$$p^{(2)} = \frac{1}{1 + \exp(-\gamma_{c}^{(2)} d_{t-1} + \gamma_{v}^{(2)} v_{t-1})},$$

where the former is in the CP and the latter is in the TVP model, respectively.
The persistence parameter estimate for the variance equation in the single-regime model is $0.988 (= 0.065 + 0.923)$, which indicates that the volatility is very persistent. The persistence parameter estimates for the MS-GARCH-CP model are $0.949 (= 0.117 + 0.832)$ in a high-variance regime and $0.813 (= 0.106 + 0.707)$ in a low-variance regime. These facts show that a volatility shock in a high-variance regime is more persistent than one in a low-variance regime. For the MS-GARCH-TVP model, the parameter estimates are $0.943 (= 0.123 + 0.820)$ in a high-variance regime and $0.812 (=0.106 + 0.706)$ in a low-variance regime, so that the persistence of volatility in the two MS-GARCH models resemble each other.

For the MS-GARCH-CP model, the parameter estimates of transition probabilities specified as logistic functions in equation (11) are $\gamma^{(11)}_0 = 7.130$ and $\gamma^{(22)}_0 = 7.298$, and thus the estimates of transition probabilities are $p^{(11)} = 0.999$ and $p^{(22)} = 0.999$. Therefore, the unconditional probability of regime 1 is approximately 0.5 according to Hamilton’s (1994) formula, that is, equation (2), so that the probabilities of regime 1 and 2 seem to be the same.

For the MS-GARCH-TVP model, the parameter estimates of the transition probability functions are different between the two regimes. The coefficient of the duration that varies the transition probability $p^{(11)}_d$ is $\gamma^{(11)}_d = -0.457$. The coefficient of the duration in the logistic function of $p^{(22)}_d$ shows a positive value, $\gamma^{(22)}_d = 2.784$; however, it is not significant at the 5% level. In addition, the trading volume decreases both $p^{(11)}_v$ and $p^{(22)}_v$ according to the estimates of $\gamma^{(11)}_v = -0.145$ and $\gamma^{(22)}_v = -0.068$. These findings imply that the duration decreases volatility, mainly reducing $p^{(11)}$, while the transactions lead to a shift of one regime to another.

Table 4 reports the results of Wald parameter tests — whether the parameters of the variance equations of the MS-GARCH models reported in table 3 are the same in both regimes. The statistics reject the null hypotheses for most of the parameters, whereas the test statistic of the parameters in the transition probability functions for the MS-GARCH-CP model does not reject the null hypothesis, and supports the fact that the transition probabilities
are the same in both regimes. In addition, the hypothesis that the coefficients of duration in the transition probability functions of the MS-GARCH-TVP model are the same in both regimes is not rejected at the 5% level. Thus, it can be confirmed again that the coefficient of duration is not significant in the low-variance regime.

Table 4  Parameter Equality Tests for Markov Switching Models

| Null Hypothesis | MS-GARCH-CP | | | MS-GARCH-TVP | | |
|-----------------|-------------|-------------|-------------|-------------|-------------|
|                 | Wald        | p-value     | Wald        | p-value     |
| \( H_0 \): \( h^{(1)} = h^{(2)} \) | 1.559.690   | 0.0000      | 1.693.715   | 0.0000      |
| \( H_0 \): \( \alpha^{(1)} = \alpha^{(2)} \) | 21.53250    | 0.0000      | 37.85316    | 0.0000      |
| \( H_0 \): \( \beta^{(1)} = \beta^{(2)} \) | 210.6775    | 0.0000      | 159.5718    | 0.0000      |
| \( H_0 \): \( \gamma_o^{(11)} = \gamma_o^{(22)} \) | 1.282155    | 0.2575      | 10.59543    | 0.0011      |
| \( H_0 \): \( \gamma_d^{(11)} = \gamma_d^{(22)} \) | -           | -           | 3.531118    | 0.0602      |
| \( H_0 \): \( \gamma_e^{(11)} = \gamma_e^{(22)} \) | -           | -           | 22.08158    | 0.0000      |

Notes: The Wald statistic on constant transition probabilities shows that the probabilities for the MS-GARCH-CP model are not different from each other. In addition, the test for duration effects in the MS-GARCH-TVP could not reject the null hypothesis at the 5% level.

Table 5  Ljung-Box Statistics for Squared Residuals

|                  | Single Regime | Markov Switching | | | | |
|------------------|---------------|------------------|-------------|-------------|-------------|
|                  |               |                  | MS-GARCH-CP | MS-GARCH-TVP | |
|                  | \( u_i^2 \)   | \( \varepsilon_i^2 \) | \( u_i^2 \) | \( \varepsilon_i^2 \) | |
| LB(10)           | 48,257        | 836.84           | 48,371      | 231.80      | 48,373      | 223.05      |
| LB(20)           | 75,159        | 1,079.7          | 75,334      | 246.17      | 75,338      | 236.81      |
| LB(50)           | 111,615       | 1,171.8          | 111,913     | 340.39      | 111,918     | 328.78      |
| LB(100)          | 144,328       | 1,316.2          | 144,712     | 455.06      | 144,717     | 453.05      |
| LB(200)          | 190,733       | 1,506.4          | 191,238     | 536.25      | 191,243     | 544.88      |

Notes: This table reports Ljung-Box statistics for the squared residuals drawn from the empirical models. LB(\( p \)) denotes the Ljung-Box statistic for serial correlation of the squared residuals out to \( p \) lags. In the table, \( u_i \) indicates a ordinary residual series, and \( \varepsilon_i \), a standardized residual series, i.e., \( \varepsilon_i = u_i / \sqrt{h_i} \).
Figure 2  Sample Autocorrelation Functions of Squared Residuals

Notes: This figure shows sample autocorrelation functions of the squared residuals drawn from the empirical models. The top panel displays sample ACFs of squared AR(6) residuals (\( u_t \)), and the bottom, squared standardized residuals (\( \epsilon_t \)). The bottom axis scale denotes the number of lags, and the left axis scale, sample ACF value on the both panels.

Table 5 reports Ljung-Box (LB) tests for the squared residuals and squared standardized residuals of the three models. All the LB statistics of the three models indicate strong and persistent autocorrelation for the volatilities. However, the statistics for squared standardized residuals of the MS-GARCH models are dramatically reduced, and represent significantly smaller values in contrast to the single-regime GARCH model. These findings suggest that the Markov-switching volatility models are very effective to reduce the autocorrelation of the volatility and to characterize the nonlinearity of the return. The autocorrelation functions (ACFs) of squared residuals in figure 2 also support these findings. To say in more detail, it could be identified that
the ACFs of squared standardized residuals are very reduced on the bottom panel in contrast to the ACFs of squared residuals drawn from an ordinary AR(6) process on the top panel. In addition, it may be visually found that autocorrelations from MS-GARCH models decrease on the bottom panel in contrast to single regime GARCH.

Figure 3 displays the smoothed probabilities of the regime 1 and the 2-SD confidence bands drawn from the MS-GARCH-CP and MS-GARCH-TVP models. Smoothed inferences for regime 1 (i.e., the high-variance regime) are calculated using an algorithm developed by Kim (1993). With our notation, this algorithm can be written as

$$P(S_i = 1 | \Omega_N) = P(S_i = 1 | \Omega_i) \times \left[ p^{(1)} \cdot \frac{P(S_{i+1} = 1 | \Omega_N)}{p_{i+1}} + (1 - p^{(1)}) \cdot \frac{1 - P(S_{i+1} = 1 | \Omega_N)}{1 - p_{i+1}} \right]. \quad (16)$$

The smoothed probability of the regime 1 $P(S_i = 1 | \Omega_N)$ is found by iterating on (16) backward for $i = N - 1, N - 2, \ldots, 1$. This iterating started with $P(S_N = 1 | \Omega_N)$, which is obtained from optimal inference at each point of transaction $i$ in the sample

$$P(S_i = 1 | \Omega_i) = \frac{p_i f_i^{(1)}}{p_i f_i^{(1)} + (1 - p_i) f_i^{(2)}}, \quad (17)$$

for $i = N$ (Hamilton, 1994). In equations (16) and (17), $p_i, p_{i+1}, p^{(1)}$, and $f_i^{(j)}$ are known from the estimation process. In order to calculate the smoothed probability for the MS-GARCH-TVP model, the time-varying transition probability $p_{i+1}^{(1)}$ should substitute for the constant probability $p^{(1)}$ in equation (16).

The top panel of Figure 3 contains a plot of the smoothed probability series of regime 1 for the MS-GARCH-CP model, and the second panel, the 2-SD confidence band. On a comparison of the two plots, the smoothed probability that the world is in a high-variance regime seems to be properly
Figure 3  Smoothed Regime 1 Probability and 2-S.D. Band

Notes: The top panel displays a smoothed probability series — that the return process is in regime 1 (the high-variance regime) for transaction \( i \) according to the MS-GARCH-CP model. The smoothed probability is inferred from the entire sample \( (P(S_i = 1|\Omega_i)) \). The second panel contains the plot of a 2-S.D. \( (\pm 2\sqrt{h}) \) band from the estimation for the MS-GARCH-CP model. The third and the fourth panels show the plots of smoothed regime 1 probability and 2-S.D. band, respectively, for the MS-GARCH-TVP model.

Inferred. For comparison of the two volatility models, the plots of the smoothed probability and 2-SD band drawn from the MS-GARCH-TVP model are displayed in the third and fourth panels, respectively. Although the broad patterns are similar, the smoothed probabilities for the two models
Figure 4  Zoomed Plots in a 2-S.D. Band, Residuals, and Smoothed Probability

Notes: This figure enlarges a part of the entire sample, to compare the estimated volatilities for the Markov-switching models with those for the single regime GARCH(1,1) model. The top panel displays a 2-S.D. band and the residual series drawn from a single regime GARCH(1,1) model. The middle and the bottom panels show the plots of those and smoothed regime 1 probabilities drawn from the MS-GARCH-CP and MS-GARCH-TVP model, respectively. On all the plots, the 2-S.D. bands appear as solid lines (left axis), the residual series, as a dotted line (left axis), and the smoothed probabilities, as a dashed line (right axis).

significantly differ from each other because of the reaction to duration and volume in the MS-GARCH-TVP model. However, the volatilities for the two MS-GARCH models are moving identically. According to these findings, we suggest that they are useful for duration and volume to explain the movement of volatility, rather than to forecast.

Figure 4 enlarges a particular part of the entire sample and shows the 2-SD
bands and the residual series for a close inspection. The top panel contains the plot of these series for the single-regime GARCH(1,1) model to be compared with those for the MS-GARCH models. The middle and bottom panels contain the plots of those with the smoothed probabilities for the MS-GARCH-CP and MS-GARCH-TVP models, respectively. The single-regime GARCH confidence intervals shrink very gradually in response to error shock, whereas the confidence intervals for the MS-GARCH models quickly return to the trend level with a fall in regime 1 probability. This finding implies that the volatilities of Markov-switching models respond more sensitively to new information than those of the single-regime model.

5. CONCLUSIONS

This paper specifies two-state Markov-switching volatility models and investigates the volatility behavior of ultra-high-frequently observed returns on JGB futures transactions. In addition, we test the effects of duration and volume on transition probabilities with a time-varying probability model. The data go through a two-stage preparatory procedure before we use the diurnally adjusted UHF data to estimate the empirical models.

Our main findings and implications are as follows: First, the MS-GARCH models are very effective to reduce the autocorrelation of volatility, since the LB statistics of the squared standardized residuals of the models are dramatically reduced and present significantly smaller values in contrast to the single-regime GARCH model. This finding suggests that Markov-switching volatility models properly characterize the nonlinearity of returns, so that the models improve the statistical properties of ultra-high-frequent volatility in contrast to single-regime models. Second, the volatilities of the MS-GARCH models respond more sensitively to new information than those of the single-regime model. This finding shows that Markov-switching volatility models capture abrupt changes in volatility more effectively. Third, the duration decreases volatility, mainly reducing $p_i^{(1)}$ in the time-varying
transition probability model, while the trading volume decreases both $p_{11}^{(11)}$ and $p_{22}^{(22)}$ so that the transactions lead to a shift from one regime to another. According to Engle (2000), the former result indicates that no trade is interpreted as no news so that volatility is reduced. The latter supports Admati and Pfleiderer (1988) in which discretionary liquidity traders can allocate their trades across different periods. Thus, their trading is relatively more concentrated in periods closer to the realization of their demands. Synthesizing these findings, the Markov-switching models properly exhibit the behavior of heterogeneous traders faced with heterogeneous risks — either high variance or low variance.

REFERENCES


Two-State Markov Switching Volatility Model for Ultra-High-Frequency Data of JGB Futures


