



HAL
open science

Applications of the Characteristic Function Based Continuum GMM in Finance

Rachidi Kotchoni

► **To cite this version:**

Rachidi Kotchoni. Applications of the Characteristic Function Based Continuum GMM in Finance. Computational Statistics and Data Analysis, 2012, <http://www.sciencedirect.com/science/article/pii/S0167947310003269>. 10.1016/j.csda.2010.08.011 . hal-00867795

HAL Id: hal-00867795

<https://hal.science/hal-00867795>

Submitted on 30 Sep 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Applications of the Characteristic Function Based Continuum GMM in Finance

Rachidi Kotchoni*
CRÉA - Université Laval, Canada

August, 2010

Abstract

A review of the theoretical properties of the GMM with a continuum of moment conditions is presented. Numerical methods for its implementation are discussed. A simulation study based on the stable distribution and an empirical application based on the autoregressive variance Gamma model are performed. Using the Alcoa price data, the findings suggest that investors require a positive premium for bearing the expected risk while a negative penalty is attached to unexpected risk.

Keywords: Autoregressive Variance Gamma Model, Continuum of Moments Conditions, Simulation, Stable Distribution.

*Correspondance: CRÉA /Pavillon Paul-Comtois /2425, rue de l'Agriculture, Local#4424-F /Université Laval /Québec (QC) G1V 0A6 /Phone: 418 656 2131 ext 3940 /Fax: 418 656 7821 /E-mail: rachidi.kotchoni@fsaa.ulaval.ca.

1 Introduction

For many interesting financial econometric models, the characteristic function (CF) is available in closed form while the likelihood function is not, for example stable distributions and discretely sampled continuous time processes. Exceptionally, a discrete sample from a square root diffusion model admits a closed form conditional likelihood expression. Unfortunately, this expression takes the form of an infinite sum that must be truncated in practice. Certain discrete time models (e.g. the variance gamma model) also have known closed form likelihood functions that are not convenient for numerical optimization. In these situations, the use of the CF for inference is an attractive alternative. In fact, two random variables have the same distribution if and only if their CF coincides on the whole real line. This suggests that an inference method that adequately exploits the information content of the CF has the potential to achieve the same level of efficiency as a likelihood-based approach. One such inference method proposed by Carrasco and Florens (2000) for IID models exploits the whole continuum of moment conditions based on the difference between the empirical and theoretical characteristic functions. Carrasco et al. (2007a) extend the method to deal with Markov and dependent models. Other leading works in this area include Singleton (2001), Knight and Yu (2002), Knight, Satchell and Yu (2002) and Chacko and Viceira (2003). A good review of this literature is provided by Yu (2004).

The goal of this paper is to make the Generalized Method of Moments with a continuum of moments conditions (CGMM) based on the CF accessible to applied researchers. Our focus will be on the approach proposed by Carrasco and Florens (2000) and its extension by Carrasco et al. (2007a). First of all, we review the theory underlying the CGMM. We recall the main assumptions that are useful for the consistency and asymptotic normality of the CGMM estimator. Next, we discuss in detail the important steps of the implementation of the CGMM in practice. Finally, we provide a simulation study with the stable distribution and an empirical application with the autoregressive variance gamma model.

The stable distribution has been introduced in finance to fit the asymmetry and fat tails observed empirically in the distributions of assets returns (Mandelbrot, 1963, or McCulloch, 1986). In its common parameterization, it has a stability parameter $\alpha \in (0, 2]$, a skewness parameter $\beta \in [-1, 1]$, a scale parameter $\sigma > 0$ and a location parameter $\mu \in \mathbb{R}$. The moments of order larger than α do not exist for the stable distribution when $\alpha < 2$. When $\alpha = 2$, all the moments exist but the asymmetry parameter β is no longer identifiable. Closed form expressions for stable densities are available only in a few cases. For example, the case $\alpha = 2$ reduces to a normal distribution $N(\alpha, 2\sigma_0^2)$. When $\alpha = 1$ and $\beta_0 = 0$, we obtain the Cauchy distribution whereas the case $\alpha = 1/2$ and $\beta_0 = 1$ results in to the so-called Levy distribution. An identity established by Zorotalev (1986) and commented in Weron (1996) allows one to deduce the density of the case $\alpha = 1/2$ and $\beta_0 = -1$ from the previous one. But as pointed out by Nolan (2009), the knowledge of the likelihood function at isolated values of the parameter space is not helpful when one is trying to fit the model to real data. This difficulty has often led researchers to rely on numerical approximations of the likelihood of the stable distribution. For example,

McCulloch (1998) discusses an approximate maximum likelihood procedure for symmetric stable distributions while Nolan (1997) proposes alternative numerical procedures for $\alpha > 0.1$. Mittnik et al. (1999) and Paoletta (2007) propose Fast Fourier Transform algorithms to approximate the likelihood function. An alternative quantile based approach is also discussed in McCulloch (1986).

In the current paper, a CGMM procedure that can be used without imposing any restriction on the parameter space is presented. Monte Carlo simulations show that the CGMM outperforms the standard GMM that uses a finite number of moment conditions based on the CF. However, the variance of the estimators cannot be computed analytically when the vector of parameters is close to the non-identification region (that is, when α is close to 2). One then has to rely on Monte Carlo simulations to build confidence intervals. As the primary goal of the current paper is to illustrate the implementation of the CGMM, we leave the simulation comparison of the latter with the likelihood-based approaches for future investigations.

The fact that the asymmetry and fat-tailedness of the stable distribution vanish when its variance exists ($\alpha = 2$) limits its use for the purpose of modeling assets returns. A simple way to circumvent this limitation consists in modeling the variance of the returns as a Gamma variable. This yields the variance gamma models. The symmetric variance gamma model has been proposed by Madan and Seneta (1990). Madan, Carr and Chang (1998) extend the basic model to include asymmetry. These two models unfortunately assume that the variance is IID. Here we relax this assumption by assuming that the variance follows the autoregressive gamma process studied in Gouriéroux and Jasiak (2005). The resulting model for assets returns is termed the “*autoregressive variance gamma model*”. We propose an estimation strategy in two steps. In the first step, we fit the autoregressive gamma model to a consistent estimator of the daily integrated variance used as a proxy for the true daily variance. Next, we estimate a relationship between returns and volatility that allows us to disentangle the risk premium from the leverage effect. An empirical application with the Alcoa stock listed in the Dow Jones Industrials shows that investors require a positive premium for bearing expected risk while a negative premium is attached to unexpected risk.

The rest of the paper is organized as follows. The next section reviews the main theoretical results on the CGMM. In section 3, we discuss the numerical aspects of its implementation. In section 4, we present a simulation study of the performance of the CGMM to estimate the stable distribution. In section 5, we present and estimate the autoregressive variance gamma model both with simulated and real data. Section 6 concludes. Some technical derivations are left in the appendix.

2 The CGMM: a Brief Theoretical Review

In this section, we present the theoretical framework underlying the CGMM estimation. The first subsection reviews the IID framework while the second subsection deals with the dependent case. In the third subsection, we discuss the assumptions needed for the CGMM estimator to have good asymptotic properties.

2.1 The CGMM in the IID Case

Let (x_1, \dots, x_T) be an IID sample of an m -dimensional vector process whose CF is given by $E^{\theta_0} (e^{i\tau'x_t}) = \varphi(\tau, \theta_0)$, where θ_0 is a finite dimensional parameter that fully characterizes the distribution of $\{x_t\}$ and $\tau \in \mathbb{R}^m$ is the Fourier transformation variable. By definition of $\varphi(\tau, \theta_0)$, the following set of moment functions can be considered for the purpose of estimating the parameter θ_0 :

$$h_t(\tau, \theta_0) = e^{i\tau'x_t} - \varphi(\tau, \theta_0), \text{ for all } \tau \in \mathbb{R}^m. \quad (1)$$

Note that these moment functions are indexed by $\tau \in \mathbb{R}^m$, and hence we have a continuum of moment conditions. Since the CF contains the same information as the likelihood function, an efficient use of the whole continuum of moment conditions permits us to achieve the maximum likelihood efficiency (see Carrasco and Florens, 2000).

As in Feuerverger and McDunnough (1981b), Singleton (2001) or Chacko and Viceira (2003), one may choose to estimate θ_0 using GMM based on a discrete subset of the continuum (1). More precisely, let $\{h_t(\tau_k, \theta_0)\}_{k=1}^q$ be a discrete subset of q moments conditions drawn from (1), and define the vector $g_t(\theta_0)$ by:

$$g_t(\theta_0) = (\text{Re } h_t(\tau_1, \theta_0), \dots, \text{Re } h_t(\tau_q, \theta_0), \text{Im } h_t(\tau_1, \theta_0), \dots, \text{Im } h_t(\tau_q, \theta_0))'.$$

The standard GMM estimator of θ_0 is computed as:

$$\hat{\theta}_{\text{GMM}} = \arg \min_{\theta} \hat{g}(\theta_0)' \hat{S}^{-1} \hat{g}(\theta_0),$$

where $\hat{g}(\theta_0) = \frac{1}{T} \sum_{t=1}^T g_t(\theta_0)$, $\hat{S} = \frac{1}{T} \sum_{t=1}^T g_t(\hat{\theta}^1) g_t(\hat{\theta}^1)'$ and $\hat{\theta}^1$ is a first step estimator of θ_0 . See Hansen (1982) for the properties of this estimator.

Feuerverger and McDunnough (1981b) claim that the asymptotic variance of the resulting estimator can be made arbitrarily close to the Cramer-Rao bound by selecting the grid (τ_1, \dots, τ_q) sufficiently refined and extended in \mathbb{R}^m . This confirms that the maximum likelihood efficiency can be achieved only by using the whole continuum of the moment function. However, as one refines and extends the grid of values of τ , the discrete set of moment conditions converges to the continuous moment function $h_t(\tau, \theta) = e^{i\tau'x_t} - \varphi(\tau, \theta)$, $\tau \in \mathbb{R}^m$, while the covariance matrix \widehat{W} converges to the covariance operator associated with that moment function. Moreover, one should note that $2q \leq T$ is a necessary condition for the covariance matrix \widehat{W} to be invertible.

Different methods that continuously match the empirical CF to its theoretical counterpart has been proposed as far back as in Press (1972) and Paulson et al. (1975), but the ideal objective function has been introduced more recently by Carrasco and Florens (2000). That objective function is given by a quadratic form:

$$Q = \left\langle K^{-1/2} \widehat{h}_T(\cdot, \theta), K^{-1/2} \widehat{h}_T(\cdot, \theta) \right\rangle, \quad (2)$$

where $\langle \cdot, \cdot \rangle$ is a scalar product on the Hilbert space of square integrable functions, $\widehat{h}_T(\tau, \theta) = \frac{1}{T} \sum_{t=1}^T h_t(\tau, \theta)$ and K is a linear operator.

To be more precise, let π be an arbitrary finite measure on \mathbb{R}^m and $L^2(\pi)$ denote the Hilbert space of complex valued functions that are square integrable with respect to π , that is:

$$\mathbf{L}^2(\pi) = \{f : \mathbb{R}^m \rightarrow \mathbb{C} \text{ such that } \int f(\tau)\overline{f(\tau)}\pi(\tau)d\tau < \infty\},$$

where \bar{z} is the complex conjugate of z for all $z \in \mathbb{C}$. Interestingly, the CF-based moment function $h_t(\tau, \theta)$ is always bounded in modulus for all random variables, and thus belongs to $L^2(\pi)$ for all $\theta \in \Theta$ and any choice of π . In the objective function (2), the scalar product $\langle \cdot, \cdot \rangle$ may thus be defined as:

$$\langle f, g \rangle = \int f(\tau)\overline{g(\tau)}\pi(\tau)d\tau.$$

Note that π is an arbitrary measure that has nothing to do with the data generating process of x_t . In practice, it is customary to set $\pi(\tau) = \exp(-\tau'\tau)$ in order to be able to compute (2) using Hermitian quadratures.

Carrasco and Florens (2000) show that the maximum likelihood efficiency is achieved when K is the asymptotic covariance operator associated with the moment function $h_t(\tau, \theta_0)$. For any function $f \in \mathbf{L}^2(\pi)$, we have $Kf(\tau) = \int k(s, \tau)f(s)\pi(s)ds$ where:

$$k(s, \tau) = E \left[h_t(s, \theta) \overline{h_t(\tau, \theta)} \right]. \quad (3)$$

The function $k(s, \tau)$ is known as the “kernel of K ” in the literature on linear operators. It can be shown that $Kf \in \mathbf{L}^2(\pi)$ for all $f \in \mathbf{L}^2(\pi)$. Some basic properties of the covariance operator K are given in Appendix A. It is shown in Section 3.1 that Hermitian quadratures can be used to obtain a matrix approximation of K .

In practice, one has to use the empirical counterpart K_T of K . The operator K_T is the one obtained by replacing the kernel $k(s, \tau)$ by a consistent estimator. An estimator of $k(s, \tau)$ is obtained by replacing the expectation operator in (3) by an empirical average:

$$\widehat{k}_T(s, \tau, \widehat{\theta}^1) = \frac{1}{T} \sum_{t=1}^T h_t(s, \widehat{\theta}^1) \overline{h_t(\tau, \widehat{\theta}^1)}, \quad (4)$$

where $\widehat{\theta}^1$ is any consistent first step estimator of θ_0 . An example of a first step estimator is given by:

$$\widehat{\theta}^1 = \arg \min_{\theta} \left\langle \widehat{h}_T(\cdot, \theta), \widehat{h}_T(\cdot, \theta) \right\rangle.$$

In the specific case of IID data, an estimator of the kernel that does not use a first step estimator is given by:

$$\widehat{k}_T(s, \tau) = \frac{1}{T} \sum_{t=1}^T \left(e^{is'x_t} - \widehat{\varphi}_T(s) \right) \overline{\left(e^{i\tau'x_t} - \widehat{\varphi}_T(\tau) \right)}, \quad (5)$$

where $\widehat{\varphi}_T(s) = \frac{1}{T} \sum_{t=1}^T e^{is'x_t}$. Carrasco and Florens (2000) show that both (4) and (5) are consistent for $k(s, \tau)$.

Unfortunately, Carrasco and Florens (2000) point out that K_T is not invertible on the whole $\mathbf{L}^2(\pi)$ space, mainly because the inverse of its theoretical counterpart K exists only on a dense subset of $\mathbf{L}^2(\pi)$. They show that one way to circumvent this problem is to work with the Tikhonov-type generalized inverse:

$$K_{T,\lambda}^{-1} = (K_T^2 + \lambda_T I_T)^{-1} K_T,$$

where I_T is the identity operator and $\lambda_T \in [0, 1]$ is a regularization parameter that is a function of the sample size T . Other types of regularized inverse can also be used (e.g. spectral cut-off, Landweber-Friedman). The Tikhonov scheme is preferred here because it applies a smooth transformation to the spectrum of K_T . The feasible CGMM estimator is given by:

$$\hat{\theta}_T(\lambda_T) = \arg \min \hat{Q}_{T,\lambda}, \quad (6)$$

where $\hat{Q}_{T,\lambda} = \left\langle K_{T,\lambda}^{-1/2} \hat{h}_T(\cdot, \theta), K_{T,\lambda}^{-1/2} \hat{h}_T(\cdot, \theta) \right\rangle$. It is shown in Carrasco and Florens (2000) that the maximum likelihood efficiency is achieved when λ_T converges to zero at a certain rate as the sample size diverges to infinity. The assumptions underlying these results are reviewed in Section 2.3. Throughout this paper, λ without the subscript denotes the regularization parameter viewed as a variable with respect to which we may want to minimize the mean square error (MSE) of $\hat{\theta}_T(\lambda)$ while λ_T denotes the optimal λ for a given sample size T . Finally, the subscript is sometimes removed for notational simplicity, for example when we write $K_{T,\lambda}^{-1}$ instead of K_{T,λ_T}^{-1} .

2.2 The CGMM with Dependent Data

When $\{x_t\}$ is Markov instead of being IID, it may not be possible to identify θ_0 from the marginal CF. In this case, Carrasco et al. (2007a) propose to use the moment function based on the conditional CF:

$$h_t(\tau, \theta) = \left[e^{i\tau'_1 x_{t+1}} - \varphi(\theta, \tau_1, x_t) \right] e^{i\tau'_2 x_t}, \quad (7)$$

where $\varphi(\tau'_1, \theta, x_t) = E^\theta(e^{i\tau'_1 x_{t+1}} | x_t)$, $\tau = (\tau'_1, \tau'_2)' \in \mathbb{R}^{2m}$. In the above expression, the manifold $\{e^{i\tau'_2 x_t}, \tau_2 \in \mathbb{R}^m\}$ is used as an instrument. Carrasco et al. (2007a) show that these instruments are optimal given the Markov assumption.

There also exist many interesting situations where the process $\{x_t\}$ is mixing instead of being Markov or IID. In a typical stochastic volatility model for instance, the joint process of the observed return and the latent variance is Markov but the return process alone is not. In that case, the idea is to use the moment function built from the joint CF:

$$h_t(\tau, \theta) = e^{i\tau' Y_t} - E^\theta(e^{i\tau' Y_t}), \quad \tau \in \mathbb{R}^{mp}, \quad (8)$$

where $Y_t = (x'_t, x'_{t-1}, \dots, x'_{t-p+1})'$. In theory, the larger the p the more efficient the estimator. However in practice, the quest for efficiency must be balanced with the computing cost. For more discussion on this point, see Feuerverger (1990), Carrasco and Florens (2002), Jiang and Knight (2002), Yu (2004) and Carrasco et al. (2007a).

The objective function of the CGMM for Markov and dependent models has the same expression as in (2), except that the kernel of the asymptotic covariance operator K associated with the moments conditions is now given by:

$$k(s, \tau) = E \left[h_t(s, \theta) \overline{h_t(\tau, \theta)} \right] + \sum_{j=1}^{\infty} E \left[h_t(s, \theta) \left(\overline{h_{t-j}(\tau, \theta)} + \overline{h_{t+j}(\tau, \theta)} \right) \right]. \quad (9)$$

Note that the moment function (1) of the IID model is also IID while the moment function (7) of the Markov model is a martingale difference sequence. By definition, a martingale difference sequence is uncorrelated at all lags. Because of this, $k(s, \tau)$ reduces to (3) in the Markov case and can thus be estimated by (4). On the other hand, the moment conditions described by (8) are autocorrelated even if the process $\{x_t\}$ is Markov. In the latter case, $k(s, \tau)$ may be estimated as in Newey and West (1987) or Andrews and Monahan (1992) using the Bartlett kernel:

$$\widehat{k}_T(s, \tau, \widehat{\theta}^1) = \frac{1}{T} \sum_{t=1}^T h_t(s, \widehat{\theta}^1) \overline{h_t(\tau, \widehat{\theta}^1)} + \sum_{j=1}^{J_T} \left(1 - \frac{j-1}{J_T} \right) \sum_{t=1}^T h_t(s, \widehat{\theta}^1) \left(\overline{h_{t-j}(\tau, \widehat{\theta}^1)} + \overline{h_{t+j}(\tau, \widehat{\theta}^1)} \right),$$

where $\widehat{\theta}^1$ is a consistent first step estimator of θ_0 and J_T is a bandwidth that is increasing in T . Again, the operator K_T with kernel $\widehat{k}_T(s, \tau, \widehat{\theta}^1)$ is not invertible on the whole reference space, and the feasible CGMM estimator is defined in the same fashion as in (6). We refer the reader to Carrasco et al. (2007a) for a comprehensive discussion on the CGMM with dependent data.

In the sequel, we shall focus on the IID and Markov cases and use the generic notation $h_t(\tau, \theta)$, $\tau \in \mathbb{R}^d$, where $d = m$ for moment conditions of type (1) and $d = 2m$ for moment conditions of type (7).

2.3 Basic Assumptions of the CGMM

To derive the theoretical properties of the CGMM estimator, we assume the same regularity conditions as in Carrasco and Kotchoni (2010a).

Assumption 1: The measure π involved in the definition of the scalar product $\langle \cdot, \cdot \rangle$ is strictly positive on \mathbb{R}^d and admits all its moments.

Assumption 2: The equation

$$E^{\theta_0} [h_t(\tau, \theta)] = 0 \text{ for all } \tau \in \mathbf{R}^d,$$

holds π almost everywhere, where E^{θ_0} denotes the expectation with respect to the distribution of x_t for $\theta = \theta_0$, and has a unique solution θ_0 which is an interior point of a compact set Θ .

Assumption 3: The moment function $h_t(\tau, \theta)$ is three times continuously differentiable with respect to θ .

Assumption 4: For all θ , $E^{\theta_0} [h_T(\cdot, \theta)]$ and its first three derivatives with respect to θ belong to the range of K^β for some $\beta \geq 1/2$.

Assumption 5: The random variable x_t is stationary and satisfies $x_t = x(\theta_0, \varepsilon_t, Z_{t-1})$ where $x(\cdot, \varepsilon_t, Z_{t-1})$ is three times continuously differentiable with respect to θ_0 , ε_t is IID white noise whose distribution is known and does not depend on θ_0 , and Z_{t-1} can only contain lagged values of x_t .

Assumption 1 ensures that the norm associated with the scalar product $\langle \cdot, \cdot \rangle$ is well defined while Assumption 2 is a global identification requirement. The CGMM estimator is still well defined if Assumption 3 is weaker, for example if $h_t(\tau, \theta)$ is only once continuously differentiable, but the derivation of some of the asymptotic properties of the estimator becomes difficult. Assumption 4 ensures that the limit of the objective function as T goes to infinity is well defined. The real number β in this assumption is the level of regularity of $E^{\theta_0} [h_T(\cdot, \theta)]$ with respect to the operator K , that is, the largest real number such that $\|K^{-\beta} E^{\theta_0} [h_T(\cdot, \theta)]\| < \infty$.

Under assumptions 1 and 2, the estimator of the covariance operator satisfies in the IID and Markov case:

$$\|K_T - K\| = O_p(T^{-1/2}),$$

where the notation $X_T = O_p(T^{-b})$ means that the random sequence $T^b X_T$ have a non-degenerate limiting distribution. The regularized inverse $K_{T,\lambda}^{-1/2}$ has the property that for any function f in the range of $K^{1/2}$, $K_{T,\lambda}^{-1/2} f$ converges to $K^{-1/2} f$ as T goes to infinity and λ_T goes to zero at some rate. In the IID and Markov case, assumptions 1 to 4 ensure that the CGMM estimator satisfies:

$$T^{1/2} \left(\widehat{\theta}_T(\lambda_T) - \theta_0 \right) \rightarrow N(0, I_{\theta_0}^{-1}),$$

as T and $\lambda_T^2 T$ go to infinity and λ_T goes to zero, where $I_{\theta_0}^{-1}$ denotes the inverse of the Fisher Information Matrix.

Assumption 5 is not crucial for the good properties of the CGMM. It has been used in Carrasco and Kotchoni (2010a) to derive the properties of the optimal sequence of regularization parameters λ_T . A similar assumption is also used in Gourieroux, Monfort and Renault (1993) to derive the properties of indirect inference estimators.

3 The CGMM in Practice

In this section, we discuss two numerical methods to evaluate the objective function of the CGMM. The first method is based on Gauss-Hermite quadratures while the second uses Monte Carlo integration. We discuss how to compute the variance of the CGMM estimator and review the simulation based selection of the regularization parameter λ_T .

3.1 Computing the Objective Function by Quadrature Method

In implementing the CGMM procedure, the main challenge is the accurate computation of the multiple integrals embedded in its objective function:

$$\widehat{Q}_{T,\lambda} = \int_{\mathbb{R}^d} K_{T,\lambda}^{-1/2} \widehat{h}_T(\tau, \theta) \overline{K_{T,\lambda}^{-1/2} \widehat{h}_T(\tau, \theta)} \pi(\tau) d\tau.$$

To start with, let us consider in the univariate case ($d = 1$) an arbitrary function $f(\tau, \theta)$ that is continuously differentiable up to order $2n$. Then $f(\tau, \theta)$ can be approximated well by a polynomial function of τ , that is:

$$f(\tau, \theta) = \sum_{k=0}^{2n-1} a_k(\theta) \tau^k + \varepsilon(\tau, \theta),$$

where the residual $\varepsilon(\tau, \theta)$ is negligible for large n . In that case, the weighting function $\pi(s) = \exp(-s^2)$ is quite convenient to work with. As shown by Carrasco and Florens (2000), the choice of the weighting function is irrelevant for the asymptotic properties of the CGMM estimator. However, the function $\pi(s) = \exp(-s^2)$ has the nice feature that it puts little weight on extreme values of $h_t(s, \theta)$, which is a desirable feature in a finite sample. More importantly, this choice of $\pi(s)$ allows one to approximate the objective function using Gauss-Hermite quadratures.

Let us first consider the computation of $K_T \widehat{h}_T(\tau, \theta)$. We have:

$$K_T \widehat{h}_T(\tau, \theta) = \int \widehat{k}_T(\tau, s) \widehat{h}_T(s, \theta) \exp(-s^2) ds. \quad (10)$$

The Gauss-Hermite quadrature amounts to finding n points (s_1, \dots, s_n) and weights $(\omega_1, \dots, \omega_n)$ such that:

$$\int P(s) \exp\{-s^2\} ds = \sum_{k=1}^n \omega_k P(s_k),$$

for any polynomial function $P(\cdot)$ of order smaller or equal to $2n - 1$. See e.g. Liu and Pierce (1994). For the function $f(\tau, \theta)$, we have:

$$\left| \int f(s) \exp\{-s^2\} ds - \sum_{k=1}^n \omega_k f(s_k) \right| = \left| \int \varepsilon(\tau, \theta) \exp(-s^2) ds \right|.$$

If $K_T \widehat{h}_T(\tau, \theta)$ is analytic as a function of τ , the residual $\left| \int \varepsilon(\tau, \theta) \exp(-s^2) ds \right|$ can be made arbitrarily small by increasing the number of quadrature points n . Note that the choice of the quadrature points and weights does not depend on the particular function $f(\tau, \theta)$. The quadrature points and weights are determined by solving:

$$\int s^l \exp\{-s^2\} ds = \sum_{k=1}^n \omega_k s_k^l \text{ for all } l = 1, \dots, 2n - 1.$$

Applying this quadrature method to (10) yields:

$$K_T \widehat{h}_T(\theta) \approx \widehat{W}_T \widehat{h}_T(\theta), \quad (11)$$

where \widehat{W}_T is the matrix with (j, k) elements $W_{jk} = \omega_k \widehat{k}_T(s_j, s_k)$, and:

$$\widehat{h}_T(\theta) = \left(\widehat{h}_T(s_1, \theta), \dots, \widehat{h}_T(s_n, \theta) \right)'$$

For any given level of precision, the matrix \widehat{W}_T can be looked at as the best finite dimensional reduction of the operator K_T . The resulting approximation of $K_{T,\lambda}^{-1/2}$ is:

$$K_{T,\lambda}^{-1/2} \approx \left(\widehat{W}_T^2 + \lambda_T I \right)^{-1/2} \widehat{W}_T^{1/2},$$

that is:

$$K_{T,\lambda}^{-1/2} \widehat{h}_T(\theta) \approx \left(\widehat{W}_T^2 + \lambda_T I \right)^{-1/2} \widehat{W}_T^{1/2} \widehat{h}_T(\theta). \quad (12)$$

Substituting for $K_{T,\lambda}^{-1/2} \widehat{h}_T(\theta)$ in the objective function of the CGMM yields:

$$\widehat{Q}_{T,\lambda} \approx \sum_{k=1}^n \omega_k \left[K_{T,\lambda}^{-1/2} \widehat{h}_T(s_k, \theta) \right] \left[\overline{K_{T,\lambda}^{-1/2} \widehat{h}_T(s_k, \theta)} \right], \quad (13)$$

where $K_{T,\lambda}^{-1/2} \widehat{h}_T(s_k, \theta)$ is the k^{th} element of the vector $K_{T,\lambda}^{-1/2} \widehat{h}_T(\theta)$ given in (12). It is important to note that $K_{T,\lambda}^{-1/2}$ is estimated using a first step estimator $\widehat{\theta}^1$ whose presence in the formulas is hidden for simplicity.

In theory, the extension of the above quadrature method to the multivariate case is straightforward. When $\tau \in \mathbf{R}^d$, the d -dimensional set of multivariate quadrature points is given by the Cartesian product:

$$D = \left\{ \tau = (\tau_{(1)}, \dots, \tau_{(d)}) : \tau_{(i)} \in \{s_1, \dots, s_n\} \text{ for all } i = 1 \text{ to } d \right\},$$

where $\{s_1, \dots, s_n\}$ is the set of n univariate quadrature points with weights $\{\omega_1, \dots, \omega_n\}$, and $\tau_{(i)}$ is the i^{th} coordinate of τ . Associated with each $\tau \in D$ is the weight:

$$\bar{\omega}(\tau) = \omega(\tau_1) \omega(\tau_2) \dots \omega(\tau_d),$$

where $\omega(\tau_i) = \omega_k$ if $\tau_{(i)} = s_k$, $i = 1, \dots, d$.

The multivariate Gauss-Hermite quadrature has the undesirable feature that $Card(D) = n^d$. This raises a ‘‘curse of dimensionality’’ because the size of the matrix \widehat{W}_T is precisely n^d while we need to take n quite large ($n \approx 10$) to accurately evaluate the objective function of the CGMM. Because \widehat{W}_T must be inverted at each iteration of the optimization algorithm, the CGMM becomes virtually infeasible by the quadrature method when $d \geq 3$. We shall thus limit ourselves to the case $d \leq 2$ in the sequel, leaving the discussion on the large d case for further research.

3.2 Computing the Objective Function by Monte Carlo Integration

This approach relies on the alternative formula of the CGMM objective function provided in Carrasco et al. (2007a):

$$\widehat{Q}_{T,\lambda} = v(\theta)' \left[\lambda_T I_T + \widehat{C}_T^2 \right]^{-1} \overline{v(\theta)}, \quad (14)$$

where $\widehat{C}_T \equiv \widehat{C}_T(\widehat{\theta}^1)$ is the square matrix of size T with (t, l) element $c_{t,l}/(T - \dim(\theta))$, I_T is the identity matrix of size T , and $v(\theta) = (v_1, \dots, v_T)'$ with:

$$\begin{aligned} v_t &= \int \overline{h_t(\tau, \widehat{\theta}^1)} h_T(\tau, \theta) \pi(\tau) d\tau \text{ and} \\ c_{t,l} &= \int \overline{h_t(\tau, \widehat{\theta}^1)} h_l(\tau, \widehat{\theta}^1) \pi(\tau) d\tau. \end{aligned}$$

The main drawback of the above expressions lies in that it involves the inverse of the matrix \widehat{C}_T which has size T . However, this should be balanced by at least one computational advantage: the integrals embedded in v_t and $c_{t,l}$ can be approximated by Monte Carlo. If we set $\pi(\tau)$ to be the multivariate standard normal density and $(\tau^{(1)}, \dots, \tau^{(M)})$ be M values of τ simulated according to $\pi(\tau)$, the Monte Carlo approximations of v_t and $c_{t,l}$ are:

$$\begin{aligned} \widetilde{v}_t &\approx \frac{1}{M} \sum_{k=1}^M \overline{h_t(\tau^{(k)}, \widehat{\theta}^1)} h_T(\tau^{(k)}, \theta) \text{ and} \\ \widetilde{c}_{t,l} &\approx \frac{1}{M} \sum_{k=1}^M \overline{h_t(\tau^{(k)}, \widehat{\theta}^1)} h_l(\tau^{(k)}, \widehat{\theta}^1). \end{aligned}$$

For numerical efficiency, it is recommended to simulate the set $(\tau^{(1)}, \dots, \tau^{(M)})$ only once at the beginning of the estimation process and supply this as a fixed array to the code that evaluates the objective function of the CGMM.

A comparison between the performance of the Monte Carlo integration and that of the Hermitian quadrature is not a trivial task. In fact, this requires us to define how one balances the computing cost against the statistical efficiency. For the models considered in this paper, the quadrature method is preferred because it requires inverting a matrix of smaller size. This feature is favored because we want to be able to replicate the estimation procedure several times at a moderate cost.

3.3 Computing the Variance of the CGMM Estimator

The asymptotic variance of the optimal CGMM estimator is derived in Carrasco and Florens (2000):

$$\begin{aligned} AVar(\widehat{\theta}) &= Var \left[\sqrt{T} (\widehat{\theta} - \theta_0) \right] \\ &= \left\langle K^{-1/2} E \left(\widehat{G}_t(\cdot, \theta) \right), K^{-1/2} E \left(\widehat{G}_t(\cdot, \theta) \right) \right\rangle^{-1}, \end{aligned} \quad (15)$$

where $\widehat{G}_t(\tau, \theta) = \frac{\partial \widehat{h}_t(\tau, \theta)}{\partial \theta}$ is a column vector of length q whose i^{th} element is $\widehat{G}_{t,i}(\tau, \theta) = \frac{\partial \widehat{h}_t(\tau, \theta)}{\partial \theta_i}$, and for every two vector functions f and g , we have: $\langle f, g \rangle_{i,j} = \langle f_i, g_j \rangle$. The asymptotic variance (15) is consistently estimated by:

$$\widehat{AVar}(\widehat{\theta}) = \left\langle K_{T,\lambda}^{-1/2} \widehat{G}_T(\tau, \widehat{\theta}), K_{T,\lambda}^{-1/2} \widehat{G}_T(\tau, \widehat{\theta}) \right\rangle^{-1}, \quad (16)$$

where $\widehat{G}_T(\tau, \widehat{\theta}) = \frac{1}{T} \sum_{t=1}^T \widehat{G}_t(\tau, \widehat{\theta})$. The above formula is convenient to work with when the scalar products are evaluated by quadrature methods. Define:

$$\begin{aligned} \widehat{G}_{T,i}(\theta) &= \left(\widehat{G}_{T,i}(\tau_1, \theta), \dots, \widehat{G}_{T,i}(\tau_N, \theta) \right)' \text{ and} \\ K_{\lambda T}^{-1/2} \widehat{G}_{T,i}(\theta) &= \left(K_{T,\lambda}^{-1/2} \widehat{G}_{T,i}(\tau_1, \theta), \dots, K_{T,\lambda}^{-1/2} \widehat{G}_{T,i}(\tau_N, \theta) \right)', \end{aligned}$$

where $N = n^d$ and $\widehat{G}_{T,i}(\tau, \widehat{\theta}) = \frac{1}{T} \sum_{t=1}^T \widehat{G}_{t,i}(\tau, \widehat{\theta})$. Then we have:

$$K_{T,\lambda}^{-1/2} \widehat{G}_{T,i}(\theta) = \left(\widehat{W}_T^2 + \lambda_T I \right)^{-1/2} \widehat{W}_T^{1/2} \widehat{G}_{T,i}(\theta),$$

where \widehat{W} is defined in (11). The (i, j) element of $\widehat{AVar}(\widehat{\theta})^{-1}$ can then be computed as:

$$\left(\widehat{AVar}(\widehat{\theta})^{-1} \right)_{i,j} = \sum_{k=1}^N \omega_k \left(K_{T,\lambda}^{-1/2} \widehat{G}_{T,i}(\theta) \right)_k \left(K_{T,\lambda}^{-1/2} \widehat{G}_{T,j}(\theta) \right)_k,$$

where $\left(K_{T,\lambda}^{-1/2} \widehat{G}_{T,i}(\theta) \right)_k$ is the k^{th} coordinate of $K_{T,\lambda}^{-1/2} \widehat{G}_{T,i}(\theta)$.

Carrasco et al. (2007a) establish the following alternative expression for $\widehat{AVar}(\widehat{\theta})$:

$$\widehat{AVar}(\widehat{\theta}) = \left(\frac{1}{T - \dim(\theta)} V(\widehat{\theta})' \left[\lambda_T I_T + \widehat{C}_T^2 \right]^{-1} \overline{V(\widehat{\theta})} \right)^{-1}, \quad (17)$$

where \widehat{C} is the same square matrix as in (14), $V(\widehat{\theta})$ is the (T, q) matrix with (t, i) element:

$$V_{t,i} = \int \overline{h_t(\tau, \widehat{\theta})} \widehat{G}_{T,i}(\tau, \widehat{\theta}) \pi(\tau) d\tau,$$

Formula (17) is best suited when the Monte Carlo integration is used to evaluate the scalar products. In this case, $V_{t,i}$ is approximated by:

$$\widetilde{V}_{t,i} \approx \frac{1}{M} \sum_{k=1}^M \overline{h_t(\tau^{(k)}, \widehat{\theta})} \widehat{G}_{T,i}(\tau^{(k)}, \widehat{\theta}),$$

where $(\tau^{(1)}, \dots, \tau^{(M)})$ are M values of τ simulated according to the multivariate normal density $\pi(\tau)$.

For the optimization algorithm to converge, it is crucial to simulate the set $(\tau^{(1)}, \dots, \tau^{(M)})$ only once at the beginning of the estimation process and supply this as a fixed array to the code that evaluates the objective function of the CGMM.

3.4 Data-driven Selection of the Regularization Parameter

The CGMM estimator is consistent for any reasonable choice of the regularization parameter λ_T . In most applications, an arbitrary choice of λ_T between 10^{-6} and 10^{-2} works quite well. However, if the spectrum of the empirical covariance operator is severely discontinuous, such an arbitrary choice is not advised. To get close to the optimal CGMM in the MSE sense, Carrasco and Kotchoni (2010a) propose two simulation based methods to select the λ_T . The first method uses the higher-order closed form approximation of the MSE whereas the second method relies on the Monte Carlo simulations of the MSE. We briefly review the second method here.

Let $\lambda_T(\theta_0)$ be the optimal value of the regularization parameter when θ_0 is the true parameter of interest and T is the sample size. We define $\lambda_T(\theta_0)$ as:

$$\lambda_T(\theta_0) = \arg \min_{\lambda \in [0,1]} E \left[\left\| \widehat{\theta}_T(\lambda; \theta_0) - \theta_0 \right\|^2 \right], \quad (18)$$

where we note that the expected squared 2-norm $E \left[\left\| \widehat{\theta}_T(\lambda; \theta_0) - \theta_0 \right\|^2 \right]$ is the trace of the MSE matrix of $\widehat{\theta}_T(\lambda; \theta_0)$, and $\widehat{\theta}_T(\lambda; \theta_0)$ is the CGMM estimator computed from an arbitrary sample of size T generated from the true distribution, and using λ as the regularization parameter. To approximate this MSE, assume that we can draw samples of size T from the true data generating process of $\{x_t\}$, and let $\widehat{\theta}_T^j(\lambda; \theta_0)$ denote the CGMM estimator of θ_0 computed using the j^{th} independently simulated sample. A good estimator of $E \left[\left\| \widehat{\theta}_T(\lambda; \theta_0) - \theta_0 \right\|^2 \right]$ is given by:

$$\frac{1}{M} \sum_{j=1}^M \left\| \widehat{\theta}_T^j(\lambda; \theta_0) - \theta_0 \right\|^2.$$

For a sufficiently large value of M , the Law of Large Numbers ensures that this criterion converges to its expectation.

If θ_0 were feasible, (18) would suggest an estimator of $\lambda_T(\theta_0)$ of the form:

$$\widehat{\lambda}_{TM}(\theta_0) = \arg \min_{\lambda \in [0,1]} \frac{1}{M} \sum_{j=1}^M \left\| \widehat{\theta}_T^j(\lambda; \theta_0) - \theta_0 \right\|^2. \quad (19)$$

Since θ_0 is not known, a feasible Monte Carlo approach simply consists of replacing θ_0 with a consistent first step estimator $\widehat{\theta}^1$ in (19), that is, choosing the optimal regularization parameter according to:

$$\widehat{\lambda}_{TM}(\widehat{\theta}^1) = \arg \min_{\lambda \in [0,1]} \frac{1}{M} \sum_{j=1}^M \left\| \widehat{\theta}_T^j(\lambda; \widehat{\theta}^1) - \widehat{\theta}^1 \right\|^2. \quad (20)$$

Carrasco and Kotchoni (2010a) establish under Assumption 1 to 5 that as M and T go to infinity, we have:

$$\frac{\widehat{\lambda}_{TM}(\widehat{\theta}^1)}{\lambda_T(\theta_0)} - 1 = O_p(T^{-1/2}) + O_p(M^{-1/2}).$$

Finally, the optimal feasible CGMM estimator is $\widehat{\theta}_T^{(2)} = \widehat{\theta}_T(\widehat{\lambda}_{TM}^*; \theta_0)$, that is, the second step estimator of θ_0 computed with the actual data by using the point estimate of the optimal regularization parameter $\widehat{\lambda}_{TM}^* = \widehat{\lambda}_{TM}(\widehat{\theta}^1)$. In practice, the MSE must be simulated using common random numbers to ensure the comparability of the results across different values of the regularization parameter.

In the sequel, we propose two illustrative implementations of the CGMM.

4 Estimating the Stable Distribution by CGMM: a Simulation Study

In econometrics, the stable distribution is a way to depart from the usual normality assumption in case the latter seems too restrictive. This family is rich enough to capture heavy tails as well as asymmetry, as pointed out by Mandelbrot (1963) or McCulloch (1986). However, the stable distribution does not admit a closed form likelihood function. This has led researchers to investigate alternative inference methods. CF based inference has been used in Paulson, Holcomb and Leitch (1975) and Feuerverger and McDunnough (1981a), while a regression-based approach is presented in Koutrouvelis (1980). Garcia, Renault and Veredas (2006) have resorted to indirect inference. Cornea and Davidson (2009) proposed a refined bootstrap method for hypothesis testing.

The stable distribution has been represented under different parameterizations in the literature. The most widely advocated parameterization is presented in the next subsection.

4.1 Parameterization and Simulation of the Stable Distribution

The standard stable distribution has two parameters: a stability parameter $\alpha \in (0, 2]$, and a skewness parameter $\beta \in [-1, 1]$. A random variable Z is said to follow the standard stable distribution if and only if its CF is given by:

$$E[\exp(i\tau Z)] = \exp\{-|\tau|^\alpha [1 + i\beta_0 \text{sign}(\tau) g(\tau, \alpha)]\},$$

where $g(\tau, \alpha) = -\tan \frac{\alpha\pi}{2}$ if $\alpha \neq 1$ and $g(\tau, \alpha) = \frac{2}{\pi} \ln |\tau|$ if $\alpha = 1$. A random variable X follows an α -stable distribution if and only if it is linked to the standard variable Z by:

$$X = \begin{cases} \sigma Z + \mu_0, & \alpha \neq 1 \\ \sigma Z + \mu_0 + \frac{2}{\pi}\beta_0\sigma \ln \sigma, & \alpha = 1 \end{cases}.$$

The CF of X is given by:

$$E[\exp(i\tau X)] = \exp\{i\mu_0\tau - \sigma_0^\alpha |\tau|^\alpha [1 + i\beta_0 \text{sign}(\tau) g(\tau, \alpha)]\}, \quad (21)$$

where μ_0 is a location parameter and σ_0 is the scale parameter. The notation $X \sim S_\alpha(\beta_0, \sigma_0, \mu_0)$ is often used to mean that the random variable X has a α -stable distribution with CF (21).

A method to simulate from $S_\alpha(\beta_0, \sigma_0, \mu_0)$ is explained in Weron (1996). To start with, one draws two independent uniforms v and w in $[0, 1]$ and calculates $V = \pi(u - 1/2)$ and $W = -\ln w$. Then $Z \sim S_\alpha^0(\beta_0, 1, 0)$ and $X \sim S_\alpha(\beta_0, \sigma_0, \mu_0)$ are obtained as follows:

If $\alpha \neq 1$, one computes:

$$Z = S_{\alpha, \beta_0} \frac{\sin(\alpha V + \alpha B_{\alpha, \beta_0})}{(\cos V)^{1/\alpha}} \left(\frac{\cos((1 - \alpha)V - \alpha B_{\alpha, \beta_0})}{W} \right)^{-1+1/\alpha}, \quad (22)$$

where $B_{\alpha, \beta_0} = \frac{\arctan(\beta_0 \tan \frac{\alpha\pi}{2})}{\alpha}$ and $S_{\alpha, \beta_0} = (1 + \beta_0^2 \tan^2 \frac{\alpha\pi}{2})^{\frac{1}{2\alpha}}$. Then we have

$$X = \sigma_0 Z + \mu_0 \sim S_\alpha(\beta_0, \sigma_0, \mu_0).$$

If $\alpha = 1$, one computes instead:

$$Z = \frac{2}{\pi} \left[\left(\frac{\pi}{2} + \beta_0 V \right) \tan V - \beta \log \left(\frac{W \cos V}{\frac{\pi}{2} + \beta_0 V} \right) \right], \quad (23)$$

which means:

$$X = \sigma_0 Z + \mu_0 + \frac{2}{\pi} \beta_0 \sigma_0 \ln \sigma_0 \sim S_1(\beta_0, \sigma_0, \mu_0).$$

As our inferences are based on Monte Carlo simulation, we review other existing parameterizations and their simulation strategies in Appendix B.

4.2 Monte Carlo Comparison of GMM and CGMM

With a method to simulate data from the stable distribution in hand, we can now evaluate by Monte Carlo the ability of the CGMM to identify the true parameters from a finite sample. The implementation of the CGMM is quite involved, compared to the standard GMM which is straightforward to implement and easy to understand. Hence to convince the finance practitioner to use the CGMM rather than the GMM with a finite number of moment conditions, we need to provide evidence showing that the former significantly outperforms the latter. To this end, we consider a stable $AR(1)$ model specified as:

$$y_t = \rho_0 + \rho_1 y_{t-1} + \varepsilon_t, \quad (24)$$

where $\varepsilon_t \sim S_\alpha(\beta_0, \sigma_0, 0)$ is IID. Note that this amount to say that $y_t \sim S_\alpha(\beta_0, \sigma_0, \mu_t)$ with $\mu_t = \rho_0 + \rho_1 y_{t-1}$. The parameters of the model are gathered in $\theta = (\rho_0, \rho_1, \alpha, \beta_0, \sigma_0)'$.

To estimate θ , the following continuum of moment conditions is considered:

$$h_t(\tau, \theta) = [e^{i\tau_1 y_t} - \varphi_t(\tau_1, \theta)] e^{i\tau_2 y_{t-1}}, \quad (25)$$

where $\tau = (\tau_1, \tau_2) \in \mathbb{R}^2$ and:

$$\varphi_t(\tau_1, \theta) = \exp \left\{ i(\rho_0 + \rho_1 y_{t-1})\tau_1 - \sigma^\alpha |\tau_1|^\alpha \left[1 - i\beta \text{sign}(\tau_1) \tan \frac{\alpha\pi}{2} \right] \right\}. \quad (26)$$

Indeed, we impose in advance that $\alpha > 1$, which forces the first moment of ε_t to be finite.

We consider the following two vectors of true parameters in our simulations: $\theta_{01} = (0, 0.1, 1.5, 0, 0.5)$ and $\theta_{02} = (0, 0.1, 1.95, 0, 0.5)$, the specificity of θ_{02} being that it is quite close to the non-identification region of β_0 . To ease the numerical optimizations, the following transformations are imposed on the parameter space:

$$\begin{aligned}\alpha_0 &= 1 + \frac{\exp(\tilde{\alpha}_0)}{1 + \exp(\tilde{\alpha}_0)} \in]1, 2] \text{ for all } \tilde{\alpha}_0 \in \mathbb{R}, \\ \beta_0 &= \frac{2 \exp(\tilde{\beta}_0)}{1 + \exp(\tilde{\beta}_0)} - 1 \in [-1, 1] \text{ for all } \tilde{\beta}_0 \in \mathbb{R} \text{ and} \\ \sigma_0 &= \exp(\tilde{\sigma}_0) > 0 \text{ for all } \tilde{\sigma}_0 \in \mathbb{R}.\end{aligned}$$

After these transformations, the new objective function of the CGMM is written in terms of the unconstrained parameters $\tilde{\alpha}$, $\tilde{\beta}$, $\tilde{\sigma}_0$ and μ_0 .

The Monte Carlo experiments are conducted in two steps for the CGMM. First, we run a small scale simulation (100 replications) for the purpose of estimating the optimal λ as in (20). In this small scale simulation, we compute the objective function with $N = 64$ Hermitian quadrature points in \mathbb{R}^2 . The first simulated sample is used to compute the following first step CGMM estimator:

$$\hat{\theta}_T^1 = \arg \min_{\theta} \left\| \hat{h}_T(\cdot, \theta) \right\|^2. \quad (27)$$

For each λ_k and each simulated sample, we compute the second step CGMM estimator as:

$$\hat{\theta}_T^{(j)}(\lambda_k) = \arg \min_{\theta} \left\| K_{T, \lambda_k}^{-1/2} \hat{h}_T(\cdot, \theta) \right\|^2,$$

where:

$$K_{T, \lambda_k}^{-1/2} \left(\hat{\theta}_T^1 \right) = \left(K_T^2 \left(\hat{\theta}_T^1 \right) + \lambda_k I \right)^{-1/2} K_T^{1/2} \left(\hat{\theta}_T^1 \right),$$

for $\lambda_k \in \{10^{-7}, 5 \times 10^{-7}, \dots, 5 \times 10^{-4}\}$. The selection of λ_k is based on the criterion:

$$\lambda_T = \arg \min_{\lambda_k} \frac{1}{M} \sum \left\| \hat{\theta}_T^{(j)}(\lambda_k) - \hat{\theta}_T^1 \right\|^2.$$

Panel 1 shows the plot of the empirical MSE as a function of λ_k . On the grid that we consider, the MSE is minimized at $\lambda_T = 5 \times 10^{-7}$.

[Fig. 1 about here]

Fig. 1. MSE of $\hat{\theta}_T(\lambda_k)$ of as function of λ_k .

In the second step, the selected λ_T is used in a larger scale simulation to assess the performance of the CGMM. We draw $M = 1000$ samples of size $T = 500$ and estimate $\theta_{0i}, i = 1, 2$. To speed up the simulations, we reduce the number of quadrature points to $N = 36$.

To implement the standard GMM, we also consider 36 moment conditions computed on an evenly spaced grid of $\tau = (\tau_1, \tau_2)$ lying within the range $(-3, 3) \times (-3, 3) \subset \mathbb{R}^2$. We avoid the decomposition of the moment conditions between their real and imaginary part by using the same scalar product as for the CGMM. More precisely, we define the vector $g_t(\theta_0)$ by:

$$g_t(\theta_0) = (h_t(\tau_1, \theta_0), \dots, h_t(\tau_{36}, \theta_0))',$$

and compute the GMM estimator of θ_0 as:

$$\hat{\theta}_{\text{GMM}} = \arg \min_{\theta} \hat{g}_T(\theta_0)' \hat{S}_T^{-1}(\hat{\theta}^{(1)}) \overline{\hat{g}_T(\theta_0)},$$

where

$$\begin{aligned} \hat{g}_T(\theta_0) &= \frac{1}{T} \sum_{t=1}^T g_t(\theta_0), \\ \hat{S}_T(\hat{\theta}^{(1)}) &= \frac{1}{T} \sum_{t=1}^T g_t(\hat{\theta}^{(1)}) \overline{g_t(\hat{\theta}^{(1)})'}, \end{aligned}$$

and $\hat{\theta}^{(1)}$ is the first step GMM estimator of θ_0 given by:

$$\hat{\theta}^{(1)} = \arg \min_{\theta} \hat{g}_T(\theta_0)' \overline{\hat{g}_T(\theta_0)}.$$

Finally, the variance of the GMM estimator is estimated by:

$$\text{Var}(\hat{\theta}_{\text{GMM}}) = \left(\frac{\partial \hat{g}_T(\hat{\theta}_{\text{GMM}})}{\partial \theta} \hat{S}_T^{-1}(\hat{\theta}_{\text{GMM}}) \overline{\frac{\partial \hat{g}_T(\hat{\theta}_{\text{GMM}})}{\partial \theta}} \right)^{-1}. \quad (28)$$

In principle, one can consider a more refined grid on an extended range but this quickly results in a badly scaled or singular covariance matrix \hat{S}_T . Also, note that the 36 Hermitian quadrature points used for the CGMM fall within the considered range. The following table shows some statistical properties of $\hat{\theta}_T(\lambda_T)$ when the true parameter is $\theta_{01} = (0, 0.1, 1.5, 0, 0.5)$.

		ρ_0	ρ_1	α_0	β_0	σ_0
True Values		0	0.1	1.5	0	0.5
CGMM	Mean Bias	0.0044	-0.0032	0.0072	0.0075	-0.0033
	Median Bias	0.0007	-0.0026	0.0066	0.0078	-0.0038
	Emp. Std. Dev.	0.0869	0.0482	0.1033	0.2133	0.0305
	Ana. Std. Dev.	0.1137	0.0496	0.1027	0.2597	0.0243
	Root-MSE	0.0870	0.0483	0.1035	0.2134	0.0306
	IC1(95%)	-0.0032	0.0925	1.4982	-0.0112	0.4940
	IC2(95%)	0.0120	0.1010	1.5163	0.0262	0.4993
GMM	Mean Bias	0.0025	-0.0048	0.1200	0.0263	-0.0221
	Median Bias	0.0027	-0.0077	0.1236	0.0281	-0.0226
	Emp. Std. Dev.	0.0612	0.0415	0.0988	0.2179	0.0252
	Ana. Std. Dev.	0.0525	0.0246	0.0693	0.2144	0.0210
	Root-MSE	0.0613	0.0418	0.1555	0.2195	0.0335
	IC1(95%)	-0.0029	0.0916	1.6114	0.0072	0.4757
	IC2(95%)	0.0078	0.0988	1.6287	0.0454	0.4801

Table 1: Simulation Results for the Stable Distribution with $\alpha = 1.5$ (far from 2).

$M = 1000$ replications, $T = 500$.

In this table, “Emp. Std. Dev” is the standard deviation of the simulated empirical distribution of $\hat{\theta}_T(\lambda_T)$, while “Ana. Std. Dev.” is the average standard deviation computed according to the analytical formulas (17) and (28). The expressions of the gradients involved in this formula are given in Appendix C. Interestingly, the standard deviations computed in these two ways are quite close. IC1(95%) and IC2(95%) are respectively the lower and upper bound of the 95% confidence interval for the true mean of the empirical distribution, assuming normality for the empirical mean of the estimates:

$$\begin{aligned}
 IC1(95\%) &= \bar{\hat{\theta}}_i - 1.96 * \hat{s}_{\hat{\theta}_i} / \sqrt{M} \text{ and} \\
 IC2(95\%) &= \bar{\hat{\theta}}_i + 1.96 * \hat{s}_{\hat{\theta}_i} / \sqrt{M},
 \end{aligned}$$

where $\hat{\theta}_i$ is the i^{th} element of $\hat{\theta}$, $\bar{\hat{\theta}}_i$ and $\hat{s}_{\hat{\theta}_i}$ are respectively the empirical mean and standard deviation of $\hat{\theta}_i$ and M is the number of independently simulated copies of $\hat{\theta}_i$. Indeed, the graphs of Panel 2 suggest that the distributions of the estimators are close to normality when the true vector of parameters is θ_{01} .

By looking at Table 1, we see that the GMM displays larger estimation biases but smaller variances compared to the CGMM. The bias is quite large for the GMM estimator of α_0 while being non-negligible for the corresponding estimators of β_0 and σ_0 . Accordingly, the MSEs of the GMM estimators are smaller only for the parameters that govern the linear part of the stable AR(1) model, that is, ρ_0 and ρ_1 . The superiority of the CGMM over the GMM clearly shows up when one compares the 95% confidence intervals. Indeed, the confidence regions drawn from the CGMM estimates contain the

true values for all the parameters except for σ_0 which is very slightly biased downward. By contrast, only the confidence interval of ρ_0 encompasses the true value for the GMM.

[Fig. 2 about here]

Fig. 2. Estimation of Stable Distribution by GMM and CGMM. Empirical distributions of the estimators for 1000 Monte Carlo replications when the true parameter is $\theta_{01} = (0, 0.1, 1.5, 0, 0.5)$.

The following table displays the Monte Carlo statistics when the true parameter is $\theta_{02} = (0, 0.1, 1.95, 0, 0.5)$.

		ρ_0	ρ_1	α	β_0	σ_0
True Values		0	0.1	1.95	0	0.5
CGMM	Mean Bias	0.0020	-0.0043	0.0081	-0.0516	-0.0011
	Median Bias	0.0023	-0.0020	0.0366	-0.0146	-0.0019
	Emp. Std. Dev.	0.0383	0.0583	0.0536	0.5752	0.0211
	Root-MSE	0.0383	0.0584	0.0542	0.5772	0.0211
	IC1(95%)	-0.0014	0.0906	1.9534	-0.1019	0.4971
	IC2(95%)	0.0054	0.1008	1.9628	-0.0012	0.5008
GMM	Mean Bias	-0.0110	-0.0284	0.0421	0.2452	-0.0399
	Median Bias	-0.0108	-0.0275	0.0500	0.9357	-0.0396
	Emp. Std. Dev.	0.0364	0.0375	0.0132	0.8900	0.0186
	Root-MSE	0.0380	0.0470	0.0441	0.9232	0.0440
	IC1(95%)	-0.0142	0.0683	1.9910	0.1673	0.4585
	IC2(95%)	-0.0078	0.0749	1.9933	0.3232	0.4617

Table 2: Simulation Results for the Stable Distribution with $\alpha = 1.95$ (close to 2). $M = 1000$ replications, $T = 500$.

Table 2 shows that the GMM estimators are more biased and have smaller variances compared to the corresponding CGMM estimators for all but one parameter. As a result, the 95% confidence intervals associated with the GMM estimates are misleading. For both procedures, the estimator of β_0 is highly volatile due to the fact that the objective function is unable to identify this parameter when α_0 is close to 2. In that region of the parameter space, the empirical gradient is so badly conditioned that it is not possible to compute the variance analytically. In conducting inferences in empirical applications, this problem is circumvented by resorting to Monte Carlo simulations of resampling methods. It is worth mentioning that compared to the GMM, the CGMM estimator of β is less affected by this weak identification of β . The CGMM procedure can be made more accurate by increasing the number of quadrature points used to approximate its objective function. Unfortunately, increasing the number of moment conditions in the GMM raises numerical

problems that are not addressed by the standard theory. This allows to conclude that overall, the CGMM outperforms the standard GMM.

Panels 3 show the graphs of the simulated empirical distributions of the estimators. The distribution of $\widehat{\alpha}_0$ is highly skewed to the left due to the fact that α_0 is close to the frontier of the parameter space. Also, the distribution of $\widehat{\beta}_0$ is diffuse as a result of β_0 being close to the non identification region.

[Fig. 3 about here]

Fig. 3. Estimation of Stable Distribution by GMM and CGMM. Empirical distributions of the estimators for 1000 Monte Carlo replications when the true parameter is

$$\theta_{02} = (0, 0.1, 1.95, 0, 0.5).$$

5 Fitting the Autoregressive Variance Gamma Model to Assets Returns

The basic variance gamma model has been proposed by Madan and Seneta (1990). A random variable r_t is said to follow a symmetric variance gamma distribution if:

$$r_t|V_t \sim N(\delta, \sigma^2 V_t), \text{ with } V_t \stackrel{IID}{\sim} \text{Gamma}(1/\gamma, 1/\gamma).$$

The density of V_t is given by:

$$f_V(v) = \frac{v^{1/\gamma-1}}{\gamma^{1/\gamma} \Gamma(1/\gamma)} \exp(-v/\gamma),$$

where $\Gamma(1/\gamma) = \int_0^\infty u^{1/\gamma-1} e^{-u} du$. It can be easily verified that $E(V_t) = 1$.

Unlike the stable distribution, all the conditional and unconditional moments of r_t exist. It can be verified that $E[r_t] = \delta$ and $E[(r_t - \delta)^2] = \sigma^2$. The kurtosis of r_t is given by:

$$\frac{E[(r_t - \delta)^4]}{E[(r_t - \delta)^2]^2} = 3(1 + \gamma),$$

which shows that the distribution of r_t is more fat-tailed than the normal distribution whenever $\gamma > 0$. To introduce skewness into this basic set up, Madan, Carr and Chang (1998) express the mean of r_t as a linear function of V_t :

$$r_t|V_t \sim N(\delta_0 + \delta_1 V_t, \sigma^2 V_t), \text{ with } V_t \sim \text{Gamma}(1/\gamma, 1/\gamma),$$

where $\gamma > 0$. If r_t is a series of returns, the parameter δ_1 captures the so-called leverage effect while δ_0 measures the risk premium. Note that when $\delta_1 = -\delta_0$, the leverage effect offsets the risk premium so that the conditional mean of r_t is zero, but the skewness is nonzero unless $\delta_1 = 0$.

Many studies have diagnosed patterns like persistence and clustering in the time series properties of the volatility of assets returns. Unfortunately, the basic Variance Gamma model assumes that V_t follows an IID process. In an effort to correctly measure the volatility, Engle (1982) and Bollerslev (1986) introduces respectively the ARCH and GARCH models that usually have good filtering properties. Not surprisingly, these models are among the most popularly used to capture time-varying volatility. Multivariate ARCH tests statistics are proposed for example in Duchesne (2006) while Bauwens et al. (2010) use multivariate GARCH models to design a value-at-risk based intradaily dynamic asset allocation.

In the stochastic volatility literature, the variance is often specified as a latent state variable that determines the distribution of the returns. For example, Jacquier et al. (1994) postulate the following model:

$$\begin{aligned} r_t &= \sqrt{V_t} \varepsilon_t \text{ and} \\ \log V_t &= a + b \log V_{t-1} + u_t, \end{aligned}$$

where ε_t and u_t are uncorrelated and $r_t|V_t \sim N(0, V_t)$. This model may be viewed as a discrete time version of Hull and White (1987). It has been extended in Jacquier et al. (2004) to allow for correlation between ε_t and u_t . Other famous examples in continuous time include Stein and Stein (1991) and Heston (1993). A class of nonlinear stochastic volatility models is proposed in Yu et al. (2006).

In the next subsection, we extend the basic variance Gamma model to account for time dependence in the volatility process.

5.1 The Autoregressive Variance Gamma Model

The Autoregressive Variance Gamma Model (ARVG) is a stochastic volatility model in which the return process r_t is a function of its expected variance $E[V_t|V_{t-1}]$ and the variance innovation $V_t - E[V_t|V_{t-1}]$:

$$r_t = \mu_0 + \mu_1 \sqrt{E[V_t|V_{t-1}]} + \delta (V_t - E[V_t|V_{t-1}]) + \sqrt{V_t} \varepsilon_t, \quad (29)$$

where $\varepsilon_t \stackrel{IID}{\sim} N(0, 1)$ is uncorrelated with past, current and future realizations of V_t , $\mu_1 \geq 0$ and $\delta \leq 0$. In turn, V_t follows an Autoregressive Gamma process with conditional density:

$$\begin{aligned} f(V_t | \{V_{t-k}\}_{k=1}^{\infty}) &= f(V_t | V_{t-1}) \\ &= \sum_{j=0}^{\infty} \frac{V_t^{j+q-1} c^{j+q}}{\Gamma(j+q)} \exp(-cV_t) p_j(V_{t-1}), \end{aligned} \quad (30)$$

with $(\kappa, \beta, \sigma) > 0$, $c = \frac{2\kappa}{\sigma^2(1-e^{-\kappa})}$, $q = \frac{2\kappa\beta}{\sigma^2}$ and $p_j(V_{t-1})$ is a Poisson weight given by:

$$p_j(V_{t-1}) = \frac{(ce^{-\kappa}V_{t-1})^j}{j!} \exp(-ce^{-\kappa}V_{t-1}).$$

The term $\mu_1 \sqrt{E[V_t|V_{t-1}]}$ in the expression of the return aims to capture the premium that investors require for bearing the expected risk while $\delta(V_t - E[V_t|V_{t-1}])$ is a penalty attached to the unexpected risk. Indeed, this model is specified in spirit of French et al. (1987) who were the first to find in a famous empirical study that the risk premium on a financial asset is positively related to the expected volatility while the unexpected return is negatively related to the unexpected volatility.

The postulated distribution for V_t is also known as the non-centered Chi-square. When $2\kappa\beta > \sigma^2$, the process V_t can be viewed as discrete observations of the following Square-Root diffusion (see Feller, 1951):

$$dV_s = \kappa(\beta - V_s) ds + \sigma\sqrt{V_s}dW_s$$

Its discretized version has been used in Gouriéroux and Jasiak (2005) to model for inter-trade durations. The conditional CF of V_t is given by:

$$E[e^{i\tau V_t}|V_{t-1}] = \left(1 - \frac{i\tau}{c}\right)^{-q} \exp\left(\frac{i\tau e^{-\kappa} V_{t-1}}{1 - \frac{i\tau}{c}}\right). \quad (31)$$

By looking at the expression above, we see that the autoregressive Gamma family nests the univariate Wishart autoregressive process of Gouriéroux et al. (2009). Carrasco et al. (2007a) use the above conditional CF in a simulation study to estimate the parameters of the Square-Root diffusion. Their results show that the CGMM compares favorably to Maximum Likelihood Estimation, Quasi-Maximum Likelihood Estimation and Efficient Method of Moments. From (30), we note that the likelihood function needs to be truncated for this model while its evaluation is time consuming.

The expressions of the conditional expectation and variance of V_t are the following:

$$\begin{aligned} E[V_t|V_{t-1}] &= \beta(1 - e^{-\kappa}) + e^{-\kappa}V_{t-1} \text{ and} \\ Var[V_t|V_{t-1}] &= \frac{1}{c} [\beta(1 - e^{-\kappa}) + 2e^{-\kappa}V_{t-1}]. \end{aligned}$$

To assess the potential of the ARVG model to capture asymmetry and fat tails in the distribution of stock returns, we examine below the third and fourth conditional moments of r_t . We have:

$$E[(r_t - E[r_t|V_{t-1}])^3|V_{t-1}] = \delta^3 E[(V_t - E[V_t|V_{t-1}])^3|V_{t-1}].$$

As V_t is positively skewed like any Gamma distribution, r_t has a time varying negative skewness whenever $\delta < 0$. It is difficult to tell whether r_t is fat-tailed in general. However, it can be shown that when $\delta = 0$ the conditional kurtosis of r_t is:

$$\frac{E[(r_t - E[r_t|V_{t-1}])^4|V_{t-1}]}{Var[r_t|V_{t-1}]^2} = 3 + \frac{3Var[V_t|V_{t-1}]}{Var[r_t|V_{t-1}]^2}.$$

We present a method to simulate the ARVG model below.

5.2 A Preliminary Simulation Study

In this subsection, we run some preliminary Monte Carlo simulations that will help us assess the scope of our empirical results. Assuming that V_t is observed, the ARVG model can be estimated in two steps. In the first step, one estimates an Autoregressive Gamma model for V_t by CGMM based on the moment function:

$$h_t(\tau, \theta_1) = (\exp(i\tau_1 V_t) - E[\exp(i\tau_1 V_t) | V_{t-1}]) \exp(i\tau_2 V_{t-1}), \quad (32)$$

where $E[\exp(i\tau_1 V_t) | V_{t-1}]$ is given by (31), $\tau = (\tau_1, \tau_2)$ and $\theta_1 = (\kappa, \beta, \sigma^2)$. The specific details of the implementation of this first step are explained in the empirical section. In the second step, the remaining set of parameters $\theta_2 = (\mu_0, \mu_1, \delta)$ is estimated by Gaussian maximum likelihood based on the distribution of ε_t conditional on V_{t-1} as postulated in (29). We have:

$$\widehat{\varepsilon}_t = V_t^{-1/2} \left[r_t - \mu_0 - \mu_1 \sqrt{\widehat{E}[V_t | V_{t-1}]} - \delta \left(V_t - \widehat{E}[V_t | V_{t-1}] \right) \right] \sim N(0, 1), \quad (33)$$

where $\widehat{E}[V_t | V_{t-1}] = \widehat{\beta} (1 - e^{-\widehat{\kappa}}) + e^{-\widehat{\kappa}} V_{t-1}$.

In all the simulations, we set the true values of the parameters to:

$$\kappa = 0.02; \beta = 0.05; \sigma^2 = 2.25 \times 10^{-4}; \mu_0 = 0; \mu_1 = 0.05; \delta = -0.05$$

Our method to simulate V_t is inferred from the Poisson-Mixing-Gamma representation of its density given in (30) (see Devroye, 1986). The simulation algorithm is initialized to the unconditional mean $V_0 = \beta$ or by drawing V_0 from the stationary Gamma distribution with density given by:

$$f(v) = \frac{v^{q-1}}{\Gamma(q)} \left(\frac{2\kappa}{\sigma^2} \right)^q \exp\left(\frac{-2\kappa}{\sigma^2} v \right).$$

At $t = 1$, one draws an integer j_0 from the Poisson distribution with parameter $ce^{-\kappa} V_0$. The current realization V_1 of the autoregressive Gamma process is then drawn from the Gamma distribution with density $f_{j_0}(v)$ given by:

$$f_{j_0}(v) = \frac{v^{j_0+q-1} c^{j_0+q}}{\Gamma(j_0+q)} \exp(-cv).$$

At $t = 2$, one draws again an integer j_1 from the Poisson distribution with parameter $ce^{-\kappa} V_1$. The new realization V_2 of the autoregressive Gamma process is now drawn from the Gamma distribution with density $f_{j_1}(v)$, and so forth. At an arbitrary step t , the realization V_t is drawn from the Gamma distribution with density $f_{j_{t-1}}(v)$, where j_{t-1} is a draw from the Poisson distribution with parameter $ce^{-\kappa} V_{t-1}$.

Let (V_0, V_1, \dots, V_T) be the simulated path for the variance process. As r_t depends on two consecutive realizations of V_t , its simulation starts at $t = 1$. We generate a sample of size T of the return process using the equation:

$$r_t = \mu_0 + \mu_1 \sqrt{\beta(1 - e^{-\kappa}) + e^{-\kappa} V_{t-1}} + \delta [V_t - \beta(1 - e^{-\kappa}) - e^{-\kappa} V_{t-1}] + \sqrt{V_t} \varepsilon_t,$$

for $t = 1, \dots, T$, where ε_t is an IID draw from the standard normal distribution.

Using the method just described, we simulate $M = 1000$ samples of size $T = 500$ and estimate the ARVG model. Table 3 below shows the summary statistics of the Monte Carlo simulations. In the estimation process, we constrained $\hat{\mu}_1$ to be positive and $\hat{\delta}$ to be negative.

	$\hat{\kappa}$	$\hat{\theta}_1$	$\hat{\sigma}^2$	$\hat{\mu}_0$	$\hat{\theta}_2$	$\hat{\delta}$
True	0.02	0.05	2.25×10^{-4}	0	0.05	-0.05
Mean	0.0286	0.0504	2.22×10^{-4}	-0.0216	0.1488	-1.0851
Median	0.0259	0.0496	2.23×10^{-4}	-0.0043	0.0393	-0.0000
Mode	0.0253	0.0473	2.25×10^{-4}	-0.0019	0.0565	-0.5323
Std. Dev.	0.0132	0.0079	1.69×10^{-5}	0.0433	0.1981	1.7315
Root-MSE	0.0157	0.0079	1.72×10^{-5}	0.0483	0.2213	2.0173
IC1(95)	0.0092	0.0375	1.88×10^{-4}	-0.1296	0.0000	-5.9421
IC2(95)	0.0578	0.0685	2.54×10^{-4}	0.0264	0.6447	-0.0000

Table 3: Monte Carlo statistics ($M = 1000$ replications, $T = 500$).

The first result is that $\hat{\theta}_2$ is relatively more volatile than $\hat{\theta}_1$. In fact, the Student-t statistics would predict that μ_1 and δ are not significantly different from zero although in reality they are. However, part of the variance of $\hat{\theta}_2$ may be due to the fact that it is estimated conditional on $\hat{\theta}_1$. Second, the mean of $\hat{\theta}_1$ is a good predictor of its true counterpart. The same cannot be said for $\hat{\theta}_2$ whose median is much closer to the truth than the mean. The graphs of Panel 4 indicate that the modes of the distributions of $\hat{\mu}_0$, $\hat{\mu}_1$ and $\hat{\delta}$ are even closer to the truth than their medians. Unfortunately, the mode of the distribution of $\hat{\delta}$ is still ten times larger in magnitude than the true δ . Finally, the 95% confidence regions contain the true values of the parameters although the intervals are quite wide and uninformative for μ_0 , μ_1 and δ .

Overall, the results are encouraging for the portion of the ARVG model that has been estimated by CGMM. In the sequel, we present an estimation strategy that relies on the availability of high frequency data.

[Panel 4 about here]

Fig. 4. Estimation of the ARVG model. Empirical distributions of the estimators for 1000 Monte Carlo replications and true parameter $(\kappa, \beta, \sigma^2) = (0.02, 0.05, 2.25 \times 10^{-4})$ and $(\mu_0, \mu_1, \delta) = (0, 0.05, -0.05)$

5.3 Estimating the ARVG Model from High Frequency Data

Assuming that the ARVG model describes the observed dynamic of daily returns, this section explains why and how one can construct a proxy for V_t using high frequency data.

Let us consider an arbitrary asset whose instantaneous log-price p_s follows a Brownian diffusion with drift:

$$dp_s = m(s, \sigma_s) ds + \sigma_s dW_s,$$

where W_s is a standard Brownian motion possibly correlated with σ_s . It is further assumed that σ_s itself follows a positive diffusion. If we normalize a trading day to be one period, then daily returns are given by:

$$r_t \equiv p_t - p_{t-1} = \int_{t-1}^t m(s, \sigma_s) ds + \int_{t-1}^t \sigma_s dW_s, \quad t = 0, 1, \dots$$

It is seen that r_t satisfies:

$$\begin{aligned} E \left[r_t | \{\sigma_s\}_{s=0}^T \right] &= \int_{t-1}^t m(s, \sigma_s) ds + E \left[\int_{t-1}^t \sigma_s dW_s | \{\sigma_s\}_{s=0}^T \right] \text{ and} \\ \text{Var} \left[r_t | \{\sigma_s\}_{s=0}^T \right] &= IV_t, \end{aligned}$$

where $IV_t = \int_{t-1}^t \sigma_s^2 ds$ is the integrated variance and $\{\sigma_s\}_{s=0}^T$ is the information set that contains the whole continuous path of the spot volatility.

Interestingly, the ARVG model also satisfies:

$$\begin{aligned} E[r_t | V_t, V_{t-1}] &= \mu_0 + \mu_1 \sqrt{E[V_t | V_{t-1}]} + \delta (V_t - E[V_t | V_{t-1}]) \text{ and} \\ \text{Var}[r_t | V_t, V_{t-1}] &= V_t. \end{aligned}$$

The last equation emphasizes that V_t is the conditional variance of the daily return r_t and is thus related to the daily integrated variance in any continuous time framework. Our strategy to estimate the ARVG model from high frequency data consists in assuming the following intuitive matching:

$$E \left[r_t | \{\sigma_s\}_{s=0}^T \right] \equiv E[r_t | V_t, V_{t-1}] \text{ and} \quad (34)$$

$$\text{Var} \left[r_t | \{\sigma_s\}_{s=0}^T \right] \equiv \text{Var}[r_t | V_t, V_{t-1}]. \quad (35)$$

Equation (34) matches the conditional expectations while (35) matches the conditional variances. The last equation implies that $V_t \equiv \int_{t-1}^t \sigma_s^2 ds$ so that the conditioning information set of the right-hand side is narrower than for the left-hand side. Our approximation should thus be viewed as a linear projection of the left-hand side onto (IV_t, IV_{t-1}) .

To construct a proxy for IV_t , let us assume that in each trading day we observe $m+1$ equidistant prices. These prices can be used to compute exactly m high frequency returns $r_{t,1}, r_{t,2}, \dots, r_{t,m}$, that is:

$$r_{t,j} = p_{t-1+j/m} - p_{t-1+(j-1)/m}.$$

Jacod (1994), Jacod and Protter (1998) and Barndorff-Nielsen and Shephard (2002) show that for large m , the realized variance $RV_t^{(m)} = \sum_{j=1}^m r_{t,j}^2$ is a fairly good proxy for IV_t . In practice however, the observed prices are contaminated with market microstructure noise

which causes the naive realized variance to be a biased estimator of IV_t . The following estimator proposed by Barndorff-Nielsen et al. (2008) is known to be consistent for IV_t even in the presence of microstructure noise:

$$K_{H,t}^{BNHLS} = \gamma_{t,0} + \sum_{h=1}^H \left(1 - \frac{h-1}{H}\right) (\gamma_{t,h} + \gamma_{t,-h}),$$

where $\gamma_{t,h} = \sum_{j=1}^m r_{t,j} r_{t,j-h}$. To further reduce the variance of $K_{H,t}^{BNHLS}$, we will use the following shrinkage estimator proposed in Carrasco and Kotchoni (2010b):

$$K_{H,t}^{\varpi} = \varpi K_{H,t}^{BNHLS} + (1 - \varpi) \widehat{IV}_t,$$

where

$$\widehat{IV}_t = \gamma_{t,0} + \gamma_{t,1} + \gamma_{t,-1} + \frac{1}{T} \sum_{s=1}^T \sum_{l=2}^{L+1} (\gamma_{s,l} + \gamma_{s,-l}).$$

The shrinkage weight ϖ is chosen so as to minimize the marginal variance of $K_{H,t}^{\varpi}$:

$$\varpi_t^* = \arg \min_{\varpi} E \left[(K_{H,t}^{\varpi} - IV_t)^2 \right].$$

It is easy to show that $\varpi_t^* = \frac{Cov[\widehat{IV}_t, \widehat{IV}_t - K_{H,t}^{BNHLS}]}{Var[\widehat{IV}_t - K_{H,t}^{BNHLS}]}$, which we estimate in the simplest possible way from the data by:

$$\widehat{\varpi}^* = \frac{\sum_{t=1}^T \left(\widehat{IV}_t - K_{H,t}^{BNHLS} \right) \widehat{IV}_t}{\sum_{t=1}^T \left(\widehat{IV}_t - K_{H,t}^{BNHLS} \right)^2}.$$

We present an empirical application in the next section.

5.4 Empirical Application

The data are the transaction prices of Alcoa, a stock listed in the Dow Jones Industrials. The prices are observed every minute from January 1, 2002 to December 31, 2007 ($T = 1510$ trading days). In a typical trading day, the market is open from 9:30 am to 4:00 pm, and this results in $m = 390$ observations per day. There are a few missing observations (fewer than 5 missing data per day) which we filled in using the previous tick method, which amounts to replacing any missing price by the most recent price.

The estimation takes place in several steps. First of all, we compute the first step CGMM estimator of $\theta_1 = (\kappa, \beta, \sigma^2)$ based on the moment function (32):

$$\widehat{\theta}_1^{(1)} = \arg \min_{\theta_1} \left\| \widehat{h}_T(\cdot, \theta_1) \right\|^2.$$

Second, we use $\widehat{\theta}_1^{(1)}$ to estimate the covariance operator $K_T \left(\widehat{\theta}_1^{(1)} \right)$ associated with the moment function. We use the ad-hoc value $\lambda = 10^{-6}$ for the regularization of the inverse of $K_T \left(\widehat{\theta}_{1,T}^{(1)} \right)$, that is:

$$K_{T,10^{-6}}^{-1/2} \left(\widehat{\theta}_1^{(1)} \right) = \left(K_T^2 \left(\widehat{\theta}_1^{(1)} \right) + 10^{-6} \times I \right)^{-1/2} K_T^{1/2} \left(\widehat{\theta}_1^{(1)} \right).$$

The second step estimator of θ_1 is thus:

$$\widehat{\theta}_1^{(2)} = \arg \min_{\theta_1} \left\| K_{T,10^{-6}}^{-1/2} \widehat{h}_T(\cdot, \theta_1) \right\|^2,$$

where $K_{T,10^{-6}}^{-1/2} \equiv K_{T,10^{-6}}^{-1/2} \left(\widehat{\theta}_1^{(1)} \right)$. The objective functions of the CGMM are computed in each cases using 100 Hermitian quadrature points in \mathbb{R}^2 .

Third, we estimate the variance of $\widehat{\theta}_1^{(2)}$. Unfortunately, the analytical expression (16) is unusable because the gradient of $\widehat{h}_T(\tau, \theta_1)$ (given in appendix D) is extremely badly scaled. This is due to the fact that the likelihood function of Gamma distributions (like some Student distributions) are very flat around the true value of the degree of freedom parameter. As a result, the derivative of the objective function with respect to the degree of freedom parameter is very small relatively to the derivatives with respect to the remaining parameters. In other words, the gradient matrix is so badly scaled that it is numerically singular. The problem is even more severe in the Autoregressive Gamma model because the degree of freedom is $q = \frac{2\kappa\beta}{\sigma^2}$, that is, a function of all three parameters of interest. However, this numerical singularity does not imply that the model is not identified. Note that even in the discrete GMM, it is possible to have an over-identified set of restrictions that is first order under-identified (see Dovonon and Renault, 2009). We can thus resort to the bootstrap to evaluate the variance of $\widehat{\theta}_1^{(2)}$. The experiment is conducted as follows.

We use the initial sample of size $T = 1510$ to compute 1509 moment functions: $\{h_t(\tau, \theta_1)\}_{t=2}^T$. According to the specified model, $h_t(\tau, \theta_1)$ is a martingale difference sequence. Hence the bootstrap procedure can be performed as if the set of moment functions were independent. We thus draw 500 moment functions with equal probability and replacement from the above set to get $\left\{ \widetilde{h}_j^{(b)}(\tau, \theta_1) \right\}_{j=1}^{500}$, for $b = 1, 2, \dots, B = 1000$. Each sample $\left\{ \widetilde{h}_j^{(b)}(\tau, \theta_1) \right\}_{j=1}^{500}$ is then used to compute an estimator $\widehat{\theta}_{1,b}$ for $\theta_1 = (\kappa, \beta, \sigma^2)$. Finally, $\widehat{\theta}_{1,b}$ is used together with the realizations of V_t on which $\left\{ \widetilde{h}_j^{(b)}(\tau, \theta_1) \right\}_{j=1}^{500}$ depend to compute an estimator $\widehat{\theta}_{2,b}$ for $\theta_2 = (\mu_0, \mu_1, \delta)$. In computing $\widehat{\theta}_{2,b}$, the constraints $\mu \geq 0$ and $\delta \leq 0$ are explicitly imposed. Likewise, $(\kappa, \beta, \sigma^2) > 0$ is imposed in the estimation of θ_1 . The following table summarizes the empirical distributions of $\widehat{\theta}_{1,b}$ and $\widehat{\theta}_{2,b}$.

	$\widehat{\theta}_{1,b}$			$\widehat{\theta}_{2,b}$		
	$\widehat{\kappa}_b$	$\widehat{\beta}_b$	$\widehat{\sigma}_b^2$	$\widehat{\mu}_{0,b}$	$\widehat{\mu}_{1,b}$	$\widehat{\delta}_b$
Mean	0.8825	0.0004	0.0006	-0.0032	0.1665	-3.2039
Mode	0.8650	0.0004	0.0006	-0.0005	0.0664	-1.8341
Std. Dev.	0.0281	2.3×10^{-5}	7.8×10^{-5}	0.0046	0.2422	5.3613
IC1(95)	0.8135	0.0003	0.0005	-0.0154	0.0000	-19.807
IC2(95)	0.8917	0.0004	0.0008	0.0008	0.8009	0.0000

Table 4: Bootstrap statistics ($B = 1000$ samples of size $T = 500$).

The empirical results are qualitatively similar to what we have seen in the preliminary simulation study. In particular, $\widehat{\theta}_{1,b}$ is less volatile than $\widehat{\theta}_{2,b}$ and this shows up as wider confidence regions for $\widehat{\theta}_{2,b}$. We have learned from the simulation exercise that the mean of $\widehat{\theta}_{1,b}$ is a good guess of θ_1 while the mode of $\widehat{\theta}_{2,b}$ is a good guess of θ_2 . The estimated risk premium parameter is $Mode(\widehat{\mu}_{1,b}) = 0.0664$ while the leverage parameter is about $Mode(\widehat{\delta}_b) = -1.83$. The latter estimate is probably exaggerated in regard of the results of Table 3. In light of this, the results of Table 4 support the conclusion of French et al. (1987) according to which the expected return is positively correlated with the expected risk while the unexpected return is negatively correlated with the unpredictable risk.

6 Conclusions

The goal of this paper is to illustrate how to implement the CGMM. To start with, we briefly review the useful theoretical properties of the CGMM estimator. Next, we present in detail some helpful numerical methods for its implementation. Finally, we apply the estimation method to the stable distribution and the autoregressive variance Gamma model.

When the parameter α of the stable distribution is close to 2, the asymmetry parameter β becomes hard to identify. As a result, the gradient of the moment function is numerically singular and one has to rely on Monte Carlo simulations for inference on the identifiable parameters. When α is far from 2, the gradient of the moment function is of full rank and the standard errors of the estimators can be computed using the standard analytical formulas. Overall, the parameters of the stable distribution can be reliably estimated by the CGMM.

In the autoregressive Gamma model, the variances of the estimators cannot be computed analytically because the gradient of the moment is numerically singular. This problem is due to the fact that the objective function is extremely flat around the true values of the parameters, and can be linked to the difficulties inherent to the estimation of the degree of freedom parameter in Gamma distributions or Student distributions. We avoid this problem by generating the empirical distributions of the estimates by resampling from the original sample. The empirical application with the Alcoa stock suggests that expected return is positively correlated with expected risk while unexpected return is negatively correlated with unpredictable risk.

7 Acknowledgement

The author thanks Marine Carrasco, Russell Davidson and two anonymous referees for helpful comments.

References

- [1] Andrew, D.W.K. and J.C. Monahan (1992) “An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator”, *Econometrica*, vol. 60, 953-966.
- [2] Barndorff-Nielsen, O. E. and N. Shephard (2002) “Estimating quadratic variation using realized variance”, *Journal of Applied Econometrics*, vol. 17, 457-477.
- [3] Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and N. Shephard (2008) “Designing realized kernels to measure the ex-post variation of equity prices in the presence of noise”, *Econometrica*, vol. 76, 1481-1536.
- [4] Bauwens, L., Ben Omrane, W. and E. Rengifo (2010) “Intradaily dynamic portfolio selection”, *Computational Statistics and Data Analysis*, vol. 54, 2400-2418.
- [5] Bollerslev, T. (1986) “Generalized Autoregressive Conditional Heteroscedasticity”, *Journal of Econometrics*, vol. 31, 307-327.
- [6] Carrasco M., Chernov M., Florens J. P. and E. Ghysels (2007) “Efficient estimation of general dynamic models with a continuum of moment conditions”, *Journal of Econometrics*, vol. 140, 529-573.
- [7] Carrasco, M. and J. P. Florens (2000) “Generalization of GMM to a continuum of moment conditions”, *Econometric Theory*, vol. 16, 797-834.
- [8] Carrasco, M. and J. P. Florens (2002) “Efficient GMM Estimation Using the Empirical Characteristic Function”, *Working Paper*, Université de Montréal.
- [9] Carrasco, M., J. P. Florens, and E. Renault (2007) “Linear Inverse Problems in Structural Econometrics: Estimation based on spectral decomposition and regularization”, in the *Handbook of Econometrics*, vol. 6, edited by J.J. Heckman and E.E. Leamer.
- [10] Carrasco, M. and R. Kotchoni (2010a) “Efficient Estimation Using the Characteristic Function”, *Working paper*, Université de Montréal.
- [11] Carrasco, M. and R. Kotchoni (2010b) “Shrinkage Realized Kernels”, *Working paper*, Université de Montréal.
- [12] Chacko, G. and L. Viceira (2003) “Spectral GMM estimation of continuous-time processes”, *Journal of Econometrics*, vol. 116, 259-292.
- [13] Chambers, J. M., Mallows, C. L. and B.W. Stuck (1976), “A Method for Simulating Stable Random Variables”, *Journal of the American Statistical Association*, vol. 71, 340-344. Corrections 82 (1987):704, 83 (1988): 581.
- [14] Cornea, A. and R. Davidson (2009) “A Refined Bootstrap for Heavy Tailed Distributions”, *Working paper*, McGill University.

- [15] Devroye, L. (1986) “*Non-Uniform Random Variate Generation*”, Springer-Verlag.
- [16] Dovonon, P. and E. Renault (2009) “GMM Overidentification Test with First Order Underidentification”, *Working Paper*, University of North Carolina at Chapel Hill.
- [17] Duchesne, P. (2006) “Testing for multivariate autoregressive conditional heteroskedasticity using wavelets”, *Computational Statistics & Data Analysis*, vol. 51, 2142-2163.
- [18] Engle, R. F. (1982) “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of UK Inflation”, *Econometrica*, vol. 50, 987-1008.
- [19] Feller, W. (1951) “Two Singular Diffusion Problems”, *Annals of Mathematics*, vol. 54, 173-182.
- [20] Feuerverger, A. (1990) “An efficiency result for the empirical characteristic function in stationary time-series models”, *Canadian Journal of Statistics*, vol. 18, 155-161.
- [21] Feuerverger, A. and P. McDunnough (1981a) “On Efficient Inference in Symmetry Stable Laws and Processes”, *Csorgo, M., ed. Statistics and Related Topics*. New York: North Holland, 109-122.
- [22] Feuerverger, A. and P. McDunnough (1981b) “On the Efficiency of Empirical Characteristic Function Procedures”, *J. R. Statist. Soc. B*, vol. 43, 20-27.
- [23] French, K. R., Schwert, G. W. and R. F. Stambaugh (1987) “Expected Stock Returns and Volatility”, *Journal of Financial Economics* 19 (1987) 3-29 North-Holland.
- [24] Garcia, R., Renault, E. and D. Veredas (2006) “Estimation of Stable Distributions by Indirect Inference”, *CORE discussion paper 2006/112*.
- [25] Gouriéroux, C. and J. Jasiak (2005) “Autoregressive Gamma Processes”, *Journal of Forecasting*, vol. 25, 129–152. Published online in Wiley InterScience (www.interscience.wiley.com) DOI: 10.1002/for.978.
- [26] Gouriéroux, C., Jasiak, J. and R. Sufana (2009) “The Wishart Autoregressive Process”, *Journal of Econometrics*, vol. 150 (2) 167-181.
- [27] Gouriéroux, C., Monfort, A. and E. Renault (1993) “Indirect Inference,” *Journal of Applied Econometrics*, vol. 8, s85-s118.
- [28] Hansen, L. (1982) “Large Sample Properties of Generalized Method of Moments Estimators”, *Econometrica*, vol. 50, 1029-1054.
- [29] Heston, S. (1993) “A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options”, *The Review of Financial Studies*, vol. 6, number 2, 327-343.

- [30] Hull, J. and A. White (1987) “The pricing of options on assets with stochastic volatility”, *Journal of Finance*, vol. 3, 281-300.
- [31] Jacod, J. (1994) “Limit of Random Measures Associated with the Increments of a Brownian Semi-martingale”, No 120, *Laboratoire de Probabilités, Université Pierre et Marie Curie, Paris*.
- [32] Jacod, J. and P. Protter (1998) “Asymptotic error distribution for the Euler method for stochastic Differential Equations”, *Annals of Probability*, vol. 26, 267-307.
- [33] Jacquier, E., Polson, N., and P. E. Rossi (1994) “Bayesian analysis of stochastic volatility models (with discussion)”, *Journal of Business and Economic Statistics*, vol. 12 (4), 371-417.
- [34] Jacquier, E., Polson, N. G., and P. E. Rossi (2004) “Bayesian analysis of stochastic volatility models with fat-tails and correlated errors”, *Journal of Econometrics*, vol. 122, 185-212.
- [35] Jiang, G. and J. Knight (2002) “Estimation of Continuous Time Processes via the Empirical Characteristic Function”, *Journal of Business and Economic Statistics*, vol. 20, 198-212.
- [36] Knight, J. L. and J. Yu (2002) “The empirical characteristic function in time series estimation”, *Econometric Theory*, vol. 18, 691-721.
- [37] Knight, J. L., Satchell, S. E. and J. Yu (2002) “Estimation of the stochastic volatility model by the empirical characteristic function method”, *Australian New Zealand J. Statist.*, vol. 44:319–335.
- [38] Koutrouvelis, I. A. (1980) “Regression-type Estimation of the Parameters of Stable Laws”, *J. Am. Statist. Assoc.*, vol. 75, 918-928.
- [39] Liu, Q. and D. A. Pierce (1994) “A Note on Gauss-Hermite Quadrature”, *Biometrika*, Vol. 81, No. 3, 624-629.
- [40] Madan, D. B., Carr, P., and E. C. Chang (1998) “The variance gamma process and option pricing”, *European Finance Review*, vol. 2, 79-105.
- [41] Madan, D. B. and E. Senata (1990) “The Variance Gamma Model for Share Market Returns”, *Journal of Business*, vol. 63, no. 4.
- [42] Mandelbrot, B. (1963) “The variations of certain speculative prices”, *Journal of Business*, vol. 36, 394-419.
- [43] McCulloch, J.H. (1986) “Simple Consistent Estimators of Stable Distribution Parameters”, *Communications in Statistics - Simulation and Computation*, vol. 15, 1109-1136.

- [44] McCulloch, J. H. (1998) “Numerical approximation of the symmetric stable distribution and density” In *R. J. Adler, R. E. Feldman, & M. S. Taqqu (eds.), A Practical Guide to Heavy Tails*, 489–500. Boston: Birkhäuser.
- [45] Mittnik, S., Rachev, S. T., Doganoglu, T. and D. Chenyao (1999) “Maximum likelihood estimation of stable paretian models”, *Mathematical and Computer Modelling*, vol. 29, 276-293.
- [46] Newey, W. and K. West (1987) “A simple positive definite, heteroskedasticity and autocorrelation consistent covariance matrix”, *Econometrica*, vol. 55, 703-708.
- [47] Nolan, J. P. (1997) “Numerical calculation of stable densities and distribution functions”, *Communication Statistics: Stochastic Models*, vol. 13, 759-774.
- [48] Nolan, J. P. (2009) “*Stable Distributions: Models for Heavy Tailed Data*”, Birkhauser, Boston,. In progress.
- [49] Paolella, M. (2007) “*Intermediate Probability: A Computational Approach*”, Wiley, Chichester.
- [50] Paulson, A. S., Holcomb W. E. and R. A. Leitch (1975) “The estimation of the parameters of the stable laws”, *Biometrika*, vol. 62(1), 163-170.
- [51] Press, S. J. (1972). “Estimation of univariate and multivariate stable distributions”, *J. Am. Statist. Assoc.*, vol. 67, 842-846.
- [52] Singleton, K. (2001) “Estimation of Affine Pricing Models Using the Empirical Characteristic Function”, *Journal of Econometrics*, vol. 102, 111-141.
- [53] Stein, E. M. and J. C. Stein (1991) “Stock Price Distribution with Stochastic Volatility: an Analytic Approach”, *The Review of Financial Studies*, vol. 4 number 4, 727-752.
- [54] Weron, R. (1996) “On the Chambers-Mallows-Stuck method for simulating skewed stable random variables”, *Statistics & Probability Letters*, vol. 28, 165-171.
- [55] Yu, J. (2004) “Empirical Characteristic Function Estimation and Its Applications”, *Econometric Reviews*, vol. 23, No 2, 93-123.
- [56] Yu, J., Yang, Z. and X. Zhang (2006) “A class of nonlinear stochastic volatility models and its implications for pricing currency options”, *Computational Statistics & Data Analysis*, vol. 51, 2218-2231
- [57] Zolotarev, V. (1986), “One-Dimensional Stable Distributions”, *American Mathematical Society, Providence, RI*. Russian original, 1983.

Appendix A: Some Basic Properties of the Covariance Operator

For formal proofs of the results mentioned in this appendix, see Carrasco et al. (2007a,b). In the sequel, $\widehat{h}_t(\tau, \theta)$ is the moment function defined in (1) or (7), K is the covariance operator with kernel given by (3) and (9), and Φ_β is the subset of $L^2(\pi)$ defined in Assumption 4.

Definition 1 *The range of K denoted $R(K)$ is the set of functions g such that $Kf = g$ for some f in $L^2(\pi)$.*

Proposition 2 *$R(K)$ is a subspace of $L^2(\pi)$.*

Note that the kernel functions $k(s, \cdot)$ and $k(\cdot, r)$ are elements of $L^2(\pi)$ because

$$|k(s, r)|^2 = \left| E \left[h_t(\theta, s) \overline{h_t(\theta, r)} \right] \right|^2 \leq 4, \quad \forall (s, r) \in \mathbb{R}^{2p}.$$

Thus for any $f \in L^2(\pi)$, we have

$$\begin{aligned} |Kf(s)|^2 &= \left| \int k(s, r) f(r) \pi(r) dr \right|^2 \leq \int |k(s, r) f(r)|^2 \pi(r) dr \\ &\leq 4 \int |f(r)|^2 \pi(r) dr < \infty, \end{aligned}$$

implying

$$\|Kf\|^2 = \int |Kf(s)|^2 \pi(s) ds < \infty \Rightarrow Kf \in L^2(\pi).$$

Definition 3 *The null space of K denoted $N(K)$ is the set of functions f in $L^2(\pi)$ such that $Kf = 0$.*

The covariance operator K associated with a moment function based on the CF is such that $N(K) = \{0\}$ (See Carrasco et al. 2007a,b, for a proof).

Definition 4 *ϕ is an eigenfunction of K associated with eigenvalue μ if and only if $K\phi = \mu\phi$.*

Proposition 5 *Suppose $\mu_1 \geq \mu_2 \geq \dots \geq \mu_j \geq \dots$ are the eigenvalues of K . Then the sequence $\{\mu_j\}$ satisfies: (i) $\mu_j > 0$ for all j , (ii) $\mu_1 < \infty$ and $\lim_{j \rightarrow \infty} \mu_j = 0$.*

Remark. The covariance operator associated with the CF-based moment function is necessarily compact.

Proposition 6 *Every $f \in L^2(\pi)$ can be decomposed as: $f = \sum_{j=1}^{\infty} \langle f, \phi_j \rangle \phi_j$.*

As a consequence, we have:

$$K^{-\beta} f = \sum_{j=1}^{\infty} \langle f, \phi_j \rangle K^{\beta} \phi_j = \sum_{j=1}^{\infty} \mu_j^{-\beta} \langle f, \phi_j \rangle \phi_j, \beta \in \mathbb{R}.$$

Because $\lim_{j \rightarrow \infty} \mu_j = 0$, $\|K^{-1} f\|^2 = \sum_{j=1}^{\infty} \mu_j^{-1} |\langle f, \phi_j \rangle|^2$ is clearly not finite for all f .

Proposition 7 *If $0 < \beta_1 \leq \beta_2$, then $\Phi_{\beta_2} \subset \Phi_{\beta_1}$.*

We recall that Φ_{β} is the set of functions such that $\|K^{-\beta} f\| < \infty$. In fact, $f \in R(K^{\beta_2}) \Rightarrow K^{-\beta_2} f$ exist and $\|K^{-\beta_2} f\|^2 = \sum_{j=1}^{\infty} \mu_j^{-2\beta_2} |\langle f, \phi_j \rangle|^2 < \infty$. Thus if $f \in R(K^{\beta_2})$, we have:

$$\begin{aligned} \|K^{-\beta_1} f\|^2 &= \sum_{j=1}^{\infty} \mu_j^{2(\beta_2 - \beta_1)} \mu_j^{-2\beta_2} |\langle f, \phi_j \rangle|^2 \\ &\leq \mu_1^{2(\beta_2 - \beta_1)} \sum_{j=1}^{\infty} \mu_j^{-2\beta_2} |\langle f, \phi_j \rangle|^2 < \infty, \end{aligned}$$

$\Rightarrow K^{-\beta_1} f$ exists $\Rightarrow f \in R(K^{\beta_1})$. This means $R(K) \subset R(K^{1/2})$ so that the function $K^{-1/2} f$ is defined on a wider subset of $L^2(\pi)$ compared to $K^{-1} f$. When $f \in \Phi_1$, $\langle K^{-1/2} f, K^{-1/2} f \rangle = \langle K^{-1} f, f \rangle$. But when $f \in \Phi_{\beta}$ for $1/2 \leq \beta < 1$, the quadratic form $\langle K^{-1/2} f, K^{-1/2} f \rangle$ is well defined while $\langle K^{-1} f, f \rangle$ is not.

Appendix B: Other Parameterizations of the Stable Distribution and Simulation Strategies

The CF of the parameterization (21) which we refer to in this appendix as $S_\alpha^0(\beta_0, \sigma_0, \mu_0)$ is discontinuous around $\alpha = 1$ whenever $\beta_0 \neq 0$. To circumvent this, Zorotalev (1986) proposed to parameterize:

$$\mu_1 = \begin{cases} \mu_0 + \beta_0 \sigma_0^\alpha \tan \frac{\alpha\pi}{2}, & \alpha \neq 1 \\ \mu_0, & \alpha = 1 \end{cases}.$$

This results in the following expression for the CF which is continuous with respect to all the parameters:

$$\begin{aligned} & E[\exp(i\tau X)] \tag{36} \\ &= \begin{cases} \exp\{i\mu_1\tau - \sigma_0^\alpha [|\tau|^\alpha - i\tau\beta_0 (|\tau|^{\alpha-1} - 1) \tan \frac{\alpha\pi}{2}]\}, & \alpha \neq 1 \\ \exp\{i\mu_1\tau - \sigma_0 |\tau| [1 + i\frac{2}{\pi}\beta_0 \text{sign}(\tau) \ln |\tau|]\}, & \alpha = 1 \end{cases} \end{aligned}$$

Let us call this new parameterization $S_\alpha^1(\beta_0, \sigma_0, \mu_1)$.

Using (21) as starting point, Nolan (1997) proposed:

$$\mu_2 = \begin{cases} \mu_0 + \beta_0 \sigma_0 \tan \frac{\alpha\pi}{2}, & \alpha \neq 1 \\ \mu_0 + \frac{2}{\pi}\beta_0 \sigma_0 \ln \sigma_0, & \alpha = 1 \end{cases}.$$

This yields another continuous representation of the CF:

$$\begin{aligned} & E[\exp(i\tau X)] \tag{37} \\ &= \begin{cases} \exp\{i\mu_2\tau - \sigma_0^\alpha |\tau|^\alpha [1 + i\beta_0 \text{sign}(\tau) (|\sigma_0\tau|^{1-\alpha} - 1) \tan \frac{\alpha\pi}{2}]\}, & \alpha \neq 1 \\ \exp\{i\mu_2\tau - \sigma_0 |\tau| [1 + i\frac{2}{\pi}\beta_0 \text{sign}(\tau) \ln |\sigma_0\tau|]\}, & \alpha = 1 \end{cases} \end{aligned}$$

The parameterization (37) will be referred to as $S_\alpha^2(\beta_0, \sigma_0, \mu_2)$. An important feature of this parameterization is that $\frac{X-\mu_2}{\sigma_0} \sim S_\alpha^2(\beta_0, 1, 0)$, no matter the value of α . This is true for the two other parameterizations only when $\alpha \neq 1$.

An alternative parameterization $S_\alpha^3(\beta_0, \sigma_0, \mu_3)$ tied to the data simulation method of Chambers, Mallows and Stuck (1976) is got by setting:

$$\mu_3 = \begin{cases} \mu_0 + \beta_0 \sigma_0 \tan \frac{\alpha\pi}{2}, & \alpha \neq 1 \\ \mu_0, & \alpha = 1 \end{cases}.$$

This is identical to $S_\alpha^2(\beta_0, \sigma_0, \mu_2)$ for the case $\alpha \neq 1$. As pointed out by Nolan (2009), these small changes in parameterization have caused much confusions in the literature. For instance, some papers build their theoretical framework on the parameterization $S_\alpha^0(\beta_0, \sigma_0, \mu_0)$ but simulate the data under the parameterization $S_\alpha^3(\beta_0, \sigma_0, \mu_3)$.

Another important parameterization proposed in Zorotalev (1986) allows one to derive an integral representation of the probability distribution function of α -stable random variables (See Zolotarev, 1986, Remark 1, page 78 or Weron, 1996).

Different simulation strategies for the stable distribution

A method to simulate from the parameterization $S_\alpha^0(\beta_0, \sigma_0, \mu_0)$ is presented in Weron (1996). To start with, one draws two independent uniforms v and w in $[0,1]$ and calculate: $V = \pi(u - 1/2)$ and $W = -\ln w$. Then $Z \sim S_\alpha^0(\beta_0, 1, 0)$ and $X \sim S_\alpha^0(\beta_0, \sigma_0, \mu_0)$ are obtained as follows:

- If $\alpha \neq 1$, one computes:

$$Z = S_{\alpha, \beta_0} \frac{\sin(\alpha V + \alpha B_{\alpha, \beta_0})}{(\cos V)^{1/\alpha}} \left(\frac{\cos((1-\alpha)V - \alpha B_{\alpha, \beta_0})}{W} \right)^{-1+1/\alpha},$$

where $B_{\alpha, \beta_0} = \frac{\arctan(\beta_0 \tan \frac{\alpha\pi}{2})}{\alpha}$ and $S_{\alpha, \beta_0} = (1 + \beta_0^2 \tan^2 \frac{\alpha\pi}{2})^{\frac{1}{2\alpha}}$. Then we have

$$\begin{aligned} X &= \sigma_0 Z + \mu_0 \sim S_\alpha^0(\beta_0, \sigma_0, \mu_0), \\ X &= \sigma_0 Z + \mu_1 - \beta_0 \sigma_0^\alpha \tan \frac{\alpha\pi}{2} \sim S_\alpha^1(\beta_0, \sigma_0, \mu_1) \text{ and} \\ X &= \sigma_0 \left(Z - \beta_0 \tan \frac{\alpha\pi}{2} \right) + \mu_2 \sim S_\alpha^2(\beta_0, \sigma_0, \mu_2). \end{aligned}$$

- If $\alpha = 1$, one computes instead:

$$Z = \frac{2}{\pi} \left[\left(\frac{\pi}{2} + \beta_0 V \right) \tan V - \beta \log \left(\frac{W \cos V}{\frac{\pi}{2} + \beta_0 V} \right) \right].$$

We then have:

$$\begin{aligned} X &= \sigma_0 Z + \mu_0 + \frac{2}{\pi} \beta_0 \sigma_0 \ln \sigma_0 \sim S_1^0(\beta_0, \sigma_0, \mu_0), \\ X &= \sigma_0 Z + \mu_1 + \frac{2}{\pi} \beta_0 \sigma_0 \ln \sigma_0 \sim S_1^1(\beta_0, \sigma_0, \mu_1) \text{ and} \\ X &= \sigma_0 Z + \mu_2 \sim S_1^2(\beta_0, \sigma_0, \mu_2). \end{aligned}$$

The simulation strategy of Z for the case $\alpha = 1$ is quite standard in the literature. However, other methods (than the one above) have been used in the literature for the case $\alpha \neq 1$. We show the link between (22) and two of them below. To start with, note that:

$$\begin{aligned} \sin(\alpha V + \alpha B_{\alpha, \beta_0}) &= \sin \alpha V \sin \alpha B_{\alpha, \beta_0} + \cos \alpha V \cos \alpha B_{\alpha, \beta_0} \text{ and} \\ \cos((1-\alpha)V - \alpha B_{\alpha, \beta_0}) &= \cos(1-\alpha)V \cos \alpha B_{\alpha, \beta_0} + \sin(1-\alpha)V \sin \alpha B_{\alpha, \beta_0}. \end{aligned}$$

Also, due to $\sin a = \cos a \tan a$ for all a , we have: $\sin \alpha B_{\alpha, \beta_0} = \beta_0 \tan \frac{\alpha\pi}{2} \cos \alpha B_{\alpha, \beta_0}$ so that:

$$\cos^2 \alpha B_{\alpha, \beta_0} = 1 - \sin^2 \alpha B_{\alpha, \beta_0} = 1 - \beta_0^2 \tan^2 \frac{\alpha\pi}{2} \cos^2 \alpha B_{\alpha, \beta_0}.$$

The last equation implies:

$$\cos^2 \alpha B_{\alpha, \beta_0} = \frac{1}{1 + \beta_0^2 \tan^2 \frac{\alpha\pi}{2}}.$$

Replacing this in the expressions of $\sin(\alpha V + \alpha B_{\alpha, \beta_0})$ and $\cos((1 - \alpha)V - \alpha B_{\alpha, \beta_0})$ yields:

$$\begin{aligned} S_{\alpha, \beta_0} &= (\cos \alpha B_{\alpha, \beta_0})^{-1/\alpha}, \\ \sin(\alpha V + \alpha B_{\alpha, \beta_0}) &= \frac{\beta_0 \tan \frac{\alpha\pi}{2} \sin \alpha V + \cos \alpha V}{\sqrt{1 + \beta_0^2 \tan^2 \frac{\alpha\pi}{2}}} \text{ and} \\ \cos((1 - \alpha)V - \alpha B_{\alpha, \beta_0}) &= \left(\frac{\cos(1 - \alpha)V + \beta_0 \tan \frac{\alpha\pi}{2} \sin(1 - \alpha)V}{\sqrt{1 + \beta_0^2 \tan^2 \frac{\alpha\pi}{2}}} \right)^{-1+1/\alpha}. \end{aligned}$$

Putting these expressions together in Equation (22) yields:

$$Z = \frac{\beta_0 \tan \frac{\alpha\pi}{2} \sin \alpha V + \cos \alpha V}{(\cos V)^{1/\alpha}} \left(\frac{\cos(1 - \alpha)V + \beta_0 \tan \frac{\alpha\pi}{2} \sin(1 - \alpha)V}{W} \right)^{-1+1/\alpha}. \quad (38)$$

Adding $-\beta \tan \frac{\alpha\pi}{2}$ to the above expression yields the formula of Chambers, Mallows and Stuck (1976). To get the alternative expression of Nolan (2009), Theorem 1.19, it suffices to substitute for $S_{\alpha, \beta_0} = (\cos \alpha B_{\alpha, \beta_0})^{-1/\alpha}$ in Equation (22). This yields:

$$Z = \frac{\sin(\alpha V + \alpha B_{\alpha, \beta_0})}{(\cos \alpha B_{\alpha, \beta_0} \cos V)^{1/\alpha}} \left(\frac{\cos((1 - \alpha)V - \alpha B_{\alpha, \beta_0})}{W} \right)^{-1+1/\alpha}. \quad (39)$$

The evaluation of (39) for values very close to $\alpha = 1$ may raise some numerical problems. By avoiding the division by $\cos \alpha B_{\alpha, \beta_0}$, the expressions (22) and (38) are more numerically stable and accurate (see Nolan, 2009).

Appendix C: Gradients of the moment function (25)

$$\begin{aligned}
\frac{\partial h_t(\tau, \theta)}{\partial \rho_0} &= -i\tau_1 \varphi_t(\tau_1, \theta) e^{i\tau_2 r_{t-1}}, \\
\frac{\partial h_t(\tau, \theta)}{\partial \rho_1} &= -i\tau_1 y_{t-1} \varphi_t(\tau_1, \theta) e^{i\tau_2 r_{t-1}}, \\
\frac{\partial h_t(\tau, \theta)}{\partial \alpha} &= \sigma^\alpha |\tau_1|^\alpha \left\{ \log(\sigma |\tau_1|) \left[1 - i\beta \text{sign}(\tau_1) \tan \frac{\alpha\pi}{2} \right] - \frac{i\pi\beta \text{sign}(\tau_1)}{2 \cos^2 \frac{\alpha\pi}{2}} \right\} \varphi_t(\tau_1, \theta) e^{i\tau_2 r_{t-1}}, \\
\frac{\partial h_t(\tau, \theta)}{\partial \beta} &= -\sigma^\alpha |\tau_1|^\alpha i \text{sign}(\tau_1) \tan \frac{\alpha\pi}{2} \varphi_t(\tau_1, \theta) e^{i\tau_2 r_{t-1}} \text{ and} \\
\frac{\partial h_t(\tau, \theta)}{\partial \sigma} &= \alpha \sigma^{\alpha-1} |\tau_1|^\alpha \left[1 - i\beta \text{sign}(\tau_1) \tan \frac{\alpha\pi}{2} \right] \varphi_t(\tau_1, \theta) e^{i\tau_2 r_{t-1}}.
\end{aligned}$$

Appendix D: Gradients of the moment function (32)

The moment function satisfies:

$$\frac{\partial h_t(\tau, \theta_1)}{\partial \theta_1} = \frac{\partial \varphi(\theta, V_{t-1})}{\partial \theta_1} \exp(i\tau_2 V_{t-1}),$$

where $\varphi(\theta, V_{t-1})$ is the CF of V_t conditional on V_{t-1} given by:

$$\varphi(\theta_1, V_{t-1}) = \left(1 - \frac{i\tau\sigma^2(1 - e^{-\kappa})}{2\kappa}\right)^{-\frac{2\kappa\beta}{\sigma^2}} \exp\left(\frac{2i\kappa\tau e^{-\kappa}V_{t-1}}{2\kappa - i\tau\sigma^2(1 - e^{-\kappa})}\right).$$

We now compute $\frac{\partial \varphi(\theta, V_{t-1})}{\partial \theta_1}$.

Derivative with respect to β :

$$\begin{aligned} \frac{\partial \varphi(\theta_1, V_{t-1})}{\partial \beta} &= \frac{-2\kappa}{\sigma^2} \ln\left(1 - \frac{i\tau\sigma^2(1 - e^{-\kappa})}{2\kappa}\right) \exp\left(\frac{2\kappa i\tau e^{-\kappa}V_{t-1}}{2\kappa - i\tau\sigma^2(1 - e^{-\kappa})}\right) \\ &\quad \times \left(1 - \frac{i\tau\sigma^2(1 - e^{-\kappa})}{2\kappa}\right)^{-\frac{2\kappa\beta}{\sigma^2}}. \end{aligned}$$

Derivative with respect to κ :

$$\begin{aligned} \frac{\partial \varphi(\theta_1, V_{t-1})}{\partial \kappa} &= \exp\left(\frac{2\kappa i\tau e^{-\kappa}V_{t-1}}{2\kappa - i\tau\sigma^2(1 - e^{-\kappa})}\right) \left(1 - \frac{i\tau\sigma^2(1 - e^{-\kappa})}{2\kappa}\right)^{-\frac{2\kappa\beta}{\sigma^2}} \\ &\quad \times \left[\frac{-2\beta}{\sigma^2} \left(\ln\left(1 - \frac{i\tau\sigma^2(1 - e^{-\kappa})}{2\kappa}\right) + \frac{i\tau\sigma^2(1 - e^{-\kappa} - \kappa e^{-\kappa})}{2\kappa - i\tau\sigma^2(1 - e^{-\kappa})} \right) \right. \\ &\quad \left. - \frac{2i\tau e^{-\kappa}V_{t-1}}{2\kappa - i\sigma^2\tau(1 - e^{-\kappa})} \left(\frac{i\sigma^2\tau(\kappa - 1 + e^{-\kappa}) - 2\kappa^2}{2\kappa - i\sigma^2\tau(1 - e^{-\kappa})} \right) \right]. \end{aligned}$$

Derivative with respect to σ^2 :

$$\begin{aligned} \frac{\partial \varphi(\theta_1, V_{t-1})}{\partial \sigma^2} &= \exp\left(\frac{2\kappa i\tau e^{-\kappa}V_{t-1}}{2\kappa - i\tau\sigma^2(1 - e^{-\kappa})}\right) \left(1 - \frac{i\tau\sigma^2(1 - e^{-\kappa})}{2\kappa}\right)^{-\frac{2\kappa\beta}{\sigma^2}} \\ &\quad \times \left[\frac{2\kappa\beta}{\sigma^4} \left(\frac{i\tau\sigma^2(1 - e^{-\kappa})}{2\kappa - i\tau\sigma^2(1 - e^{-\kappa})} + \ln\left(1 - \frac{i\sigma^2\tau(1 - e^{-\kappa})}{2\kappa}\right) \right) \right. \\ &\quad \left. + \frac{i\tau(1 - e^{-\kappa})}{2\kappa - i\tau\sigma^2(1 - e^{-\kappa})} \frac{2i\kappa\tau e^{-\kappa}V_{t-1}}{2\kappa - i\tau\sigma^2(1 - e^{-\kappa})} \right]. \end{aligned}$$