Markov Switching GARCH Diffusion

Carol Alexander
ICMA Centre, University of Reading

Emese Lazar
ICMA Centre, University of Reading

Abstract[§]

GARCH option pricing models have the advantage of a well-established econometric foundation. However, multiple states need to be introduced as single state GARCH and even Lévy processes are unable to explain the term structure of the moments of financial data. We show that the continuous time version of the Markov switching GARCH(1,1) process is a stochastic model where the volatility follows a switching process. The continuous time switching GARCH model derived in this paper, where the variance process jumps between two or more GARCH volatility states, is able to capture the features of implied volatilities in an intuitive and tractable framework.

Author Details:

Professor Carol Alexander
Chair of Risk Management and Director of Research,
Email: c.alexander@icmacentre.rdg.ac.uk
Tel: +44 (0)1183 786431 (ICMA Centre) Fax: +44 (0)1189 314741

Emese Lazar
Lecturer in Finance,
Email: e.lazar@icmacentre.rdg.ac.uk
Tel: +44 (0)1183 786768 (ICMA Centre) Fax: +44 (0)1189 314741

Contents

I   INTRODUCTION

Multi-state GARCH models have been extensively studied in the works of Vlaar and Palm (1993), Cai (1994), Ding and Granger (1996), Gray (1996), Bauwens, Bos and van Dijk (1999), Bai, Russell and Tiao (2001, 2003), Roberts (2001), Klaassen (2002), Haas, Mittnik and Paolella (2004a, 2004b), Alexander and Lazar (2005b, 2006) and others. There is considerable evidence that these models offer better fit in the physical measure than single state GARCH processes. Also, multi-state GARCH has desirable features such as time variation and term structure in higher conditional moments and multiple leverage and mean-reverting mechanisms that are not a feature of single state GARCH models. We know that these properties are essential for any model that pertains to be consistent with the empirical properties of higher moments in financial returns and the behaviour of implied volatility (see Garcia, Ghysels and Renault, 2005). Even Lévy processes with a single state are inconsistent with the observed term structure of higher moments and switching models are needed to capture this behaviour, as shown by Konikov and Madan (2002).

Intuitively, a model with multiple mean-reverting volatility states will be appropriate if different types of shocks induce different responses. For instance, a rumour may have an enormous effect but die out very quickly whereas an announcement of important changes to economic policy may have a more persistent shock that also raises the general level of volatility. Single state GARCH models have only one mean-reversion mechanism and if there are in fact two types of shocks then the model cannot differentiate between them. If the model is estimated on high frequency data, the identified mean-reversion is likely to be rapid, whereas with low frequency data the model will capture the slower mean-reverting effect. There can also be differing degrees of leverage effect, i.e. asymmetry in the volatility response to price shocks. For instance, Alexander and Lazar (2005b) show that when equity markets are in a volatile regime the leverage effect is more pronounced. So if in fact there are regime specific leverage and mean-reversion effects, the coefficients in a single state GARCH process can represent only an average of these.

Since multi-state GARCH models provide a closer fit to both the conditional and the unconditional returns densities than single state GARCH processes, it is worthwhile to examine a continuous time pricing model with a multiple state GARCH processes as its econometric foundation. Nelson (1990) proved that the continuous time limit of a single state GARCH process is a mean-reverting stochastic volatility model that is similar to the popular model of Heston (1993). More recently Alexander and Lazar (2005a) derived the limit of weak GARCH processes but no continuous limit has been derived for multi-state GARCH processes. Switching models have been used for option pricing, but almost all models assume that the volatility is constant given the ruling state.

This paper introduces a new model for pricing options with two or more state dependent time-varying volatility processes where volatility jumps up or down between these mean-reverting states. To our

knowledge this is the first development of a continuous time GARCH option pricing model with multiple mean-reversion and leverage mechanisms, and we call it the continuous Markov switching GARCH model. The model has a strong micro foundation because it is the continuous time analogue of the Markov switching (MS) GARCH process of Haas, Mittnik and Paolella (2004b), which is the most tractable MS GARCH model because individual variances do not depend on lagged values of other individual variances. As a special case we consider the normal mixture (NM) GARCH model, where the conditional state probabilities are state-free and thus at each point in time the selection of the state is random and does not depend on the previous state. We show that because of this assumption the continuous time limit of the NM GARCH process falls outside the class of Lévy processes.

The remainder of this paper is organized as follows: Section II reviews the large research literature on option pricing using GARCH, other pricing models based on Markov switching and general models with price and volatility jumps. Section III defines the weak Markov Switching GARCH process, the weak formulation being necessary so that the GARCH component aggregates in time, and derives its continuous time limit. Section IV explores the properties of the continuous MS GARCH models by discussing the risk premium for state uncertainty, considers a replicating portfolio and discusses the model's discretization. Section V summarizes and concludes.

II   LITERATURE REVIEW

A path-breaking paper by Duan (1995) was the first to derive a discrete time GARCH option pricing model. The pricing is based on Monte Carlo simulations and utilizes the equivalence of the variance processes under the real-world and risk-neutral measures. Heston and Nandi (2000) derive a closed-form solution but only for a specific model. More recently Barone-Adesi, Engle and Mancini (2005) also use simulations for pricing but assume the variance processes and the parameters are different under the two measures. Siu, Tong and Yang (2004) use conditional Esscher transforms to choose the equivalent martingale measure that minimises the relative entropy between the real-world and the risk-neutral distribution. Elliott, Siu and Chan (2006) derive a discrete time pricing model, also using Esscher transforms, for MS GARCH processes based on the Hamilton and Susmel (1994) parameterization.

Bollen (1998) and Bollen, Gray and Whaley (2000) consider more general discrete time Markov switching models for option pricing, computing lattice-based European and American option prices. Later Hardy (2001) provided closed-form solutions for European options. Duan, Popova and Ritchken (2002) consider a model in which changes in the state in the Markov switching process are determined by an impact function for the innovations. Interestingly, the authors find that under certain assumptions, the limit of this model when the number of states goes to infinity is one of the asymmetric discrete time GARCH models. More recently Smith (2002), Hwang, Satchell and Valls Pereira (2004) and Kalimpalli and Susmel (2004) have all considered discrete time Markov switching in discrete time stochastic volatility

models. Discrete time Markov switching models do not converge as the time step decreases (Klein, 2002). However, they do have a continuous time counterpart: the continuous time Markov chain.[1]

Continuous time Markov switching models for option pricing were introduced by Naik (1993), who derived a closed form solution for simple European calls and puts when the underlying price dynamics are governed by a two state process with constant drift and volatility in each state. This idea was extended to mean-variance hedging by Di Masi, Kabanov and Runggaldier (1994); to pricing American and exotic options by Guo (2001a, 2001b) and Guo and Zhang (2004); and to more than two states by Jobert and Rogers (2006). Elliot et al. (1995) introduced a martingale process for the evolution of the state probability which forms the basis of other continuous time Markov switching models for option pricing, including those derived by Buffington and Elliott (2002) and Yao, Zhang and Zhou (2006). Continuous time Markov chains have been also combined with non-normal distributions in Konikov and Madan (2002), Albanese, Jaimungal and Rubisov (2003) and Elliott and Osakwe (2006), who price options using a two-state variance gamma process; Chourdakis (2000) prices simple and exotic options using a switching Lévy parameterization and Edwards (2005) considers a process that has different specifications (not just different parameters) in two different states.

Option pricing models with jumps have been identified as a very useful tool to describe option price behaviour – see Merton (1976), Bakshi, Cao and Chen (1997), Naik and Lee (1990), Jones (2003) and Johannes (2004). Jumps can be present in the price and/or the volatility processes; the main difference between these is that price jumps have a short-term effect and reduce the level of the variance of the diffusion, whilst variance jumps have a longer term effect, decreasing the variability and strengthening the autocorrelation in the variance. Bates (2000) and Pan (2002) show that jumps in the price process alone are not enough to explain major crashes in financial returns and do not generate enough negative skewness in the return density, which necessitates the use of volatility jumps.

Naik (1993) was the first to discuss deterministic jumps in the variance process; several other papers followed: Duffie, Pan and Singleton (2000) have considered models that have stochastic jumps in both processes with variance jumps being negatively correlated with price jumps. Interestingly, they obtain that jump times are clustered and this can be interpreted as an indication for multi-state processes. Eraker, Johannes and Polson (2003) show that models without volatility jumps are misspecified and cannot capture the correct dynamics for volatility. Later Eraker (2004) found that positive volatility jumps alone, whilst very important, still cannot explain the 1987 crash. He suggests the use of a more flexible model that allows for two-way jumps in the variance and the continuous Markov switching GARCH model derived in this paper does indeed display this feature.

---

[1] Other continuous time models as well can be discretized using Markov chains; for instance any stochastic volatility model can be approximated by a continuous time Markov chain, as shown by Kushner (1990) and Chourdakis (2002), and this methodology can be used for pricing options as for example in Chourdakis (2004).

Jumps are not only used in continuous time modelling; they have also been treated in a GARCH framework based on a discrete time analysis. Jorion (1988) was the first to discuss ARCH models with jumps in the mean equation. Maheu and McCurdy (2004) and Daal and Yu (2005) extended this framework to GARCH models where the residual is decomposed into a normal and a jump component, the normal part following a GARCH process. A more interesting model is that of Duan, Ritchken and Sun (2005a), where the GARCH variance process characterizes not only the normal component but the entire residual including the jump term. In another paper, Duan, Ritchken and Sun (2005b) show that their model converges to a continuous-time model with jumps in both the price and variance processes, but with diffusion in the price process only. If restricted to a normal GARCH, their limit model gives the continuous time limit GARCH model derived by Corradi (2000) because they use the same limiting assumptions for the parameters. Finally, Klüppelberg, Lindner and Maller (2004) introduced a continuous time process that features the properties of GARCH where the residuals follow a Lévy process but this is not an exact limit of the discrete time GARCH.

III   WEAK MARKOV SWITCHING GARCH

This section motivates the use of weak GARCH processes in the MS GARCH model, discusses the properties of the model and derives our theoretical results. All proofs are given in the appendix.

III.1   Motivation for Weak GARCH Processes

The GARCH(1,1) process, defined by Bollerslev (1986) as the generalized version of Engle's (1982) ARCH model, is given by the following:

$$(1) \qquad y_t = \mu + \varepsilon_t \text{ where } y_t = \frac{S_t - S_{t-1}}{S_{t-1}} \cong \ln\left(S_t / S_{t-1}\right)$$

$$(2) \qquad h_t = \omega + \alpha\varepsilon_t^2 + \beta h_{t-1}$$

The subscript t here stands for the time that the process becomes known; this means that h_t is the conditional variance for ε_{t+1}^2 and it is revealed at time t. The strong definition states that:

$$E\left(\varepsilon_{t+1}\big|I_t\right) = 0 \text{ and } E\left(\varepsilon_{t+1}^2\big|I_t\right) = h_t \text{ with } I_t = \left\{\varepsilon_t, \varepsilon_{t-1}, \ldots\right\}$$

However, since this process is not aggregating in time, it is more correct to study the continuous time limit of the weak GARCH(1,1) process defined by Drost and Nijman (1993) that is time aggregating. Only such a model would ensure that a GARCH process could be defined for any time step and it will have:

$$(3) \qquad \begin{aligned} E\left(\varepsilon_{t+1}\varepsilon_{t-i}^r\right) &= 0 \quad i \geq 0 \quad r = 0,1,2 \\ E\left(\left(\varepsilon_{t+1}^2 - h_t\right)\varepsilon_{t-i}^r\right) &= 0 \quad i \geq 0 \quad r = 0,1,2 \end{aligned}$$

In this model h_t is not the conditional variance but the best linear predictor (BLP) of the squared residuals. Alexander and Lazar (2005a) prove that as the time interval between realisations decreases to zero the data generating process (1) – (3) will converge to a stochastic volatility model characterized by mean-reversion in the variance process and non-zero correlation between the variance and the returns processes. This is a more general model than Nelson's (1990) limit, in which the correlation between returns and variance must be zero.

To analyse the continuous limit of multi-state GARCH processes we need to define a weak multi-state GARCH(1,1) with a GARCH component that aggregates in time. It is known that Markov switching models do not aggregate in time, although they do have counterparts given by continuous Markov chains (see Cox and Miller, 1987). This lack of time aggregation means that the continuous time model derived here is not the exact limit of the MS GARCH process; rather, it is its continuous time counterpart. Our derivation is based on the Markov Switching GARCH formulation introduced by Haas, Mittnik and Paolella (2004b) as it is the most tractable and intuitive MS GARCH model in the extant literature.

III.2 Definition and Properties of Weak MS GARCH

The state at time t is denoted by s_t and this can take K values, $i = 1,\ldots, K$. Besides (1), the model has K distinct GARCH processes:

(4)
$$h_{i,t} = \omega_i + \alpha_i \varepsilon_t^2 + \beta_i h_{i,t-1} \qquad i = 1,\ldots, K$$

and the strong MS GARCH process assumes:[2]

$$E\left(\varepsilon_{t+1} \middle| I_t\right) = 0, \; E\left(\varepsilon_{t+1} \middle| I_t, s_{t+1} = i\right) = \mu_i \; \text{and} \; E\left(\varepsilon_{t+1}^2 - \mu_i^2 \middle| I_t, s_{t+1} = i\right) = h_{i,t}$$

However, the weak (i.e. time aggregating) MS GARCH process assumes:[3]

(5)
$$E\left(\left(\varepsilon_{t+1} - \mu_i\right)\varepsilon_{t-j}^r \middle| s_t = i\right) = 0 \quad j \geq 0; \quad r = 0, 1, 2$$

(6)
$$E\left(\left(\varepsilon_{t+1}^2 - \mu_i^2 - h_{i,t}\right)\varepsilon_{t-j}^r \middle| s_t = i\right) = 0 \quad j \geq 0; \quad r = 0, 1, 2$$

and hence:

(7)
$$E\left(\varepsilon_{t+1}\varepsilon_{t-j}^r\right) = 0 \quad j \geq 0; \quad r = 0, 1, 2$$

(8)
$$E\left(\left(\varepsilon_{t+1}^2 - h_t\right)\varepsilon_{t-j}^r\right) = 0 \quad j \geq 0; \quad r = 0, 1, 2$$

---

[2] Note that Haas, Mittnik and Paolella (2004b) use zero conditional means for the residuals in different regimes, arguing that non-zero means would contradict Timmermann (2000)'s results. Timmermann (2000) showed empirically that when the residuals have non-zero regime-specific means then the returns are autocorrelated. Haas et al. (2004b) state that due to the autocorrelation in the returns, fitting a GARCH model in this situation is not appropriate. However, we argue that it is only autocorrelation in the residuals, and not autocorrelation in returns, that would make a GARCH fit undesirable.

[3] Here $h_{i,t}$ no longer denotes the conditional variance, but the best linear predictor of the squared returns in state i.

Let $\mathbf{p}_t$ denote the time-dependent state indicator vector, i.e.:

(9)
$$\mathbf{p}_t = (p_{i,t})_i \quad \text{where} \quad p_{i,t} = \begin{cases} 1 & \text{if } s_t = i \text{ (i.e. the system is in state } i\text{)} \\ 0 & \text{otherwise} \end{cases}$$

This is a vector of zeros, except for a single unit element corresponding to the ruling state. In other words, the state indicator follows a discrete time Markov chain with state space $\{e_1, ..., e_K\}$ of unit vectors and where $\mathbf{p}_t = e_i$ in state i. The transition probabilities are:

(10)
$$q_{ij} := P(s_t = i \mid s_{t-1} = j) = P(s_t = i \mid s_{t-1} = j, s_{t-2} = l, ...)$$

and the transition matrix is:

$$\mathbf{Q} = (q_{ij}) \quad \text{with} \quad \sum_{i=1}^{K} q_{ij} = 1 \quad \text{and} \quad q_{ij} \geq 0$$

The conditional expectation of the state indicator vector is defined as:

(11)
$$\tilde{\mathbf{p}}_t = E(\mathbf{p}_t \mid I_t) = P(s_t = i \mid I_t)$$

This gives the state probabilities conditional on an arbitrary information set I_t. This can be the realized state at and before a previous point in time or the distribution of the state vector at any previous point in time. Based on (10) we have the following Markov updating formula for the state probabilities:

(12)
$$\tilde{\mathbf{p}}_t = \mathbf{Q} \, \tilde{\mathbf{p}}_{t-1}$$

The unconditional expectation of the state vector gives the unconditional probabilities of the states:

(13)
$$\pi_i := E(p_{i,t}) = P(s_t = i) \quad \text{where} \quad \sum_{i=1}^{K} \pi_i = 1; \quad \boldsymbol{\pi} = (\pi_1, ..., \pi_K)' = E(\mathbf{p}_t)'$$

Because the residuals have zero expected value, we have:

(14)
$$\sum_{i=1}^{K} \pi_i \mu_i = 0$$

Also, it can be shown that the following relationship holds:

(15)
$$\mathbf{Q} \, \pi = \pi$$

For instance, when there are only two possible states:

(16)
$$\pi_1 = (1 - p_{22}) / (2 - p_{11} - p_{22})$$

We define the weighted GARCH process to be the weighted average of the component GARCH processes, i.e.:

(17)
$$h_t = \sum_{i=1}^{K} \pi_i \left(h_{i,t} + \mu_i^2 \right)$$

This is the best linear predictor of the squared returns when there is no information on the state process. Its unconditional expectation can be expressed as the weighted sum of the unconditional expectations of the GARCH terms and this will equal the unconditional variance of the residuals:

$$(18) \qquad E\left(h_t\right) = E\left(\varepsilon_t^2\right) = \sum_{i=1}^{K} \pi_i \left(E\left(h_{i,t}\right) + \mu_i^2\right)$$

To prove our theoretical results in the next section we shall need to decompose this variance, knowing the relative contribution of each GARCH component. To this end we define:

$$(19) \qquad z_i = E\left(h_t\right)^{-1} E\left(h_{i,t}\right)$$

$$(20) \qquad \eta_{i,t+1} = z_i \varepsilon_{t+1}^2 - h_{it}$$

It is easy to see that $E\left(\eta_{i,t}\right) = 0$. Re-writing (20) in the form

$$\eta_{i,t+1} = z_i \left(\varepsilon_{t+1}^2 - h_t\right) + \left(z_i h_t - h_{i,t}\right)$$

we see that $\eta_{i,t}$ captures two time varying effects for each component, viz.: the deviation of the squared return from the weighted GARCH process, and the difference between the individual GARCH process and a proportion z_i of the weighted GARCH process.

III.3 Convergence Results

Our results concern the convergence of a discrete time model to its continuous time counterpart. Thus we need to re-write the MS-GARCH model using a notation that facilitates this analysis. In the following the pre-subscript stands Δ for the step-length used. That is, time is indexed as $k\Delta$, with $k = 1, 2,...$ where Δ is used for a time series with step-length Δ. We must consider the normalized variance processes obtained by dividing by the step length Δ, because we need to compare the variances for processes of different frequencies. For the same purpose we shall divide the Δ-step squared error process by Δ in the GARCH formulation.

Thus, for an arbitrary step-length Δ the MS-GARCH model is specified as follows:

$$(21) \qquad {}_\Delta y_{k\Delta} = \Delta\mu + {}_\Delta\varepsilon_{k\Delta} \text{ where } {}_\Delta y_{k\Delta} = \frac{S_{k\Delta} - S_{(k-1)\Delta}}{S_{(k-1)\Delta}} \cong \ln\left(S_{k\Delta} / S_{(k-1)\Delta}\right)$$

$$(22) \qquad {}_\Delta h_{i,k\Delta} = {}_\Delta\omega_i + {}_\Delta\alpha_i \, {}_\Delta\varepsilon_{i,k\Delta}^2 / \Delta + {}_\Delta\beta_i \, {}_\Delta h_{i,(k-1)\Delta} \qquad i = 1,\ldots, K$$

$$(23) \qquad E\left({}_\Delta\varepsilon_{(k+1)\Delta}\left({}_\Delta\varepsilon_{(k-j)\Delta}^r - \mu_i\right)\big|\, {}_\Delta s_{k\Delta} = i\right) = 0 \quad j \geq 0 \quad r = 0,1,2$$

$$(24) \qquad E\left(\left({}_\Delta\varepsilon_{(k+1)\Delta}^2 / \Delta - \mu_i^2 - {}_\Delta h_{k\Delta}\right) {}_\Delta\varepsilon_{(k-j)\Delta}^r \big|\, {}_\Delta s_{k\Delta} = i\right) = 0 \quad j \geq 0 \quad r = 0,1,2$$

The state indicator vector is:

(25) $\quad _\Delta\mathbf{P}_{k\Delta} = \left(_\Delta p_{i,k\Delta} \right)_i \;$ where $\; _\Delta p_{i,k\Delta} = \begin{cases} 1 & \text{if } _\Delta s_{k\Delta} = i \text{ (i.e. the system is in state } i) \\ 0 & \text{otherwise} \end{cases}$

The unconditional expectation of this does not depend on the observation interval, i.e. $E\left(_\Delta p_t \right) = \pi$. The conditional expectation of the state indicator vector is defined as:

(26) $\quad\quad\quad _\Delta\tilde{\mathbf{p}}_{k\Delta} = E\left(\mathbf{p}_{k\Delta} \mid I_{k\Delta} \right) = P\left(_\Delta s_{k\Delta} = i \mid I_{k\Delta} \right)$

The transition matrix is given by:

(27) $\quad _\Delta q_{ij} := P\left(_\Delta s_{k\Delta} = i \mid _\Delta s_{(k-1)\Delta} = j \right) = P\left(_\Delta s_{k\Delta} = i \mid _\Delta s_{(k-1)\Delta} = j, _\Delta s_{(k-2)\Delta} = l, \ldots \right)$

$$_\Delta\mathbf{Q} = \left(_\Delta q_{ij} \right) \text{ with } \sum_{i=1}^{K} {_\Delta q_{ij}} = 1 \text{ and } {_\Delta q_{ij}} \geq 0$$

We have the following properties:

(28) $$\sum_{i=1}^{K} \pi_i \, \mu_i = 0$$

(29) $$_\Delta\mathbf{Q}\,\pi = \pi$$

(30) $$_\Delta\tilde{\mathbf{p}}_{k\Delta} = {_\Delta\mathbf{Q}} \; _\Delta\tilde{\mathbf{p}}_{(k-1)\Delta}$$

The weighted GARCH process and its unconditional expectation are:

(31) $$_\Delta h_{k\Delta} = \sum_{i=1}^{K} \pi_i \left(_\Delta h_{i,k\Delta} + \mu_i^2 \right)$$

(32) $$E\left(_\Delta h_{k\Delta} \right) = E\left(_\Delta \epsilon_{k\Delta}^2 \right) = \sum_{i=1}^{K} \pi_i \left(E\left(_\Delta h_{i,k\Delta} \right) + \mu_i^2 \right)$$

Also, we have:

(33) $$z_i = E\left(_\Delta h_{i,k\Delta} \right) / E\left(_\Delta h_{k\Delta} \right)$$

(34) $$_\Delta\eta_{i,(k+1)\Delta} = z_i \; _\Delta\epsilon_{(k+1)\Delta}^2 - {_\Delta h_{i,k\Delta}}$$

with $E\left(_\Delta\eta_{i,k\Delta} \right) = 0$, and:

(35) $$E\left(_\Delta\epsilon_{(k+1)\Delta} \; _\Delta\epsilon_{(k-j)\Delta}^r \right) = 0 \quad j \geq 0; \quad r = 0, 1, 2$$

(36) $$E\left(\left(_\Delta\epsilon_{(k+1)\Delta}^2 - {_\Delta h_{k\Delta}} \right) _\Delta\epsilon_{(k-j)\Delta}^r \right) = 0 \quad j \geq 0; \quad r = 0, 1, 2$$

The above formulation allows us to derive the following results, which are necessary for the continuous version of the weak MS GARCH model:

Theorem 1: The class of weak MS GARCH processes has a GARCH component that is closed under temporal aggregation.

Proposition 1: The weak MS GARCH model has the following convergence rates:

$$\omega_i = \lim_{\Delta \downarrow 0} \Delta^{-1} {}_\Delta\omega_i; \quad \theta_i = \lim_{\Delta \downarrow 0} \Delta^{-1}\left(1 - \left(z_i^{-1} {}_\Delta\alpha_i + {}_\Delta\beta_i\right)\right); \quad 0 < \omega_i, \theta_i < \infty$$

III.4   The Need for Markov Switching

In this section consider the normal mixture (NM) GARCH model that is a reduced form of Markov Switching GARCH model where the transition matrix is given by ${}_\Delta\mathbf{Q} = \boldsymbol{\pi} \cdot \mathbf{1'}$, i.e. a matrix with rank 1. In this case the error term is assumed to follow a conditional normal mixture distribution with mixing law $\boldsymbol{\pi}$:

$$(37) \qquad {}_\Delta\varepsilon_{k\Delta} \Big| I_{(k-1)\Delta} \sim NM\left(\pi_1, ..., \pi_K; \mu_1, ..., \mu_K; {}_\Delta h_{1,(k-1)\Delta}, ..., {}_\Delta h_{K,(k-1)\Delta}\right)$$

with the conditional density functions $f_\varepsilon\left(\varepsilon_t\right) = \sum_{i=1}^{K} \pi_i \varphi_{\varepsilon i}\left(\varepsilon_t\right)$ where

$$\varphi_{\varepsilon i}\left(\varepsilon_t\right) = \frac{1}{\sigma_{it-1}\sqrt{2\pi}} \exp\left(-\frac{\left(\varepsilon_t - \mu_i\right)^2}{2\sigma_{it-1}^2}\right).$$

At each point in time we have one ruling state and in each state the returns follow a normal distribution. It is important to differentiate between two different types of conditional distributions. First, when only past data is known, as above, the conditional distribution of the residuals is a normal mixture. But when we are conditioning on past information <u>and</u> the prevailing state as well, then the conditional distribution is normal:

$$(38) \qquad {}_\Delta\varepsilon_{k\Delta} \Big| \left\{I_{(k-1)\Delta}, {}_\Delta s_{(k-1)\Delta} = i\right\} \sim N\left(\Delta\mu_i, \Delta {}_\Delta h_{i,(k-1)\Delta}\right)$$

Proposition 2: The variance of the continuous limit of the NM GARCH is not a Lévy process.

The problem with the NM-GARCH model is that it switches states too often. The history of the ruling states has no effect on the next state and the conditional probability of the switches does not depend on the step length.

III.5   Continuous Markov Switching GARCH

We now discuss the continuous time version of MS GARCH, which has the advantage over NM GARCH that it limits the number of switches between states when the step length converges to zero. As Δ decreases the transition matrix for MS GARCH converges to the identity matrix and it is possible to attain this without modifying the unconditional probability vector. This ensures that at smaller steps the conditional probability of jumps decreases whilst the probability of jumps over a given period remains unchanged.

We make the following definitions: The continuous state process, a continuous time Markov chain and the continuous variance processes, provided these exist, are:

(39) $$s(t) = \lim_{\Delta \downarrow 0} {}_\Delta s_t \quad \text{where} \quad {}_\Delta s_t = {}_\Delta s_{k\Delta} \quad \text{for } k\Delta \leq t < (k+1)\Delta$$

(40) $$V_i(t) = V(t)\big|\{s(t) = i\} = \lim_{\Delta \downarrow 0} {}_\Delta h_{i,t} \quad \text{where} \quad {}_\Delta h_{i,t} = {}_\Delta h_{i,k\Delta} \text{ for } k\Delta \leq t < (k+1)\Delta$$

The continuous time state probability vector and its expectation are:

$$\mathbf{p}(t) = \lim_{\Delta \downarrow 0} {}_\Delta \mathbf{p}_t \quad \text{where} \quad {}_\Delta \mathbf{p}_t = {}_\Delta \mathbf{p}_{k\Delta} \quad \text{for } k\Delta \leq t < (k+1)\Delta$$

(41) $$\tilde{\mathbf{p}}(t) = \lim_{\Delta \downarrow 0} {}_\Delta \tilde{\mathbf{p}}_t \quad \text{where} \quad {}_\Delta \tilde{\mathbf{p}}_t = {}_\Delta \tilde{\mathbf{p}}_{k\Delta} \quad \text{for } k\Delta \leq t < (k+1)\Delta$$

and the continuous time transition rate (or generator) matrix is:

(42) $$\mathbf{\Lambda} = (\lambda_{ij}) = \lim_{\Delta \downarrow 0} \Delta^{-1}({}_\Delta \mathbf{Q} - \mathbf{I}) \text{ with } \lambda_{ij} > 0, \ i \neq j \text{ and } \sum_{i=1}^{K} \lambda_{ij} = 0$$

whence $\mathbf{\Lambda}\boldsymbol{\pi} = \mathbf{0}$ where $\mathbf{0}$ is a vector of zeros.[4] The off-diagonal elements of the transition rate matrix are positive and the diagonal elements are negative and equal in absolute value to the sum of the other elements in the same column. By definition[5]

$$_\Delta \mathbf{Q} = \mathbf{I} + \Delta\,\mathbf{\Lambda} + o(\Delta)$$

In other words, during a time interval of length Δ:

$$P(\text{one jump from state } j \text{ to state } i) = \Delta\lambda_{ij} + o(\Delta)$$

$$P(\text{no jump from state } j \text{ to state } i) = 1 - \Delta\lambda_{ij} + o(\Delta)$$

$$P(\text{more than one jump from state } j \text{ to state } i) = o(\Delta)$$

It follows that if the continuous time Markov chain $\{s(t)\}$ has holding times $\{H_k\}$ (i.e. the duration of the time periods spent in a state) and jump times $\{J_k\}$ (i.e. the points in time at which the chain switches to a different state) where $J_k = \sum_{i=1}^{k} H_i$ and $J_0 = 0$, then the holding times are independent and follow an exponential distribution and the jump times follow a Poisson distribution, i.e.:

$$P(\text{the chain jumps from state } j \text{ in an interval of length } t) = P(H_k \leq t) = 1 - \exp(-\lambda_j t)$$

$$P(k \text{ jumps from state } j \text{ in an interval of length } t) = P(J_k \leq t) = \frac{(\lambda_j t)^k}{k!}\exp(-\lambda_j t)$$

where $\lambda_j = \sum_{i \neq j} \lambda_{ij}$. The proof can be found in the classical literature on continuous time Markov chains, for example, in Ethier and Kurtz (1986).

---

[4] On the other hand, the transition matrix of the NM-GARCH remains unchanged when the step length decreases; it is inflexible. Because of this the conditional probability of jumps remains unchanged, forcing the number of jumps over a given period of time to converge to infinity. This way the transition rate matrix will not exist – and this is why the NM GARCH model does not have a continuous time version.

[5] Here $f(x) = o(g(x))$ means that $\lim(f(x)/g(x)) = 0$ as $x \rightarrow 0$.

To derive the continuous time MS GARCH we shall use the weak convergence results of Stroock and Varadhan (1996) and for this we must assume that the following limits exist and are finite:

$$(43) \qquad \alpha = \lim_{\Delta \downarrow 0} \Delta^{-1} {}_{\Delta}\alpha \ ; \ \psi = \lim_{\Delta \downarrow 0} \Delta^{-1} \left(\mathbf{1} - {}_{\Delta}\beta \right)$$

with $\alpha_i \geq 0$, $\psi_i > 0$ and $\mathbf{1}$ is a vector of ones. Here ${}_{\Delta}\omega$, ${}_{\Delta}\alpha$ and ${}_{\Delta}\beta$ are $K \times 1$ vectors of the GARCH coefficients in the discrete MS GARCH process. Note that under the condition that ${}_{\Delta}\alpha$ converges to a positive constant at rate $\sqrt{\Delta}$, the variance processes would not converge.

Consider the first two conditional moments and the conditional skewness and kurtosis:[6]

$$_{\Delta}\mu_{i,k\Delta} = E\left(\Delta^{-1} {}_{\Delta}\varepsilon_{(k+1)\Delta} \middle| {}_{\Delta}I_{k\Delta}, {}_{\Delta}s_{k\Delta} = i \right)$$

$$_{\Delta}\sigma^2_{i,k\Delta} = E\left(\Delta^{-1} \left({}_{\Delta}\varepsilon_{(k+1)\Delta} - \Delta \, {}_{\Delta}\mu_{i,k\Delta} \right)^2 \middle| {}_{\Delta}I_{k\Delta}, {}_{\Delta}s_{k\Delta} = i \right)$$

$$_{\Delta}\tau_{i,k\Delta} = E\left(\left(\Delta^{3/2} {}_{\Delta}\sigma^3_{i,k\Delta} \right)^{-1} \left({}_{\Delta}\varepsilon_{(k+1)\Delta} - \Delta \, {}_{\Delta}\mu_{i,k\Delta} \right)^3 \middle| {}_{\Delta}I_{k\Delta}, {}_{\Delta}s_{k\Delta} = i \right)$$

$$_{\Delta}\eta_{i,k\Delta} = E\left(\left(\Delta^2 {}_{\Delta}\sigma^4_{i,k\Delta} \right)^{-1} \left({}_{\Delta}\varepsilon_{(k+1)\Delta} - \Delta \, {}_{\Delta}\mu_{i,k\Delta} \right)^4 \middle| {}_{\Delta}I_{k\Delta}, {}_{\Delta}s_{k\Delta} = i \right)$$

Now the conditional variance and the BLP series for the squared residuals will converge to the same process:

$$V_i(t) = \lim_{\Delta \downarrow 0} {}_{\Delta}\sigma^2_{i,t} \text{ where } {}_{\Delta}\sigma^2_{i,t} = {}_{\Delta}\sigma^2_{i,k\Delta} \text{ for } k\Delta \leq t < (k+1)\Delta.$$

if the following limits exist:

$$\varepsilon(t) = \lim_{\Delta \downarrow 0} {}_{\Delta}\varepsilon_t \quad \text{where} \quad {}_{\Delta}\varepsilon_t = {}_{\Delta}\varepsilon_{k\Delta} \text{ for } k\Delta \leq t < (k+1)\Delta$$

$$\mu_i(t) = \mu + \lim_{\Delta \downarrow 0} {}_{\Delta}\mu_{i,t} \quad \text{where} \quad {}_{\Delta}\mu_{i,t} = {}_{\Delta}\mu_{i,k\Delta} \text{ for } k\Delta \leq t < (k+1)\Delta$$

Thus BLP of the squared residuals in a given state is 'close' to the corresponding conditional variance process, and it is reasonable that as the step length Δ converges to zero

$$(44) \qquad \lim_{\Delta \downarrow 0} \Delta^{-1/2} \left({}_{\Delta}\sigma^2_{i,t} - {}_{\Delta}h_{i,t} \right) < \infty$$

This is because the BLP process becomes more and more informative as the time step decreases and so it should converge fast to the conditional variance.

We also have:

$$E\left(\Delta^{-1} {}_{\Delta}\varepsilon^2_{(k+1)\Delta} \middle| I_{k\Delta}, {}_{\Delta}s_{k\Delta} = i \right) = {}_{\Delta}\sigma^2_{i,k\Delta} + \Delta \, {}_{\Delta}\mu^2_{i,k\Delta}$$

---

[6] Recall that we divide by Δ when computing the conditional mean and variance series because these are additive in time.

and at least one of the processes $_\Delta\mu_{i,k\Delta}$ and $_\Delta\sigma^2_{i,k\Delta} - _\Delta h_{i,k\Delta}$ must be different from zero, otherwise the process is a semi-strong MS-GARCH and looses the aggregation property. Additionally, the conditional expectation of the second moment and the kurtosis in a given state must be positive. The above construction allows us to prove the following main result:

Theorem 2: If the limits given in (42), (41) and (43) exist then the continuous time counterpart of the Markov switching GARCH process is:

(45)
$$\frac{dS}{S} = \mathbf{p}'\mathbf{\mu}\, dt + \sqrt{V} dB$$
$$dV = \mathbf{p}' d\mathbf{V} + dJ \quad \text{with} \quad dJ = d\mathbf{p}'\mathbf{V}$$
$$d\mathbf{V} = \left(\mathbf{\omega} + \mathbf{\alpha} V - \mathbf{\psi} \odot \mathbf{V}\right) dt$$
$$d\tilde{\mathbf{p}} = \mathbf{\Lambda}\tilde{\mathbf{p}} dt$$

and we have dropped the time dependence of the processes here for ease of notation.[7]

However, this model is not appropriate for pricing equity or commodity options since it does not capture asymmetric volatility responses to market shocks. We also study the asymmetric extensions of MS GARCH; these are based on the AGARCH model of Engle and Ng (1993) and the GJR model of Glosten et al. (1993). Interestingly, it is only the MS AGARCH, and not the MS GJR-GARCH model that has state-dependent leverage effects in the limit, as the following corollary shows:

Corollary 1: The continuous time limit of the Markov switching AGARCH process with

$$_\Delta h_{i,(k+1)\Delta} = {}_\Delta\omega_i + {}_\Delta\alpha_i \left(\Delta^{-1}{}_\Delta\varepsilon_{(k+1)\Delta} - {}_\Delta\xi_i\right)^2 + {}_\Delta\beta_i\, {}_\Delta h_{i,k\Delta} \qquad i = 1,\dots, \text{K}$$

is given by (45) with the third equation replaced by:

$$dV = \left(\mathbf{\omega} + \mathbf{\alpha} \odot \mathbf{\xi} \odot \mathbf{\xi} + \mathbf{\alpha} V - \mathbf{\psi} \odot V\right) dt$$

where $\mathbf{\xi} = \left(\xi_i\right)_i = \lim_{h\downarrow 0} {}_h\mathbf{\xi}$. However, the model given by the GJR parameterisation with

$$_\Delta h_{i,(k+1)\Delta} = {}_\Delta\omega_i + {}_\Delta\alpha_i\, {}_\Delta\varepsilon^2_{(k+1)\Delta} / \Delta + {}_\Delta\xi_i\, {}_\Delta d^-_{(k+1)\Delta}\, {}_\Delta\varepsilon^2_{(k+1)\Delta} / \Delta + {}_\Delta\beta_i\, {}_\Delta h_{i,k\Delta} \qquad i = 1,\dots, \text{K}$$

where $_\Delta d^-_t = 1$ if $\varepsilon_t < 0$, and 0 otherwise, has the same limit as the MS-GARCH model.

The details of the proof are similar to those in the proof of Theorem 2 and are therefore omitted from the appendix for the sake of brevity. The reason why the GJR parameterisation has no asymmetric response in its limit is that whilst $_\Delta d^-_t$ is finite, it does not converge when $\Delta \downarrow 0$. However, it converges to the model in Theorem 2 under the condition $0 = \lim_{\Delta\downarrow 0} {}_\Delta\xi_i$, for all i.

---

[7] The notation $\odot$ stands for the element-by-element product.

IV   PROPERTIES OF THE CONTINUOUS MS GARCH MODEL

This section describes the stochastic returns and variance processes of the MS GARCH model, quantifies the state risk premium and then expresses the model in the risk neutral measure.

IV.1   Stochastic Returns and Variance Processes

The continuous MS GARCH model (45) has the following interpretation: The returns follow a stochastic process where the drift depends on the ruling state and the diffusion is determined by the realized variance $V(t)$. The vector of state-variances is a stochastic process with state dependent mean-reversion. The change in the realized variance is given by two factors. First it depends on the change in the variance process of the ruling state and secondly, if there is a switch in the states, it depends on the difference between the variances of the new and old states. If there is no switch then $d\mathbf{p}(t)$ is the zero vector; otherwise it will have the values –1 and 1 for the jth and ith elements, respectively, if we switch from state j to state i.

This way the size of the jump in the volatility process (i.e. the square root of the variance process) will vary stochastically with time, as depicted in Figure 1 for the case K = 2. Each state has its own long-term volatility. Following a jump from the low volatility state 1 to the high volatility state 2 the individual volatility in state 1 generally increases.

Figure 1: Realised Volatility Jumps between 2 States
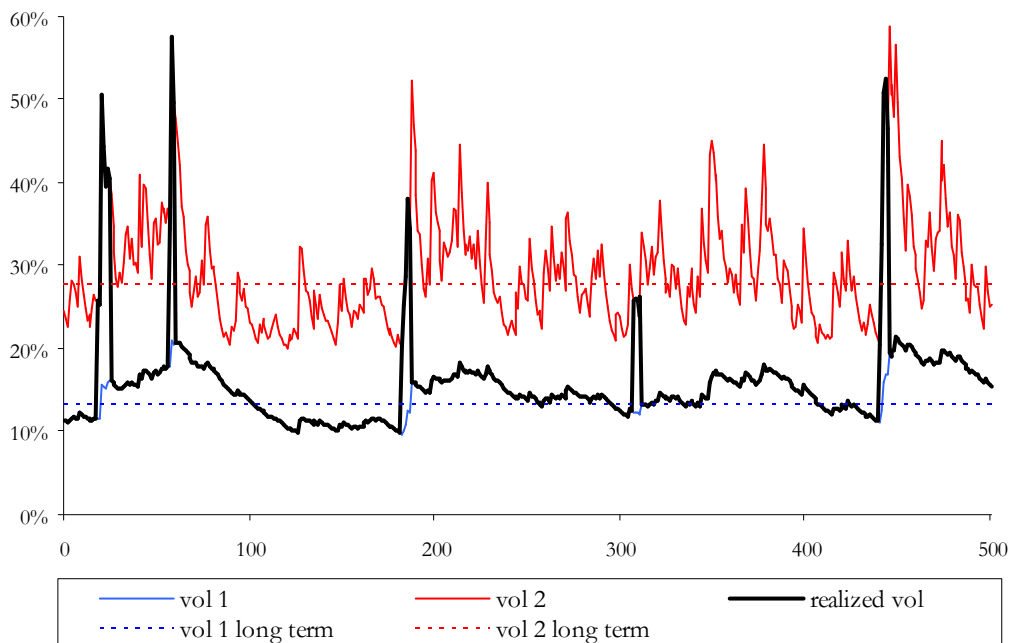
Note that the drift in this model changes at the same time as the volatility jumps because the first equation in (45) can be rewritten as (this should not be confused with a jump in the price):

$$\frac{dS(t)}{S(t)} = \mu(t)dt + \sigma(t)dB \text{ with } d\mu(t) = d\mathbf{p}(t)'\boldsymbol{\mu}$$

The jump times are stochastic, being governed by the state process (a continuous Markov chain), the jump probabilities are given by the conditional expectation of the state vector, and the state probability dynamics are given by the last equation. On setting $\overline{h}_i = \lim_{\Delta \downarrow 0} {}_\Delta y_i$ with

$${}_\Delta y_i = E\left[{}_\Delta h_{i,t}\right] = \left(1 - {}_\Delta\beta_i\right)^{-1}\left({}_\Delta\omega_i + {}_\Delta\alpha_i \, {}_\Delta\overline{s}^2\right) = \left[\Delta^{-1}\left(1 - {}_\Delta\beta_i\right)\right]^{-1}\left(\Delta^{-1}{}_\Delta\omega_i + \Delta^{-1}{}_\Delta\alpha_i \, {}_\Delta\overline{s}^2\right) \text{ and}$$

$${}_\Delta\overline{s}^2 = E\left[{}_\Delta\mathrm{e}_{i,t}^2 / \Delta\right]$$

we see that each of the deterministic GARCH components converges to a different constant steady state forward variance, given by:

$$\overline{h}_i = E\left[h_i(t)\right] = \psi_i^{-1}\left(\omega_i + \alpha_i\overline{s}^2\right)$$

where ψ_i is the speed of mean reversion and $\overline{s}^2 = \lim_{\Delta \downarrow 0} {}_\Delta\overline{s}^2 = E\left[\sum_{i=1}^{K}\pi_i\left(h_i(t) + \mu_i^2\right)\right] = \sum_{i=1}^{K}\pi_i\left(\overline{h}_i + \mu_i^2\right)$

gives the unconditional variance. We get that:

$$\overline{s}^2 = \left(1 - \sum_{i=1}^{K}\psi_i^{-1}\pi_i\alpha_i\right)^{-1}\sum_{i=1}^{K}\pi_i\left(\psi_i^{-1}\omega_i + \mu_i^2\right)$$

For the normal mixture AGARCH model, the speed of mean-reversion remains unchanged but the unconditional variance to which the component variances mean-revert becomes:

$$\overline{h}_i = \psi_i^{-1}\left(\omega_i + \alpha_i\varphi_i^2 + \alpha_i\overline{s}^2\right)$$

where

$$\overline{s}^2 = \left(1 - \sum_{i=1}^{K}\psi_i^{-1}\pi_i\alpha_i\right)^{-1}\sum_{i=1}^{K}\pi_i\left(\psi_i^{-1}\omega_i + \psi_i^{-1}\alpha_i\varphi_i^2 + \mu_i^2\right)$$

For the normal mixture GJR model, since the continuous model is the same as for the GARCH, the above properties remain the same.

We conclude that the continuous MS GARCH model has a rich and flexible structure, admitting state-dependent mean-reversion through different volatility components and capturing long-term smile and skew effects through state-dependent volatility and drift and volatility, respectively.

IV.2   Hedging State Uncertainty

In this section we examine the price and variance risk premia in the continuous MS GARCH model, and in the case when there are only two states we quantify the hedge ratios for state uncertainty and specify the dynamics of a contingent claim.

Proposition 3: The state-dependent drift has no effect on the price risk premium.

However, the continuous MS GARCH model does not have the standard volatility risk premium that arises in models where a second Brownian motion drives the volatility and which represents the additional volatility drift under the real-world measure compared with the risk-free measure. The volatility risk-premium in the continuous MS GARCH model has a different interpretation since the volatility process has no diffusion, only a jump process. The source of uncertainty is the state process, so we should call this a state risk premium rather than a volatility risk premium.

Denote a contingent claim with price at time t by $f(t) = f(t, S(t), V(t))$ or, for brevity,

$$f = f(t, S, V)$$

as it should be understood that the claim price, the underlying price S and its variance V are all time dependent. Since the jump in the variance process is independent of the Brownian motion of the price process, Ito's lemma gives:

$$df = f_t dt + f_S dS + 0.5 f_{SS} S^2 V dt + f_V dV + \left[f - f^- \right]$$

where subscripts denote partial derivatives and the last term is the change in the derivative's value due to the jump in the variance process; in the following this will be denoted by df^J. Given the underlying process specified by (45) the claim price dynamics can be further expressed as:

(46)
$$df = \xi dt + f_S \sqrt{V} S dB + df^J$$
$$\xi = f_t + f_S \mathbf{p}'\mathbf{\mu} S + f_V \left(\mathbf{p}'(\mathbf{\omega} + \mathbf{\alpha} V - \mathbf{\phi} \odot \mathbf{V}) + d\mathbf{p}'\mathbf{V} \right) + 0.5 f_{SS} V S^2$$

Not surprisingly, it is impossible to replicate the claim price with positions only in the underlying only. The risk of volatility jumps cannot be hedged with the underlying alone, due to the independence of the Brownian and the jumps. For simplicity consider a two-state world and take two claims with prices f_1 and f_2.

Proposition 4: The state uncertainty hedged portfolio Π that is long one unit of the underlying and short A_1 and A_2 units of the two claims has price:

$$\Pi = S - A_1 f_1 - A_2 f_2$$

where the hedge ratios are given by:

(47)
$$A_1 = df^{J_2} \left(\Delta_1 df^{J_2} - \Delta_2 df^{J_1} \right)^{-1} \text{ and } A_2 = df^{J_1} \left(\Delta_2 df^{J_1} - \Delta_1 df^{J_2} \right)^{-1}$$

Note that perfect hedging with the underlying and a second derivative is possible only in two-state models. In this case we can be certain about the size of jump in the variance process even if the time of the jump is not known. If we have more than two states then it is not just the timing of the jump that is uncertain, but the size of the jumps as well. Hence the change in the value of the claim explained by the jump would depend on the new state. This way it can be seen that additional claims will be needed for hedging (indeed the number of claims needed for hedging is $K-1$, where K is the number of states). With two states only, the new state is always the 'other' state, so hedging can be done using the underlying and a second derivative.

In practice to hedge a long call in a two-state model, $1/A_1$ units of the underlying need to be sold short and A_2/A_1 calls (of different type than the original) need to be shorted or puts bought. The hedge ratios (47) can be reformulated as:

$$A_1 = \left(\Delta_1 - f_{12}\Delta_2 \right)^{-1} \text{ and } A_2 = \left(\Delta_2 - f_{12}^{-1}\Delta_1 \right)^{-1}$$

where $f_{12} = \dfrac{df^{J_1}}{df^{J_2}}$.

Hence the state uncertainty leads to hedge ratios that are not equal to one over delta in each state. Instead the delta in each state is augmented by a multiple of the delta in the other state. The multiple f_{12} is the change in value of the claim due to the variance jump when jumping from state 1 to state 2, relative to the value change when jumping from state 2 to state 1. Since the value change is positive when jumping from state 1 to state 2, but negative when jumping from state 2 to state 1, $f_{12} < 0$.

Theorem 3: In the continuous MS GARCH model a claim with price $f(t)$ has the following dynamics:

$$df = \left(r(f - Sf_S) - \lambda_s df^J + f_S \boldsymbol{p}'\boldsymbol{\mu}S \right)dt + f_S\sqrt{V}SdB + df^J$$

where

$$\lambda_s = \left(f_{i,t} + f_{i,V}\left(\boldsymbol{p}'(\boldsymbol{\omega} + \boldsymbol{\alpha}V - \boldsymbol{\psi}\odot\mathbf{V}) + d\boldsymbol{p}'\mathbf{V} \right) + 0.5f_{i,SS}\sigma V S^2 + r\left(Sf_{i,S} - f_i \right) \right)\left(df^{J_i} \right)^{-1} \quad i = 1,2$$

is the 'market price of state risk'. This way, the market is arbitrage-free but incomplete.

There are two sources of risk: the Brownian motion in the return and the state variable. Therefore the market is incomplete, and so there are no unique prices for the options. The state risk is not diversifiable and thus it is priced. We define the risk-neutral measure Q with:

(48)

$$\frac{dQ}{dP} = \exp\left(-\int_0^t \frac{\boldsymbol{p}(s)'\boldsymbol{\mu}-r}{\sigma(s)} dB(s) - \frac{1}{2}\int_0^t \left(\frac{\boldsymbol{p}(s)'\boldsymbol{\mu}-r}{\sigma(s)} \right)^2 ds - \int_0^t \lambda(s)\eta(s)ds + \sum_{0<s\le t} \ln\left(1+\eta(s)\right)\Delta N(s) \right)$$

with respect to which B^* will be a Brownian motion. Then, under the risk-neutral measure the underlying process can be rewritten as:

(49)

$$\frac{dS}{S} = rdt + \sqrt{V}dB^*$$
$$dV = \boldsymbol{p}'d\mathbf{V} + dJ \quad with \quad dJ = d\boldsymbol{p}'\mathbf{V}$$
$$d\mathbf{V} = \left(\boldsymbol{\omega}+\boldsymbol{\alpha}V - \boldsymbol{\psi}\odot\mathbf{V}\right)dt$$
$$d\tilde{\boldsymbol{p}} = \boldsymbol{\Lambda}^*\tilde{\boldsymbol{p}}dt$$

where $\boldsymbol{\Lambda}^* = \boldsymbol{\Lambda}\odot\left(\mathbf{1}(1+\boldsymbol{\eta})'\right)$, $\mathbf{1}$ is a vector of ones and $\eta(t) = \boldsymbol{p}(t)'\boldsymbol{\eta}$, $\lambda(t) = \boldsymbol{p}(t)'\boldsymbol{\lambda}$ with $\boldsymbol{\eta} = (\eta_1, \eta_2)'$ and $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)'$. The option pricing can be based on either a non-recombining tree or simulation for the variance process.

IV.3 Discretizing the Continuous MS GARCH Model

When discretizing this process, first we substitute the continuous time model with a discrete time one where post-subscript denotes time and we approximate ω by $_h\omega h^{-1}$, α by $_h\alpha h^{-1}$ and ψ by $\left(1-_h\beta\right)h^{-1}$. All differences are forward differences. One important approximation has to be made. We know that

$$\left(dB(t)\right)^2 = dt \quad so \quad \left(B(t+h)-B(t)\right)^2 \approx h.$$

Also, $_h\varepsilon_{t+h} = \frac{S_{t+h}-S_t}{S_t} - \mu h = {}_h\sigma_t\left(B(t+h)-B(t)\right)$ so $\sigma^2(t)dt$ can be approximated by:

$$_h\sigma_t^2 h \approx {}_h\sigma_t^2\left(B(t+h)-B(t)\right)^2 = \left({}_h\sigma_t\left(B(t+h)-B(t)\right)\right)^2 = {}_h\varepsilon_{t+h}^2$$

This leads to the following discretization:

$$\frac{S_{t+h}-S_t}{S_t} = \mu h + {}_h\sigma_t\left(B(t+h)-B(t)\right)$$
$$_h\sigma_{t+h}^2 - {}_h\sigma_t^2 = \omega h + \alpha\,{}_h\sigma_t^2 h - \left(1-\beta_h\right)\sigma_t^2 h$$

which can be rewritten as:

$$\frac{S_{t+h}-S_t}{S_t} = \mu h + {}_h\varepsilon_{t+h}$$
$$_h\sigma_{t+h}^2 = \omega + \alpha\,{}_h\varepsilon_{t+h}^2 h^{-1} + \beta\,{}_h\sigma_t^2$$

which is the same as the original GARCH.

V   SUMMARY AND CONCLUSIONS

We consider the continuous time version of the MS GARCH of Haas, Mittnik and Paolella (2004b). For this model, with both symmetric and asymmetric GARCH variance components, showed that the continuous time analogue is a stochastic model where the variance jumps between several time-varying processes, but it has no diffusion. Also, we derive the state risk premium and hedge ratios for contingent claims priced with this model.

Stochastic volatility models where the variance follows a diffusion are considered to be successful tools for option pricing. Some might say that much is lost by not having a diffusion in the variance process in our limit model; however, much is gained by having a multi-state variance process. In discrete time GARCH the squared residuals introduce uncertainty to the variance, a stochastic feature that is, sadly, lost in continuous time when multiple states are considered. The reason for this leads to the limit of stochastic modelling, namely that, with some exceptions, Brownian motions and jump processes are the only basic processes used in financial modelling – and these prove to be not enough to keep the stochasticity of discrete time GARCH models in continuous time.

In line with the results of Corradi (2000), Wang(2002), Brown, Wang and Zhao (2002) and Duan et al. (2005b) we have made assumptions under which the normal GARCH process can be extended to continuous time. Nelson's (1990) result that the limit is a stochastic volatility model is more intuitive; however, his assumptions cannot be generalized to GARCH models with more than one volatility state, whilst it is recognised that such models are required for explaining the empirical behaviour of asset returns and implied volatility. Our assumptions, that both α/h and $(1 - \beta)/h$ converge as the time-step goes to zero, are also more intuitive than assuming that only their sum and α^2/h converge and they are essential for the existence of the limit of multi-state GARCH models. Since the strong versions of MS GARCH are not aggregating in time, we have given and used the weak definition that is not sensitive to the choice of step length.

Also, we showed that the normal mixture GARCH model, which is a reduced form of MS GARCH with state-independent conditional probabilities, has no continuous-time limit. This model has infinite variation in the variance, and no tools are available yet to model such processes.

There is a growing literature on models that have jumps in the volatility process. These have been shown to offer a good time-series and option pricing fit. Some of the problems with the models considered in the literature are that they do not model jump clustering and they do not allow for different speeds of mean-reversions at different levels of volatility (Duan, Ritchken and Sun, 2005a), but our framework allows for both these features.

Option pricing with normal mixture and Markov Switching GARCH processes is likely to be an important area for future research. These processes can be consistent with a volatility surface that has smile and skew effects that change with maturity but still persist longer than the central limit theorem would imply, as the conditional skewness and kurtosis are time varying and the unconditional skewness and kurtosis is non-zero. Perhaps the most important feature of the model is that is captures regime dependent behaviour of the volatility surface, with different mean reversion and leverage effects in different market regimes. Very few other volatility models are able to capture such intricacies in the behaviour of implied volatility.

<div align="center">APPENDIX</div>

Proof of Theorem 1

We present the proof for $n = 2$. The proof for general n-period time intervals follows by induction.

From (4) we get that:

$$_1 h_{i,2k+1} = {_1}\omega_i + \left(z_i^{-1}\,{_1}\alpha_i + {_1}\beta_i \right) z_i\,{_1}\varepsilon_{2k+1}^2 - {_1}\beta_i\,{_1}\eta_{i,2k+1}$$

This can be rearranged as:

$$z_i\,{_1}\varepsilon_{2k+2}^2 = {_1}\omega_i\left(1 + z_i^{-1}\,{_1}\alpha_i + {_1}\beta_i\right) + \left(z_i^{-1}\,{_1}\alpha_i + {_1}\beta_i\right)^2 z_i\,{_1}\varepsilon_{2k}^2 + {_1}\eta_{i,2k+2} + \left(z_i^{-1}\,{_1}\alpha_i\right){_1}\eta_{i,2k+1} - {_1}\beta_i\left(z_i^{-1}\,{_1}\alpha_i + {_1}\beta_i\right){_1}\eta_{i,2k}$$

Repeating this equation for time step $2k+1$ and summing yields:

$$z_i\,{_2}\varepsilon_{2(k+1)}^2 = 2\left({_1}\omega_i\left(1 + z_i^{-1}\,{_1}\alpha_i + {_1}\beta_i\right) + \left(z_i^{-1}\,{_1}\alpha_i + {_1}\beta_i\right)^2 z_i\,{_2}\varepsilon_{2k}^2 \big/ 2 + z_i\,{_1}\varepsilon_{2k+2}\,{_1}\varepsilon_{2k+1} - z_i\left(z_i^{-1}\,{_1}\alpha_i + {_1}\beta_i\right)^2 {_1}\varepsilon_{2k}\,{_1}\varepsilon_{2k-1}\right)$$

$$+ \left({_1}\eta_{i,2k+2} + \left(1 + z_i^{-1}\,{_1}\alpha_i\right){_1}\eta_{i,2k+1} + \left(z_i^{-1}\,{_1}\alpha_i - {_1}\beta_i\left(z_i^{-1}\,{_1}\alpha_i + {_1}\beta_i\right)\right){_1}\eta_{i,2k} - {_1}\beta_i\left(z_i^{-1}\,{_1}\alpha_i + {_1}\beta_i\right){_1}\eta_{i,2k-1}\right)$$

where we have used the following relationship for returns:

$$_2\varepsilon_{2(k+1)}^2 = \left({_1}\varepsilon_{2k+2} + {_1}\varepsilon_{2k+1}\right)^2 = {_1}\varepsilon_{2k+2}^2 + {_1}\varepsilon_{2k+1}^2 + 2\,{_1}\varepsilon_{2k+2}\,{_1}\varepsilon_{2k+1}$$

Letting

$$v_{i,2(k+1)} = 2 z_i\,{_1}\varepsilon_{2k+2}\,{_1}\varepsilon_{2k+1} - 2 z_i\left(z_i^{-1}\,{_1}\alpha_i + {_1}\beta_i\right)^2 {_1}\varepsilon_{2k}\,{_1}\varepsilon_{2k-1}$$

$$+ \left({_1}\eta_{i,2k+2} + \left(1 + z_i^{-1}\,{_1}\alpha_i\right){_1}\eta_{i,2k+1} + \left(z_i^{-1}\,{_1}\alpha_i - {_1}\beta_i\left(z_i^{-1}\,{_1}\alpha_i + {_1}\beta_i\right)\right){_1}\eta_{i,2k} - {_1}\beta_i\left(z_i^{-1}\,{_1}\alpha_i + {_1}\beta_i\right){_1}\eta_{i,2k-1}\right)$$

we can now write

$$z_i\,{_2}\varepsilon_{2(k+1)}^2 = 2\,{_1}\omega_i\left(1 + z_i^{-1}\,{_1}\alpha_i + {_1}\beta_i\right) + \left(z_i^{-1}\,{_1}\alpha_i + {_1}\beta_i\right)^2 z_i\,{_2}\varepsilon_{kh}^2 + v_{i,2(k+1)}$$

and it can be shown that:

$$E\left(v_{i,2k}\right) = 0$$

(50)

$$E\left(v_{i,2(k+1)} v_{i,2(k+1-l)}\right) = 0 \quad \text{for } l > 1$$

$$Corr\left(v_{i,2(k+1)}, v_{i,2k}\right) = \frac{-\lambda_i}{1 + \lambda_i^2}$$

for some $\lambda_i \in (0, 1)$. We now define $w_{i,2(k+1)}$ by:

(51)
$$w_{i0} = v_{i0} \quad \text{and} \quad w_{i,2(k+1)} = v_{i,2(k+1)} + \lambda_i\, w_{i,2k}$$

Obviously $E\left(w_{i,2k}\right) = 0$ and it can be shown that

$$Corr\left(w_{i,2(k+1)}, w_{i,2(k+1-l)}\right) = 0 \ \text{ for } l \geq 1.$$

Rewriting (51) as

$$v_{i,2(k+1)} = w_{i,2(k+1)} - \lambda_i w_{i,2k}$$

we have:

$$(52) \quad z_{i\ 2}\varepsilon^2_{2(k+1)} - w_{i,2(k+1)} = 2\omega_i\left(1 + z_i^{-1}{}_1\alpha_i + {}_1\beta_i\right) + \left(\left(z_i^{-1}{}_1\alpha_i + {}_1\beta_i\right)^2 - \lambda_i\right)z_{i\ 2}\varepsilon^2_{2k} + 2\lambda_i z_{i\ 2}\varepsilon^2_{2k} - \lambda_i w_{i,2k}$$

Now we define

$$_2h_{i,2k} = z_{i\ 2}\varepsilon^2_{2(k+1)} - w_{i,2(k+1)}$$

Hence (52) becomes:

$$_2h_{i,2k} = {}_1\omega_i\left(1 + z_i^{-1}{}_1\alpha_i + {}_1\beta_i\right) + \left(\left(z_i^{-1}{}_1\alpha_i + {}_1\beta_i\right)^2 - \lambda_i\right)z_{i\ 2}\varepsilon^2_{2k} + \lambda_{i\ 2}h_{i,2(k-1)}$$

and this is the updating formula for the new two-period conditional variances of each component. Hence when $n = 2$ the new parameters are

$$_2\omega_i = {}_1\omega_i\left(1 + z_i^{-1}{}_1\alpha_i + {}_1\beta_i\right)$$

$$_2\alpha_i = z_i\left(\left(z_i^{-1}{}_1\alpha_i + {}_1\beta_i\right)^2 - {}_2\beta_i\right)$$

and $_2\beta_i \in (0,1)$ is the solution to $Corr\left(v_{i,t}, v_{i,t-2}\right) = \dfrac{-{}_2\beta_i}{1 + {}_2\beta_i^2}$.

In general, after annualisation, the parameters of $_nh_{i,t}$, for n > 1 integer, will be given by:

$$_n\omega_i = {}_1\omega_i\left(1 - \left(z_i^{-1}{}_1\alpha_i + {}_1\beta_i\right)^n\right)\left(1 - z_i^{-1}{}_1\alpha_i - {}_1\beta_i\right)^{-1}$$

$$_n\alpha_i = z_i\left(\left(z_i^{-1}{}_1\alpha_i + {}_1\beta_i\right)^n - {}_n\beta_i\right)$$

and $_n\beta_i \in (0,1)$ is the solution to $Corr\left(v_{i,t}, v_{i,t-n}\right) = \dfrac{-{}_n\beta_i}{1 + {}_n\beta_i^2}$ where

$$v_{i,t} = \left(\sum_{p=0}^{2n-1}\sum_{k=\max(0,p+1-n)}^{\min(p,n)} c_k\eta_{i,t-p}\right) / n + 2z_i\left(\sum_{0 \le r < s < n} {}_1\varepsilon_{t-r}\,{}_1\varepsilon_{t-s} - \left(z_i^{-1}{}_1\alpha_i + {}_1\beta_i\right)^n \sum_{0 \le r < s < n} {}_1\varepsilon_{t-n-r}\,{}_1\varepsilon_{t-n-s}\right) / n$$

and

$$c_k = \begin{cases} 1 & \text{if } k = 0 \\ z_i^{-1}{}_1\alpha_i\left(z_i^{-1}{}_1\alpha_i + {}_1\beta_i\right)^{k-1} & \text{if } 1 \le k \le n-1 \\ -{}_1\beta_i\left(z_i^{-1}{}_1\alpha_i + {}_1\beta_i\right)^{n-1} & \text{if } k = n \end{cases}$$

Also, it follows trivially that: $_n\mu_i = {}_1\mu_i$, $_n\pi_i = {}_1\pi_i$ and $_n\mathbf{Q} = {}_1\mathbf{Q}^n$ ☐

Proof of Proposition 1

It can be seen that $\left(z_i^{-1}{}_\Delta\alpha_i + {}_\Delta\beta_i\right)^{1/\Delta}$ is a constant between 0 and 1, so we denote this by $e^{-\theta_i}$, $\theta_i > 0$.

Therefore $z_i^{-1}{}_\Delta\alpha_i + {}_\Delta\beta_i = e^{-\theta_i\Delta}$ and so:

$$\lim_{\Delta \downarrow 0} \Delta^{-1} \left(1 - \left(z_i^{-1} {}_\Delta \alpha_i + {}_\Delta \beta_i \right) \right) = \lim_{\Delta \downarrow 0} \Delta^{-1} \left(1 - e^{-\theta_i \Delta} \right) = \theta_i$$

Also, we have that ${}_\Delta \omega_i \left(1 - e^{-\theta_i \Delta} \right)^{-1}$ is a positive constant that we denote by $\omega_i \theta_i^{-1}$ where $\omega_i > 0$, so that

${}_\Delta \omega_i = \omega_i \left(1 - e^{-\theta_i \Delta} \right) \theta_i^{-1}$. As a result:

$$\lim_{\Delta \downarrow 0} \Delta^{-1} {}_\Delta \omega_i = \omega_i \lim_{\Delta \downarrow 0} \Delta^{-1} \left(1 - e^{-\theta_i \Delta} \right) \theta_i^{-1} = \omega_i \qquad \square$$

Proof of Proposition 2

We prove that the variance of the limiting model does not satisfy the stochastic continuity condition, i.e. we show that:

$$\forall t \; \exists \; \eta \ni \lim_{l \downarrow 0} P \left(\left| V(t+l) - V(t) \right| \geq \eta \right) > 0$$

To prove this assume that $K = 2$, the unconditional individual variances are $\overline{h}_1$ and $\overline{h}_2$ and the mixing law is $(\pi, 1 - \pi)'$ with $0 < \pi < 1$. That is, if we are in state 1 the probability of a jump in the volatility is $1 - \pi$ and if we are in state 2 the probability of a jump in the volatility is π. At a given point in time t, not knowing the ruling state, the probability of a switch in the state is $2\pi(1 - \pi)$. The expected size of the jump over an arbitrary time interval l is $\left| \overline{h}_2 - V_1(t) \right|$ if in state 1 and $\left| \overline{h}_1 - V_2(t) \right|$ if in state 2.

Let $\eta = \min \left\{ \left| \overline{h}_2 - V_1(t) \right|, \left| \overline{h}_1 - V_2(t) \right| \right\}$. Then

$$P \left(\left| V(t+l) - V(t) \right| \geq \eta \right) = 2\pi (1 - \pi).$$

The above probability is the same for any l so its limit as the step length converges to zero is $2\pi(1 - \pi)$, which is strictly positive. Hence the stochastic continuity condition for a Lévy process is not satisfied for the variance, and we conclude that the continuous-time limit of the variance process of the NM-GARCH does not exist.

It can be seen that the variation and quadratic variation of the variance process is infinite not just when the classical definition of convergence is used, but also when the limit in probability is applied. Consider the quadratic variation of the variance using probability limits.[8] Take an arbitrary time interval $[a, b]$ and a step length Δ. Construct a partition for this interval $\left({}_\Delta t_i \right)_{i=0,\ldots,{}_\Delta n}$, ${}_\Delta t_i = a + i\Delta$ where

${}_\Delta n = \lfloor (b-a) / \Delta \rfloor$ is the number of steps in the interval. Define the quadratic variation as:

$$V_h \left([a,b] \right) = p \lim_{\Delta \downarrow 0} {}_\Delta S \; \text{with} \; {}_\Delta S = \sum_{i=1}^{{}_\Delta n} \left| V \left({}_\Delta t_i \right) - V \left({}_\Delta t_{i-1} \right) \right|^2$$

Then

---

[8] This is important since Lévy processes have finite quadratic variation when convergence in probability is used.

$$E\left(\left| V\left(_\Delta t_i \right) - V\left(_\Delta t_{i-1} \right) \right|^2 \right) > 2\pi \left(1-\pi \right) \left| \overline{h}_1 - \overline{h}_2 \right|^2$$

$$E\left(S_\Delta \right) > 2 _\Delta n \pi \left(1-\pi \right) \left| \overline{h}_1 - \overline{h}_2 \right|^2$$

The variance of $\left| V\left(_\Delta t_i \right) - V\left(_\Delta t_{i-1} \right) \right|^2$ is finite. Also,

$$Var\left(_\Delta S \right) = {}_\Delta n \; Var\left(\left| V\left(_\Delta t_i \right) - V\left(_\Delta t_{i-1} \right) \right|^2 \right)$$

We have by the central limit theorem:

$$_\Delta Z = \left(_\Delta S - E\left(_\Delta S \right) \right) / \sqrt{Var\left(_\Delta S \right)} \overset{a}{\sim} N\left(0,1 \right)$$

where the normal distribution is an approximation. For an $M > 0$ we have that:

$$\mathrm{P}\left(_\Delta S > M \right) = \mathrm{P}\left(_\Delta Z > \left(M - E\left(S_\Delta \right) \right) / \sqrt{Var\left(S_\Delta \right)} \right)$$

$$\geq \mathrm{P}\left(_\Delta Z > \left(M - 2 _\Delta n \pi \left(1-\pi \right) \left| \overline{h}_1 - \overline{h}_2 \right|^2 \right) / \sqrt{_\Delta n \; Var \left| V\left(_\Delta t_i \right) - V\left(_\Delta t_{i-1} \right) \right|^2} \right)$$

$$\rightarrow \mathrm{P}\left(_\Delta Z > -\infty \right) = 1$$

Thus, we again obtain that the variance of the limit of normal mixture GARCH models is not a Lévy process. □

Proof of Theorem 2

Consider first the returns process. For the drift per unit time we have:

$$E\left(\Delta^{-1} \left(\frac{S_{(k+1)\Delta} - S_{k\Delta}}{S_{k\Delta}} \right) \middle| I_{k\Delta}, {}_\Delta s_{k\Delta} = i \right) = \mu + E\left(\Delta^{-1} {}_\Delta \epsilon_{(k+1)\Delta} \mid I_{k\Delta}, {}_\Delta s_{k\Delta} = i \right) = \mu + {}_\Delta \mu_{i,k\Delta}$$

and the square drift per unit time is

$$E\left(\Delta^{-1} \left(\frac{S_{(k+1)\Delta} - S_{k\Delta}}{S_{k\Delta}} \right)^2 \middle| I_{k\Delta}, {}_\Delta s_{k\Delta} = i \right) = E\left(\Delta^{-1} \left(\Delta\mu + {}_\Delta \epsilon_{(k+1)\Delta} \right)^2 \middle| I_{k\Delta}, {}_\Delta s_{k\Delta} = i \right)$$

$$= E\left(\Delta\mu^2 + \Delta^{-1} {}_\Delta \epsilon_{(k+1)\Delta}^2 + 2\mu {}_\Delta \epsilon_{(k+1)\Delta} \middle| I_{k\Delta}, {}_\Delta s_{k\Delta} = i \right) = E\left(\Delta^{-1} {}_\Delta \epsilon_{(k+1)\Delta}^2 \middle| I_{k\Delta}, {}_\Delta s_{k\Delta} = i \right) + \mathrm{o}\left(1 \right)$$

$$= {}_\Delta \sigma_{i,k\Delta}^2 + \Delta {}_\Delta \mu_{i,k\Delta}^2 + \mathrm{o}\left(1 \right) = {}_\Delta h_{i,k\Delta} + \left({}_\Delta \sigma_{i,k\Delta}^2 - {}_\Delta h_{i,k\Delta} \right) + \mathrm{o}\left(1 \right) = {}_\Delta h_{i,k\Delta} + \mathrm{o}\left(1 \right)$$

so the conditional first and second moments per unit time converge to $\mu_i\left(t \right)$ and $V_i\left(t \right)$ if $s\left(t \right) = i$ as $\Delta \downarrow 0$. For the state-variances we can write:

$$E\left(\Delta^{-1} \left(_\Delta h_{i,(k+1)\Delta} - {}_\Delta h_{i,k\Delta} \right) \middle| I_{kh}, {}_\Delta s_{k\Delta} = j \right)$$

$$= \Delta^{-1} {}_\Delta \omega_i + \Delta^{-1} {}_\Delta \alpha_i {}_\Delta h_{j,k\Delta} + \Delta^{-1} \left({}_\Delta \beta_i - 1 \right) {}_\Delta h_{i,k\Delta} + \mathrm{o}\left(1 \right)$$

We need the limit to exist and be finite when $h \to 0$ for $i = 1, \ldots, K$ and this forces us to have separate limits for the α and β parameters. The variances and covariances of the individual variance components converge to zero:

$$E\left(\Delta^{-1}\left({}_{\Delta}h_{i,(k+1)\Delta} - {}_{\Delta}h_{i,k\Delta}\right)\left({}_{\Delta}h_{j,(k+1)\Delta} - {}_{\Delta}h_{j,k\Delta}\right)\Big| I_{k\Delta}, {}_{\Delta}s_{k\Delta} = l\right) = o(1)$$

The covariance between the returns and the changes in the variances converges as follows:

$$E\left(\Delta^{-1}\left(\frac{S_{(k+1)\Delta} - S_{k\Delta}}{S_{k\Delta}}\right)\left({}_{\Delta}h_{i,(k+1)\Delta} - {}_{\Delta}h_{i,k\Delta}\right)\Bigg| I_{k\Delta}, {}_{\Delta}s_{k\Delta} = j\right)$$

$$= E\left(\Delta^{-1}\left(\Delta\mu + \varepsilon_{(k+1)\Delta}\right)\left({}_{\Delta}\omega + \Delta^{-1}{}_{\Delta}\alpha\,{}_{\Delta}\varepsilon^2_{(k+1)\Delta} + \left({}_{\Delta}\beta - 1\right){}_{\Delta}h_{i,k\Delta}\right)\Big| I_{k\Delta}, {}_{\Delta}s_{k\Delta} = j\right) = o(1)$$

Finally, to derive the process for the conditional probabilities we have:

$$E\left(\Delta^{-1}\left({}_{\Delta}\tilde{\mathbf{P}}_{(k+1)\Delta} - {}_{\Delta}\tilde{\mathbf{P}}_{k\Delta}\right)\Big| I_{kh}\right) = \Delta^{-1}\left({}_{\Delta}\tilde{\mathbf{P}}_{(k+1)\Delta} - {}_{\Delta}\tilde{\mathbf{P}}_{k\Delta}\right) = \Delta^{-1}\left({}_{\Delta}\mathbf{Q} - \mathbf{I}\right){}_{\Delta}\tilde{\mathbf{P}}_{k\Delta}$$

$$E\left(\Delta^{-1}\left({}_{\Delta}P_{i,(k+1)\Delta} - {}_{\Delta}P_{i,k\Delta}\right)\left({}_{\Delta}P_{j,(k+1)\Delta} - {}_{\Delta}P_{j,k\Delta}\right)\Big| I_{k_\Delta}\right) =$$

$$= \Delta\left(\Delta^{-1}\left({}_{\Delta}\mathbf{Q}_i - e_i\right) - \mathbf{\Lambda}_i\right)\left(\Delta^{-1}\left({}_{\Delta}\mathbf{Q}_j - e_j\right) - \mathbf{\Lambda}_j\right){}_{\Delta}\tilde{P}_{i,k\Delta}\,{}_{\Delta}\tilde{P}_{j,k\Delta} + o(1) = o(1)$$

where e_i is the ith row of the identity matrix $\mathbf{I}$ and the subscript i for $\mathbf{\Lambda}$ symbolizes the ith row of the transition rate matrix. □

Proof of Proposition 3

The price risk premium, λ_e is such that:

$$\exp(-rt)E(S(t)) = \exp(\lambda_e t)S(0)$$

and for the continuous MS GARCH model:

(53)
$$S(t) = S(0)\exp\left(\int_0^t \mathbf{p}(s)'\mu\, ds - \frac{1}{2}\int_0^t V(s)\, ds + \int_0^t \sqrt{V(s)}\, dB(s)\right)$$

Now

$$E\left(\exp\left(\int_0^t \mathbf{p}(s)'\mu\, ds\right)\right) = \exp\left(E\left(\int_0^t \mathbf{p}(s)'\mu\, ds\right)\right) = \exp\left(\int_0^t \pi'\mu\, ds\right) = \exp\left(\pi'\mu t\right) = \exp(\mu t)$$

since π, the unconditional expectation of $\mathbf{p}$, also determines the realized frequency of the regimes. But

$$E\left[\exp\left(\int_0^t \sigma(s)\, dB(s)\right)\right] = \exp\left(0.5\int_0^t \pi'\overline{\mathbf{h}}\, ds\right) = \exp\left(0.5\,\pi'\overline{\mathbf{h}}\, t\right)$$

and:

$$E\left[\exp\left(\int_0^t \sigma^2(s)\, ds\right)\right] = \exp\left(\int_0^t \pi'\overline{\mathbf{h}}\, ds\right) = \exp\left(\pi'\overline{\mathbf{h}}\, t\right)$$

so that $E(S(t)) = S(0)\exp(\mu t)$. Therefore $\lambda_e = \mu - r$ as in standard geometric Brownian motion.

Proof of Proposition 4

We have:

$$d\Pi = dS - A_1 df_1 - A_2 df_2$$

This can be further expressed as:

$$d\Pi = \zeta dt + \left(1 - \left(A_1 f_{1,S} + A_2 f_{2,S}\right)\right)\sqrt{V}S dB - A_1 df^{J_1} - A_2 df^{J_2}$$

where

(54)
$$\begin{aligned}
\zeta = &-\left(A_1 f_{1,t} + A_2 f_{2,t}\right) + \left(1 - \left(A_1 f_{1,S} + A_2 f_{2,S}\right)\right)\mathbf{p}'\boldsymbol{\mu}S \\
&-\left(A_1 f_{1,V} + A_2 f_{2,V}\right)\left(\mathbf{p}'\left(\boldsymbol{\omega} + \boldsymbol{\alpha}V - \boldsymbol{\psi}\odot\mathbf{V}\right) + d\mathbf{p}'\mathbf{V}\right) - 0.5\left(A_1 f_{1,SS} + A_2 f_{2,SS}\right)VS^2
\end{aligned}$$

This portfolio is risk-free if, in addition to the term in *dB* being zero, the change in portfolio value does not depend on the state. Hence the following two conditions must hold:

$$A_1\Delta_1 + A_2\Delta_2 = 1$$
$$A_1 df^{J_1} + A_2 df^{J_2} = 0$$
$$\text{where } \Delta_i = f_{i,S}, i = 1,2$$

and solving for A_1 and A_2 gives the result.

Proof of Theorem 3

To avoid arbitrage opportunities, the following must be satisfied:

$$d\Pi = r\Pi dt$$

That is,

$$\zeta dt = r\left(S - A_1 f_1 - A_2 f_2\right)dt$$

Using (54) this can be further expressed as:

$$\left(\left(df^{J_1}\right)^{-1}f_{1,t} - \left(df^{J_2}\right)^{-1}f_{2,t}\right) - \left(\left(df^{J_1}\right)^{-1}\Delta_1 - \left(df^{J_2}\right)^{-1}\Delta_2 - \left(\left(df^{J_1}\right)^{-1}f_{1,S} - \left(df^{J_2}\right)^{-1}f_{2,S}\right)\right)\mathbf{p}'\boldsymbol{\mu}S$$
$$+\left(\left(df^{J_1}\right)^{-1}f_{1,V} - \left(df^{J_2}\right)^{-1}f_{2,V}\right)\left(\mathbf{p}'\left(\boldsymbol{\omega} + \boldsymbol{\alpha}V - \boldsymbol{\psi}\odot\mathbf{V}\right) + d\mathbf{p}'\mathbf{V}\right) + 0.5\left(\left(df^{J_1}\right)^{-1}f_{1,SS} - \left(df^{J_2}\right)^{-1}f_{2,SS}\right)VS^2$$
$$= r\left(-S\left(\left(df^{J_1}\right)^{-1}\Delta_1 - \left(df^{J_2}\right)^{-1}\Delta_2\right) + \left(df^{J_1}\right)^{-1}f_1 - \left(df^{J_2}\right)^{-1}f_2\right)$$

We obtain that for each derivative the following expression is equal with the same constant:

$$-\left(df^{J_i}\right)^{-1}\left[f_{i,t} + f_{i,V}\left(\mathbf{p}'\left(\boldsymbol{\omega} + \boldsymbol{\alpha}V - \boldsymbol{\psi}\odot\mathbf{V}\right) + d\mathbf{p}'\mathbf{V}\right) + 0.5 f_{i,SS}VS^2 + r\left(Sf_{i,S} - f_i\right)\right]$$

We denote this by λ_s. The result follows by noticing that ξ can be expressed as:

$$\xi = f_t + f_V\left(\mathbf{p}'\left(\boldsymbol{\omega} + \boldsymbol{\alpha}V - \boldsymbol{\psi}\odot\mathbf{V}\right) + d\mathbf{p}'\mathbf{V}\right) + 0.5 f_{SS}VS^2 + f_S\mathbf{p}'\boldsymbol{\mu}S$$
$$= r\left(f - Sf_S\right) - \lambda_s df^J + f_S\boldsymbol{p}'\boldsymbol{\mu}S$$

References:

Albanese, C., Jaimungal, S. and Rubisov, D.H. (2003): 'A Two-state Jump Model', *Quantitative Finance* 3, 145-154

Alexander, C. and Lazar, E. (2005a): 'On the Continuous Limit of GARCH' *ICMA Centre Discussion Paper in Finance* 2005-13, available from www.ismacentre.rdg.ac.uk/dp

Alexander, C. and Lazar, E. (2005b): 'Asymmetries and Volatility Regimes in the European Equity Markets' *ICMA Centre Discussion Paper in Finance* 2005-14, available from www.ismacentre.rdg.ac.uk/dp

Alexander, C. and Lazar, E. (2006): 'Normal Mixture GARCH(1,1): Applications to Exchange Rate Modelling', *Journal of Applied Econometrics* 21, 307-336

Bai, X., Russell, J.R. and Tiao, G.C. (2001): 'Beyond Merton's Utopia (I): Effects of Non-normality and Dependence on the Precision of Variance Using High-frequency Financial Data', *University of Chicago, GSB Working Paper* July 2001

Bai, X., Russell, J.R. and Tiao, G.C. (2003): 'Kurtosis of GARCH and Stochastic Volatility Models with Non-normal Innovations' *Journal of Econometrics* 114(2), 349-360

Bakshi, G., Cao, C. and Chen, Z. (1997): 'Empirical Performance of Alternative Option Pricing Models', *Journal of Finance* 52, 2003-2049

Barone-Adesi, G., Engle, R. and Mancini, L. (2005): 'GARCH Options in Incomplete Markets', Proceedings of the ASMDA Conference 2005

Bates, D.S. (2000): 'Post-'87 Crash Fears in the S&P 500 Futures Option Market', *Journal of Econometrics* 94, 181-238

Bauwens, L., Bos, C.S. and van Dijk, H.K. (1999): 'Adaptive Polar Sampling with an Application to a Bayes measure of Value-at-Risk', *Tinbergen Institute Discussion Paper* TI 99-082/4

Bollen, N.P.B. (1998): 'Valuing Options in Regime-Switching Models', *Journal of Derivatives* 6, 38-49

Bollen, N.P.B., Gray, S.F. and Whaley, R.E. (2000): 'Regime-Switching in Foreign Exchange Rates: Evidence from Currency Option Prices', *Journal of Econometrics* 94, 239-276

Bollerslev, T. (1986): 'Generalized Autoregressive Conditional Heteroskedasticity', *Journal of Econometrics* 31, 309-328

Brown, L. D., Y. Wang and L.H. Zhao (2002) On the Statistical Equivalence at Suitable Frequencies of GARCH and Stochastic Volatility Models with Suitable Diffusion Models, *University of Pennsylvania,* Working paper

Buffington, J. and Elliott, R.J. (2002): 'American Options with Regime Switching', *International Journal of Theoretical and Applied Finance* 5, 497-514

Cai, J. (1994): 'A Markov Model of Switching-Regime ARCH', *Journal of Business and Economic Statistics* 12(3), 309-316

Chourdakis, K. (2000): 'Stochastic Volatility and Jumps Driven by Continuous Time Markov Chains', *University of London Queen Mary Working Paper* No. 430

Chourdakis, K. (2002): 'Continuous Time Regime Switching Models and Applications in Estimating Processes with Stochastic Volatility and Jumps', *University of London Queen Mary Working Paper* No. 464

Chourdakis, K. (2004): 'Non-Affine Option Pricing', *Journal of Derivatives* 11(3), 10-25

Corradi, V. (2000): 'Reconsidering the Continuous Time Limit of the GARCH(1,1) Process', *Journal of Econometrics* 96, 145-153

Cox, D.R. and Miller, H.D. (1987): *The Theory of Stochastic Processes*, Chapman and Hall

Daal, E. and Yu, J.S. (2005): 'A Comparison of Mixed GARCH-Jump Models with Skewed t-Distribution for Asset Returns', *SSRN Paper*, available from http://ssrn.com/abstract=670502

Di Masi, G.B., Kabanov, Y.M. and Runggaldier, W.J. (1994): 'Mean-Variance Hedging of Options on Stocks with Markov Volatilities', *Theory of Probability and Its Applications* 39, 173-181

Ding, Z. and Granger, C.W.J. (1996): 'Modeling Volatility Persistence of Speculative Returns: A New Approach', *Journal of Econometrics* 73, 185-215

Drost, F.C. and Nijman, T.E. (1993): 'Temporal Aggregation of GARCH Processes', *Econometrica* 61(4), 909-927

Duan, J. (1995): 'The GARCH Option Pricing Model', *Mathematical Finance* 5, 13-32

Duan, J, Popova, I. and Ritchken, P. (2002): 'Option Pricing under Regime Switching', *Quantitative Finance* 2, 1-17

Duan, J., Ritchken, P. and Sun, Z. (2005a): 'Jump Starting GARCH: Pricing and Hedging Options with Jumps in Returns and Volatilities, *Working Paper*, available from www.rotman.utoronto.ca/~jcduan/

Duan, J., Ritchken, P. and Sun, Z. (2005b): 'Approximating GARCH-Jump Models, Jump-Diffusion Processes, and Option Pricing', *Mathematical Finance* (forthcoming)

Duffie, D., Pan, J. and Singleton, K. (2000): 'Transform Analysis and Asset Pricing for Affine Jump-Diffusion', *Econometrica* 68(6), 1343-1376

Edwards, C. (2005): 'Derivative Pricing Models with Regime Switching: A General Approach', *Journal of Derivatives* Fall 2005, 41-47

Elliott, R.J. and Osakwe, C.-J.U. (2006): 'Option Pricing for Pure Jump Processes with Markov Switching Compensators', *Finance and Stochastics* 10(2), 250-275

Elliott, R.J., Siu, T.K. and Chan, L. (2006): 'Option Pricing for GARCH Models with Markov Switching', *International Journal of Theoretical and Applied Finance* 9(6), 825-841

Engle, R.F. (1982): 'Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation', *Econometrica* 50(4), 987-1007

Engle, R. F. and Ng, V. K. (1993): 'Measuring and Testing the Impact of News on Volatility', *Journal of Finance* 48, 1749-1778

Eraker, B. (2004): 'Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices', *The Journal of Finance* 59(3), 1367-1403

Eraker, B., Johannes, M. and Polson, N. (2003): 'The Impact of Jumps in Volatility and Returns', *The Journal of Finance* 58(3), 1269-1300

Ethier, S.N. and Kurtz, T.G. (1986): *Markov Processes: Characterization and Convergence*, John Wiley and Sons

Garcia, R, E. Ghysels and E. Renault (2005): 'The Econometrics of Option Pricing', forthcoming in Aït-Sahalia, Y. and L.P. Hansen (Eds.) *Handbook of Financial Econometrics*, Elsevier-North Holland, Amsterdam.

Glosten, L. R, Jagannathan, R. and Runkle, D. E. (1993): 'On the Relationship Between Expected Value and the Volatility of Excess Returns on Stocks', *Journal of Finance* 48, 1779-1801

Gray, S.F. (1996): 'Modeling the Conditional Distribution of Interest Rates as a Regime-Switching Process', *Journal of Financial Economics* 42, 27-62

Guo, X. (2001a): 'An Explicit Solution to an Optimal Stopping Problem with Regime Switching', *Journal of Applied Probability* 38(2), 464-481

Guo, X. (2001b): 'Information and Option Pricings', *Quantitative Finance* 1(1), 38-44

Guo, X. and Zhang, Q. (2004): 'Closed-Form Solutions for Perpetual American Put Options with Regime Switching', SIAM *Journal of Applied Mathematics* 64(6), 2034-2049

Haas, M., Mittnik, S. and Paolella, M.S. (2004a): 'Mixed Normal Conditional Heteroskedasticity', *Journal of Financial Econometrics* 2(2), 211-250

Haas, M., Mittnik, S. and Paolella, M.S. (2004b): 'A New Approach to Markov-Switching GARCH Models', *Journal of Financial Econometrics* 2(4), 493-530

Hamilton, J.D. and Susmel, R. (1994): 'Autoregressive Conditional Heteroscedasticity and Changes in Regime', *Journal of Econometrics* 64, 307-333

Hardy, M.R. (2002): 'A Regime-Switching Model of Long-Term Stock Returns', *North American Actuarial Journal* 5(2), 41-53

Heston, S. and Nandi, S. (2000): 'A Closed-form GARCH Option Valuation Model', *Review of Financial Studies* 13, 585-625

Hwang, S., Satchell, S.E. and Valls Pereira, P.L. (2004): 'How Persistent if Volatility? An Answer with Stochastic Volatility Models with Markov Regime Switching State Equations', *Proceedings of The Econometric Society 2004 Latin American Meetings* 198

Jobert, A. and Rogers, L.C.G. (2006): 'Option Pricing with Markov-Modulated Dynamics', SIAM *Journal on Control and Optimization* 44(6), 2063-2078

Johannes, M. (2004): 'The Statistical and Economic Role of Jumps in Continuous-time Interest Rate Models', *The Journal of Finance* 59(1), 227-260

Jones, C.S. (2003): 'The Dynamics of Stochastic Volatility: Evidence from Underlying and Option Markets', *Journal of Econometrics* 116, 181-224

Jorion, P. (1988): 'On Jump Processes in the Foreign Exchange and Stock Markets', *The Review of Financial Studies* 1(4), 427-445

Kalimpalli, M. and Susmel, R. (2004): 'Regime-Switching Stochastic Volatility and Short-Term Interest Rates', *Journal of Empirical Finance* 11(3), 309-329

Klaassen, F. (2002): 'Improving GARCH Volatility Forecasts with Regime-Switching GARCH', *Empirical Economics* 27, 363-394

Klein, G.E. (2002): Reply to 'A Regime-Switching Model of Long-Term Stock Returns' by M. Hardy, April 2001, *North American Actuarial Journal* 6(1), 171-176

Klüppelberg, C., Lindner, A. and Maller, R. (2004): 'A Continuous Time GARCH Process Driven by Lévy Process: Stationarity and Second Order Behaviour', *Journal of Applied Probability* 43(3), 601-622

Konikov, M. and Madan, D.B. (2002): 'Option Pricing using Variance Gamma Markov Chains', *Review of Derivatives Research* 5(1), 81-115

Kushner, H.J. (1990): 'Numerical Methods for Stochastic Control Problems in Continuous Time', SIAM *Journal of Control and Optimization* 28(5), 999-1048

Maheu, J.M. and McCurdy, T.H. (2004): 'News Arrival, Jump Dynamics, and Volatility Components for Individual Stock Returns', *The Journal of Finance* 59(2), 755-793

Merton, R.C. (1976): 'Option Pricing when Underlying Stock Returns are Discontinuous', *Journal of Financial Economics* 3, 125-144

Naik, V. (1993): 'Option Valuation and Hedging Strategies with Jumps in the Volatility of Asset Returns', *The Journal of Finance* 48(5), 1969-1984

Naik, V. and Lee, M. (1990): 'General Equilibrium Pricing of Options on the Market Portfolio with Discontinuous Returns', *Review of Financial Studies* 3, 493-521

Nelson, D.B. (1990): 'ARCH Models as Diffusion Approximations', *Journal of Econometrics* 45, 7-38

Pan, J. (2002): 'The Jump-risk Premia Implicit in Options: Evidence from an Integrated Time-series Study', *Journal of Financial Economics* 63, 3-50

Roberts, M.C. (2001): 'Commodities, Options, and Volatility: Modelling Agricultural Futures Prices', *Working Paper, The Ohio State University*

Siu, T.K., Tong, H. and Yang. H. (2004): 'On Pricing Derivatives under GARCH Models: A Dynamic Gerber-Shiu Approach', *North American Actuarial Journal* 8(3), 17-31

Smith, D.R. (2002): 'Markov-Switching and Stochastic Volatility Diffusion Models of Short-Term Interest Rates', *Journal of Business and Economic Studies* 20(2), 183-197

Stroock, D.W. and Varadhan, S.R.S. (1996): *Multidimensional Diffusion Processes*, Springer-Verlag

Timmermann, A. (2000): 'Moment of Markov Switching Models', *Journal of Econometrics* 96, 75-111

Vlaar, P.J.G. and Palm, F.C. (1993): 'The Message in Weekly Exchange Rates in the European Monetary System: Mean Reversion, Conditional Heteroscedasticity, and Jumps', *Journal of Business & Economic Statistics* 11(3), 351-60

Wang, Y. (2002) Asymptotic Nonequivalence of GARCH Models and Diffusions. *The Annals of Statistics* 30, 754-783

Yao, D.D., Zhang, Q. and Zhou, X.Y. (2006): 'A Regime-Switching Model for European Options', in Yan, H., Yin, G. and Zhang, Q.: *Stochastic Processes, Optimization and Control Theory. Applications in Financial Engineering, Queuing Networks and Manufacturing Systems*, Springer, 281-300