STATISTICAL ANALYSIS OF SOME FINANCIAL TIME SERIES MODELS

Fangfang Wang

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics and Operations Research (Statistics).

Chapel Hill
2009

Approved by,

Eric Ghysels, Advisor

Chuanshu Ji, Co-Advisor

Amarjit Budhiraja, Committee Member

Mariano Croce, Committee Member

Eric Renault, Committee Member

ABSTRACT

FANGFANG WANG: Statistical analysis of some financial time series models
(Under the direction of Professor Eric Ghysels)

The aim of this dissertation is to study the dynamics of asset returns under both the physical measure and the risk neutral measure. It consists of two different research topics.

The first topic is primarily concerned with a specific class of volatility component models. This family of models have received much attention recently, not only because of their ability to capture complex dynamics via a parsimonious parameter structure, but also because it is believed that they can handle well structural breaks or non-stationarity in asset price volatility. The first part of the dissertation focuses on their probabilistic properties and statistical inference on these models is discussed as well.

The second topic pertains to the distributional approximations of risk neutral distribution of asset returns for the purpose of option pricing. Risk neutral measures are a key ingredient of financial derivative pricing. Much effort has been devoted to characterizing the risk neutral distribution pertaining to the underlying asset. The rest of the dissertation studies the Generalized Hyperbolic family of distributions and examines their applications in option pricing.

ACKNOWLEDGEMENTS

I would like to take this opportunity to extend my heartfelt gratitude and deepest appreciation to all who were involved in the completion of this dissertation.

First, I would like to express my sincere gratitude to my dissertation advisor Professor Eric Ghysels for his encouragement, confidence and always reliable support. Throughout the dissertation he continually stimulated my analytical thinking and greatly assisted me with scientific writing.

I would also like to thank Professor Chuanshu Ji who also supervised my master thesis, for leading me into the area of financial statistics. With his unselfish kindness he has always been helping me through these years.

Furthermore, I am very grateful to the members of my committee: Professor Amarjit Budhiraja, Professor Mariano Croce, Professor Eric Renault with whom I had many extremely fruitful discussions that led me to improve this work. I also thank Professor Edward Carlstein and Professor Per Mykland for their insightful suggestions and comments on my work.

A special thanks goes to Professor Stephen Figlewski, editor of the Journal of Derivatives for his helpful comments and constructive suggestions related to Chapter 2 and Professor Bob Dittmar for providing the data in the empirical work.

Thanks always to my parents, for their unwavering support of all my endeavors, academic or otherwise.

I enjoyed my time at the University of North Carolina at Chapel Hill and I am thankful to all the faculty members in the Department of Statistics and Operations Research. I also thank my friends for their support and care throughout the years.

PREFACE

The aim of this dissertation is to study the dynamics of asset returns under both the physical measure and the risk neutral measure. It consists of two independent research topics. The first analyzes a specific class of recently introduced volatility component models. The second pertains to distributional approximations of risk neutral densities for the purpose of derivative pricing. The dissertation is made up of three chapters, which are three self-contained essays. The technical details are provided in the Appendix.

Chapter 1 is concerned with the statistical analysis of various volatility component models. This family of models have received much attention recently, not only because of their ability to capture complex dynamics via a parsimonious parameter structure, but also because it is believed that they can handle well structural breaks or non-stationarity in asset price volatility. This chapter focuses on studying the distributional properties of recently proposed volatility component models. Sufficient conditions for the existence or/and uniqueness of weakly/strictly stationary (ergodic) solutions with mixing property to the volatility component models are derived. There is a clear need for such an analysis, since any discussion about non-stationarity presumes we know when component models are stationary. As it turns out, this is not the case and the purpose of the study is to rectify this. The necessary conditions under which these models could structure non-stationarity are presented. This chapter also includes the sampling behavior of the maximum likelihood estimates of the volatility component models and their local consistency and asymptotic normality are establishes as well.

Chapter 2 and Chapter 3 pertain to derivative pricing and the characterization of the risk neutral distribution of underlying asset. Chapter 2 focuses on the class of normal inverse Gaussian (NIG) distributions and study its performance in approximating risk neutral density. The appeal of the NIG class of distributions is that they are skewed and leptokurtic which meet the stylized feature of asset returns, and it is analytically tractable in terms of the moment estimation. The reason to consider the method of moments estimation in this study is that the first four moments are what we care about in many risk management applications and the risk neutral moments could be formulated by a portfolio of European options. One strength of this approach is that we link the pricing of individual derivatives to the moments of the risk neutral distribution, which has an intuitive appeal in terms of how volatility, skewness and kurtosis of the risk neutral distribution can explain the behavior of derivative prices. It is shown, through numerical and empirical evidence, that the NIG distribution outperforms other existing methods in approximating the risk neutral distribution. Chapter 3 extends the work in Chapter 2. In Chapter 3, a more general class of distributions, the generalized hyperbolic (GH) family of distributions, are introduced in the context of risk management and various subclasses of the GH distribution which possess four-parameter characterization are investigated, including the NIG distribution, the variance gamma (VG) distribution, the generalized skewed t (GST) distribution. By analyzing their skewness, kurtosis and tail behavior, the NIG distribution and the VG distribution stand out of the others. In the numerical calibration, we follow the approach described in Chapter 2 to estimate parameters. Once again, the numerical results indicate that the NIG distribution outperforms all the other distributions studied in this chapter.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Statistical Analysis for Volatility Component Models

1.1 Introduction

Asset price volatility is persistent and several models have been proposed to capture this salient stylized fact. The ARCH class models originated by Engle (1982) is the most popular. The basic structure of ARCH is very much similar to ARMA, the appearance is deceiving. Indeed, there is a considerable literature on the stationarity, mixing and moment properties of various ARCH-type models, see e.g. Carrasco and Chen (2002), He and Teräsvirta (1999).

The prime focus has been on the GARCH(p,q) model - in particular GARCH(1,1) - originated by Bollerslev (1986). Yet, empirical evidence suggests that volatility dynamics is better described by component models. Engle and Lee (1999) introduced a GARCH model with a long and short run component, and several others have proposed related two-factor volatility models, see e.g. Ding and Granger (1996), Alizadeh et al. (2002), Chernov et al. (2003) and Adrian and Rosenberg (2008) among many others. The volatility component model of Engle and Lee (1999) decomposed the equity conditional variance as the sum of the short-run (transitory) and long-run (trend) components.

The appeal of component models is their ability to capture complex dynamics via a parsimonious parameter structure. Yet, there is also another reason why component models are becoming more popular, and this is again motivated by empirical evidence. Several studies have reported evidence of so called structural breaks in asset price volatility, see for example Inclan and Tiao (1994), Chen and Gupta (1997), Kokoszka and Leipus (2000), Horvath et al. (2001), Andreou and Ghysels (2002), Berkes et al. (2004), Kulperger and Yu (2005), Horvath et al. (2006), and among others.

To address the non-stationarity in the data, it has been suggested that such breaks should be captured by the long run component. Alternatively, locally stable GARCH models have been considered to handle non-stationarity - see e.g. Dahlhaus and Rao (2006). This chapter focuses exclusively on component models. For some component models, like the restricted GARCH(2,2) model of Engle and Lee (1999) which consist of two GARCH(1,1) components, the literature has not well covered the conditions that characterize non-stationarity issues of the components. Moreover, the component models that have been suggested recently are not of the additive ARCH-type, but instead consist of a multiplicative structure. The first to suggest a component structure that accommodates non-stationarity of volatility is Engle and Rangel (2008), later extended by Engle et al. (2008). These component models, also known as Spline-GARCH and GARCH-MIDAS respectively, feature a multiplicative decomposition of the conditional variance into a short-run (high-frequency) and long-run (low-frequency) components. The high-frequency volatility component in both models is driven by a GARCH(1,1) process which mean-reverts to one. The low-frequency component picks up the non-stationarity. The difference between the two models is the specification of the low-frequency volatility. The Spline-GARCH model formulates the low-frequency volatility in a non-parametric framework. Exponential quadratic Spline is used to estimate the long memory structure of low-frequency volatility so that the unconditional variance is time varying. This makes the model much more flexible but at the cost of losing the mean-reverting property.

The economic implications of component models and their empirical application have been studied intensively in Engle and Lee (1999), Engle and Rangel (2008), Engle et al. (2008). This chapter revisits the component models from a statistical perspective and attempts to explore the stationarity and mixing properties of the underlying processes. There is a clear need for such an analysis, since any discussion about non-stationarity presumes we know when component models are stationary. As it turns out, this is not the case and the purpose of the chapter is to rectify this.

Although most of our focus is on the aforementioned multiplicative models, we start with filling a gap in the literature pertaining to additive component models, that is the original Engle and Lee model. The dynamic structure of the conditional variance in their model can be reduced to a restricted GARCH(2,2) model with certain coefficients negative, which, to some extent, distinguishes itself from the classic GARCH model. Hence, the existing regularity conditions for GARCH models need to be extended to handle the constrained additive component models. Under certain regularity conditions on the parameters, the transitory component mean-reverts to zero and the trend converges to the unconditional variance but at a much slower rate. While, the resulting volatility process is covariance stationary, as pointed out by Engle and Lee (1999), the mapping from component models to GARCH involves nonlinear transformations of the parameter space.

The GARCH-MIDAS model of Engle et al. (2008) modified the dynamics of low-frequency volatility as a stochastic component "by smoothing realized volatility in the spirit of MIDAS (mixed data sampling, see e.g. Ghysels et al. (2004)) regression and MIDAS filtering" so that it can incorporate directly data sampled at lower frequency (say, monthly or quarterly) than the asset returns (sampled at a daily basis). The GARCH-MIDAS model has two basic specifications. In terms of the structure of low-frequency volatility, they are classified as: (1) GARCH-MIDAS model with fixed time span realized volatility (RV) where the low-frequency component is constant within a fixed time span, say a month or a quarter but the high-frequency component is varying from day to day; (2) GARCH-MIDAS model with rolling window realized volatility (RV) where both low-frequency and high-frequency components change at a daily basis.

In particular, we are looking for regularity conditions under which the models could admit covariance stationary or strictly stationary ergodic solutions with/without β-mixing property. By linking the models with multivariate stochastic difference equations, we study the covariance stationary property through a reversed martingale argument and the strict stationarity property in terms of the top Lyapounov exponent. The dilemma is how to evaluate theoretically the top Lyapounov exponents which are defined on (i) a sequence of i.i.d. matrices with certain negative entries and (ii) a sequence of strictly stationary ergodic matrices with positive entries. In addition, we derive the locally consistent estimates of the GARCH-MIDAS model with rolling window realized volatility specification and study their asymptotic behaviors by means of Cramér-Wold device.

The rest of this chapter is organized as follows: we revisit the volatility component model of Engle and Lee in section 1.2, and present the conditions under which it is strictly stationary ergodic and β-mixing. Section 1.3 focuses on the stationarity properties of the two GARCH-MIDAS specifications. The consistent estimates with asymptotic behaviors of GARCH-MIDAS model with rolling window RV are studied in section 1.4. Section 1.5 gives the concluding remarks. In Appendix A.1, we list the theorems and lemmas cited from others' work for easy reference.

1.2 Volatility component model of Engle and Lee

The volatility component model of Engle and Lee (1999) structures the daily return r_t as

$$\begin{aligned} r_t &= \sqrt{h_t}\varepsilon_t \\ h_t &= \tau_t + g_t \\ g_t &= \alpha(r_{t-1}^2 - \tau_{t-1}) + \beta g_{t-1} \\ \tau_t &= \omega + \rho\tau_{t-1} + \phi(r_{t-1}^2 - h_{t-1}) \end{aligned} \tag{1.1}$$

where $\varepsilon_t \overset{iid}{\sim} N(0,1)$ and the parameter space is

$$\mathcal{P} = \{(\alpha, \beta, \omega, \rho, \phi) \in (\mathbb{R}^{5+})^\circ : \alpha + \beta < \rho < 1, \phi < \beta\}$$

which ensures the conditional variance h is nonnegative (see Engle and Lee (1999) for the proof of nonnegativity of h).

According to the model, the conditional variance is the sum of long-run (trend) variance τ and the short-run (transitory) variance g. The condition $0 < \alpha + \beta < \rho < 1$ guarantees that the short-run volatility mean-reverts to zero at a geometric rate of $\alpha + \beta$ and long-run volatility converges to $\omega/(1-\rho)$ with a much slower rate.

Engle and Lee (1999) provided sufficient conditions for the covariance stationarity of $\{r_t\}$ with parameter space $\mathcal{P}$ by linking it to an ARMA(2,2) process, i.e.

$$\begin{aligned} r_t^2 = \omega(1-\alpha-\beta) + (\alpha+\beta+\rho)r_{t-1}^2 - (\rho\alpha + \rho\beta)r_{t-2}^2 \\ + \eta_t - (\rho+\beta-\phi)\eta_{t-1} - [(\phi(\alpha+\beta) - \beta\rho)]\eta_{t-2} \end{aligned}$$

where $\eta_t = r_t^2 - h_t$ (see Engle and Lee (1999)). Here we shall present conditions for strict stationarity and β-mixing. For the time being, we assume the process to extend infinitely into the past. Later, we will consider the scenario of closing the system by assigning an initial distribution at time point 0.

The volatility component model of Engle and lee is also referred to as the restricted GARCH(2,2) model because the dynamics of conditional variance h can be cast into the framework of a GARCH(2,2) process as

$$\begin{aligned} r_t &= \sqrt{h_t}\varepsilon_t \\ h_t &= \alpha_0 + \alpha_1 r_{t-1}^2 + \alpha_2 r_{t-2}^2 + \beta_1 h_{t-1} + \beta_2 h_{t-2} \end{aligned} \tag{1.2}$$

where $\alpha_0 = \omega(1-\alpha-\beta) > 0, \alpha_1 = \phi + \alpha > 0$, $\alpha_2 = -(\phi(\alpha+\beta)+\alpha\rho) < 0$, $\beta_1 = \rho + \beta - \phi > 0$, and $\beta_2 = \phi(\alpha+\beta) - \rho\beta < 0$. The distinct feature of this 'new' model is its similarity to a GARCH(2,2) setting but of having negative coefficients (α_2 and β_2 are negative). So the existing results about classic GARCH(2,2) model Bougerol and Picard (1992) can not be applied to the volatility component model of Engle and Lee.

Introducing $Y_t = (h_{t+1}, h_t, r_t^2)'$, $B = (\alpha_0, 0, 0)'$, and

$$A(\varepsilon_t) = \begin{pmatrix} \beta_1 + \alpha_1\varepsilon_t^2 & \beta_2 & \alpha_2 \\ 1 & 0 & 0 \\ \varepsilon_t^2 & 0 & 0 \end{pmatrix}$$

the restricted GARCH(2,2) process (1.2) of Engle and Lee is equivalent to the solution to a stochastic difference equation defined through

$$Y_t = A(\varepsilon_t)Y_{t-1} + B. \tag{1.3}$$

with iid coefficients.

There is a vast literature on the existence/uniqueness of the strictly stationary solution to the stochastic difference equation of the form

$$Y_t = A_t Y_{t-1} + B_t, \;\; t \in \mathbb{Z} \tag{1.4}$$

where Y_t and B_t are $\mathbb{R}^n$-valued random vectors, A_t is a $\mathbb{R}^{n\times n}$-valued random matrix, and $\{(A_t, B_t), t \in \mathbb{Z}\}$ is a strictly stationary ergodic sequence. Vervaat (1979) and Brandt (1986) analyzed the stochastic difference equation for the scaler case, i.e. $n = 1$ with assumption that the coefficients are iid and strictly stationary ergodic respectively. Bougerol and Picard (1992) studied the problem with A_t and B_t being iid. Glasserman and Yao (1995) extended the results for the general strictly stationary ergodic sequence. For the vector case, the problem of strictly stationary ergodic solution to (1.4) is closely related to the associated top Lyapounov exponent which is defined as

Definition 1.1. *Let $\{A_t, t \in \mathbb{Z}\}$ be a strictly stationary and ergodic sequence of $\mathbb{R}^{n\times n}$-valued random matrices, such that $E\log^+\|A_0\| < \infty$. Then the top Lyapounov exponent associated with $\{A_t, t \in \mathbb{Z}\}$ is defined as*

$$\gamma = \inf_{t\in\mathbb{N}} E(\frac{1}{t+1}\log\|A_tA_{t-1}\dots A_0\|).$$

Combining subadditive ergodic theory of Kingman (1973) due to the sub-multiplicativity of matrix norm and the work of Furstenberg and Kesten (1960), we could derive a well-known property of the top Lyapounov exponent which is stated as

Theorem 1.1 (Furstenberg and Kesten (1960), Kingman (1973))**.** *If $\{A_t, t \in \mathbb{Z}\}$ is a strictly stationary ergodic sequence of $\mathbb{R}^{n\times n}$-valued random matrices, such that $E\log^+\|A_0\| < \infty$, then*

$$-\infty \le \gamma < \infty$$

$$\lim_{t\to\infty}\frac{1}{t+1}\log\|A_tA_{t-1}\dots A_0\| = \gamma \textit{ almost surely}$$

and

$$\lim_{t\to\infty}\frac{1}{t+1}E\log\|A_tA_{t-1}\dots A_0\| = \gamma$$

The top Lyapounov exponent is independent of the choice of underlying matrix norm $\|.\|$ since all the norms on the finite norm space are equivalent. For ease of analysis, we consider the Frobenius norm in particular throughout this chapter. Next proposition gives a sufficient condition for the strict stationarity of the restricted GARCH(2,2) model of Engle and Lee when we assume the whole system starts from the negative infinity.

Proposition 1.1. *For the volatility component model of Engle and Lee with the parameter space* $\mathcal{P}$, $\{r_t, h_t\}$ *is strictly stationary ergodic if* $\alpha < \phi$, $2\alpha + \beta + \phi < \rho < 5\alpha + \beta$ *and* $\alpha + \beta + \rho < 1$.

The proof of Proposition 1.1 needs the following lemmas.

Lemma 1.1. *Let* $\{F_t, t \in \mathbb{Z}\}$ *be a sequence of iid random matrices such that* $P(F_t F_{t-1} \dots F_0 \geq 0) = 1$ *for any* t. *Suppose that* $E(\log^+ \|F_0\|) < \infty$ *and* $\rho(E(F_0)) < 1$. *Then the top Lyapounov exponent associated with this sequence is strictly negative.*

Proof of Lemma 1.1. *Define* $M_k = F_{-1} \dots F_{-k}$, *then* $E(M_k) = F^k$ *where* $F = E(F_0)$. *Under the assumption that* $\rho(F) < 1$, $\sum_k F^k < \infty$. *Since* F_j *is iid and* $P(F_t F_{t-1} \dots F_0 \geq 0) = 1$, *by Fubini's theorem, we have* $\sum_k M_k < \infty$ *almost surely. Therefore, almost surely* $\lim_{k\to\infty} M_k = 0$, *or* $\lim_{k\to\infty} \|M_k\| = 0$. *Let* $\tilde{F}_k$ *be the transpose of* F_{-k}. *Since* $\|\tilde{F}_k \dots \tilde{F}_1\| = \|F_{-1} \dots F_{-k}\|$, *the top Lyapounov exponent associated with* $\{F_t, t \in \mathbb{Z}\}$ *is strictly negative, following from Lemma 3.4 of Bougerol and Picard (1992) (See Lemma A.2 in Appendix).*

Next we establish, through the following lemma, the conditions under which the product of matrices $A(\varepsilon_t)A(\varepsilon_{t-1}) \dots A(\varepsilon_0)$ is nonnegative almost surely for any t.

Lemma 1.2. *Suppose that* $\alpha < \phi$, *and* $2\alpha + \beta + \phi < \rho < 5\alpha + \beta$. *If further express* h_t *in model (1.2) with parameter space* $\mathcal{P}$ *as an infinite distributed lag of* r_t^2, *then all the coefficients are positive, i.e.*

$$h_t = \omega^* + \sum_{k=0}^{\infty} \phi_k r_{t-k-1}^2$$

with $\omega^* \geq 0, \phi_k \geq 0 \; \forall k$.

Proof of Lemma 1.2. *Let* Z_1 *and* Z_2 *be the roots of* $Z^2 - \beta_1 Z - \beta_2$. *WLOG, assume* $|Z_1| \geq |Z_2|$. *By theorem 2 of Nelson and Cao Nelson and Cao (1992) (see Appendix), to show* $\omega^* \geq 0, \phi_k \geq 0$ *it is equivalent to prove that*

(1) Z_1, Z_2 *are real, and* $|Z_1| < 1$ $|Z_2| < 1$;

(2) $\alpha_0/(1 - Z_1 - Z_2 + Z_1 Z_2) \geq 0$;

(3) $\alpha_1 Z_1 + \alpha_2 > 0$ *and* $\alpha_1 Z_2 + \alpha_2 \neq 0$;

(4) $\phi_k \geq 0$ *for* $k = 0, 1, 2$.

Conditions (1) & (2) have been checked by Engle and Lee (see Appendix of Engle and Lee (1999)). We only need to justify conditions (3) & (4) under the restrictions specified. Since

$$\begin{aligned}\alpha_1 Z_1 + \alpha_2 &= \tfrac{\phi+\alpha}{2}[\rho + \beta - \phi + \sqrt{(\rho-\beta-\phi)^2 + 4\alpha\phi} - 2\tfrac{\alpha\phi+\alpha\rho+\phi\beta}{\phi+\alpha}] \\ &= \tfrac{\phi+\alpha}{2}[\rho - (\beta + 2\alpha + \phi) + \sqrt{(\rho-\beta-\phi)^2 + 4\alpha\phi} \\ &\quad -\tfrac{2\alpha}{\phi+\alpha}(\rho - \alpha - \beta)]\end{aligned}$$

Note that under the restrictions, $\rho - (\beta + 2\alpha + \phi) > 0$, $\frac{2\alpha}{\phi+\alpha} < 1$, *and the polynomial* $g(\phi) = (\rho-\beta-\phi)^2 + 4\alpha\phi - (\rho-\alpha-\beta)^2 = \phi^2 - 2\phi(\rho-\beta-2\alpha) + 2\alpha(\rho-\beta) - \alpha^2 > 0$ *due to the fact that* $\Delta = (\rho-\beta-2\alpha)^2 - 2\alpha(\rho-\beta) - \alpha^2 = (\rho-\beta-5\alpha)(\rho-\beta-\alpha) < 0$. *Therefore,* $\alpha_1 Z_1 + \alpha_2 > 0$. *Meanwhile,*

$$\begin{aligned}\alpha_1\beta_1/2 + \alpha_2 &= (\phi+\alpha)(\rho+\beta-\phi)/2 - (\phi\alpha + \phi\beta + \alpha\rho) \\ &= -\frac{1}{2}[\phi(\phi-\alpha) + \phi(5\alpha+\beta-\rho) + \alpha(\rho-\beta-\phi)] < 0\end{aligned}$$

thus $\alpha_1 Z_2 + \alpha_2 \neq 0$.

Next to check condition (4). Since

$$\begin{aligned}\phi_0 &= \alpha_1 > 0 \\ \phi_1 &= \beta_1\alpha_1 + \alpha_2 \\ &= (\phi+\alpha)(\rho+\beta-\phi) - (\phi\alpha + \phi\beta + \alpha\rho) \\ &= \phi(\rho-\phi-\alpha) + \alpha(\beta-\phi) > 0 \\ \phi_2 &= \beta_1\phi_1 + \beta_2\phi_0 \\ &= (\rho+\beta-\phi)\phi_1 + [\phi(\alpha+\beta) - \rho\beta](\phi+\alpha) \\ &= (\beta-\phi)\phi_1 + \phi(\phi+\alpha)(\alpha+\beta) + \rho\phi(\rho-2\alpha-\beta-\phi) > 0\end{aligned}$$

Condition (4) is also satisfied. Therefore $\omega^* \geq 0$ *and* $\phi_k \geq 0\ \forall k$.

Proof of Proposition 1.1. *According to Theorem 3.1 of Glasserman and Yao (see Appendix),* *the statement is true if*

$$E(\log \|A(\varepsilon_0)\|)^+ < \infty \text{ and } \gamma < 0.$$

Under Frobenius norm,

$$\|A(\varepsilon_0)\|^2 = (\beta_1 + \alpha_1 \varepsilon_t^2)^2 + (\beta_2)^2 + (\alpha_2)^2 + 1 + (\varepsilon_t^2)^2 > 1$$

so $E(\log \|A(\varepsilon_0)\|)^+ < \infty$.

Define $M_{t,k} = A(\varepsilon_t)A(\varepsilon_{t-1})\dots A(\varepsilon_{t-k})$, *then* $E(M_{t,k}) = M^k$ *where*

$$M = \begin{pmatrix} \beta_1 + \alpha_1 & \beta_2 & \alpha_2 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

The eigenvalues of M *is* $0, \alpha + \beta$ *and* ρ, *from where we know* $\rho(M) < 1$ *by assumption. Using Lemma 1.2, it could be derived that each component of* $M_{t,k}$ *is nonnegative. Further applying Lemma 1.1, the top Lyapounov exponent* γ *associated with* $\{A(\varepsilon_t), t \in \mathbb{Z}\}$ *is strictly negative.*

In Proposition 1.1, the model is assumed to extend infinitely into the past. Next we consider the system (1.2) starting from time 0 with initial values g_0 and τ_0 defined on the probability space $\{\Omega, \mathcal{F}, \mathcal{P}\}$ such that $P(0 < \tau_0 < \infty) = P(0 < \tau_0 + g_0 < \infty) = 1$. Now the process (1.3) can be viewed as a time-homogeneous Markov process, which puts us in the setting of the polynomial random coefficient autoregressive model mentioned in Carrasco and Chen (2002). Starting from there, we could derive the mixing property of volatility component model.

Based on the work of Mokkadem (1990), Carrasco and Chen (2002) studied the conditions for the stationarity, mixing and moment properties of various ARCH-type models. Again, we consider Theorem 4.3 of Mokkadem (1990) or Theorem 1 of Carrasco and Chen (2002) (see Appendix), and we have the following,

Proposition 1.2. *Consider the volatility component model of Engle and Lee with the parameter space* $\mathcal{P}$, *with* $\alpha < \phi$, $2\alpha + \beta + \phi < \rho < 5\alpha + \beta$, $\alpha + \beta + \rho < 1$ *and the distribution induced by* $\tau_0 + g_0$ *invariant, then* $E[h_t] < \infty$, $E[r_t^2] < \infty$, $\{r_t, h_t\}$ *is strictly stationary and* β*-mixing with exponential decay.*

Proof of Proposition 1.2. *By Theorem 4.3 of Mokkadem (1990) or Theorem 1 of Carrasco and Chen (2002) (see Appendix), as long as assumptions (A.1-A.5) are verified, the statement is true. (A.1) and (A.2) are satisfied straightforwardly. Hence, we need to check (A.3), (A.4) and (A.5).*

- *Assumption (A.3): Note*

$$A(0) = \begin{pmatrix} \beta_1 & \beta_2 & \alpha_2 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

and its characteristic function is $det(\lambda I_3 - A(0)) = \lambda(\lambda^2 - \beta_1\lambda - \beta_2)$. *Let* $f(\lambda) = \lambda^2 - \beta_1\lambda - \beta_2$. *Since*

$$\begin{aligned} \beta_1^2 + 4\beta_2 &= (\rho - \phi - \beta)^2 + 4\alpha\phi > 0, \\ f(\beta_1/2) &= -(\beta_1^2/4 + \beta_2) < 0, \\ f(0) &= -\beta_2 > 0, \\ f(1) &= (1-\beta)(1+\phi-\rho) - \phi\alpha > 0, \end{aligned}$$

hence $\rho[A(0)] < 1$.

- *Assumption (A.4): From the proofs of Lemma 1.1 and Proposition 1.1,*

$$\sum_{k=1}^{\infty} A(\varepsilon_t)A(\varepsilon_{t-1})\dots A(\varepsilon_{t-k})B < \infty$$

almost surely and

$$A(\varepsilon_t)A(\varepsilon_{t-1})\dots A(\varepsilon_{t-k})$$

converges almost surely to the 0 matrix.

- *Assumption (A.5): Define* $V(y) = |y_1| + a|y_2| + a|y_3|$ *for* $y = (y_1, y_2, y_3)' \in \mathbb{R}^3$, *where* $a = \frac{1-(\alpha_1+\beta_1)}{4} > 0$ *(since* $\alpha_1 + \beta_1 = \alpha + \beta + \rho < 1$ *by assumption). Let* $\pi = \frac{1+\alpha_1+\beta_1}{2} < 1$ *and* $B > 0$ *be such that* $\frac{\alpha_0+1}{B} < 1 - \pi$. *Note*

$$E[V(Y_t)|Y_{t-1}] = E[h_{t+1} + ah_t + ar_t^2|Y_{t-1}]$$

$$= E[\alpha_0 + (\beta_1 + a)h_t + (\alpha_1 + a)r_t^2 + \beta_2 h_{t-1} + \alpha_2 r_{t-1}^2 | Y_{t-1}]$$

$$= \alpha_0 + (\beta_1 + \alpha_1 + 2a)h_t^2 + \beta_2 h_{t-1} + \alpha_2 r_{t-1}^2$$

$$\leq \alpha_0 + \pi V(Y_{t-1})$$

Define $K = \{k \in \mathbb{R}^3 : V(k) \leq B\}$*, then* $E[V(Y_t)|Y_{t-1} = y]$ *is bounded for* $y \in K$*. On* K^c*,*

$$E[V(Y_t)|Y_{t-1} = y] \leq \alpha_0 + \pi V(y) \leq (\frac{\alpha_0 + 1}{B} + \pi)V(y) - 1$$

Assumption (A.5) is also satisfied.

1.3 Stationarity of GARCH-MIDAS process

The spline-GARCH model of Engle and Rangel (2008) and the GARCH-MIDAS model of Engle et al. (2008) assume the conditional volatility to be the product of long-run and short-run volatility. To be specific, the spline-GARCH model is defined through the following three equations

$$\begin{aligned} r_t &= \mu + \sqrt{\tau_t g_t}\varepsilon_t \\ g_t &= (1 - \alpha - \beta) + \alpha \tfrac{(r_{t-1} - \mu)^2}{\tau_{t-1}} + \beta g_{t-1} \\ \tau_t &= c \exp(w_0 t + \textstyle\sum_{i=1}^k w_i (t - t_{i-1})^2 1_{\{t > t_{i-1}\}}) \end{aligned}$$

where

- $\varepsilon_t \overset{iid}{\sim} N(0, 1)$
- $\{0 = t_0 < t_1 < t_2 < \ldots < t_k = T\}$ is a partition of the time horizon T in k equally spaced intervals.

The high-frequency component g follows a mean-reverting unit GARCH(1,1) process. The low-frequency component τ is deterministic, and it equals the unconditional variance, ie $E(r_t - \mu)^2 = \tau_t$ from where we could see the conditional volatility process is not mean-reverting and is not stationary as well.

GARCH-MIDAS model, as an extension of spline-GARCH model, keeps the structure of short-run component g but modifies the long-run component τ as stochastic. According to the way the low-frequency component is structured, GARCH-MIDAS model has two basic specifications: GARCH-MIDAS model with fixed time span realized volatilities (RV) and GARCH-MIDAS model with rolling window realized volatility (RV).

For the fixed time span RV setting, the dynamics of long-run and short-run components are specified as

$$\begin{aligned} r_{i,t} &= \mu + \sqrt{\tau_t g_{i,t}}\varepsilon_{i,t}, \quad 2 \leq i \leq N_t, t \in \mathbb{Z} \\ g_{i,t} &= (1-\alpha-\beta) + \alpha\frac{(r_{i-1,t}-\mu)^2}{\tau_t} + \beta g_{i-1,t} \\ \tau_t &= m + \theta \sum_{k=1}^{K} \varphi_k(\omega) RV_{t-k}, \quad RV_t = \sum_{i=1}^{N_t} r_{i,t}^2 \end{aligned} \tag{1.5}$$

where

- r_{it} is the log return on day i of period (say month, quarter, etc.) t.
- N_t is the number of days in period t, but in this chapter we assume $N_t = N$(a predetermined number) for any t.
- $\varepsilon_{i,t} \overset{iid}{\sim} N(0,1) \quad \forall i, t$.
- $E(g_{1,t}|\mathcal{F}_{t-1}) = 1$, which is equivalent to $E(g_{i,t}|\mathcal{F}_{t-1}) = 1 \quad (1 \leq i \leq N_t)$, an assumption used in Engle et al. (2008).
- $\varphi_k(\omega)$ are nonnegative functions of ω such that $\sum_{k=1}^{N} \varphi_k(\omega) = 1$.
- $\alpha > 0$,$\beta > 0$, $\alpha + \beta < 1$, $\theta > 0$ $m > 0$.

For the rolling window RV setting, the long-run component dynamics is simplified,

$$\begin{aligned} r_t &= \mu + \sqrt{\tau_t g_t}\varepsilon_t, t \in \mathbb{Z} \\ g_t &= (1-\alpha-\beta) + \alpha\frac{(r_{t-1}-\mu)^2}{\tau_{t-1}} + \beta g_{t-1} \\ \tau_t &= m + \theta \sum_{k=1}^{K} \varphi_k(\omega) RV_{t-k}, RV_t = \sum_{j=0}^{N-1} r_{t-j}^2 \end{aligned} \tag{1.6}$$

where

- r_t is the log return on day t,

- $\varepsilon_t \overset{iid}{\sim} N(0,1)$,

- N is the length of a certain period of interest with value predetermined,

- $\varphi_k(\omega)$ are nonnegative functions of ω such that $\sum_{k=1}^{N} \varphi_k(\omega) = 1$,

- $\alpha > 0, \beta > 0$, $\alpha + \beta < 1$, $\theta > 0$, $m > 0$.

The appeal of GARCH-MIDAS model is that the structure of long-run component is stochastic which makes it possible to study the statistical property of the conditional volatility process.

1.3.1 GARCH-MIDAS model with fixed time span RV

In our analysis, we assume that $\mu = 0$. Hence Model (1.5) becomes

$$\begin{aligned} r_{i,t} &= \sqrt{\tau_t g_{i,t}} \varepsilon_{i,t}, \quad 1 \le i \le N, t \in \mathbb{Z} \\ g_{i,t} &= (1-\alpha-\beta) + \alpha \frac{(r_{i-1,t})^2}{\tau_t} + \beta g_{i-1,t} \\ \tau_t &= m + \theta \sum_{k=1}^{K} \varphi_k(\omega) RV_{t-k}, \quad RV_t = \sum_{i=1}^{N} r_{i,t}^2 \end{aligned} \tag{1.7}$$

Proposition 1.3. *Suppose that $\alpha > 0$, $\beta > 0$, $\alpha + \beta < 1$, $\theta > 0$ and $m > 0$, $\{r_{i,t}\}$ defined in (1.7) is a White Noise if* $0 < \theta < 1/N$.

Proof of Proposition 1.3. *To show that $\{r_{i,t}\}$ is a White Noise, we need to verify the following three conditions:*

(i) $E(r_{i,t}) = 0$

(ii) $Cov(r_{i,t}, r_{j,s}) = 0$ *for* $j \neq i$ *or* $t \neq s$

(iii) $Var(r_{i,t})$ *is a finite constant.*

(i) is true since $E(r_{i,t}) = E(\sqrt{\tau_t g_{i,t}} \varepsilon_{i,t}) = 0$ and (ii) also holds due to the property of $\varepsilon_{i,t}$. Now we need to check the third condition.

For ease of reference, let $\eta \equiv \alpha + \beta$, $\Psi_{i-1,t} \equiv \alpha \varepsilon_{i-1,t}^2 + \beta$. Then $g_{i,t} = 1 - \eta + \Psi_{i-1,t} g_{i-1,t}$

and

$$
\begin{aligned}
E_{t-1}[\tau_t g_{i,t}] &= \tau_t[(1-\eta) + \eta E_{t-1} g_{i-1,t}] \\
&\vdots \\
&= \tau_t[(1-\eta^{i-1}) + \eta^{i-1} E_{t-1} g_{1,t}] \\
&= \tau_t
\end{aligned}
$$

where $E_{t-1}[.]$ *is equivalent to* $E[.|F_{N,t-1}]$.

It follows that $E_{t-s}[\tau_t g_{i,t}] = E_{t-s}[\tau_t]$ *for* $s \geq 1$, *and*

$$
Var_{t-s}[r_{i,t}] = E_{t-s}[\tau_t g_{i,t} \varepsilon_{i,t}^2] = E_{t-s}[\tau_t g_{i,t}] = E_{t-s}[\tau_t]
$$

therefore,

$$
Var[r_{i,t}] = Var[E_{t-s}(r_{i,t})] + E[Var_{t-s}(r_{i,t})] = E[\tau_t]
$$

Next we need to show that $E[\tau_t]$ *exists and is finite. Notice that*

$$
\begin{aligned}
E_{t-K-1}[\tau_t] &= m + \theta \sum_{k=1}^{K} \varphi_k(\omega) E_{t-K-1}[RV_{t-k}] \\
&= m + \theta \sum_{k=1}^{K} \varphi_k(\omega) E_{t-K-1}[\textstyle\sum_{i=1}^{N} r_{i,t-k}^2] \\
&= m + \theta N \sum_{k=1}^{K} \varphi_k(\omega) E_{t-K-1}(\tau_{t-k})
\end{aligned} \tag{1.8}
$$

Introduce $Y_t = (\tau_t, \tau_{t-1}, \ldots, \tau_{t-K+1})^T$. *(1.8) is equivalent to*

$$
E_{t-K-1}(Y_t) = A E_{t-K-1}(Y_{t-1}) + B \tag{1.9}
$$

where

$$
A = \begin{pmatrix}
N\theta\varphi_1 & N\theta\varphi_2 & \ldots & N\theta\varphi_{K-1} & N\theta\varphi_K \\
1 & 0 & \ldots & 0 & 0 \\
0 & 1 & \ldots & 0 & 0 \\
\vdots & & & & \\
0 & 0 & \ldots & 1 & 0
\end{pmatrix}, B = \begin{pmatrix} m \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}
$$

Moreover, we have

$$E_{t-s}(Y_t) = AE_{t-s}(Y_{t-1}) + B, \forall s \geq K+1 \tag{1.10}$$

by iteration,

$$E_{t-s}(Y_t) = A^{s-K}E_{t-s}(Y_{t-s+K}) + (I + A + \ldots + A^{s-K-1})B \tag{1.11}$$

Set $t = s$, *equ(1.11) becomes*

$$E_0(Y_s) = A^{s-K}E_0(Y_K) + (I + A + \ldots + A^{s-K-1})B$$

Since if $0 < \theta < 1/N$, $\lim_{s\to\infty} A^s = 0$*(*∗*), then* $\lim_{s\to\infty} E_0(Y_s) = (I - A)^{-1}B$. *It follows that* $E(Y_s)$ *is finite (elementary-wise) when* s *is sufficiently large. Together with equ(1.10), we know* $E(Y_s)$ *is finite for every* s. *Fix* t, *and let* s *go to infinity in (1.11). By the property of reversed martingale, we have*

$$E(Y_t) = \lim_{s\to\infty} E_{t-s}(Y_t) = (I - A)^{-1}B = \frac{m}{1 - N\theta}\iota$$

where ι *is a vector of 1's, and* $Var[r_{it}] = E[\tau_t] = \frac{m}{1-N\theta}$.

Now we need to verify (∗*):* $\lim_{s\to\infty} A^s = 0$ *if* $0 < \theta < 1/N$. *Note*

$$f(\lambda) = det(\lambda I_K - A) = \lambda^K - N\theta\varphi_1\lambda^{K-1} - N\theta\varphi_2\lambda^{K-2} - \ldots - N\theta\varphi_K$$

Since

$$|f(\lambda)| \geq 1 - N\theta\varphi_1 - \ldots - N\theta\varphi_K = 1 - N\theta > 0 \ \ if \ |\lambda| \geq 1,$$

$\rho(A) = \max_j |\lambda_j|$ *should be strictly less than* 1 *which implies that* $\lim_{s\to\infty} A^s = 0$.

1.3.2 GARCH-MIDAS model with rolling window RV

For the rolling window RV setting, again, we consider the model without drift, i.e., $\mu = 0$. Model (1.6) becomes

$$\begin{aligned} r_t &= \sqrt{\tau_t g_t}\varepsilon_t, t \in \mathbb{Z} \\ g_t &= (1-\alpha-\beta) + \alpha\frac{(r_{t-1})^2}{\tau_{t-1}} + \beta g_{t-1} \\ \tau_t &= m + \theta \textstyle\sum_{k=1}^{K} \varphi_k(\omega) RV_{t-k}, RV_t = \sum_{j=0}^{N-1} r_{t-j}^2 \end{aligned} \tag{1.12}$$

Further, the dynamics of r_t^2 could be reduced to

$$r_t^2 = m g_t \varepsilon_t^2 + \theta g_t \varepsilon_t^2 \textstyle\sum_{l=1}^{N+K-1} c_l r_{t-l}^2 \tag{1.13}$$

where c_l's are certain combinations of φ_k's and they satisfy

$$\sum_{l=1}^{N+K-1} c_l = N \sum_{k=1}^{K} \varphi_k(\omega) = N.$$

Under the assumptions $\alpha > 0, \beta > 0$ and $\alpha + \beta < 1$, model (1.13) can be linked to a multivariate stochastic difference equation with strictly stationary ergodic coefficients through Markovian representation (Akaike (1974), Vervaat (1979)). In other words, the stationarity property of the process $\{r_t^2, t \in \mathbb{Z}\}$ is equivalent to the existence of stationary solution to the following stochastic difference equation

$$Y_t = A_t(\tilde{c})Y_{t-1} + B_t. \tag{1.14}$$

where

$$Y_t = (r_t^2, r_{t-1}^2, \ldots, r_{t-N-K+2}^2)',$$

$$A_t(\tilde{c}) = \begin{pmatrix} \theta g_t \varepsilon_t^2 c_1 & \dots & \theta g_t \varepsilon_t^2 c_{N+K-2} & \theta g_t \varepsilon_t^2 c_{N+K-1} \\ 1 & \dots & 0 & 0 \\ 0 & \dots & 0 & 0 \\ \vdots & & & \\ 0 & \dots & 1 & 0, \end{pmatrix} \tag{1.15}$$

$$B_t = (m g_t \varepsilon_t^2, 0, \dots, 0)',$$

$$\tilde{c} = (c_1, c_2, \dots, c_{N+K-1})'.$$

Again, we are put in the setting of model (1.4) with strictly stationary ergodic coefficients. If we could find conditions to meet the assumptions in Theorem 3.1 of Glasserman and Yao Glasserman and Yao (1995), then model (1.15) will have a unique strictly stationary solution. But the problem is how to evaluate the top Lyapounov exponent associated with the stationary ergodic matrices. We approach this problem in three steps: (1) $K = 1, N = 1$ (2) $K = 1, N > 1$ (3) $K > 1$ and $N \geq 1$ due to the complicated structure of $A_t(\tilde{c})$.

When $KN = 1$, $A_t(\tilde{c})$ is just a scaler and the top Lyapounov exponent is easy to compute. The sufficient condition of stationary solution comes directly from Theorem 1 of Brandt (1986) or Theorem 3.1 of Glasserman and Yao (1995) (see Appendix).

Proposition 1.4. *When $KN = 1$, under the assumptions that $\alpha > 0$,$\beta > 0$, $\alpha + \beta < 1$, $\theta > 0$ and $m > 0$, model (1.12) has a unique strictly stationary ergodic solution if $\theta < 1$.*

Proof of Proposition 1.4. *When $KN = 1$, r_t^2 defined in model (1.12) is reduced to $r_t^2 = m g_t \varepsilon_t^2 + \theta g_t \varepsilon_t^2 r_{t-1}^2$. Notice that when $\alpha > 0, \beta > 0, \alpha + \beta < 1$, $\{g_t \varepsilon_t^2, t \in \mathbb{Z}\}$ is strictly stationary ergodic. If $0 < \theta < 1$,*

$$E \log(\theta g_0 \varepsilon_0^2) \leq \log E(\theta g_0 \varepsilon_0^2) = \log \theta < 0,$$

$$E \log(m g_0 \varepsilon_0^2) \leq \log E(m g_0 \varepsilon_0^2) = \log m < \infty$$

the conclusion follows from Theorem 1 of Brandt (1986) or Theorem 3.1 of Glasserman and Yao (1995) (see Appendix) directly.

When $K = 1$ and $N > 1$, the weight function vanishes and $A_t(\tilde{c})$ is simplified as

$$A_t(\tilde{1}) = \begin{pmatrix} \theta g_t \varepsilon_t^2 & \theta g_t \varepsilon_t^2 & \dots & \theta g_t \varepsilon_t^2 & \theta g_t \varepsilon_t^2 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & & & & \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix} \doteq A_t. \tag{1.16}$$

Introduce

$$H = \begin{pmatrix} 1 & 1 & \dots & 1 & 1 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ \vdots & & & & \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix}, G = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & & & & \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}$$

and define $M(a) = aH + G$, then

$$A_t = M(\theta g_t \varepsilon_t^2).$$

Matrix of this type is encountered a lot when one expresses an autoregressive model using a Markovian representation. The next lemma gives the basic properties of matrix $M(a)$ with a positive.

Lemma 1.3. *For matrix $M(a)$ with $a > 0$, we have the following properties:*

1. *Let $f(\lambda)$ be the characteristic function of $M(a)$. For any positive number k, $\rho(M(a)) < k$ if $f(k) > 0$;*

2. *$\rho(M(a))$ is increasing in a and it is a concave function of a.*

Proof of Lemma 1.3. *1. If $|\lambda| \geq k$,*

$$\begin{aligned} |f(\lambda)| &= |\lambda^N - a\lambda^{N-1} - a\lambda^{N-2} - \dots - a\lambda^2 - a\lambda - a| \\ &\geq |\lambda^N|(1 - \frac{a}{|\lambda|} - \frac{a}{|\lambda^2|} - \dots - \frac{a}{|\lambda^N|}) \\ &\geq k^N(1 - \frac{a}{k} - \frac{a}{k^2} - \dots - \frac{a}{k^N}) \\ &= f(k) > 0 \end{aligned}$$

therefore, $\rho(M(a)) < k$.

2. *Note* A *is nonnegative and irreducible. By Perron-Frobenius theory,* $\rho(A)$ *is the maximal positive root of* $f(\lambda) = det(\lambda I - A)$. *It is simple and* $\rho(A) \geq |\lambda|$ *for each root* λ *of* $f(\lambda) = 0$

   For ease of reference, we use $\lambda(a)$ *or* λ *for* $\rho(A)$. *Since* $f(\lambda) = \lambda^N - a\lambda^{N-1} - a\lambda^{N-2} - \ldots - a\lambda^2 - a\lambda - a = 0$,

$$a = \frac{\lambda^N}{\lambda^{N-1} + \lambda^{N-2} + \ldots + \lambda^2 + \lambda + 1} = \lambda - 1 + g(\lambda) \tag{1.17}$$

   where $g(\lambda) = \frac{1}{h(\lambda)}$ *and* $h(\lambda) = \lambda^{N-1} + \lambda^{N-2} + \ldots + \lambda^2 + \lambda + 1$

   Since λ *is a smooth function of* a, *to prove* λ *is a concave function of* a *is equivalent to show that* $\frac{d^2\lambda(a)}{da^2} < 0$. *Taking derivative on both sides of (1.17) with respect to* a, *we could have*

$$1 = (1 + g^{'})\lambda^{'} \tag{1.18}$$

   where $g^{'} = \frac{dg(\lambda)}{d\lambda}$ *and* $\lambda^{'} = \frac{d\lambda(a)}{da}$

   Furthermore

$$0 = (1 + g^{'})\lambda^{''} + g^{''}(\lambda^{'})^2 \tag{1.19}$$

   On the other hand, put $f(\lambda) = 0$ *as* $F(\lambda, a) = 0$. *By implicit function theorem,*

$$\lambda^{'} = -\frac{F_a}{F_\lambda}$$

   where $F_a = \frac{\partial F}{\partial a} = -h(\lambda) < 0$ *and* $F_\lambda > 0$ *(since* λ *is the largest root of* f *and* f *goes to* ∞ *as* λ *goes to* ∞ *for fixed* a*). Hence* $\lambda^{'} > 0$ *and* $1 + g^{'} > 0$.

   To show $\lambda^{''} < 0$, *it is sufficient to show that* $g^{''} = \frac{2(h^{'}(\lambda))^2 - h(\lambda)h^{''}(\lambda)}{h^3(\lambda)} > 0$ *or* $\Delta = 2(h^{'}(\lambda))^2 - h(\lambda)h^{''}(\lambda) > 0$.

   Note

$$\begin{aligned} h(\lambda) &= \frac{\lambda^N - 1}{\lambda - 1} \\ h^{'}(\lambda) &= \frac{N\lambda^{N-1}}{\lambda - 1} - \frac{\lambda^N - 1}{(\lambda - 1)^2} \end{aligned}$$

$$h^{''}(\lambda) = \frac{N(N-1)\lambda^{N-2}}{\lambda - 1} - \frac{2N\lambda^{N-1}}{(\lambda-1)^2} + \frac{2(\lambda^N - 1)}{(\lambda-1)^3}$$

therefore,

$$\Delta = \frac{N\lambda^{N-2}[(N-1)\lambda^{N+1} - (N+1)\lambda^N + (N+1)\lambda - (N-1)]}{(\lambda-1)^3} \tag{1.20}$$

Define

$$D(\lambda) = (N-1)\lambda^{N+1} - (N+1)\lambda^N + (N+1)\lambda - (N-1),$$

then

$$D^{'}(\lambda) = (N-1)(N+1)\lambda^N - (N+1)N\lambda^{N-1} + (N+1)$$

and

$$D^{''}(\lambda) = (N-1)N(N+1)\lambda^{N-2}(\lambda - 1).$$

Note $D(1) = D^{'}(1) = D^{''}(1) = 0$ *and* $D^{''} < 0$ *for* $0 < \lambda < 1$, *while on* $\lambda > 1$, $D^{''} > 0$, *which implies that* $D^{'} > 0$ *except* $\lambda = 1$. *Going one step further, we have* $D > 0$ *on* $\lambda > 1$ *and* $D < 0$ *on* $0 < \lambda < 1$, *which means* $\Delta > 0$ *on both* $\lambda > 1$ *and* $0 < \lambda < 1$. *By continuity,* $\Delta > 0$ *for* $\lambda > 0$. *It finishes the proof.*

Proposition 1.5. *For* $K = 1$ *and* $N > 1$, *if* $\beta^2 + 2\alpha\beta + 3\alpha^2 < 1$ *and* $\theta < \frac{\eta^{N-1}}{1+\eta+\ldots+\eta^{N-1}}$ *where* $\eta = \alpha + \beta$, *the top Lyapounov exponent* γ *associated with* A_t *(defined in (1.16)) is negative.*

Proof of Proposition 1.5. *Note when* $\beta^2 + 2\alpha\beta + 3\alpha^2 < 1$, *under Frobenius norm*

$$\begin{aligned} \|A_0\| &= \sqrt{tr(A_0^* A_0)} \\ &= \sqrt{N\theta^2 g_0^2 \varepsilon_0^4 + N - 1} \geq 1 \end{aligned}$$

$$\begin{aligned} E(\log\|A_0\|)^+ &= E\log\sqrt{N\theta^2 g_0^2 \varepsilon_0^4 + N - 1} \\ &\leq \frac{1}{2}\log[N\theta^2 E(g_0^2\varepsilon_0^4) + N - 1] < \infty \end{aligned}$$

Furthermore,

$$A_t = g_t(\theta\varepsilon_t^2 H + \frac{1}{g_t}G) \le g_t(\theta\varepsilon_t^2 H + \frac{1}{\eta}G).$$

Let $\tilde{A}_t \doteq \theta\varepsilon_t^2 H + \frac{1}{\eta}G$, *we have*

$$\|A_t A_{t-1} \dots A_0\| \le g_t g_{t-1} \dots g_0 \|\tilde{A}_t \tilde{A}_{t-1} \dots \tilde{A}_0\|.$$

It follows that

$$\gamma \le E \log g_0 + \lim_t \frac{1}{1+t} E \log \|\tilde{A}_t \tilde{A}_{t-1} \dots \tilde{A}_0\|.$$

Let $\tilde{\gamma}$ *be the top Lyapounov exponent associated with sequence* $\{\tilde{A}_t, t \in \mathbb{Z}\}$, *then* $\gamma \le \tilde{\gamma}$.

Since $\tilde{A}_t$*'s are iid and nonnegative, according to Lemma 1.1, if* $\rho[E(\tilde{A}_0)] < 1$, *then* $\gamma \le \tilde{\gamma} < 0$.

Note $E(\tilde{A}_0) = \frac{1}{\eta}M(\theta\eta)$ *and* $\rho[E(\tilde{A}_0)] < 1$ *is equivalent to* $\rho(M(\theta\eta)) < \eta$. *Its sufficient condition is*

$$f(\eta) = det(\eta I_N - M(\theta\eta)) > 0,$$

by Lemma 1.3 which is satisfied if $\theta < \frac{\eta^{N-1}}{1+\eta+\dots+\eta^{N-1}}$.

Proposition 1.6. *When* $K > 1$ *and* $N \ge 1$, *the top Lyapounov exponent associated with* $A_t(\tilde{c})$ *defined in (1.15) is negative if* $\beta^2 + 2\alpha\beta + 3\alpha^2 < 1$ *and* $\theta < \frac{\eta^{K+N-2}}{1+\eta+\dots+\eta^{K+N-2}}$ *where* $\eta = \alpha + \beta$.

Proof of Proposition 1.6. *Under the Frobenius norm,*

$$\begin{aligned}\|A_0\| &= \sqrt{tr(A_0^* A_0)} \\ &= \sqrt{\theta^2 g_0^2 \varepsilon_0^4 (c_1^2 + \dots + c_{N+K-1}^2) + N + K - 2} \ge 1.\end{aligned}$$

And when $\beta^2 + 2\alpha\beta + 3\alpha^2 < 1$,

$$\begin{aligned}E(\log \|A_0(\tilde{c})\|)^+ &= E \log \sqrt{\theta^2 g_0^2 \varepsilon_0^4 (c_1^2 + \dots + c_{N+K-1}^2) + N + K - 2} \\ &\le \frac{1}{2} \log[\theta^2 (c_1^2 + \dots + c_{N+K-1}^2) E(g_0^2 \varepsilon_0^4) + N + K - 2] < \infty\end{aligned}$$

The top Lyapounov exponent associated with $A_t(\tilde{c})$ *is*

$$\gamma(\tilde{c}) = \lim_{t\to\infty} \frac{1}{t+1} E \log \|A_t(\tilde{c})A_{t-1}((\tilde{c}))\dots A_0(\tilde{c})\|$$

Define

$$g_n(\tilde{c}) = \|A_t(\tilde{c})A_{t-1}(\tilde{c})\dots A_0(\tilde{c})\|^2$$

Since g_n *is a polynomial in* $\tilde{c}$ *and all the entries in the matrices are nonnegative, the coefficients of* c_j $(1 \le j \le K+N-1)$ *are positive which implies that, for every* n, $g_n(\tilde{c})$ *is nondecreasing in each* c_j. *In other words,* $g(\tilde{c}) \le g(\tilde{1})$. *It follows from Proposition 1.5 that*

$$\gamma(\tilde{c}) \le \gamma(\tilde{1}) < 0 \text{ if } \theta < \frac{\eta^{K+N-2}}{1+\eta+\dots+\eta^{K+N-2}}.$$

Combining the above results, we have

Proposition 1.7. *Suppose that* $\alpha > 0, \beta > 0$, $\alpha+\beta < 1$, $\theta > 0$, $m > 0$, *and* $KN > 1$. *The sufficient condition for the existence and uniqueness of a strictly stationary ergodic solution to model (1.12) is* $\beta^2 + 2\alpha\beta + 3\alpha^2 < 1$ *and* $\theta < \frac{\eta^{K+N-2}}{1+\eta+\dots+\eta^{K+N-2}}$ *where* $\eta = \alpha + \beta$.

Proof of Proposition 1.7. *Under the assumption* $\beta^2 + 2\alpha\beta + 3\alpha^2 < 1$,

$$\begin{aligned} E(\log \|A_0\|)^+ &= E \log \sqrt{\theta^2 g_0^2 \varepsilon_0^4 (c_1^2 + \dots + c_{N+K-1}^2) + N + K - 2} \\ &\le \frac{1}{2} \log[\theta^2 (c_1^2 + \dots + c_{N+K-1}^2) E(g_0^2 \varepsilon_0^4) + N + K - 2] < \infty \\ E(\log \|B_0\|)^+ &= E(\log m + \log g_0 + \log \varepsilon_0^2)^+ < \infty \end{aligned}$$

Further $\gamma < 0$ *is derived from Proposition 1.5 and 1.6 if* $\theta < \frac{\eta^{K+N-2}}{1+\eta+\dots+\eta^{K+N-2}}$. *Applying Theorem 3.1 of Glasserman and Yao (1995) (see Appendix), there exists a unique strictly stationary ergodic solution to model (1.12).*

Corollary 1.1. *The GARCH-MIDAS model with rolling window RV (with* $\mu = 0$*) has a unique strictly stationary ergodic solution if*

1. $\theta < 1$ *when* $KN = 1$
2. *or* $\theta < \frac{\eta^{K+N-2}}{1+\eta+\dots+\eta^{K+N-2}}$ *and* $\beta^2 + 2\alpha\beta + 3\alpha^2 < 1$ *when* $KN > 1$.

The resulting process is nonanticipative (or causal). In addition, the low-frequency volatility component τ is strictly stationary ergodic as well.

1.4 Asymptotic properties of GARCH-MIDAS model

The last property in section 1.3 tells us that GARCH-MIDAS model with rolling-window RV (1.12) has a unique strictly stationary ergodic solution under certain regularity conditions. In this section, we will follow this line and study the consistency and asymptotic behavior of maximum likelihood estimates (MLE) of this model.

The parameter space we will consider in this section is

$$\begin{aligned}\mathcal{U} = &\{\Phi = (\alpha, \beta, m, \theta, \omega)' \in \mathcal{R}^5 : \alpha > 0, \beta > 0, m > 0 \\ &(\alpha+\beta)^2 + 2\alpha^2 < 1, 0 < \theta < \frac{\eta^{K+N-2}}{1+\eta+\ldots+\eta^{K+N-2}}\}\end{aligned}$$

Suppose that $\Phi_0 = (\alpha_0, \beta_0, m_0, \theta_0, \omega_0)$ is the true parameter such that $\Phi_0 \in \mathcal{U}$. Given a sequence of $\{r_t, 1 \le t \le T\}$ where $T \gg N + K$ which are generated by the following dynamics

$$\begin{aligned} r_t &= \sqrt{g_t(\Phi_0)\tau_t(\Phi_0)}\varepsilon_t, \quad t \in \mathbb{Z} \\ g_t(\Phi_0) &= (1-\alpha_0-\beta_0) + \alpha_0 \tfrac{r_{t-1}^2}{\tau_{t-1}(\Phi_0)} + \beta_0 g_{t-1}(\Phi_0) \\ \tau_t(\Phi_0) &= m_0 + \theta_0 \textstyle\sum_{k=1}^K \varphi_k(\omega_0) RV_{t-k}, RV_t = \sum_{j=0}^{N-1} r_{t-j}^2 \end{aligned} \tag{1.21}$$

the MLE of Φ_0 (denoted as $\hat{\Phi}_T$) is the minimizer of

$$L_T(\Phi) = \frac{1}{T}\sum_{t=N+K}^{T}[\log g_t(\Phi) + \log \tau_t(\Phi) + \frac{r_t^2}{g_t(\Phi)\tau_t(\Phi)}]$$

For ease of reference, we use $\{\phi_i, 1 \le i \le 5\}$ to refer to the parameter set $\{\alpha, \beta, m, \theta, \omega\}$ when there is no confusion. Introduce

$$l_t(\Phi) \equiv \log g_t(\Phi) + \log \tau_t(\Phi) + \frac{r_t^2}{g_t(\Phi)\tau_t(\Phi)}.$$

The gradient of $L_T(\Phi)$ is

$$\nabla L_T(\Phi) = \frac{1}{T}\sum_{t=N+K}^{T} \nabla l_t(\Phi) = \frac{1}{T}\sum_{t=N+K}^{T} (s_t^{\alpha}, s_t^{\beta}, s_t^{m}, s_t^{\theta}, s_t^{\omega})'(\Phi)$$

with $s_t^{\phi_i}(\Phi) = \frac{\partial l_t(\Phi)}{\partial \phi_i}, i = 1, \ldots, 5$. The hessian matrix of $L_T(\Phi)$ is

$$H(L_T)(\Phi) = (\frac{\partial^2 L_T(\Phi)}{\partial \phi_i \partial \phi_j})_{1 \leq i,j \leq 5} = \frac{1}{T}\sum_{t=N+K}^{T} H(l_t)(\Phi) \tag{1.22}$$

As a convention, if a function is expressed without specifying Φ, we assume that it is evaluated at the true parameter Φ_0.

The following main result establishes the existence and uniqueness of the consistent and asymptotically normal estimator $\hat{\Phi}_T$.

Proposition 1.8. *Assume $\{r_t, 1 \leq t \leq T\}$ is generated from model (1.21) with $\Phi_0 \in \mathcal{U}$. Then there exists a fixed open neighborhood $N(\Phi_0) \subset \overline{N(\Phi_0)} \subset \mathcal{U}$ of Φ_0 such that with probability tending to 1 as T goes to ∞, $L_T(\Phi)$ has a unique minimum $\hat{\Phi}_T$ in $N(\Phi_0)$ such that*

$$\hat{\Phi}_T \xrightarrow{P} \Phi_0$$

and

$$\sqrt{T}(\hat{\Phi}_T - \Phi_0) \Rightarrow N(0, \Sigma_I^{-1}\Sigma_S\Sigma_I^{-1})$$

where $\Sigma_I = E(H(l_1)), \Sigma_S = E(\nabla l_1 \nabla l_1')$.

The next proposition gives the consistent estimate of the asymptotic covariance matrix $\Sigma_I^{-1}\Sigma_S\Sigma_I^{-1}$.

Proposition 1.9. *With the same regularity conditions as Proposition 1.8, we have*

$$\hat{\Sigma}_{I(T)}^{-1}\hat{\Sigma}_{S(T)}\hat{\Sigma}_{I(T)}^{-1} \xrightarrow{P} \Sigma_I^{-1}\Sigma_S\Sigma_I^{-1}$$

where $\hat{\Sigma}_{I(T)} = \frac{1}{T}\sum_{t=N+K}^{T} H(l_t)(\hat{\Phi}_T)$ and $\hat{\Sigma}_{S(T)} = \frac{1}{T}\sum_{t=N+K}^{T} \nabla l_t \nabla l_t'(\hat{\Phi}_T)$.

1.4.1 Proofs of Proposition 1.8 and Proposition 1.9

To establish the consistency and asymptotic normality of $\hat{\Phi}_T$, we need the following helpful lemmas.

Lemma 1.4. *Let $\{X_n, \mathcal{F}_n : n \geq 1\}$ be a strictly stationary ergodic martingale difference sequence such that $\sigma^2 = E(X_1^2) < \infty$. Then*

$$\frac{1}{\sqrt{n}} \sum_{j=1}^{n} X_j \Rightarrow N(0, \sigma^2).$$

Proof of Lemma 1.4. *Define $X_{nj} = \frac{X_j}{\sigma\sqrt{n}}$, $1 \leq j \leq n$. Note for any ε,*

$$P(\max_{j\leq n} |X_{nj}| > \varepsilon) \leq P(\sum_{j\leq n} X_{nj}^2 I(|X_{nj}| > \varepsilon) > \varepsilon^2) \leq \frac{1}{\varepsilon^2} E(\sum_{j\leq n} X_{nj}^2 I(|X_{nj}| > \varepsilon))$$

$$\max_{j\leq n} X_{nj}^2 \leq \varepsilon^2 + \sum_{j\leq n} X_{nj}^2 I(|X_{nj}| > \varepsilon)$$

Since

$$\begin{aligned} E(\sum_{j\leq n} X_{nj}^2 I(|X_{nj}| > \varepsilon)) &= \frac{1}{\sigma^2 n} \sum_{j\leq n} E(X_j^2 I(|X_j| > \varepsilon\sigma\sqrt{n})) \\ &= \frac{1}{\sigma^2} E(X_1^2 I(|X_1| > \varepsilon\sigma\sqrt{n})) \to 0 \text{ as } n \to \infty \end{aligned}$$

due to the fact that $P\{|X_1| > \varepsilon\sigma\sqrt{n}\} \to 0$ as $n \to \infty$. Therefore $\{\max_{j\leq n} |X_{nj}|\}$ is uniformly bounded in L_2 norm and $\max_{j\leq n} |X_{nj}| \xrightarrow{P} 0$.

Note also that

$$\sum_j X_{nj}^2 = \frac{1}{\sigma^2 n} \sum_j X_j^2 \to 0 \text{ almost surely, as } n \to \infty$$

by Birkhoff's Ergodic Theorem. It follows from martingale central limit theorem of McLeish (1974) that $\frac{1}{\sqrt{n}} \sum_{j=1}^{n} X_j \Rightarrow N(0, \sigma^2)$.

Lemma 1.4 presents a fact for a one-dimensional situation. To extend it to higher dimensions, we need to use Cramér-Wold Device of Bilingsley (1995). Moreover, we could derive the following result.

Lemma 1.5. *Under the assumptions in Proposition 1.8,*

$$\sqrt{T}\nabla L_T(\Phi_0) \Rightarrow N(0, \Sigma_S)$$

where

$$\Sigma_S = E(\nabla l_1 \nabla l_1^{'}) = \begin{pmatrix} E(s_1^{\alpha 2}) & E(s_1^{\alpha} s_1^{\beta}) & E(s_1^{\alpha} s_1^{m}) & E(s_1^{\alpha} s_1^{\theta}) & E(s_1^{\alpha} s_1^{\omega}) \\ * & E(s_1^{\beta 2}) & E(s_1^{\beta} s_1^{m}) & E(s_1^{\beta} s_1^{\theta}) & E(s_1^{\beta} s_1^{\omega}) \\ * & * & E(s_1^{m2}) & E(s_1^{m} s_1^{\theta}) & E(s_1^{m} s_1^{\omega}) \\ * & * & * & E(s_1^{\theta 2}) & E(s_1^{\theta} s_1^{\omega}) \\ * & * & * & * & E(s_1^{\omega 2}) \end{pmatrix} \quad (1.23)$$

Remark 1.1. Σ_S *is symmetric. We only display its upper triangular part here for brevity. In the rest of chapter, we will express a symmetric matrix this way.*

Proof of Lemma 1.5. *According to Cramér-Wold Device, it is sufficient to show that for any* $t = (t_1, t_2, t_3, t_4, t_5)^{'} \in \mathbb{R}^5$,

$$\sqrt{T}t^{'}\nabla L_T(\Phi_0) \Rightarrow t^{'}Z$$

where $Z \sim N(0, \Sigma_S)$.

Notice that

$$\sqrt{T}t^{'}\nabla L_T(\Phi_0) = \frac{1}{\sqrt{T}} \sum_{t=N+K}^{T} t_1 s_t^{\alpha} + t_2 s_t^{\beta} + t_3 s_t^{m} + t_4 s_t^{\theta} + t_5 s_t^{\omega}.$$

Let

$$s_t = t_1 s_t^{\alpha} + t_2 s_t^{\beta} + t_3 s_t^{m} + t_4 s_t^{\theta} + t_5 s_t^{\omega}.$$

The strictly stationary ergodic solutions g_t *and* r_t *are measurable functions of* $\{\varepsilon_j : 1 \leq j \leq t\}$, *so is* τ_t. *It follows that* s_t^{α},s_t^{β},s_t^{m},s_t^{θ},s_t^{ω} *are also measurable functions of* $\{\varepsilon_j : 1 \leq j \leq t\}$.

Therefore $\{s_t\}$ is a strictly stationary and ergodic process (due to Stout (1974)).

$$\begin{aligned} E(s_t|\mathcal{F}_{t-1}) &= t_1E(s_t^{\alpha}|\mathcal{F}_{t-1}) + t_2E(s_t^{\beta}|\mathcal{F}_{t-1}) + t_3E(s_t^{m}|\mathcal{F}_{t-1}) \\ &\quad +E(s_t^{\theta}|\mathcal{F}_{t-1}) + E(s_t^{\omega}|\mathcal{F}_{t-1}) \\ &= 0 \\ E(s_t^2) &\leq t_1^2E(s_t^{\alpha 2}) + t_2^2E(s_t^{\beta 2}) + t_3^2E(s_t^{m2}) + t_4^2E(s_t^{\theta 2}) + t_5^2E(s_t^{\omega 2}) \end{aligned} \tag{1.24}$$

Now we need to show that $E(s_t^2) < \infty$. Since

$$\frac{\partial l_t(\Phi)}{\partial \phi_i} = (1 - \frac{r_t^2}{g_t(\Phi)\tau_t(\Phi)})(\frac{\partial \tau_t/\partial \phi_i}{\tau_t}(\Phi) + \frac{\partial g_t/\partial \phi_i}{g_t}(\Phi)),$$

evaluated at the true parameters

$$s_t^{\phi_i} = \frac{\partial l_t}{\partial \phi_i} = (1-\varepsilon_t^2)(\frac{\partial \tau_t/\partial \phi_i}{\tau_t} + \frac{\partial g_t/\partial \phi_i}{g_t}), i = 1, \ldots, 5.$$

Note also

$$s_t^{\phi_i 2} \leq 2(1-\varepsilon_t^2)^2[(\frac{\partial \tau_t/\partial \phi_i}{\tau_t})^2 + (\frac{\partial g_t/\partial \phi_i}{g_t})^2].$$

For $i = 1$, ie $\phi_1 = \alpha$, $\frac{\partial \tau_t}{\partial \alpha} = 0$, and $\frac{\partial g_t}{\partial \alpha} = \sum_{j=1}^{\infty} \beta^{j-1} g_{t-j}\varepsilon_{t-j}^2 - \frac{1}{1-\beta}$, we have

$$E(s_t^{\alpha 2}) \leq \frac{2E(1-\varepsilon_t^2)^2}{(1-\alpha_0-\beta_0)^2}[(\frac{1}{(1-\beta_0)^2}) + \sum_{j=1}^{\infty} \beta_0^{2j-2} E g_1^2 E\varepsilon_1^4] < \infty.$$

For $i = 2$, ie $\phi_2 = \beta$, $\frac{\partial \tau_t}{\partial \beta} = 0$, and $\frac{\partial g_t}{\partial \beta} = \sum_{j=1}^{\infty} \beta^{j-1} g_{t-j} - \frac{1}{1-\beta}$, we have

$$E(s_t^{\beta 2}) \leq \frac{2E(1-\varepsilon_t^2)^2}{(1-\alpha_0-\beta_0)^2}[(\frac{1}{(1-\beta_0)^2}) + \sum_{j=1}^{\infty} \beta_0^{2j-2} E g_1^2 E\varepsilon_1^4] < \infty.$$

For $i = 3$, ie $\phi_3 = m$, $\frac{\partial \tau_t/\partial m}{\tau_t} = \frac{1}{\tau_t}$, $\frac{\partial g_t}{\partial m} = -\alpha_0 g_{t-1}\varepsilon_{t-1}^2(\frac{\partial \tau_{t-1}/\partial m}{\tau_{t-1}}) = -\alpha_0 g_{t-1}\varepsilon_{t-1}^2(\frac{1}{\tau_{t-1}})$, we have

$$E(s_t^{m2}) \leq 2E(1-\varepsilon_t^2)^2[\frac{1}{m_0^2} + \frac{E(g_1^2)E(\varepsilon_1^4)}{(1-\alpha_0-\beta_0)^2 m^2}] < \infty.$$

For $i=4$, *ie* $\phi_4=\theta$, $\frac{\partial\tau_t/\partial\theta}{\tau_t}=\frac{1}{\theta_0}(1-\frac{m_0}{\tau_t})$, $\frac{\partial g_t}{\partial\theta}=-\alpha_0 g_{t-1}\varepsilon_{t-1}^2(\frac{\partial\tau_{t-1}/\partial\theta}{\tau_{t-1}})$, *we have*

$$E(s_t^{\theta 2})\leq 2E(1-\varepsilon_t^2)^2[\frac{1}{\theta_0^2}+\frac{E(g_1^2)E(\varepsilon_1^4)}{(1-\alpha_0-\beta_0)^2\theta_0^2}]<\infty.$$

For $i=5$, *ie* $\phi_5=\omega$, $\frac{\partial\tau_t/\partial\omega}{\tau_t}=\frac{\sum\varphi_k^{'}RV_{t-k}}{\tau_t}\leq\frac{\max_k\varphi_k^{'}(\omega_0)}{\min_k\varphi_k(\omega_0)}$ *(without loss of generality, we could assume* $\{\varphi_k,1\leq k\leq K\}$ *are all positive) and* $\frac{\partial g_t}{\partial\omega}=-\alpha_0 g_{t-1}\varepsilon_{t-1}^2(\frac{\partial\tau_{t-1}/\partial\omega}{\tau_{t-1}})$. *We have*

$$E(s_t^{\omega 2})\leq 2E(1-\varepsilon_t^2)^2[1+\frac{E(g_1^2)E(\varepsilon_1^4)}{(1-\alpha_0-\beta_0)^2}](\frac{\max_k\varphi_k^{'}(\omega_0)}{\min_k\varphi_k(\omega_0)})^2<\infty.$$

Therefore $E(s_1^2)<\infty$. *Applying Lemma 1.4, we get*

$$\sqrt{T}t^{'}\frac{\partial}{\partial\phi}L_T(\Phi_0)\Rightarrow N(0,t^{'}\Omega t)\quad\forall t\in\mathbb{R}^5.$$

The following lemma evaluates the probabilistic property of the Hessian matrix of L_T with value taken at $\Phi=\Phi_0$.

Lemma 1.6. *Under the assumptions in Proposition 1.8*

$$H(L_T)(\Phi_0)\xrightarrow{P}\Sigma_I$$

where

$$\Sigma_I=E(H(l_1))=\begin{pmatrix} E(\frac{\partial^2 l_1(\Phi_0)}{\partial\alpha^2}) & E(\frac{\partial^2 l_1(\Phi_0)}{\partial\alpha\partial\beta}) & E(\frac{\partial^2 l_1(\Phi_0)}{\partial\alpha\partial m}) & E(\frac{\partial^2 l_1(\Phi_0)}{\partial\alpha\partial\theta}) & E(\frac{\partial^2 l_1(\Phi_0)}{\partial\alpha\partial\omega}) \\ * & E(\frac{\partial^2 l_1(\Phi_0)}{\partial\beta^2}) & E(\frac{\partial^2 l_1(\Phi_0)}{\partial\beta\partial m}) & E(\frac{\partial^2 l_1(\Phi_0)}{\partial\beta\partial\theta}) & E(\frac{\partial^2 l_1(\Phi_0)}{\partial\beta\partial\omega}) \\ * & * & E(\frac{\partial^2 l_1(\Phi_0)}{\partial m^2}) & E(\frac{\partial^2 l_1(\Phi_0)}{\partial m\partial\theta}) & E(\frac{\partial^2 l_1(\Phi_0)}{\partial m\partial\omega}) \\ * & * & * & E(\frac{\partial^2 l_1(\Phi_0)}{\partial\theta^2}) & E(\frac{\partial^2 l_1(\Phi_0)}{\partial\theta\partial\omega}) \\ * & * & * & * & E(\frac{\partial^2 l_1(\Phi_0)}{\partial\omega^2}) \end{pmatrix} \tag{1.25}$$

Proof of Lemma 1.6. *Introduce* $D_T=(d_{i,j}^T)_{1\leq i,j\leq 5}=H(L_T)(\Phi_0)$ *and each element in* Σ_I *is denoted by* σ_{ij}^2. *We need to show that*

$$\lim_{T\to\infty}P(\|D_T-\Sigma_I\|>\epsilon)=0\quad\forall\epsilon>0$$

where $\|.\|$ is an arbitrary matrix norm.

All norms on the finite dimensional norm space are equivalent, which implies that all the matrix norms on $\mathbb{C}^{n\times n}$ should be equivalent. Thus, we only need to show the result is true for Frobenious norm. Under Frobenious norm,

$$\|D_T - \Sigma_I\|^2 = trace[(D_T - \Sigma_I)^*(D_T - \Sigma_I)] = \sum_{i,j=1}^{5} (d_{i,j}^T - \sigma_{i,j}^2)^2$$

Note

$$d_{i,j}^T = \frac{1}{T} \sum_{t=N+K}^{T} \frac{\partial^2 l_t(\Phi_0)}{\partial \phi_i \partial \phi_j},$$

and $\frac{\partial^2 l_t(\Phi_0)}{\partial \phi_i \partial \phi_j}$ is a measurable function of $\{\varepsilon_s, s \leq t\}$, hence is strictly stationary ergodic. By Birkhoff's ergodic theorem,

$$d_{i,j}^T \xrightarrow{P} \sigma_{i,j}^2$$

ie.

$$P(\|D_T - \Sigma_I\| > \epsilon) \leq \sum_{i,j=1}^{5} P(|d_{i,j}^T - \sigma_{i,j}^2| > \frac{\epsilon}{\sqrt{5}}) \to 0.$$

Therefore,

$$D_T \xrightarrow{P} \Sigma_I.$$

Next, we want to show the third derivatives of L_T is locally bounded in a 'weak' sense, i.e.,

Lemma 1.7. *Let $N(\Phi_0)$ be an arbitrary open set of Φ_0 such that $N(\Phi_0) \subset \overline{N(\Phi_0)} \subset \mathcal{U}$. Then there exists a random variable c_T which satisfies*

$$\max_{i,j,k=1,\ldots,5} \sup_{\Phi \in N(\Phi_0)} |\frac{\partial^3 L_T(\Phi)}{\partial \phi_i \partial \phi_j \partial \phi_k}| \leq c_T$$

and

$$c_T \xrightarrow{P} c \text{ for some constant } c.$$

Proof of Lemma 1.7.

$$\frac{\partial^3 L_T(\Phi)}{\partial \phi_i \partial \phi_j \partial \phi_k} = \frac{1}{T} \sum_{t=N+K}^{T} \frac{\partial^3 l_t(\Phi)}{\partial \phi_i \partial \phi_j \partial \phi_k}$$

Note $|\frac{\partial^3 l_t(\Phi)}{\partial\phi_i\partial\phi_j\partial\phi_k}|$ *is continuous in* Φ*, there exists an open neighborhood* $N(\Phi_0)$ *of* Φ_0 *such that* $\overline{N(\Phi_0)} \subset \mathcal{U}$ *and further, there exists a point* $\tilde{\Phi}_t^{i,j,k} \in \overline{N(\Phi_0)}$ *such that*

$$|\frac{\partial^3 L_T(\Phi)}{\partial\phi_i\partial\phi_j\partial\phi_k}| \leq \frac{1}{T}\sum_{t=N+K}^{T} w_t^{i,j,k}$$

where

$$w_t^{i,j,k} = |\frac{\partial^3 l_t(\tilde{\Phi}_t^{i,j,k})}{\partial\phi_i\partial\phi_j\partial\phi_k}|.$$

Therefore

$$\max_{i,j,k=1,\ldots,5} \sup_{\phi\in N(\phi_0)} |\frac{\partial^3 L_T(\Phi_0)}{\partial\phi_i\partial\phi_j\partial\phi_k}| \leq \frac{1}{T}\sum_{t=N+K}^{T}(\sum_{i,j,k=1}^{5} w_t^{i,j,k}).$$

Further, let $w_t = \sum_{i,j,k=1}^{5} w_t^{i,j,k}$*. Since* $\{w_t\}$ *is a strictly stationary ergodic sequence, by Birkhoff's ergodic theorem,*

$$\frac{1}{T}\sum_{t=N+K}^{T} w_t \xrightarrow{P} E(w_1).$$

With the above established results, we can complete the proof of Proposition 1.8.

Proof of Proposition 1.8. *Combine the lemmas 1.5, 1.6, and 1.7, and apply Lemma 1 of Jensen and Rahbek (2004) (see Appendix). The existence and uniqueness of the consistent and asymptotic normal estimator* $\hat{\Phi}_T$ *are ensured.*

The proof of Proposition 1.9 needs one more lemma.

Lemma 1.8. *Let* $\{x_n(\theta), n = 1, 2, \ldots\}$ *be a sequence of random variables defined on probability space* $\{\Omega, \mathcal{F}, P\}$ *such that* x_n *is uniformly continuous in* θ *and for each fixed* θ*,* $x_n(\theta) \xrightarrow{P} x(\theta)$*. Suppose that* $\hat{\theta}_n \xrightarrow{P} \theta_0$*, then*

$$x_n(\hat{\theta}_n) \xrightarrow{P} x(\theta_0).$$

Proof of Lemma 1.8. *For any* $\varepsilon > 0$,

$$\begin{aligned} P(|x_n(\hat{\theta}_n) - x(\theta_0)| > \varepsilon) \leq & P(|x_n(\theta_0) - x(\theta_0)| > \varepsilon) + P(|\hat{\theta}_n - \theta_0| > \varepsilon) \\ & + P(|x_n(\hat{\theta}_n) - x_n(\theta_0)| > \varepsilon, |\hat{\theta}_n - \theta_0| < \varepsilon) \end{aligned}$$

The conclusion follows from the inequality immediately.

Proof of Proposition 1.9. *For each* $\Phi \in \overline{N(\Phi_0)}$, $H(l_t)(\Phi), \nabla l_t \nabla l_t^{'}(\Phi)$ *are strictly stationary ergodic,*

$$\frac{1}{T}\sum H(l_t)(\Phi) \xrightarrow{P} E(H(l_1)(\Phi)),$$

and

$$\frac{1}{T}\sum \nabla l_t \nabla l_t^{'}(\Phi) \xrightarrow{P} E(\nabla l_1 \nabla l_1^{'}(\Phi))$$

due to Birkhoff's ergodic theorem. Also consider the fact that $\hat{\Phi}_T \xrightarrow{P} \Phi$, *and* $H(l_t)(\Phi), \nabla l_t \nabla l_t^{'}(\Phi)$ *are uniformly continuous in* $\Phi \in \overline{N(\Phi_0)}$. *Therefore,*

$$\frac{1}{T}\sum H(l_t)(\hat{\Phi}_T) \xrightarrow{P} E(H(l_1))$$

and

$$\frac{1}{T}\sum \nabla l_t \nabla l_t^{'}(\hat{\Phi}_T) \xrightarrow{P} E(\nabla l_1 \nabla l_1^{'}).$$

Applying continuous mapping theorem,

$$\hat{\Sigma}_{I(T)}^{-1}\hat{\Sigma}_{S(T)}\hat{\Sigma}_{I(T)}^{-1} \xrightarrow{P} \Sigma_I^{-1}\Sigma_S\Sigma_I^{-1}.$$

1.5 Conclusion

This chapter focused on the distributional properties of two volatility component models: the restricted GARCH(2,2) model of Engle and Lee, the GARCH-MIDAS model of Engle, Ghysels and Sohn. We presented necessary conditions under which these models were able to characterize the nonstationarity of the financial returns. The restricted GARCH(2,2) model structured the conditional variance as the sum of low-frequency and high-frequency stochastic components. It was shown that, under certain regularity conditions on the parameter space, it was strictly stationary ergodic and β-mixing. In the GARCH-MIDAS model, the conditional volatility was characterized as the multiplicative effects of low-frequency and high-frequency stochastic components. It was an extension of Spline-GARCH model of Engle and Rangel where the low-frequency volatility was fitted by an exponential quadratic spline, a deterministic structure. For GARCH-MIDAS model with fixed time span realized volatility, we showed that it could admit a covariance stationary solution in a specific parameter space. We also derived sufficient conditions for the existence and uniqueness of strictly stationary ergodic solution to the GARCH-MIDAS model with rolling window realized volatility. Further, this chapter showed that its maximum likelihood estimates were locally consistent and asymptotically normal. Its asymptotic variance-covariance matrix and associated consistent estimate were also specified.

CHAPTER 2

The Normal Inverse Gaussian Distribution and the Pricing of Derivatives

2.1 Introduction

In an arbitrage-free world the price of a derivative contract is the discounted expectation of the future payoff under a so-called risk neutral measure. Hence, the pricing formula has three key ingredients: the risk free rate, the contract specification - i.e. payoff function, and the data generating process of the underlying asset. This chapter pertains to the latter, namely the specification of the risk neutral distribution (henceforth RND) of the future prices of an underlying asset.

Several approaches have been developed to characterize or estimate the risk neutral probability measure in literature. Broadly speaking they can be characterized as: (1) modelling the shape of the RND directly [See Jackwerth and Rubinstein (1996), Melick and Thomas (1997), Rubinstein (1994), Figlewski and Gao (1999), among others], (2) differentiating the pricing function twice with respect to strike price to arrive at the RND of the underlying [see Breeden and Litzenberger (1978), Aït-Sahalia and Lo (1998), Longstaff (1995), among others], or (3) specifying a parametric stochastic process driving the price of the underlying asset and the change of probability measure [see Bates (1991), Bates (1996), Chernov and Ghysels (2000), among others]. These approaches range from purely nonparametric (e.g. Rubinstein (1994), Aït-Sahalia and Lo (1998)) to parametric (all papers cited above in (3)). For a more recent comprehensive literature review, see e.g. Figlewski (2007).

We suggest a flexible class of densities combined with data-driven moment estimators, i.e.

option-based estimators for variance, skewness and kurtosis. Our approach has several advantages. Purely nonparametric techniques are flexible and robust, yet they typically are extremely data intensive as they try to capture the entire shape of an unknown density [See Aït-Sahalia and Lo (1998), Pagan (1999), Broadie et al. (2000), Ghysels et al. (1997)]. We only need good estimates of the variance, skewness and kurtosis - which can be obtained from options data as suggested by Bakshi et al. (2003). With the first four moments (the mean is determined by the risk free rate) available we propose a method to obtain directly the risk neutral probability measure. Our approach is most directly related to some existing approaches. One consists of modelling the shape of the risk neutral density directly via Gram-Charlier series expansions (henceforth GCSE). There are two types of Gram-Charlier series expansion discussed in literature: A-type GCSE, applied in the context of derivative pricing by Madan and Milne (1994) and C-type GCSE applied to option pricing by Rompolis and Tzavalis (2007). Also related to GCSE is the Edgeworth expansion[1] - applied to reconstruct risk neutral densities by Rubinstein (1998).

We adopt an approach, suggested in a different context by Eriksson et al. (2004), and use the class of Normal Inverse Gaussian (henceforth NIG) densities to approximate an unknown RND. The appeal of NIG distributions is that they are characterized by the first four moments: mean, variance, skewness and kurtosis. These are the moments we care about in many risk management applications - including derivative pricing. Hence, once the four moments are given, we can fill in the blanks with the NIG and obtain the entire distribution.[2]

The use of the NIG family has several advantages over A-GCSE and C-GCSE. A-type Gram-Charlier expansion can result in negative probabilities with unsuitable - or as it is often called infeasible outside the domain of positive definiteness - combinations of skewness and kurtosis. Using recent empirical evidence from Conrad et al. (2007a) we find that for most traded options in the US feature skewness and kurtosis outside the admissible region for A-GCSE. We show this translates into serious errors for pricing derivatives. The C-GCSE approximation - on the other hand - yields nonnegative probabilities, yet is very cumbersome in terms of computations. The NIG class is - in comparison - easy to compute and is a proper density. Moreover, the NIG

[1] See Appendix A.2 for more details on the Gram-Charlier expansion and the Edgeworth expansion.

[2] The NIG family of distribution has recently also been suggested to price synthetic CDO contracts, see Kalemanova et al. (2007) and further references therein.

family has the nice properties that it is flexible and the parameters can be solved in a closed form by means of cumulants of the distribution, which facilitates parameter estimation.

The rest of this chapter is structured as follows. In section 2.2 we briefly review the NIG class of densities, and present the main results obtained by using a method of moments estimation approach. Section 2.3 describes the NIG approximation errors by comparing them with other approximation methods focusing on regions of unimodality and positive definiteness. Section 2.4 appraises via a calibration exercise the NIG density when used for pricing derivatives. We also provide an empirical illustration in section 2.5, while section 2.6 concludes this chapter.

2.2 The NIG class of distributions: Properties and option-based estimation

The Normal Inverse Gaussian (henceforth NIG) distribution is characterized via a normal inverse Gaussian mixing distribution. Formally stated, let Y be a random variable that follows an inverse Gaussian probability law (IG) discussed in Seshadri (1993):

$$\mathcal{L}(Y) = IG(\delta, \sqrt{\alpha^2 - \beta^2})$$

Furthermore, if X conditional on Y is normally distributed with mean $\mu + \beta Y$ and variance Y, namely $\mathcal{L}(X|Y) = N(\mu + \beta Y, Y)$, then the unconditional density X is NIG:

$$\mathcal{L}(X) = NIG(\alpha, \beta, \mu, \delta).$$

The density function for the NIG family is defined as follows:

$$f_{NIG}(x; \alpha, \beta, \mu, \delta) = \frac{\alpha}{\pi} \exp(\delta\sqrt{\alpha^2 - \beta^2} - \beta\mu) \frac{K_1(\alpha\delta\sqrt{1 + (\frac{x-\mu}{\delta})^2})}{\sqrt{1 + (\frac{x-\mu}{\delta})^2}} \exp(\beta x) \tag{2.1}$$

where $x \in \mathbb{R}$, $\alpha > 0$, $\delta > 0$, $\mu \in \mathbb{R}$, $0 < |\beta| < \alpha$ and $K_1(.)$ is the modified Bessel function of the third kind with index 1 (see Abramowitz (1974)). The Gaussian distribution is obtained as a limiting case, namely when $\alpha \to \infty$.

The NIG class of densities has the following two properties, namely (1) a scaling property:

$$\mathcal{L}_{NIG}(X) = NIG(\alpha, \beta, \mu, \delta) \Leftrightarrow \mathcal{L}_{NIG}(cX) = NIG(\alpha/c, \beta/c, c\mu, c\delta),$$

and (2) a closure under convolution property:

$$NIG(\alpha, \beta, \mu_1, \delta_1) * NIG(\alpha, \beta, \mu_2, \delta_2) = NIG(\alpha, \beta, \mu_1 + \mu_2, \delta_1 + \delta_2).$$

Another parameterization used in this chapter is obtained by setting $\bar{\alpha} = \delta\alpha$ and $\bar{\beta} = \delta\beta$. This representation is a scale-invariant parameterization denoted as $\overline{NIG}(\bar{\alpha}, \bar{\beta}, \mu, \delta)$ with density:

$$f_{\overline{NIG}}(x; \bar{\alpha}, \bar{\beta}, \mu, \delta) = \frac{\bar{\alpha}}{\pi\delta} \exp(\sqrt{\bar{\alpha}^2 - \bar{\beta}^2} - \frac{\bar{\beta}\mu}{\delta}) \frac{K_1(\bar{\alpha}\sqrt{1 + (\frac{x-\mu}{\delta})^2})}{\sqrt{1 + (\frac{x-\mu}{\delta})^2}} \exp(\frac{\bar{\beta}}{\delta} x) \tag{2.2}$$

2.2.1 Moment estimators for the NIG class of densities

The method of moments estimation applied to the NIG class consists of constructing a non-linear system of equations for the four parameters in the NIG distribution. In particular, one sets the first and second cumulant, the skewness and the excess kurtosis equal to their empirical counterparts.

The following two theorems, taken from Eriksson et al. (2004), yield the expression for the parameters in the class of NIG probability distributions in terms of its mean, variance, skewness and excess kurtosis.

Theorem 2.1. *Suppose that random variable X is $\overline{NIG}(\bar{\alpha}, \bar{\beta}, \mu, \delta)$ distributed and its mean, variance, skewness and excess kurtosis are denoted as $\mathcal{M}$, $\mathcal{V}$, $\mathcal{S}$ and $\mathcal{K}$, respectively. Then the parameters are related to the moments by*

$$\bar{\alpha} = 3(4\rho^{-1} + 1)(1 - \rho^{-1})^{-1/2}\mathcal{K}^{-1} \tag{2.3}$$

$$\bar{\beta} = sgn(\mathcal{S}) \left\{ 3(4\rho^{-1} + 1)(\rho - 1)^{-1/2}\mathcal{K}^{-1} \right\} \tag{2.4}$$

$$\mu = \mathcal{M} - sgn(\mathcal{S}) \left\{ 3\rho^{-1}(4\rho^{-1} + 1)(\mathcal{K}^{-1}\mathcal{V}) \right\}^{1/2} \tag{2.5}$$

$$\delta = \left\{ 3(4\rho^{-1} + 1)(1 - \rho^{-1})\mathcal{K}^{-1}\mathcal{V} \right\}^{1/2} \tag{2.6}$$

where $\rho = 3\mathcal{K}\mathcal{S}^{-2} - 4 > 1$ *and* $sgn(\cdot)$ *is the sign function.*

Theorem 2.2. *Given a NIG*$(\alpha, \beta, \mu, \delta)$ *distributed random variable. If its sample mean, sample variance, sample skewness and sample excess kurtosis are* $\hat{\mathcal{M}}$, $\hat{\mathcal{V}}$, $\hat{\mathcal{S}}$ *and* $\hat{\mathcal{K}}$ *respectively, and* $3\hat{\mathcal{K}} > 5\hat{\mathcal{S}}^2 > 0$, *then the method of moments estimators for the parameters are*

$$\hat{\alpha}_{MM} = 3\hat{\rho}^{1/2}(\hat{\rho} - 1)^{-1}\hat{\mathcal{V}}^{-1/2}|\hat{\mathcal{S}}|^{-1} \tag{2.7}$$

$$\hat{\beta}_{MM} = 3(\hat{\rho} - 1)^{-1}\hat{\mathcal{V}}^{-1/2}\hat{\mathcal{S}}^{-1} \tag{2.8}$$

$$\hat{\mu}_{MM} = \hat{\mathcal{M}} - 3\hat{\rho}^{-1}\hat{\mathcal{V}}^{1/2}\hat{\mathcal{S}}^{-1} \tag{2.9}$$

$$\hat{\delta}_{MM} = 3\hat{\rho}^{-1}(\hat{\rho} - 1)^{1/2}\hat{\mathcal{V}}^{1/2}|\hat{\mathcal{S}}|^{-1} \tag{2.10}$$

where $\hat{\rho} = 3\hat{\mathcal{K}}\hat{\mathcal{S}}^{-2} - 4 > 1$.

2.2.2 Moments of Risk Neutral Distribution

Bakshi et al. (2003) show that the risk neutral moments of $\tau-$period return $R_t(\tau) = \ln(S_{t+\tau}) - \ln(S_t)$ evaluated at time t can be written in terms of these payoffs. We use their methodology and a sample of out-of-the-money (OTM) calls and puts to estimate the higher moments of the risk neutral density function of log-price. Specifically, Bakshi et al. (2003) show that the price of contracts at time t on variance $VAR(t, \tau)$, skewness $SKEW(t, \tau)$ and kurtosis $KURT(t, \tau)$ of $R_t(\tau)$ can be calculated as

$$VAR(t,\tau) = e^{r\tau}V(t,\tau) - \mu^2(t,\tau) \tag{2.11}$$

$$SKEW(t,\tau) = \frac{e^{r\tau}W(t,\tau) - 3\mu(t,\tau)e^{r\tau}V(t,\tau) + 2\mu(t,\tau)^3}{[e^{r\tau}V(t,\tau) - \mu(t,\tau)^2]^{3/2}} \tag{2.12}$$

$$KURT(t,\tau) = \frac{e^{r\tau}X(t,\tau) - 4\mu(t,\tau)e^{r\tau}W(t,\tau) + 6e^{r\tau}\mu(t,\tau)^2V(t,\tau) - 3\mu(t,\tau)^4}{[e^{r\tau}V(t,\tau) - \mu(t,\tau)^2]^2} \tag{2.13}$$

where $V(t, \tau)$, $W(t, \tau)$, $X(t, \tau)$, and $\mu(t, \tau)$ are given by

$$V(t,\tau) = \int_{S_t}^{\infty} \frac{2(1 - \ln(K/S_t))}{K^2} C(t,\tau;K)dK + \int_0^{S_t} \frac{2(1 - \ln(K/S_t))}{K^2} P(t,\tau;K)dK \tag{2.14}$$

$$W(t,\tau) = \int_{S_t}^{\infty} \frac{6\ln(K/S_t) - 3(\ln(K/S_t))^2)}{K^2} C(t,\tau;K)dK + \int_{0}^{S_t} \frac{6\ln(K/S_t) - 3(\ln(K/S_t))^2}{K^2} P(t,\tau;K)dK \tag{2.15}$$

$$X(t,\tau) = \int_{S_t}^{\infty} \frac{12(\ln(K/S_t))^2 - 4(\ln(K/S_t))^3)}{K^2} C(t,\tau;K)dK + \int_{0}^{S_t} \frac{12(\ln(K/S_t))^2 - 4(\ln(K/S_t))^3}{K^2} P(t,\tau;K)dK \tag{2.16}$$

$$\mu(t,\tau) = e^{r\tau} - 1 - e^{r\tau}V(t,\tau)/2 - e^{r\tau}W(t,\tau)/6 - e^{r\tau}X(t,\tau)/24 \tag{2.17}$$

while $C(t,\tau;K)$ and $P(t,\tau;K)$ are the prices of European calls and puts written on the underlying stock with strike price K and expiration τ periods from time t. Bakshi et al. (2003) call $V(t,\tau)$ the price of the volatility contract on the underlying security, while $W(t,\tau)$ and $X(t,\tau)$ are the prices of the cubic contract and quartic contract respectively.

Thus, the risk neutral mean of $\ln(S_{t+\tau})$ conditional at time t is

$$M(t,\tau) = \mu(t,\tau) + \ln(S_t) \tag{2.18}$$

while the conditional variance, skewness and kurtosis of $\ln(S_{t+\tau})$ under risk neutral measure are exactly $VAR(t,\tau)$, $SKEW(t,\tau)$ and $KURT(t,\tau)$. As equations (3.28), (3.29) and (3.30) show, the procedure involves using a weighted sum of (out-of-the-money) options across varying strike prices to construct the prices of payoffs related to the second, third and fourth moments of returns. These prices are then used to construct estimates of the mean, variance, skewness and kurtosis of the risk neutral density function.

2.3 The NIG approximation and its relation to A-type Gram-Charlier expansions

In this section we discuss the NIG approximation and how well it approximates a function of random variables compared to the A-type Gram-Charlier and Edgeworth expansions. We do this by considering the region for which A-type Gram-Charlier expansion produces positive definite distribution and compare this region with the similar region produced by the normal inverse Gaussian distribution, since it has been shown in Barton and Dennis (1952) that the region, in terms of skewness/kurtosis combinations, covered by positive definite of A-type Gram-Charlier expansion is larger than the region of unimodality and also that of Edgeworth expansion. Hence, the A-type Gram-Charlier and Edgeworth expansions may easily lead to negative probabilities - a common problem encountered in practice.

Figure 1 has six panels, three pairs - with each pair representing data points superimposed on two regions. The data points are skewness and kurtosis daily estimates extracted from S & P 500 index options. These estimates are obtained by applying the formulas appearing in subsection 2.2.2, using the method of Bakshi et al. (2003). The details are discussed in Conrad et al. (2007a) who use data on out of the money (OTM) puts and calls, with at least two OTM puts and two OTM calls to calculate the moments on daily basis. The time to maturity is kept roughly constant at one month (see also later, section 2.5). The three pairs represent data for three different years: 1999, 2000 and 2003. The plots on the left in Figure 1 are the data and the NIG admissible region. All data points below the line are admissible, all those above are not. We note that the majority of data points yield proper NIG densities. The plots on the right in Figure 1 provide a close-up (note the scale ends at 8) in order to display the admissible region for Gram-Charlier expansion. The region is obtained via the dialytic method of Sylvester [see, for instance Wang (2001)] for finding the common zeros for A-type Gram-Charlier expansion.[3] Similar computations are reported in Shenton (1951), Barton and Dennis (1952) and Draper and Tierney (1972).

It is clear from Figure 1 that Gram-Charlier expansion almost never works. Only perhaps one data point - one day that is - falls below the curve. All other data points, as we can see from the left plots fall far beyond the region. As we will be discussed later this will seriously affect option pricing, favoring the use of NIG density approximation.

[3]The computations and plots were generated with Maple software.

2.4 Numerical Calibration

In this section we appraise via a calibration exercise the NIG density when used for pricing derivatives. The data generating process is the Heston model. We compare the risk neutral density estimation using the NIG distribution, Edgeworth and Gram-Charlier expansions with the true risk neutral density.[4] We find that the NIG approximation outperforms the Edgeworth and A-type Gram-Charlier expansions and achieves accuracy similar to the C-type Gram-Charlier expansion. Then we give a toy example which illustrates the pricing of European calls and a butterfly trading strategy. We find that the NIG approximation is closer to the pricing under true risk neutral measure than Edgeworth expansion and A-type Gram-Charlier expansion.

2.4.1 Density Approximations

We generate the stock price S_t and its volatility V_t using Heston model under risk-neutralized pricing probability (see Heston (1993)):

$$\begin{aligned} dS_t &= S_t(rdt + \sqrt{V_t}dW_t^1) \\ dV_t &= \kappa(\theta - V_t)dt + \sigma\sqrt{V_t}dW_t^2 \end{aligned} \tag{2.19}$$

where W_t^1 and W_t^2 are two correlated Brownian motions with correlation coefficient ρ. The conditional characteristic function of $X_{t+\tau} = \ln(S_{t+\tau})$ $(\tau > 0)$ under the risk neutral measure conditional at time t with $S_t = s$ and $V_t = v$ is given by Heston (1993) as

$$\Phi_t(u;\tau) = \exp(C(u;\tau) + D(u;\tau)v + iu\ln(s)) \tag{2.20}$$

with

$$C(u;\tau) = ru\tau i + \frac{\kappa\theta}{\sigma^2}[(\kappa - \rho\sigma ui + d)\tau - 2\ln(\frac{1 - ge^{d\tau}}{1 - g})] \tag{2.21}$$

$$D(u;\tau) = \frac{\kappa - \rho\sigma ui + d}{\sigma^2}\frac{1 - e^{d\tau}}{1 - ge^{d\tau}} \tag{2.22}$$

[4] Edgeworth expansions are a special case of Gram-Charlier and included here as well.

where $g = (\kappa - \rho\sigma ui + d)/(\kappa - \rho\sigma ui - d)$ and $d = \sqrt{(\rho\sigma ui - \kappa)^2 + \sigma^2(u^2 + ui)}$ if $\sigma > 0$. In particular, when $\sigma = 0$, $C(u;\tau)$ and $D(u;\tau)$ become

$$C(u;\tau) = ru\tau i + \frac{\theta(u^2 + ui)}{2}(\frac{1 - e^{-\kappa\tau}}{\kappa} - \tau) \tag{2.23}$$

$$D(u;\tau) = \frac{u^2 + ui}{2\kappa}(e^{-\kappa\tau} - 1) \tag{2.24}$$

The true conditional density function $f_t(x;\tau)$ can be derived from $\Phi_t(u;\tau)$ through the inverse Fourier transform, ie

$$f_t(x;\tau) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-iux}\Phi_t(u;\tau)du. \tag{2.25}$$

The European call prices are explicitly calculated using the formula from Heston (1993)

$$Call(t, s, v, K) = sP_1 - e^{-r\tau}KP_2 \tag{2.26}$$

where P_1, P_2 are respectively given in Appendix A.2.2.

To numerically evaluate the NIG performance , we use the same parameter settings as as Rompolis and Tzavalis (2007) who study Gram-Charlier approximations. Namely, we let $r = 0.05$, $\kappa = 1.62$, $\theta = 0.04$, $\sigma = 0.44$ and $\rho = -0.76$. Starting from an arbitrary time t_0, without loss of generality, we assume $t_0 = 0$ with $S_0 = 1080$ and $V_0 = 0.026$, we can generate a cross-section set of European call options with time to maturity $\tau = 0.21$ (in year) and strike prices spanning the interval $[820, 1260]$ at every 20 points. Further, the put prices are calculated using call-put parity.

Note that the estimation of mean, variance, skewness and kurtosis involves the evaluation of integrals (3.28), (3.29) and (3.30). To increase the accuracy of moments estimation, we interpolate the call/put prices within $[820, 1260]$ using a cubic spline and beyond this range, we perform the linear extrapolation (for reference, see Shimko (1993), Campa et al. (1998), Dennis and Mayhew (2002), Jiang and Tian (2005), Rompolis and Tzavalis (2007)). Note that this is not exactly what is done with real data, where typically only discrete sums are taken.[5] Once we have estimated the conditional moments of log-price, we are able to compute the

[5]See Conrad et al. (2007a) for further discussion and also Dennis and Mayhew (2002) who study the effect of discretization bias and show that its magnitude is typically not important.

RND approximation by the NIG distribution, Edgeworth expansion and A-type Gram-Charlier expansion. The true risk neutral density function $f_t(x;\tau)$ are evaluated by truncating the integral (2.25) at ± 100 and we also truncate the integration (A.3) at 100.

Figure 2 plots the true density curve and the curves generated by the NIG distribution, Edgeworth and A-type Gram-Charlier expansions, where one can see that the NIG density curve is very close to the true density curve especially at the two tails while Edgeworth and A-type Gram-Charlier expansions produce negative densities. We see that the NIG and true density are almost identical. In contrast, the other two approximations produce negative probabilities and feature humps that are not present in the true density. These phenomena are due to the fact that this realistic parameter setting for the Heston model is outside the feasible range for the Edgeworth and A-type Gram-Charlier expansions.

We also calibrate numerically the performance of various density estimations via the mean absolute error $L^1(f) = \int |f(x) - \hat{f}(x)|dx$ and the mean squared error $L^2(f) = \int (f(x) - \hat{f}(x))^2 dx$, where f is the true density and $\hat{f}$ is its estimate. Panel A of Table 1 reports the value of measurements L^1 and L^2 for NIG, Edgeworth, A-type Gram-Charlier and C-type Gram-Charlier expansions, respectively.[6] Again, we can see that the NIG approximation outperforms the Edgeworth and A-type Gram-Charlier expansions, while it achieves accuracy similar to the C-type Gram-Charlier expansion. The appeal of the NIG approximation, however, compared to C-type Gram-Charlier expansion is that it is easy to implement and it requires very little computational work.

2.4.2 Derivative pricing

Given that one has estimated the risk neutral density, we can proceed and price derivatives. We first consider the pricing of European call options through the true risk neutral density, NIG approximation, Edgeworth and A-type Gram-Charlier expansions. In Figure 3 and Figure 4, we plot the pricing of in-the-money (ITM), at-the-money (ATM) and out-of-the-money (OTM) call options with associated relative pricing errors which are defined as

$$(\hat{p} - p)/p$$

[6] The values with respect to C-type Gram-Charlier expansion are taken from Rompolis and Tzavalis (2007)

where p is the price under the true risk neutral measure and $\hat{p}$ is the price using the approximation. The ATM & ITM calls are not very much mispriced (see Panel c and d of Figure 3 and Figure 4) even though the risk neutral densities are severely misspecified (see Figure 2). However, for the deep OTM calls, Edgeworth and A-type Gram-Charlier expansions perform poorly. When we compare the scale of the various graphs we observe that relative pricing errors in the latter case exceed 10000. Obviously, these are relatively cheap options, with large pricing errors. What is important, however, is that the NIG approximation does not feature such pricing errors.

Next, we look at the butterfly trading strategy with payoff function

$$g(S_T; K, a) = (S_T - K + a)1_{(K-a \leq S_T \leq K)} + (K + a - S_T)1_{(K < S_T \leq K+a)}$$

In our numerical analysis, we take K such that $\log(K) = 6.60, 6.70, 6.80, 6.90$ $(< \log(S_0))$ and $a = 50$. Panel B of Table 1 reports the pricing using the true risk neutral density and the approximations where one could see that NIG pricing is closer to the true RND pricing than the other two approximations. The appeal of a butterfly strategy is that it singles out a particular area of the payoff space. Whenever the probability of that area is misspecified, we should expect serious mispricing. This is indeed what happens.

Figure 5 plots the pricing of butterfly strategy using various pricing tools and the relative pricing errors. Again one notes that the NIG approximation outperforms Edgeworth and A-type Gram-Charlier expansions. In comparison, the NIG density has mild pricing errors since its density approximation is more accurate.

2.5 Empirical illustration

The data used in this chapter is the same as in Conrad et al. (2007a) and similar to that used in Figlewski (2007). Our data on option prices is provided (through Wharton Research Data Services) from Optionmetrics. We use S&P 500 index option price data for all out-of-the-money calls and puts. In estimating the moments, we use equal numbers of out-of-the-money (OTM) calls and puts for each stock for each day. Thus, if there are n OTM puts with closing prices available on day t we require n OTM call prices. If there are N ($> n$) OTM call prices available on day t, we use the n OTM calls which have the most similar distance from stock to strike as the OTM puts for which we have data.

In the empirical illustration we average the daily moment estimates on a monthly basis, instead of plotting densities for specific days - since we do not really attempt to model the daily variation of the moments.[7] Figure 6 provides a time series plot of the S&P 500 as well as the risk neutral densities estimated with our NIG approximation for two dates, using three month contracts in March 2000 and August 2003. In both cases we superimpose the Gaussian distribution obtained with the same mean and variance. In both cases, the Normal is very different from the NIG, as expected. In the March 2000 case, we note a skewed distribution, while it is centered in the 2003 case.

2.6 Concluding remarks

In this chapter, we introduce the Normal Inverse Gaussian family to approximate the risk neutral distribution. Computational results indicate that NIG approximation is more efficient than Gram-Charlier series expansions by providing smaller approximation errors in comparison with the A-type Gram-Charlier expansion and by being less computationally burdensome than the C-type Gram-Charlier expansion. There are various expansions possible, notably the use of NIG approximations to compute stochastic discount factors, as further discussed in Conrad et al. (2007a). It should also be noted that the NIG density is not the only one with a four parameter characterization. In the future work we plan to explore alternative densities. In particular, the generalized hyperbolic distribution (GH) is a more general class of continuous probability distributions also defined as a normal variance-mean mixture where the mixing distribution is the generalized inverse Gaussian distribution.

[7] The next step is to model the dynamics of $\hat{\alpha}_{MM}$, $\hat{\beta}_{MM}$, $\hat{\mu}_{MM}$ and $\hat{\delta}_{MM}$ through time. We leave this as a topic for future research.

CHAPTER 3

Some Useful Densities for Risk Management and their Properties

3.1 Introduction

How to characterize the conditional distribution of asset returns is an important issue in risk management. This is because it is a key ingredient of option pricing, while in portfolio management the choice of equity distribution is more important than the quantification of risk level (see Embrechts et al. (2002) for more details). In the Black-Scholes option pricing model, the conditional distribution of asset returns is assumed to be normal, which is solely determined by the first two moments. Hence, it fails to account for the skewness and kurtosis of the financial time series which turn out to be very informative factors in modeling returns. The generalized hyperbolic (GH) distribution is derived from the normal distribution. It is a normal variance-mean mixture where the mixture is a generalized inverse Gaussian (GIG) distribution, so that it admits more degrees of freedom than the normal distribution and it could provide more realistic description of real data. It is therefore of interest to explore the GH family of distributions in the context of risk management.

The GH distribution was introduced by Barndorff-Nielsen (1977) for studying the aeolian sand deposits. It was first applied to a financial context by Eberlein and Keller (1995). This family of distributions are closed under linear transformations, which is a desirable property in portfolio management. Hence, they are infinitely divisible as well, which yield lévy processes by subordinating to Brownian Motions. They are not closed under convolution in general however, except for the normal inverse Gaussian (NIG) distribution which forms a subclass of the GH family of distributions. In addition, the GH distribution is skewed and leptokurtic, which meets the stylized feature of most financial returns.

The GH distribution considered in this chapter is characterized by five parameters. We will be talking in detail about their tail behavior and moments of higher order, which are the properties relevant to financial modeling. We further narrow down to a subclass of four-parameter distributions, so that there exists a one-to-one mapping from the parameter space to the space spanned by mean, variance, skewness and (excess) kurtosis. We only focus on the first four moments in that they contain most of the information we need and the analysis is less tedious than when considering even higher moments. The idea of keeping the number of moments small and building densities on this has been suggested in various papers, see Madan et al. (1998), Theodossiou (1998), Aas and Haff (2006), Eriksson et al. (2009), and among others.

The subclasses of four-parameter distributions we are interested in are the normal inverse Gaussian (NIG) distribution, the variance gamma (VG) distribution, and the generalized skewed t (GST) distribution. We consider these distributions because of their desirable tail behavior and analytical tractability in terms of moment estimation. The NIG distribution first appeared in the work of Barndorff-Nielsen (1997). Eriksson et al. (2009) recently studied this class of distributions in option pricing. They argue that, compared with the Edgeworth expansion and the Gram-Charlier expansion which are common approaches to approximate an unknown distribution, the NIG distribution is superior to aforementioned expansions when approximating the risk neutral distribution of asset returns, and provides less pricing error as well when it comes to derivative pricing. We revisit the class of NIG distributions in this chapter, and explore their properties as a subclass of the GH distribution. The VG distribution was introduced by Madan and Seneta (1990) for studying the dynamics of market returns. However, the distribution considered in the work of Madan and Seneta (1990) is a normal variance mixture with a three-parameter characterization, which is slightly different from our setting. This distribution was further studied in the context of option pricing by Madan et al. (1998) where it was extended to a three-parameter stochastic process by subordinating to a Brownian motion. We are among the few papers discussing the VG distribution as a limiting case of the GH distribution and formulating it with four parameters. There exists much literature on the GST distribution as well, see McDonald and Newey (1988), Theodossiou (1998), Prause (1999), Barndorff-Nielsen and Shephard (2001), Jones and Faddy (2003), Aas and Haff (2006), and among others. Nevertheless, the definitions are not consistent and slightly different from paper to paper. The GST distribution considered in this chapter is derived from the GH distribution, which is the same as the one studied in Aas and Haff (2006). The GST distribution is the only member in the GH family that possesses exponential/polynomial tails. Several other competing definitions of the (generalized) skewed t distribution could be found in McDonald and Newey (1988), Theodossiou (1998) but they fail to handle well the substantial skewness.

In this chapter, we systematically analyze the class of NIG distributions, the class of VG distributions, and the class of GST distributions. We focus on their tail behavior and the range of skewness and kurtosis for the purpose of risk management. We further use these distributions to model the risk neutral density of asset returns. The A-type Gram-Charlier expansion and the Edgeworth expansion[1] are considered as alternative approximating densities in this study as well. As in Eriksson et al. (2009), the parameters are estimated via the method of moments and the risk neutral moments of asset returns are evaluated based on option prices. To be specific, the risk neutral moments are formulated by a portfolio of the out-of-money(OTM) European Call/Put options indexed by their strikes (Bakshi et al. (2003)). The characterization of risk neutral distribution of asset returns is then directly linked to the options written upon it. However, we are unable to tell which approximating density is best in that the risk neutral distribution is unknown to us and it is then infeasible to calibrate how close the approximating density is to the true risk neutral distribution. In order to judge the performance of various approximating densities, we consider a heuristic option pricing model, affine jump-diffusion model, which could yield a closed form expression for density function. We further look into the issue of option pricing under the true density and approximating ones as well so as to compare the pricing errors numerically.

The rest of this chapter is outlined as follows: we start with a review on the GH family of distributions in section 3.2, and then introduce various subclasses of the GH family and talk about their properties. In section 3.3, parameter estimation via the method of moments is presented and the ranges of skewness and kurtosis for various distributions are derived as

[1]See Appendix A.2 for more details on the Gram-Charlier expansion and the Edgeworth expansion.

well. We illustrate the option-based moment estimation in section 3.4. Section 3.5 focuses on evaluating the performance of the GH family of distributions in option pricing. And we give the concluding remarks in section 3.6.

3.2 The Generalized Hyperbolic Distribution

The generalized hyperbolic (GH) distribution can be considered as a normal variance-mean mixture where the mixture is a generalized inverse Gaussian (GIG) distribution.

Suppose that Y is a GIG-distributed random variable, or $Y \sim GIG(a, b, p)$. Namely, its density function is

$$f(y; a, b, p) = \frac{(a/b)^{p/2}}{2K_p(\sqrt{ab})} y^{p-1} \exp[-\frac{1}{2}(ay + b/y)], \quad y > 0$$

where

$$K_p(z) = \frac{1}{2}\int_0^\infty y^{p-1} \exp[-\frac{1}{2}z(y + 1/y)]dy, \quad z > 0 \tag{3.1}$$

is a modified Bessel function of the third kind with index p. The parameter space of the GIG distribution is

$$\{a > 0, b > 0, p = 0\} \bigcup \{a > 0, b \geq 0, p > 0\} \bigcup \{a \geq 0, b > 0, p < 0\}$$

The GIG distribution can be reduced to a gamma distribution if $a > 0, b = 0, p > 0$, while it becomes an inverse gamma distribution if $a = 0, b > 0, p < 0$.

A GH random variable is constructed by allowing for the mean and the variance of a normal random variable GIG distributed. Namely,

Definition 3.1. *A random variable X is generalized hyperbolic (GH) distributed, or $X \sim GH(\alpha, \beta, \mu, b, p)$ if it has the same law as*

$$X \overset{\mathcal{L}}{=} \mu + \beta Y + \sqrt{Y} Z \tag{3.2}$$

where $Y \sim GIG(\alpha^2 - \beta^2, b^2, p)$, $Z \sim N(0,1)$, *and* Y *is independent of* Z. *The parameter space for a GH distribution is*

$$\{(\alpha, \beta, b, p, \mu) : \alpha > |\beta|, b > 0, p \in \mathbb{R}, \mu \in \mathbb{R}\}.$$

In particular, its density function is

$$f_{GH}(x; \alpha, \beta, \mu, b, p) = \frac{\alpha^{1/2-p}(\alpha^2 - \beta^2)^{p/2} e^{(x-\mu)\beta}}{\sqrt{2\pi} b K_p(b\sqrt{\alpha^2 - \beta^2})} K_{p-1/2}\left(\alpha b \sqrt{1 + \frac{(x-\mu)^2}{b^2}}\right)\left(1 + \frac{(x-\mu)^2}{b^2}\right)^{p/2-1/4} \tag{3.3}$$

Or,

$$f_{GH}(x; \alpha, \beta, \mu, b, p) = \frac{\bar{\gamma}^p \bar{\alpha}^{1/2-p} e^{(x-\mu)\beta}}{\sqrt{2\pi} b K_p(\bar{\gamma})} K_{p-1/2}\left(\bar{\alpha} \sqrt{1 + \frac{(x-\mu)^2}{b^2}}\right)\left(1 + \frac{(x-\mu)^2}{b^2}\right)^{p/2-1/4}$$

where $\bar{\alpha} = b\alpha, \bar{\beta} = b\beta, \gamma = \sqrt{\alpha^2 - \beta^2}$ *and* $\bar{\gamma} = b\gamma = \sqrt{\bar{\alpha}^2 - \bar{\beta}^2}$.

The GH distribution is characterized through five parameters. Among them, β controls skewness. We will say the GH distribution is non-skewed if $\beta = 0$. μ is a location parameter, and p pertains to how fat the tails are.

Due to the scaling property of the GIG distribution (ie., if $Y \sim GIG(a, b, p)$, then $tY \sim GIG(a/|t|, |t|b, p)$ for $t \neq 0$), the GH distribution is closed under linear transformations. Suppose that X is $GH(\alpha, \beta, \mu, b, p)$ distributed. $tX + l$ belongs to the GH family as well with parameters $(\alpha/|t|, \beta/|t|, |t|\mu + l, tb, p)$ for any $t \neq 0$. Therefore, the set

$$\{\mu + \beta Y + \sigma\sqrt{Y}Z : Y \sim GIG(\alpha^2 - \beta^2, b^2, p), Z \sim N(0,1), Y \perp Z, \alpha > |\beta|, b > 0, p \in \mathbb{R}, \mu \in \mathbb{R}, \sigma > 0\}$$

and the set

$$\{\mu + \beta Y + \sqrt{Y}Z : Y \sim GIG(\alpha^2 - \beta^2, b^2, p), Z \sim N(0,1), Y \perp Z, \alpha > |\beta|, b > 0, p \in \mathbb{R}, \mu \in \mathbb{R}\}$$

are equivalent under the linear transform, which implies that it is sufficient to model the GH distribution with five parameters. It also follows that the GH distribution is infinitely divisible, a property that results in the existence of GH Lévy processes by subordinating to Brownian motions. However, the GH family is not closed under convolution in general except when

$p = -1/2$, which is what is called the normal inverse Gaussian (NIG) distribution.

Various subclasses of the GH family could be derived by allowing the parameter(s) to assume special values within the parameter space. The most commonly used distributions which form subclasses of the GH distribution are: (1) the symmetric GH distribution, which is related to the GH distribution by $SGH(\alpha, \mu, b, p) \doteq GH(\alpha, 0, \mu, b, p)$; (2) the hyperbolic distribution, $H(\alpha, \beta, \mu, b) \doteq GH(\alpha, \beta, \mu, b, 1)$; (3) the normal inverse Gaussian distribution, $NIG(\alpha, \beta, \mu, b) \doteq GH(\alpha, \beta, \mu, b, -1/2)$.

Notice that the parameter space of the GH distribution excludes the sets $\{\alpha > |\beta|, b = 0, p > 0\}$ and $\{\alpha = |\beta|, b > 0, p < 0\}$ which, however, are permitted to the GIG distribution. If we take values on the boundary of parameter space, we would be deriving various limiting cases of the GH distribution. To be specific, we would arrive at the variance gamma (VG) distribution by taking b to 0 in equation (3.3). That α approaches to β produces the generalized skewed t (GST) distribution. In particular, by allowing $p = -b^2/2$ in the GST distribution, we have the skewed t distribution. If further assume $\beta = 0$ it is reduced to the noncentral student t distribution with b^2 degrees of freedom. It becomes the central student t distribution if both $\beta = 0$ and $\mu = 0$. The Cauchy distribution could also be regarded as a limiting case of the GH distribution in that it could be derived from the GH distribution by $\alpha \to |\beta| = 0$ and $p = -1/2$. In addition, if we send some of parameters in the GH distribution to infinity, we will end up with the normal distribution, the gamma distribution, the inverse gamma distribution, etc. (see Eberlein and von Hammerstein (2002) for more details).

3.2.1 Tail Behavior

It is well known that the returns of most financial assets exhibit semi-heavy tails. We will look into the tail properties of the GH family of distributions and argue that the NIG distribution, the VG distribution and the GST distribution provide a good fit to the financial time series in terms of the tail behavior.

Based on the fact that $K_p(z) \sim \sqrt{\frac{\pi}{2}} z^{-1/2} e^{-z}$ as $z \to \infty$ (Jorgensen (1982)), we have the following statement regarding the tails of the GH distribution. A similar result could be found in Aas and Haff (2006).

Theorem 3.1 (Tails of the GH distribution)**.** *Given a GH*$(\alpha, \beta, \mu, b, p)$ *distribution with* $\alpha > |\beta|, b > 0, p \in \mathbb{R}, \mu \in \mathbb{R}$, *when* $|x|$ *gets larger, its density function is of the form*

$$f_{GH}(x; \alpha, \beta, \mu, b, p) \sim c|x|^{p-1} \exp(-\alpha|x| + \beta x)$$

where c *is a constant.*

The two tails of the GH distribution are asymmetric in that the right tail behaves like $cx^{p-1} \exp(-\alpha x + \beta x)$ while the left tail is $c|x|^{p-1} \exp(\alpha x + \beta x)$. Right tail is heavier if $\beta > 0$ but left tail is heavier if $\beta < 0$. Hence, the GH distribution is semi-heavy tailed and it has moments of any order.

Corollary 3.1. *As special cases of the GH distribution, the symmetric GH distribution, the hyperbolic distribution and the NIG distribution have the following tail behavior: for sufficiently large* x*,*

(i) $f_{SGH}(x; \alpha, \mu, b, p) \doteq f_{GH}(x; \alpha, 0, \mu, b, p) \sim c|x|^{p-1} \exp(-\alpha|x|)$

(ii) $f_H(x; \alpha, \beta, \mu, b) \doteq f_{GH}(x; \alpha, \beta, \mu, b, 1) \sim c \exp(-\alpha|x| + \beta x)$

(iii) $f_{NIG}(x; \alpha, \beta, \mu, b) \doteq f_{GH}(x; \alpha, \beta, \mu, b, -1/2) \sim c|x|^{-3/2} \exp(-\alpha|x| + \beta x)$

The tails of the GH limiting distributions are derived by taking limits in the density function (3.3).

Theorem 3.2 (Tails of the VG distribution)**.** *Suppose that* $\alpha > |\beta|, p > 0, \mu \in \mathbb{R}$. *The GH distribution is reduced to the VG distribution if* $b \to 0$, *and the tails become*

$$f_{VG}(x; \alpha, \beta, \mu, p) \doteq f_{GH}(x; \alpha, \beta, \mu, 0, p) \sim c|x|^{p-1} \exp(-\alpha|x| + \beta x).$$

In particular, the VG distribution possesses moments of arbitrary order.

Fix $\beta \neq 0, \mu \in \mathbb{R}, b > 0, p < 0$ and take α to $|\beta|$ in (3.3), we have

Theorem 3.3 (Tails of the GST distribution)**.** *Suppose that* $|\beta| > 0, \mu \in \mathbb{R}, b > 0, p < 0$. *The GST distribution is derived from the GH distribution by* $\alpha \to |\beta|$. *It has tails*

$$f_{GST}(x; \beta, \mu, b, p) \doteq f_{GH}(x; |\beta|, \beta, \mu, b, p) \sim c|x|^{p-1} \exp(-|\beta x| + \beta x).$$

The GST law is not semi-heavy tailed, because one tail is polynomial while the other is exponential. Namely,

1. when $\beta > 0$, right tail $\sim cx^{p-1}$, left tail $\sim c|x|^{p-1}e^{2\beta x}$.

2. when $\beta < 0$, right tail $\sim cx^{p-1}e^{2\beta x}$, left tail $\sim c|x|^{p-1}$.

Therefore it could not have moments of arbitrary order. The r^{th} moment exists if and only if $r < -p$. Particularly, when $p = -b^2/2$, it is the skewed t distribution with tails

$$f_{GST}(x; |\beta|, \beta, \mu, b, -b^2/2) \sim c|x|^{-b^2/2-1} \exp(-|\beta x| + \beta x).$$

Let both α and β go to 0 while freeze the other parameters (i.e., keep $\mu \in \mathbb{R}, b > 0, p < 0$ fixed), we have

Theorem 3.4 (Tails of the noncentral student t distribution)**.** *When $\alpha = \beta = 0, \mu \in \mathbb{R}, b > 0, p < 0$, we have a noncentral student t distribution with $-2p$ degrees of freedom. Its tails behave like*

$$f_{GH}(x; 0, 0, \mu, b, p) \sim c|x|^{2p-1}.$$

The r^{th} moment exists if and only if $r < -2p$.

Therefore, the Cauchy distribution, as a special case of the noncentral student t distribution (i.e. when $p = -\frac{1}{2}$), has tails

$$f_{GST}(x; 0, 0, \mu, b, -\frac{1}{2}) \sim c|x|^{-2}.$$

The normal distribution can be regarded as a limiting case of the GH law as well if we assume $b > 0$, $\alpha \to \infty$ and $\lim_{\alpha\to\infty} \frac{b}{\alpha} = \sigma^2$ for some σ, under which circumstances, the GH distribution is reduced to the normal distribution with mean $\mu + \beta\sigma^2$ and variance σ^2. The tails of the normal law are proportional to $e^{-x^2/2}$.

Remark 3.1. *Among all the limiting distributions, only the VG distribution is semi-heavy tailed and possesses moments of arbitrary order, which are desirable properties in modeling financial returns. Although the GST distribution does not have moments of any order, its tails are a* *mixture of polynomial and exponential, which distinguishes the GST law from the others and makes it special.*

Since the main task of this chapter is to model financial returns and build density upon skewness and (excess) kurtosis, we focus on a subclass of the GH family which have a four-parameter characterization, so that there exists a bijection from the parameter space to the space spanned by the first four moments. Besides, we require the distributions to be skewed (i.e., $\beta \neq 0$) and semi-heavy tailed. The subclasses of the GH family which could serve this purpose are the NIG distribution, the VG distribution, and the GST distribution.

3.3 Parameter Estimation via the Method of Moments

This section is primarily concerned with parameter estimation of the NIG distribution, the VG distribution, and the GST distribution via the method of moments. First, we will present some general results regarding the GH family of distributions.

3.3.1 A General Case

For a centered GH distribution (i.e., $\mu = 0$), its moments of arbitrary order could be expanded as an infinitely series of Bessel functions of third kind with gamma weights (see Barndorff-Nielsen and Stelzer (2005) for more details).

Lemma 3.1 (Moments of GH law)**.** *For a $GH(\alpha, \beta, 0, b, p)$ distributed random variable X, its n^{th} moment, denoted by m_n, equals*

$$m_n \doteq EX^n = \frac{2^{\lceil \frac{n}{2} \rceil} \bar{\gamma}^p b^{2\lceil \frac{n}{2} \rceil} \beta^m}{\sqrt{\pi} K_p(\bar{\gamma}) \bar{\alpha}^{p+\lceil \frac{n}{2} \rceil}} \sum_{k=0}^{\infty} \frac{2^k \bar{\beta}^{2k} \Gamma(k + \lceil \frac{n}{2} \rceil + \frac{1}{2})}{\bar{\alpha}^k (2k+m)!} K_{p+k+\lceil \frac{n}{2} \rceil}(\bar{\alpha}) \tag{3.4}$$

where $m \equiv n(mod\ 2)$.

Equation (3.4) could be further simplified for the first four moments.

Theorem 3.5. *The first four moments of a $GH(\alpha, \beta, 0, b, p)$ random variable can be explicitly expressed as*

$$m_1 = \frac{b\beta K_{p+1}(\bar{\gamma})}{\gamma K_p(\bar{\gamma})} \tag{3.5}$$

$$m_2 = \frac{bK_{p+1}(\bar{\gamma})}{\gamma K_p(\bar{\gamma})} + \frac{\beta^2 b^2 K_{p+2}(\bar{\gamma})}{\gamma^2 K_p(\bar{\gamma})} \tag{3.6}$$

$$m_3 = \frac{3\beta b^2 K_{p+2}(\bar{\gamma})}{\gamma^2 K_p(\bar{\gamma})} + \frac{\beta^3 b^3 K_{p+3}(\bar{\gamma})}{\gamma^3 K_p(\bar{\gamma})} \tag{3.7}$$

$$m_4 = \frac{\beta^4 b^4 K_{p+4}(\bar{\gamma})}{\gamma^4 K_p(\bar{\gamma})} + \frac{6\beta^2 b^3 K_{p+3}(\bar{\gamma})}{\gamma^3 K_p(\bar{\gamma})} + \frac{3b^2 K_{p+2}(\bar{\gamma})}{\gamma^2 K_p(\bar{\gamma})} \tag{3.8}$$

Proof: *It follows from the fact that*

$$K_v(z) = \frac{z^v}{x^v} \sum_{k=0}^{\infty} \frac{1}{2^k k!} \frac{y^{2k}}{x^k} K_{v+k}(x)$$

Note that given a $GH(\alpha, \beta, \mu, b, p)$ random variable X, $X - \mu$ is $GH(\alpha, \beta, 0, b, p)$ distributed. Therefore, the mean, variance, skewness and excess kurtosis of X could be derived from Theorem 3.5 directly.

Theorem 3.6 (Mean, Variance, Skewness and Excess Kurtosis of the GH law)**.**

Suppose that X is $GH(\alpha, \beta, \mu, b, p)$ distributed. Then its mean M, variance V, skewness S and excess kurtosis K can be put as

$$M = \mu + \frac{\beta b K_{p+1}(\bar{\gamma})}{\gamma K_p(\bar{\gamma})} \tag{3.9}$$

$$V = \frac{bK_{p+1}(\bar{\gamma})}{\gamma K_p(\bar{\gamma})} + \frac{b^2\beta^2 K_{p+2}(\bar{\gamma})}{\gamma^2 K_p(\bar{\gamma})} - \frac{b^2\beta^2 K_{p+1}^2(\bar{\gamma})}{\gamma^2 K_p^2(\bar{\gamma})} \tag{3.10}$$

$$S = \frac{m_3 - 3m_1m_2 + 2m_1^3}{Var(X)^{3/2}} \tag{3.11}$$

$$K = \frac{m_4 - 4m_1m_3 + 12m_1^2m_2 - 6m_1^4 - 3m_2^2}{Var(X)^2} \tag{3.12}$$

with $m_i, i = 1, 2, 3, 4$ defined in Theorem 3.5.

For a general GH distribution, it is hard to extract the parameters from equations (3.9), (3.10), (3.11) and (3.12) in terms of the moments. However, it is easy for the NIG distribution, the VG distribution and the GST distribution.

3.3.2 The Normal Inverse Gaussian Distribution

The GH distribution is reduced to a NIG distribution when $p = -1/2$. The mixture of the NIG distribution is inverse Gaussian or $IG(b, \sqrt{\alpha^2 - \beta^2})$[2]. Replacing p by $-1/2$, and considering the facts that $K_{1/2}(z) = \sqrt{\frac{\pi}{2}} z^{-1/2} e^{-z}$ and $K_p(z) = K_{-p}(z)$, we could derive the density function for the NIG law, which is

$$f_{NIG}(x; \alpha, \beta, \mu, b) = \frac{\alpha}{\pi} \frac{e^{\gamma b + \beta(x-\mu)}}{\sqrt{1 + (\frac{x-\mu}{b})^2}} K_1 \left(\alpha b \sqrt{1 + (\frac{x-\mu}{b})^2} \right)$$

Next, we restate Theorem 3.6 for the NIG distribution.

Corollary 3.2 (Mean, Variance, Skewness and Excess Kurtosis of the NIG law)**.**

Given a NIG(α, β, μ, b) distributed random variable, its mean M, variance V, skewness S and excess kurtosis K can be related to the parameters in the following way:

$$\begin{aligned} M =& \mu + \frac{\beta b}{\gamma} \\ V =& \frac{b\alpha^2}{\gamma^3} \\ S =& \frac{3\beta}{\alpha\sqrt{b\gamma}} \\ K =& \frac{3(4\beta^2 + \alpha^2)}{b\alpha^2\gamma} \end{aligned}$$

Particularly, we have $3K > 5S^2$.

Proof: *Consider*

$$K_{n+1/2}(z) = K_{1/2}(z) \left(1 + \sum_{i=1}^{n} \frac{(n+i)!}{i!(n-i)!} 2^{-i} z^{-i} \right)$$

From Corollary 3.2, one could know that the range of skewness and excess kurtosis implied by the NIG distribution is $\{(K, S^2) : 3K > 5S^2\}$. We call it feasible domain of the NIG distribution. Next, we will give the method of moments estimation for the NIG distribution analytically.

[2] $Y \sim IG(\delta, \gamma)$ if its density function is $f(y; \delta, \gamma) = (\frac{\delta^2}{2\pi y^3})^{1/2} \exp(\delta\gamma) \exp[-\frac{1}{2}(\gamma^2 y + \delta^2/y)]$

Theorem 3.7. *Suppose that for a given NIG*(α, β, μ, b) *random variable, its sample mean, sample variance, sample skewness and sample excess kurtosis are* $\hat{M}$, $\hat{V}$, $\hat{S}$, $\hat{K}$ *and* $3\hat{K} > 5\hat{S}^2$. *The method of moments estimation for the parameters are*

$$\hat{\alpha} = 3\frac{\sqrt{D + \hat{S}^2}}{D}\hat{V}^{-1/2} \tag{3.13}$$

$$\hat{\beta} = \frac{3\hat{S}}{D}\hat{V}^{-1/2} \tag{3.14}$$

$$\hat{\mu} = \hat{M} - \frac{3\hat{S}}{D + \hat{S}^2}\hat{V}^{1/2} \tag{3.15}$$

$$\hat{b} = \frac{3\sqrt{D}}{D + \hat{S}^2}\hat{V}^{1/2} \tag{3.16}$$

where $D = 3\hat{K} - 5\hat{S}^2 > 0$.

Proof: *See Eriksson et al. (2004).*

3.3.3 The Variance Gamma Distribution

Keep $\alpha > 0, \mu \in \mathbb{R}, p > 0$ fixed and take b to 0. We will arrive at the VG distribution with the mixture $Gamma(p, \frac{2}{\alpha^2 - \beta^2})$[3]. The density function of $VG(\alpha, \beta, \mu, p)$ is

$$f_{VG}(x; \alpha, \beta, \mu, p) = \frac{(\alpha^2 - \beta^2)^p}{\sqrt{\pi}\Gamma(p)(2\alpha)^{p-1/2}} e^{(x-\mu)\beta} |x - \mu|^{p-1/2} K_{p-1/2}(|x - \mu|\alpha)$$

which is obtained by using the fact that $K_p(z) \sim 2^{p-1}\Gamma(p)z^{-p}$ as $z \to 0$ if $p > 0$ (Jorgensen (1982)).

Next lemma presents some useful properties regarding the Bessel function of third kind, which could be helpful in deriving the method of moments estimation for the VG distribution and the GST distribution.

Lemma 3.2. *The Bessel function of third kind defined in (3.1) has the following properties:*

[3] $X \sim Gamma(k, \theta)$ if the density function is $f(x; k, \theta) = x^{k-1} \frac{e^{-x/\theta}}{\theta^k \Gamma(k)}$ where $x > 0, k > 0, \theta > 0$

for any $k \in \mathbb{Z}^+$,

$$\lim_{b \to 0} \frac{b^k K_{p+k}(\bar{\gamma})}{K_p(\bar{\gamma})} = \frac{2^k \Gamma(p+k)}{\gamma^k \Gamma(p)} \quad \text{if} \quad p > 0 \tag{3.17}$$

$$\lim_{\gamma \to 0} \frac{K_{p+k}(\bar{\gamma})}{\gamma^k K_p(\bar{\gamma})} = \frac{b^k \Gamma(-p-k)}{2^k \Gamma(-p)} \quad \text{if} \quad p < 0 \tag{3.18}$$

where $\bar{\gamma} = b\gamma$.

Let b approach 0 in Theorem 3.6. We have the following statement:

Corollary 3.3 (Mean, Variance, Skewness and Excess Kurtosis of the VG law)**.** *The mean* M*, variance* V*, skewness* S *and excess kurtosis* K *of a* $VG(\alpha, \beta, \mu, p)$ *random variable are*

$$\begin{aligned} M &= \mu + \frac{\beta p}{\eta} \\ V &= \frac{p}{\eta^2}(\eta + \beta^2) \\ S &= \frac{\beta(3\eta + 2\beta^2)}{(\eta + \beta^2)^{3/2} p^{1/2}} \\ K &= \frac{3(\eta^2 + 4\eta\beta^2 + 2\beta^4)}{p(\eta + \beta^2)^2} \end{aligned}$$

where $\eta = \frac{\alpha^2 - \beta^2}{2} > 0$, $\alpha > 0, p > 0, \mu \in \mathbb{R}$ *and* $2K > 3S^2$.

Proof: *It follows from equation (3.17) of Lemma 3.2 and the dominant convergence theorem.*

Therefore, the feasible domain of the VG distribution is $\{(K, S^2) : 2K > 3S^2\}$. Next theorem presents the method of moments estimation for the VG distribution.

Theorem 3.8 (Method of moments estimation for the VG law)**.** *Suppose that, for a given* $VG(\alpha, \beta, \mu, p)$ *distribution, the sample mean, sample variance, sample skewness and sample excess kurtosis are* $\hat{M}$, $\hat{V}$, $\hat{S}$ *and* $\hat{K}$ *such that* $2\hat{K} > 3\hat{S}^2$*. If further we assume* $\hat{S} \neq 0$*, the method of moments estimation for the VG parameters are*

$$\hat{\alpha} = \frac{2\sqrt{R}(3+R)}{\sqrt{\hat{V}}|\hat{S}|(1-R^2)} \tag{3.19}$$

$$\hat{\beta} = \frac{2R(3+R)}{\sqrt{\hat{V}}\hat{S}(1-R^2)} \tag{3.20}$$

$$\hat{p} = \frac{2R(3+R)^2}{\hat{S}^2(1+R)^3} \tag{3.21}$$

$$\hat{\mu} = \hat{M} - \frac{2\sqrt{\hat{V}}R(3+R)}{\hat{S}(1+R)^2} \tag{3.22}$$

where, letting $C = 3\hat{S}^2/2\hat{K}$, R is the unique solution to the following equation:

$$(C-1)R^3 + (7C-6)R^2 + (7C-9)R + C = 0$$

and $0 < R < 1$. If $\hat{S} = 0$, we have

$$\alpha = 6/(\hat{V}\hat{K}), \quad \beta = 0, \quad p = 3/\hat{K}, \quad \mu = \hat{M}$$

Proof: *See Appendix A.3.1.*

3.3.4 The Generalized Skewed T Distribution

The GST distribution is derived by allowing $\alpha \to |\beta|$. Therefore, it has density function as

$$f_{GST}(x;\beta,\mu,b,p) = \frac{2^{p+1/2}b^{-2p}}{\sqrt{\pi}\Gamma(-p)} e^{(x-\mu)\beta} \left(\frac{|\beta|}{\sqrt{(x-\mu)^2+b^2}}\right)^{-p+1/2} K_{-p+1/2}(|\beta|\sqrt{(x-\mu)^2+b^2})$$

where $\beta \in \mathbb{R}, \mu \in \mathbb{R}, b > 0$, and $p < -4$ so that the 4^{th} moment exists. The mixing distribution of the GST law is inverse gamma or $InvGamma(-p, b^2/2)$[4]. Its mean, variance, skewness and excess kurtosis are derived by sending γ to 0 in Theorem 3.6 and using (3.18) of Lemma 3.2.

Corollary 3.4 (Mean, Variance, Skewness and Excess Kurtosis of the GST law)**.** *The mean M, variance V, skewness S and excess kurtosis K of a GST distribution can be explicitly expressed as follows*

$$M = \mu + \frac{b^2\beta}{v-2}$$

$$V = \frac{b^2}{v-2} + \frac{2b^4\beta^2}{(v-2)^2(v-4)}$$

$$S = \left[6(v-2) + \frac{16b^2\beta^2}{v-6}\right] \frac{b\beta(v-4)^{1/2}}{[(v-2)(v-4)+2b^2\beta^2]^{3/2}}$$

[4] $Y \sim InvGamma(k,\theta)$ $(k > 0, \theta > 0)$ if the density function is $f(y;k,\theta) = \theta^k(1/y)^{k+1}\frac{e^{-\theta/y}}{\Gamma(k)}$

$$K = \left[\frac{8b^4\beta^4(5v-22)}{(v-6)(v-8)} + \frac{16b^2\beta^2(v-2)(v-4)}{v-6} + (v-2)^2(v-4) \right] \frac{6}{[(v-2)(v-4)+2b^2\beta^2]^2}$$

where $v = -2p > 0$.

Theorem 3.9 (Method of moments estimation for the GST law)**.** *Given the sample mean* $\hat{M}$*, sample variance* $\hat{V}$*, sample skewness* $\hat{S}$ *and sample excess kurtosis* $\hat{K}$*, the method of moments estimation for a* $GST(\beta, \mu, b, p)$ *distribution is*

$$\hat{\beta} = sig(\hat{S})\sqrt{\frac{\rho(1+\rho)(v-4)}{2\hat{V}}} \tag{3.23}$$

$$\hat{\mu} = \hat{M} - sig(\hat{S})\sqrt{\frac{\rho(v-4)\hat{V}}{2(1+\rho)}} \tag{3.24}$$

$$\hat{b} = \sqrt{\frac{\hat{V}(v-2)}{1+\rho}} \tag{3.25}$$

$$\hat{p} = -v/2 \tag{3.26}$$

where $\rho(\geq 0)$ *and* $v(>8)$ *are solutions to the following system of equations:*

$$\begin{aligned} 2\rho[3(v-6)+4(v-4)\rho]^2 - \hat{S}^2(v-4)(v-6)^2(1+\rho)^3 = 0 \\ 12(v-4)(5v-22)\rho^2 + 48(v-4)(v-8)\rho + 6(v-6)(v-8) - \hat{K}(v-4)(v-6)(v-8)(1+\rho)^2 = 0 \end{aligned} \tag{3.27}$$

A sufficient condition under which (3.27) could have solutions is $\hat{S}^2 < \min(3/2, \hat{K}/3)$*, while the necessary condition is* $\hat{S}^2 < \min(3/2, \hat{K})$.

Particularly, when $\hat{S} = 0$*, the estimates are reduced to*

$$\beta = 0, \quad \mu = \hat{M}, \quad b = \sqrt{\hat{V}(2+6/\hat{K})}, \quad p = -(2+3/\hat{K}).$$

Proof: *See Appendix A.3.2.*

Although we do not give an explicit expression for the feasible domain of the GST distribution, it is clear, from Theorem 3.9, that this region is enclosed in the set $\{(K, S^2) : S^2 < \min(3/2, K)\}$ and it contains $\{(K, S^2) : S^2 < \min(3/2, K/3)\}$. Moreover, from Figure 7 and Figure 8, one could see that these two regions are sufficient to characterize the feasible domain of the GST distribution. We will talk about Figure 7 and Figure 8 in more details in the next section.

Remark 3.2. *From the discussions above, one could see that among the three distributions, the VG distribution admits the largest possible combinations of skewness and (excess) kurtosis. The NIG law comes next but it is still far bigger than that of the GST distribution (see Figure 7 panel (a) and Figure 8).*

3.4 Moments of risk neutral distribution

Since we are interested in modeling the asset returns with the NIG distribution, the VG distribution and the GST distribution, it is natural to think about the risk neutral distribution of asset returns which plays an important role in derivative pricing.

Suppose that $\{S_t\}$ is a given price process. Bakshi et al. (2003) point out that the risk neutral moments of $\tau-$period return $R_t(\tau) = \ln(S_{t+\tau}) - \ln(S_t)$ evaluated at time t can be written in terms of a sample of out-of-the-money (OTM) call and put options. Specifically, Bakshi et al. (2003) show that the arbitrage-free prices of volatility contract, cubic contract and quartic contract at time t can be formulated as

$$\begin{aligned} V(t,\tau) = E_t^Q(e^{-r\tau}R_t(\tau)^2) = &\int_{S_t}^{\infty} \frac{2(1-\ln(K/S_t))}{K^2} C(t,\tau;K)dK \\ &+ \int_0^{S_t} \frac{2(1-\ln(K/S_t))}{K^2} P(t,\tau;K)dK \end{aligned} \tag{3.28}$$

$$\begin{aligned} W(t,\tau) = E_t^Q(e^{-r\tau}R_t(\tau)^3) = &\int_{S_t}^{\infty} \frac{6\ln(K/S_t) - 3(\ln(K/S_t))^2)}{K^2} C(t,\tau;K)dK \\ &+ \int_0^{S_t} \frac{6\ln(K/S_t) - 3(\ln(K/S_t))^2}{K^2} P(t,\tau;K)dK \end{aligned} \tag{3.29}$$

$$\begin{aligned} X(t,\tau) = E_t^Q(e^{-r\tau}R_t(\tau)^4) = &\int_{S_t}^{\infty} \frac{12(\ln(K/S_t))^2 - 4(\ln(K/S_t))^3)}{K^2} C(t,\tau;K)dK \\ &+ \int_0^{S_t} \frac{12(\ln(K/S_t))^2 - 4(\ln(K/S_t))^3}{K^2} P(t,\tau;K)dK \end{aligned} \tag{3.30}$$

where Q denotes risk neutral measure, r is risk-free rate, while $C(t,\tau;K)$ and $P(t,\tau;K)$ are the prices of European calls and puts written on the underlying asset with strike price K

and expiration τ periods from time t. Therefore, the risk neutral moments (mean, variance, skewness, and excess kurtosis) of $\ln(S_{t+\tau})$ conditional on the information up to time t are constructed as

$$Mean(t,\tau) = \mu(t,\tau) + \ln(S_t) \tag{3.31}$$

$$Var(t,\tau) = e^{r\tau}V(t,\tau) - \mu^2(t,\tau) \tag{3.32}$$

$$Skew(t,\tau) = \frac{e^{r\tau}W(t,\tau) - 3\mu(t,\tau)e^{r\tau}V(t,\tau) + 2\mu(t,\tau)^3}{[e^{r\tau}V(t,\tau) - \mu(t,\tau)^2]^{3/2}} \tag{3.33}$$

$$EKurt(t,\tau) = \frac{e^{r\tau}X(t,\tau) - 4\mu(t,\tau)e^{r\tau}W(t,\tau) + 6e^{r\tau}\mu(t,\tau)^2V(t,\tau) - 3\mu(t,\tau)^4}{[e^{r\tau}V(t,\tau) - \mu(t,\tau)^2]^2} - 3 \tag{3.34}$$

where

$$\mu(t,\tau) = e^{r\tau} - 1 - e^{r\tau}V(t,\tau)/2 - e^{r\tau}W(t,\tau)/6 - e^{r\tau}X(t,\tau)/24$$

As equations (3.28), (3.29) and (3.30) indicate, the procedure involves using a weighted sum of (out-of-the-money) options across varying strike prices to construct the prices of payoffs related to the second, third and fourth moments of returns. These prices are then used to construct estimates of the mean, variance, skewness and excess kurtosis of the risk neutral density function. We plot the daily skewness and kurtosis extracted from the daily S&P 500 index options from January 1996 to December 2005 in Figure 7, where each dot represents a combination of squared skewness and kurtosis on each day and the integrals in equations (3.28), (3.29) and (3.30) are evaluated by trapezoid approximation (see Conrad et al. (2007b) for more details).

In Figure 7 we also plot, on top of the dots, the feasible domains of the NIG distribution and the VG distribution and the regions implied by the sufficient and necessary conditions of the GST distribution. Also available is the feasible domain of the A-type Gram-Charlier expansion. Eriksson et al. (2009) studied the A-type Gram-Charlier expansion and the Edgeworth expansion as alternative approximating densities to the NIG distribution. Since the Edgeworth expansion admits a smaller feasible region than the A-type Gram-Charlier expansion (see Barton and Dennis (1952) for more detail), we only draw the feasible domain of the A-type Gram-Charlier expansion in Figure 7. However, we could barely see the regions concerning the GST distribution and the feasible domain of the A-type Gram-Charlier expansion from plots in the left column of Figure 7, which illustrate the whole time series from 1996 through 2005. By zooming in the lower left corner of each plot, we obtain the associated plots in the right column, which give us a clear idea of the region concerning the GST distribution and the feasible domain of the A-type Gram-Charlier expansion.

It is striking to see that most of the options traded in the US produce a combination of skewness and kurtosis outside the feasible regions of the GST distribution and the A-type Gram-Charlier expansion (outside the feasible region of the Edgeworth expansion as well). The GST distribution admits a larger region than the two expansions though. The feasible domains of the NIG distribution and the VG distribution could cover half of the data points, and hence they are superior to the GST law, the Edgeworth expansion and the A-type Gram-Charlier expansion in terms of the feasible region. Although the VG distribution produces a larger region than the NIG distribution, the difference is not significant as shown in Figure 7.

We further look into the risk neutral skewness and kurtosis for 1999, 2000, 2003 in particular and provide the information together with the feasible domains of the NIG distribution, the VG distribution and the GST distribution in Figure 8, where one could clearly see that most of the data points are within the feasible regions of the NIG distribution and the VG distribution and that of the GST distribution is too tight.

3.5 Density Approximation and Option Pricing

That the NIG distribution and the VG distribution produce larger feasible domains than the GST distribution, the A-type Gram-Charlier expansion and the Edgeworth expansion makes the NIG distribution and the VG distribution more flexible in modeling the risk neutral distribution. However, we are unable to numerically assess how good they are and how close they are to the true risk neutral distribution in that the latter is unknown to us. In this section, we will consider a heuristic option pricing model – affine jump-diffusion model – under the risk neutral measure. Assuming that the underlying asset is generated from the affine model, we could obtain a closed-form expression for the true risk neutral distribution as well as the approximating densities whose parameters are estimated using the algorithm described in the previous section. We could further investigate option pricing under the true density and the approximating densities as well in order to compare their pricing errors.

3.5.1 The model

let S be the price process and $Y = \ln(S)$. We assume that Y is generated from the following affine model under risk neutral measure:

$$\begin{aligned} dY_t &= (r - \lambda_J \bar{\mu} - \tfrac{1}{2}V_t)dt + \sqrt{V_t}dW_t^1 + J_t dN_t \\ dV_t &= \kappa(\theta - V_t)dt + \sigma\rho\sqrt{V_t}dW_t^1 + \sigma\sqrt{1-\rho^2}\sqrt{V_t}dW_t^2 \\ N_t &\sim Poisson(\lambda_J) \\ J_t &\sim N(\mu_J, \sigma_J^2) \end{aligned} \tag{3.35}$$

where $\bar{\mu} = \exp(\mu_J + \frac{1}{2}\sigma_J^2) - 1$ is the mean jump size of the price process and W^1, W^2 are two independent standard Brownian motions.

For $u \in \mathbb{C}$, the conditional characteristic function of log price, $E(e^{uY_T}|\mathcal{F}_t)$, is equal to, according to Duffie et al. (2000),

$$\Psi(u; t, T, x_t) \doteq \exp(\alpha(u, T-t) + \beta(u, T-t)v_t + uy_t) \tag{3.36}$$

where

$$\beta(u,\tau) = -\frac{a(1-e^{-\gamma\tau})}{2\gamma - (\gamma+b)(1-e^{-\gamma\tau})}$$

$$\begin{aligned} \alpha(u,\tau) =& ru\tau - \kappa\theta\left(\frac{\gamma+b}{\sigma^2}\tau + \frac{2}{\sigma^2}\ln\left[1 - \frac{\gamma+b}{2\gamma}(1-e^{-\gamma\tau})\right]\right) \\ & - \lambda_J\tau(1+\bar{\mu}u) + \lambda_J\tau\exp(\mu_J u + \frac{1}{2}\sigma_J^2 u^2) \end{aligned}$$

and $b = \sigma\rho u - \kappa, a = u(1-u), \gamma = \sqrt{b^2 + a\sigma^2}$, and $x_t \doteq (y_t, v_t)$.

Therefore, the density function of Y_T conditional on the information up to time t is

$$f(y; t, T, x_t) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-iuy}\Psi(iu; t, T, x_t)du$$

Moreover, the arbitrage-free price of a plain vanilla European call option with expiration time

T and strike price K at time t can be explicitly expressed as

$$
\begin{aligned}
C_t &= E(e^{-r(T-t)}(S_T - K)^+ | \mathcal{F}_t) \\
&= P_1(K, t, T, x_t) - K P_2(K, t, T, x_t)
\end{aligned}
$$

with

$$
P_1(K, t, T, x_t) = \frac{1}{2} s_t - \frac{e^{-r(T-t)}}{\pi} \int_0^\infty Im \left[\frac{e^{iv \ln(K)} \Psi(1 - iv; t, T, x_t)}{v} \right] dv
$$

$$
P_2(K, t, T, x_t) = \frac{1}{2} e^{-r(T-t)} - \frac{e^{-r(T-t)}}{\pi} \int_0^\infty Im \left[\frac{e^{iv \ln(K)} \Psi(-iv; t, T, x_t)}{v} \right] dv
$$

where Im denotes the imaginary part of a complex number.

3.5.2 Numerical Analysis

To numerically evaluate how close the approximating densities to the true density, we will consider the affine model with structural parameters estimated from three different data sets.

We start with the parameters estimated by Bakshi et al. (1997). Namely, $r = 5\%$, $\kappa = 1.62$, $\theta = 0.04$, $\sigma = 0.44$, $\rho = -0.76$ and $\lambda_J = \mu_J = \sigma_J = 0$ where the jump component is suppressed to zero[5]. These values are obtained based on the S&P 500 index call option from June 1988 through May 1991. The starting values of state variables are $S_0 = 1080$ and $V_0 = 0.026$, based on which we can generate a set of European calls with time to maturity $\tau = 0.21$ (in years) and strike prices spanning the interval $[820, 1260]$ at every 20 points. Further, the put prices are calculated using put-call parity.

The second set of parameters we will use is $r = 5.814\%$, $\kappa = 0.6901$, $\theta = 0.0096$, $\sigma = 0.0615$, $\rho = -0.0183$ and $\lambda_J = \mu_J = \sigma_J = 0$. They are estimated by Chernov and Ghysels (2000) using the S&P 500 index and the SPX European options traded on the index from November 1985 to October 1994. Again, the model is evaluated by assuming that no jump occurs. In our numerical analysis, the starting values of the state variables are $s_0 = 1.1804, v_0 = 0.0102$, based

[5] Rompolis and Tzavalis (2007) and Eriksson et al. (2009) considered these values as well.

on which we generate a portfolio of calls and puts with moneyness ranging from 0.87 to 1.13 at every 0.001 points and time to maturity varying from 10 days, 22 days to 44 days.

The last set of parameters are $r = 3.19\%$, $\rho = -0.79$, $\theta = 0.014$, $\kappa = 3.99$, $\sigma = 0.27$, $\lambda_J = 0.11$, $\mu_J = -0.14$, and $\sigma_J = 0.15$. They are estimated by Duffie et al. (2000) based on the S&P 500 index option of Nov 2, 1993. The estimated volatility on that day is $v_0 = 0.0089$. We then derive the value $s_0 = 0.6453$ by simulating a sample path of 1500 observations from model (3.35)[6]. After dropping the first 1000 observations, we look for a pair of (s_t, v_t) with v_t closest to 0.0088. The moneyness ranges from 0.74 to 1.17 at every 0.01 points. We also consider three different time-to-maturities, i.e., 17 days, 45 days, 80 days in the numerical analysis.

The values of structural parameters (annualized) are summarized in the following table:

Parameter	r	κ	θ	σ	ρ	λ_J	μ_J	σ_J
I	5%	1.62	0.04	0.44	-0.76	0	0	0
II	5.814%	0.6901	0.0096	0.0615	-0.0183	0	0	0
III	3.19%	3.99	0.014	0.27	-0.79	0.11	-0.14	0.15

We evaluate the accuracy of various approximating densities through L_1 and L_2 norms which are defined as

$$L_1(f) = \int |f(x) - \hat{f}(x)| dx$$

and

$$L_2(f) = \int (f(x) - \hat{f}(x))^2 dx$$

where f is the true density and $\hat{f}$ is its estimator. These two norms measure the average distance between the true density and the estimated one. L_2 norm is more sensitive than L_1 norm if $\hat{f}$ deviates from f dramatically.

We report the mean, variance, skewness and excess kurtosis across different time to maturities estimated from the three sets of parameters in Table 2, Table 3 and Table 4 respectively. The values of L_1 and L_2 norms are available in the three tables as well. Besides, in Table 2, we also list the estimated L_1 and L_2 norms for the C-type Gram-Charlier expansion which are

[6]To simulate the jump, we follow the algorithm described in Chapter 6 of Cont and Tankov (2004). Namely, we first generate the number of jumps, N, which is Poisson distributed with parameter $\lambda_J * 1500$, then simulate the jump times – $\{U_i, i = 1, 2, \ldots, N\}$ independently and uniformly distributed along $[0, T]$ – and jump sizes – N i.i.d. normal variables with mean μ_J and variance σ_J^2.

taken from Rompolis and Tzavalis (2007).

From Table 2, one could see that when the underlying asset is generated by a diffusion model (no jump included), the NIG distribution, the VG distribution and the GST distribution could achieve a similar accuracy to the C-type Gram-Charlier expansion and they are better than the Edgeworth expansion and the A-type Gram-Charlier expansion. The reason behind this is, as shown in Panel (a) of Figure 9, the Edgeworth expansion and the A-type Gram-Charlier expansion produce negative values, which is caused by the fact that the skewness and kurtosis are outside the feasible domains of the two expansions. Panel (b) of Figure 9 gives the information regarding the true risk neutral density based on the given starting values of state variables and the NIG approximation, the VG approximation, the GST approximation. They all perform pretty well overall. But still we have an impression that the VG distribution and the NIG distribution are a little bit better than the GST distribution. In addition, the VG distribution overestimates the density while NIG underestimates the true density. The deviation is not at all dramatic though.

Table 3 is produced based on Parameter II, where the estimated skewness and excess kurtosis reported in the first four columns are pretty mild and they fall into the feasible domain of the Edgeworth expansion. All the approximating densities provide very good fits to the true risk neutral density. As time to maturity gets longer, the approximations become even more accurate.

However, things get different in Table 4 where the underlying model includes a jump component. The reported skewness and excess kurtosis based on Parameter III are significantly bigger than what are in Table 2 and Table 3 and they are outside the feasible domain of the GST distribution. The Edgeworth expansion and the A-type Gram-Charlier expansion are not valid density functions at all provided the skewness and excess kurtosis. The true risk neutral density in this scenario is more skewed and leptokurtic. Only the NIG distribution and the VG distribution are capable of matching the risk neutral moments. The VG distribution, however, fails to provide a satisfactory approximation compared with the NIG distribution, as read from the L_1 and L_2 norms. Table 4 gives us a clear idea that the NIG distribution outperforms all the other approximating densities and it increases the accuracy as time-to-maturity grows. Figure 10 is based on Parameter III as well, where one could see the true density curve together with its approximations using the NIG distribution, the VG distribution and the Edgeworth expansion. Consistent with the findings in Table 4, the NIG approximation is superior to all the others and it gets better when time to maturity gets longer.

We further look into the issue of option pricing. Figure 11 is concerned with the pricing of the at-the-money (ATM) and out-of-the-money (OTM) European calls. The underlying asset is generated from model (3.35) with Parameter III and 17 days to maturity where the most serious mispricing of the true risk neutral density is observed. Clearly, the NIG distribution provides mcuh less approximation error than the VG distribution and the Edgeworth expansion. However, it is worth noting that the NIG distribution underprices the ATM calls and overprices the contracts when the moneyness is bigger than 1.03, although the mispricing is not significant.

To make more apparent the difference between the NIG pricing and the VG pricing, we consider a (balanced) butterfly trading strategy based on the European call options discussed in Figure 11. The payoff function of the butterfly trading strategy is

$$g(S_T; K) = (S_T - K + a)1_{(K-a \leq S_T \leq K)} + (K + a - S_T)1_{(K < S_T \leq K+a)}.$$

In the numerical analysis, we allow the strike price, K, to range from 0.6 through 0.8 (or from -0.511 to -0.223 for log(strike)), and a to take values .05, .02, and .005 respectively. Although the payoff approaches 0 as a tends to 0, still one could see from Figure 12 that the NIG pricing is closer to the pricing under the true density than the VG pricing, in particular when log(strike) is around -0.45 and -0.4 where the VG approximation deviates severely from the true density.

3.6 Conclusion

This chapter is concerned with the problem of modeling asset returns with a density function built upon mean, variance, skewness and kurtosis. We explore the GH family of distributions in the financial context and in particular we focus on the NIG distribution, the VG distribution and the GST distribution. We provide numerical evidence that the NIG distribution outperforms all the other candidate approximating densities discussed in this chapter. In terms of the feasible region spanned by skewness and kurtosis, the NIG distribution and the VG distribution are superior to the GST distribution, the Edgeworth expansion and the A-type Gram-Charlier expansion. But when the risk neutral density of asset returns comes from an affine jump-diffusion model, the NIG distribution provides a much better fit than the VG distribution.

However, the feasible region of the NIG distribution is still not large enough. In Figure 7, half of the data points are above the solid line, and in particular the points along the upper boundary form a straight line. This fact prompts us to think about a new family of distributions which could accommodate a wider range of skewness and kurtosis and we leave this as a topic for future research.

Appendix

A.1 Appendix to Chapter 1

In this appendix, we present the cited theorems and lemmas throughout Chapter 1 for readers' quick reference.

Theorem A.1 (Theorem 3.1 of Glasserman and Yao (1995)). *Suppose* $\{(A_t, B_t), t \in \mathbb{Z}\}$ *is a strictly stationary ergodic process and one of the conditions*

$$E(\log \|A_0\|)^+ < \infty, \gamma < 0, E(\log \|B_0\|)^+ < \infty$$

or

$$P(A_t \dots A_0 = 0) > 0 \ \textit{for some} \ n \geq 0$$

is satisfied, then

$$y_t = \sum_{j=1}^{\infty} [A_t \dots A_{t-j+1}] B_{t-j}, \ \ t \in \mathbb{Z} \tag{A.1}$$

converges almost surely. It is the only strictly stationary ergodic solution of $Y_t = A_t Y_{t-1} + B_t$.

Theorem A.2 (Lemma 3.4 of Bougerol and Picard (1992)). *Let* $\{F_t, t \in \mathbb{Z}\}$ *be a strictly stationary ergodic sequence of* $\mathbb{R}^{n \times n}$*-valued random matrices and suppose that* $E(\log^+ \|F_0\|) < \infty$ *and that*

$$\lim_{t \to \infty} \|F_t F_{t-1} \dots F_1\| = 0.$$

Then the top Lyapounov exponent associated with this sequence is strictly negative.

Theorem A.3. *(Theorem 4.3 of Mokkadem (1990); Theorem 1 of Carrasco and Chen (2002)) Given a polynomial random coefficient vector autoregressive model defined as*

$$Y_t = A(\varepsilon_t) Y_{t-1} + B(\varepsilon_t) \tag{A.2}$$

where $\{Y_t, t \in \mathbb{Z}^+\}$ *is a sequence of* $\mathbb{R}^m$*-valued random process,* $\{\varepsilon_t\}$ *is a* $\mathbb{R}^p-$*valued iid sequence,* $A(.)$ *is a* $m \times m$ *matrix-valued polynomial function and* $B(.)$ *is a* $m \times 1$ *vector-valued polynomial function. If further it satisfies the following assumptions:*

(A.1) The marginal probability distribution of ε_t *is absolutely continuous with respect to Lebesgue measure on* $\mathbb{R}^p$ *and zero is in the interior of its support.*

(A.2) $A(.)$ *and* $B(.)$ *are measurable with respect to the sigma-field generated by* ε_t*.*

(A.3) The spectral radius of $A(0)$*, denoted by* $\rho[A(0)]$*, is less than 1.*

(A.4) The series $\sum_{k=1}^{\infty}[A(\varepsilon_t)A(\varepsilon_{t-1})\dots A(\varepsilon_{t-k})]B(\varepsilon_{t-k-1})$ *converges almost surely. The sequence* $A(\varepsilon_t)A(\varepsilon_{t-1})\dots A(\varepsilon_{t-k})$ *converges (as* $k \to \infty$*) to the 0 matrix almost surely.*

(A.5) There exists a positive function V *on* $\mathbb{R}^m$*, a compact set* K *of* $\mathbb{R}^m$ *with nonempty interior, and some positive numbers* $\delta > 0, \nu > 0$*, and* $0 < \lambda < 1$ *such that*

(i) $E[V(Y_t)|Y_{t-1} = y] \le \lambda V(y) - \nu$ *if* $x \notin K$

(ii) $E[V(Y_t)|Y_{t-1} = y] \le \delta$ *if* $x \in K$

Then $\{Y_t\}$ *is Markov geometrically ergodic and* $E[V(Y_t)] < \infty$*. Moreover, if* Y_0 *is initialized from an invariant distribution,* $\{Y_t\}$ *is strictly stationary and* β*-mixing with exponential decay.*

Theorem A.4. *(Theorem 2 of Nelson and Cao (1992)) For a GARCH(2,q) as below*

$$r_t = \sigma_t \varepsilon_t$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{2} \beta_i \sigma_{t-i}^2 + \sum_{j=1}^{q} \alpha_j r_{t-j}^2$$

where ε_t*'s are iid and* $E(\varepsilon_t) = 0, var(\varepsilon_t) = 1$*. Let* z_1 *and* z_2 *be the roots of* $1 - \sum_{i=1}^{2} \beta_i z^{-i}$ *such that* $|z_2| \le |z_1| \le 1$ *and if* $z_1 = -z_2$*, we take* $z_1 > 0$*. Suppose further that* $1 - \sum_{i=1}^{2} \beta_i z^i$ *and* $\sum_{j=1}^{q} \alpha_j z^{j-1}$ *have no common roots. If we write* σ_t^2 *in ARCH(*∞*) form:*

$$\sigma_t^2 = \omega^\star + \sum_{k=0}^{\infty} \phi_k r_{t-k-1}^2,$$

then $\omega^\star \geq 0$ *and* $\phi_k \geq 0 (\forall k)$ *if and only if*

(B.1) $\omega^\star = \omega/(1 - z_1 - z_2 + z_1 z_2) \geq 0$

(B.2) z_1 *and* z_2 *are real*

(B.3) $z_1 > 0$

(B.4) $\sum_{j=0}^{q-1} z_1^{-j} \alpha_{j+1} > 0$

(B.5) $\phi_k \geq 0$ *for* $k = 0$ *to* q.

Theorem A.5. *(Lemma 1 of Jensen and Rahbek (2004)) Consider* $L_T(\Phi)$*, which is a function of the observations* $\{X_t\}_{1 \leq t \leq T}$ *and the parameter* $\Phi \in O \subseteq \mathbb{R}^k$*. Let* Φ_0 *be is an interior point of* O*. Assume that* $L_T(.) : \mathbb{R}^k \to \mathbb{R}$ *is three times continuously differentiable in* Φ *and that*

(A.1) As $T \to \infty$*,* $\sqrt{T} \nabla L_T(\Phi_0) \Rightarrow N(0, \Sigma_S)$*,* $\Sigma_S > 0$.

(A.2) As $T \to \infty$*,* $H(L_T)(\Phi_0) \xrightarrow{P} \Sigma_I > 0$.

(A.3) $\max_{i,j,h=1,\ldots,k} \sup_{\Phi \in N(\Phi_0)} |\frac{\partial^3 L_T(\Phi)}{\partial \phi_i \partial \phi_j \partial \phi_k}| \leq c_T$

where $N(\Phi_0)$ *is a neighborhood of* Φ_0 *and* $0 \leq c_T \xrightarrow{P} c$*,* $0 < c < \infty$*. Then there exists a fixed open neighborhood* $U(\Phi_0) \subseteq N(\Phi_0)$ *of* Φ_0 *such that*

(B.1) With probability tending to one as $T \to \infty$*, there exists a minimum point* $\hat{\Phi}_T$ *of* $L_T(\Phi)$ *in* $U(\Phi_0)$*. In particular,* $\hat{\Phi}_T$ *is unique and solves* $\nabla L_T(\hat{\Phi}_T) = 0$

(B.2) As $T \to \infty$*,* $\hat{\Phi}_T \xrightarrow{P} \Phi_0$.

(B.3) As $T \to \infty$*,* $\sqrt{T}(\hat{\Phi}_T - \Phi_0) \Rightarrow N(0, \Sigma_I^{-1} \Sigma_S \Sigma_I^{-1})$

A.2 Appendix to Chapter 2

A.2.1 Edgeworth and Gram-Charlier series expansions

The central idea of the Edgeworth expansion and the A-type Gram-Charlier expansion is to expand the risk-neutral density around the normal distribution using Hermite polynomials. Consider a random variable X with mean μ, standard deviation σ, skewness S and excess kurtosis K and let $Z = \frac{X-\mu}{\sigma}$, the true density function $f(x)$ of X can be expanded as

$$f(x) = g(x)\left[1 + \sum_{m=3}^{\infty} \frac{1}{m!} E_{X,m} H_m(z)\right]$$

where $g(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{1}{2}z^2)$, $H_m(.)$ are the Hermite polynomials and $E_{X,m}$ is the series expansion coefficient defined as $E_{X,m} = E(H_m(Z))$. If truncate the infinite series at some finite order, we will have the Edgeworth expansion and the A-type Gram-Charlier expansion. In particular, in term of the Edgeworth expansion, the true density $f(x)$ of X can be expressed as

$$f(x) = g(x)\left[1 + \frac{1}{6}S(z^3 - 3z) + \frac{1}{24}K(z^4 - 6z^2 + 3) + \frac{1}{72}S^2(z^6 - 15z^4 + 45z^2 - 15)\right]$$

If drop the last term within the bracket, it yields the A-type Gram-Charlier expansion. Namely,

$$f(x) = g(x)\left[1 + \frac{1}{6}S(z^3 - 3z) + \frac{1}{24}K(z^4 - 6z^2 + 3)\right]$$

The C-type Gram-Charlier expansion is an improvement over the A-type Gram-Charlier expansion. It relaxes the dependence on the Gaussian density and produces positive estimate of density function. The C-type Gram-Charlier expansion implies that the true density of X is expanded as

$$f(x) = \frac{\exp[\sum_{m=1}^{\infty} \frac{1}{m}\delta_m H_m(z)]}{\int \exp[\sum_{m=1}^{\infty} \frac{1}{m}\delta_m H_m(z)]dx}$$

where δ_m is the m^{th} order series coefficient of the C-type Gram-Charlier expansion. [1]

[1] See Rompolis and Tzavalis (2005) and Rompolis and Tzavalis (2007) for more technical details

A.2.2 Call option pricing using Heston model

For European option written on the asset prices generated from model (2.19), Heston (1993) gave an explicit formula to calculate the call prices:

$$Call(t, s, v, K) = sP_1 - e^{-r\tau} KP_2$$

where for $j = 1, 2$,

$$P_j = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty Re[e^{-iu \log(K)\Phi_t^{(j)}(u;\tau)}/(iu)]du \tag{A.3}$$

and $\Phi_t^{(j)}(u;\tau)$ is defined as

$$\Phi_t^{(j)}(u;\tau) = \exp(C^{(j)}(u;\tau) + D^{(j)}(u;\tau)v + iu\ln(s)) \tag{A.4}$$

with

(i) when $\sigma > 0$,

$$\begin{aligned}
C^{(j)}(u;\tau) &= ru\tau i + \frac{\kappa\theta}{\sigma^2}[(b_j - \rho\sigma ui + d)\tau - 2\ln(\frac{1 - ge^{d\tau}}{1 - g})] \\
D^{(j)}(u;\tau) &= \frac{b_j - \rho\sigma ui + d}{\sigma^2} \frac{1 - e^{d\tau}}{1 - ge^{d\tau}} \\
g &= \frac{b_j - \rho\sigma ui + d}{b_j - \rho\sigma ui - d} \\
d &= \sqrt{(\rho\sigma ui - b_j)^2 + \sigma^2(u^2 - 2u_j ui)}
\end{aligned}$$

(ii) when $\sigma = 0$,

$$\begin{aligned}
C^{(j)}(u;\tau) &= ru\tau i + \frac{\kappa\theta(u^2 - 2u_j ui)}{2b_j}(\frac{1 - e^{-b_j\tau}}{b_j} - \tau) \\
D^{(j)}(u;\tau) &= \frac{u^2 - 2u_j ui}{2b_j}(e^{-b_j\tau} - 1)
\end{aligned}$$

and

$$b_1 = \kappa - \rho\sigma, u_1 = 1/2; b_2 = \kappa, u_2 = -1/2 \tag{A.5}$$

Remark A.1. *One could see that the characteristic function (2.20) is $\Phi_t^{(2)}(u;\tau)$ of (A.4).*

Table 1: Option pricing: Comparison of NIG with Gram-Charlier and Edgeworth

Panel A reports the values of L_1 and L_2 which measure how close the approximations by NIG, Gram-Charlier and Edgeworth expansions are to the true density function on average. The mean absolute error L_1 and the mean percentage error L_2 are defined as

$$L_1(f) = \int |f(x) - \hat{f}(x)| dx \text{ and } L_2(f) = \int (f(x) - \hat{f}(x))^2 dx$$

where f is the true density and $\hat{f}$ is its estimate. And they are derived based on Heston model (2.19). Panel B illustrates the pricing of Butterfly trading strategy with payoff function

$$g(S_T; K) = (S_T - K + 50)1_{(K-50 \leq S_T \leq K)} + (K + 50 - S_T)1_{(K < S_T \leq K+50)}$$

using the true risk neutral density, NIG law, Edgeworth and A-type Gram-Charlier expansions when K takes values at $e^{6.60}, e^{6.70}, e^{6.80}$ and $e^{6.90}$. The underlying stock prices are generated using Heston model (2.19) with parameters given in section 2.4.1.

Panel A: Density comparisons

	NIG	Edgeworth	A-type GCSE	C-type GCSE
L_1 Norm	0.0690	0.1620	0.2818	0.0781
L_2 Norm	0.0244	0.0660	0.2794	0.0153

Panel B: Butterfly Trading Strategy

$\ln(K)$	True RND	NIG	Edgeworth	A-type GCSE
6.60	0.0238	0.0298	0.0148	0.0064
6.70	0.1723	0.1824	0.3654	0.3150
6.80	0.9367	0.9022	0.8019	1.4414
6.90	4.3490	4.3874	4.0484	1.9144

Figure 1: Admissible regions for 1 month TTM S&P 500 index options

The figure shows the feasible region of NIG distribution and the region of positive definiteness for A-type Gram-Charlier in terms of the kurtosis and the squared skewness. The areas below the curves are admissible regions. The region of positive definiteness is obtained via the dialytic method of Sylvester(see, for instance Shenton (1951), Barton and Dennis (1952)). Superimposed are the kurtosis/squared skewness from 1 month time to maturity S&P 500 index options (daily data) for 1999 (top), 2000 (middle) and 2003 (lower).

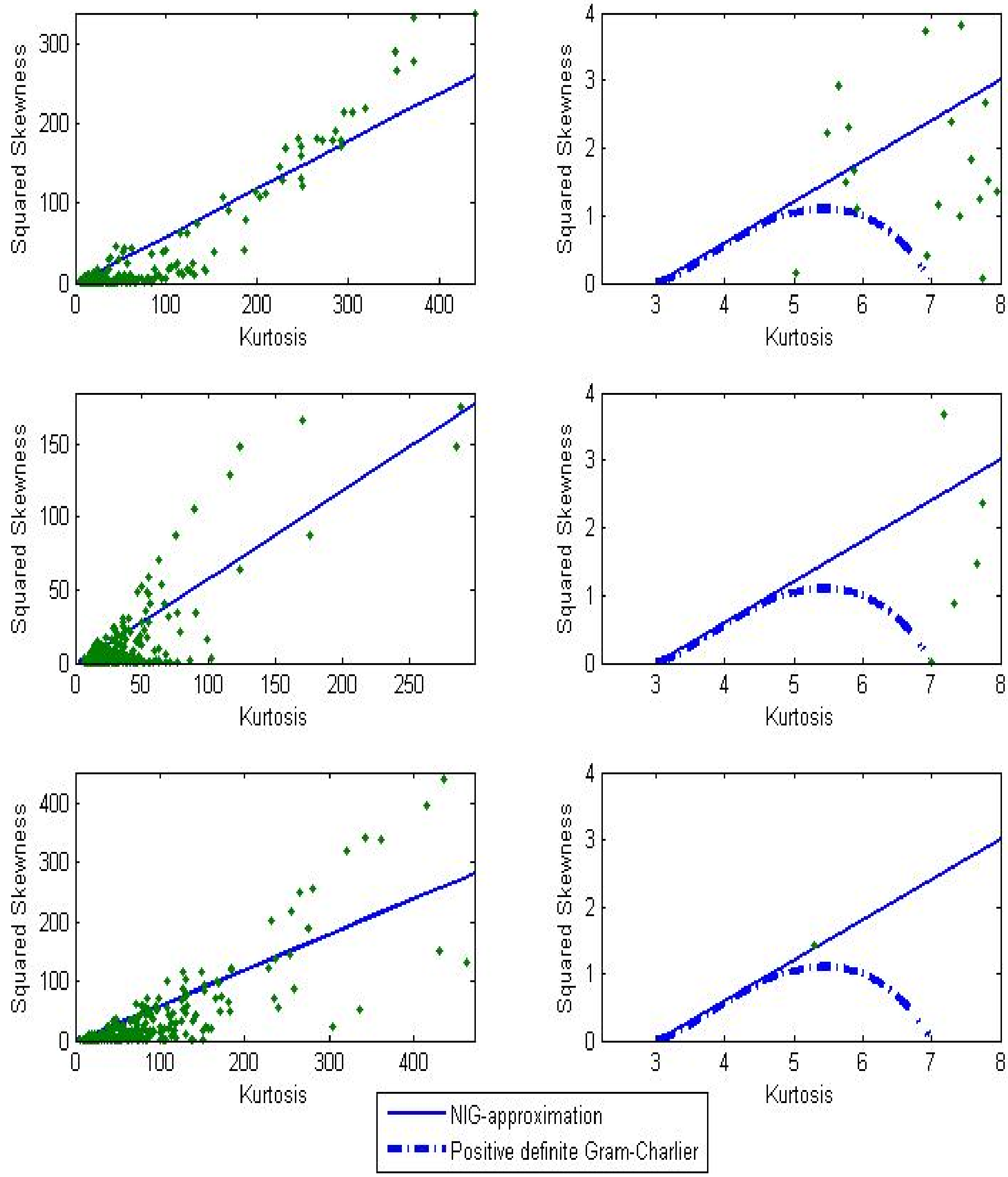


Figure 2: Risk Neutral Density of Heston Model

The figure plots the true probability density curve of the log price $\ln(S_{\tau+t_0})$ (with $\tau = 0.21$ in year) conditional at time $t_0 = 0$ together with its approximations by the NIG distribution, Edgeworth and A-type Gram-Charlier expansions. The stock price S_t and its volatility V_t are generated using Heston model (2.19) under risk-neutralized pricing probability, namely:

$$\begin{aligned} dS_t &= S_t(rdt + \sqrt{V_t}dW_t^1) \\ dV_t &= \kappa(\theta - V_t)dt + \sigma\sqrt{V_t}dW_t^2 \end{aligned}$$

where W_t^1 and W_t^2 are two correlated Brownian motions with correlation coefficient ρ. In our numerical evaluation, we set up the parameters the same as Rompolis and Tzavalis (2007), ie $r = 0.05$, $\kappa = 1.62$, $\theta = 0.04$, $\sigma = 0.44$ and $\rho = -0.76$. Starting from time 0 with $S_0 = 1080, V_0 = 0.026$, we generate a cross-section set of European calls with time to maturity τ and strike prices spanning the interval $[820, 1260]$ at every 20 points. The put prices are calculated using call-put parity. And the risk neutral moments are computed using the methodology of Bakshi et al. (2003).

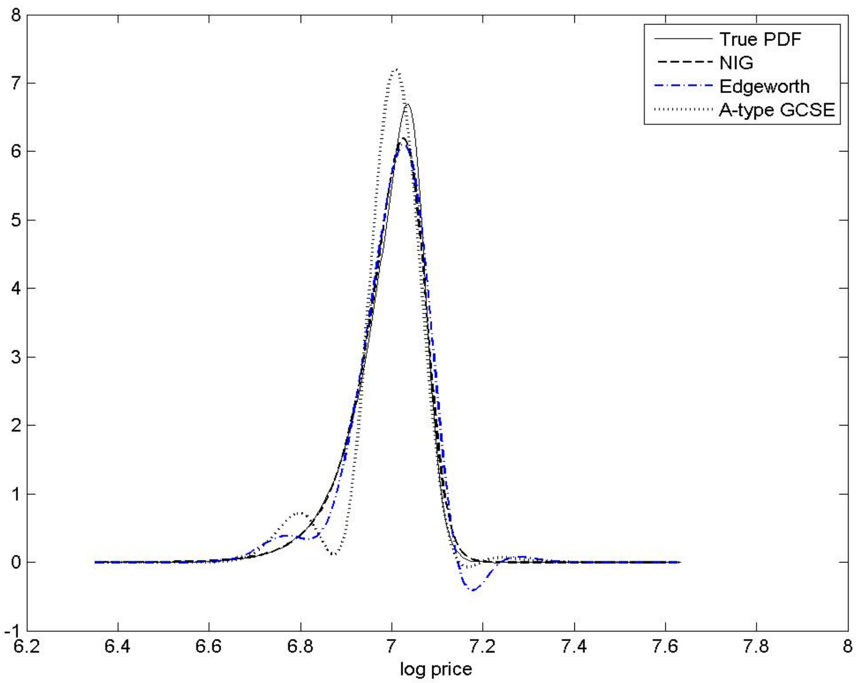
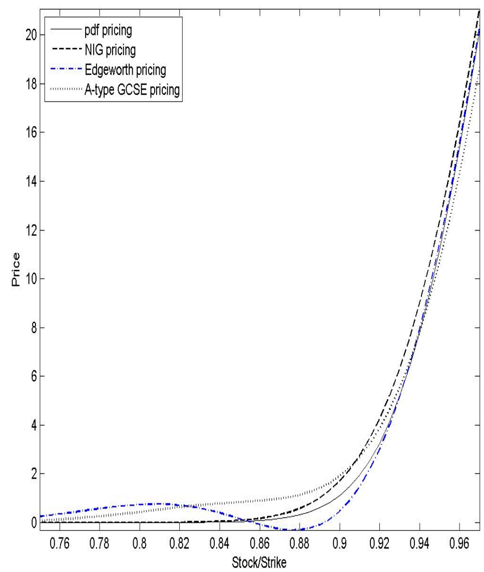
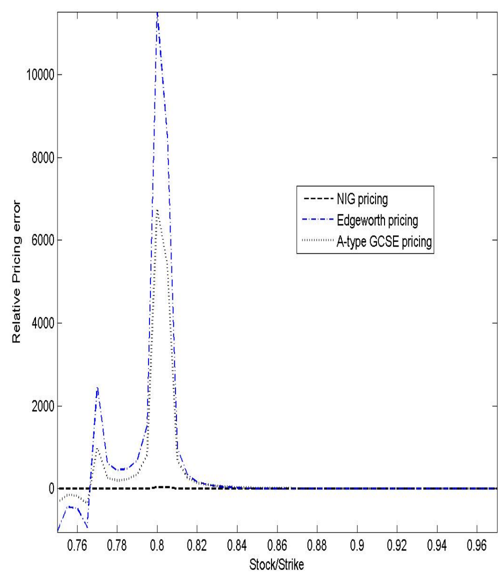


Figure 3: Pricing of European calls: OTM & ATM

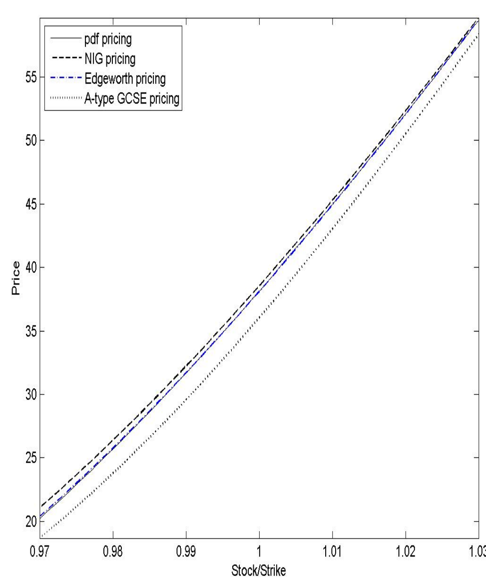
The figure displays the pricing of the European calls under the true risk neutral measure, NIG approximation, Edgeworth and A-type Gram-Charlier expansions, and the associated relative pricing errors which are defined as $(\hat{p} - p)/p$ where p is the price under the true risk neutral measure and $\hat{p}$ is the price using the approximation. The strike prices range from $1080/0.970 - 1080/0.750$ in Panel (a,b) and from $1080/1.030 - 1080/0.970$ in Panel (c,d). The underlying stock prices are generated using Heston model (2.19) with parameters specified in section 2.4.1.



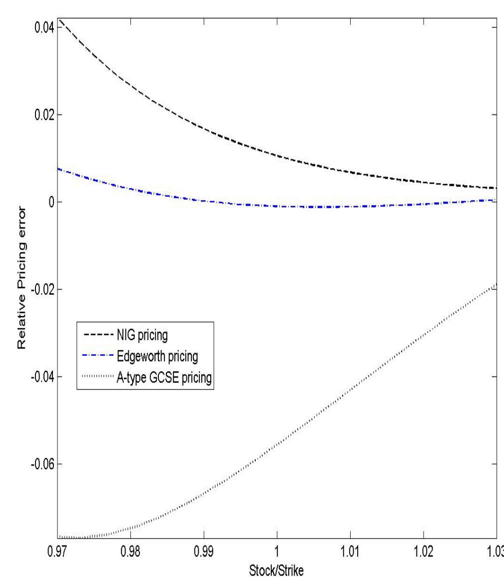
(a) OTM



(b) Relative Pricing Error: OTM



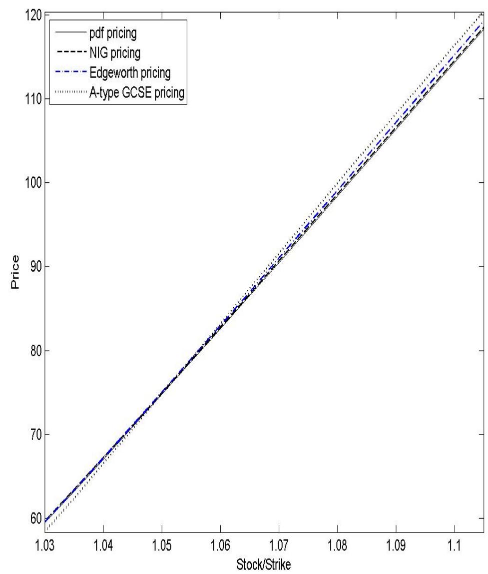
(c) ATM



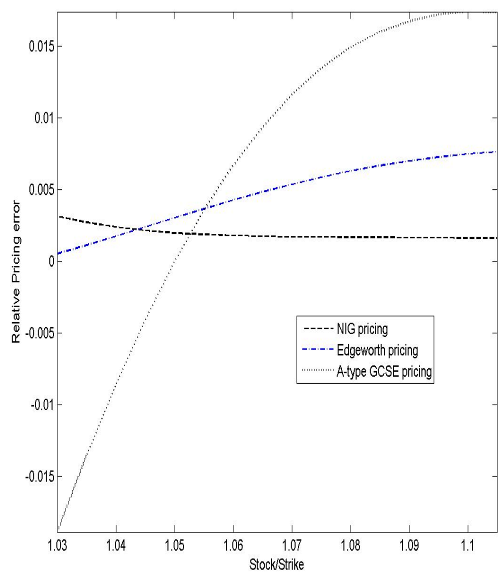
(d) Relative Pricing Error: ATM

Figure 4: Pricing of European calls: ITM

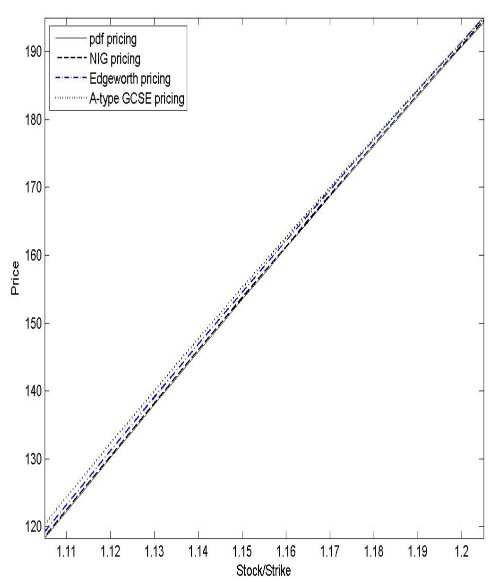
The figure displays the pricing of the European calls under the true risk neutral measure, NIG approximation, Edgeworth and A-type Gram-Charlier expansions, and the associated relative pricing errors which are defined as $(\hat{p} - p)/p$ where p is the price under the true risk neutral measure and $\hat{p}$ is the price using the approximation. The strike prices range from $1080/1.105 - 1080/1.030$ in Panel (a,b) and from $1080/1.205 - 1080/1.105$ in Panel (c,d). The underlying stock prices are generated using Heston model (2.19) with parameters specified in section 2.4.1.



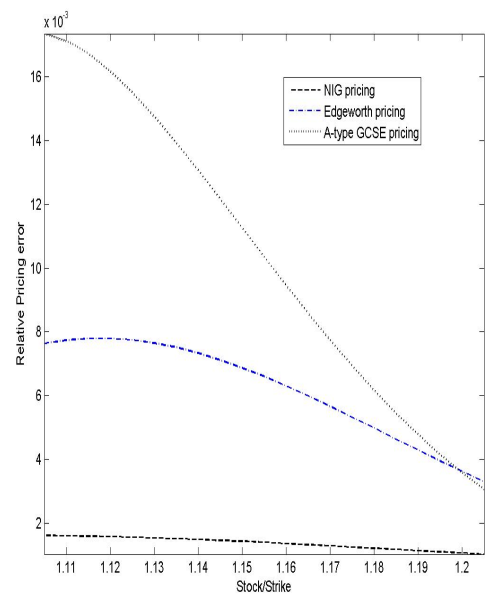
(a) ITM



(b) Relative Pricing Error: ITM



(c) ITM



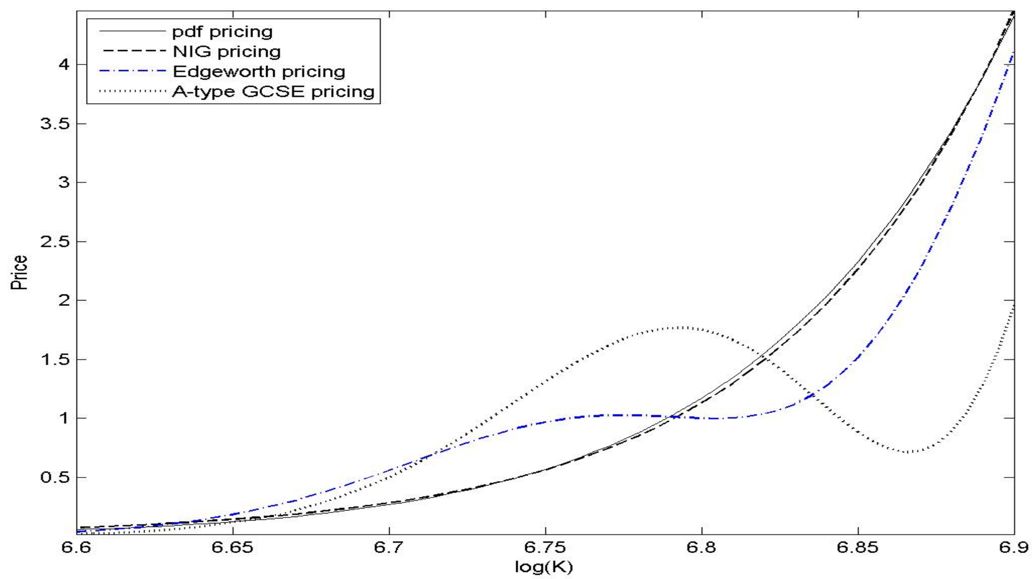
(d) Relative Pricing Error: ITM

Figure 5: Pricing of Butterfly Trading Strategy

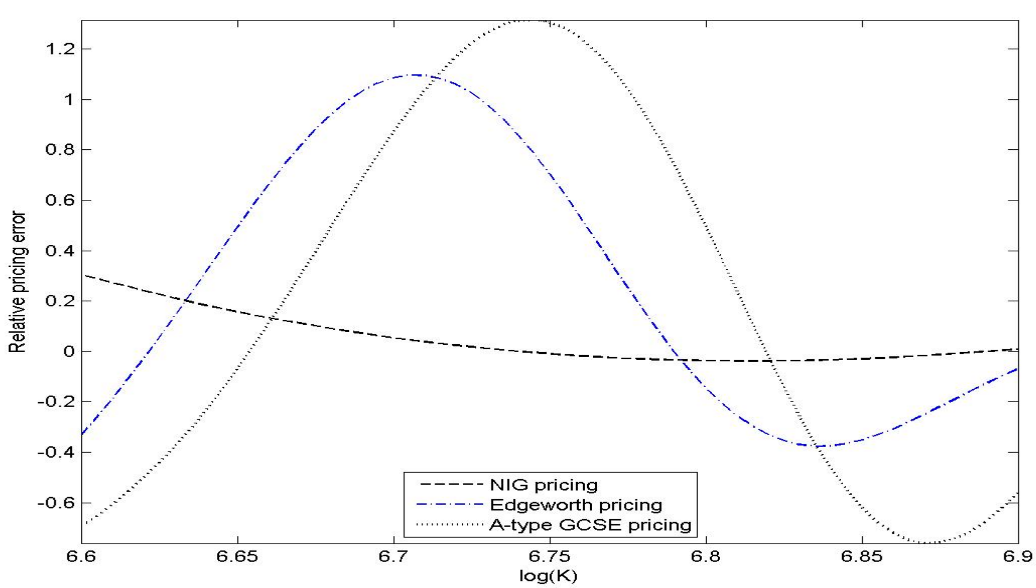
Panel (a) plots the pricing of Butterfly trading strategy with payoff function

$$g(S_T; K) = (S_T - K + 50)1_{(K-50 \leq S_T \leq K)} + (K + 50 - S_T)1_{(K < S_T \leq K+50)}$$

using the true risk neutral density, NIG law, Edgeworth and A-type Gram-Charlier expansions when K takes values from $e^{6.6}$ to $e^{6.9}$. The underlying stock prices are generated using Heston model (2.19) with parameters given in section 2.4.1. The associated relative pricing errors are provided in Panel (b).



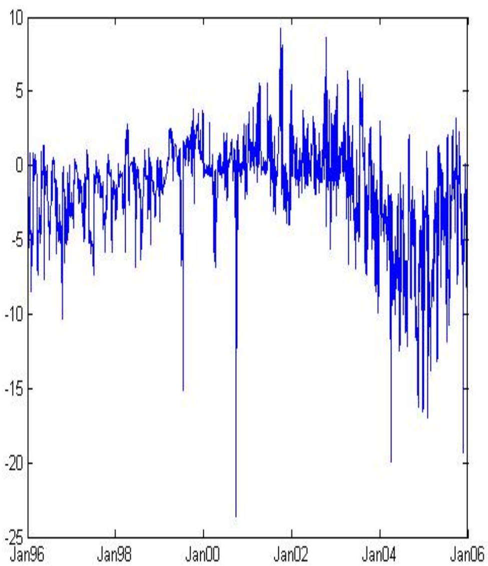
(a) Pricing



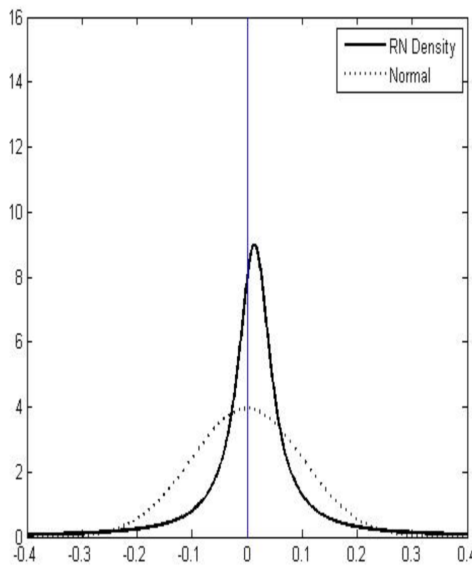
(b) Relative Pricing Error

Figure 6: Risk neutral densities for S&P 500 Index using SPX Option contracts and the NIG Approximation

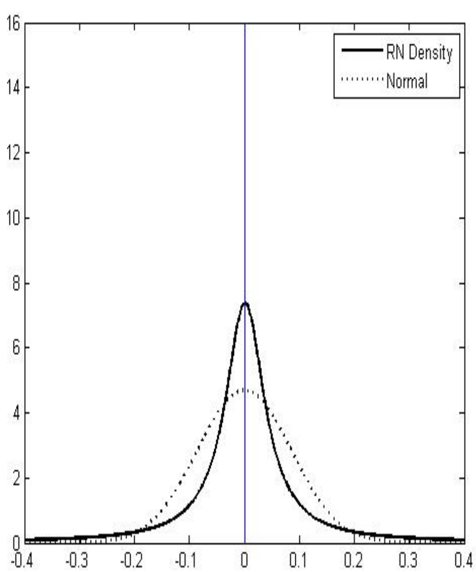
The figure provides a time series plot of S&P 500 and the risk neutral densities approximated by the NIG law, using three month contracts in March 2000 and August 2003. We use daily option price data extracted from Optionmetrics through WRDS for all OTM calls and puts for all stocks from 1996-2005. Closing prices are constructed as midpoint averages of the closing bid and ask prices. We eliminate option prices below 50 cents. In estimating the moments, we use equal numbers of OTM calls and puts for each stock for each day.



(a) S&P 500 time series of returns



(b) 3/2000



(c) 8/2003

A.3 Appendix to Chapter 3

A.3.1 Proof of Theorem 3.8

First, we need to show that, given $0 < C < 1$, there exists a unique $0 < R < 1$ satisfying

$$(C-1)R^3 + (7C-6)R^2 + (7C-9)R + C = 0$$

Note that for $0 < R < 1$,

$$f(R) = \frac{R(3+R)^2}{(1+6R+R^2)(1+R)} \tag{A.6}$$

is continuous and strictly increasing in R with range $(0,1)$. This is because

$$f'(R) = \frac{(R-1)(R+3)(R^2-6R-3)}{(1+6R+R^2)^2(1+R)^2} > 0$$

for $0 < R < 1$.

Next from corollary 3.3, we have

$$M = \mu + \frac{\beta p}{\eta} \tag{A.7}$$

$$V = \frac{p}{\eta^2}(\eta + \beta^2) \tag{A.8}$$

$$S = \frac{\beta(3\eta + 2\beta^2)}{(\eta+\beta^2)^{3/2}p^{1/2}} \tag{A.9}$$

$$K = \frac{3(\eta^2 + 4\eta\beta^2 + 2\beta^4)}{p(\eta+\beta^2)^2} \tag{A.10}$$

where $\eta = \frac{\alpha^2-\beta^2}{2} > 0,\ \alpha > 0, p > 0, \mu \in \mathbb{R}$.

(i) If $\beta = 0$, the above equations are simplified as

$$M = \mu$$

$$V = p/\eta$$

$$S = 0$$

$$K = 3/p$$

hence $\alpha = \sqrt{6/(KV)}, \mu = M, p = 3/K$.

(ii) If $\beta \neq 0$, from (A.8)

$$p^{1/2} = \eta\sqrt{V/(\eta + \beta^2)}$$

Combined with (A.9), we have

$$S = \frac{\beta(3\eta + 2\beta^2)}{(\eta + \beta^2)^{3/2}\eta\sqrt{V/(\eta + \beta^2)}} = \frac{\beta(3\eta + 2\beta^2)}{(\eta + \beta^2)\eta\sqrt{V}} \tag{A.11}$$

Introduce $\rho = \frac{\beta}{\alpha}$ ($|\rho| < 1$). $\beta = \alpha\rho$, $\eta = \alpha^2(1-\rho^2)/2$. From (A.11), we have

$$\begin{aligned} S &= \frac{\alpha\rho(3\alpha^2(1-\rho^2)/2 + 2(\alpha\rho)^2)}{(\alpha^2(1-\rho^2)/2 + (\alpha\rho)^2)(\alpha^2(1-\rho^2)/2)\sqrt{V}} = \frac{2\rho(3+\rho^2)}{\sqrt{V}\alpha(1-\rho^4)} \\ \alpha &= \frac{2\rho(3+\rho^2)}{\sqrt{V}S(1-\rho^4)} \end{aligned}$$

It follows from (A.8) and (A.10) that

$$K = \frac{3(\eta^2 + 4\eta\beta^2 + 2\beta^4)}{V\eta^2(\eta + \beta^2)} \tag{A.12}$$

Considering $\beta^2 = \alpha^2 R$, $\eta = \alpha^2(1-R)/2$ and $\alpha^2 = \frac{4R(3+R)^2}{VS^2(1-R^2)^2}$ where $R = \rho^2 < 1$ in (A.12), we have

$$\frac{3S^2}{2K} = \frac{R(3+R)^2}{(1+6R+R^2)(1+R)} \tag{A.13}$$

Further define $C = \frac{3S^2}{2K}$ where $0 < C < 1$. There exists a unique $0 < R < 1$ satisfying (A.13). It follows immediately that

$$\begin{aligned} \rho &= sig(S)\sqrt{R} \\ \alpha &= \frac{2\rho(3+\rho^2)}{\sqrt{V}S(1-\rho^4)} = \frac{2\sqrt{R}(3+R)}{\sqrt{V}|S|(1-R^2)} \\ \beta &= \rho\alpha = \frac{2R(3+R)}{\sqrt{V}S(1-R^2)} \end{aligned}$$

$$p = \frac{V(1-R)^2\alpha^2}{2(1+R)} = \frac{2R(3+R)^2}{S^2(1+R)^3}$$
$$\mu = M - \frac{2\beta p}{\alpha^2-\beta^2} = M - \frac{2\sqrt{V}R(3+R)}{S(1+R)^2}$$

Therefore, given sample mean $\hat{M}$, sample variance $\hat{V}$, sample skewness $\hat{S}$ and sample excess kurtosis $\hat{K}$ such that $2\hat{K} > 3\hat{S}^2 > 0$, then the method of moments estimation of the VG parameters are

$$\hat{\alpha} = \frac{2\sqrt{R}(3+R)}{\sqrt{\hat{V}}|\hat{S}|(1-R^2)}$$
$$\hat{\beta} = \frac{2R(3+R)}{\sqrt{\hat{V}}\hat{S}(1-R^2)}$$
$$\hat{p} = \frac{2R(3+R)^2}{\hat{S}^2(1+R)^3}$$
$$\hat{\mu} = \hat{M} - \frac{2\sqrt{\hat{V}}R(3+R)}{\hat{S}(1+R)^2}$$

A.3.2 Proof of Theorem 3.9

Note that

$$M = \mu + \frac{b^2\beta}{v-2} \tag{A.14}$$
$$V = \frac{b^2}{v-2} + \frac{2b^4\beta^2}{(v-2)^2(v-4)} \tag{A.15}$$
$$S = \left[6(v-2) + \frac{16b^2\beta^2}{(v-6)}\right] \frac{b\beta(v-4)^{1/2}}{[(v-2)(v-4)+2b^2\beta^2]^{3/2}} \tag{A.16}$$
$$K = \left[\frac{8b^4\beta^4(5v-22)}{(v-6)(v-8)} + \frac{16b^2\beta^2(v-2)(v-4)}{(v-6)} + (v-2)^2(v-4)\right] \frac{6}{[(v-2)(v-4)+2b^2\beta^2]^2} \tag{A.17}$$

where $v = -2p > 0$.

(i) if $\beta = 0$ (iff $S = 0$),

$$M = \mu$$
$$V = b^2/(v-2)$$

$$S = 0$$

$$K = 6/(v-4)$$

hence

$$\mu = M, \quad b = \sqrt{V(2 + 6/K)}, \quad p = -(2 + 3/K).$$

(ii) if $\beta \neq 0$ (iff $S \neq 0$). Introduce $\rho = \frac{2b^2\beta^2}{(v-2)(v-4)}$. So $0 < \rho < b^2\beta^2/12$. Meanwhile (A.15), (A.16) and (A.17) become

$$\begin{aligned} V &= \frac{(1+\rho)b^2}{v-2} \\ S^2 &= \left(3 + \frac{4(v-4)\rho}{v-6}\right)^2 \frac{2\rho}{(v-4)(1+\rho)^3} \\ K &= \left[\frac{2(5v-22)(v-4)\rho^2}{(v-6)(v-8)} + \frac{8(v-4)\rho}{v-6} + 1\right] \frac{6}{(1+\rho)^2(v-4)} \end{aligned}$$

therefore

$$\begin{aligned} b^2 &= \frac{V(v-2)}{1+\rho} \\ \beta^2 &= \frac{\rho(1+\rho)(v-4)}{2V} \end{aligned}$$

where ρ, v are solved from

$$\begin{aligned} S^2 &= \left(3 + \tfrac{4(v-4)\rho}{v-6}\right)^2 \tfrac{2\rho}{(v-4)(1+\rho)^3} \\ K &= \left[\tfrac{2(5v-22)(v-4)\rho^2}{(v-6)(v-8)} + \tfrac{8(v-4)\rho}{v-6} + 1\right] \tfrac{6}{(1+\rho)^2(v-4)} \end{aligned} \tag{A.18}$$

or

$$\begin{aligned} 2\rho[3(v-6) + 4(v-4)\rho]^2 - S^2(v-4)(v-6)^2(1+\rho)^3 = 0 \\ 12(v-4)(5v-22)\rho^2 + 48(v-4)(v-8)\rho + 6(v-6)(v-8) - K(v-4)(v-6)(v-8)(1+\rho)^2 = 0 \end{aligned} \tag{A.19}$$

Further, we have

$$b = \sqrt{\frac{V(v-2)}{1+\rho}}$$

$$
\begin{aligned}
\beta &= sig(S)\sqrt{\frac{\rho(1+\rho)(v-4)}{2V}} \\
\mu &= M - \frac{b^2\beta}{v-2} \\
p &= -v/2
\end{aligned}
$$

Replace M, V, S, K by their empirical counterparts. We have the method of moments estimation for the GST parameters.

Next, we need to verify the conditions under which (A.19) have solutions. Introduce $x = \rho(>0)$ and $y = v - 8(>0)$. Thus (A.18) becomes

$$x^3\left[\frac{32(y+4)}{(y+2)^2} - S^2\right] + 3x^2\left[\frac{16}{y+2} - S^2\right] + 3x\left[\frac{6}{y+4} - S^2\right] - S^2 = 0 \tag{A.20}$$

$$x^2\left[\frac{12(5y+18)}{(y+2)y} - K\right] + 2x\left[\frac{24}{y+2} - K\right] + \left[\frac{6}{y+4} - K\right] = 0 \tag{A.21}$$

Notice that

$$\frac{12(5y+18)}{(y+2)y} > \frac{32(y+4)}{(y+2)^2} > \frac{24}{y+2} > \frac{16}{y+2} > \frac{6}{y+4} > 0$$

To solve a unique $x > 0$ from (A.21), one should have

$$\frac{12(5y+18)}{(y+2)y} > K > \frac{6}{y+4} > 0 \tag{A.22}$$

and

$$x_1(y;K) \equiv \frac{\sqrt{B^2 - AC} - B}{A} > 0$$

where $A = \frac{12(5y+18)}{(y+2)y} - K > 0, B = \frac{24}{y+2} - K$ and $C = \frac{6}{y+4} - K < 0$. To solve a unique $x > 0$ from (A.20) (denoted by $x_2(y;S)$), the necessary and sufficient condition is

$$\frac{6}{y+4} \geq S^2 \tag{A.23}$$

Together with (A.22) and (A.23), we know that y should be within the following set

$$\mathcal{D} = \{y : \left(\frac{6}{K} - 4\right) \vee 0 < y < \left(\frac{30}{K} - 1 + \sqrt{1 + \frac{156}{K} + \frac{900}{K^2}}\right) \wedge \left(\frac{6}{S^2} - 4\right)\}$$

with $S^2 < \min(3/2, K)$.

Next we need to show that x_1 and x_2 cross in $\mathcal{D}$. Define

$$\begin{aligned} f(x; y, K) \equiv & x^2 \left[\frac{12(5y+18)}{(y+2)y} - K \right] + 2x \left[\frac{24}{y+2} - K \right] + \frac{6}{y+4} - K \\ = & Ax^2 + 2Bx + C \\ g(x; y, S) \equiv & x^3 \left[\frac{32(y+4)}{(y+2)^2} - S^2 \right] + 3x^2 \left[\frac{16}{y+2} - S^2 \right] + 3x \left[\frac{6}{y+4} - S^2 \right] - S^2 \\ = & Dx^3 + 3Ex^2 + 3Fx - S^2 \end{aligned}$$

where

$$D = \frac{32(y+4)}{(y+2)^2} - S^2 > 0, E = \frac{16}{y+2} - S^2 > 0, F = \frac{6}{y+4} - S^2 > 0.$$

Since $\lim_{y \to 0} x_1(y; K) = 0$ if $K \geq 3/2$, and $\lim_{y \to 6/K-4} x_1(y; K) = 0$ if $K < 3/2$, we have $g(x_1(y; K); y, S) < 0$.

Note also that

$$S^2 < K - \frac{6}{y+4} \Leftrightarrow \frac{6}{y+4} < K - S^2 \tag{A.24}$$

And

$$g^{'}(x) - f^{'}(x) = 3Dx^2 + 2(3E - A)x + 3F - 2B$$

where

$$\begin{aligned} 3F - 2B &= 3\left(\frac{6}{y+4} - S^2\right) - 2\left(\frac{24}{y+2} - K\right) \\ &= 2K - 3S^2 - \left(\frac{48}{y+2} - \frac{18}{y+4}\right) \\ 3E - A &= 3\left(\frac{16}{y+2} - S^2\right) - \left(\frac{12(5y+18)}{(y+2)y} - K\right) \\ &= K - 3S^2 - \left(\frac{12(5y+18)}{(y+2)y} - \frac{48}{y+2}\right) \end{aligned}$$

Since the condition $\frac{12(5y+18)}{(y+2)y} - \frac{48}{y+2} < K - 3S^2$ implies both (A.24) and $\frac{48}{y+2} - \frac{18}{y+4} < 2K - 3S^2$, it follows that $g^{'} > f^{'}$ if $K > 3S^2$, and y satisfies (A.22), (A.23) and

$$0 < \frac{12(5y+18)}{(y+2)y} - \frac{48}{y+2} < K - 3S^2. \tag{A.25}$$

Further we have $g(x_1(y;K);y,S) > 0$, which implies that x_1 and x_2 cross in $\mathcal{D}$. Therefore the necessary condition that we can solve x, y from (A.20) and (A.21) is $S^2 < \min(3/2, K/3)$. A sufficient condition is $S^2 < \min(3/2, K)$.

Table 2: Comparison of various approximating densities: I

We compare the performance of various approximating densities by assuming that the underlying asset is generated from the affine jump-diffusion model (3.35) under the risk neutral measure, namely,

$$\begin{aligned} dY_t &= (r - \lambda_J \bar{\mu} - \tfrac{1}{2} V_t)dt + \sqrt{V_t} dW_t^1 + J_t dN_t \\ dV_t &= \kappa(\theta - V_t)dt + \sigma\rho\sqrt{V_t} dW_t^1 + \sigma\sqrt{1-\rho^2}\sqrt{V_t} dW_t^2 \\ N_t &\sim Poisson(\lambda_J) \\ J_t &\sim N(\mu_J, \sigma_J^2) \end{aligned}$$

with $\bar{\mu} = \exp(\mu_J + \frac{1}{2}\sigma_J^2) - 1$ and W^1, W^2 are two independent Brownian motions. The values of structural parameters are taken from Bakshi et al. (1997) (Parameter I), ie., $r = 5\%$, $\kappa = 1.62$, $\theta = 0.04$, $\sigma = 0.44$ $\rho = -0.76$ and $\lambda_J = \mu_J = \sigma_J = 0$. The starting values of state variables are $S_0 = 1080$ and $V_0 = 0.026$. The strike prices span the interval $[820, 1260]$ at every 20 points and the time to maturity is $\tau = 0.21$ (in year). The estimated mean, variance, skewness and excess kurtosis are 6.9921, 0.0064, -1.2193 and 3.1090, respectively. The values of L_1 and L_2 norms are reported in the following table where the values regarding the C-type Gram-Charlier series expansion (C-type GCSE) are taken from Rompolis and Tzavalis (2007).

	VG	GST	NIG	Edgeworth	A-type GCSE	C-type GCSE
L_1	0.0659	0.0836	0.0690	0.1620	0.2818	0.0781
L_2	0.0257	0.0395	0.0244	0.0660	0.2794	0.0153

Table 3: Comparison of various approximating densities: II

We compare the performance of various approximating densities by assuming that the underlying asset is generated from the affine jump-diffusion model (3.35) under the risk neutral measure, namely,

$$\begin{aligned} dY_t &= (r - \lambda_J \bar{\mu} - \tfrac{1}{2}V_t)dt + \sqrt{V_t}dW_t^1 + J_t dN_t \\ dV_t &= \kappa(\theta - V_t)dt + \sigma\rho\sqrt{V_t}dW_t^1 + \sigma\sqrt{1-\rho^2}\sqrt{V_t}dW_t^2 \\ N_t &\sim Poisson(\lambda_J) \\ J_t &\sim N(\mu_J, \sigma_J^2) \end{aligned}$$

with $\bar{\mu} = \exp(\mu_J + \frac{1}{2}\sigma_J^2) - 1$ and W^1, W^2 are two independent Brownian motions. Here we consider the values of parameters from Chernov and Ghysels (2000) (Parameter II): $r = 5.814\%$, $\kappa = 0.6901$, $\theta = 0.0096$, $\sigma = 0.0615$ $\rho = -0.0183$ and $\lambda_J = \mu_J = \sigma_J = 0$. The starting values of the state variables are $s_0 = 1.1804, v_0 = 0.0102$ and the moneyness ranges from 0.87 to 1.13 at every 0.001 points.

	Mean	Var	Skew	ExKurt	NIG	Edgeworth	A-GCSE	VG	GST
				Time to maturity: 0.039683 years (10 days)					
L_1	0.1680	0.0004	0.0081	0.2655	0.0423	0.0443	0.0443	0.0435	0.0411
L_2					0.0240	0.0259	0.0259	0.0258	0.0223
				Time to maturity: 0.087302 years (22 days)					
L_1	0.1705	0.0009	0.0062	0.1978	0.0287	0.0298	0.0298	0.0294	0.0280
L_2					0.0074	0.0078	0.0078	0.0078	0.0069
				Time to maturity: 0.1746 years (44 days)					
L_1	0.1751	0.0017	-0.0022	0.1381	0.0168	0.0172	0.0172	0.0171	0.0164
L_2					0.0018	0.0019	0.0019	0.0019	0.0017

Table 4: Comparison of various approximating densities: III

We compare the performance of various approximating densities by assuming that the underlying asset is generated from the affine jump-diffusion model (3.35) under the risk neutral measure, namely,

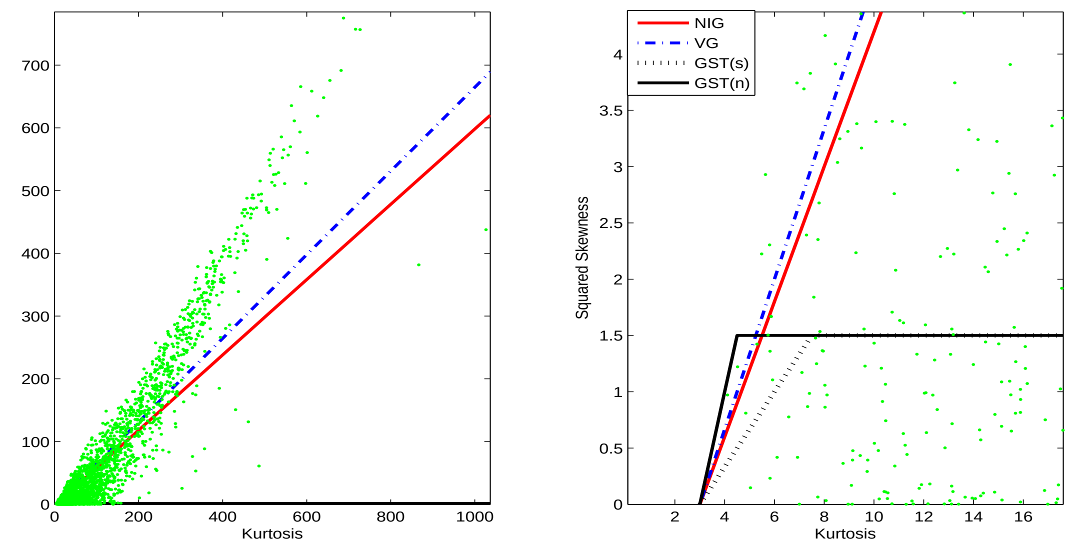
$$\begin{aligned} dY_t &= (r - \lambda_J \bar{\mu} - \tfrac{1}{2}V_t)dt + \sqrt{V_t}dW_t^1 + J_t dN_t \\ dV_t &= \kappa(\theta - V_t)dt + \sigma\rho\sqrt{V_t}dW_t^1 + \sigma\sqrt{1-\rho^2}\sqrt{V_t}dW_t^2 \\ N_t &\sim Poisson(\lambda_J) \\ J_t &\sim N(\mu_J, \sigma_J^2) \end{aligned}$$

with $\bar{\mu} = \exp(\mu_J + \frac{1}{2}\sigma_J^2) - 1$ and W^1, W^2 are two independent Brownian motions.
The structural parameters are taken from Duffie et al. (2000) (Parameter III), ie $r = 3.19\%, \rho = -0.79, \theta = 0.014, \kappa = 3.99, \sigma = 0.27, \lambda_J = 0.11, \mu_J = -0.14, \sigma_J = 0.15$ and the starting values of the state variables are $s_0 = 0.6453, v_0 = 0.0089$. The moneyness spans from 0.74 to 1.17 at every 0.01 points.

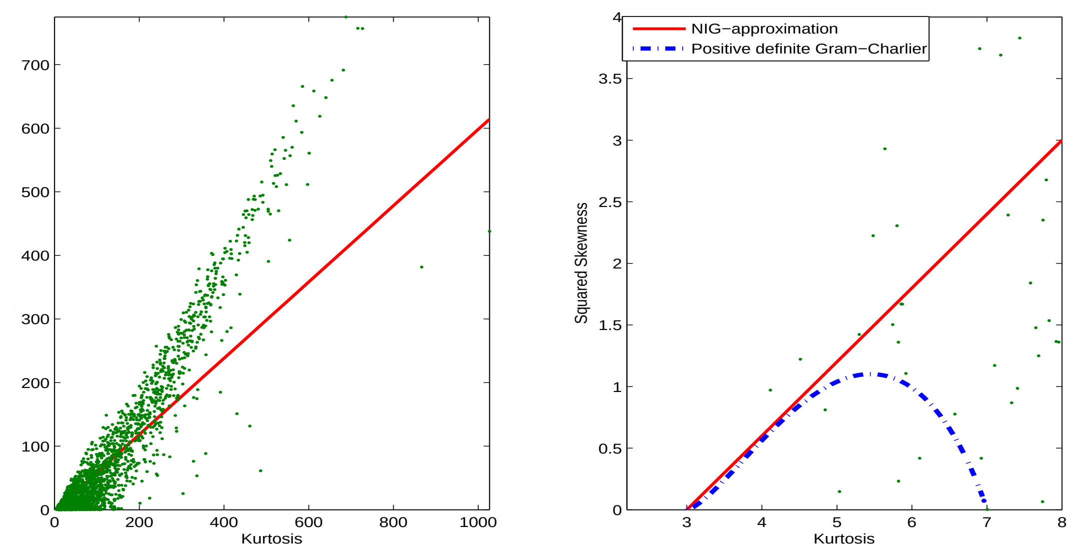
	Mean	Var	Skew	ExKurt	NIG	Edgeworth	A-GCSE	VG	GST
				Time to maturity: 0.06746 years (17 days)					
L_1	-0.4363	0.0009	-3.1160	27.4994	0.4286	1.8116	3.0822	0.7800	NA
L_2					2.7125	19.9109	72.9905	29.1625	NA
				Time to maturity: 0.17857 years (45 days)					
L_1	-0.4336	0.0026	-2.1075	9.5410	0.1417	0.4827	0.9717	0.4072	NA
L_2					0.1508	0.7848	4.4325	3.0147	NA
				Time to maturity: 0.31746 years (80 days)					
L_1	-0.4303	0.0049	-1.6960	5.0476	0.0718	0.2408	0.4602	0.1558	NA
L_2					0.0244	0.1887	0.7646	0.2086	NA

Figure 7: The feasible domains of various approximating densities: 1996-2005

The areas under the curves are feasible domains of the NIG distribution, the VG distribution and the A-type Gram-Charlier expansion. For the GST distribution, its feasible domain is bounded by the areas associated with the 'sufficient condition' and the 'necessary condition'. Each dot represents a combination of squared skewness and kurtosis extracted from daily S&P 500 index options from 1996 to 2005.



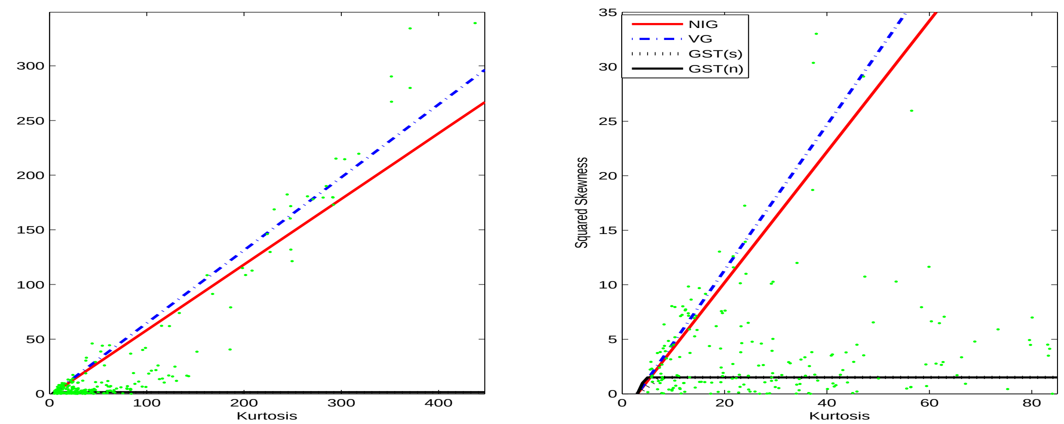
(a) 1996-2005



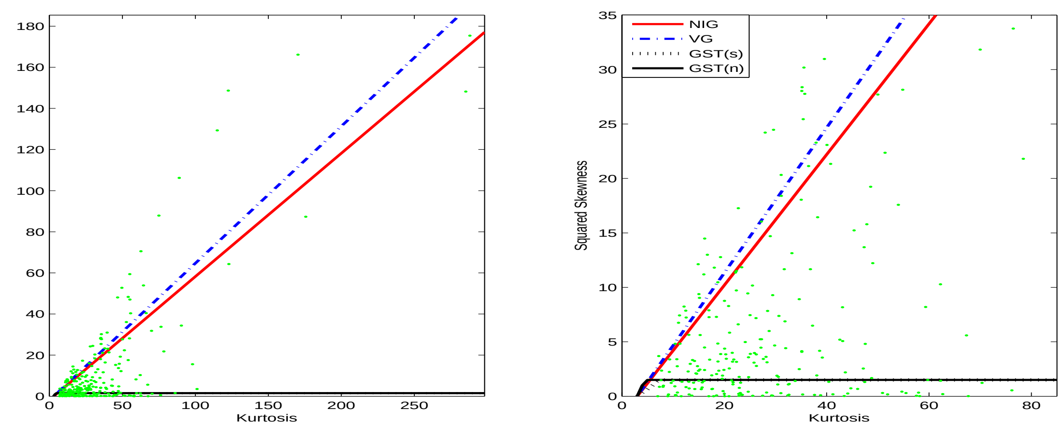
(b) 1996-2005

Figure 8: The feasible domains of various approximating densities: 1999, 2000, 2003

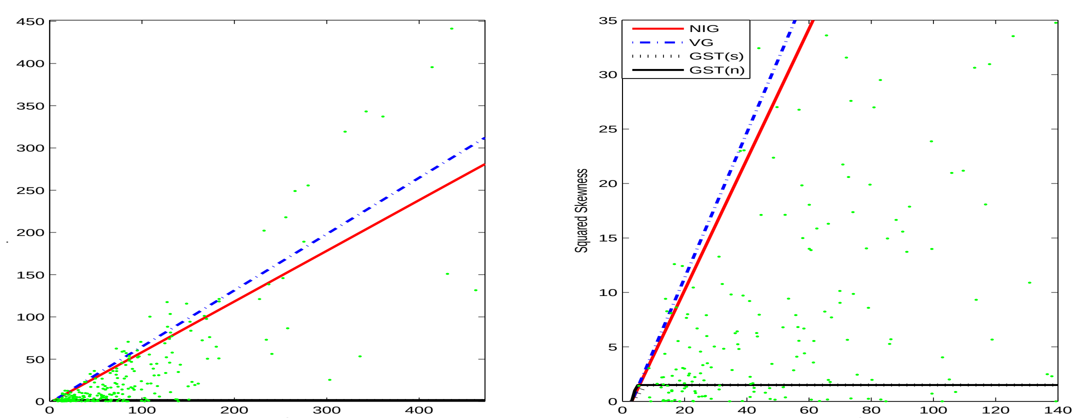
The areas under the curves are feasible domains of the NIG distribution and the VG distribution. For the GST distribution, its feasible domain is bounded by the areas associated with the 'sufficient condition' and the 'necessary condition'. Each dot represents a combination of squared skewness and kurtosis extracted from daily S&P 500 index options for 1999, 2000 and 2003, respectively.



(a) 1999



(b) 2000

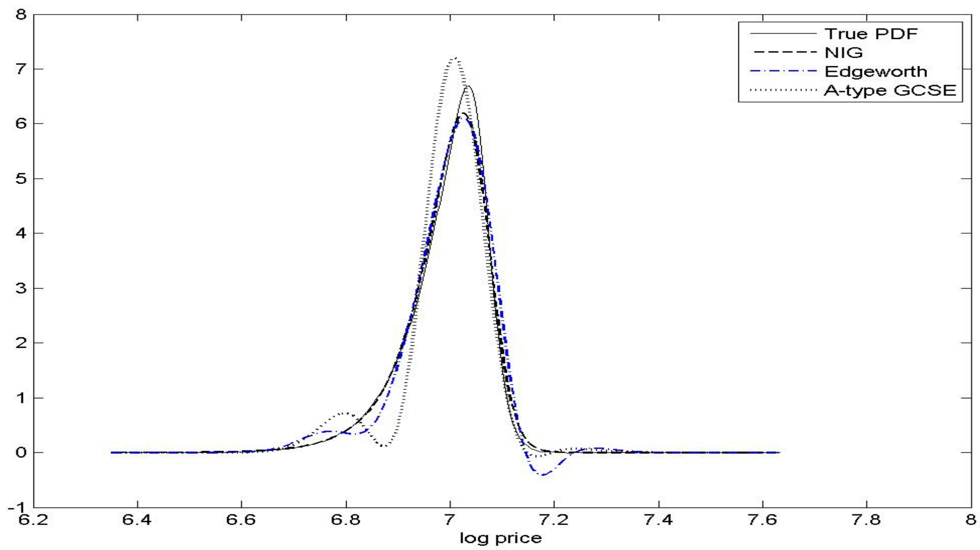


(c) 2003

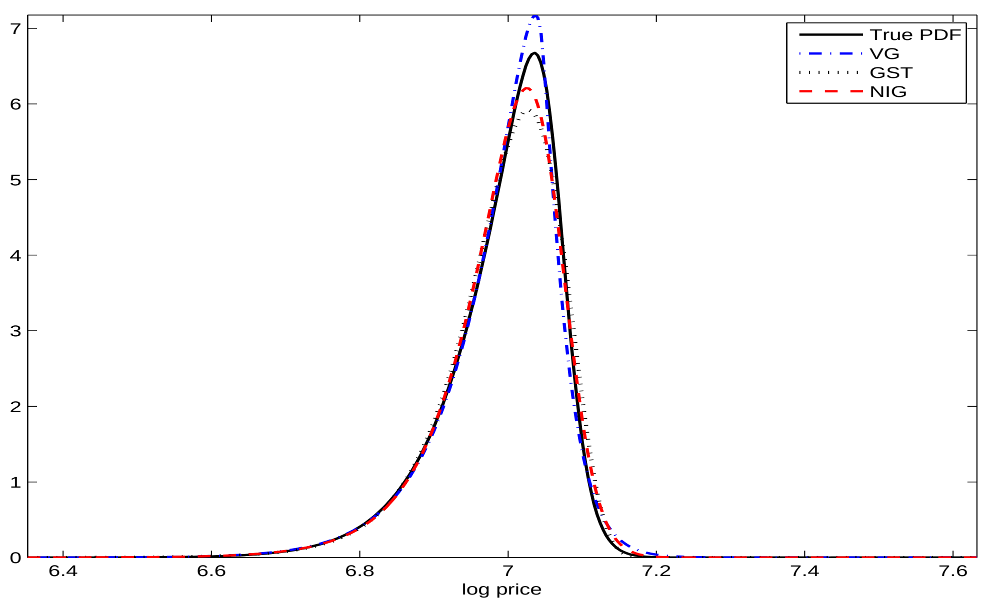
Figure 9: Risk Neutral Density approximations: I

The plots below show the true risk neutral density curve together with its approximations by the NIG, VG, and GST distributions, the Edgeworth expansion and the A-type Gram-Charlier expansion. The underlying asset is generated from model (3.35) with parameters taken from Bakshi et al. (1997) (Parameter I) where there is no jump assumed. The parameter estimation via the method of moments for the NIG, VG and GST laws are:

1. NIG: $(\hat{\alpha}, \hat{\beta}, \hat{\mu}, \hat{b}) = (36.32, -24.09, 7.08, 0.10)$
2. VG: $(\hat{\alpha}, \hat{\beta}, \hat{\mu}, \hat{p}) = (27.78, -11.54, 7.04, 1.45)$
3. GST: $(\hat{\beta}, \hat{\mu}, \hat{b}, \hat{p}) = (-504.49, 7.25, 0.11, -13.32)$



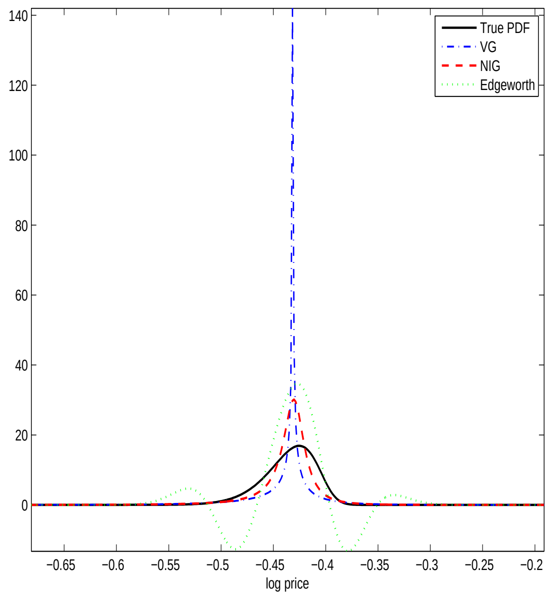
(a)



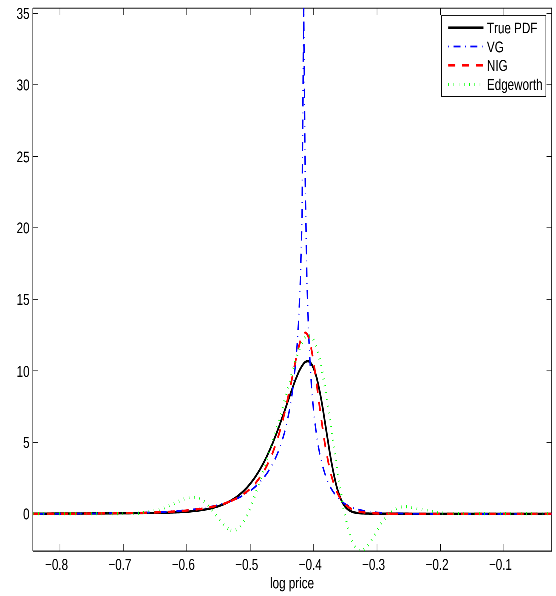
(b)

Figure 10: Risk Neutral Density approximations: III

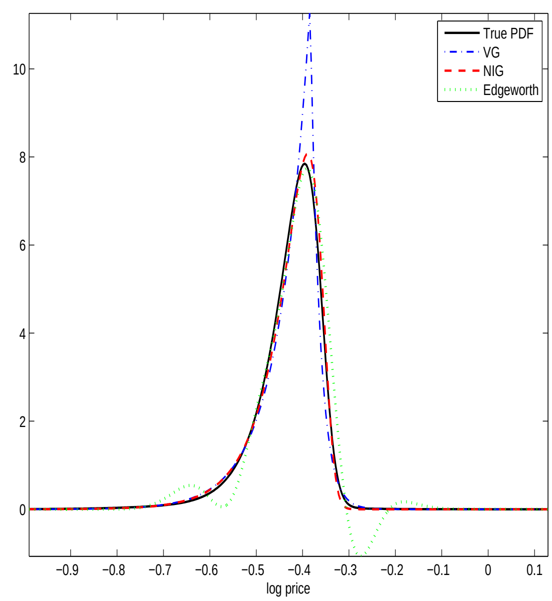
The plots below show the true risk neutral density curves together with their approximations by the NIG, VG distributions and the Edgeworth expansion considering different time to maturities. The underlying asset is generated from model (3.35) with parameters taken from Duffie et al. (2000) (Parameter III) where jump is assumed.



(a) 17 days to maturity



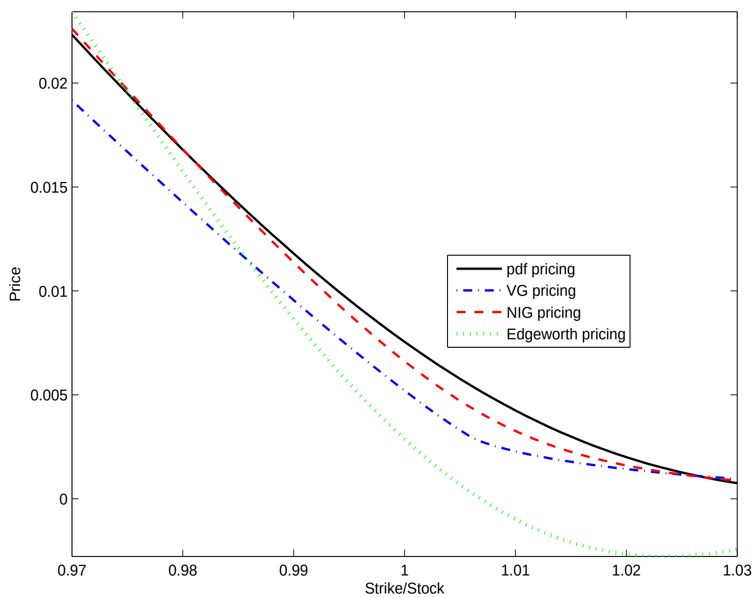
(b) 45 days to maturity



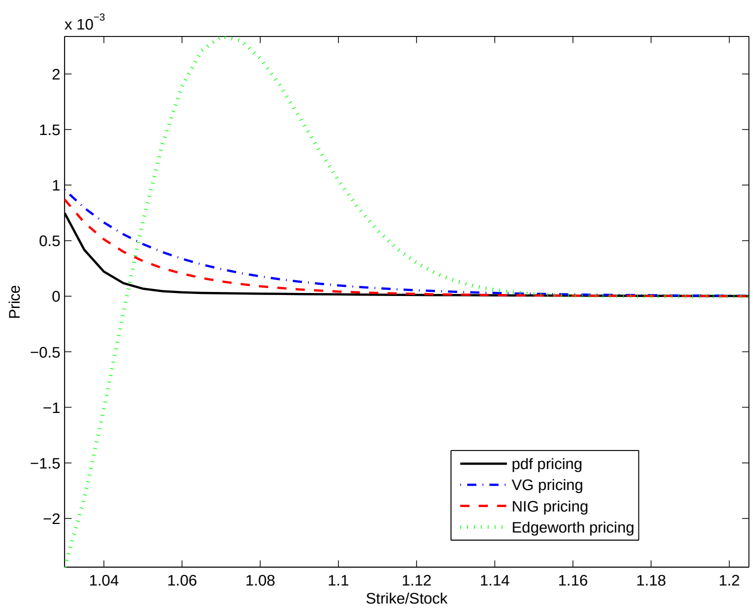
(c) 80 days to maturity

Figure 11: Pricing of European call options: III

The figure displays the pricing of the European calls under the risk neutral measure, and using the NIG law, the VG law and the Edgeworth expansion. The underlying stock prices are generated from option pricing model (3.35) with parameters taken from Duffie et al. (2000) (Parameter III) and the time to maturity is 17 days.



(a) ATM



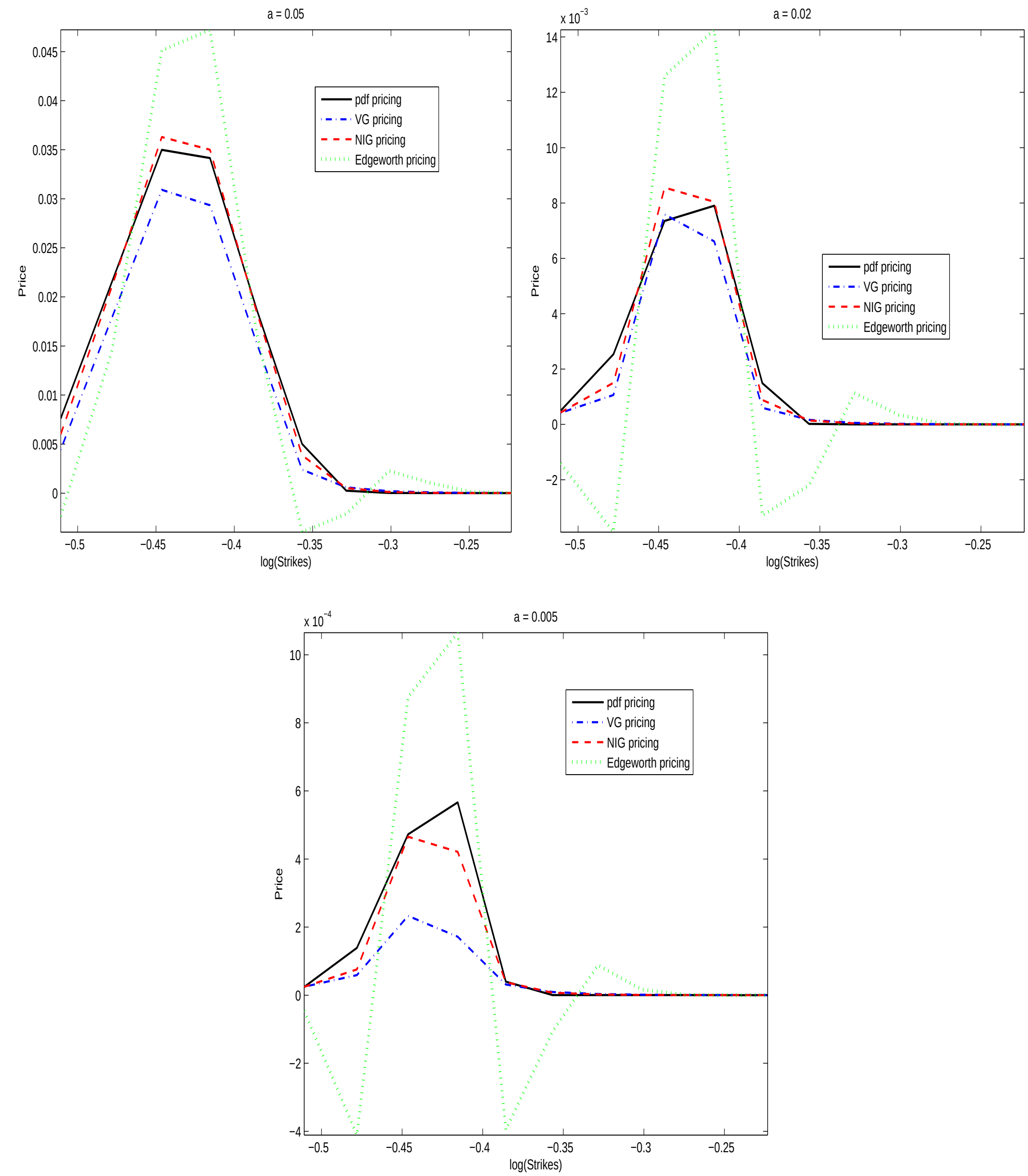
(b) OTM

Figure 12: Pricing of Butterfly trading strategy: III

The figure plots the pricing of balanced Butterfly trading strategy with payoff function

$$g(S_T; K) = (S_T - K + a)1_{(K-a \le S_T \le K)} + (K + a - S_T)1_{(K < S_T \le K+a)}$$

using the true risk neutral density, the NIG law, the VG law, and the Edgeworth expansion. The underlying stock prices are generated from affine model (3.35) with parameters taken from Duffie et al. (2000) (Parameter III). And T is equal to 17 days.



Bibliography

Aas, K. and Haff, I. (2006), "The Generalized Hyperbolic Skew Student's t-Distribution," *Journal of Financial Econometrics*, 4, 275–309.

Abramowitz, M. (1974), *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables*, Dover Publications, Incorporated.

Adrian, T. and Rosenberg, J. (2008), "Stock Returns and Volatility: Pricing the Short-Run and Long-Run Components of Market Risk," Journal of Finance (forthcoming).

Aït-Sahalia, Y. and Lo, A. (1998), "Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices," *Journal of Finance*, 53, 499–547.

Akaike, H. (1974), "Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes," *Annals of the Institute of Statistical Mathematics*, 26, 363–387.

Alizadeh, S., Brandt, M., and Diebold, F. (2002), "Range-Based Estimation of Stochastic Volatility Models," *The Journal of Finance*, 57, 1047–1091.

Andreou, E. and Ghysels, E. (2002), "Detecting multiple breaks in financial market volatility dynamics," *Journal of Applied Econometrics*, 17, 579–600.

Bakshi, G., Cao, C., and Chen, Z. (1997), "Empirical performance of alternative option pricing models," *Journal of Finance*, 2003–2049.

Bakshi, G., Kapadia, N., and Madan, D. (2003), "Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options," *Review of Financial Studies*, 16, 101–143.

Barndorff-Nielsen, O. (1977), "Exponentially decreasing distributions for the logarithm of particle size," *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 401–419.

— (1997), "Normal Inverse Gaussian Distributions and Stochastic Volatility Modelling," *Scandinavian Journal of Statistics*, 24, 1–13.

Barndorff-Nielsen, O. and Shephard, N. (2001), "Normal modified stable processes," .

Barndorff-Nielsen, O. and Stelzer, R. (2005), "Absolute Moments of Generalized Hyperbolic Distributions and Approximate Scaling of Normal Inverse Gaussian Levy Processes," *Scandinavian Journal of Statistics*, 32, 617–637.

Barton, D. and Dennis, K. (1952), "The Conditions Under Which Gram-Charlier and Edgeworth Curves are Positive Definite and Unimodal," *Biometrika*, 39, 425–427.

Bates, D. (1991), "The Crash of'87: Was It Expected? The Evidence from Options Markets," *Journal of Finance*, 46, 1009–1044.

— (1996), "Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options," *Review of Financial Studies*, 9, 69–107.

Berkes, I., Gombay, E., Horváth, L., and Kokoszka, P. (2004), "Sequential changepoint detection in GARCH(p,q) models," *Econometric Theory*, 20, 1140–1167.

Bilingsley, P. (1995), *Probability and Measure*, John Wiley and Sons, Series in Probability and Mathematical Statistics, New York.

Bollerslev, T. (1986), "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

Bougerol, P. and Picard, N. (1992), "Strict Stationarity of Generalized Autoregressive Processes," *The Annals of Probability*, 20, 1714–1730.

Brandt, A. (1986), "The Stochastic Equation $Y_{n+1} = A_n Y_n + B_n$ with Stationary Coefficients," *Advances in Applied Probability*, 18, 211–220.

Breeden, D. and Litzenberger, R. (1978), "Prices of State-Contingent Claims Implicit in Option Prices," *Journal of Business*, 51, 621.

Broadie, M., Detemple, J., Ghysels, E., and Torrès, O. (2000), "American options with stochastic dividends and volatility: A nonparametric investigation," *Journal of Econometrics*, 94, 53–92.

Campa, J., Chang, P., and Reider, R. (1998), "Implied exchange rate distributions: evidence from OTC option markets," *Journal of International Money and Finance*, 17, 117–160.

Carrasco, M. and Chen, X. (2002), "Mixing and moment properties of various GARCH and stochastic volatility models," *Econometric Theory*, 18, 17–39.

Chen, J. and Gupta, A. (1997), "Testing and Locating Variance Changepoints With Application to Stock Prices," *Journal of the American Statistical Association*, 92, 739–747.

Chernov, M. and Ghysels, E. (2000), "A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation," *Journal of Financial Economics*, 56, 407–458.

Chernov, M., Ronald Gallant, A., Ghysels, E., and Tauchen, G. (2003), "Alternative models for stock price dynamics," *Journal of Econometrics*, 116, 225–257.

Conrad, J., Dittmar, R., and Ghysels, E. (2007a), "Skewness and the Bubble," Discussion Paper, University of Michigan and UNC.

— (2007b), "Skewness and the Bubble," Tech. rep., Working Paper, University of North Carolina at Chapell Hill.

Cont, R. and Tankov, P. (2004), *Financial Modelling with Jump Processes*, Chapman & Hall/CRC.

Dahlhaus, R. and Rao, S. (2006), "Statistical inference for time-varying ARCH processes," *Annals of Statistics*, 34, 1075–1114.

Dennis, P. and Mayhew, S. (2002), "Risk-Neutral Skewness: Evidence from Stock Options," *Journal of Financial and Quantitative Analysis*, 37, 471–494.

Ding, Z. and Granger, C. (1996), "Modeling volatility persistence of speculative returns: A new approach," *Journal of Econometrics*, 73, 185–215.

Draper, N. and Tierney, D. (1972), "Regions of positive and unimodal series expansion of the Edgeworth and Gram-Charlier approximations," *Biometrika*, 59, 463.

Duffie, D., Pan, J., and Singleton, K. (2000), "Transform Analysis and Asset Pricing for Affine Jump-diffusions," *Econometrica*, 68, 1343–1376.

Eberlein, E. and Keller, U. (1995), "Hyperbolic distributions in finance," *Bernoulli*, 281–299.

Eberlein, E. and von Hammerstein, E. (2002), "Generalized hyperbolic and inverse Gaussian distributions: limiting cases and approximation of processes," *FDM Preprint*, 80.

Embrechts, P., McNeil, A., and Straumann, D. (2002), "Correlation and dependence in risk management: properties and pitfalls," *Risk management: value at risk and beyond*, 176–223.

Engle, R. (1982), "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.

Engle, R., Ghysels, E., and Sohn, B. (2008), "On the Economic Sources of Stock Market Volatility," University of North Carolina at Chapel Hill, Manuscript.

Engle, R. and Lee, G. (1999), "A Permanent and Transitory Component Model of Stock Return Volatility," *R. Engle and H. White (ed.) Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W. J. Granger, Oxford University Press*, 475–497.

Engle, R. and Rangel, J. (2008), "The Spline-GARCH Model for Low-Frequency Volatility and Its Global Macroeconomic Causes," *Review of Financial Studies*, 21, 1187–1222.

Eriksson, A., Forsberg, L., and Ghysels, E. (2004), "Approximating the Probability Distribution of Functions of Random Variables: A New Approach," Discussion paper CIRANO.

Eriksson, A., Ghysels, E., and Wang, F. (2009), "The Normal Inverse Gaussian Distribution and the Pricing of Derivatives," *Journal of Derivatives.*

Figlewski, S. (2007), "The Implied Risk Neutral Density for the U.S. Market Portfolio," Discussion Paper, NYU Stern.

Figlewski, S. and Gao, B. (1999), "The adaptive mesh model: a new approach to efficient option pricing," *Journal of Financial Economics*, 53, 313–351.

Furstenberg, H. and Kesten, H. (1960), "Products of Random Matrices," *The Annals of Mathematical Statistics*, 31, 457–469.

Ghysels, E., Patilea, V., Renault, É., and Torrès, O. (1997), "Nonparametric Methods and Option Pricing," in *Statistics in Finance*, eds. Hand, D. and Jacka, S., London: Edward Arnold, pp. 261–282.

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2004), *The MIDAS Touch: Mixed Data Sampling Regression Models*, CIRANO.

Glasserman, P. and Yao, D. (1995), "Stochastic Vector Difference Equations with Stationary Coefficients," *Journal of Applied Probability*, 32, 851–866.

He, C. and Teräsvirta, T. (1999), "Properties of moments of a family of GARCH processes," *Journal of Econometrics*, 92, 173–192.

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

Horvath, L., Kokoszka, P., and Teyssière, G. (2001), "Empirical process of the squared residuals of an ARCH sequence," *Annals of Statistics*, 29, 445–469.

Horvath, L., Kokoszka, P., and Zhang, A. (2006), "Monitoring constancy of variance in conditionally heteroskedastic time series," *Econometric Theory*, 22, 373–402.

Inclan, C. and Tiao, G. (1994), "Use of cumulative sums of squares for retrospective detection of changes of variance," *Journal of the American Statistical Association*, 89, 913–923.

Jackwerth, J. and Rubinstein, M. (1996), "Recovering Probability Distributions from Option Prices," *Journal of Finance*, 51, 1611–1631.

Jensen, S. and Rahbek, A. (2004), "Asymptotic Inference For Nonstationary Garch," *Econometric Theory*, 20, 1203–1226.

Jiang, G. and Tian, Y. (2005), "The Model-Free Implied Volatility and Its Information Content," *Review of Financial Studies*, 18, 1305–1342.

Jones, M. and Faddy, M. (2003), "A skew extension of the t-distribution, with applications," *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 159–174.

Jorgensen, B. (1982), *Statistical properties of the generalized inverse Gaussian distribution*, Springer-Verlag.

Kalemanova, A., Schmid, B., and Werner, R. (2007), "The Normal Inverse Gaussian Distribution for Synthetic CDO Pricing," *Journal of Derivatives*, 14, 80–93.

Kingman, J. (1973), "Subadditive Ergodic Theory," *The Annals of Probability*, 1, 883–899.

Kokoszka, P. and Leipus, R. (2000), "Change-point estimation in ARCH models," *Bernoulli*, 6, 513–540.

Kulperger, R. and Yu, H. (2005), "High moment partial sum processes of residuals in GARCH models and their applications," *Annals of Statistics*, 33, 2395–2422.

Longstaff, F. (1995), "Option pricing and the martingale restriction," *Review of Financial Studies*, 8, 1091–1124.

Madan, D., Carr, P., and Chang, E. (1998), "The Variance Gamma Process and Option Pricing," *European Finance Review*, 2, 79–105.

Madan, D. and Milne, F. (1994), "Contingent Claims Valued and Hedged by Pricing and Investing in a basis," *Mathematical Finance*, 4, 223–245.

Madan, D. and Seneta, E. (1990), "The Variance Gamma (VG) Model for Share Market Returns," *Journal of Business*, 63, 511.

McDonald, J. and Newey, W. (1988), "Partially adaptive estimation of regression models via the generalized t distribution," *Econometric Theory*, 4, 428–457.

McLeish, D. (1974), "Dependent central limit theorems and invariance principles," *Ann. Probab*, 2, 620–628.

Melick, W. and Thomas, C. (1997), "Recovering an Asset's Implied PDF from Option Prices: An Application to Crude Oil during the Gulf Crisis," *Journal of Financial and Quantitative Analysis*, 32, 91–115.

Mokkadem, A. (1990), "Propriétés de mélange des processus autorégressifs polynomiaux," *Annales de l'I. H. P. Probabilités et statistiques*, 26, 219–260.

Nelson, D. and Cao, C. (1992), "Inequality constraints in the univariate GARCH model," *Journal of Business and Economic Statistics*, 10, 229–235.

Pagan, A. (1999), *Nonparametric Econometrics*, Cambridge University Press.

Prause, K. (1999), *The generalized hyperbolic model: Estimation, financial derivatives, and risk measures*, Freiburg (Breisgau).

Rompolis, L. and Tzavalis, E. (2005), "The C-type Gram-Charlier Series Expansion of the risk neutral density implied by option prices," MIMEO, University of Accouting and finance, Athens University of Economics and Business.

— (2007), "Retrieving risk neutral densities based on risk neutral moments through a Gram–Charlier series expansion," *Mathematical and Computer Modelling*, 46, 225–234.

Rubinstein, M. (1994), "Implied Binomial Trees," *Journal of Finance*, 49, 771–818.

— (1998), "Edgeworth Binomial Trees," *Journal of Derivatives*, 5, 20–27.

Seshadri, V. (1993), *The Inverse Gaussian Distribution: A Case Study in Exponential Families*, Oxford University Press.

Shenton, L. (1951), "Efficiency of the method of moments and the Gram-Charlier Type A distribution," *Biometrika*, 38, 58–73.

Shimko, D. (1993), "Bounds of probability," *Risk*, 6, 33–37.

Stout, W. (1974), *Almost sure convergence*, New York: Academic Press.

Theodossiou, P. (1998), "Financial data and the skewed generalized T distribution," *Management Science*, 44, 1650–1661.

Vervaat, W. (1979), "On a stochastic difference equation and a representation of non-negative infinitely divisible random variables," *Advances in Applied Probability*, 11, 750–783.

Wang, D. (2001), *Estimation Theory Methods and Practice in Mathematics and Mathematics-Mechanisation*, Shangdong Education Publishing House, Jinan.