The Mean Variance Mixing GARCH (1,1) model

-a new approach to estimate conditional skewness.*

Anders Eriksson Lars Forsberg

First draft: November 2003
This version: May 21, 2004

Abstract

Here we present a general framework for a GARCH (1,1) type of process with innovations with a probability law of the mean- variance mixing type, therefore we call the process in question the mean variance mixing GARCH (1,1) or MVM GARCH (1,1). One implication is a GARCH model with skewed innovations and constant mean dynamics. This is achieved without using a location parameter to compensate for time dependence that affects the mean dynamics. From a probabilistic viewpoint the idea is straightforward. We just construct our stochastic process from the desired behavior of the cumulants. Further we provide explicit expressions for the unconditional second to fourth cumulants for the process in question. In the paper we present a specification of the MVM-GARCH process where the mixing variable is of the inverse Gaussian type. On the basis on this assumption we can formulate a maximum likelihood based approach for estimating the process closely related to the approach used to estimate an ordinary GARCH (1,1). Under the distributional assumption that the mixing random process is an inverse Gaussian i.i.d process the MVM-GARCH process is then estimated on log return data from the Standard and Poor 500 index. An analysis for the conditional skewness and kurtosis implied by the process is also presented in the paper.

*Anders Eriksson (corresponding author), Department of Information Science - Division of Statistics, University of Uppsala. email:anders.eriksson@dis.uu.se. Lars Forsberg, Department of Information Science - Division of Statistics, University of Uppsala. email:lars.forsberg@dis.uu.se.
The authors thank *The Jan Wallander and Tom Hedelius Research Foundation* and *The Swedish Foundation for International Cooperation in Research and Higher Education, (STINT)* for financial support.

1 Introduction

The introduction of the Autoregressive Conditional Heteroscedastic (ARCH) process by Engle (1982) was among a lot of other things a new very powerful tool in the modeling of financial data in general and stock returns in particular His suggested process was different in relation to earlier conventional time series models in that it instead of the assumption of constant variances allowed the conditional variances to change through time as functions of past errors. The ARCH process can be written as a product of two entities $\sqrt{h_t}\varepsilon_t$, where h_t is the variance function i.e. it specifies the dependence structure of the variance. ε_t is an i.i.d sequence of standard normal random variables. However in general the work of improving the univariate ARCH models to be more coherent with financial data has taken two main directions.

1. The first approach in trying to improve the ARCH model starts out in an alternative specification of the variance function h_t. One celebrated improvement was introduced in Bollerslev (1986) where the Generalized Autoregressive Conditional Heteroscedastic (GARCH) process was presented. Further we have the Integrated GARCH (IGARCH) Engle and Bollerslev (1986) and the exponential GARCH (EGARCH) Nelson (1991) where the focus is on the respecification on variance equation. Other examples are the suggestions to introduce different kinds of asymmetry in the variance equation see Glosten, Jaganathan, and Runkle (1993) and Zakoian (1994). Another important extension of the ARCH model is the ARCH in mean model (ARCH-M) model introduced in Engle, Lilien, and Robins (1987) which extends the ARCH model to allow the conditional variance to affect the mean.

2. The second approach has taken the direction of finding a more realistic assumption regarding the ε_t i.i.d sequence. Examples on this research direction is Bollerslev (1987) where the ε_t are assumed to follow a student t i.i.d sequence. Other examples are when ε_t follows a symmetric normal inverse Gaussian i.i.d sequence first mentioned in Barndorff-Nielsen (1997) and explicitly formulated in Andersson (2001) and Jensen and Lunde (2001). In Jensen and Lunde (2001) the issue of skewed innovations with GARCH type errors is investigated. In the context of the EGARCH see Nelson (1991), the generalized error distribution (GED) was introduced as an assumption for the ε_t sequence.

In this paper we have chosen to elaborate with the distributional assumptions regarding the ARCH process. The main reason for doing this are the difficulties of obtaining a process which exhibits conditional skewness by just dealing with the variance equation. These difficulties occur since the ARCH process from the beginning was defined to be a symmetrical process and the variance equation defined as a scaling of a symmetrical distribution and therefore it only affects even moments.

A relevant question to ask yourself in context of above discussion: Do returns on financial assets really exhibit conditional skewness? The question is worthwhile asking since skewness in the stochastic process for returns have many implications; asset returns Harvey and Siddique (2000a), portfolio construction Kraus and Litzberger (1976) and of course risk management, this since skewness affects the tail mass. Further there is some literature that suggests that skewness helps to explain the market risk premium, Harvey and Siddique (2000b). The answer on this question will probably depend on the type of asset that you are investigating for the moment. However there are some empirical investigations that suggests that some types of financial returns are skewed at least unconditionally see for instance Peiró (1999), Badrinith and Chatterjee (1988) and Simkowitz and Beedles (1980).

This paper presents a stochastic process with innovations related to a GARCH(1,1) process where the conditional probability measure exhibits skewness and excess kurtosis. Some of the more interesting earlier attempts to present such a process is:

1. Hansen (1994) suggested the Autoregressive Conditional Density (ACD) estimator. Here both the variance and skewness are indexed by time. As a starting point a skewed version of the student t probability law with separate time dependence structure for the conditional skewness is used.

2. Harvey and Siddique (1999) presented a model which they claimed to be an alternative parametrization of the Hansen model. The time-varying skewness is obtained by solving a non-linear equations system linking the first and third conditional moments.

3. Lee and Tse (1991) suggested an approach based on an approximative probability measure using a Gram Charlier series see Charlier (1905) of order 4. This scaled with a GARCH type variance equation. This results in a process with time-varying skewness.

The first two papers have the common feature of stating a skewness parameter as a function of the conditioning information set. One problem becomes to choose which function that captures the time dependence in the third moment best. This is not always an easy task. In the third paper the conditional variance determines the conditional skewness but the probability measure used is only an approximation. Therefore it will not be a valid probability measure. That is for certain combinations of skewness and excess kurtosis the approximation will not always be a valid probability measure or run into problems with multi modality. This is a result of the characteristics of Gram Charlier expansions, see for instance Draper and Tierny (1972).

This paper is organized as follows. Section 2 contains a motivation for why the conditional skewness could be sufficiently modeled by a GARCH variance function. This section is finalized in a conjecture that states how the time dependence for moments higher than two should be specified. Section 3 presents the general process and the conditional and unconditional moments of the suggested process. Section 4 contains the MVM $GARCH(1,1)$ process under inverse Gaussian distributional assumptions. In connection with this we present an approach for achieving maximum likelihood estimates of the process. We also estimate the process suggested and interpret the results in the context of conditional skewness and excess kurtosis. Section 5 consists of a description of two possible extensions from the general framework presented in this paper. The paper is ended with some concluding remarks in section 6.

2 Motivation

In this section we focus on the behavior of the time dependence for powers of stock returns data. The main purpose of this examination is to motivate that it is enough to model the time dependence for the second moment in order to model the time dependence for higher moments. In other words this means that there is no need to specify a particular function of the conditional information set to model the conditional skewness, the same function that models the time dependence in conditional variance can be used for the time dependence in conditional skewness.

The data used is daily log returns on the Standard and Poor 500 index obtained from the Ecowin v.3.1 database. The range for the data is 1 JAN 1997 to 1 JAN 2000. For descriptive statistics concerning this data see table 5. In order to continue the investigation we make following two definitions.

Definition 2.1 *The returns are given by* $r_t = \ln(S_t) - \ln(S_{t-1})$ *where* S_t *the index value at time point* t*. Further* r_t^a *denotes the a:th power of the return series. For example* r_t^2 *is the squared return series.*

Definition 2.2 ($GARCH(1,1)$ **process)** *A* $GARCH\ (1,1)$ *process is defined by:*

$$Y_t = m + \sigma_t Z_t$$

where $\mathcal{L}(Z_t) = N(0,1)$ Z_t *is a i.i.d sequence of random variables.* $\sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2 + \alpha_2 \sigma_{t-1}^2$ *where* α_0, α_1 *and* $\alpha_2 \in \mathbb{R}^+$ *and* $m \in \mathbb{R}$

For further insight in this process see Bollerslev (1986) The scheme for this investigation can be divided into four parts which can be viewed below.

1. We calculate a correlogram for r_t^a when $a = \{1, 2, 3, 4\}$.
2. We estimate a GARCH (1,1) model on the data and obtain the $GARCH(1,1)$ variance series, denoted as $\hat{\sigma}_t^2$.
3. Construct the standardized series $\tilde{r}_t = r_t / \sqrt{\hat{\sigma}_t^2}$.
4. Calculate correlogram for $\tilde{r}_t^a$ when $a = \{1, 2, 3, 4\}$.

The results from this investigation can be viewed on the forthcoming pages, where table 1 and 2 contains correlogram output data for r_t^a and $\tilde{r}_t^a$. Estimation results for the $GARCH(1,1)$ process are presented in table 3.

Table 1: Correlogram non-standardized returns

	$\tilde{r_t}$		$\tilde{r_t}^2$		$\tilde{r_t}^3$		$\tilde{r_t}^4$	
Lags	AC	Prob	AC	Prob	AC	Prob	AC	Prob
1	-0.07	0.08	0.21	0.21	-0.20	0.00	0.15	0.00
2	0.02	0.19	0.12	0.08	0.07	0.00	0.04	0.00
3	-0.07	0.07	0.05	0.01	-0.03	0.00	0.01	0.00
4	-0.04	0.09	0.04	0.02	0.01	0.00	0.00	0.00
5	-0.05	0.08	0.14	0.14	-0.14	0.00	0.10	0.00
6	0.04	0.09	0.01	-0.05	0.00	0.00	-0.01	0.00
7	-0.06	0.07	0.10	0.09	-0.01	0.00	0.01	0.00
8	0.05	0.06	0.07	0.03	-0.01	0.00	0.01	0.00
9	0.01	0.10	0.03	0.00	-0.01	0.00	0.00	0.00
10	0.06	0.07	0.02	-0.02	-0.01	0.00	0.00	0.01

Prob denotes the p-value for the null hypothesis of zero autocorrelation.

Table 2: Correlogram standardized returns

	$\tilde{r_t}$		$\tilde{r_t}^2$		$\tilde{r_t}^3$		$\tilde{r_t}^4$	
Lags	AC	Prob	AC	Prob	AC	Prob	AC	Prob
1	-0.01	0.75	-0.02	0.58	-0.02	0.52	0.00	0.99
2	0.03	0.69	0.08	0.12	0.08	0.09	0.04	0.64
3	-0.06	0.41	-0.04	0.14	-0.01	0.18	-0.01	0.82
4	-0.06	0.24	-0.02	0.22	-0.02	0.26	0.00	0.92
5	-0.03	0.28	0.01	0.34	-0.03	0.33	0.00	0.97
6	0.04	0.27	-0.03	0.39	0.00	0.46	-0.01	0.99
7	-0.07	0.15	0.04	0.41	-0.02	0.54	0.00	1.00
8	0.05	0.14	0.01	0.50	0.00	0.64	-0.01	1.00
9	0.02	0.18	0.00	0.60	0.00	0.74	-0.01	1.00
10	0.07	0.10	-0.02	0.66	0.00	0.81	-0.01	1.00

Prob denotes the p-value for the null hypothesis of zero autocorrelation.

Table 3: GARCH estimates

	Estimate	Standard error
$\hat{m}$	$1.11 \cdot 10^{-3}$	$4.52 \cdot 10^{-4}$
$\hat{\alpha_0}$	$1.52 \cdot 10^{-5}$	$4.66 \cdot 10^{-6}$
$\hat{\alpha_1}$	0.111	0.020
$\hat{\alpha_2}$	0.786	0.047

Under the assumption that the calculated correlograms can be viewed as estimators of correlation structures it seems obvious that by modeling the second moment we model the correlation structure in higher powers in the data. That suggests that the dependence in higher moments is sufficiently modeled by a time dependent scaling i.e. a $GARCH(1,1)$.

Conjecture 2.1 (Regarding stock returns data) *A time dependent scaling of a Gaussian random variable, i.e. a GARCH, is sufficient to model the time dependence in moments equal and higher than two. This is valid at least on the lower frequencies such as daily.*

This conjecture is the basis for the formulation of the mean variance mixing $GARCH(1,1)$ process.

3 The mean variance mixing $GARCH(1,1)$ process.

Here we present the mean variance mixing GARCH process. For the sake of probabilistic stringency when dealing with stochastic processes we need to make some definitions concerning the filtered probability space on which our stochastic process evolves.

Definition 3.1 (General probability space) *State a general probability space* $(\Omega, \mathcal{F}, \mathbb{P})$. *Where* Ω *is the set of all possible outcomes,* $\mathcal{F}$ *is the sigma field associated with the probability space containing all sets for which we want to make a statement on.* $\mathbb{P}$ *is the probability measure that generates the probability that such a set in* $\mathcal{F}$ *will occur.*

A sigma field can be defined as a family of subsets of Ω closed under any countable collection of set operations. For a more detailed discussion about the construction of sigma fields see Billingsley (1995) p. 30-32.

Definition 3.2 (Filtration) *Define a general filtration* $\mathbb{F} = (\mathcal{F}_t)_{t \in \mathbb{T}}$, *associated with the above probability space, where* $\mathbb{T} = \{0, 1, 2, ..., T\}$ *Where* $\mathcal{F}_t$ *is characterized by being an increasing sequence of sigma fields.*

For a more in depth analysis of sigma filtration and the construction of stochastic processes we refer to chapter one in Karatzas and Shreve (1991).

Definition 3.3 (Stochastic process) *Consider a stochastic process* $Y = (Y_t)$ *defined on the stochastic basis (or filtered probability space),denoted with the following pentet* $(\Omega, \mathcal{F}, \mathbb{P}, \mathbb{F}, \mathbb{T})$. *Recall that each* Y_t *is* $\mathcal{F}_{t-1}$ *measurable for each* $t \in \mathbb{T}$.

On the above filtered probability space we will construct our stochastic process. Let us now be a little more specific in how to define our stochastic process. Consider the following construction for a MVM GARCH (1,1) process.

Definition 3.4 (The $MVM\ GARCH(1,1)$ **process)** *The MVM GARCH (1,1) stochastic process is defined by:*

$$Y_t = \mu + \lambda V_t + \sigma_t \sqrt{V_t} Z_t$$

where $\mathcal{L}\ (V_t)$ *is a probability measure with the domain on* $\mathbb{R}^+$ *with finite moments* $\mathcal{L}\,(Z_t) = N(0,1)$ $\lambda, \mu \in \mathbb{R}$. *Both* V_t *and* Z_t *are i.i.d sequences of random variables.*

The GARCH equation is given by:

$$\sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2 + \alpha_2 \sigma_{t-1}^2$$

where α_0, α_1 *and* $\alpha_2 \in \mathbb{R}^+$

The above defined process will in the case of a λ parameter different from zero exhibit skewness which comes clear when looking at the conditional cumulants presented below. Further we know that

$$\mathcal{L}\left(Y_t | \sigma_t^2, V_t\right) = N(\mu + \lambda V_t, \sigma_t^2 V_t)$$

One drawback with the suggested process is that standardization of data becomes unfeasible under the parametrization suggested here. This is so since we have conditional mean of the process that contains a stochastic variable.

Below we state the conditional cumulants for the $MVM\ GARCH(1,1)$ process. The main reason to include these is to characterize the conditional behavior of the process. From the below proposition we can see that the conditional first moment of the process is remaining constant. However the higher order conditional cumulants are functions of the GARCH equation. So conditionally the time-dependence is just a function of the scaling of the second moment of the symmetrical part of the $MVM\ GARCH(1,1)$ process.

Proposition 3.1 (Conditional cumulants)

$$\kappa_1^{Y_t}|\mathcal{F}_{t-1} = \mu + \lambda\kappa_1^{V_t} \tag{3.1}$$
$$\kappa_2^{Y_t}|\mathcal{F}_{t-1} = E\left[\sigma_t^2|\mathcal{F}_{t-1}\right]\kappa_1^{V_t} + \lambda^2\kappa_2^{V_t} \tag{3.2}$$
$$\kappa_3^{Y_t}|\mathcal{F}_{t-1} = 3\lambda\kappa_2^{V_t}E\left[\sigma_t^2|\mathcal{F}_{t-1}\right] + \lambda^3\kappa_3^{V_t} \tag{3.3}$$
$$\kappa_4^{Y_t}|\mathcal{F}_{t-1} = \lambda^4\kappa_4^{V_t} + 6\lambda^2\kappa_3^{V_t}E\left[\sigma_t^2|\mathcal{F}_{t-1}\right] + 3E\left[\sigma_t^4|\mathcal{F}_{t-1}\right]\left(\kappa_2^{V_t} + \left(\kappa_1^{V_t}\right)^2\right) \tag{3.4}$$

where the $\kappa_i^{Y_t}$ *denotes the i:te cumulant of the process* Y_t *conditional on the filtration* $\mathcal{F}_{t-1}$*. Further is* $E\left[\sigma_t^2|\mathcal{F}_{t-1}\right] = \sigma_t^2$ *and* $E\left[\sigma_t^4|\mathcal{F}_{t-1}\right] = \sigma_t^4$ *which not should be mixed up with the unconditional expectation denoted as* $E(\sigma_t^2)$ *and* $E(\sigma_t^4)$*.*

Proof see Appendix A

One implication from the conditional cumulants is that the conditional skewness not can change sign. It will, which will be shown below when we derive the unconditional cumulants, always have the same sign as the unconditional skewness.

3.1 Unconditional Cumulants for the MVM $GARCH(1,1)$ stochastic process

One important step in order to characterize the MVM $GARCH(1,1)$ stochastic process is to calculate the unconditional cumulants of the process. Below we present the second, third and fourth cumulant. From these expression we can derive conditions on the parameters for the existence of the first four cumulants in the unconditional distribution for the MVM $GARCH$ process.

Theorem 3.1 (unconditional second cumulant) *The unconditional second cumulant of the* MVM $GARCH(1,1)$ *process is*

$$\kappa_2^{Y_t} = \left(\frac{\alpha_0 E\left(V_t\right) + \left(1-\alpha_2\right)E\left(\mu+\lambda V_t\right)^2}{1-\alpha_1 E\left(V_t\right) - \alpha_2} - E\left(\mu+\lambda V_t\right)^2\right) + \lambda^2 var(V_t) \tag{3.5}$$

Proof: see Appendix B

Remark 3.1 (existence criteria second cumulant) *From theorem 3.1 we can conclude that the second cumulant for the MVM GRACH (1,1) process exits if*

$$\alpha_1 E\left(V_t\right) + \alpha_2 > 1$$

Corollary 3.1 (unconditional third cumulant) *The unconditional third cumulant of the MVM GARCH(1,1) process is*

$$\kappa_3^{Y_t} = 3\lambda\left(\frac{\alpha_0 E\left(V_t\right) + \left(1-\alpha_2\right)E\left(\mu+\lambda V_t\right)^2}{1-\alpha_1 E\left(V_t\right)-\alpha_2} - E\left(\mu+\lambda V_t\right)^2\right)\frac{var(V_t)}{E\left(V_t\right)} + \lambda^3\kappa_3^{V_t}$$

Proof of unconditional third cumulant. The proof follows directly from proposition 3.1 together with theorem 3.1. ■

Remark 3.2 *From theorem 3.1 and corollary 3.1 we can see that if the unconditional second cumulant exists the unconditional third cumulant also exists. There are not added any existence conditions when moving from the second to the third cumulant.*

Theorem 3.2 (unconditional fourth cumulant) *The unconditional fourth cumulant of the MVM GARCH(1,1) process is*

$$\kappa_4^{Y_t} = \lambda^4\kappa_4^{V_t} + 6\lambda^2\kappa_3^{V_t}E\left(\sigma_t^2\right) + 3\left(\kappa_2^{V_t} + \left(\kappa_1^{V_t}\right)^2\right)E(\sigma_t^4)$$

where the expectations are given by:

$$E\left(\sigma_t^2\right) = \left(\tfrac{\alpha_0 E(V_t)+(1-\alpha_2)E(\mu+\lambda V_t)^2}{1-\alpha_1 E(V_t)-\alpha_2} - E\left(\mu+\lambda V_t\right)^2\right)\tfrac{1}{E(V_t)}$$

$$E\left(\sigma_t^4\right) = \frac{\alpha_0^2+2\alpha_0\alpha_1\frac{\alpha_0 E(V_t)+(1-\alpha_2)E(\mu+\lambda V_t)^2}{(1-\alpha_1 E(V_t)-\alpha_2)}+2\left(\alpha_0\alpha_2+\alpha_1\alpha_2 E(\mu+\lambda V_t)^2\right)E\left(\sigma_t^2\right)+\alpha_1^2\varkappa}{1-\left(2\alpha_1\alpha_2 E(V_t)+\alpha_2^2\right)-3\alpha_1^2 E\left(V_t^2\right)}$$

$$\varkappa = E\left(\mu+\lambda V_t\right)^4 + 6(\mu^2 E(V_t^2) + 2\mu\lambda E(V_t^3) + \lambda^2 E(V_t^4))E\left(\sigma_t^2\right)$$

Proof: see Appendix C

Remark 3.3 (existence criteria fourth cumulant) *From theorem 3.2 we can conclude that the fourth cumulant for the MVM GRACH (1,1) process exits if*

$$\left(2\alpha_1\alpha_2 E\left(V_t\right) + \alpha_2^2\right) + 3\alpha_1^2 E\left(V_t^2\right) > 1$$

4 An application with V_t assumed to follow an inverse Gaussian probability measure.

In this section we present a formulation of the $MVM\ GARCH(1,1)$ process under distributional assumptions that imply closed form expression of the conditional density functions. This opens the path for formulating a maximum likelihood scheme as the one presented in Bollerslev (1986), of course we are being addressed to use some kind of numerical procedure in order to achieve maximum likelihood estimates. Here we assume that the V_t is an i.i.d sequence of inverse Gaussian (IG) random variables. This makes it possible to define the mean variance mixing inverse Gaussian GARCH or MVM $IG(\delta, \gamma)$ GARCH(1,1). There are two main reasons for choosing the inverse Gaussian probability measure. First it is the fact that the resulting real valued probability law results in a closed form conditional density function. The second reason is that for this probability measure (real valued) all moments exits in contrast to for instance the student t probability measure. This together with the fact that earlier work suggests the inverse Gaussian as an appropriate mixing probability measure for financial modeling. See for instance Jensen and Lunde (2001), Andersson (2001) and Forsberg and Bollerslev (2002).

Definition 4.1 (The MVM IG GARCH (1,1) process) *The MVM IG*(δ, γ) *GARCH(1,1) stochastic process is defined by:*

$$Y_t = \mu + \lambda V_t + \sigma_t \sqrt{V_t} Z_t \tag{4.6}$$

where $\mathcal{L}\ (V_t) = IG(\gamma, \delta)$ *and* $\mathcal{L}\,(Z_t) = N(0,1)$ $\mu, \lambda \in \mathbb{R}$ V_t *and* Z_t *are sequences of i.i.d random variables.*

The GARCH equation is given by:

$$\sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2 + \alpha_2 \sigma_{t-1}^2 \tag{4.7}$$

where α_0, α_1 *and* $\alpha_2 \in \mathbb{R}^+$

In order to fully characterize the stochastic process we need some knowledge of the inverse Gaussian probability law. The density and the Laplace transform of this law an be found below.

Definition 4.2 (Inverse Gaussian probability law) *A stochastic variable is said to be distributed as a inverse Gaussian stochastic variable if its probability measure can written as:*

$$f(v;\delta,\gamma) = \tfrac{\delta e^{\delta\gamma}}{\sqrt{2\pi}} v^{-\frac{3}{2}} \exp\left(-\tfrac{1}{2}\left(\delta^2 v^{-1} + \gamma^2 v\right)\right)$$

where $v, \delta, \gamma \in \mathbb{R}^+$

The Laplace transform of the above density function:

$$\varphi(s;\delta,\gamma) = \exp(\delta\left(\gamma - \sqrt{(\gamma^2 - 2s)}\right))$$

For more details concerning this probability law see Sheshardi (1993)

Below a simulation of an MVM IG process can be found. The skewness of the process is easily seen since the downwards spikes have larger magnitude than the upwards spikes just as stock return data sometimes behaves.

[Insert figure 1 somewhere here]

4.1 Formulation of a likelihood for the $MVM\ IG(\delta,\gamma)\ GARCH(1,1)$ process

In this section we intend to sketch a scheme to estimate the $MVM\ IG(\delta,\gamma)\ GARCH(1,1)$ process. The main objective when it comes to formulate a likelihood estimation is to derive the unconditional (w.r.t to the mixing variable here $IG(\delta,\gamma)$) probability law. This is feasible when the mixing variable is an inverse Gaussian variable. This is of course not true in general since it implies restrictions on the density function of the mixing variable. It is for instance not possible to achieve this kind of closed form expression when the mixing variable is a log-normal random variable.

Proposition 4.1 (log likelihood of T observations)

$$L(\lambda, \sigma_t^2, \delta, \gamma, \mu; y_t) = T\{\ln(\delta) + \delta\gamma\} + \frac{1}{2}\sum_{t=1}^{T}\{\ln[(\gamma^2 + \frac{\lambda^2}{\sigma_t^2})/(\delta^2 + \frac{(x_t-\mu)^2}{\sigma_t^2})] - \ln\sigma_t^2\}$$
$$+ \sum_{t=1}^{T}\{\ln K_1\left(\sqrt{(\gamma^2 + \frac{\lambda^2}{\sigma_t^2})(\delta^2 + \frac{(x_t-\mu)^2}{\sigma_t^2})}\right) + \frac{\lambda(x_t-\mu)}{\sigma_t^2}\}$$

where $x, \mu, \lambda \in \mathbb{R}$ *and* $, \delta, \gamma \in \mathbb{R}^+$ $\sigma_t^2 = \alpha_0 + \alpha_1 y_{t-1}^2 + \alpha_2\sigma_{t-1}^2$ *where* α_0, α_1 *and* $\alpha_2 \in \mathbb{R}^+$
$K_1(.)$ *denotes the modified Bessel function of third order and index one*

Proof of log likelihood. A proof of the unconditional density expression can be found in Eriksson and Forsberg (2004). This in combination with standard operations with the natural logarithm gives the proof of the log likelihood. ∎

4.2 The Data

The data used in the maximum likelihood estimation is log return data from the Standard and Poor 500 index. The data is from two periods in time. The first data set is from the period 1 JAN 1987 to the 1 JAN 1990 and the second data set is from 1 JAN 1997 to 1 JAN 2000 obtained from the ECOWIN data base. Descriptive statistics for these two time series can be viewed below. The 87 to 90 data set includes the infamous crash observation of 19th of October. Observe that we have no illusion to be able to model this kind extreme observation. However, it is interesting to see how the process behaves in such an extreme situation.

Table 4:

JAN 87- JAN 90	
Mean	0.000
Median	0.000
Maximum	0.0910
Minimum	-0.229
Standard deviation	0.0149
Skewness	-5.402
Excess kurtosis	83.037
Kolmogorov-Smirnov	0.481
Probability	0.000
Observations	714

Table 5:

JAN 97- JAN 00	
Mean	0.000
Median	0.000
Maximum	0.0499
Minimum	-0.0711
Standard deviation	0.0120
Skewness	-0.552
Excess kurtosis	3.997
Kolmogorov-Smirnov	0.482
Probability	0.000
Observations	708

4.3 Estimation

The process that we choose to estimate is the $MVM\ IG(\delta, 1)$ GARCH(1,1) process. The covariance matrix for the estimated parameters is obtained using that $T^{-1}(\sum_{t=1}^{T} \frac{\partial L_t}{\partial \theta} \frac{\partial L_t}{\partial \theta'})^{-1}$ is a consistent estimate of the asymptotic covariance matrix. This is done using the analytical gradient. Further details regarding the estimation procedure can be found in appendix D. The results of the estimation can be examined in the tables below. It is important to notice that this empirical investigation serves as an example of how the process may behave estimated on stock returns. A more extensive investigation focused on the empirical implications is to be regarded as further work.

Table 6: ML Estimates
JAN 87- JAN 90

Parameter	Estimate	Standard error
$\hat{\alpha_0}$	$6.89 \cdot 10^{-6}$	$7.99 \cdot 10^{-3}$
$\hat{\alpha_1}$	0.0715	$9.98 \cdot 10^{-7}$
$\hat{\alpha_2}$	0.898	$1.48 \cdot 10^{-4}$
$\hat{\lambda}$	-0.00165	$1.99 \cdot 10^{-4}$
$\hat{\delta}$	0.738	$9.03 \cdot 10^{-5}$
$\hat{\mu}$	0.00191	$1.04 \cdot 10^{-4}$
$-L$	-3098.42	

Table 7: ML Estimates
JAN 97- JAN 00

Parameter	Estimate	Standard error
$\hat{\alpha_0}$	$1.32 \cdot 10^{-5}$	$4.90 \cdot 10^{-4}$
$\hat{\alpha_1}$	0.0911	$9.77 \cdot 10^{-8}$
$\hat{\alpha_2}$	0.875	$7.88 \cdot 10^{-6}$
$\hat{\lambda}$	-0.00238	$1.28 \cdot 10^{-5}$
$\hat{\delta}$	0.788	$2.98 \cdot 10^{-6}$
$\hat{\mu}$	0.00217	$7.81 \cdot 10^{-6}$
$-L$	-2967.61	

In order to comment the estimation results in the context of skewness and excess we state a lemma describing the coefficient of conditional skewness denoted as $S(Y_t|\mathcal{F}_{t-1})$ and ditto for excess kurtosis, $K(Y_t|\mathcal{F}_{t-1})$.

Lemma 4.1 (conditional skewness and excess kurtosis) *The coefficient of conditional skewness for the for the $MVM\ IG(\delta, 1)GARCH(1, 1)$ process is given by the following expression:*

$$S(Y_t|\mathcal{F}_{t-1}) = (\phi_1 + \phi_2\sigma_t^2)/(SE(Y_t|\mathcal{F}_{t-1}))^3$$

where $\phi_1 = \lambda^3\delta$ and $\phi_2 = 3\lambda\delta$
The corresponding expression for excess kurtosis is given by:

$$K(Y_t|\mathcal{F}_{t-1}) = (\varpi_1 + \varpi_2\sigma_t^2 + \varpi_3\sigma_t^4)/(VAR(Y_t|\mathcal{F}_{t-1}))^2$$

where $\varpi_1 = 15\lambda^4\delta$, $\varpi_2 = 18\lambda^2\delta$ *and* $\varpi_3 = 3\delta(1+\delta)$

Proof of conditional skewness and excess kurtosis. The proof follows directly from proposition 3.1 together with the Laplace transform in definition 4.2 and the definition of the coefficient of skewness and kurtosis see page 85 in Kendall and Stuart (1952). ∎

From the above lemma we can conclude that the conditional skewness is a linear function of the σ_t^2 parameter in the $MVM\ IG(\delta, 1)\ GARCH(1, 1)$ process, standardized with the cube of the conditional standard error. Further we see that the conditional excess kurtosis is a quadratic function in the σ_t^2 parameter, this time standardized with the square of the variance. A table with the calculated parameters and the corresponding standard error for the conditional skewness and kurtosis is presented below The estimates and standard error for the conditional moment parameters are obtained using Taylor expansion, see for instance page 353 ff in Cramér (1945)

Table 8: Conditional moment parameters
JAN 87- JAN 90

Parameter	Estimate	Standard error
$\hat{\phi}_1$	$-3.32 \cdot 10^{-9*}$	$1.20 \cdot 10^{-9}$
$\hat{\phi}_2$	$-3.67 \cdot 10^{-3*}$	$4.41 \cdot 10^{-4}$
$\hat{\varpi}_1$	$8.23 \cdot 10^{-11*}$	$3.97 \cdot 10^{-11}$
$\hat{\varpi}_2$	$3.62 \cdot 10^{-5}$	$8.86 \cdot 10^{-4}$
$\hat{\varpi}_3$	3.85^*	$6.71 \cdot 10^{-4}$

Table 9: Conditional moment parameters
JAN 97- JAN 00

Parameter	Estimate	Standard error
$\hat{\phi}_1$	$-1.07 \cdot 10^{-8*}$	$4.55 \cdot 10^{-9}$
$\hat{\phi}_2$	$-5.63 \cdot 10^{-3*}$	$8.02 \cdot 10^{-4}$
$\hat{\varpi}_1$	$3.81 \cdot 10^{-10}$	$2.17 \cdot 10^{-10}$
$\hat{\varpi}_2$	$8.05 \cdot 10^{-5}$	$8.88 \cdot 10^{-4}$
$\hat{\varpi}_3$	4.23^*	$6.13 \cdot 10^{-4}$

* Denotes results significant different from zero on the five percent level, obtained using asymptotic standard errors

The results for the conditional kurtosis imply that only the square of the GARCH variance equation helps to explain the time dependence. That only the σ_t^4 parameter is significant should not be a total surprise since σ_t^4 contains the same information as σ_t^2. Although there is nothing in general that speaks against a specification were the conditional kurtosis can have a terms linear in σ_t^2 . Further are both the intercept parameter and the parameter in front of the GARCH variance equation contributing to the conditional skewness. The intercept could indicate that there is some kind of default conditional skewness. The practical impact of the default conditional skewness. depends on the magnitude of σ_t^2. If σ_t^2 is sufficiently big the ϕ_1 parameter can for practical purposes be neglected. That is to say that the default skewness plays its biggest role in periods of low volatility.

With the estimation results from table 6 and 7 we can compute the unconditional cumulants implied by the estimated parameters. This is done using the Laplace transform in definition 4.2 together with theorem 3.1, 3.2 and corollary 3.1, We obtain a table for the unconditional cumulants implied by the estimation results. This table is presented below.

Table 10: Unconditional moments

	JAN 87- JAN 90	JAN 97- JAN 00
$\kappa_2^{Y_t}$	$1.042 \cdot 10^{-3}$	$2.58 \cdot 10^{-4}$
$S(Y_t)$	-0.154	-0.498
$K(Y_t)$	8.520	11.262

$S(Y_t)$ and $K(Y_t)$ denotes the coefficient of skewness and excess kurtosis. The unconditional moments imply that the estimated process has finite moments, at least up to the fourth. We can see that for the 97-00 period, the obtained unconditional skewness coincide with the empirical skewness . This result tells us that the conditional skewness is implied by a process which have fairly the same and skewness as the empirical ditto.

5 Possible extensions

In this section we present some interesting possible extensions of the MVM GARCH (1,1) process.

1. The first extensions is to develop methods to deal with the issue of temporal aggregation along the same lines as with a symmetrical GARCH (1,1) process (see Drost and Nijman (1993) and Drost and Werker (1996)). This could have implications in risk management issues such as value at risk and expected shortfall.

2. The other extension is to address the risk premium issue in the same manner as Engle, Lilien, and Robins (1987). That is specifying a more flexible version of the ARCH-M process. In a more formal way:

   Definition 5.1 (The MVM GARCH (1,1) in mean process) *The MVM GARCH (1,1) in mean stochastic process is defined by:*

$$Y_t = \underbrace{\mu + \lambda_1 V_t + \lambda_2 f(\sigma_t^2)}_{Riskpremium} + \sigma_t \sqrt{V_t} Z_t$$

where $\mathcal{L}\ (V_t)$ *is a probability measure with the domain on* $\mathbb{R}^+$ *with finite moments* $\mathcal{L}\,(Z_t) = N(0,1)$ $\lambda_1, \lambda_2, \mu \in \mathbb{R}$ *Both* V_t *and* Z_t *are i.i.d sequences of random variables.* $f(.)$ *is arbitrary function. The GARCH equation is given by:*

$$\sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2 + \alpha_2 \sigma_{t-1}^2$$

where α_0, α_1 *and* $\alpha_2 \in \mathbb{R}^+$

One thing that this process implies is the possibility to test the impact to the mean of the GARCH variance equation versus the stochastic mean dynamics. However this process requires further work since we meet new problems of both mathematical and statistical nature as well from the perspective of finance theory.

Both of the above possible extensions is to be regarded as work in progress.

6 Some concluding remarks

Here we have presented a new discrete time stochastic process which exhibits skewness and models the time dependence in the variance (and higher moments) according to a GARCH (1,1) scheme. We motivated this specification by doing an empirical investigation of powers of return data. The conjecture is that time dependence in higher moments can be modeled as a function of the time dependence in the variance. Further we derive some theoretical results concerning the suggested process, we derived the unconditional moments of the process. We also presented a specification of the process with the possibility to use maximum likelihood estimations methods. This method is then used in order to estimate the process on stock return data. The results is analyzed in the context of conditional skewness and kurtosis. We also present some of the possible extensions from the suggested process.

References

Andersson, J. (2001): "On the Normal Inverse Gaussian Stochastic Volatility Model," *Journal of Business and Economic Statistics*, 19, 44–54.

Badrinith, S., and S. Chatterjee (1988): "On Measuring Skewness and Elongation in Common Stock Return Distributions: The case of the Market Index," *Journal of Business*, 61(4), 451–472.

Barndorff-Nielsen, O., and P. Blaesild (1981): "Hyperbolic Distributions and Ramifications: Contribution to Theory and Application," in *Statistical Distributions in Scientific Work*, ed. by G. P. C. Taillie, and B. Baldessari, pp. 19–44, Reidel. Dordrecht.

Barndorff-Nielsen, O. E. (1997): "Normal Inverse Gaussian Distributions and Stochastic Volatility Modelling," *Scandinavian Journal of Statistics*, 24, 1–13.

Billingsley, P. (1995): *Probability and Measure*. John Wiley and sons, New York.

Bollerslev, T. (1986): "Generalized Autoregressive Conditional Heteroscedasticity," *Journal of Econometrics*, 31, 307–327.

——— (1987): "A Conditionally Heteroscedastic Time Series Model For Speculative Prices and Rates of Return," *Review of Economics and Statistics*, 69(3), 542–546.

Charlier, C. V. (1905): "Uber Die Darstellung Willkurlicher Funktionen," *Arkiv fur Matematik, Astronomi och Fysik*, 9(20).

Cramér, H. (1945): *Mathematical methods of Statistics*. Almqvist & Wiksell, Stockholm.

Draper, N., and D. Tierny (1972): "Regions of Positive and Unimodal Series Expansion of the Edgeworth and Gram-Charlier Approximations," *Biometrika*, 59(2).

Drost, F. C., and T. E. Nijman (1993): "Temporal Aggregation of GARCH Processes," *Econometrica*, 61(4), 909–927.

Drost, F. C., and B. J. Werker (1996): "Closing the Gap: Continuous Time GARCH Modeling," *Journal of Econometrics*, 74, 31–57.

Engle, R., and T. Bollerslev (1986): "Modelling the Persistence in Conditional Variances," *Econometric Reviews*, 5, 81–87.

ENGLE, R. F. (1982): "Autoregressive Conditional Hetroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50(4), 987–1007.

ENGLE, R. F., D. M. LILIEN, AND R. P. ROBINS (1987): "Estimating Time-Varying Risk Premia in the Term Structure the ARCH-M Model," *Econometrica*, 55, 391–407.

ERIKSSON, A., AND L. FORSBERG (2004): "An Extension of the Generalized Hyperbolic Family of Probability Laws," *Working paper, Uppsala University.*

FORSBERG, L., AND T. BOLLERSLEV (2002): "Bridging the gap between the distribution of realized (ECU) volatility and ARCH modelling (of the Euro): the GARCH-NIG model," *Journal of Applied Econometrics*, 17, 535–548.

GLOSTEN, L. R., R. JAGANATHAN, AND D. E. RUNKLE (1993): "On the Relation Between the Expected Value and the Volatility of Nominal Excess Return on Stocks," *Journal of Finance*, 48, 1779–1802.

HANSEN, B. E. (1994): "Autoregressive Conditional Density Estimation," *International Economic Review*, 35, 705–730.

HARVEY, C. R., AND A. SIDDIQUE (1999): "Autoregressive Conditional Skewness," *Journal of Financial and Quantitative Analysis*, 34(3), 465–487.

——— (2000a): "Conditional Skewness in Assets Pricing Tests," *Journal of Finance*, 55(3), 1263–1295.

——— (2000b): "Conditional Skewness in Assets Pricing Tests," *Research in Banking and Finance*, 1, 25–58.

JENSEN, M., AND A. LUNDE (2001): "The NIG ARCH Model a Fat-Tailed, Stochastic and Autoregressive Conditional Heteroscedastic Volatility Model," *Econometrics Journal*, 5, 319–342.

KARATZAS, I., AND S. E. SHREVE (1991): *Brownian Motion and Stochastic Calculus.* Sprnger Verlag, New York.

KENDALL, M. G., AND A. STUART (1952): *The Advanced Theory of Statistics Vol.1.* Griffin, London.

KRAUS, A., AND R. H. LITZBERGER (1976): "Skewness preference and the valuation of risk assets," *Journal of Finance*, 31(4), 1085–1100.

LEE, T. K., AND Y. TSE (1991): "Term Structure of Interest rates in the Singapore Asian Dollar market," *Journal of Applied Econometrics*, 6, 143–152.

NELSON, D. B. (1991): "Conditional Heteroscedasticity in Asset Returns: A New Approach," *Econometrica*, 64(2), 347–370.

PEIRÓ, A. (1999): "Skewness in finacial returns," *Journal of Banking and Finance*, 23, 847–862.

SHESHARDI, V. (1993): *The Inverse Gaussian Distribution - a Case Study in Exponential Families*. Oxford university press, Oxford.

SIMKOWITZ, M. A., AND W. L. BEEDLES (1980): "Assymetric Stable Distributed Security Returns," *Journal of American Statistical Association*, 75(370), 306–312.

ZAKOIAN, J.-M. (1994): "Threshold Heteroskedastic Models," *Journal of Economic Dynamics and Control*, 18, 931–955.

Appendix

A Proof of proposition 3.1

Proof. *i)* $\kappa_1^{Y_t}|\mathcal{F}_{t-1}$

$$\begin{aligned}
\kappa_1^{Y_t}|\mathcal{F}_{t-1} &= E\left[Y_t\right]|\mathcal{F}_{t-1} \\
&= \mu + E\left(\lambda V_t|\mathcal{F}_{t-1}\right) + E\left[\sigma_t\sqrt{V_t}Z_t\right]|\mathcal{F}_{t-1} \\
&= \mu + E\left[\lambda V_t\right]|\mathcal{F}_{t-1} \\
&= \mu + \lambda E\left(V_t\right) \text{ (i..i.d assumption } V_t) \\
&= \mu + \lambda \kappa_1^{V_t}
\end{aligned}$$

ii) $\kappa_2^{Y_t}|\mathcal{F}_{t-1}$

$$\begin{aligned}
\kappa_2^{Y_t}|\mathcal{F}_{t-1} &= E\left[Y_t - \kappa_1^{Y_t}\right]^2|\mathcal{F}_{t-1} \\
&= E\left[\left(\lambda V_t - \lambda\kappa_1^{V_t}\right)^2 + 2\left(\lambda V_t - \lambda\kappa_1^{V_t}\right)\sigma_t\sqrt{V_t}Z_t + \sigma_t^2 V_t Z_t^2\right]|\mathcal{F}_{t-1} \\
&= E\left[\lambda^2\left(V_t - \kappa_1^{V_t}\right)^2 + \sigma_t^2 V_t Z_t^2\right]|\mathcal{F}_{t-1} \text{ (i.i.d assumption } Z_t \text{ and } V_t) \\
&= E\left[\sigma_t^2|\mathcal{F}_{t-1}\right]\kappa_1^{V_t} + \lambda^2\kappa_2^{V_t} \text{ (i.i.d assumption } Z_t \text{ and } V_t)
\end{aligned}$$

iii) $\kappa_3^{Y_t}|\mathcal{F}_{t-1}$

$$\begin{aligned}
\kappa_3^{Y_t}|\mathcal{F}_{t-1} &= E\left(\left[Y_t - \kappa_1^{Y_t}\right]^3|\mathcal{F}_{t-1}\right) \\
&= E\left[\left(\lambda V_t - \lambda\kappa_1^{V_t}\right)^3 + 3\left(\lambda V_t - \lambda\kappa_1^{V_t}\right)^2\sigma_t\sqrt{V_t}Z_t + 3\left(\lambda V_t - \lambda\kappa_1^{V_t}\right)\sigma_t^2 V_t Z_t^2 + \left(\sigma_t\sqrt{V_t}Z_t\right)^3\right]|\mathcal{F}_{t-1} \\
&= E\left[\left(\lambda V_t - \lambda\kappa_1^{V_t}\right)^3 + 3\left(\lambda V_t - \lambda\kappa_1^{V_t}\right)\sigma_t^2 V_t Z_t^2\right]|\mathcal{F}_{t-1} \quad \text{(i.i.d assumption } Z_t \text{ and } V_t) \\
&= E\left[\lambda^3\left(V_t - E\left(V_t\right)\right)^3 + 3\lambda\sigma_t^2\left(V_t^2 - V_t E\left(V_t\right)\right)\right]|\mathcal{F}_{t-1} \\
&= 3\lambda\kappa_2^{V_t}E\left[\sigma_t^2|\mathcal{F}_{t-1}\right] + \lambda^3\kappa_3^{V_t}
\end{aligned}$$

$iv)\ \kappa_4^{Y_t}|\mathcal{F}_{t-1}$

$$\begin{aligned}
\kappa_4^{Y_t}|\mathcal{F}_{t-1} &= E\left(\left[\left(Y_t - \kappa_1^{Y_t}\right)^4 - 3\left(\kappa_2^{Y_t}\right)^2\right]|\mathcal{F}_{t-1}\right) \\
&= E\left[\left(\lambda V_t - \lambda\kappa_1^{V_t}\right)^4 + 6\left(\lambda V_t - \lambda\kappa_1^{V_t}\right)^2\sigma_t^2 V_t Z_t^2 + \sigma_t^4 V_t^2 Z_t^4 - 3\left(\kappa_2^{Y_t}\right)^2\right]|\mathcal{F}_{t-1} \quad \text{(i.i.d assumption)} \\
&= \lambda^4\kappa_4^{V_t} + 6\lambda^2\kappa_3^{V_t} E\left[\sigma_t^2|\mathcal{F}_{t-1}\right] + 3\left(\kappa_2^{V_t} + \left(\kappa_1^{V_t}\right)^2\right) E\left[\sigma_t^4|\mathcal{F}_{t-1}\right]
\end{aligned}$$

■

B Proof of theorem 3.1

Proof. Assume the following stochastic process is stationary.

$$Y_t = (\mu + \lambda V_t) + \sigma_t\sqrt{V_t}Z_t \tag{B.8}$$

$\mu, \lambda \in \mathbb{R}$, $\sigma_t \in \mathbb{R}^+$, $\mathcal{L}\ (V_t) = D^+\ \mathcal{L}\,(Z_t) = N(0,1)$ both V_t and Z_t are assumed to be *i.i.d.* Further D^+ denotes a probability measure defined on $\mathbb{R}^+$ with finite moments.

The GARCH equation is given by:

$$\sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2 + \alpha_2\sigma_{t-1}^2$$

The second conditional cumulant is denoted as:

$$\kappa_2^{Y_t}|\mathcal{F}_{t-1} = E\left(\sigma_t^2|\mathcal{F}_{t-1}\right)\kappa_1^{V_t} + \lambda^2\kappa_2^{V_t} \tag{B.9}$$

and its unconditional counterpart is denoted as:

$$\kappa_2^{Y_t} = E\left(\sigma_t^2\right)\kappa_1^{V_t} + \lambda^2\kappa_2^{V_t} \tag{B.10}$$

i) Determine the relation between $E\left(Y_t^2\right)$ and $E\left(\sigma_t^2\right)$.

$$
\begin{aligned}
E\left(Y_t^2\right) &= E\left(\mu+\lambda V_t\right)^2 + E\left(\sigma_t^2\right)E\left(V_t\right) \\
&\Leftrightarrow \\
\frac{E\left(Y_t^2\right)-E\left(\mu+\lambda V_t\right)^2}{E\left(V_t\right)} &= E\left(\sigma_t^2\right)
\end{aligned}
\tag{B.11}
$$

ii) Determine the expression for $E\left(\sigma_t^2\right)$.

$$
E\left(Y_t^2\right)-E\left(\mu+\lambda V_t\right)^2=\alpha_0 E\left(V_t\right)+\alpha_1 E\left(V_t\right)E\left(Y_{t-1}^2\right)+\alpha_2\left(E\left(Y_{t-1}^2\right)-E\left(\mu+\lambda V_t\right)^2\right)
$$

$$
E\left(Y_t^2\right)=\frac{\alpha_0 E\left(V_t\right)+\left(1-\alpha_2\right)E\left(\mu+\lambda V_t\right)^2}{\left(1-\alpha_1 E\left(V_t\right)-\alpha_2\right)}
\tag{B.12}
$$

Which implies the following expression for $E\left(\sigma_t^2\right)$

$$
E\left(\sigma_t^2\right)=\frac{\frac{\alpha_0 E\left(V_t\right)+\left(1-\alpha_2\right)E\left(\mu+\lambda V_t\right)^2}{\left(1-\alpha_1 E\left(V_t\right)-\alpha_2\right)}-E\left(\mu+\lambda V_t\right)^2}{E\left(V_t\right)}
\tag{B.13}
$$

iii) Use equation B.10 together with B.13 this yields:

$$
\begin{aligned}
\kappa_2^{Y_t} &= E\left(\sigma_t^2\right)\kappa_1^{V_t}+\lambda^2\kappa_2^{V_t} \\
&= \left(\frac{\alpha_0 E\left(V_t\right)+\left(1-\alpha_2\right)E\left(\mu+\lambda V_t\right)^2}{1-\alpha_1 E\left(V_t\right)-\alpha_2}-E\left(\mu+\lambda V_t\right)^2\right)+\lambda^2 var(V_t)
\end{aligned}
$$

■

C Proof of theorem 3.2

Proof. Just like in the proof of the unconditional second cumulant we have stationary process:

$$Y_t = (\mu + \lambda V_t) + \sigma_t \sqrt{V_t} Z_t \tag{C.14}$$

$\mu, \lambda \in \mathbb{R}$, $\sigma_t \in \mathbb{R}^+$, $\mathcal{L}\,(V_t) = D$ $\mathcal{L}\,(Z_t) = N(0,1)$ both $i.i.d$ V_t and Z_t.

where D denotes probability measure defined on $\mathbb{R}^+$ with finite moments.

The GARCH equation is given by:

$$\sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2 + \alpha_2 \sigma_{t-1}^2 \tag{C.15}$$

The fourth cumulant is denoted as:

$$\kappa_4^{Y_t} | \mathcal{F}_{t-1} = \lambda^4 \kappa_4^{V_t} + 6\lambda^2 \kappa_3^{V_t} E\left[\sigma_t^2\right] + 3E\left[\sigma_t^4 | \mathcal{F}_{t-1}\right] \kappa_2^{V_t} \tag{C.16}$$

and its unconditional counterpart is denoted as

$$\kappa_4^{Y_t} = \lambda^4 \kappa_4^{V_t} + 6\lambda^2 \kappa_3^{V_t} E\left(\sigma_t^2\right) + 3\left(\kappa_2^{V_t} + \left(\kappa_1^{V_t}\right)^2\right) E\left[\sigma_t^4\right] \tag{C.17}$$

i) determine the relation between $E\left(Y_t^4\right)$ and $E\left(\sigma_t^4\right)$.

$$\begin{aligned} E\left(Y_t^4\right) &= E(Z_t^4) E\left(V_t^2\right) E\left(\sigma_t^4\right) + E\left(\mu + \lambda V_t\right)^4 + 6E\left(V_t\left(\mu + \lambda V_t\right)\right)^2 E\left(\sigma_t^2\right) \\ &= 3E\left(V_t^2\right) E\left(\sigma_t^4\right) + E\left(\mu + \lambda V_t\right)^4 + 6E\left(V_t\left(\mu + \lambda V_t\right)\right)^2 E\left(\sigma_t^2\right) \end{aligned}$$

$$E\left(\sigma_t^4\right) = \frac{E\left(Y_t^4\right) - E\left(\mu + \lambda V_t\right)^4 - 6E\left(V_t\left(\mu + \lambda V_t\right)\right)^2 E\left(\sigma_t^2\right)}{3E\left(V_t^2\right)} = \frac{E\left(Y_t^4\right) - \varkappa}{3E\left(V_t^2\right)} \tag{C.18}$$

ii) Take the square of the GARCH equation $\left(\sigma_t^2\right)^2$ and determine a expression for $E\left(\sigma_t^4\right)$

$$\sigma_t^4 = \alpha_0^2 + 2\alpha_0\alpha_1 Y_{t-1}^2 + 2\alpha_0\alpha_2\sigma_{t-1}^2 + \alpha_1^2 Y_{t-1}^4 + 2\alpha_1\alpha_2\left(\sigma_{t-1}^2 Y_{t-1}^2\right) + \alpha_2^2\sigma_{t-1}^4$$

note that the following is true:

$$\sigma_{t-1}^2 Y_{t-1}^2 = \sigma_{t-1}^2 \left((\mu + \lambda V_{t-1}) + \sigma_{t-1} \sqrt{V_{t-1}} Z_{t-1} \right)^2$$
$$= (\mu + \lambda V_{t-1})^2 \sigma_{t-1}^2 + \sigma_{t-1}^4 V_{t-1} Z_{t-1}^2 + \sigma_{t-1}^2 (\mu + \lambda V_{t-1}) \sigma_{t-1} \sqrt{V_{t-1}} Z_{t-1}$$

This yields that $E(\sigma_t^4)$ can be written:

$$E\left(\sigma_t^4\right) = \alpha_0^2 + 2\alpha_0\alpha_1 E\left(Y_{t-1}^2\right) + \alpha_1^2 E\left(Y_{t-1}^4\right) + 2\left(\alpha_0\alpha_2 + \alpha_1\alpha_2(\mu + \lambda V_t)^2\right) E\left(\sigma_{t-1}^2\right)$$
$$+ E\left(\sigma_{t-1}^4\right)\left(2\alpha_1\alpha_2 E(V_t) + \alpha_2^2\right)$$

Substitute $E(\sigma_t^4)$ and $E\left(\sigma_{t-1}^4\right)$ with the expression in C.18 using the stationarity assumption concerning the process Y_t and solve for $E(Y_t^4)$. This procedure gives:

$$E\left(Y_t^4\right) = \frac{3E\left(V_t^2\right)\left(\alpha_0^2 + 2\alpha_0\alpha_1 E\left(Y_{t-1}^2\right) + 2\left(\alpha_0\alpha_2 + \alpha_1\alpha_2 E(\mu + \lambda V_t)^2\right) E\left(\sigma_{t-1}^2\right)\right)}{\left(1 - \left(2\alpha_1\alpha_2 E(V_t) + \alpha_2^2 + 3E\left(V_t^2\right)\alpha_1^2\right)\right)} \tag{C.19}$$
$$+ \varkappa \frac{\left(1 - \left(2\alpha_1\alpha_2 E(V_t) + \alpha_2^2\right)\right)}{\left(1 - \left(2\alpha_1\alpha_2 E(V_t) + \alpha_2^2 + 3E\left(V_t^2\right)\alpha_1^2\right)\right)}$$

Now use C.19 in C.18 to determine $E(\sigma_t^4)$.

$$E\left(\sigma_t^4\right) = \frac{\alpha_0^2 + 2\alpha_0\alpha_1 E\left(Y_{t-1}^2\right) + 2\left(\alpha_0\alpha_2 + \alpha_1\alpha_2 E(\mu + \lambda V_t)^2\right) E\left(\sigma_{t-1}^2\right) + \alpha_1^2 \varkappa}{\left(1 - \left(2\alpha_1\alpha_2 E(V_t) + \alpha_2^2 + 3E\left(V_t^2\right)\alpha_1^2\right)\right)}$$

Using expression C.17 together with the above expression and expression B.13 the unconditional fourth cumulant for the process can be written as:

$$\kappa_4^{Y_t} = \lambda^4 \kappa_4^{V_t} + 6\lambda^2 \kappa_3^{V_t} E\left(\sigma_t^2\right) + 3\left(\kappa_2^{V_t} + \left(\kappa_1^{V_t}\right)^2\right) E(\sigma_t^4)$$

where

$E(\sigma_t^2) = \left(\frac{\alpha_0 E(V_t) + (1-\alpha_2) E(\mu + \lambda V_t)^2}{1 - \alpha_1 E(V_t) - \alpha_2} - E(\mu + \lambda V_t)^2\right) \frac{1}{E(V_t)}$

$E(\sigma_t^4) = \frac{\alpha_0^2 + 2\alpha_0\alpha_1 \frac{\alpha_0 E(V_t) + (1-\alpha_2) E(\mu + \lambda V_t)^2}{(1 - \alpha_1 E(V_t) - \alpha_2)} + 2\left(\alpha_0\alpha_2 + \alpha_1\alpha_2 E(\mu + \lambda V_t)^2\right) E\left(\sigma_t^2\right) + \alpha_1^2 \varkappa}{1 - \left(2\alpha_1\alpha_2 E(V_t) + \alpha_2^2\right) - 3\alpha_1^2 E\left(V_t^2\right)}$

$\varkappa = E(\mu + \lambda V_t)^4 + 6(\mu^2 E(V_t^2) + 2\mu\lambda E(V_t^3) + \lambda^2 E(V_t^4)) E(\sigma_t^2)$ ■

D Estimation procedure

The MVM IG(δ,1) GARCH (1,1) process is estimated by solving a minimization problem with non-linear constraints, i.e that is minimizing minus the log likelihood. The non-linear constraints main purpose is to make sure that the optimization routine not diverge away into unrealistic estimations of the conditional mean and conditional variance.It also makes the possible parameter space significantly smaller. The optimization was obtained using Gauss- Newton line search methods in `fmincon` routine in the Matlab program package with analytical gradient and non linear constraints.

Definition D.1 (Outline of estimation procedure) *The estimation problem can be defined as follows:*

$$\min\{-L(\theta; y_t, T)\} \textit{ under the constraint that } \begin{cases} -0.1 \leq \mu + \delta\lambda \leq 0.1 \\ \quad 0 \leq \delta(\alpha_0 + \lambda^2) \leq 0.01 \\ \quad 0 \leq \delta\alpha_1 \leq 2 \\ \quad 0 \leq \delta\alpha_2 \leq 2 \end{cases}$$

where $\theta = (\alpha_0, \alpha_1, \alpha_2, \lambda, \delta, \mu)$ *and* $L(\theta; y_t, T)$ *is defined in proposition 4.1*

Lemma D.1 (Analytical gradient) *The analytical gradient was obtained using the formulas for differentiation for the log of modified Bessel functions which can be found in Barndorff-Nielsen and Blaesild (1981). In the expressions for the analytical gradient the modified Bessel function of third order and index one and zero is denoted as* $K_1(.)$ *and* $K_0(.)$*.*

$$\frac{\partial L}{\partial \alpha_0} = \sum_{t=1}^{T} \Bigg\{ \frac{1}{2} \left[\frac{\left(\frac{\lambda^2\delta^2 + (x_t-\mu)^2\gamma^2}{\sigma_t^4} + \frac{2(x_t-\mu)^2\lambda^2}{\sigma_t^6} \right)}{\left(\delta^2 + \frac{(x_t-\mu)^2}{\sigma_t^2}\right)\left(\gamma^2 + \frac{\lambda^2}{\sigma_t^2}\right)} + \frac{\left(\frac{\lambda^2\delta^2 + (x_t-\mu)^2\gamma^2}{\sigma_t^4} + \frac{2(x_t-\mu)^2\lambda^2}{\sigma_t^6} \right)}{\sqrt{\left(\delta^2 + \frac{(x_t-\mu)^2}{\sigma_t^2}\right)\left(\gamma^2 + \frac{\lambda^2}{\sigma_t^2}\right)}} \frac{K_0\left(\sqrt{\left(\delta^2 + \frac{(x_t-\mu)^2}{\sigma_t^2}\right)\left(\gamma^2 + \frac{\lambda^2}{\sigma_t^2}\right)}\right)}{K_1\left(\sqrt{\left(\delta^2 + \frac{(x_t-\mu)^2}{\sigma_t^2}\right)\left(\gamma^2 + \frac{\lambda^2}{\sigma_t^2}\right)}\right)} \right]$$

$$+ \frac{1}{\sigma_t^4} \frac{(x_t-\mu)^2\gamma^2 - \lambda^2\delta^2}{\left(\gamma^2 + \frac{\lambda^2}{\sigma_t^2}\right)\left(\delta^2 + \frac{(x_t-\mu)^2}{\sigma_t^2}\right)} + \frac{\lambda(\mu - x_t)}{\sigma_t^4} - \frac{1}{\sigma_t^2} \Bigg\}$$

$$\frac{\partial L}{\partial \alpha_1} =$$

$$\sum_{t=1}^{T}\Big\{\frac{x_{t-1}^2}{2}\left[\frac{\left(\frac{\lambda^2\delta^2+(x_t-\mu)^2\gamma^2}{\sigma_t^4}+\frac{2(x_t-\mu)^2\lambda^2}{\sigma_t^6}\right)}{\left(\delta^2+\frac{(x_t-\mu)^2}{\sigma_t^2}\right)\left(\gamma^2+\frac{\lambda^2}{\sigma_t^2}\right)}+\frac{\left(\frac{\lambda^2\delta^2+(x_t-\mu)^2\gamma^2}{\sigma_t^4}+\frac{2(x_t-\mu)^2\lambda^2}{\sigma_t^6}\right)}{\sqrt{\left(\delta^2+\frac{(x_t-\mu)^2}{\sigma_t^2}\right)\left(\gamma^2+\frac{\lambda^2}{\sigma_t^2}\right)}}\frac{K_0\left(\sqrt{\left(\delta^2+\frac{(x_t-\mu)^2}{\sigma_t^2}\right)\left(\gamma^2+\frac{\lambda^2}{\sigma_t^2}\right)}\right)}{K_1\left(\sqrt{\left(\delta^2+\frac{(x_t-\mu)^2}{\sigma_t^2}\right)\left(\gamma^2+\frac{\lambda^2}{\sigma_t^2}\right)}\right)}\right]$$

$$+\frac{x_{t-1}^2}{\sigma_t^4}\frac{(x_t-\mu)^2\gamma^2-\lambda^2\delta^2}{\left(\gamma^2+\frac{\lambda^2}{\sigma_t^2}\right)\left(\delta^2+\frac{(x_t-\mu)^2}{\sigma_t^2}\right)}+x_{t-1}^2\frac{\lambda(\mu-x_t)}{\sigma_t^4}-\frac{x_{t-1}^2}{\sigma_t^2}\Big\}$$

$$\frac{\partial L}{\partial \alpha_2} =$$

$$\sum_{t=1}^{T}\Big\{\frac{\sigma_{t-1}^2}{2}\left[\frac{\left(\frac{\lambda^2\delta^2+(x_t-\mu)^2\gamma^2}{\sigma_t^4}+\frac{2(x_t-\mu)^2\lambda^2}{\sigma_t^6}\right)}{\left(\delta^2+\frac{(x_t-\mu)^2}{\sigma_t^2}\right)\left(\gamma^2+\frac{\lambda^2}{\sigma_t^2}\right)}+\frac{\left(\frac{\lambda^2\delta^2+(x_t-\mu)^2\gamma^2}{\sigma_t^4}+\frac{2(x_t-\mu)^2\lambda^2}{\sigma_t^6}\right)}{\sqrt{\left(\delta^2+\frac{(x_t-\mu)^2}{\sigma_t^2}\right)\left(\gamma^2+\frac{\lambda^2}{\sigma_t^2}\right)}}\frac{K_0\left(\sqrt{\left(\delta^2+\frac{(x_t-\mu)^2}{\sigma_t^2}\right)\left(\gamma^2+\frac{\lambda^2}{\sigma_t^2}\right)}\right)}{K_1\left(\sqrt{\left(\delta^2+\frac{(x_t-\mu)^2}{\sigma_t^2}\right)\left(\gamma^2+\frac{\lambda^2}{\sigma_t^2}\right)}\right)}\right]$$

$$+\frac{\sigma_{t-1}^2}{\sigma_t^4}\frac{(x_t-\mu)^2\gamma^2-\lambda^2\delta^2}{\left(\gamma^2+\frac{\lambda^2}{\sigma_t^2}\right)\left(\delta^2+\frac{(x_t-\mu)^2}{\sigma_t^2}\right)}+\sigma_{t-1}^2\frac{\lambda(\mu-x_t)}{\sigma_t^4}-\frac{\sigma_{t-1}^2}{\sigma_t^2}\Big\}$$

$$\frac{\partial L}{\partial \delta} = \frac{T}{\delta}+T\gamma-\sum_{t=1}^{T}\Big\{\delta\left(\sqrt{\frac{\gamma^2+\frac{\lambda^2}{\sigma_t^2}}{\delta^2+\frac{(x_t-\mu)^2}{\sigma_t^2}}}\frac{K_0\left(\sqrt{\left(\delta^2+\frac{(x_t-\mu)^2}{\sigma_t^2}\right)\left(\gamma^2+\frac{\lambda^2}{\sigma_t^2}\right)}\right)}{K_1\left(\sqrt{\left(\delta^2+\frac{(x_t-\mu)^2}{\sigma_t^2}\right)\left(\gamma^2+\frac{\lambda^2}{\sigma_t^2}\right)}\right)}+\frac{2}{\delta^2+\frac{(x_t-\mu)^2}{\sigma_t^2}}\right)$$

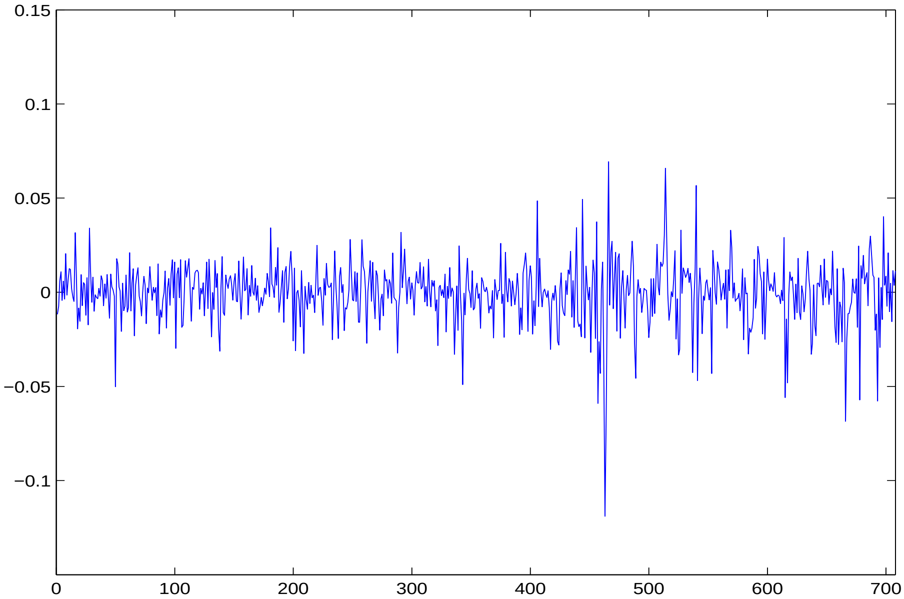
$$\frac{\partial L}{\partial \gamma} = T\delta-\sum_{t=1}^{T}\gamma\sqrt{\frac{\delta^2+\frac{(x_t-\mu)^2}{\sigma_t^2}}{\gamma^2+\frac{\lambda^2}{\sigma_t^2}}}\frac{K_0\left(\sqrt{\left(\gamma^2+\frac{\lambda^2}{\sigma_t^2}\right)\left(\delta^2+\frac{(x_t-\mu)^2}{\sigma_t^2}\right)}\right)}{K_1\left(\sqrt{\left(\gamma^2+\frac{\lambda^2}{\sigma_t^2}\right)\left(\delta^2+\frac{(x_t-\mu)^2}{\sigma_t^2}\right)}\right)}$$

$$\frac{\partial L}{\partial \lambda} = -\sum_{t=1}^{T}\Big\{\frac{\lambda}{\sigma_t^2}\frac{K_0\left(\sqrt{\left(\delta^2+\frac{(x_t-\mu)^2}{\sigma_t^2}\right)\left(\gamma^2+\frac{\lambda^2}{\sigma_t^2}\right)}\right)}{K_1\left(\sqrt{\left(\delta^2+\frac{(x_t-\mu)^2}{\sigma_t^2}\right)\left(\gamma^2+\frac{\lambda^2}{\sigma_t^2}\right)}\right)}\sqrt{\frac{\left(\delta^2+\frac{(x_t-\mu)^2}{\sigma_t^2}\right)}{\left(\gamma^2+\frac{\lambda^2}{\sigma_t^2}\right)}}+\frac{(x_t-\mu)}{\sigma_t^2}\Big\}$$

$$\frac{\partial L}{\partial \mu} = \sum_{t=1}^{T}\Big\{\frac{(x_t-\mu)}{\sigma_t^2}\left(\frac{K_0\left(\sqrt{\left(\delta^2+\frac{(x_t-\mu)^2}{\sigma_t^2}\right)\left(\gamma^2+\frac{\lambda^2}{\sigma_t^2}\right)}\right)}{K_1\left(\sqrt{\left(\delta^2+\frac{(x_t-\mu)^2}{\sigma_t^2}\right)\left(\gamma^2+\frac{\lambda^2}{\sigma_t^2}\right)}\right)}\sqrt{\frac{\left(\gamma^2+\frac{\lambda^2}{\sigma_t^2}\right)}{\left(\delta^2+\frac{(x_t-\mu)^2}{\sigma_t^2}\right)}}+\frac{2}{\left(\delta^2+\frac{(x_t-\mu)^2}{\sigma_t^2}\right)}\right)-\frac{\lambda}{\sigma_t^2}\Big\}$$

E Figures

Figure 1: Simulation of an MVM IG process



$MVM\ IG(0.9, 1)$ process with $\lambda = -0.0020$, $\mu = 0.0019$, $\alpha_0 = 2.0 \cdot 10^{-5}$, $\alpha_1 = 0.1$ and $\alpha_2 = 0.85$