Continuous Time Stochastic Volatility Models with Regime Shifts

Seungmoon Choi*

School of Economics, University of Seoul

Di Yuan†

College of Finance and Statistics, Hunan University

November, 2015

Abstract

In this paper, we estimate continuous time stochastic volatility models considered by Aït-Sahalia and Kimmel (2007) in a more general setting. Regime-switching Heston, GARCH, and CEV models where all parameters are allowed to vary depending on the state of the economy are suggested and examined. Hamilton (1989) algorithm is used to compute likelihood and conduct maximum likelihood estimation. In doing so, it is necessary to know the true transition probability density function (TPDF) of a stochastic volatility model. But the true TPDF is unavailable for all models considered in this paper. So, we adopted the irreducible approach established by Aït-Sahalia (2008) and Choi (2015b) to obtain closed-form log-TPDFs of those models and then to get TPDFs. Using S&P 500 and VIX data for the stock price and volatility proxy, respectively, we investigate which model can describe the dynamics of data better. We found strong evidence of regime shifts for all models. The CEV model is statistically preferred to the other two nested models in explaining the evolutions of the data. Moreover, the transition probability of the regime variable changes with the stock price rather than its volatility.

KEY WORDS: Continuous-Time Stochastic Volatility Model; Regime-Switching; Maximum likelihood estimation

*Contact: Seungmoon Choi, School of Economics, University of Seoul, Email: schoi22@uos.ac.kr

†College of Finance and Statistics, Hunan University, Email: di_yuan@hnu.edu.cn

1 Introduction

Since Black and Scholes (1973)'s and Merton (1973)'s pioneering work on option pricing theory, researchers including Black (1976) have questioned about validity of a geometric Brownian motion employed by them to describe the dynamics of stock prices. Modeling stock prices with a geometric Brownian motion implies that logarithmic stock price is normally distributed with constant mean and variance, which does not seem to reflect real data observed in the market. In the light of this, stochastic volatility models have been proposed by, for example, Scott (1987), Hull and White (1987), Stein and Stein (1991), and Heston (1993) to get a more realistic models to explain some features of data better. Recently, Aït-Sahalia and Kimmel (2007) estimated several continuous time stochastic volatility models by maximum likelihood (ML) estimation method. Chacko and Viceira (2003) and Jones (2003) also estimated some stochastic volatility models by using GMM and MCMC methods, respectively.

Looking at the data, however, existing stochastic volatility models do not appear to be able to fit the data well. Going through the dot-come bubble growth and bust period and ongoing global financial and the Euro debt crises, it seems to be too strong to assume that only one data generating process can explain the movements of stock prices. So, it might be more reasonable to ask whether there has been any changes in the underlying data generating process governing dynamics of stock price and its volatility. To do this, it is better to let the data answer to this question in a more general settings. One popular way to address this issue is to incorporate regime-switching features into a model.

We estimate the stochastic volatility models examined by Aït-Sahalia and Kimmel (2007) in a more general setting. Regime-switching Heston, GARCH, and CEV models where all parameters are allowed to vary depending on the state of the economy are estimated by MLE in this article. Hamilton (1989) algorithm is used to compute likelihood and conduct MLE. In doing so, it is necessary to know the true transition probability density function (TPDF) of a stochastic volatility model. But the true TPDF is unavailable for all models considered in this paper. So, we adopted the irreducible approach established by Aït-Sahalia (2008) and Choi (2015b) to obtain closed-form log-TPDF of those models and then to get TPDF.

After Aït-Sahalia (2002) makes a breakthrough in finding closed-form approximate TPDF of a univariate time-homogeneous diffusion, many researchers have extended the results to more general cases such as TPDF of a univariate time-inhomogeneous diffusion by Egorov, Li, and Xu (2003), TPDF of univariate time homogeneous diffusion models driven by Lévy processes by Schaumburg (2001), log-TPDF of multivariate time homogeneous diffusions by Aït-Sahalia (2008), TPDF of multivariate time homogeneous jump diffusion models by Yu (2007), log-TPDF of multivariate time inhomogeneous diffusion models by Choi (2013), Bayesian setting by DiPietro (2001) and Stramer, Bognar, and Schneider (2010), TPDF of damped diffusion by Li (2010), TPDF of multivariate diffusion models by Choi (2015b), and TPDF of multivariate time inhomogeneous jump diffusion models by Choi (2015a).

In the literature, regime-switching diffusion models have been used to model interest rates (Naik and Lee (1998), Liechty and Roberts (2001), and Choi (2009)). Because the true TPDF of a diffusion process is unknown in general, people often discretized continuous-time diffusion models to estimate them (Ang and Bekaert (2002), Bansal and Zhou (2002), and Dai, Singleton, and Yang (2007)). In a recent paper, Durham and Park (2013) applied some regime-switching stochastic volatility models to stock price and volatility data. Their models are different from ours and they also discretized diffusion models for estimation. Moreover, discretization of the continuous time diffusion is well known to yield biased and inefficient estimates (Lo (1988)). Therefore, we can expect to obtain more precise estimates for the parameters of our models making use of quite accurate TPDF expansions.

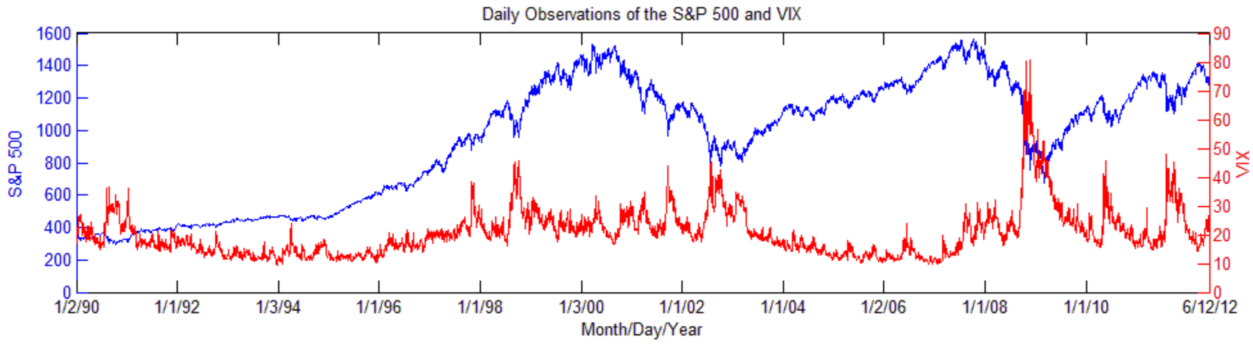
Using S&P 500 and VIX data for the stock price and to generate volatility proxy, respectively, we investigate which model can describe the dynamics of data better. We found strong evidence of regime shifts for all models. The CEV model is statistically preferred to the other two nested models in explaining dynamics of data. Moreover, the transition probability of regime variable changes with the stock price rather than its volatility.

Organization of this paper is as follows. Discussing some features of data which motivated this article in the next section, three regime-switching Heston, GARCH and CEV models will be introduced in Section 3. The ideas of Hamilton algorithm and approximation method of getting the TPDF of diffusion models are presented in Section 4. In Section 5, all estimation results are exhibited and discussed. Then we conclude.

2 Data and Motivation

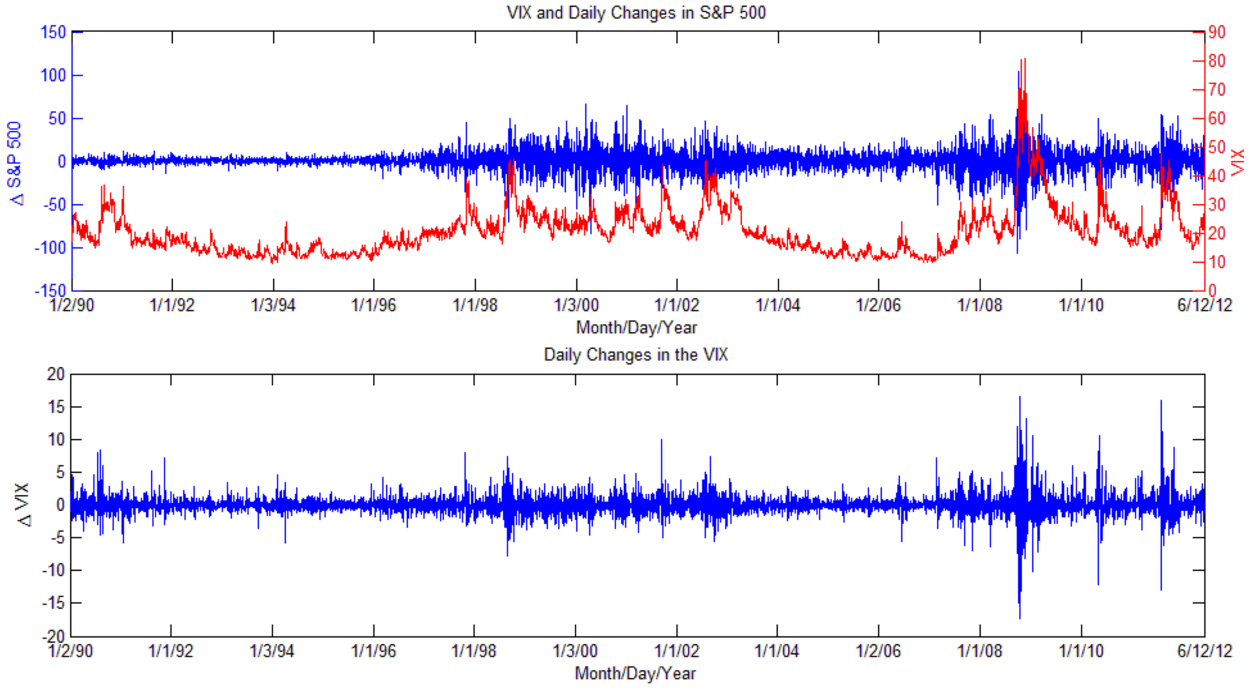
Daily observations of the Standard & Poor's 500 (S&P 500) Index and the Chicago Board Options Exchange (CBOE) Volatility Index, VIX from January 2, 1990 till June 12, 2012 have been downloaded from DataStream. These two series are depicted in Figure 1, where the left y-axis corresponds to the S&P 500 data plot in blue and the right y-axis is the line plot in red for the VIX data. On the top panel of Figure 2, daily changes in the S&P 500 and the VIX data are plotted. Here, the left y-axis and right y-axis are for the changes in the S&P 500 and the VIX, respectively Daily changes in the VIX are graphed on the bottom panel in Figure 2.

Figure 1. Daily S&P 500 and VIX from Jan/2/90 till June/12/12



From visual inspection of Figure 1, we can see that the VIX tends to increase more when the S&P 500 decreases, which is well known as the leverage effect of asset prices. That is, the S&P 500 appears to be negatively correlated with the VIX. Another well documented phenomenon that can be observed from Figure 2 is volatility clustering which terms the behavior that large changes in a variable tend to be followed by large changes for a while and small changes tend to be followed by small changes for the same variable. Volatility clustering can be seen not only in the S&P 500 but also in the VIX.

Figure 2: Daily Changes in S&P 500 and VIX, and VIX from Jan/2/90 till Jun/12/12



Persistency of volatility in the S&P 500 can be explained by the second factor in a stochastic volatility model[1] because when the level of VIX is high the S&P 500 looks to be more volatile. On the other hand, volatility clustering is exhibited in the VIX as well. Moreover, the long run mean level of the VIX may not be unique. It seems to be too strong to assume that the dynamics of data are governed by only one data generating process. One way to check whether or not it is the case is to allow the underlying data generating process to change over time. This motivates to adopt a regime-switching stochastic volatility model so that the data can say whether or not there exists an additional regime.

[1] The ARCH (Engle (1982)) and GARCH (Bollerslev (1986) models provide alternative way to explain volatility clustering.

Table 1: Summary Statistics

	Observations	Minimum	Maximum	Mean	Std. Dev	Skewness Coeff	Degree of Excess
S&P 500	5856	295.4600	1565.1500	954.3139	370.4182	-0.3649	-1.2632
VIX	5856	9.3100	80.8600	20.5184	8.1869	1.9440	6.7626
ln (S&P 500)	5856	5.6885	7.3557	6.7631	0.4736	-0.7453	-0.9174
IV	5856	0.0087	0.6538	0.0488	0.0490	4.6274	33.6339

3 Model

Regime-switching is incorporated with three different continuous time stochastic volatility models. One of these is the CEV model that encompasses the other two, the Heston and GARCH, models. Starting with the Heston model, in what follows, we present how regime shifts are added to stochastic volatility models.

3.1 Heston Model with Regime Shifts

Even if the Black-Scholes-Merton model (Black and Scholes (1973) and Merton (1973)) has made a huge contribution to option pricing theory, they make a strong assumption that stock prices follow a geometric Brownian motion. In such a case, a stock price, conditioning on the past observation, is lognormally distributed with constant mean and variance. In order to improve upon the Black-Scholes-Merton model by relaxing assumptions about the dynamics of asset prices so that it can reflect what can be observed in data better Heston (1993) proposed the Heston model. Under the risk-neutral measure Q, the Heston model for a stock price, S_t and its local variance V_t is written as

$$d\begin{pmatrix} S_t \\ V_t \end{pmatrix} = \begin{bmatrix} rS_t \\ \kappa' \left(\gamma' - V_t\right) \end{bmatrix} dt + \begin{bmatrix} \sqrt{\left(1-\rho^2\right) V_t} S_t & \rho\sqrt{V_t} S_t \\ 0 & \sigma\sqrt{V}_t \end{bmatrix} d\begin{pmatrix} W_{1t}^Q \\ W_{2t}^Q \end{pmatrix}, \tag{1}$$

where W_{1t}^Q and W_{2t}^Q are independent Brownian motions and r is the risk-free interest rate[2]. The variance, V_t follows the Feller (1952)'s square root process which was also used by Cox, Ingersoll, and Ross (1985). The variable V_t is always greater than zero as long as Feller's condition $2k'\gamma' \geq \sigma^2$ holds. Rewriting (1) in terms of $s_t = \ln S_t$ and V_t we get

$$d\begin{pmatrix} s_t \\ V_t \end{pmatrix} = \begin{bmatrix} r - \frac{1}{2}V_t \\ \kappa'(\gamma' - V_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)V_t} & \rho\sqrt{V_t} \\ 0 & \sigma\sqrt{V}_t \end{bmatrix} d\begin{pmatrix} W_{1t}^Q \\ W_{2t}^Q \end{pmatrix}, \tag{2}$$

by Ito's lemma.

The market price of risk is specified as $\left[\lambda_1\sqrt{(1-\rho^2)V_t}, \lambda_2\sqrt{V}_t\right]^T$ following Aït-Sahalia and Kimmel (2007). Then the joint dynamics of s_t and V_t under the objective measure P are by the Girsanov theorem

$$d\begin{pmatrix} s_t \\ V_t \end{pmatrix} = \begin{bmatrix} r + bV_t \\ \kappa(\gamma - V_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)V_t} & \rho\sqrt{V_t} \\ 0 & \sigma\sqrt{V}_t \end{bmatrix} d\begin{pmatrix} W_{1t}^P \\ W_{2t}^P \end{pmatrix}, \tag{3}$$

where $b = \lambda_1(1-\rho^2) + \lambda_2\rho - \frac{1}{2}$, $\kappa = \kappa' - \lambda_2\sigma$, and $\gamma = \left(\frac{\kappa'}{\kappa' - \lambda_2\sigma}\right)\gamma'$.

We will use a proxy for V_t that is described in Section 5.1 to estimate (3) by maximum likelihood estimation (MLE) method. In this case only the parameters $\kappa, \gamma, \sigma, \rho, r,$ and β can be identified. Because both of λ_1 and λ_2 cannot be identified, we will assume $\lambda_2 = 0$ as in Aït-Sahalia and Kimmel (2007). Then (3) can be rewritten as

$$d\begin{pmatrix} s_t \\ V_t \end{pmatrix} = \begin{bmatrix} r + \left[\lambda_1(1-\rho^2) - \frac{1}{2}\right]V_t \\ \kappa(\gamma - V_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)V_t} & \rho\sqrt{V_t} \\ 0 & \sigma\sqrt{V}_t \end{bmatrix} d\begin{pmatrix} W_{1t}^P \\ W_{2t}^P \end{pmatrix}. \tag{4}$$

Notice that the parameter vector $\theta = (\kappa, \gamma, \sigma, \rho, r, \lambda_1)$. The parameter κ tells the speed of mean reversion of V_t and γ is its long run mean level. And ρ measures the correlation between the innovations to stock price and stochastic volatility and it can explain leverage skewness effects.

Let us incorporate regime shifts into the dynamics of (s_t, V_t) by using regime index R_t to capture possible regime-switching behavior of the data. Depending on the state of the

[2] We assume that the dividend yield of the stock is zero. Or r can be interpreted as the risk-free rate minus the dividend yield.

economy, there are two regimes such that $R_t = L$ or H. Therefore, for the Heston model there are two different sets of parameters, $(\kappa_L, \gamma_L, \sigma_L, \rho_L, r_L, \lambda_{1L})$ and $(\kappa_H, \gamma_H, \sigma_H, \rho_H, r_H, \lambda_{1H})$.

$$
\begin{aligned}
d\begin{pmatrix} s_t \\ V_t \end{pmatrix} &= \begin{bmatrix} r_{R_t} + \left[\lambda_{1R_t}\left(1-\rho_{R_t}^2\right) - \frac{1}{2}\right] V_t \\ \kappa_{R_t}\left(\gamma_{R_t} - V_t\right) \end{bmatrix} dt \\
&\quad + \begin{pmatrix} \sqrt{\left(1-\rho_{R_t}^2\right) V_t} & \rho_{R_t}\sqrt{V_t} \\ 0 & \sigma_{R_t}\sqrt{V_t} \end{pmatrix} d\begin{pmatrix} W_{1t}^P \\ W_{2t}^P \end{pmatrix}.
\end{aligned}
$$

The regime index R_t follows a continuous time first order Markov chain with two states. The infinitesimal matrix of the Markov chain is

$$
\Pi = \begin{pmatrix} \pi_{LL} & \pi_{HL} \\ \pi_{LH} & \pi_{HH} \end{pmatrix} = \begin{pmatrix} -\pi_{LH} & \pi_{HL} \\ \pi_{LH} & -\pi_{HL} \end{pmatrix},
$$

where the intensity parameter π_{ij} is the rate of the probability at which the process switches ($\pi_{ij} > 0$ for $i \neq j$). Assuming that regime shift can occur at most once during each period Δ, the transition matrix is

$$
P^{\Delta} = \frac{1}{\pi_{LH} + \pi_{HL}} \begin{pmatrix} \pi_{HL} + \pi_{LH} e^{-\Delta(\pi_{LH}+\pi_{HL})} & \pi_{HL}\left(1 - e^{-\Delta(\pi_{LH}+\pi_{HL})}\right) \\ \pi_{LH}\left(1 - e^{-\Delta(\pi_{LH}+\pi_{HL})}\right) & \pi_{LH} + q\pi_{HL} e^{-\Delta(\pi_{LH}+\pi_{HL})} \end{pmatrix} \tag{5}
$$

and unconditional probabilities are $\left(\frac{\pi_{HL}}{\pi_{LH}+\pi_{HL}}, \frac{\pi_{LH}}{\pi_{LH}+\pi_{HL}}\right)$. Reparametrizing (5), the transition matrix and the corresponding unconditional probabilities are respectively

$$
P = \begin{pmatrix} p_{LL} & p_{HL} \\ p_{LH} & p_{HH} \end{pmatrix} \text{ and } \left(\frac{1 - p_{HH}}{2 - p_{LL} - p_{HH}}, \frac{1 - p_{LL}}{2 - p_{LL} - p_{HH}}\right).
$$

Therefore the parameter vector to be estimated for the regime-switching Heston model is $\theta = (\kappa_L, \gamma_L, \sigma_L, \rho_L, r_L, \lambda_{1L}, \kappa_H, \gamma_H, \sigma_H, \rho_H, r_H, \lambda_{1H}, p_{LL}, p_{HH})$.

We also consider the case where the transition probability is dependent on the state variables. Time-varying transition probabilities are specified as

$$\begin{aligned}
p_{LL} &= P\left(R_t = L | R_{t-\Delta} = L\right) = F\left(c_L + d_L s_{t-\Delta} + e_L V_{t-\Delta}\right) \\
p_{LH} &= P\left(R_t = H | R_{t-\Delta} = L\right) = 1 - F\left(c_L + d_L s_{t-\Delta} + e_L V_{t-\Delta}\right) \\
p_{HH} &= P\left(R_t = H | R_{t-\Delta} = H\right) = F\left(c_H + d_H s_{t-\Delta} + e_H V_{t-\Delta}\right) \\
p_{HL} &= P\left(R_t = L | R_{t-\Delta} = H\right) = 1 - F\left(c_H + d_H s_{t-\Delta} + e_H V_{t-\Delta}\right).
\end{aligned}$$

In the literature a logistic function, $F(x) = e^x / (1 + e^x)$ (Diebold, Lee, and Weinbach (1994) and Dai, Singleton, and Yang (2007)), the cumulative standard normal distribution function (Gray (1996)) or both (Choi (2009)) have been adopted for $F(x)$. As Choi (2009) found out, it is not expected to make much difference in the estimation results whichever function is used for $F(\cdot)$. Because $F(\cdot)$ is a strictly increasing function and the argument of the function is a linear function of s_t and V_t, if $d > 0$ $(e > 0)$,then the probability of staying in the same regime in the next period increases as s_t (V_t) increases. When $d_L = d_H = e_L = e_H = 0$, the transition probability is time-invariant.

For each continuous time stochastic volatility model studied in this article, depending on the number of regimes, if additional parameter is used for the initial state probability or not, and whether or not the transition probability is constant, four different cases are estimated. R1 is for the single-regime model. If we use unconditional probability (additional parameter) for the probability of the initial state for a regime-switching model with time-invariant transition matrix it is R2-1 (R2-2). Finally, regime-switching model with time-varying transition matrix is R2TVTP[3]. Thus, with regard to the Heston model, MLE is implemented for Heston-R1, Heston-R2-1, Heston-R2-2, and Heston-R2TVTP models.

3.2 CEV Model with Regime Shifts

Jones (2003) points out that the Heston model is unlikely to be able to explain the evolution of the instantaneous volatility, $\sqrt{V_t}$ well. The reason is because the Heston model implies that the volatility of the instantaneous volatility is constant, whereas the volatility of the instantaneous volatility seems to be more volatile when it is relatively higher as can be seen in the bottom panel of Figure 2. To overcome this problem he considers the constant

[3] Additional parameter is required since the transition probability varies with the state variables.

elasticity of variance (CEV) model where an additional parameter β is used for the power of V_t in the diffusion process of the variance. That is, under the risk-neutral measure Q, the underlying asset price S_t and its volatility V_t follow the dynamics

$$d\begin{pmatrix} S_t \\ V_t \end{pmatrix} = \begin{bmatrix} rS_t \\ \kappa'(\gamma' - V_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)V_t}S_t & \rho\sqrt{V_t}S_t \\ 0 & \sigma V_t^{\beta} \end{bmatrix} d\begin{pmatrix} W_{1t}^Q \\ W_{2t}^Q \end{pmatrix}. \quad (6)$$

This model encompasses the Heston model, where $\beta = 1/2$ and the GARCH model, where $\beta = 1$ to be discussed in the next section. The diffusion process of V_t in (6) was introduced by Chan, Karolyi, Longstaff, and Sanders (1992) and has been popularly employed to model interest rates since then. The CEV model for the volatility also has been used by Lewis (2000) and Chacko and Viceira (2003).

In terms of the logarithmic stock price, (6) is rewritten as

$$d\begin{pmatrix} s_t \\ V_t \end{pmatrix} = \begin{bmatrix} r - \frac{1}{2}V_t \\ \kappa'(\gamma' - V_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)V_t} & \rho\sqrt{V_t} \\ 0 & \sigma V_t^{\beta} \end{bmatrix} d\begin{pmatrix} W_{1t}^Q \\ W_{2t}^Q \end{pmatrix}.$$

Using the same assumption for the market prices of risks, $\left[\lambda_1\sqrt{(1-\rho^2)Y_t}, 0\right]^T$ as in the Heston model, under the physical measure, dynamics of the state variables are represented as

$$d\begin{pmatrix} s_t \\ V_t \end{pmatrix} = \begin{bmatrix} r + bV_t \\ \kappa(\gamma - V_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)V_t} & \rho\sqrt{V_t} \\ 0 & \sigma V_t^{\beta} \end{bmatrix} d\begin{pmatrix} W_{1t}^P \\ W_{2t}^P \end{pmatrix}, \quad (7)$$

where $b = \lambda_1(1-\rho^2) - \frac{1}{2}$, $\kappa = \kappa'$, and $\gamma = \gamma'$. The true TPDF is unavailable for this model.

Finally, giving regime-switching structure to (7), the regime-switching CEV model is

$$d\begin{pmatrix} s_t \\ V_t \end{pmatrix} = \begin{bmatrix} r + \left(\lambda_{1R_t}\left(1-\rho_{R_t}^2\right) - \frac{1}{2}\right)V_t \\ \kappa_{R_t}\left(\gamma_{R_t} - V_t\right) \end{bmatrix} dt + \begin{bmatrix} \sqrt{\left(1-\rho_{R_t}^2\right)V_t} & \rho_{R_t}\sqrt{V_t} \\ 0 & \sigma_{R_t}V_t^{\beta_{R_t}} \end{bmatrix} d\begin{pmatrix} W_{1t}^P \\ W_{2t}^P \end{pmatrix},$$

where the regime determining variable R_t follows the aforementioned continuous time first order Markov chain with two states. In addition to those parameters from Section 3.1 two more parameters, β_L and β_H need to be estimated. Restriction, $1/2 \leq \beta \leq 1$ will be imposed for the uniqueness of option prices following Aït-Sahalia and Kimmel (2007). As in the regime-switching Heston model, both the time constant and time varying transition matrix of R_t will be investigated. Therefore, we consider four different specifications, CEV-R1, CEV-R2-1, CEV-R2-2, and CEV-R2TVTP.

3.3 GARCH Model with Regime Shifts

Another interesting model that will be examined is the GARCH model (Nelson (1990) and Meddahi (2001)). In this setting, the stock price and its variance follow

$$d\begin{pmatrix} S_t \\ V_t \end{pmatrix} = \begin{bmatrix} rS_t \\ \kappa'(\gamma' - V_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)V_t} S_t & \rho\sqrt{V_t}S_t \\ 0 & \sigma V_t \end{bmatrix} d\begin{pmatrix} W_{1t}^Q \\ W_{2t}^Q \end{pmatrix} \tag{8}$$

under the risk-neutral measure. As discussed above, this model is nested by the CEV model. To ensure positivity of the variance V_t, $\kappa'\gamma' \geq 0$ is required. As before, expressing (8) in (s_t, V_t)

$$d\begin{pmatrix} s_t \\ V_t \end{pmatrix} = \begin{bmatrix} r - \frac{1}{2}V_t \\ \kappa'(\gamma' - V_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)V_t} & \rho\sqrt{V_t} \\ 0 & \sigma V_t \end{bmatrix} d\begin{pmatrix} W_{1t}^Q \\ W_{2t}^Q \end{pmatrix}.$$

With the same market price specification, $\left[\lambda_1\sqrt{(1-\rho^2)Y_t}, 0\right]^T$ as above, under the measure P,

$$d\begin{pmatrix} s_t \\ V_t \end{pmatrix} = \begin{bmatrix} r + bV_t \\ \kappa(\gamma - V_t) \end{bmatrix} dt + \begin{bmatrix} \sqrt{(1-\rho^2)V_t} & \rho\sqrt{V_t} \\ 0 & \sigma V_t \end{bmatrix} d\begin{pmatrix} W_{1t}^P \\ W_{2t}^P \end{pmatrix}, \tag{9}$$

where $b = \lambda_1(1-\rho^2) - \frac{1}{2}$, $\kappa = \kappa'$, and $\gamma = \gamma'$. Combining regime-switching with (9) yields

$$d\begin{pmatrix} s_t \\ V_t \end{pmatrix} = \begin{bmatrix} r + \left(\lambda_{1R_t}\left(1-\rho_{R_t}^2\right) - \frac{1}{2}\right)V_t \\ \kappa_{R_t}\left(\gamma_{R_t} - V_t\right) \end{bmatrix} dt + \begin{bmatrix} \sqrt{\left(1-\rho_{R_t}^2\right)V_t} & \rho_{R_t}\sqrt{V_t} \\ 0 & \sigma_{R_t}V_t \end{bmatrix} d\begin{pmatrix} W_{1t}^P \\ W_{2t}^P \end{pmatrix}.$$

The dynamics of regime variable R_t and the associated transition matrix are illustrated in Section 3.1. We have exactly the same set of parameters as in the Heston model. Similarly to other two cases, four different models are said to be GARCH-R1, GARCH-R2-1, GARCH-R2-2, and GARCH-R2TVTP.

4 Estimation Method

Maximum likelihood estimation method is employed to estimate parameters. The likelihood of the data is computed by the Hamilton algorithm (Hamilton (1989)). In doing so, it is necessary to know the transition probability density function (TPDF) of our stochastic volatility models. However, the true TPDF of most diffusion processes including our models are unknown. Utilizing Aït-Sahalia (2008) and Choi (2015b) we can obtain a TPDF expansion of a multivariate time-homogeneous diffusion in a closed-form. Let us first see how we can get the approximate TPDF.

4.1 Approximate Transition Probability Density Function

An m-dimensional diffusion process,

$$dX_t = \mu\left(X_t;\theta\right)dt + \sigma\left(X_t;\theta\right)dW_t$$

is said to be reducible if there exists a transformation by which we can transform a diffusion into a unit diffusion, where the volatility is the identity matrix. If a multivariate diffusion is reducible, the log-likelihood function can be found by using the reducible method that includes Hermite expansion and Kolmogorov method. However, if it is not reducible (irreducible), the reducible method is not applicable but the irreducible method can be adopted. The irreducible method is general enough to be applied to an arbitrary multivariate diffusion models including reducible diffusions.

A necessary and sufficient condition for the reducibility of an m-dimensional diffusion process X_t is

$$\frac{\partial \sigma_{ij}^{-1}\left(x;\theta\right)}{\partial x_k} = \frac{\partial \sigma_{ik}^{-1}\left(x;\theta\right)}{\partial x_j},$$

for each x in the domain of X_t and triplet $\{i,j,k\} \subset \{1,2,\cdots,m\}$ such that $k > j$. Here $\sigma_{ij}^{-1}\left(x;\theta\right)$ is the (i,j) element of the inverse matrix of the volatility $\sigma\left(x;\theta\right)$. Checking this condition, all of the stochastic volatility models considered in this paper turn out to be irreducible.

The key idea of finding the approximate log-TPDF of an irreducible diffusions is to postulate the form of the log-likelihood expansion of X_t as the one obtained from the reducible case:

$$\begin{aligned} & l_X^{(K)}(\Delta, x|x_0;\theta) \\ = \; & -\frac{m}{2}\ln(2\pi\Delta) - D_\upsilon(x;\theta) + \frac{C_X^{(-1)}(\Delta, x|x_0;\theta)}{\Delta} + \sum_{k=0}^{K} C_X^{(k)}(\Delta, x|x_0;\theta)\frac{\Delta^k}{k!}. \end{aligned} \tag{10}$$

And use the fact that the true log-TPDF satisfies the Kolmogorov forward and backward equations. The Kolmogorov forward equation is

$$\begin{aligned} \frac{\partial l_X(\Delta, x|x_0;\theta)}{\partial \Delta} = \; & -\sum_{i=1}^{m}\frac{\partial \mu_i(x;\theta)}{\partial x_i} + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\partial^2 \upsilon_{ij}(x;\theta)}{\partial x_i \partial x_j} - \sum_{i=1}^{m}\mu_i(x;\theta)\frac{\partial l_X(\Delta, x|x_0;\theta)}{\partial x_i} \\ & + \sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\partial \upsilon_{ij}(x;\theta)}{\partial x_i}\frac{\partial l_X(\Delta, x|x_0;\theta)}{\partial x_j} + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\upsilon_{ij}(x;\theta)\frac{\partial^2 l_X(\Delta, x|x_0;\theta)}{\partial x_i \partial x_j} \\ & + \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\partial l_X(\Delta, x|x_0;\theta)}{\partial x_i}\upsilon_{ij}(x;\theta)\frac{\partial l_X(\Delta, x|x_0;\theta)}{\partial x_j}. \end{aligned} \tag{11}$$

Replacing $l_X(\Delta, x|x_0;\theta)$ with $l_X^{(K)}(\Delta, x|x_0;\theta)$ in (11) and matching the terms with the same orders of Δ yield the partial differential equations (PDEs) of the coefficients $C_X^{(k)}(\Delta, x|x_0;\theta)$, $k \geq -1$. They are

$$-2C_X^{(-1)}(\Delta, x|x_0;\theta) = \sum_{i=1}^{m}\sum_{j=1}^{m}\upsilon_{ij}(x;\theta)\frac{\partial C_X^{(-1)}(\Delta, x|x_0;\theta)}{\partial x_i}\frac{\partial C_X^{(-1)}(\Delta, x|x_0;\theta)}{\partial x_j}$$

$$-\sum_{i=1}^{m}\sum_{j=1}^{m}\upsilon_{ij}(x;\theta)\frac{\partial C_X^{(-1)}(\Delta, x|x_0;\theta)}{\partial x_i}\frac{\partial C_X^{(0)}(\Delta, x|x_0;\theta)}{\partial x_j} = G_X^{(0)}(\Delta, x|x_0;\theta)$$

and for all $k \geq 1$

$$C_X^{(k)}(\Delta, x|x_0;\theta) - \frac{1}{k}\sum_{i=1}^{m}\sum_{j=1}^{m}\upsilon_{ij}(x;\theta)\frac{\partial C_X^{(-1)}(\Delta, x \mid x_0;\theta)}{\partial x_i}\frac{\partial C_X^{(k)}(\Delta, x \mid x_0;\theta)}{\partial x_j} = G_X^{(k)}(\Delta, x \mid x_0;\theta),$$

where

$$\begin{aligned} G_X^{(0)}(\Delta,x\mid x_0;\theta) &= \frac{m}{2}-\sum_{i=1}^{m}\mu_i(x;\theta)\frac{\partial C_X^{(-1)}(\Delta,x\mid x_0;\theta)}{\partial x_i}+\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\partial \upsilon_{ij}(x;\theta)}{\partial x_i}\frac{\partial C_X^{(-1)}(\Delta,x\mid x_0;\theta)}{\partial x_j} \\ &-\sum_{i=1}^{m}\sum_{j=1}^{m}\upsilon_{ij}(x;\theta)\frac{\partial C_X^{(-1)}(\Delta,x\mid x_0;\theta)}{\partial x_i}\frac{\partial D_\upsilon(x;\theta)}{\partial x_j}+\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\upsilon_{ij}(x;\theta)\frac{\partial^2 C_X^{(-1)}(\Delta,x\mid x_0;\theta)}{\partial x_i\partial x_j}, \end{aligned}$$

$$\begin{aligned} G_X^{(1)}(\Delta,x\mid x_0;\theta) &= -\sum_{i=1}^{m}\mu_i(x;\theta)\frac{\partial C_X^{(0)}(\Delta,x\mid x_0;\theta)}{\partial x_i}+\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\partial \upsilon_{ij}(x;\theta)}{\partial x_i}\frac{\partial C_X^{(0)}(\Delta,x\mid x_0;\theta)}{\partial x_j} \\ &-\sum_{i=1}^{m}\sum_{j=1}^{m}\upsilon_{ij}(x;\theta)\frac{\partial C_X^{(0)}(\Delta,x\mid x_0;\theta)}{\partial x_i}\frac{\partial D_\upsilon(x;\theta)}{\partial x_j}+\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\upsilon_{ij}(x;\theta)\frac{\partial^2 C_X^{(0)}(\Delta,x\mid x_0;\theta)}{\partial x_i\partial x_j} \\ &+\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\upsilon_{ij}(x;\theta)\frac{\partial C_X^{(0)}(\Delta,x\mid x_0;\theta)}{\partial x_i}\frac{\partial C_X^{(0)}(\Delta,x\mid x_0;\theta)}{\partial x_j} \\ &-\sum_{i=1}^{m}\frac{\partial \mu_i(x;\theta)}{\partial x_i}+\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\partial^2 \upsilon_{ij}(x;\theta)}{\partial x_i\partial x_j}+\sum_{i=1}^{m}\mu_i(x;\theta)\frac{\partial D_\upsilon(x;\theta)}{\partial x_i}-\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\partial \upsilon_{ij}(x;\theta)}{\partial x_i}\frac{\partial D_\upsilon(x;\theta)}{\partial x_j} \\ &-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\upsilon_{ij}(x;\theta)\left[\frac{\partial^2 D_\upsilon(x;\theta)}{\partial x_i\partial x_j}-\frac{\partial D_\upsilon(x;\theta)}{\partial x_i}\frac{\partial D_\upsilon(x;\theta)}{\partial x_j}\right]. \end{aligned}$$

and when $k \geq 2$

$$\begin{aligned} G_X^{(k)}(\Delta,x\mid x_0;\theta) &= -\sum_{i=1}^{m}\mu_i(x;\theta)\frac{\partial C_X^{(k-1)}(\Delta,x\mid x_0;\theta)}{\partial x_i}+\sum_{i=1}^{m}\sum_{j=1}^{m}\frac{\partial \upsilon_{ij}(x;\theta)}{\partial x_i}\frac{\partial C_X^{(k-1)}(\Delta,x\mid x_0;\theta)}{\partial x_j} \\ &-\sum_{i=1}^{m}\sum_{j=1}^{m}\upsilon_{ij}(x;\theta)\frac{\partial C_X^{(k-1)}(\Delta,x\mid x_0;\theta)}{\partial x_i}\frac{\partial D_\upsilon(x;\theta)}{\partial x_j}+\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\upsilon_{ij}(x;\theta)\frac{\partial^2 C_X^{(k-1)}(\Delta,x\mid x_0;\theta)}{\partial x_i\partial x_j} \\ &+\sum_{i=1}^{m}\sum_{j=1}^{m}\upsilon_{ij}(x;\theta)\frac{\partial C_X^{(0)}(\Delta,x\mid x_0;\theta)}{\partial x_i}\frac{\partial C_X^{(k-1)}(\Delta,x\mid x_0;\theta)}{\partial x_j} \\ &+\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\upsilon_{ij}(x;\theta)\left[\sum_{h=1}^{k-2}\binom{k-1}{h}\frac{\partial C_X^{(h)}(\Delta,x\mid x_0;\theta)}{\partial x_i}\frac{\partial C_X^{(k-1-h)}(\Delta,x\mid x_0;\theta)}{\partial x_j}\right]. \end{aligned}$$

Here $\upsilon(x;\theta)=\sigma(x;\theta)\,\sigma(x;\theta)^T$, where $\sigma(x;\theta)^T$ is the transpose of $\sigma(x;\theta)$.

If X_t is reducible, the PDEs of $C_X^{(k)}(\Delta,x|x_0;\theta)$ are solvable (see Choi (2015a)). Or we can convert X_t into a unit diffusion process and use Choi (2015b) to acquire the approximate TPDF. But we are no longer able to get analytic solutions of the PDEs when it is irreducible. In this case we can Taylor-expand each coefficient, $C_X^{(k)}(\Delta,x|x_0;\theta)$ around x_0 up to j_k-th order which is determined to achieve the same approximation error of $O_p\left(\Delta^{K+1}\right)$ for each coefficient. Setting $j_k = 2(K-k+1)$ would provide an approximation error $O_p\left(\Delta^{K+1}\right)$. For example, if $K = 2$, $j_{-1} = 8$, $j_0 = 6$, $j_1 = 4$, $j_2 = 2$, $j_3 = 0$. The coefficients have to be found recursively from a low order term to the next higher order term one by one in $C_X^{(j_{-1},-1)}(\Delta, x|x_0;\theta)$. Given $C_X^{(j_{-1},-1)}(\Delta, x|x_0;\theta)$, the next term $C_X^{(j_0,0)}(\Delta, x|x_0;\theta)$ needs to be computed from low order to high order up to j_0. Given $C_X^{(j_{-1},-1)}(\Delta, x|x_0;\theta)$ and $C_X^{(j_0,0)}(\Delta, x|x_0;\theta)$, the next term $C_X^{(j_1,1)}(\Delta, x|x_0;\theta)$ can be determined, and so on. Let j_k-th order Taylor expansion of $C_X^{(k)}(\Delta, x|x_0;\theta)$ be $C_X^{(k,j_k)}(\Delta, x|x_0;\theta)$.[4]

Choi (2015b) established that

$$p_X^{(K)}(\Delta, x|x_0;\theta) = \Delta^{-m/2}\exp\left[\frac{C_X^{(-1)}(\Delta, x|x_0;\theta)}{\Delta}\right](2\pi)^{-m/2}\left|\det\left[\upsilon(x;;\theta)\right]^{-1/2}\right| \times\exp\left[C_X^{(0)}(\Delta, x|x_0;\theta)\right]\sum_{k=0}^{K}c_X^{(k)}(\Delta, x|x_0;\theta)\frac{\Delta^k}{k!}. \quad (12)$$

The coefficients $c_X^{(k)}(t, x|t_0, x_0)$, $k \geq 0$ are

$$c_X^{(k)}(\Delta, x|x_0;\theta) = \sum\frac{k!}{s_1!s_2!\cdots s_k!}\left[\frac{C_X^{(1)}(\Delta, x|x_0;\theta)}{1!}\right]^{s_1}\left[\frac{C_X^{(2)}(\Delta, x|x_0;\theta)}{2!}\right]^{s_2}\cdots\left[\frac{C_X^{(k)}(\Delta, x|x_0;\theta)}{k!}\right]^{s_k},$$

where the summation is over all nonnegative integers s_j, $j = 1, \cdots, k$ satisfying $k = s_1 + 2s_2 + \cdots + ks_k$ and $C_X^{(l)}(\Delta, x|x_0;\theta)$, $l \geq -1$ are the coefficients of log-likelihood expansion (10). Replacing $C_X^{(k)}(\Delta, x|x_0;\theta)$ with $C_X^{(k,j_k)}(\Delta, x|x_0;\theta)$ in (12) results in the approximate TPDF of order $O_p\left(\Delta^{K+1}\right)$:

$$\widetilde{p}_X^{(K)}(\Delta, x|x_0;\theta) = \Delta^{-m/2}\exp\left[\frac{C_X^{(j_{-1},-1)}(\Delta, x|x_0;\theta)}{\Delta}\right](2\pi)^{-m/2}\left|\det\left[\upsilon(x;;\theta)\right]^{-1/2}\right| \times\exp\left[C_X^{(j_0,0)}(\Delta, x|x_0;\theta)\right]\sum_{k=0}^{K}c_X^{(j_k,k)}(\Delta, x|x_0;\theta)\frac{\Delta^k}{k!}.$$

4.2 Hamilton Algorithm

We observe the process $X_t = (s_t, V_t)$ at discrete time points $t = i\Delta$ where $i = 0, 1, 2, \cdots, n$. Using the Bayes' rule, the joint probability density function of the data $\left(x_{n\Delta}, x_{(n-1)\Delta}, \cdots, x_\Delta, x_0\right)$

[4] The Kolmogorov backward equation for the log-TPDF of X_t is $\frac{\partial l_X(\Delta,x|x_0;\theta)}{\partial\Delta} = \sum_{i=1}^m \mu_i(x_0;\theta)\frac{\partial l_X(\Delta,x|x_0;\theta)}{\partial x_{0i}} + \frac{1}{2}\sum_{i=1}^m\sum_{j=1}^m \upsilon_{ij}(x_0;\theta)\frac{\partial^2 l_X(\Delta,x|x_0;\theta)}{\partial x_{0i}\partial x_{0j}} + \frac{1}{2}\sum_{i=1}^m\sum_{j=1}^m\frac{\partial l_X(\Delta,x|x_0;\theta)}{\partial x_{0i}}\upsilon_{ij}(x_0;\theta)\frac{\partial l_X(\Delta,x|x_0;\theta)}{\partial x_{0j}}$. The PDEs of the coefficients in x_0 and Δ are obtainable if we use the backward equation. Making use of those PDEs, Taylor expansions of the coefficients can be found as well. Either produces the same results.

can be written as

$$\begin{aligned}
& p\left(x_{n\Delta}, x_{(n-1)\Delta}, \cdots, x_{\Delta}, x_0; \theta\right) \\
= \;& p\left(x_{n\Delta} | x_{(n-1)\Delta}, \cdots, x_{\Delta}, x_0; \theta\right) \times p\left(x_{(n-1)\Delta} | x_{(n-2)\Delta} \cdots, x_{\Delta}, x_0; \theta\right) \times \\
& \cdots \times p\left(x_{2\Delta} | x_{\Delta}, x_0; \theta\right) \times p\left(x_{\Delta} | x_0; \theta\right) \times p\left(x_0; \theta\right).
\end{aligned}$$

Ignoring the initial observation, the log-likelihood function is represented as

$$l_n(\theta) \equiv \sum_{i=1}^{T} \ln\left[p\left(x_{i\Delta} | I_{(i-1)\Delta}; \theta\right)\right], \tag{13}$$

where $I_t = \{X_s | s \leq t\}$ is the information set which consists of data through time t.

Conditional likelihood functions $p\left(x_{i\Delta} | I_{(i-1)\Delta}; \theta\right)$ for all $i = 0, 1, 2, \cdots, n$ can be calculated using the algorithm developed by Hamilton (1989). As a matter of convenience to explain the algorithm, let's define a new state variable R_t^* as follows (Hamilton (1994)):

$$\begin{aligned}
R_t^* &= 1 \text{ if } R_{t-\Delta} = L \text{ and } R_t = L \\
R_t^* &= 2 \text{ if } R_{t-\Delta} = L \text{ and } R_t = H \\
R_t^* &= 3 \text{ if } R_{t-\Delta} = H \text{ and } R_t = L \\
R_t^* &= 4 \text{ if } R_{t-\Delta} = H \text{ and } R_t = H
\end{aligned}$$

Then R_t^* follows a four-state continuous time Markov chain with the following transition matrix

$$P^* = \begin{pmatrix} p_{LL} & 0 & p_{LL} & 0 \\ p_{LH} & 0 & p_{LH} & 0 \\ 0 & p_{HL} & 0 & p_{HL} \\ 0 & p_{HH} & 0 & p_{HH} \end{pmatrix}.$$

At step $t + \Delta$, the input of the algorithm are $\hat{\xi}_{t+\Delta|t}$ and $\eta_{t+\Delta}$,

$$\hat{\xi}_{t+\Delta|t} = \begin{pmatrix} P\left(R_{t+\Delta}^* = 1 | I_t; \theta\right) \\ P\left(R_{t+\Delta}^* = 2 | I_t; \theta\right) \\ P\left(R_{t+\Delta}^* = 3 | I_t; \theta\right) \\ P\left(R_{t+\Delta}^* = 4 | I_t; \theta\right) \end{pmatrix} \text{ and } \eta_{t+\Delta} = \begin{pmatrix} p\left(x_{t+\Delta} | R_{t+\Delta}^* = 1, I_t; \theta\right) \\ p\left(x_{t+\Delta} | R_{t+\Delta}^* = 2, I_t; \theta\right) \\ p\left(x_{t+\Delta} | R_{t+\Delta}^* = 3, I_t; \theta\right) \\ p\left(x_{t+\Delta} | R_{t+\Delta}^* = 4, I_t; \theta\right) \end{pmatrix}$$

Elements of $\hat{\xi}_{t+\Delta|t}$ are the inferences about the value of $R^*_{t+\Delta}$ based on I_t and knowledge of the population parameter vector θ. The vector $\eta_{t+\Delta}$ contains conditional density functions of $X_{t+\Delta}$ given $R^*_{t+\Delta}, I_t$ and θ. For example, if $R^*_{t+\Delta} = 1$ then $p\left(x_{t+\Delta}|R^*_{t+\Delta}, I_t; \theta\right) = p\left(x_{t+\Delta}|x_t, R_t = L; \theta\right)$ by exploiting the Markov property of the diffusion process, X_t and the timing assumption that the parameters of conditional density function $X_{t+\Delta}$ depend on i_t. Notice that the transition density function $p\left(x_{t+\Delta}|x_t, R_t = L; \theta\right) = p\left(x_{t+\Delta}|x_t, ; \theta_L\right)$ is unknown in general as is the case for our models, but we can use the approximate log-TPDF explained above.

The joint likelihood of $X_{t+\Delta}$ and $R^*_{t+\Delta}$ can be computed by the element by element multiplication of $\hat{\xi}_{t+\Delta|t}$ and $\eta_{t+\Delta}$, that is

$$\hat{\xi}_{t+\Delta|t} \odot \eta_{t+\Delta} = \begin{pmatrix} P\left(R^*_{t+\Delta} = 1|I_t; \theta\right) p\left(x_{t+\Delta}|R^*_{t+\Delta} = 1, I_t; \theta\right) \\ P\left(R^*_{t+\Delta} = 2|I_t; \theta\right) p\left(x_{t+\Delta}|R^*_{t+\Delta} = 2, I_t; \theta\right) \\ P\left(R^*_{t+\Delta} = 3|I_t; \theta\right) p\left(x_{t+\Delta}|R^*_{t+\Delta} = 3, I_t; \theta\right) \\ P\left(R^*_{t+\Delta} = 4|I_t; \theta\right) p\left(x_{t+\Delta}|R^*_{t+\Delta} = 4, I_t; \theta\right) \end{pmatrix} = \begin{pmatrix} p\left(x_{t+\Delta}, R^*_{t+\Delta} = 1|I_t; \theta\right) \\ p\left(x_{t+\Delta}, R^*_{t+\Delta} = 2|I_t; \theta\right) \\ p\left(x_{t+\Delta}, R^*_{t+\Delta} = 3|I_t; \theta\right) \\ p\left(x_{t+\Delta}, R^*_{t+\Delta} = 4|I_t; \theta\right) \end{pmatrix}$$

Summing $\hat{\xi}_{t+\Delta|t} \odot \eta_{t+\Delta}$ over the variable $R^*_{t+\Delta}$, we get the value of the conditional density function given I_t,

$$p\left(x_{t+\Delta}|I_t; \theta\right) = \sum_{j=1}^{4} p\left(x_{t+\Delta}, R^*_{t+\Delta} = j|I_t; \theta\right). \tag{14}$$

The optimal inference about $R^*_{t+\Delta}$ denoted as $\hat{\xi}_{t+\Delta|t+\Delta}$ is calculated by

$$\hat{\xi}_{t+\Delta|t+\Delta} = \frac{\hat{\xi}_{t+\Delta|t} \odot \eta_{t+\Delta}}{p\left(x_{t+\Delta}|I_t; \theta\right)} = \begin{pmatrix} P\left(R^*_{t+\Delta} = 1|I_{t+\Delta}; \theta\right) \\ P\left(R^*_{t+\Delta} = 2|I_{t+\Delta}; \theta\right) \\ P\left(R^*_{t+\Delta} = 3|I_{t+\Delta}; \theta\right) \\ P\left(R^*_{t+\Delta} = 4|I_{t+\Delta}; \theta\right) \end{pmatrix} \tag{15}$$

The input of the next step, $\hat{\xi}_{t+2\Delta|t+\Delta}$ is updated by premultiplying the transition matrix P^* by $\hat{\xi}_{t+\Delta|t+\Delta}$.

The initial state probabilities of R^*_t are $(\pi_1, \pi_2, \pi_3, \pi_4) = (p_{LL}p, p_{LH}p, p_{HL}(1-p), p_{HH}(1-p))$ when $P(R_1 = L) = p$, where an additional parameter p is introduced for the probability of the initial state $R_1 = L$. In this case $P(R_1 = H) = 1 - p$. Given these starting values, if we iterate the above procedure and collect the conditional density function (14) of each step, we can compute the log-likelihood (13) to conduct MLE. For the probability of the initial state, we also use stationay probability such that $P(R_1 = L) = \frac{1-p_{HH}}{2-p_{LL}-p_{HH}}$ instead of p for the R2-1 and R2-1 models.

After computing the conditional likelihood $p(x_{t+\Delta}|I_t;\theta)$ from (14), we can also attain the probability of each state of $R^*_{t+\Delta}$ as in (15). Then the filtered probability of $R_{t+\Delta}$ at each time point is easily obtained as

$$P(R_t = L|I_t;\theta) = P(R^*_t = 1|I_t;\theta) + P(R^*_t = 3|I_t;\theta)$$

and

$$\begin{aligned} P(R_t = H|I_t;\theta) &= 1 - P(R_t = L|I_t;\theta) \\ &= P(R^*_t = 2|I_t;\theta) + P(R^*_t = 4|I_t;\theta). \end{aligned}$$

The smoothed probabilitis of each regime, $P(R_t = L|I_T;\theta)$ and $P(R_t = H|I_T;\theta)$ that use all information available have also been calculated.

5 Estimation Results and Discussion

5.1 Integrated Volatility

When MLE is carried out to estimate the parameter vector, the S&P 500 Index is used for S_t and the integrated volatility proxy for V_t is obtained by using the VIX. Exploiting Hull and White (1987)'s idea, Aït-Sahalia and Kimmel (2007) developed a way to compute the integrated volatility proxy from the observed implied variance of a short-maturity at-the-money S&P 500 Index options (See also Jones (2003)). If the Q-measure drift of V_t is of the form $a + bV_t$ as is the case for the all of our models, then the integrated volatility proxy, IV_t, is given by

$$IV_t = \frac{b\cdot\tau\cdot V_t^{imp} + a\cdot\tau}{\exp(b\cdot\tau) - 1} - \frac{a}{b}, \tag{16}$$

where V_t^{imp} is the implied volatility of the short-maturity at-the-money option and τ is time to maturity of the option.

We first estimate parameters of the following univariate CEV model for $(VIX_t/100)^2$

$$dV_t = \kappa(\gamma - V_t)\,dt + \sigma V_t^{\beta} dW_t$$

to get $a = \kappa\gamma$ and $b = -\kappa$. Because the true log-likelihood function of the CEV model is unknown, the approximate log-likelihood function was obtained using the irreducible method from Aït-Sahalia (2008) to do MLE. Estimation results are

$$\begin{aligned} dV_t &= \underset{(0.726)}{2.61^{**}}\,(\underset{(0.010)}{0.054^{**}} - V_t)\,dt + \underset{(0.011)}{1.28^{**}} X_t^{\underset{(0.002)}{0.84^{**}}} dW_t \end{aligned}$$

with standard errors in parentheses.

We use estimates of a and b and $(VIX_t/100)^2$ for V_t^{imp} in equation (16) to compute the integrated volatility such that

$$IV_t = -0.0064 + 1.1182\,(VIX_t/100)^2$$

The logarithm of S&P 500 index and the integrated volatility IV_t are used for s_t and V_t, respectively to estimate parameters. Because the log-likelihood function is not known for any of models discussed above, the irreducible method has been applied to construct an approximate joint log-likelihood function of (s_t, V_t).

5.2 Estimation Results

In table 2, we summarize the ML estimation results for the four different Heston stochastic volatility models. The first column shows estimation outcomes for the single-regime model H-R1. The second and third columns report the estimation results for two regime-switching models with time invariant transition matrix. The difference between them is the approach we use to deal with the probability of the original state. We keep the unconditional probability at the value of 0.5 for the H-R2-1 model, but set an additional parameter to estimate for the model H-R2-2. The last column presents ML estimation outcomes for the regime-switching model with varying transition matrix and an additional parameter for the probability of the original state. The asterisk by the parameter estimate implies that,

Table 2: Maximum Likelihood Estimation Results of Heston Models

Model	Heston-R1	Heston-R2-1	Heston-R2-2	Heston-R2TVTP
Regime	Single Regime	Two Regimes		
		Time Invariant Transition Matrix		Time Varying Transition Matrix
$P(R_1 = L)$	-	$\frac{1-p_{HH}}{2-p_{LL}-p_{HH}}$		p
$P(R_{t+\Delta} = i \mid R_t = i)$ with $i = L$ and H	-	p_{ii}	p_{ii}	$\frac{\exp(c_i+d_i s_t+c_i V_t)}{1+\exp(c_i+d_i s_t+c_i V_t)}$
Log-likelihood	42231.87	43094.03	43094.03	43293.30
Parameters	Maximum Likelihood Estimates (Standard Error)			
ρ_L	$-0.80^{**}(0.002)$	$-0.600^{**}(0.010)$	$-0.60^{**}(0.010)$	$-0.62^{**}(0.002)$
ρ_H	-	$-0.799^{**}(0.002)$	$-0.80^{**}(0.004)$	$-0.79^{**}(0.004)$
κ_L	$6.82^{**}(1.79)$	$5.00^{**}(1.05)$	$5.00^{**}(1.04)$	$5.00^{**}(0.76)$
κ_H	-	$7.00^{**}(0.93)$	$7.00^{**}(1.15)$	$7.00^{**}(0.75)$
γ_L	$0.046^{**}(0.017)$	$0.018^{**}(0.002)$	$0.018^{**}(0.002)$	$0.019^{**}(0.002)$
γ_H	-	$0.042^{**}(0.005)$	$0.042^{**}(0.006)$	$0.040^{**}(0.005)$
σ_L	$0.73^{**}(0.002)$	$0.249^{**}(0.006)$	$0.25^{**}(0.006)$	$0.23^{**}(0.001)$
σ_H	-	$0.558^{**}(0.001)$	$0.56^{**}(0.004)$	$0.63^{**}(0.003)$
r_L	0.0002 (0.23)	0.0002 (0.090)	0.0002 (0.091)	0.0002 (0.015)
r_H	-	0.0002 (0.099)	0.0002 (0.14)	0.0002 (0.008)
λ_L	6.93 (9.91)	6.00 (7.72)	6.000 (7.74)	6.00 (4.90)
λ_H		6.00 (6.31)	6.000 (7.87)	6.00 (4.15)
p_{LL}	-	$0.099^{**}(0.011)$	$0.099^{**}(0.012)$	-
p_{HH}	-	$0.90^{**}(0.011)$	$0.90^{**}(0.011)$	-
c_L	-	-	-	-0.20 (2.92)
c_H	-	-	-	0.170 (0.51)
d_L	-	-	-	$-1.46^{**}(0.45)$
d_H	-	-	-	$1.24^{**}(0.083)$
e_L	-	-	-	-0.014 (9.55)
e_H	-	-	-	0.010 (0.48)
p	-	-	0.50 (1.01)	0.50 (0.66)
AIC	-14.42	-14.71	-14.71	-14.78
BIC	-14.41	-14.70	-14.70	-14.76
RCM	-	27.03	27.03	6.39

Note: Maximum likelihood estimates with standard errors in parentheses for the period 01/02/1990-06/12/2012 for the four different models with the general diffusion specification are presented in this table. The standard errors are calculated using sample mean of the outer product of the score functions evaluated at the ML estimates. According to the number of regimes, the starting probability and the transition probability matrix specifications, the four models are: H-R1 (single regime),H-R2-1 (two regimes, time invariant transition matrix,unconditional probability for the probability of initial state), H-R2-2 (two regimes, time invariant transition matrix, additioanl parameter for the probability of initial state), H-R2TVTP (two regimes,time varying transition matrix with logistic function,additional parameter for the probability of initial state).

at the 5% significance level, it is different from zero.

In the single regime model H-R1, besides the market price of risk λ and the rate of return rd, the rest estimates are statistically significant. The long run average volatility $\gamma^{1/2}$ is estimated to be approximately 21.45% per year with a rate of mean reversion coefficient κ of 6.817. The correlation coefficient ρ between the innovations to the stock price and stochastic volatility is strongly negative, -0.799. The variance of volatility σ is close to 0.726.

Comparing three different regime-switching models, we find that the rate of mean reversion of Y_t is higher in Regime H, $\kappa_H > \kappa_L$. The long-run average value of Y_t is bigger in Regime H, $\gamma_H > \gamma_L$. The correlation coefficient ρ between the innovations to stock price and stochastic volatility is stronger in Regime H, $\rho_H > \rho_L$. The variance of Y_t is more volatile in Regime H, $\sigma_H > \sigma_L$. rd_L, rd_H, λ_L and λ_H are statistically insignificant. Regime H is very persistent $p_{HH} \approx 0.900$ but Regime L is not so persistent $p_{LL} \approx 0.099$. In the model of H-R2TVTP, the variable s_t appears to be much more important in explaining the time varying transition probabilities of i_t. Because $d_L < 0$ $(d_H > 0)$, p_{LL} (p_{HH}) decreases (increases) as s_t increases. Both c and e are statistically insignificant.

Analyzing the log-likelihood values from the four different models, it illustrates that the single regime model H-R1 reports the smallest value 42231.867 while the model H-R2TVTP owns the biggest one 43293.294. The models of H-R2-1 and H-R2-2 display the same log-likelihood value with very similar ML estimates. There seems to be not much difference between using unconditional probability and employing a new parameter for the original state. However, there is strong evidence of existence of high and low volatility regimes for the time varying transition probability of the regime variable.

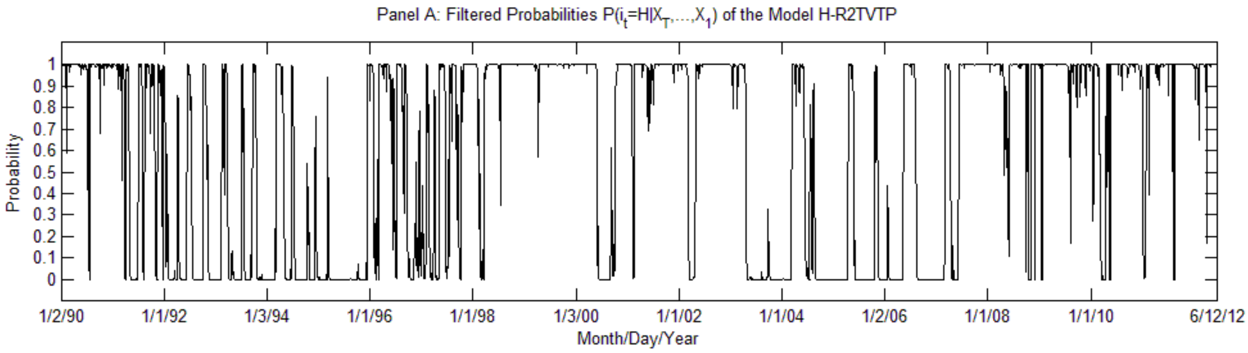
Since the parameters related to the second regime of the process are unidentified under the null hypothesis of single regime, traditional test statistics cannot be used to test whether there is one regime or not. Take the model of H-R2-1 for example, given $P(s_1 = L) = 1$, if the null hypothesis is $p_{LL} = 1$, then λ_H, κ_H, γ_H, ρ_H, σ_H, rd_H, and p_{HH} cannot be identified under the null of no regime switching. In this case, standard asymptotic distribution theory cannot be applied so that standard likelihood ratio tests and Wald tests cannot be conducted. Although some literatures (Davis (1987), Hansen (1992),Hansen (1996) and Cho and White (2007)) try to address this problem, testing for multiple regimes is still particularly challenging. Therefore, we resort to calculate Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) to compare different specifications. Although the difference is very narrow, the H-R2TVTP model reports the smallest AIC and BIC values. Another metric we employed to compare the performance of different regime-switching models is called Regime Classification Measure (RCM). It is first proposed by Ang and Bekaert (2002) and then applied by Choi (2009) for R regime case as

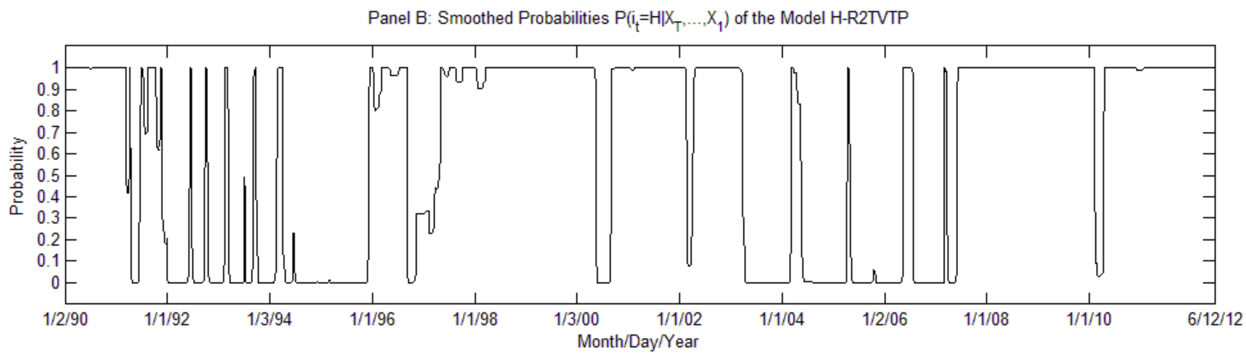
$$RCM\left(R\right) = 100R^{R}\frac{1}{n}\sum_{t=1}^{n}\left(\prod_{i=1}^{R}p_{i,t}\right),$$

where $p_{i,t} = P\left(s_t = i|I_T\right)$, $R = 2$, $100R^R$ is used to normalize the statistic to be between 0 and 100. A model that does a good (bad) job of distinguishing between regimes will make an inference about s_t being in a regime close to 1 or 0 $(1/R)$. Therefore, according to the inferred probabilities of staying in a particular regime, the closer the RCM value is to 0, the better the regime classification of a model is. It can be seen from table 3, the H-R2-1 and H-R2-2 models display very similar RCM, but there is a great improvement in the model H-R2TVTP. In the term of RCM values, it also shows strong evidence for the existence of varying transition matrix.

In order to analyze the probability of being in state H, we draw the time series of filtered probabilities and smoothed probabilities over the sample for the model H-R2TVTP.

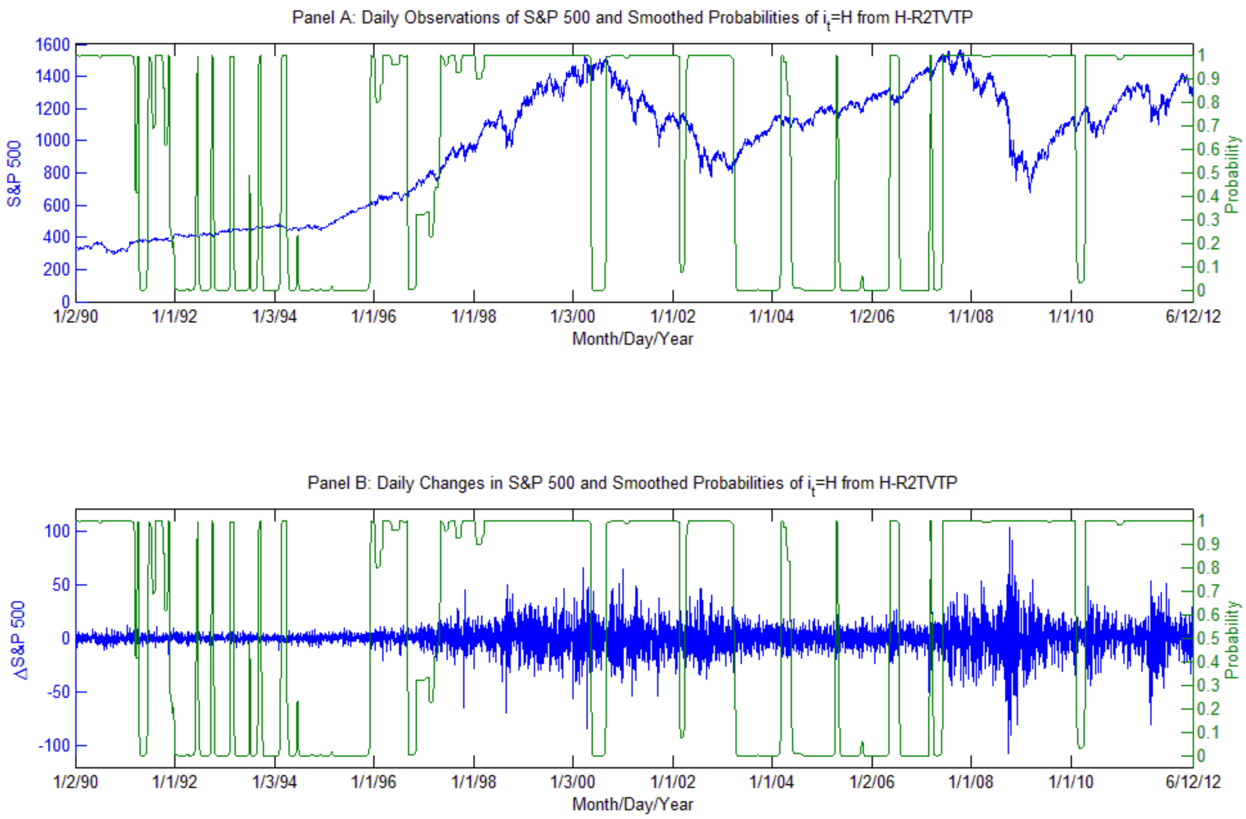
Figure 2: Regime-Switching Probabilities of the Model H-R2TVTP



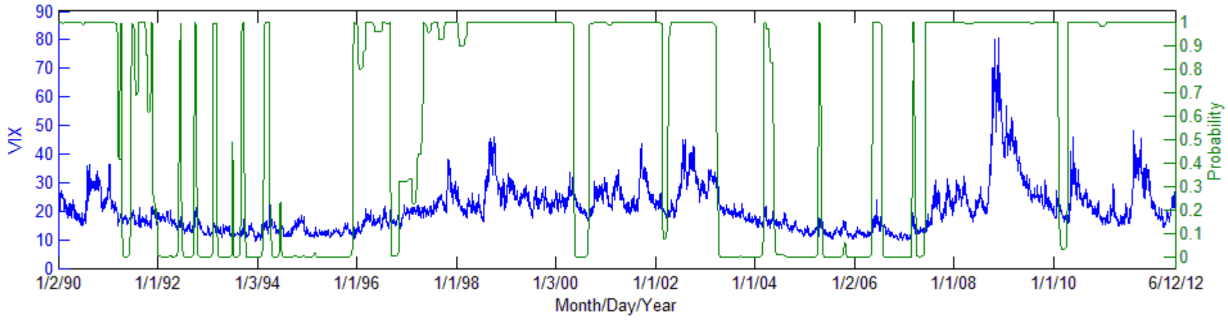


To make it more clear, we plot the time series of smoothed probabilities with S&P 500, VIX and their first difference separately.

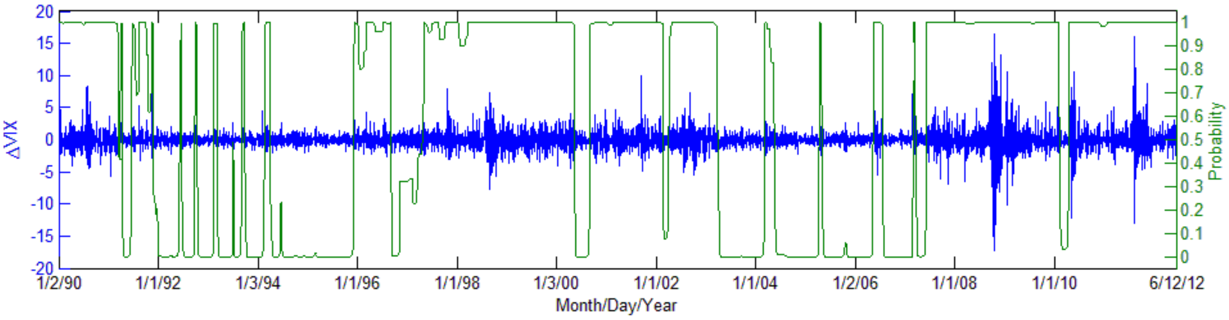
Figure 3: Regime-Switching Probabilities of the Model H-R2TVTP with S&P 500, VIX and Their First Difference



Panel C: Daily Observations of VIX and Smoothed Probabilities of i_t=H from H-R2TVTP



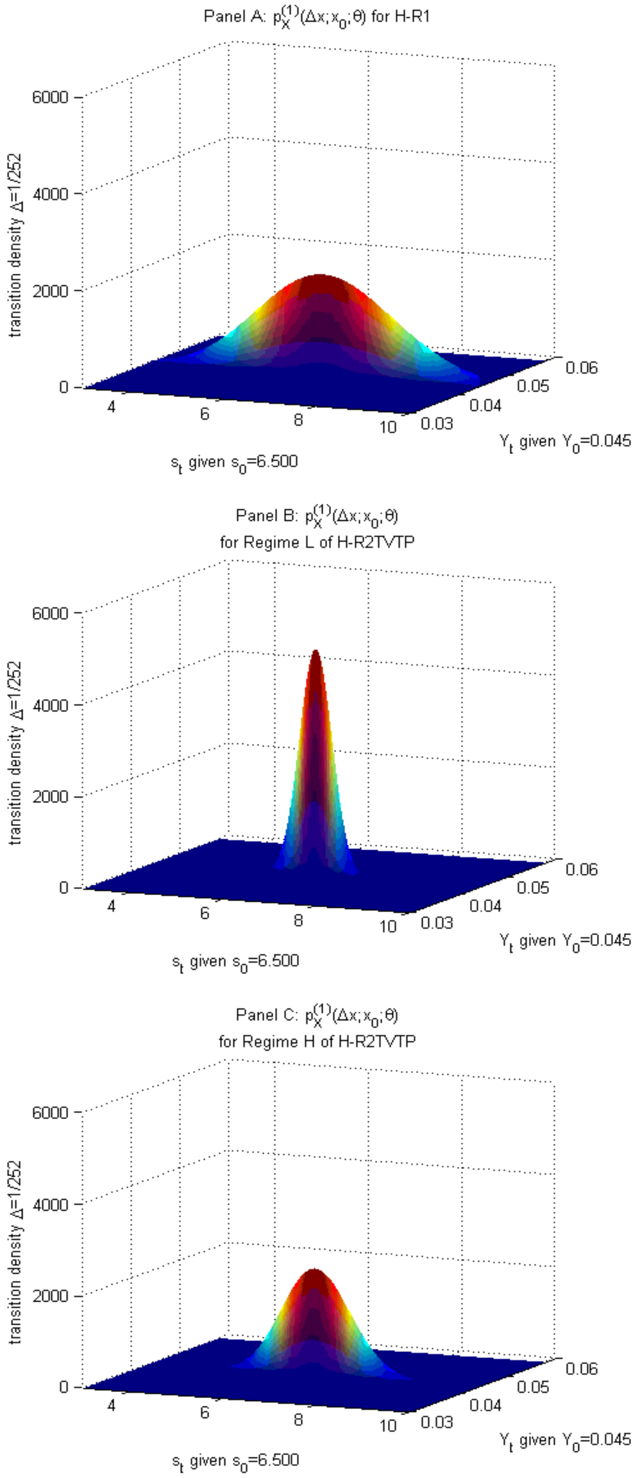
Panel D: Daily Changes in VIX and Smoothed Probabilities of i_t=H from H-R2TVTP



These illustrations display that the H-R2TVTP model identifies most high volatility periods, which can be connected with some significant events in U.S. history and financial history since 1990. The beginning high volatility regime corresponds to the Gulf War, which began from 2 August 1990 and lasted to 28 February 1991. The next state of H reflects the Black Wednesday in 1992. The high volatility regime around 1994 matches the 1994 Northridge earthquake in the Los Angeles area, caused an estimated $20 billion in damage, making it one of the costliest natural disasters in U.S. history. The following H state throughout 1996 is linked to the severe budget crisis, which forced the federal government to shutdown for several weeks at the end of 1995. The 1998-2000 high volatility regime is related to the Asian financial crisis, started in Thailand in July 1997 and then triggered the 1998 Russian financial crisis. It is also coincident with the 1998 collapse of long-term capital management, leading to an agreement among 14 financial institutions for a $3.6 billion recapitalization under the supervision of the Federal Reserve. Furthermore, the dot-com bubble covered roughly the period of 1995 to 2000, and finally burst because of an inflation report in April of 2000. The next three successive years of H state is caused by the September 11, 2001 terrorist attacks and the October 2001 invasion in Afghanistan. After that, the Iraq War was launched on March 19, 2003. The 2004 high volatility regime is linked to the 2004 Atlantic hurricane season, which impacted Florida, Charley, Frances, Ivan and Jeanne, and produced over $50 billion in damage. What's even worse, in August and September 2005, two powerful hurricanes, Hurricane Katrina and Hurricane Rita hit the Gulf Coast region. The last long H state is associated with the result of a succession of financial events. The Chinese Correction plunge of February 27, 2007 caused the Dow Jones Industrial Average dropped by 3.29%, which is the biggest one-day slide since the September 11. The subprime mortgage crisis, which has its roots in the closing years of the 20th century, became apparent between 2007 and 2008, and then rapidly evolved into a global financial crisis. On September 15, 2008 Lehman Brothers filed for bankruptcy protection. The European sovereign-debt crisis started late 2009, followed by the 2010 Flash Crash and the August 2011 stock markets fall across the world, furthering severe volatility of stock market indexes until now.

Using the ML estimates given in table 2, we plot the approximate conditional transition density functions of the stochastic differential equations given $x_0 = [s_0, Y_0]' = [6.500, 0.045]'$ for the models of H-R1 and H-R2TVTP, respectively. Since the corresponding graphs of the model H-R2-1 and H-R2-2 are very similar to those of the H-R2TVTP model, we omit them and compare only the model H-R1 and H-R2TVTP. At around the given asset price $s_0 = 6.500$ and its volatility $Y_0 = 0.045$, the conditional density function of regime L for the model H-R2TVTP is topped out with the peak of 4925.969 while that of regime H is flattened out with the top of 2316.343. Compared with those of the H-R2TVTP model, the conditional density function of the H-R1 modal is centered on a height of 2071.961. We also calculate the 95% confidence interval by delta method for each case and find that the 3D 95% confidence interval surfs form an upper layer and a lower layer just around the density functions.

Figure 4: Conditional Transition Density Functions for the Model H-R1 and H-R2TVTP



To make it more clear, we split the 3D shapes into two pieces and take the cross sections at $Y_0 = 0.045$. Then we get the following two plots of conditional transition density functions and their 95% confidence bands for the model H-R1 and H-R2TVTP separately. The different approximate conditional transition density functions show strong evidences for the existence of high and low volatility regimes and for the time varying transition probability of the regime variable.

Figure 5: Conditional Transition Density Functions and 95% Confidence Bands for the Model H-R1 and H-R2TVTP

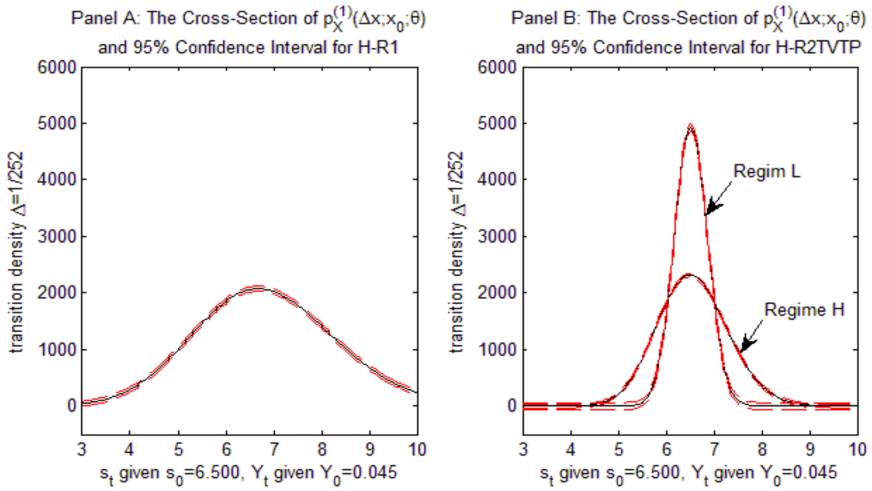


Table 3 reports the ML estimation results for the family of GARCH stochastic volatility models. The first column shows ML estimation outcomes for the single-regime model, G-R1. The rest three columns display the three different two-regime models G-R2-1(two regimes, time invariant transition matrix,unconditional probability for the probability of initial state), G-R2-2(two regimes, time invariant transition matrix, additional parameter for the probability of initial state) and G-R2TVP(two regimes,time varying transition matrix with logistic function,additional parameter for the probability of initial state) respectively. The asterisk by the parameter estimate implies that, at the 5% significance level, it is different from zero.

In the term of the single regime model G-R1, the long term value of the volatility $\gamma^{1/2}$ is approximately 17.89% per year, which is smaller than that in the H-R1 model. The speed of mean reversion coefficient κ is estimated to be 2.859, much smaller than 6.817 in the H-R1 model. The leverage effect ρ is -0.799 with the variance of volatility 2.319. However, the market price of risk λ and the rate of return rd are statistically insignificant.

To analyze the three different two-regime models, we find very similar conclusions as the class of Heston stochastic volatility models. $\kappa_H > \kappa_L$, $\gamma_H > \gamma_L$, $\rho_H > \rho_L$, $\sigma_H > \sigma_L$, rd_L,

Table 3: Maximum Likelihood Estimation Results of GARCH Models(01/02/1990-06/12/2012)

Model	GARCH-R1	GARCH-R2-1	GARCH-R2-2	GARCH-R2TVTP
Regime	Single Regime	Two Regimes		
		Time Invariant Transition Matrix		Time Varying Transition Matrix
$P(R_1 = L)$	-	$\frac{1-p_{HH}}{2-p_{LL}-p_{HH}}$		p
$P(R_{t+\Delta} = i \mid R_t = i)$ with $i = L$ and H	-	p_{ii}	p_{ii}	$\frac{\exp(c_i+d_i s_t+c_i V_t)}{1+\exp(c_i+d_i s_t+c_i V_t)}$
Log-likelihood	42061.62	42957.08	43063.94	43274.18
Parameters	Maximum Likelihood Estimates(Standard Error)			
ρ_L	$-0.80^{**}(0.000)$	$-0.76^{**}(0.000)$	$-0.76^{**}(0.000)$	$-0.70^{**}(0.000)$
ρ_H	-	$-0.81^{**}(0.000)$	$-0.80^{**}(0.000)$	$-0.80^{**}(0.000)$
κ_L	$2.86^{**}(0.66)$	$2.00^{**}(1.03)$	$2.00^{**}(0.94)$	$2.00^{**}(0.48)$
κ_H	-	$5.00^{**}(1.21)$	$5.00^{**}(1.28)$	$5.00^{**}(0.64)$
γ_L	$0.032^{**}(0.006)$	$0.030^{**}(0.012)$	$0.037^{**}(0.014)$	$0.030^{**}(0.005)$
γ_H	-	$0.063^{**}(0.013)$	$0.081^{**}(0.018)$	$0.11^{**}(0.013)$
σ_L	$2.32^{**}(0.001)$	$1.79^{**}(0.006)$	$1.79^{**}(0.005)$	$1.00^{**}(0.003)$
σ_H	-	$2.60^{**}(0.002)$	$2.60^{**}(0.003)$	$2.50^{**}(0.003)$
r_L	0.0002 (0.049)	0.0002 (0.084)	0.0002 (0.074)	0.0002 (0.009)
r_H	-	0.0002 (0.021)	0.0002 (0.044)	0.0002 (0.001)
λ_L	3.02 (3.78)	2.00 (7.40)	2.00 (6.80)	2.00 (6.27)
λ_H	-	4.00 (5.21)	4.00 (5.38)	4.00 (3.17)
p_{LL}	-	$0.099^{**}(0.013)$	$0.090^{**}(0.009)$	-
p_{HH}	-	$0.90^{**}(0.014)$	$0.900^{**}(0.010)$	-
c_L	-	-	-	-0.20 (1.66)
c_H	-	-	-	0.17 (0.22)
d_L	-	-	-	$-1.46^{**}(0.25)$
d_H	-	-	-	$1.24^{**}(0.032)$
e_L	-	-	-	-0.014 (4.28)
e_H	-	-	-	0.010 (0.20)
p	-	-	0.50 (1.01)	0.50 (0.54)
AIC	-14.36	-14.67	-14.70	-14.77
BIC	-14.36	-14.65	-14.69	-14.75
RCM		47.58	30.71	2.39

Note: Maximum likelihood estimates with standard errors in parentheses for the period 01/02/1990-06/12/2012 for the four different models with the general diffusion specification are presented in this table. The standard errors are calculated using sample mean of the outer product of the score functions evaluated at the ML estimates. According to the number of regimes, the starting probability and the transition probability matrix specifications, the four models are: G-R1 (single regime),G-R2-1 (two regimes, time invariant transition matrix,unconditional probability for the probability of initial state), G-R2-2 (two regimes, time invariant transition matrix, additioanl parameter for the probability of initial state), G-R2TVTP (two regimes,time varying transition matrix with logistic function,additional parameter for the probability of initial state).

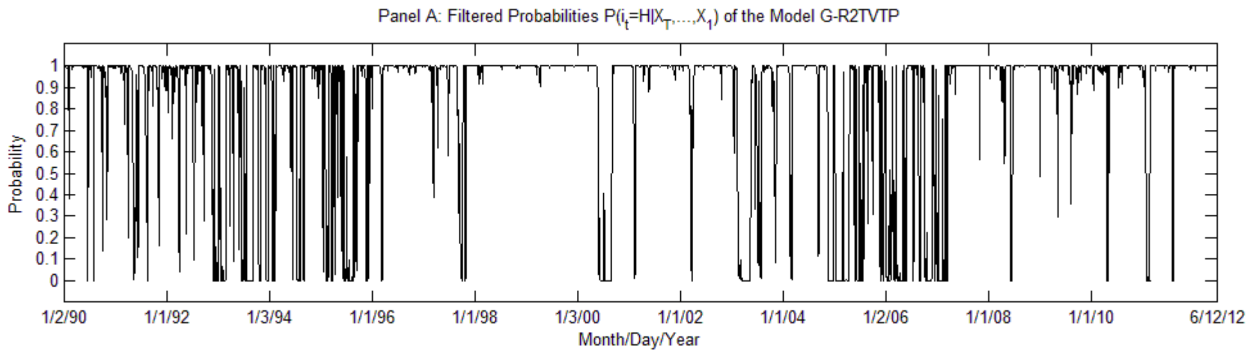
rd_H, λ_L and λ_H are still not statistically significant. Although κ_L is not significant, in fact, the p-value is 0.051, which is a bit bigger than 0.05. H State is very persistent while L state is not so steady. In the model of G-R2TVTP, time varying transition probabilities of i_t are mainly affected by the variable s_t, other than the constant term and Y_t. When $d_L < 0$ ($d_H > 0$), p_{LL} (p_{HH}) and s_t have a negative (positive) relationship. The ML estimates c and e are not significant.

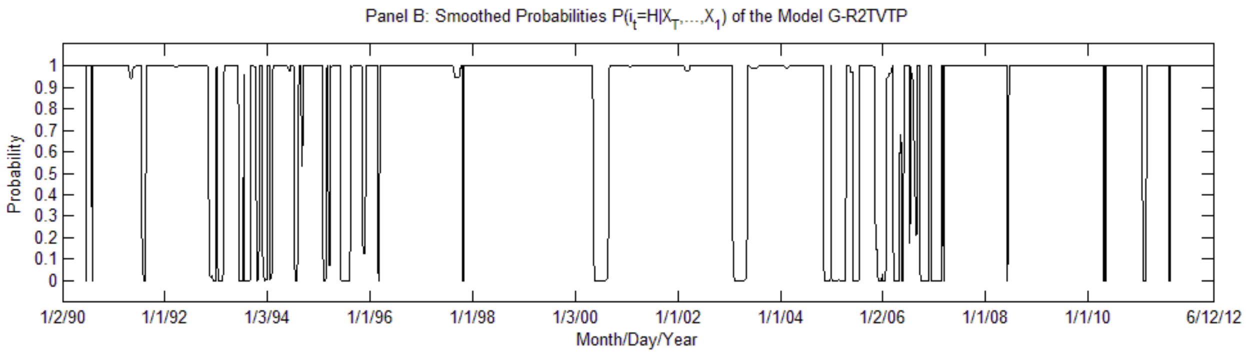
Comparing four log-likelihood values, the single regime model G-R1 reports the smallest value 42061.618. What we should pay attention to is that, in the Heston stochastic volatility model, the parameters have to satisfy the Feller condition $2\kappa\gamma \geq \sigma^2$, which is much more restricted than the condition $\kappa\gamma \geq 0$ in the G-R1 model. With Feller condition, κ, γ and σ can't vary so much as they do in the GARCH stochastic volatility model. We can find that in the G-R1 model, Feller condition is violated by $2 \times 2.859 \times 0.032 < 2.319^2$. Like the model of H-R2-1 and H-R2-2, the model of G-R2-1 and G-R2-2 also report very closed log-likelihood values and similar ML estimation outcomes. The log-likelihood value of G-R2TVTP is 43274.182 that is much bigger than the log-likelihood value of single regime model G-R1.

AIC and BIC are employed to compare the single regime model with three different two-regime models. The G-R1 model reports the biggest AIC -14.363 and BIC -14.356 while the G-R2TVTP model shows the smallest AIC -14.772 and BIC -14.751. The model of G-R2-1 and G-R2-2 display the similar AIC and BIC values. In the term of RCM, we can find strong advantage of the two regimes model with time varying transition matrix.

The filtered probabilities and smoothed probabilities of staying in H state over the sample for the model of G-R2TVTP are displayed.

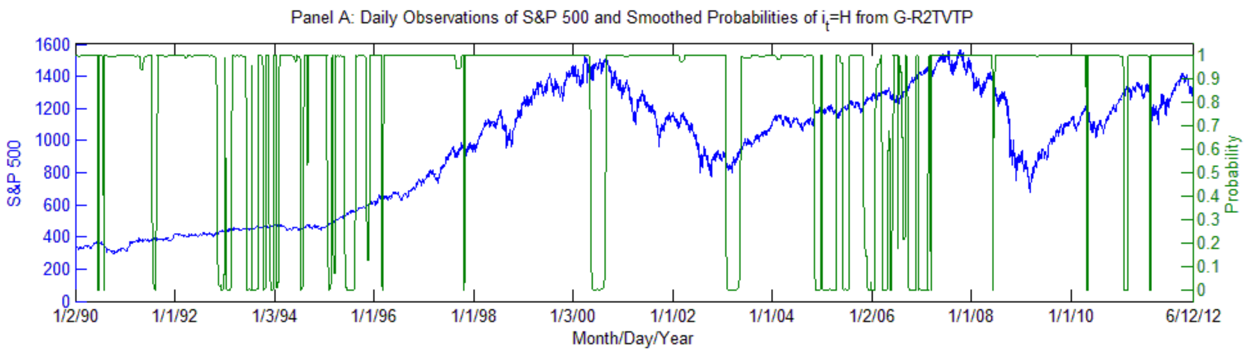
Figure 6: Regime-Switching Probabilities of the Model G-R2TVTP



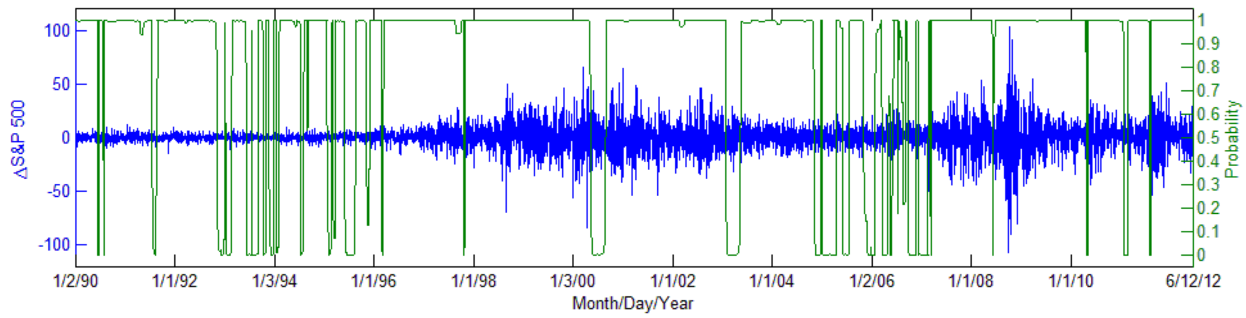


Then we plot the time series of smoothed probabilities with S&P 500, VIX and their first difference separately.

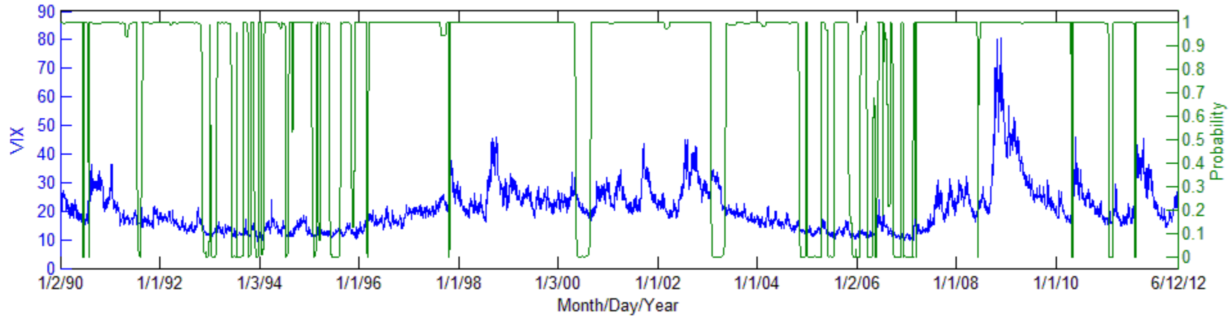
Figure 7: Regime-Switching Probabilities of the Model G-R2TVTP with S&P 500, VIX and Their First Difference



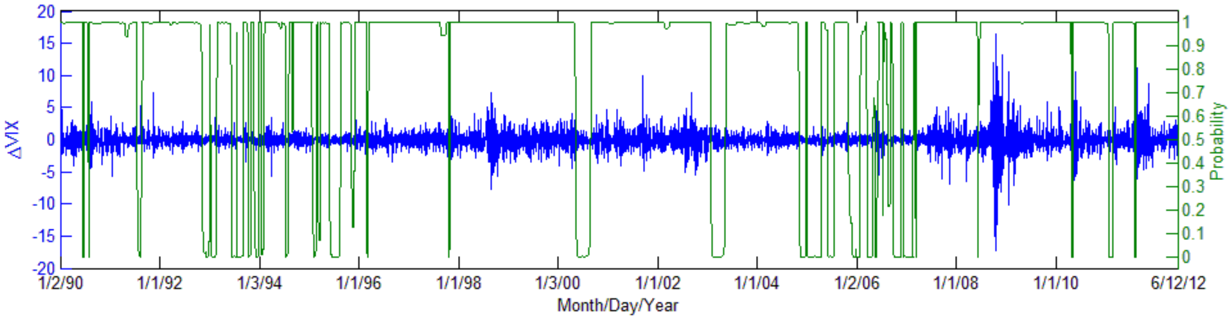
Panel B: Daily Changes in S&P 500 and Smoothed Probabilities of i_t=H from G-R2TVTP



Panel C: Daily Observations of VIX and Smoothed Probabilities of i_t=H from G-R2TVTP

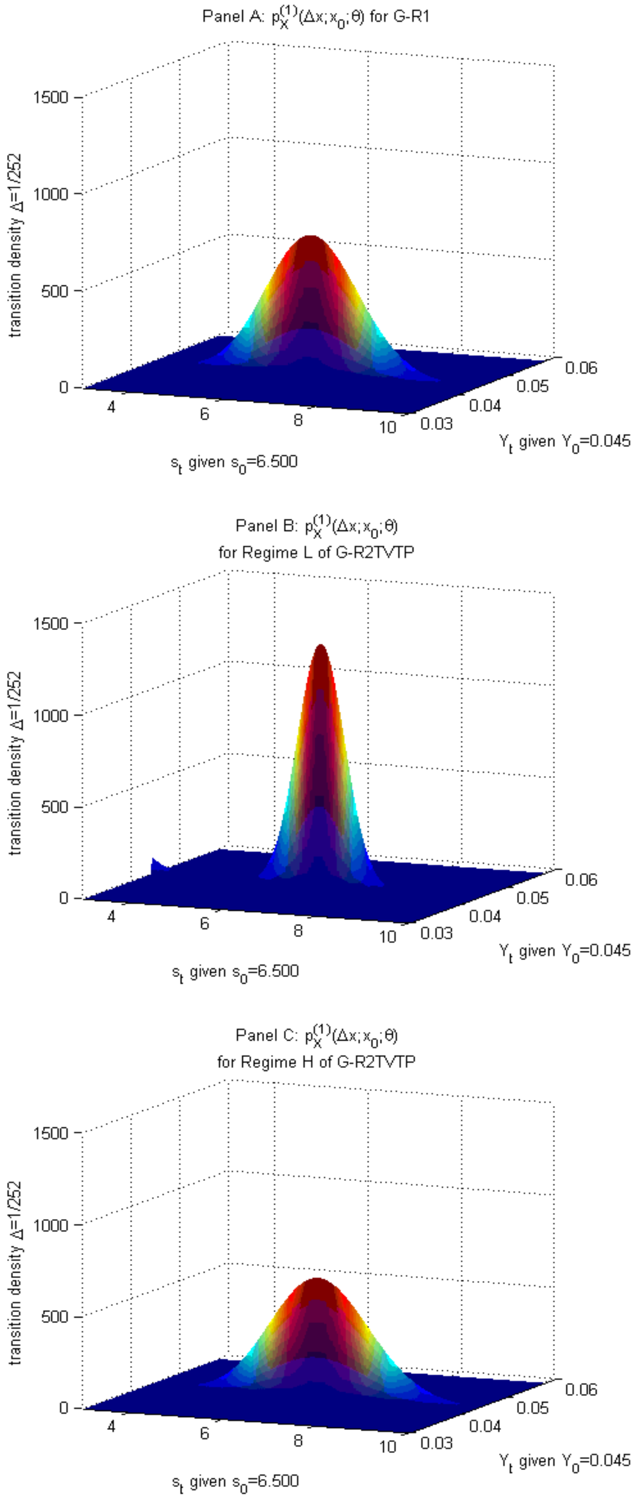


Panel D: Daily Changes in VIX and Smoothed Probabilities of i_t=H from G-R2TVTP



Although the smoothed probabilities of G-R2TVTP are not so smooth as that of the model H-R2TVTP, it still identifies the main high volatility periods of 1990-1991, 1994-1995, 1998-2000, 2008-2010 and 2010-2012. Using the ML estimates provided in table 3, we plot the approximate conditional transition density functions of the stochastic differential equations still given $x_0 = [s_0, Y_0]' = [6.500, 0.045]'$ for the model of G-R1 and G-R2TVTP, respectively. At around the point x_0, the conditional density function of regime L for the model H-R2TVTP has a peak of 1318.050 while that of regime H is with the top of 641.312 and the model G-R1 is centered on a height of 716.815.

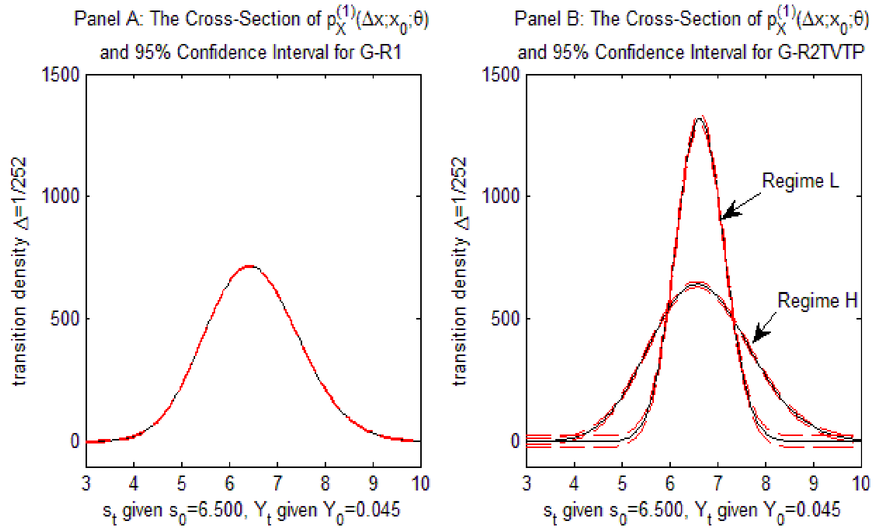
Figure 8: Conditional Transition Density Functions for the Model G-R1 and G-R2TVTP



In order to show 95% confidence bands, we plot the cross sections at $Y_0 = 0.045$. These

95% confidence interval surfs are very closed to the conditional density function for each specification.

Figure 9: Conditional Transition Density Functions and 95% Confidence Bands for the Model G-R1 and G-R2TVTP



In table 4, we display the estimation results for the model of C-R1, C-R2-1, C-R2-2 and C-R2TVTP respectively. The instantaneous standard deviation of S&P 500 $\gamma^{1/2}$ in the model C-R1 is 23.45% per year. The rate of mean reversion coefficient κ is estimated to be 6.049 with the correlation coefficient $\rho = -0.809$. The variance of volatility σ is close to 0.994, which is between 0.726 from the model G-R1 and 2.319 from the model H-R1. Of particular interest for the CEV model is the exponent β, which is estimated at 0.54, above the Heston value of 0.5 but below the GARCH value of 1. The market price of risk λ and the rate of return rd are statistically insignificant.

Comparing three different regime-switching models, we find similar results as the Heston and GARCH stochastic volatility models. $\kappa_H > \kappa_L$, $\gamma_H > \gamma_L$, $\rho_H > \rho_L$, $\sigma_H > \sigma_L$, however, the exponent $\beta_H < \beta_L$. λ_L and rd_L are statistically significant at the 5% level, but λ_H and rd_H are still insignificant.

Considering the log-likelihood values from the four different models, the single regime model C-R1 exhibits the smallest value 42397.267, which is bigger than 42061.618 from the model G-R1 and 42061.618 of the model H-R1. The main reason is still Feller condition

Table 4: Maximum Likelihood Estimation Results of CEV Models(01/02/1990-06/12/2012)

Model	CEV-R1	CEV-R2-1	CEV-R2-2	CEV-R2TVTP
Regime	Single Regime	Two Regimes		
		Time Invariant Transition Matrix		Time Varying Transition Matrix
$P(R_1 = L)$	-	$\frac{1-p_{HH}}{2-p_{LL}-p_{HH}}$		p
$P(R_{t+\Delta} = i \mid R_t = i)$ with $i = L$ and H	-	p_{ii}	p_{ii}	$\frac{\exp(c_i+d_i s_t+c_i V_t)}{1+\exp(c_i+d_i s_t+c_i V_t)}$
Log-likelihood	42397.27	43096.67	43197.40	43734.81
Parameters	Maximum Likelihood Estimates (Standard Error)			
ρ_L	$-0.81^{**}(0.001)$	$-0.71^{**}(0.004)$	$-0.71^{**}(0.004)$	$-0.70^{**}(0.002)$
ρ_H	-	$-0.80^{**}(0.002)$	$-0.80^{**}(0.002)$	$-0.83^{**}(0.001)$
κ_L	$6.05^{**}(1.30)$	$2.20^{**}(0.96)$	$2.18^{**}(0.98)$	$3.00^{**}(0.85)$
κ_H	-	$7.03^{**}(1.38)$	$7.03^{**}(1.27)$	$5.00^{**}(1.28)$
γ_L	$0.055^{**}(0.011)$	$0.020^{**}(0.006)$	$0.020^{**}(0.006)$	$0.020^{**}(0.004)$
γ_H	-	$0.045^{**}(0.007)$	$0.045^{**}(0.008)$	$0.067^{**}(0.016)$
σ_L	$0.99^{**}(0.004)$	$0.60^{**}(0.014)$	$0.60^{**}(0.014)$	$0.61^{**}(0.009)$
σ_H	-	$1.12^{**}(0.006)$	$1.12^{**}(0.007)$	$1.32^{**}(0.007)$
β_L	0.54 (0.004)	$0.74^{**}(0.004)$	$0.74^{**}(0.004)$	$0.75^{**}(0.002)$
β_H	-	$0.65^{**}(0.002)$	$0.65^{**}(0.002)$	$0.65^{**}(0.001)$
r_L	0.0002 (0.082)	0.0002 (0.091)	0.0002 (0.091)	0.0002 (0.017)
r_H	-	0.0002 (0.077)	0.0002 (0.084)	$0.0002^{**}(0.000)$
λ_L	6.07 (5.02)	3.00 (8.54)	3.00 (8.63)	2.00 (6.63)
λ_H	-	5.00 (5.75)	5.00 (5.53)	6.00 (4.15)
p_{LL}	-	$0.079^{**}(0.007)$	$0.090^{**}(0.009)$	-
p_{HH}	-	$0.90^{**}(0.008)$	$0.900^{**}(0.010)$	-
c_L	-	-	-	-0.19 (2.43)
c_H	-	-	-	0.16 (0.41)
d_L	-	-	-	$-1.44^{**}(0.37)$
d_H	-	-	-	$1.22^{**}(0.063)$
e_L	-	-	-	-0.013 (7.35)
e_H	-	-	-	0.009 (1.30)
p	-	-	0.50 (1.01)	0.50 (0.73)
AIC	-14.48	-14.71	-14.75	-14.93
BIC	-14.47	-14.70	-14.73	-14.91
RCM		34.58	33.11	7.78

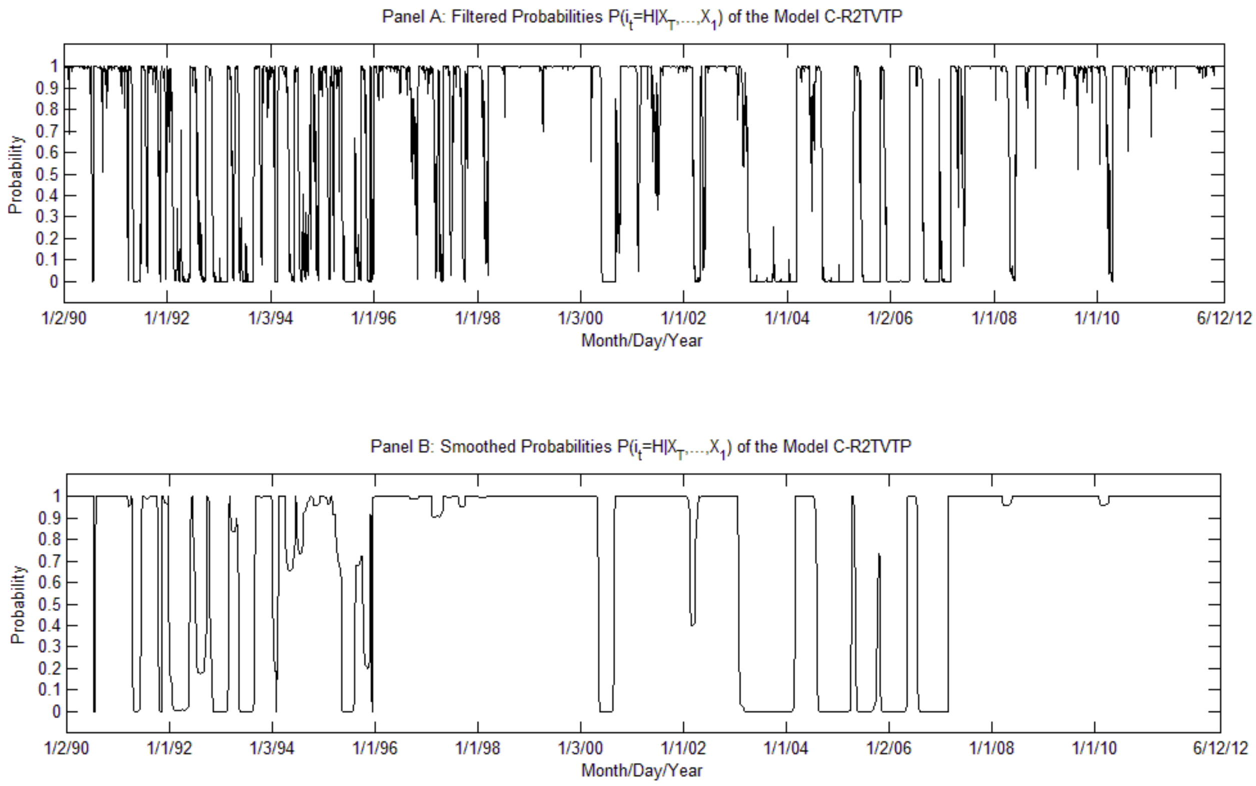
Note: Maximum likelihood estimates with standard errors in parentheses for the period 01/02/1990-06/12/2012 for the four different models with the general diffusion specification are presented in this table. The standard errors are calculated using sample mean of the outer product of the score functions evaluated at the ML estimates. According to the number of regimes, the starting probability and the transition probability matrix specifications, the four models are: C-R1 (single regime),C-R2-1 (two regimes, time invariant transition matrix,unconditional probability for the probability of initial state), C-R2-2 (two regimes, time invariant transition matrix, additioanl parameter for the probability of initial state), C-R2TVTP (two regimes,time varying transition matrix with logistic function,additional parameter for the probability of initial state).

$2\kappa\gamma \geq \sigma^2$. The models of C-R2-1 and C-R2-2 show log-likelihood values of 43096.667 and 43197.402, respectively. The largest log-likelihood value 43734.808 belongs to C-R2TVTP with time varying transition matrix.

When we move from the single regime model C-R1 to C-R2-1, C-R2-2 and C-R2TVTP, the AIC values decrease gradually from −14.477 to −14.929 and BIC declines from −14.469 to −14.905. The models of C-R2-1 and C-R2-2 report RCM 34.583 and 33.113 while the C-R2TVTP model shows RCM 7.781 that is much smaller than the previous two models. It is definitely a strong evidence for the existence of two regimes with varying transition matrix.

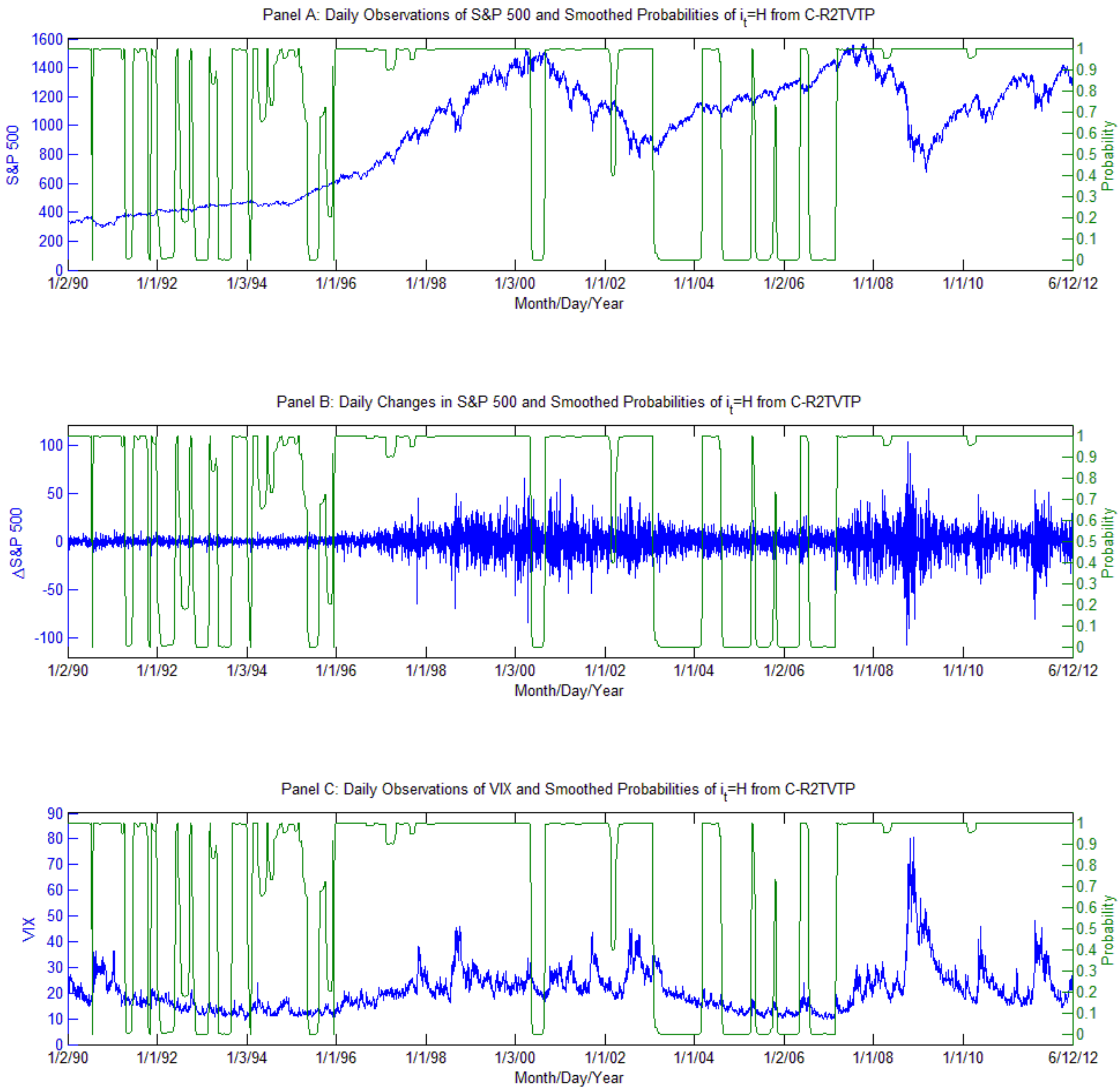
To analyze the probability of staying in state H, we draw the time series of filtered probabilities and smoothed probabilities over the sample for the model C-R2TVTP.

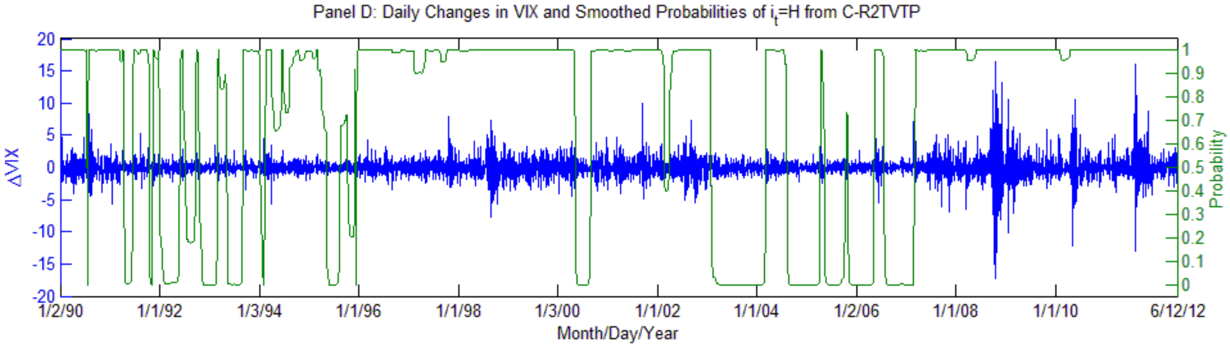
Figure 10: Regime-Switching Probabilities of the Model C-R2TVTP



Then we plot the time series of smoothed probabilities with S&P 500, VIX and their first difference separately.

Figure 11: Regime-Switching Probabilities of the Model C-R2TVTP with S&P 500, VIX and Their First Difference

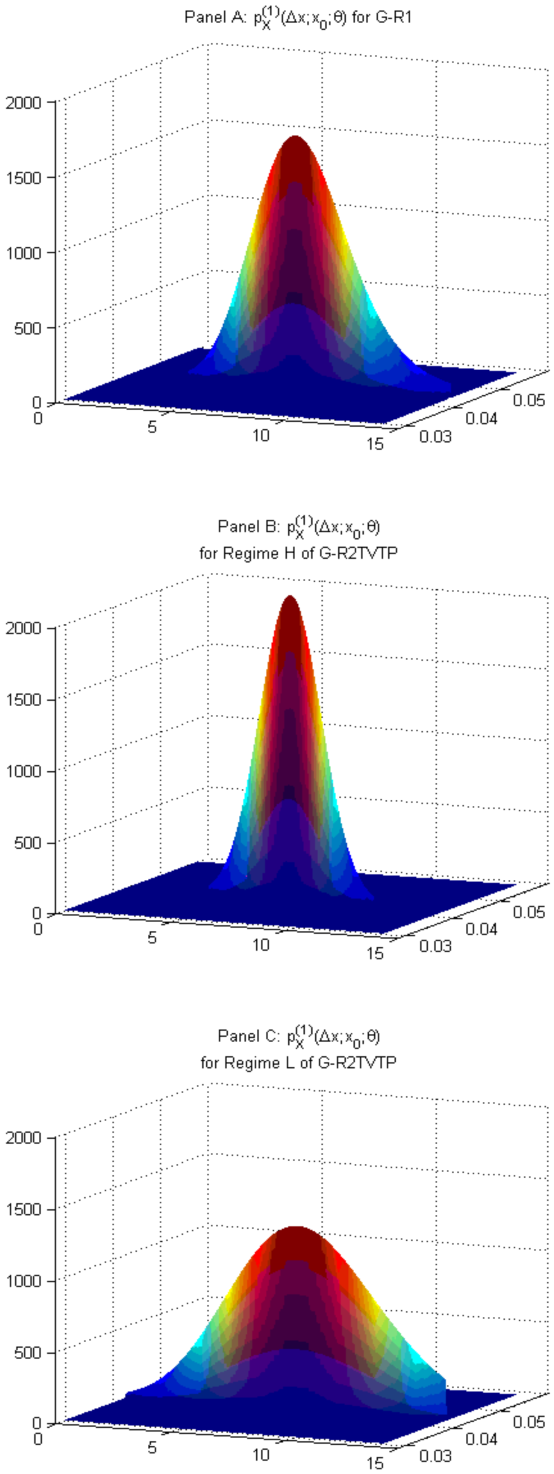




The smoothed probabilities are very smooth and therefore we can figure out most high volatility periods more clearly than the models of H-R2TVTP and G-R2TVTP. The 1990-1991 high regime state connects with the Gulf War, the 1992 period reflects the Black Wednesday, the high volatility regime around 1994 is linked to he 1994 Northridge earthquake in the Los Angeles area. The successive 1996-2003 H state is associated with the budget crisis in U.S., Asian financial crisis, Russian financial crisis, long-term capital management collapse, the dot-com bubble, September 11, 2001 terrorist attacks, as well as the October 2001 invasion in Afghanistan and 2003 Iraq War. The 2004 and 2005 high volatility regime is coincident with the 2004 Atlantic hurricane season and 2005 Hurricane Katrina and Hurricane Rita, respectively. The H state regime around 2007 matches the Chinese Correction plunge. The final long high volatility period of 2008-2012 is associated with the subprime mortgage crisis, Lehman Brothers Bankruptcy protection, the European sovereign-debt crisis, the 2010 Flash Crash and the August 2011 stock markets fall, which is still going on until now.

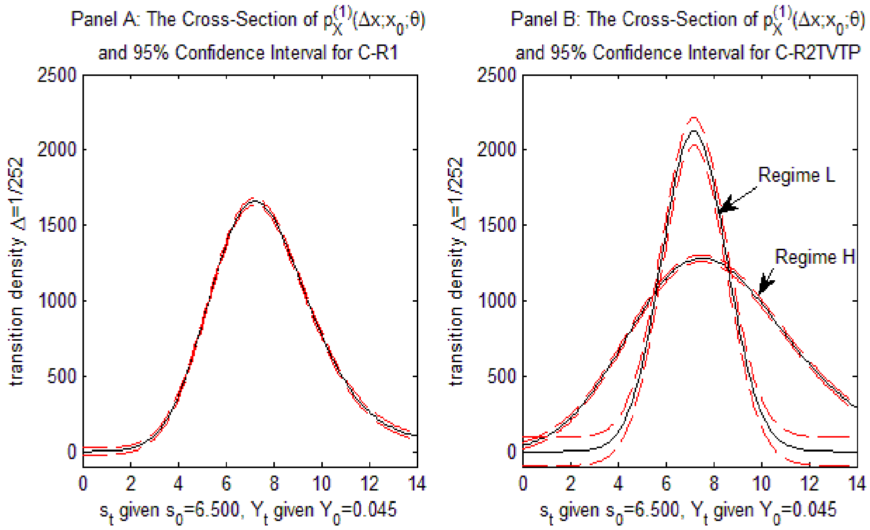
Using the ML estimates provided in table 4, given $x_0 = [s_0, Y_0]' = [6.500, 0.045]'$, the approximate conditional transition density functions of the stochastic differential equations for the C-R1 and C-R2TVTP models are plotted respectively. At around x_0 point, the conditional density function of regime L for the model C-R2TVTP peaks at the height of 2122.339, that of regime H is flattened out with the top of 1280.951 and the model C-R1 is centered on a height of 1658.754. Compared to the H-R2TVTP and G-R2TVTP models, the conditional transition density functions of the model C-R2TVTP for the two regimes are more close to each other as well as that of the single regime model.

Figure 12: Conditional Transition Density Functions for the Model C-R1 and C-R2TVTP



The cross sections at $Y_0 = 0.045$ of conditional transition density functions and their 95% confidence bands for the C-R1 model and the C-R2TVTP model are showed in Figure 13.

Figure 13: Conditional Transition Density Functions and 95% Confidence Bands for the Model G-R1 and G-R2TVTP



In order to compare three families of continuous time stochastic volatility models, we calculate likelihood ratio statistics for the nested models as Aït-Sahalia and Kimmel (2007) do. The H-R1 and G-R1 models are rejected in favor of the model C-R1 by reporting LRT statistic of 330.800 and 671.298, respectively. This result is cohere with Aït-Sahalia and Kimmel (2007). The LRT statistic for the model H-R1-1 against the model C-R1-1 is 5.282 and the corresponding p-value is 0.071. Hence, the H-R1-1 model can't be rejected. However, the statistic for the model G-R1-1 is 279.168 and cannot reject the hypothesis of the model C-R1-1. Considering the H-R1-2 and the G-R1-2 models, the LRT statistics against the C-R1-2 model are 206.752 and 266.926 respectively. So the C-R1-2 model is rejected. Moving to the regime-switching models with time varying transition matrix, the H-R2TVTP and G-R2TVTP models are rejected in favor of the model C-R2TVTP by reporting LRT statistic of 833.028 and 921.251, respectively. To conclude, the CEV model outperforms the Heston and GARCH models not only for the single regime models but also for the two-regime models.

6 Conclusions

Our estimation results show four main findings. First, the regime switching models are significantly different from the single regime models. Second, there are strong evidences for the existence of high and low volatility regimes, for the time varying transition probability of the regime variable, and for high persistence of the high regime. Third, the time varying transition probability mainly depends on the underlying asset price rather than its volatility. Fourth, except the H-R1-1 model, the other regime switching Heston and GARCH models are all rejected in favor of the regime switching CEV models. Furthermore, the regime switching CEV model with time varying transition matrix and additional parameter for the probability of initial state performs better than the other regime switching models.

References

AÏT-SAHALIA, Y. (2002): "Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach," *Econometrica*, 70(1), 223–262.

——— (2008): "Closed-Form Likelihood Expansions for Multivariate Diffusions," *Annals of Statistics*, 36(2), 906–937.

AÏT-SAHALIA, Y., AND R. KIMMEL (2007): "Maximum likelihood estimation of stochastic volatility models," *Journal of Financial Economics*, 83, 413–452.

ANG, A., AND G. BEKAERT (2002): "Regime Switches in Interest Rates," *Journal of Business and Economic Statistics*, 20(2), 163–182.

BANSAL, R., AND H. ZHOU (2002): "Term Structure of Interest Rates with Regime Shifts," LVII(5), 1997–2043.

BLACK, F. (1976): "Studies of stock price volatility changes," pp. 171–181, In: Proceedings of the 1976 Meetings of the American Statistical Association.

BLACK, F., AND M. SCHOLES (1973): "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81(3), 637–654.

CHACKO, G., AND L. M. VICEIRA (2003): "Spectral GMM Estimation of Continuous-Time Processes," *Journal of Econometrics*, 116, 259–292.

CHAN, K., KAROLYI, F. LONGSTAFF, AND A. SANDERS (1992): "An Empirical Comparison of Alternative Models of the Short-Term Interest Rate," 47(3), 1209–1227.

Cho, J. S., and H. White (2007): "Testing for Regime Switching," *Econometrica*, 75, 1671–1720.

Choi, S. (2009): "Regime-Switching Univariate Diffusion Models of the Short-Term Interest Rate," *Studies in Nonlinear Dynamics & Econometrics*, 13(1), Article 4.

——— (2013): "Closed-Form Likelihood Expansions for Multivariate Time-Inhomogeneous Diffusions," *Journal of Econometrics*, 174(2), 45–65.

——— (2015a): "Approximate Transition Probability Density Function of a Multivariate Time-inhomogeneous Jump Diffusion Process in a Closed-Form Expression," *Working paper*, University of Seoul.

——— (2015b): "Explicit Form of Approximate Transition Probability Density Functions of Diffusion Processes," *Journal of Econometrics*, 187, 57–73.

Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985): "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53(2), 385–408.

Dai, Q., K. J. Singleton, and W. Yang (2007): "Regime Shifts in a Dynamic Term Structure Model of U.S. Treasury Bond Yields," *Review of Financial Studies*, 20(5), 1669–1706.

Davis, R. B. (1987): "Hypothesis Testing When a Nuisance Parameter is Present Only under the Alternative," *Biometrika*, 74, 33–43.

Diebold, F., J.-H. Lee, and G. Weinbach (1994): "Regime Switching with Time-Varying Transition Probabilities," *in C. Hargreaves, ed., Nonstationary Time Series Analysis and Cointegration*, pp. 283–302, Oxford University Press. Oxford.

DiPietro, M. (2001): "Bayesian Inference for Discretely Sampled Diffusion Processes with Financial Applications," Ph.D. thesis, Department of Statistics, Carnegie-Mellon University.

Durham, G. B., and Y.-H. Park (2013): "Beyond Stochastic Volatility and Jumps in Returns and Volatility," *Journal of Business and Economic Statistics*, 31(1), 107–121.

Egorov, A. V., H. Li, and Y. Xu (2003): "Maximum Likelihood Estimation of Time Inhomogeneous Diffusions," *Journal of Econometrics*, 114, 107–139.

Feller, W. (1952): "The Parabolic Differential Equations and the Associated Semi-Groups of Transformations," 55, 468–519.

Gray, S. F. (1996): "Modeling the conditional distribution of interest rates as a regime-switching process," 42, 27–62.

Hamilton, J. D. (1989): "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57(2), 357–384.

——— (1994): *Time Series Analysis*. Princeton University Press.

Hansen, B. E. (1992): "The Likelihood Ratio Test Under Nonstandard Conditions: Testing the Markov Switching Model of GNP," *Journal of Applied Econometrics*, 7, S61–S82.

——— (1996): "Erratum:The Likelihood Ratio Test Under Nonstandard Conditions: Testing the Markov Switching Model of GNP," *Journal of Applied Econometrics*, 11(2), 195–198.

Heston, S. L. (1993): "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *Review of Financial Studies*, 6(2), 327–343.

Hull, J., and A. White (1987): "The pricing of options on assets with stochastic volatilities," *Journal of Finance*, 42, 281–300.

Jones, C. S. (2003): "The dynamics of stochastic volatility: evidence from underlying and options markets," *Journal of Econometrics*, 116, 181–224.

Lewis, A. (2000): *Option Valuation Under Stochastic Volatility*. Finance Press, Newport Beach.

Li, M. (2010): "A damped diffusion framework for financial modelingand closed-form maximum likelihood estimation," *Journal of Economic Dynamics and Control*, 34, 132–157.

Liechty, J. C., and G. O. Roberts (2001): "Markov Chain Monte Carlo Methods for Switching Diffusion Models," *Biometrika*, 88(2), 299–315.

Lo, A. W. (1988): "Maximum Likelihood Estimation of Generalized Ito Processes with Discretely Sampled Data," *Econometric Theory*, 4, 231–247.

Meddahi, N. (2001): "An eigenfunction approach for volatility modeling," *Working paper*, Université de Montréal.

Merton, R. C. (1973): "Theory of Rational Option Pricing," *Bell Journal of Economics and Management Science*, 4, 141–183.

Naik, V., and M. H. Lee (1998): "Yield Curve Dynamics with Discrete Shifts in Economics Regimes: Theory and Estimation," *Working paper*, University of Saskatchewan.

Nelson, D. (1990): "ARCH models as diffusion approximations," *Journal of Econometrics*, 45, 7–38.

Schaumburg, E. (2001): "Maximum Likelihood Estimation of Jump Processes," *Ph.D. thesis*, Princeton University.

Scott, L. (1987): "Option Pricing When the Variance Changes Randomly: Theory, Estimation, and an Application," *Journal of Financial and Quantitative Analysis*, 22, 419–438.

Stein, E., and J. Stein (1991): "Stock Price Distributions with Stochastic Volatility: An Analytic Approach," *Review of Financial Studies*, 4, 727–752.

Stramer, O., M. Bognar, and P. Schneider (2010): "Bayesian Inference for Discretely Sampled Markov Processes with Closed-Form Likelihood Expansions," *Journal of Financial Econometrics*, 8(4), 450–480.

Yu, J. (2007): "Closed-Form Likelihood Approximation and Estimation of Jump-Diffusions with an Applicaiton to the Realignment Risk of the Chinese Yuan," *Journal of Econometrics*, 141, 1245–1280.