GARCH-Jump Models with Regime-Switching
Conditional Volatility and Jump Intensity

Pujun Liu*

The University of Western Ontario

Preliminary May 2011

Abstract

The paper presents two types of regime-switching GARCH-jump models with autoregressive jump intensity to model the non-linearity in return series. The first model is a Markov regime-switching model which generalizes the GARCH model by distinguishing two regimes with different GARCH volatility and jump intensity levels. As the regimes are unknown to the econometricians in Markov regime-switching models which lead to difficulty in forecasting, we also build a threshold GARCH-jump model with an exogenous threshold variable. The stationary conditions and moments of returns are derived for the threshold GARCH-jump model. Using Japanese YEN-US Dollar exchange rate, we show that both types of regime-switching models work much better than the traditional GARCH model for in-sample period. Constructing realized volatility from 5-minute intraday data for evaluation, threshold GARCH-jump model makes better forecasts than the single regime autoregressive jump intensity model and the GARCH (1,1) model.

*I am deeply indebted to my supervisor Professor John Knight for his excellent guidance and support. I would like to thank Professor Dinghai Xu for providing high frequency data of Japanese Yen-US Dollar spot exchange rate. I am responsible for all errors.
1 Introduction

Jump diffusion models have received wide-spread acceptance for their ability to model both continuous small changes and infrequent large movements in financial return series since the seminal work by Press (1967) and Merton (1976). In the discrete version of the jump diffusion models, GARCH models and stochastic volatility models are used to account for volatility clustering as the diffusion part. For the jump part, jump intensity, which refers to the arrival rate of jumps, is usually assumed as a constant, partly because of the difficulty of estimation for stochastic jump intensity models. Recently, Chan and Maheu (2002) and Maheu and McCurdy (2004) model the return series as a combination of jumps and smoothly changing components, in which the conditional jump intensity is autoregressive. They find that the autoregressive parameter in the jump intensity is positively significant and quite high for individual stock returns, which supports the phenomenon of jump clustering.

While jump diffusion models present a parametric way to model abnormal or jump innovations as well as normal innovations, the harnessing of high frequency data in the last decade has also led to separate analysis of diffusive and jump components using a non-parametric approach. Daily realized volatility, constructed by Andersen and Bollerslev (1998) and Bandorff-Nielsen and Shephard (2001) using the summation of squared intraday returns, is a consistent estimator of the quadratic variation in a continuous jump-diffusion setting with a bounded jump intensity. This provides a good proxy for daily volatility after dealing with intraday pattern and microstructure noise. Moreover, Bandorff-Nielsen and Shephard (2004) present realized bipower variation constructed from high frequency data, which is consistent for the integrated variance in the same jump-diffusion setting. As the difference of quadratic variation and the integrated variance is the cumulative
squared jumps, the result renders feasible statistical tests for the presence and impact of jumps. Huang and Tauchen (2005) show that a test for jumps has good power and detection capacities using Monte Carlo analysis, and indicate strong empirical evidence for jumps that account for seven percent of stock market price variance.

Another line of literature deals with non-linearity using regime switching models. Lamoureux and Lastrapes (1990) show that the high persistence of the conditional variance using GARCH model may be overstated due to the failure of recognizing structural changes in the model. Gray (1996) develop a generalized regime switching GARCH model and find it outperforms single-regime models out-of-sample using short-term interest rate. More recently, Hillebrand (2005) show that the convergence of the sum of estimated autoregressive parameters holds for all common estimators of GARCH. Thus, in the presence of neglected parameter changes, GARCH is no longer a suitable model to measure persistence.

In this paper we model structural breaks and jumps together by building regime switching conditional jump intensity models which is based on Chan and Maheu (2002)’s autoregressive jump intensity (ARJI) model. The motivation is that although ARJI model provides good in-sample estimation, the out-of-sample forecasting ability is not good especially when the jump frequency in the out-of-sample period differs from the in-sample period. For example, when the out-of-sample period is a relatively tranquil period, using parameters estimated from the relatively volatile period will overestimate the jump frequency and lead to inaccurate forecasting. In this paper we will show that the out-of-sample forecasting performance is not as good as GARCH model for Japanese Yen-US Dollar exchange rate. Thus, it is necessary to distinguish between volatile period and tranquil period for jumps. In addition, Maheu and McCurdy (2004) plot the time-series of conditional variance components of IBM estimated by their GARJI model, in which both
GARCH component and jump component of the conditional variance are higher than normal in some periods, while in other periods both of them are less volatile. The phenomenon suggests that both smooth changes and jumps may be governed by regime changes. Furthermore, the high persistence in conditional variance may be spurious due to latent structural changes in the data generating process. Thus, we model the GARCH volatility and jump intensity process in different regimes in order to improve the out-of-sample forecasting performance.

We build two types of regime-switching GARCH-jump model. The first one follows the traditional Markov regime-switching model proposed in Hamilton (1989), which has good stationary conditions but latent regimes. The difficulty to introduce regimes into conditional jump intensity in this type of model is that the jump intensity will depend on the entire regime path from the beginning of the period to the current period as it is autoregressive, which leads to computational complexity. To circumvent this problem, the jump intensity is assumed to depend only on its current regime state. However, as unknown regimes render difficulty for forecasting in Markov regime-switching models, we also build a threshold GARCH-jump model, in which regimes are known after we observe the threshold variable at the previous period. Recently, Knight and Satchell (2010) derive sufficient and necessary conditions for the existence of a stationary distribution for a threshold AR(1) model with exogenous threshold variable. We extend their research and find stationary conditions for the threshold GARCH-jump model with regimes in both GARCH type conditional variance and jump intensity.

The paper is organized as follows. Section 2 presents Markov regime-switching GARCH-jump model and propose the estimation mechanism after a brief review of autoregressive jump intensity (ARJI) model by Chan and Maheu (2002). In section 3, we deal with threshold GARCH-jump model with an exogenous trigger and present stationary conditions. Empirical analysis is
conducted in section 4 and 5. We use Japanese Yen-US Dollar spot exchange rate for the estimation, with realized volatility constructed from 5-minute intraday data used for evaluation of forecasts of different models. Section 6 contains a brief conclusion.

2 Markov regime-switching generalized autoregressive jump intensity (RSGARJI) model

We incorporate regimes in both GARCH variance and jump intensity. For a Markov regime-switching GARCH (1,1)-jump intensity AR (1) process, which we denote by RSGARJI model, the model setting is

\[ R_t = \mu + \epsilon_{1,t} + \sum_{k=1}^{N(t)} Y_{t,k} \]

\[ \epsilon_{1,t} = z_t \sigma_t \]

\[ Y_{t,k} \sim i.i.d N(\theta, \delta^2) \]

\[ P(N(t) = j|\Phi_{t-1}) = \exp(-\lambda_t)\lambda_t^j/j! \text{for} j = 0, 1, 2... \]

In regime \( s_t \), for \( s_t = 1, 2 \),

\[ \sigma_t^2 = \omega_{s_t} + a_{s_t} \epsilon_{t-1}^2 + b_{s_t} E[\sigma_{t-1}^2|\Phi_{t-1}, s_t] \quad (1) \]

\[ \lambda_t = \alpha_{s_t} + \rho_{s_t} E[\lambda_{t-1}|\Phi_{t-1}, s_t] + \gamma_{s_t} E[\xi_{t-1}|\Phi_{t-1}, s_t] \quad (2) \]

\[ \xi_{t-1} = E[N(t - 1)|\Phi_{t-1}, S_{t-1}] - \lambda_{t-1} \]
$R_t$ denotes the time $t$ return, which is the first difference of logarithmic price. $\epsilon_{1,t}$ is a mean-zero normal innovation with an autoregressive conditional variance, and $\sum_{k=1}^{N(t)} Y_{t,k}$ is the jump innovation which is a compound Poisson process. $N(t)$ is the number of jumps and follows a poisson process with autoregressive jump intensity $\lambda_t$. $s_t$ denotes the regime at time $t$, which can take value of 1 or 2 referring to two different regimes. Smoothly changing components are represented by $\epsilon_{1,t}$, which follows a GARCH process with different parameters in different regime, corresponding to $(\omega_{s_t}, a_{s_t}, b_{s_t})$ in regime $s_t$. The jump intensity $\lambda_t$ follows an approximate AR(1) process in each regime, with parameters $(\alpha_{s_t}, \rho_{s_t}, \gamma_{s_t})$ in regime $s_t$. We use the approximate AR(1) process introduced by Chan and Maheu (2002) to model the jump intensity, as it can circumvent the problem that the likelihood function has no closed form when jump intensity follows an ARMA process. $\xi_{t-1}$ can be viewed as an approximate error term. It will be mentioned later.

When regimes are assumed to be exogenous, i.e, explanatory variables in the conditional intensity process $\sigma_t$ contain no information about $s_t$ beyond that contained in $\Phi_{t-1}$, $s_t$ is assumed to follow a first-order Markov process as in Hamilton (1989).

$$P(s_t = j | S_{t-1}) = P(s_t = j | s_{t-1}) = p_{ij}$$

(3)

The transition matrix is

$$P = \begin{bmatrix} p_{11} & 1 - p_{11} \\ 1 - p_{22} & p_{22} \end{bmatrix}$$

The value of $p_{11}$ and $p_{22}$ implies the persistence of regimes.

This is the model setting when the regimes follows a Markov process. In the following subsections, we will show why we build the model in this way. First we present the autoregressive jump intensity (ARJI) model of Chan and Maheu (2002) and discuss its properties. Then we will elaborate on the
way of constructing the regime-switching model in both conditional variance and jump intensity.

2.1 ARJI model by Chan and Maheu (2002)

Previous literature suggests that jump intensity is time-varying and may depend on its lagged values. For example, Knight and Satchell (1998) model a self-exciting jump intensity which depends on past volatility and a stochastic deviation from fundamentals. By substitution the jump intensity can be expressed as an autoregressive form together with non-negative error term. Chan and Maheu (2002) and Maheu and McCurdy (2004) generate an autoregressive conditional jump intensity (ARJI) model and derive conditional moments of the returns. Applying to several individual firms, the persistence parameter for the arrival of jump events is quite high, up to 0.924 for Texaco. Their model is a single-regime discrete-time GARCH-jump model with time dependent jump intensity with the following specification.

\[ \lambda_t = \alpha + \rho \lambda_{t-1} + \gamma \xi_{t-1} \]

\[ \xi_{t-1} = E[N(t-1)|\Phi_{t-1}] - \lambda_{t-1} \]

\( E[N(t-1)|\Phi_{t-1}] \) is the ex post assessment of the expected number of jumps given information set \( \Phi_{t-1} \), while \( \lambda_{t-1} \) is the ex ante assessment. So \( \xi_t \) can be viewed as the change in the conditional forecast of the jumps after the information set it updated. It is a martingale difference sequence with respect to \( \{\Phi_{t-1}\} \). Thus it can be viewed as an error term in the jump intensity process. If \( |\rho| < 1 \), then the jump intensity is stationary. The conditional mean of return is \( E(R_t|\Phi_{t-1}) = \theta \lambda_t \), and the conditional variance is \( Var(R_t|\Phi_{t-1}) = \sigma_t^2 + (\theta^2 + \delta^2)\lambda_t \), which is a combination of the GARCH conditional variance component and the jump component.
Maheu and McCurdy (2004) describe some empirical results found by the model. They find strong evidence of time dependence in jump intensities for both stock indices and several individual stocks. The average proportion of conditional variance explained by jumps varies from 20% to 40%, at times as much as 90%. It is much higher than Huang and Tauchen (2005)’s finding that jumps account for 7 percent of stock market price variance using S&P index.

2.2 Construction of Markov regime-Switching GARCH-jump model (RSGARJI model)

The reported high persistence of conditional variance by GARCH model, together with high level of jump clustering revealed in Maheu and McCurdy (2004), may be spurious and due to structural shifts in the data generating process, such as deterministic changes in the intercept parameter of autoregressive jump intensity process. What’s more, there are periods during which few jumps happen and other periods when jumps cluster. To solve the problem we introduce regime shifts into the conditional jump intensity. To incorporate structural changes in the data generating process, a popular approach is the Markov regime-switching model applied to dependent processes by Hamilton (1989). State 1 and state 2 refer to low jump intensity regime and high jump intensity regime respectively, and a Markov process is used to govern the switches between regimes. Within each regime the jump intensity depends on its own lagged value and some explanatory variables, such as conditional variance of the GARCH term.

We build the regime-switching model based on ARJI model for two reasons. Firstly, the autoregressive jump intensity setting can account for clustering of jumps and also incorporate shocks. By introducing regimes into GARCH-type conditional variance and jump intensity, we can explore if per-
sistence of conditional variance and jumps vary for different regimes and whether the high persistence is spurious due to structural changes. Secondly, jump diffusion models with stochastic jump intensity are hard to estimate as the likelihood function has no closed form. The ARJI model avoids this problem by assuming approximate autoregressive jump intensity structure with a filter to infer the ex post distribution of jumps.

Let \( \{s_t\} \) be a Markov Chain with 2-dimensional state space. In state 1, jumps are not so frequent, while jumps are more possible to happen in state 2. \( S_t \) denotes the regime path at time \( t \), \( \{s_t, s_{t-1}, s_{t-2}, \ldots\} \). \( \Phi_{t-1} \) refers to the information set at time \( t - 1 \). To clarify, \( \epsilon_t \) refers to \( R_t - \mu \), which is the summation of \( \epsilon_{t,t} \) and the jump part.

The first specification for the conditional jump intensity \( \lambda_t \) is

\[
P(N(t) = j|\Phi_{t-1}, S_t, x_t) = \exp(-\lambda_t)\lambda_t^j/j! \text{ for } j = 0, 1, 2\ldots
\]

(4)

\[
\lambda_t = \alpha_1 + \rho_1 \lambda_{t-1} + \gamma_1 \xi_{t-1} \text{ if } s_t = 1
\]

(5)

\[
\lambda_t = \alpha_2 + \rho_2 \lambda_{t-1} + \gamma_2 \xi_{t-1} \text{ if } s_t = 2
\]

(6)

The conditional jump intensity, \( \lambda_t = E(N(t)|\Phi_{t-1}, S_t) \), has an autoregressive form and depends on contemporaneous conditional variance and trading volume. Hereby \( \xi_{t-1} = E[N(t-1)|\Phi_{t-1}, S_{t-1}] - \lambda_{t-1} \), which also depends on the state space of regimes. It is easy to show that \( \xi_t \) is still a martingale difference sequence with respect to \( \{\Phi_{t-1}, S_{t-1}\} \), where \( \Phi_{t-1} \) is the information set up to \( t - 1 \), which is composed of past values of \( R_t \). Therefore it is a well defined residual.

However, this specification for jump intensity has its computational difficulty. Although the current regime only determines the parameters \( \alpha, \rho, \beta \) and \( \gamma \), the dependence of \( \lambda_t \) on both the current regime and its own lagged value
\( \lambda_{t-1} \) makes \( \lambda_t \) depend on the entire regime path \( \{s_t, s_{t-1}, s_{t-2}, \ldots\} \) by iterative substitution. As the regimes are latent, the inability to observe them leads to the need of integrating out all possible paths when calculating the sample likelihood. It makes the estimation practically intractable. This problem is similar to that arises in literature of regime-switching GARCH models, as noted by Klaaseen (2002), which models the GARCH type volatility as a regime-switching process.

To circumvent the problem of path dependence, the conditional jump intensity in this paper is modeled as

\[
\lambda_t = \alpha_{s_t} + \rho_{s_t} E[\lambda_{t-1}|\Phi_{t-1}, s_t] + \gamma_{s_t} E[\xi_{t-1}|\Phi_{t-1}, s_t] \tag{7}
\]

Let \( \lambda_t = E[N(t)|\Phi_{t-1}, s_t] \). Note that \( \lambda_t \) depends only on \( s_t \) instead of \( S_t \) now. The idea is inspired by Klaaseen (2002) to integrate out the regime path \( S_{t-1} \) out of the right hand side of the equation. After \( S_{t-1} \) is integrated out for conditional intensity of the last period, the right hand side only depends on the current regime \( s_t \). In addition, this is equivalent to integrate out \( s_{t-1} \), the regime at time \( t-1 \), as the lag of equation (8) implies that \( \lambda_{t-1} \) only depends on \( s_{t-1} \) and is independent of \( S_{t-2} \). \( s_t \) is included in the conditioning variables because it may contain some information about the last period regime \( s_{t-1} \).

In order for \( \lambda_t \) to be positive for all \( t \), a sufficient condition is that \( \alpha_{s_t} > 0 \), \( \rho_{s_t} - \gamma_{s_t} \geq 0 \), \( \beta_{s_t} \geq 0 \) and \( \gamma_{s_t} \geq 0 \). This specification of the conditional jump intensity removes the problem of regime path dependence, and allows the jump intensity to be autocorrelated, which can explain for the phenomenon of jump clustering around significant news events.
2.3 Properties of the model when regime is exogenous and estimation mechanism

The steady state probabilities of the regimes 1 and 2 at time $t-1$, $P(s_{t-1} = 1)$ and $P(s_{t-1} = 2)$, are derived in Hamilton (1989),

$$P(s_{t-1} = 1) = \frac{1 - p_{22}}{2 - p_{11} - p_{22}} \quad (8)$$

$$P(s_{t-1} = 2) = \frac{1 - p_{11}}{2 - p_{11} - p_{22}} \quad (9)$$

**Proposition 2.1** If the unconditional jump intensity exists for both regime 1 and 2, denoted by $\lambda_1$ and $\lambda_2$ respectively, then

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = B^{-1} \begin{bmatrix} \alpha_1 + \tau_1 \\ \alpha_2 + \tau_2 \end{bmatrix} \quad (10)$$

where $\tau_{s_t} = \beta_{s_t} E[\sigma_t^2]$ for $s_t = 1$ or 2, and

$$B = \begin{bmatrix} 1 - \rho_1 p_{11} & -\rho_1 (1 - p_{11}) \\ -\rho_2 (1 - p_{22}) & 1 - \rho_2 p_{22} \end{bmatrix} \quad (11)$$

From Proposition 2.1, for the existence of the unconditional jump intensity, the inverse of $B$, the unconditional mean of GARCH-type conditional variance needs to exist. In order for the unconditional jump intensity to be strictly positive for all $t$, the four elements of the inverse of $B$ should be positive.

$$B^{-1} = \frac{1}{1 - \rho_1 p_{11} - \rho_2 p_{22} - \rho_1 \rho_2 (1 - p_{11} - p_{22})} \begin{bmatrix} 1 - \rho_2 p_{22} & \rho_1 (1 - p_{11}) \\ \rho_2 (1 - p_{22}) & 1 - \rho_1 p_{11} \end{bmatrix} \quad (12)$$
For the unconditional jump intensity to be strictly positive, \(1 - \rho_1 p_{11} - \rho_2 p_{22} - \rho_1 \rho_2 (1 - p_{11} - p_{22}) > 0\), \(1 - \rho_1 p_{11} > 0\), and \(1 - \rho_2 p_{22} > 0\).

Then the conditional variance of \(R_t\) is

\[
\text{Var}(R_t|\Phi_{t-1}) = \sum_{s_t=1,2} P(s_t|\Phi_{t-1}) \text{Var}(R_t|s_t, \Phi_{t-1})
= \sum_{s_t=1,2} P(s_t|\Phi_{t-1}) (\sigma_t^2 + (\theta^2 + \delta^2)\lambda_t)
\]  

(13)

The estimation of the model can be conducted by maximum likelihood method based on estimation mechanism of ARJI model using an iterative algorithm. Regimes are unknown for the econometrician. The conditional density function of the return can be found by summing over the conditional density function in different regimes times their conditional probability.

\[
f(R_t|\Phi_{t-1}) = \sum_{j=0}^{\infty} f(R_t|N(t) = j, \Phi_{t-1}) P(N(t) = j|\Phi_{t-1})
\]

(14)

When there is an infinite summation in the likelihood function, we truncate it at 20. Experiments in empirical analysis shows that \(N(t)\) does not exceed 15. The first part of the right hand side of the conditional sample likelihood function, \(f(R_t|N(t) = j, \Phi_{t-1})\), can be derived as follows.

\[
f(R_t|N(t) = j, \Phi_{t-1}) = \frac{1}{\sqrt{2\pi(j\delta^2 + \sigma_t^2)}} \exp\left(-\frac{(R_t - \mu - j\theta)^2}{2(j\delta^2 + \sigma_t^2)}\right)
\]

(15)

And the second part is

\[
P(N(t) = j|\Phi_{t-1}) = P(N(t) = j|\Phi_{t-1}, s_t = 1) P(s_t = 1|\Phi_{t-1})
+ P(N(t) = j|\Phi_{t-1}, s_t = 2) P(s_t = 2|\Phi_{t-1})
\]

(16)

Then we need to get the expression of \(P(N(t) = j|\Phi_{t-1}, s_t)\) and \(P(s_t|\Phi_{t-1})\). We know from the model specification that

\[
P(N(t) = j|\Phi_{t-1}, s_t) = \exp(-\lambda_t) \lambda_t^j / j! \text{for } j = 1, 2,
\]

(17)
where $\lambda_t$ is a function of $s_t$, which is not straightforward to compute because of integrating out of regime path $S_{t-1}$ in $E[\lambda_{t-1}|\Phi_{t-1}, s_t]$.

$$\lambda_t = \alpha_{s_t} + \rho_{s_t} E[\lambda_{t-1}|\Phi_{t-1}, s_t] + \gamma_{s_t} E[\xi_{t-1}|\Phi_{t-1}, s_t]$$

(18)

As $\lambda_{t-1}$ is a function of $s_{t-1}$,

$$E[\lambda_{t-1}|\Phi_{t-1}, s_t] = \sum_{s_{t-1}=1,2} \lambda_{t-1}(s_{t-1}) P(s_{t-1}|\Phi_{t-1}, s_t)$$

(19)

$\xi_t = E[N(t)|\Phi_t, S_t] - \lambda_t$ is also a function of $s_t$. The density of the expectation part of $\xi_t$ is

$$P(N(t) = j|\Phi_t, s_t) = f(R_t|N(t) = j, \Phi_{t-1}, s_t)P(N(t) = j|\Phi_{t-1}, s_t)/f(R_t|\Phi_{t-1}, s_t)$$

(20)

Then

$$\xi_{t-1} = \sum_{j=0}^{\infty} j P(N(t-1) = j|\Phi_{t-1}, s_{t-1}) - \lambda_{t-1}$$

(21)

$$E[\xi_{t-1}|\Phi_{t-1}, s_t] = \sum_{s_{t-1}=1,2} \xi_{t-1}(s_{t-1}) P(s_{t-1}|\Phi_{t-1}, s_t)$$

(22)

After getting $P(s_{t-1}|\Phi_{t-1})$ and $P(s_{t-1}|\Phi_{t-1}, s_t)$, the sample likelihood can be resolved. According to Bayes’ rule,

$$P(s_{t-1}|\Phi_{t-1}, s_t) = \frac{P(s_{t-1}|\Phi_{t-1})P(s_t|s_{t-1}, \Phi_{t-1})}{P(s_t|\Phi_{t-1})}$$

(23)

$$P(s_t|\Phi_{t-1}) = \sum_{s_{t-1}=1,2} P(s_{t-1}|s_{t-1}, \Phi_{t-1})P(s_{t-1}|\Phi_{t-1})$$

(24)

where $P(s_{t-1}|s_{t-1}, \Phi_{t-1}) = P(s_{t-1})$ is the transition probability. The computation of ex post regime probability $P(s_{t-1}|\Phi_{t-1})$ and ex ante regime probability $P(s_t|\Phi_{t-1})$ is discussed in Hamilton (1994) in details, by applying a first-order recursive mechanism. That is,

$$P(s_{t-1}|\Phi_{t-1}) = \frac{f(R_{t-1}|s_{t-1}, \Phi_{t-2}) \sum_{s_{t-2}=1,2} (P(s_{t-2}|\Phi_{t-2})P(s_{t-1}|s_{t-2}, \Phi_{t-2}))}{f(R_{t-1}|\Phi_{t-2})}$$

(25)
And

\[ f(R_{t-1}|s_{t-1}, \Phi_{t-2}) = \sum_{j=0}^{\infty} f(R_{t-1}|N(t-1) = j, \Phi_{t-2}) P(N(t-1) = j|\Phi_{t-2}, s_{t-1}) \]  

(26)

Thus, based on its previous value \( P(s_{t-2}|\Phi_{t-2}) \), the previous densities \( f(R_{t-1}|N(t-1) = j, \Phi_{t-2}) \) and \( f(R_{t-1}|s_{t-1}, \Phi_{t-2}) \) can be computed and be used to calculate \( P(s_{t-1}|\Phi_{t-1}, s_t) \), from which \( E[\lambda_{t-1}|\Phi_{t-1}, s_t] \) and \( E[\xi_{t-1}|\Phi_{t-1}, s_t] \) can be derived. Then \( \lambda_t \) is available to compute the densities \( f(R_t|s_t, \Phi_{t-1}) \) and \( f(R_t|\Phi_{t-1}) \) of this period. Then the sample likelihood can be computed.

As the log likelihood function has a closed form expression, maximum likelihood method can be applied to estimate the model when the regime variable is exogenous.

3 Threshold GARCH-jump model with exogenous trigger

3.1 Model

One main shortcoming of the hidden Markov regime switching model is that the regimes of each period are not known to the econometricians, and this leads to difficulty for forecasting, especially when used to forecast volatility a few days later. Threshold models with observable triggers can solve this problem. In threshold models, when the trigger is below the threshold value, the economy is in regime 1, while it is in regime 2 otherwise. Thus, in each period, the regime can be observed, which is an exogenous variable. The difficulty of using threshold models is because of the very limited theory developed about the existence of stationary conditions and expressions for stationary distribution.

Recently, Knight and Satchell (2010) derive necessary and sufficient con-
ditions for the existence of a stationary distribution of threshold-AR (1) model with an exogenous trigger variable. As our GARCH-jump model is a GARCH model with AR (1) jump intensity, this conditions for threshold-AR (1) can be applied to threshold GARCH-jump model with exogenous trigger variable. The threshold GARCH-jump model is

\[ R_t = \mu + \epsilon_{1,t} + \sum_{k=1}^{N(t)} Y_{t,k} \]  

\[ \epsilon_{1,t} = z_t \sigma_t \]  

\[ Y_{t,k} \sim i.i.d N(\theta, \delta^2) \]  

\[ P(N(t) = j|\Phi_{t-1}) = \exp(-\lambda_t)\lambda_t^j/j! \text{ for } j = 0, 1, 2... \]  

\[ \sigma_t^2 = \omega_1 + a_1 \epsilon_{t-1}^2 + b_1 \sigma_{t-1}^2 \text{ if } \nu_{t-1} \leq \nu_0 \]  

\[ \lambda_t = \alpha_1 + \rho_1 \lambda_{t-1} + \gamma_1 \xi_{t-1} \text{ if } \nu_{t-1} \leq \nu_0 \]  

\[ \xi_{t-1} = E[N(t-1)|\Phi_{t-1}] - \lambda_{t-1} \]  

\[ \sigma_t^2 = \omega_2 + a_2 \epsilon_{t-1}^2 + b_2 \sigma_{t-1}^2 \text{ if } \nu_{t-1} > \nu_0 \]  

\[ \lambda_t = \alpha_2 + \rho_2 \lambda_{t-1} + \beta_2 \sigma_{t-1}^2 + \gamma_2 \xi_{t-1} \text{ if } \nu_{t-1} > \nu_0 \]  

\( \Phi_{t-1} \) denotes the information up to time \( t - 1 \), which includes the time series of \( R_t \) and threshold variable \( \nu_t \) up to time \( t - 1 \). In this model setting,
there are two regimes as well as in the previous hidden Markov regime-switching model. The parameters in the GARCH conditional variance and jump intensity depends on the threshold variable $\nu_t$. In hidden Markov model regimes in each period are unknown to the econometricians both before and after the estimation, however, in the threshold model the regimes are known to the econometricians, which is very helpful in the estimation as well as in forecasting. For example, there is no need to integrate the previous regimes out in the threshold model, which simplifies the estimation algorithm a lot.

Let $s_t = 0$ if $\nu_t \leq \nu_0$, and $s_t = 1$ if $\nu_t > \nu_0$. We assume that $s_t$ follows a i.i.d Bernoulli distribution with $P(s_t = 1) = \pi$.

### 3.2 Stationary Conditions and Moments of Returns

**Proposition 3.1** $\lambda_t$ is strictly stationary if $\ln|\rho_1(1 - \pi)| + \ln|\rho_2|\pi < 0$.

The return series is covariance-stationary if $|\rho_1(1 - \pi)| + |\rho_2|\pi < 1$ and $|(a_1 + b_1)|(1 - \pi) + |(a_2 + b_2)|\pi < 1$. The mean of return is given by

$$E(R_t) = \mu + \theta E(\lambda_t) = \mu + \theta(\frac{\alpha_1(1 - \pi) + \alpha_2\pi}{1 - \rho_1(1 - \pi) - \rho_2\pi})$$

The variance of return is given by

$$Var(R_t) = \frac{\omega_1(1 - \pi) + \omega_2\pi}{1 - (a_1 + b_1)(1 - \pi) - (a_2 + b_2)\pi}$$

$$+ \frac{(\theta^2 + \delta^2)(\alpha_1(1 - \pi) + \alpha_2\pi)(1 - b_1(1 - \pi) - b_2\pi)}{(1 - \rho_1(1 - \pi) - \rho_2\pi)(1 - (a_1 + b_1)(1 - \pi) - (a_2 + b_2)\pi)}$$

The conditional skewness and kurtosis are given by

$$Skewness(R_t|\Phi_{t-1}) = \frac{\lambda_t(\theta^3 + 3\theta\delta^2)}{(\sigma_t^2 + \lambda_t\delta_t^2 + \lambda_t\theta^2)^{3/2}} \quad (36)$$

$$Kurtosis(R_t|\Phi_{t-1}) = 3 + \frac{\lambda_t(\theta^4 + 6\theta^2\delta^2 + 3\delta^4)}{(\sigma_t^2 + \lambda_t\delta_t^2 + \lambda_t\theta^2)^2} \quad (37)$$
The derivation of the above conditional moments is from Das and Sundaram (1997). The skewness is positive if $\theta > 0$. Both skewness and kurtosis depend on the conditional jump intensity $\lambda_t$, the jump size’s mean $\theta$ and variance $\delta^2$. The conditional kurtosis is larger than 3 in the presence of jumps, as the existence of outliers leads to fatter tails in the return distribution.

The estimation of the threshold GARCH-jump model in-sample is conducted by MLE. After the threshold value $\nu_0$ is chosen, the regime of each observation is known by comparison of the threshold variable and $\nu_0$. Given $\nu_0$, the MLE estimator is obtained by maximizing the log likelihood function. Thus, both the MLE estimator and the log likelihood value are functions of $\nu_0$.

Construction of the likelihood function is similar to that of GARCH-jump model. The conditional density of returns is normal given on $j$ jumps occurring,

$$f(R_t|N(t) = j, \Phi_{t-1}) = \frac{1}{\sqrt{2\pi(j\delta^2 + \sigma_t^2)}} \exp\left(-\frac{(R_t - \mu - j\theta)^2}{2(j\delta^2 + \sigma_t^2)}\right)$$  \hspace{1cm} (38)

Then the conditional density of returns can be got by integrating out the number of jumps occurring,

$$f(R_t|\Phi_{t-1}) = \sum_{j=0}^{\infty} f(R_t|N(t) = j, \Phi_{t-1}) P(N(t) = j|\Phi_{t-1})$$  \hspace{1cm} (39)

When constructing the jump intensity, the ex post filter can be built via Bayes’ rule as,

$$P(N(t) = j|\Phi_t) = f(R(t)|N(t) = j, \Phi(t-1)) P(N(t) = j|\Phi(t-1))/f(R(t)|\Phi(t-1))$$  \hspace{1cm} (40)

for $j = 0, 1, 2, \ldots$. Then the jump intensity residual is available and the autoregressive jump intensity can be constructed. In order to find $\nu_0$ which maximizes the log likelihood value, the sample of the threshold variable is divided into 20 intervals with 19 grid points from 5 percentile point to 95 percentile point.
Table 1: Descriptive Statistics of Daily Returns of Japanese Yen

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs</td>
<td>3751</td>
<td>3500</td>
<td>251</td>
</tr>
<tr>
<td>Mean</td>
<td>-0.009</td>
<td>-0.009</td>
<td>-0.012</td>
</tr>
<tr>
<td>Std. Deviation</td>
<td>0.703</td>
<td>0.707</td>
<td>0.651</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.508</td>
<td>-0.552</td>
<td>0.277</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>7.027</td>
<td>7.176</td>
<td>3.903</td>
</tr>
<tr>
<td>Min</td>
<td>-5.630</td>
<td>-5.630</td>
<td>-1.550</td>
</tr>
<tr>
<td>Max</td>
<td>3.240</td>
<td>3.240</td>
<td>2.452</td>
</tr>
<tr>
<td>$</td>
<td>R</td>
<td>&gt; 2$</td>
<td>59</td>
</tr>
<tr>
<td>$</td>
<td>R</td>
<td>&gt; 3$</td>
<td>13</td>
</tr>
</tbody>
</table>

4 Data and estimation

4.1 Data

We use the Japanese Yen- US Dollar spot exchange rate. The in-sample period contains 3500 daily observations, which starts from January 2nd, 1990 to January 9th, 2004. The data is accessed by Wharton Research Data Service (WRDS), and is obtained from Bank of Japan. The return $R_t$ is calculated to be 100 times the log difference of exchange rate $R_t$. 251 observations from January 12th, 2004 to January 11th, 2005 are used as the out-of-sample period for forecasting purpose. Table 1 provides summary statistics for percent returns for daily Japanese Yen exchange rate according to different sample periods.

The exogenous threshold variable used in the paper is Chicago Board Option Exchange (CBOE) S&P 500 Volatility Index (VIX), which is a key measure of market expectations of near-term 30-day implied volatility built by S&P 500 stock index option prices. It is reasonable to assume that when VIX is high, the market has an expectation of high volatility of stocks. When
VIX is higher than some particular value, we assume that the market enters into a regime that is more volatile. Furthermore, as VIX is a 30-day expectation built on a market index, it is also reasonable to assume that it is exogenous of the current volatility of the return of a specific stock or exchange rate. The VIX series is also accessed by Wharton Research Data Service (WRDS) from January 2nd, 1990 to January 11th, 2005.

4.2 Estimation

Table 2 reports the MLE estimates with standard errors in the bracket, and corresponding log likelihood values using threshold GARCH (1,1)-jump AR (1) model (TS-GARJI), threshold GARCH (1,1) model (TS-GARCH), and regime switching GARCH (1,1)-jump AR (1) model (RS-GARJI), together with the results using GARCH (1,1) model. Akaike’s information criterion and Bayesian information criterion are also included to compare goodness of fit of models.

Table 2 shows that the parameters in the threshold GARJI model are significant except the intercepts in the GARCH variance term and jump intensity AR (1) term. For the regime switching GARJI model, the parameters of jump intensity in the second regime are insignificant. As the second regime is the less volatile regime with a much smaller unconditional variance, it implies that there are no jumps in the less volatile period. Therefore the regime switching GARJI model is re-estimated with only jumps in the more volatile period. Both AIC and BIC suggest that the threshold GARJI model has the best fit of data among all the models, while GARCH (1,1) model has the worst. The parameters in threshold GARJI model satisfy the stationary condition, and in each regime the summation of $a$ and $b$ are less than one. The threshold GARCH (1,1) model has a threshold value at 70%, which implies that there is chance of 30% that the conditional variance shifts to regime 2. When using threshold GARJI model, the threshold value is at 55%, implying
Table 2: Estimates and log likelihood values using different models

<table>
<thead>
<tr>
<th>Parameter</th>
<th>TS-GARJI</th>
<th>TS-GARCH</th>
<th>RS-GARJI</th>
<th>GARCH(1,1)</th>
<th>ARJI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>0.035</td>
<td>-0.006</td>
<td>0.038</td>
<td>-0.007</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.011)</td>
<td>(0.012)</td>
<td>(0.011)</td>
<td>(0.012)</td>
</tr>
<tr>
<td>$\omega_1$</td>
<td>0.003</td>
<td>0.011</td>
<td>0.006</td>
<td>0.009</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.007)</td>
<td>(0.001)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.011</td>
<td>0.033</td>
<td>0.008</td>
<td>0.041</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.008)</td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>$b_1$</td>
<td>0.966</td>
<td>0.938</td>
<td>0.984</td>
<td>0.941</td>
<td>0.968</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.008)</td>
<td>(0.031)</td>
<td>(0.006)</td>
<td>(0.008)</td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>0.009</td>
<td>0.014</td>
<td>0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.004)</td>
<td>(0.002)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$a_2$</td>
<td>0.011</td>
<td>0.062</td>
<td>0.019</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.006)</td>
<td>(0.008)</td>
<td>(0.006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$b_2$</td>
<td>0.956</td>
<td>0.920</td>
<td>0.961</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.012)</td>
<td>(0.013)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>0.006</td>
<td>0.026</td>
<td>0.017</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.024)</td>
<td>(0.008)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho_1$</td>
<td>0.970</td>
<td>0.935</td>
<td>0.901</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.024)</td>
<td>(0.098)</td>
<td>(0.048)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>0.081</td>
<td>0.964</td>
<td>0.184</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.048)</td>
<td>(0.423)</td>
<td>(0.066)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_2$</td>
<td>0.004</td>
<td>0.063</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.104)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho_2$</td>
<td>0.987</td>
<td>0.640</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.582)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.146</td>
<td>-0.077</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.058)</td>
<td>(0.111)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\theta$</td>
<td>-0.219</td>
<td>-0.265</td>
<td>-0.292</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.066)</td>
<td>(0.071)</td>
<td>(0.082)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.912</td>
<td>0.917</td>
<td>0.984</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.075)</td>
<td>(0.083)</td>
<td>(0.088)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\nu_0$</td>
<td>20.235</td>
<td>22.765</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p_{11}$</td>
<td>0.953</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.053)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p_{22}$</td>
<td>0.983</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>logLLF</td>
<td>-3426.2</td>
<td>-3563.9</td>
<td>-3426.3</td>
<td>-3571.4</td>
<td>-3437.2</td>
</tr>
<tr>
<td>AIC</td>
<td>6882.4</td>
<td>7141.8</td>
<td>6886.6</td>
<td>7150.8</td>
<td>6892.4</td>
</tr>
<tr>
<td>BIC</td>
<td>6974.8</td>
<td>7184.9</td>
<td>6991.3</td>
<td>7175.4</td>
<td>6947.8</td>
</tr>
</tbody>
</table>
that there is chance of 45% for the conditional variance and jump intensity to shift to regime 2. Including jump term better fits the data, as threshold GARJI model outperforms the threshold GARCH (1,1) model.

The mean reported in threshold GARJI and threshold GARCH model are not the mean of the return, which is the reason that it is different from the mean of return in GARCH (1,1) model. The persistence parameters in jump intensity process are high for both regimes in the threshold GARJI model. After using a threshold GARCH (1,1) model, we find that the persistence parameter in the diffusive conditional variance are lower in each regime than in the GARCH (1,1) model. However, after jumps are incorporated, the persistence parameter in the conditional variance is higher in each regime for both threshold GARJI model and regime switching GARJI model than in the GARCH (1,1) model. It implies that separating the effects of jumps and diffusive volatility makes the volatility more persistent, while the high persistence of diffusive volatility of GARCH (1,1) model may come from the reason that different regimes are not identified.

5 Forecasting

In order to measure the forecasting performance of regime switching GARCH-Jump models, we use realized volatility as the proxy ex post daily volatility. The data set for constructing realized volatility contains the five-minute transaction price for Japanese Yen-US dollar spot exchange rate from January 12th, 2004 to January 11th, 2005. We use five minute data as it is considered the highest frequency at which prices are less distorted by the market microstructure noise. Following Andersen and Bollerslev (1998), the trading day t starts from 21:00 GMT on day $t - 1$ to 21:00 GMT on day $t$, which ensures that all transactions on day t of local time take place during
this period. Weekend days and major holidays are deducted for the reason of too many missing values or slower trading pattern. 251 days are left after the deduction.

We denote the five-minute exchange rates on day \( t \) by \( P_{m,t} \), for \( m = 1, 2, ..., M \). For five-minute data we have \( M = 288 \). The five minute return \( r_{m,t} \) is constructed as \( r_{m,t} = 100(\ln P_{m,t} - \ln P_{m-1,t}) \), for \( m = 1, 2, ..., M \), and \( t = 1, 2, ..., 251 \). The realized volatility is obtained by summing up the squared intra-day 5-minute returns as \( RV_t = \sum_{m=1}^{M} r_{m,t}^2 \). Andersen and Bollerslev (1998) presents that, under a jump-diffusion semi-martingale setting of price process with bounded jump intensity \( \lambda_t \), the realized volatility is consistent for the quadratic variation of the logarithmic price process in absence of microstructure noise. As the quadratic variation consists of both the diffusive volatility and the cumulative squared jumps, the realized volatility is a good proxy for volatility when jumps are taken into consideration as in our models. Figure 1 plots the evolution of the realized volatility constructed using 5-minute intraday returns in the out-of-sample period.

We use the respective parameters estimated from in-sample period to conduct conditional variance forecasts for all the models. A rolling scheme is used, that is, the in-sample period contains 3500 observations and moves forward every 50 observations. Figure 2, Figure 3 and Figure 4 depict the out-of-sample one-step-ahead forecasts of conditional variance of GARCH(1,1) model, TSGARJI (1,1) model, and RSGARJI (1,1) model respectively. From the figures we can find that conditional variances conducted from Markov regime-switching GARCH-jump model has a bigger variation than those from GARCH(1,1) model and threshold GARCH-jump model. Although Markov regime-switching GARCH-jump model fits the data better sometimes when there is a peak in the realized volatility, it also makes some worse forecasts. When realized volatility is quite high, conditional variance from threshold GARCH-jump model cannot catch up with it. One possible reason is that
jump size could be an increasing function of volatility. As $\text{Var}(R_t|\Phi_{t-1}) = \sigma_t^2 + (\theta^2 + \delta^2)\lambda_t$, $\theta$ and $\delta$ can also play a role in determining the conditional variance. We can let them be functions of $\sigma_t$ or $\lambda_t$.

For evaluating the forecasts, we run a linear regression of realized volatility on its forecast. Then the coefficient of determination, $R^2$, provides a guide to the accuracy of volatility forecasts. We evaluate the one-day-ahead out-of-sample volatility forecasts using the following regression,

$$RV_t = c + d\text{Var}(R(t)|\Phi_{t-1}) + \text{error}_t$$

(41)

where $\text{Var}_{t-1}(R(t))$ is the out-of-sample conditional variance forecast for day $t$ of a particular model. $R^2$ can be used to evaluate forecasting models as it shows how much of the variation in realized volatility can be explained by the variation of conditional variance forecasts. Table 3 reports the $R^2$, mean square error (MSE), mean absolute error (MAE) and mean absolute percent
Figure 2: Out-of-sample conditional variance forecast using GARCH(1,1) model

Figure 3: Out-of-sample conditional variance forecast using threshold-GARJI model
error (MAPE) for different models.

From the table we find that threshold-GARJI model performs best to explain the variation of the data out-of-sample in terms of $R^2$ among the models. The performance of threshold-GARJI model is better than both GARCH(1,1) model and ARJI model in terms of mean absolute error and mean absolute percent error. The performance of the single regime ARJI model, is worse than even the GARCH (1,1) model which does not take into

<table>
<thead>
<tr>
<th></th>
<th>TS-GARJI</th>
<th>TS-GARCH</th>
<th>RS-GARJI</th>
<th>ARJI</th>
<th>GARCH (1,1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.2077</td>
<td>0.1903</td>
<td>0.1030</td>
<td>0.1499</td>
<td>0.1895</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0579</td>
<td>0.0661</td>
<td>0.0658</td>
<td>0.0578</td>
<td>0.0579</td>
</tr>
<tr>
<td>MAE</td>
<td>0.1569</td>
<td>0.1654</td>
<td>0.1697</td>
<td>0.1604</td>
<td>0.1573</td>
</tr>
<tr>
<td>MAPE</td>
<td>0.2775</td>
<td>0.2778</td>
<td>0.3042</td>
<td>0.3040</td>
<td>0.2821</td>
</tr>
</tbody>
</table>
consideration of jumps. The reason is that the out-of-sample period is a tranquil period, in which only one out of 251 observations has an absolute value that is larger than 2; however, in the in-sample period, there are 58 out of 3500 observations whose absolute values are larger than 2. Thus the ARJI model overestimates the jump part and leads to inaccurate forecasts. The threshold GARJI model has the advantage to distinguish tranquil period from volatile period, leading to more accurate forecasting performance. The performance of the Markov regime-switching GARJI model is not good, which is generally found in the exchange rate literature.

6 Conclusion

In this paper we have developed two models to model jumps and regime switching at the same time. The data generating process is assumed to be a combination of a normal innovation capturing small and smooth changes in volatility and a compound poisson process with autoregressive jump intensity to model large and abrupt changes in return. Meanwhile, we present switching regimes in our models to account for the phenomenon that the persistence of individual shocks is lower during periods of extreme volatility. Regimes are incorporated either using a hidden first-order Markov process or through an exogenous threshold variable. The empirical results indicate that jump intensity has a significant level of persistence, and regime switching GARCH-jump models outperform GARCH model in sample. We also find that the persistence of diffusive volatility is lower for a threshold GARCH (1,1) than a single regime GARCH model, which is in accordance with previous literature. However, the persistence parameter is higher in each regime for both regime switching models than a single regime GARCH model after jumps are incorporated, indicating that separating the effects of jumps and smooth
changes in return makes the GARCH type conditional variance more persistent. Out-of-sample forecasts suggest that threshold GARCH-jump model has a good ability to forecast volatility of Japanese Yen-US Dollar exchange rate and it performs better than a single regime GARCH (1,1) model.

Appendix: Proofs of Results

Proof of Proposition 2.1:
As \( \lambda_t = E[N(t)|\phi_{t-1}, s_t] \), applying the law of iterated expectations to equation (7), we get

\[
E[N(t)|s_t] = E[\lambda_t|s_t]
\]

\[
= \alpha_{s_t} + \rho_{s_t} E[E[N(t-1)|\phi_{t-2}, s_{t-1}]|s_t] + \beta_{s_t} E[\sigma^2_t|s_t] + \gamma_{s_t} E[\nu_t|s_t]
\]

\[
= \alpha_{s_t} + \rho_{s_t} E[E[N(t-1)|\phi_{t-2}, s_{t-1}, s_t]|s_t] + \beta_{s_t} E[\sigma^2_t] + \gamma_{s_t} E[\nu_t]
\]

\[
= \alpha_{s_t} + \rho_{s_t} E[N(t-1)|s_{t-1}]|s_t] + \tau_{s_t}
\]

Substituting the latter two equations into the first equation leads to following equations.

\[
E[E[N(t-1)|s_{t-1}]|s_t = 1] = E[N(t-1)|s_{t-1} = 1]P(s_{t-1} = 1|s_t = 1)
\]

\[
+ E[N(t-1)|s_{t-1} = 2]P(s_{t-1} = 1|s_t = 1)
\]

\[
E[E[N(t-1)|s_{t-1}]|s_t = 2] = E[N(t-1)|s_{t-1} = 1]P(s_{t-1} = 1|s_t = 2)
\]

\[
+ E[N(t-1)|s_{t-1} = 2]P(s_{t-1} = 1|s_t = 2)
\]

Substituting the latter two equations into the first equation leads to following equations.

\[
\lambda_1 = E[N(t)|s_t = 1]
\]

\[
= \alpha_1 + \rho_1 P(s_{t-1} = 1|s_t = 1)\lambda_1 + \rho_1 P(s_{t-1} = 2|s_t = 1)\lambda_2 + \tau_1 \quad (42)
\]
\[ \lambda_2 = E[N(t)|s_t = 2] = \alpha_2 + \rho_2 P(s_{t-1} = 1|s_t = 2)\lambda_1 + \rho_2 P(s_{t-1} = 2|s_t = 2)\lambda_2 + \tau_2 \] (43)

Organizing the above expressions for \( \lambda_1 \) and \( \lambda_2 \) leads to

\[ \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = B^{-1} \begin{bmatrix} \alpha_1 + \tau_1 \\ \alpha_2 + \tau_2 \end{bmatrix} \] (44)

\[ B = \begin{bmatrix} 1 - \rho_1 P(s_{t-1} = 1|s_t = 1) & -\rho_1 P(s_{t-1} = 2|s_t = 1) \\ -\rho_2 P(s_{t-1} = 1|s_t = 2) & 1 - \rho_2 P(s_{t-1} = 2|s_t = 2) \end{bmatrix} \] (45)

By the Bayes’ rule, the conditional probability of \( s_{t-1} \) given \( s_t \) is easy to get.

\[ P(s_{t-1}|s_t) = \frac{P(s_{t-1})P(s_t|s_{t-1})}{\sum_{s_{t-1}=1,2} P(s_{t-1})P(s_t|s_{t-1})} \] (46)

By calculation, \( P(s_{t-1} = 1|s_t = 1) = p_{11}, P(s_{t-1} = 1|s_t = 2) = 1 - p_{22}, P(s_{t-1} = 2|s_t = 1) = 1 - p_{11}, \) and \( P(s_{t-1} = 2|s_t = 2) = p_{22}. \) So the unconditional intensities are given by equations (10) and (11).

**Proof of Proposition 3.1:**

The return at time \( t \) can be divided into 3 parts, the mean \( \mu \), and \( \epsilon_{1,t} \), whose conditional variance is \( \sigma^2_{1,t} \), and the jump part, whose conditional variance is \( (\theta^2 + \delta^2)\lambda_t \).

The conditional jump intensity equation can be rewritten as

\[ \lambda_t = \alpha_1 + (\alpha_2 - \alpha_1)s_{t-1} + \rho_1 \lambda_{t-1} + (\rho_2 - \rho_1)s_{t-1}\lambda_{t-1} + \gamma_1 \xi_{t-1} + (\gamma_2 - \gamma_1)s_{t-1}\xi_{t-1} \] (47)

Knight and Satchell (2010) present a TAR(1) model with constant intercept coefficient across regimes, and the above equation is a variation of their model with different coefficients in both intercept and error term.
Let
\[ c_0 = \alpha_1 (1 - \pi) + \alpha_2 \pi ; \quad c_1 = \alpha_2 - \alpha_1 \]
\[ d_0 = \rho_1 (1 - \pi) + \rho_2 \pi ; \quad d_1 = \rho_2 - \rho_1 \]
\[ e_0 = \gamma_1 (1 - \pi) + \gamma_2 \pi ; \quad e_1 = \gamma_2 - \gamma_1 \]
\[ B_{t-1} = s_{t-1} - \pi \]

Then
\[ \lambda_t = c_0 + c_1 B_{t-1} + (d_0 + d_1 B_{t-1}) \lambda_{t-1} + (e_0 + e_1 B_{t-1}) \xi_{t-1} \]  \hspace{1cm} (48)

We have \( P(B_{t-1} = -\pi) = 1 - \pi, \) \( P(B_{t-1} = 1 - \pi) = \pi. \)

Backward substitution in (48) leads to
\[ \lambda_t = c_0 + c_1 B_{t-1} + \sum_{n=1}^{k-1} (c_0 + c_1 B_{t-n-1}) \prod_{m=1}^{n} (d_0 + d_1 B_{t-m}) + (e_0 + e_1 B_{t-1}) \xi_{t-1} + \lambda_{t-k} \prod_{m=1}^{k} (d_0 + d_1 B_{t-m}) \]

Following Quinn (1982) and Knight and Satchell (2010), and letting \( G_n(t) = \prod_{m=1}^{n} (d_0 + d_1 B_{t-m}), \)
\[ \ln G_n(t) = \sum_{m=1}^{n} \ln (d_0 + d_1 B_{t-m}) \]

and \( \frac{1}{n} \ln |G_n(t)| \to E(\ln |d_0 + d_1 B_{t-m}|) \)

Then \( G_n(t) \xi_{t-n-1} \) are geometrically bounded if \( E(\ln |d_0 + d_1 B_{t-m}|) < 0, \) i.e,
\[ \pi \ln |d_0 + d_1 (1 - \pi)| + (1 - \pi) \ln |d_0 - d_1 \pi| < 0 \]
or
\[ \pi \ln |\rho_2| + (1 - \pi) \ln |\rho_1| < 0 \]
Then equation (48) has the solution that
\[
\lambda_t = c_0 + c_1 B_{t-1} + \sum_{n=1}^{\infty} G_n(t) (c_0 + c_1 B_{t-n-1}) + (e_0 + e_1 B_{t-1}) \xi_{t-1} + \sum_{n=1}^{\infty} (e_0 + e_1 B_{t-n-1}) \xi_{t-n-1} G_n(t) \tag{49}
\]

From Quinn (1982) and Feigin and Tweedie (1985), the mean of the stationary distribution will exist, i.e., \(E(|\lambda_t|) < \infty\), if \(E(|d_0 + d_1 B_{t-1}|) < 1\).

That is, \(|\rho_1(1 - \pi) + |\rho_2| \pi < 1\)

Consequently, given \(\rho_1(1 - \pi) + \rho_2 \pi < 1\), and \(\alpha_1 < \infty, \alpha_2 < \infty\), the unconditional mean of \(\lambda_t\) exists: \(E(\lambda_t) = c_0 + c_0 \left( \sum_{n=1}^{\infty} d_n^0 \right) = \frac{c_0}{1-d_0} = \frac{\alpha_1(1-\pi) + \alpha_2 \pi}{1-\rho_1(1-\pi)-\rho_2 \pi}\)

To find the unconditional expectation of \(\sigma_t\), we rewrite the GARCH-type conditional variance as
\[
\sigma_t^2 = \omega_1 + (\omega_2 - \omega_1) s_{t-1} + a_1 \epsilon_{t-1}^2 + (a_2 - a_1) s_{t-1} \epsilon_{t-1}^2 + b_1 \sigma_{t-1}^2 + (b_2 - b_1) s_{t-1} \epsilon_{t-1}^2 \tag{50}
\]
in which \(\epsilon_{t-1}^2 = \epsilon_{1,t-1}^2 + \epsilon_{2,t-1}^2 + 2 \epsilon_{1,t-1} \epsilon_{2,t-1}\)

Let
\[
f_0 = \omega_1 (1 - \pi) + \omega_2 \pi ; f_1 = \omega_2 - \omega_1
\]
\[
g_0 = a_1 (1 - \pi) + a_2 \pi ; g_1 = a_2 - a_1
\]
\[
h_0 = b_1 (1 - \pi) + b_2 \pi ; h_1 = b_2 - b_1
\]
\[B_{t-1} = s_{t-1} - \pi\]

Then
\[
\sigma_t^2 = f_0 + f_1 B_{t-1} + g_0 \epsilon_{t-1}^2 + g_1 \epsilon_{t-1}^2 B_{t-1} + h_0 \sigma_{t-1}^2 + h_1 \sigma_{t-1}^2 B_{t-1} = f_0 + f_1 B_{t-1} + (g_0 + g_1 B_{t-1}) (\epsilon_{t-1}^2 + 2 \epsilon_{1,t-1} \epsilon_{2,t-1})
\]
\[+ [(g_0 + g_1 B_{t-1}) \epsilon_{t-1}^2 + h_0 + h_1 B_{t-1}] \sigma_{t-1}^2 \tag{51}\]
Backward substitution in (51) leads to

\[ \sigma_t = f_0 + f_1 B_{t-1} + \sum_{n=1}^{k-1} \prod_{m=1}^{n} [(g_0 + g_1 B_{t-m}) z^2_{t-1} + h_0 + h_1 B_{t-m}] (f_0 + f_1 B_{t-m}) \]

\[ + (g_0 + g_1 B_{t-1}) (e^2_{2t-1} + 2 \epsilon_1 t-1 \epsilon_{2t-1}) + \sum_{n=1}^{k-1} (g_0 + g_1 B_{t-n-1}) \]

\[ (\epsilon^2_{2,t-n-1} + 2 \epsilon_1 t-n-1 \epsilon_{2,t-n-1}) \prod_{m=1}^{n} [(g_0 + g_1 B_{t-m}) z^2_{t-m} + h_0 + h_1 B_{t-m}] \]

\[ + \sigma^2_{t-k} \prod_{m=1}^{k} [(g_0 + g_1 B_{t-m}) z^2_{t-m} + h_0 + h_1 B_{t-m}] \]

Similarly, letting \( Q_n(t) = \prod_{m=1}^{n} [(g_0 + g_1 B_{t-m}) z^2_{t-m} + h_0 + h_1 B_{t-m}] \), if \( E(\ln[(g_0 + g_1 B_{t-m}) z^2_{t-m} + h_0 + h_1 B_{t-m}]) < 0 \), that is, \((a_1 + b_1)(1-\pi) + (a_2 + b_2)\pi < 1\), equation (51) has the solution that

\[ \sigma^2_t = f_0 + f_1 B_{t-1} + \sum_{n=1}^{\infty} Q_n(t) (f_0 + f_1 B_{t-n-1}) + (g_0 + g_1 B_{t-1}) (e^2_{2t-1} + 2 \epsilon_1 t-1 \epsilon_{2t-1}) \]

\[ + \sum_{n=1}^{\infty} (g_0 + g_1 B_{t-n-1}) (\epsilon^2_{2,t-n-1} + 2 \epsilon_1 t-n-1 \epsilon_{2,t-n-1}) Q_n(t) \]

Then the unconditional expectation can be computed as

\[ E(\sigma^2_t) = f_0 + \sum_{n=1}^{\infty} E(Q_n(t)) f_0 + g_0 E(\epsilon^2_{2t-1}) + \sum_{n=1}^{\infty} g_0 E(\epsilon^2_{2t-1}) E(Q_n(t)) \] (52)

We have \( E(\epsilon^2_{2,t-1}) = E(E(\epsilon^2_{2t-1} | \lambda_{t-1})) = E((\theta^2 + \delta^2) \lambda_{t-1}) = (\theta^2 + \delta^2) E(\lambda(t-1)) \)

\[ E(Q_n(t)) = E(\prod_{n=1}^{n} [(g_0 + g_1 B_{t-m}) z^2_{t-m} + h_0 + h_1 B_{t-m}]) = (g_0 + h_0)^n \]

Taking \( E(Q_n(t)) \) and \( E(\epsilon^2_{2,t-1}) \) into equation (52),

\[ E(\sigma^2_t) = \frac{f_0}{1-g_0-h_0} + (\theta^2 + \delta^2) \frac{(\alpha_1 (1-\pi) + \alpha_2 \pi) g_0}{(1-\rho_1 (1-\pi) - \rho_2 \pi) (1-g_0-h_0)} \] (53)
Thus, given $\omega_1 < \infty$, $\omega_2 < \infty$, $\alpha_1 < \infty$, $\alpha_2 < \infty$, and $\rho_1(1 - \pi) + \rho_2\pi < 1$, $(a_1 + b_1)(1 - \pi) + (a_2 + b_2)\pi < 1$, we have $Var(R_t) = E(\sigma_t^2) + (\theta^2 + \delta^2)E(\lambda_t)$ to calculate the value of $Var(R_t)$, which is stated in the equation (36) in the proposition.
References


