Contaminated Variance-Mean Mixing Model

Thomas Fung[1,2,4], Joanna J.J. Wang[1] and Eugene Seneta[3]
[1] Department of Statistics, Macquarie University, Sydney, Australia
[2] Honorary Associate, University of Sydney, Sydney, Australia
[3] School of Mathematics and Statistics, University of Sydney, Sydney, Australia
[4] Corresponding author: Thomas Fung, e-mail: thomas.fung@mq.edu.au

Abstract

We consider the Generalised Normal Variance-Mean (GNVM) model in which the mixing random variable is Gamma distributed for financial return data. This model generalises the popular Variance-Gamma (VG) distribution. This GNVM model can be interpreted as the addition of noise to a (skew) VG base. In this presentation, we will not only discuss the parameter estimation of the general model, but also discuss how to utilise this noise contamination for desirable results. A simulation study will be used to illustrate the results.

Keywords: JAGS, Normal Variance-Mean mixture distribution, Maximum likelihood estimation, Variance Gamma distribution.

1 Introduction

Tjetjep and Seneta (2006) proposed the Generalised Normal Variance-Mean (GNVM) model that synthesises the concepts of several popular distributional models for financial returns. A random variable X_{GNVM} is said to follow a GNVM model if

$$X_{GNVM}|V \sim N(\mu + \theta V, \sigma^2 V + k) \tag{1}$$

where V is a positive random variable, $\mu, \theta \in R$; $\sigma, k \geq 0$ but not simultaneously equal to 0. When $\theta = 0$, the distribution of X is symmetric. When $k = 0$, (1) reduces to the Normal Variance-Mean (NVM) Model of Barndorff-Nielsen, Kent and Sørensen (1982):

$$X_{NVM}|V \sim N(\mu + \theta V, \sigma^2 V). \tag{2}$$

The (symmetric) Variance Gamma (VG) distribution, introduced by Madan and Seneta (1990) where $\theta = 0$, is a prime example of the NVM model with $V \sim \Gamma(\frac{1}{\nu}, \frac{1}{\nu})$, $\nu > 0$ i.e. V has a Gamma distribution with density

$$f(y) = \frac{1}{\nu^{\frac{1}{\nu}}\Gamma(\frac{1}{\nu})} y^{\frac{1}{\nu}-1} e^{-\frac{y}{\nu}}, \quad y > 0; \quad = 0, \quad \text{otherwise}; \tag{3}$$

and $EV = 1$. The distribution of X_{NVM}, as specified by (2) and (3) is thus often also called the (skew) VG distribution. In the sequel "NVM" and "GNVM" will relate to the case where $V \sim \Gamma(\frac{1}{\nu}, \frac{1}{\nu})$.

Rzetelski (2009) showed that if X_{GNVM} and X_{NVM} are two random variables that follow (1) and (2) respectively, then

$$X_{GNVM} \stackrel{d}{=} X_{NVM} + Z \tag{4}$$

where $Z \sim N(0, k)$ is independently distributed of X_{NVM}. The GNVM model can therefore be interpreted as a NVM model "contaminated" by a normal random variable as noise. The type of representation in (4) can be ascribed to Press (1967) with a compound Poisson distribution replacing X_{NVM}.

One of the foci of this paper is to use the mixing representation as in (1) to obtain maximum likelihood estimates for the GNVM model, making convenient use of the statistical package `JAGS`. The other focus is to demonstrate how one can use contamination in a positive light to make the NVM distribution more flexible. By using simulated data for (2), we will demonstrate that the tail weight of the fitted GNVM model can be manipulated to a certain extent without compromising the goodness of fit of the GNVM model to the data. "Contamination" can therefore be interpreted as a way to provide extra robustness to the NVM model via the GNVM.

2 Parameter Estimation for the GNVM model

`JAGS` is a programme for Bayesian estimation of parameters, where JAGS stands for **J**use **A**nother **G**ibbs **S**ampler (See Plummer (2003)). For a large size N of the iid sample $x_1, x_2, \ldots, x_N$, the Bayes estimators obtained are *asymptotically equivalent to maximum likelihood estimators.*

The advantage of `JAGS`, as with `WinBUGS`, is that it allows us to use a mixing representation to specify the model. Since the GNVM distribution does not possess a closed-formed density, `JAGS` enables us to obtain MLE without specifying the likelihood function explicitly, as we call on the known mixing distribution, which makes the input code straightforward.

We use the following input code of the GNVM model in `JAGS`:

```
for (i in 1:N) {
x[i] ~ dnorm(mu[i], tau[i])
mu[i] <- mu.c+theta*v[i]
tau[i] <- 1/sigma2[i]
sigma2[i] <- sigma*sigma*v[i] + k
v[i] ~ dgamma(u,u) }
mu.c ~ dunif(-10,10)
theta ~ dunif(-10,10)
sigma ~ dunif(0.0001,5)
k ~ dunif(0.0000001,1)
u ~ dunif(0.1,3.5)
nu <- 1/u
```

`JAGS` is cross-platform but `WinBUGS` is available in Windows only. Furthermore, we need the lower bound of the uniform prior for u to be quite small (as shown in the above input code) for our subsequent studies, `JAGS` seems to handle that just fine, but not `WinBUGS`. As a result, we decided to use `JAGS` instead of `WinBUGS` in this paper.

The actual data used in this section is the log daily closing price increment ("return") on Microsoft Corporation's share prices between 1 January 1996 and 21 December 2005. The data set, obtained from the wire service Bloomberg had been already cleaned for splits, has $N = 2518$ readings. We multiply the returns by 100 to improve the convergence of the Markov chains in `JAGS`. This makes the estimated μ, σ and θ a 100 times, and k 10,000 times the corresponding parameters of the original returns but the shape parameter ν remains unchanged. The NVM model (i.e. $k = 0$) was also fitted for comparison.

In `JAGS`, we asked the program to initialise the parameters and generate a single Markov chain. 50,000 and 25,000 iterations are then run for the GNVM and NVM models respectively of which the first 5,000 are treated as burn-in. This implies that we had 45,000 and 20,000 stable-state values to obtain the Bayes (equivalently maximum likelihood because of the large sample size) estimators for the GNVM and NVM models.

Assessing the data fitting capability of the two models is our secondary aim, so Deviance Information Criterion (DIC) is used to access and compare the fit. The DIC is defined as

$$DIC = 2\overline{D(\boldsymbol{\theta})} - D(\bar{\boldsymbol{\theta}}) \tag{5}$$

where $\overline{D(\boldsymbol{\theta})}$ is the posterior mean deviance and $D(\bar{\theta})$ is the deviance evaluated at the estimate of $\boldsymbol{\theta} = (\mu, \sigma, \theta, \nu, k)^T$, where deviance is defined as: -2 times the log likelihood. The model which returns the smaller DIC by at least 5 units will be said to provide a better fit, otherwise the fit between the two models will be considered as similar. The values of DIC we provide here are calculated from first principles: using (5), numerical integration was used to evaluate the density of the GNVM model for the calculation of deviance. We used a thinning of 15 and 10 on the posterior samples of the GNVM and NVM models respectively to speed up the DIC calculation.

The results of the estimated parameters and their corresponding DICs are collected in Table 1. By comparing the results in Table 1, we conclude that the fit of the GNVM model is no better than the NVM model in the Microsoft data set. This demonstrates that using k as an extra parameter will not always be beneficial. In such situations, it would be better to specify k numerically to control certain aspects of the distribution. In other words, the parameter k becomes our instrument in the GNVM model and this will be explored in more detail in the next section.

Model	μ	σ	θ	ν	k	DIC
GNVM	-0.075	2.188	0.141	1.01	0.2973	10961.96
NVM	-0.094	2.244	0.160	0.81	–	10961.91

Table 1: Estimated parameters for the GNVM and NVM models using the Microsoft data set with $N = 2518$

3 Contamination as an instrument of financial policy

Using (4), we can interpret the GNVM model as a NVM model contaminated by some additive normal noise. In this section, we try to build on that idea by exploring the possibility of using contamination as an instrument by specifying the level of contamination in the GNVM model, specifically the value of k, and fitting to three different simulated datasets from the NVM and two different skew t model on `JAGS`.

The first of the two skew t distributions was discussed in Bibby and Sørensen (2003) and is defined as $X_{st1}|V \sim N(\mu + \theta/V, \sigma^2/V)$, where $V \sim \Gamma(\frac{\eta}{2}, \frac{\eta}{2})$. The second of the two skew distributions was originally introduced by Gupta, Chang and Huang (2002), but we will define it by the mixing representation used in Fung and Seneta (2010) as $X_{st2}|(Y, V) \sim N(\mu + \theta Y/\sqrt{V}, \sigma^2/V)$, where $Y = |W|$ with $W \sim N(0, 1)$ and $V \sim \Gamma(\frac{\eta}{2}, \frac{\eta}{2})$. These two skew t distributions will be called skew t distribution Type I and Type II respectively in the sequel.

For the simulation study, we simulated 2,000 readings from each of the three distributions with parameters $(\mu, \sigma, \theta, \nu) = (0.2,\ 2,\ 0.1,\ 0.83)$ for the NVM model and $(\mu, \sigma, \theta, \eta) = (0.2, 1, 0.1, 4)$ for the two skew t models, so that both skew t

distributions have heavy tails. We then fit all data with the NVM as well as the GNVM distributions. The data fitting procedure is very similar to the one used in the last section except that we fix $k = (0.2, 0.4, 0.6, 0.8, 1)$ here. The DIC values are again computed from first principles to assess the fit of different models. The results are summarised in Table 2. Based on the DIC values for the NVM data in Table 2, the GNVM model with a $k \leq 0.6$ provides a similar fit to the correctly specified NVM model whereas the GNVM model with $k = 0.8$ and 1.0 provides an inferior fit demonstrated by their significantly larger DIC values.

Data	Model	μ	σ	θ	ν	DIC
NVM	NVM (true)	0.2	2	0.1	0.833	–
	NVM (estimated)	0.162	1.96	0.144	0.856	8143.01
	GNVM, $k = 0.2$	0.186	1.924	0.122	1.11	8141.07
	GNVM, $k = 0.4$	0.206	1.878	0.102	1.34	8143.06
	GNVM, $k = 0.6$	0.226	1.830	0.082	1.60	8145.85
	GNVM, $k = 0.8$	0.237	1.768	0.071	1.78	8149.20
	GNVM, $k = 1.0$	0.246	1.700	0.062	1.86	8156.11
Skew t Type I	NVM	0.142	1.394	0.257	0.782	6842.332
	GNVM, $k = 0.2$	0.198	1.342	0.202	1.326	6820.548
	GNVM, $k = 0.4$	0.238	1.280	0.162	2.096	6810.931
	GNVM, $k = 0.6$	0.264	1.211	0.135	3.367	6807.264
	GNVM, $k = 0.8$	0.284	1.153	0.115	5.780	6809.692
Skew t Type II	NVM	0.187	1.373	0.157	0.637	6814.426
	GNVM, $k = 0.2$	0.225	1.307	0.118	0.985	6803.562
	GNVM, $k = 0.4$	0.250	1.235	0.093	1.521	6794.699
	GNVM, $k = 0.6$	0.268	1.164	0.073	2.417	6789.303
	GNVM, $k = 0.8$	0.282	1.087	0.061	4.174	6786.192
	GNVM, $k = 1.0$	0.294	1.014	0.050	7.112	6789.498

Table 2: Estimated parameters for the GNVM and NVM models using the three simulated data set with $N = 2000$ via `JAGS`

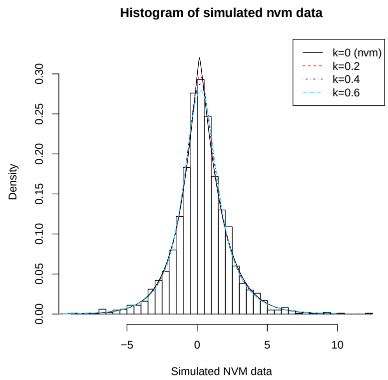

Figure 1: Histogram of the simulated NVM data with the fitted NVM and GNVM (with various fixed k) density

Further, we believe that the true advantage of adding contamination is that it allows us to modified the shape of the NMV density in a desirable way. Contamination seems to reduce the weight at the centre and reallocate it to the tails of the distribution. As evident from the fitted NVM density curves in Figure 1, the NVM model can overestimate the weight at the centre and adding contamination seems to rectify that problem. Furthermore, the fitted GNVM models tend to have heavier tails than the corresponding NVM model. This can be seen by comparing their coefficient of kurtosis as well as examining their tail behaviour.

Suppose X_{GNVM} satisfies (1) with $V \sim \Gamma(\frac{1}{\nu}, \frac{1}{\nu})$, $\nu > 0$ and $\mu_r = E(X_{GNVM} - E(X_{GNVM}))^r$ be the r-th central moment of X_{GNVM}. Then the coefficient of skewness β and kurtosis κ of X_{GNVM} are

$$\beta = \frac{\mu_3}{(\mu_2)^{3/2}} = \frac{3\theta\sigma^2\nu + 2\theta^3\nu^2}{(\sigma^2 + \theta^2\nu + k)^{\frac{3}{2}}}; \tag{6}$$

$$\kappa = \frac{\mu_4}{(\mu_2)^2} = \frac{3\sigma^4(1+\nu) + 3\theta^2\nu(2\sigma^2 + \theta^2\nu)(2\nu+1) + 6(\theta^2\nu + \sigma^2)k + 3k^2}{(\sigma^2 + \theta^2\nu + k)^2} \tag{7}$$

respectively. These are obtained by adapting the results for (4) on p. 114 of Tjetjep and Seneta (2006). The coefficients of skewness and kurtosis, defined in (6) and (7), corresponding to those fitted GNVM models which are indistinguishable as regards fit from the NVM (those with $k \leq 0.6$) from Table 2 are summarised in Table 3. As the level of contamination increases, the coefficient of kurtosis also increases. This suggests the tails of the GNVM fitted model get heavier as k increases.

A similar conclusion can be drawn from the asymptotic tail behaviour of the GNVM model, which is summarised as the following theorem.

Theorem 1. *If X_{GNVM} is a random variable which satisfies (1) and (3) with $\mu = 0$, then*

$$P(X_{GNVM} \leq x) \sim \frac{|z|^{\frac{1}{\nu}-1} e^{t_l(z+\frac{1}{2}kt_l)}}{\nu^{\frac{1}{\nu}}\Gamma(\frac{1}{\nu})(\frac{2\sigma^2}{\nu} + \theta^2)^{\frac{1}{2\nu}} t_l}, \quad \text{as } z \to -\infty; \tag{8}$$

$$P(X_{GNVM} \geq z) \sim \frac{z^{\frac{1}{\nu}-1} e^{-t_u(z-\frac{1}{2}kt_u)}}{\nu^{\frac{1}{\nu}}\Gamma(\frac{1}{\nu})(\frac{2\sigma^2}{\nu} + \theta^2)^{\frac{1}{2\nu}} t_u}, \quad \text{as } z \to \infty, \tag{9}$$

$$\text{where} \quad t_l = \frac{\sqrt{\frac{2\sigma^2}{\nu} + \theta^2} + \theta}{\sigma^2} \quad \text{and} \quad t_u = t_l - \frac{2\theta}{\sigma^2}. \tag{10}$$

Thus, the tail behaviour of the GNVM model is controlled by t_l and t_u. The closer to zero they are, the heavier the tails are. The constants t_l and t_u corresponding to those fitted GNVM models are summarised into Table 3. It is obvious that as k increases, the tails of the fitted GNVM model get heavier. We can also see from the expression of t_u and t_l that the effect of raising ν is to increase the tail weight of the GNVM and the NVM models.

Model	NVM (true)	GNVM, $k = 0.2$	GNVM, $k = 0.4$	GNVM, $k = 0.6$
β	0.125	0.194	0.186	0.167
κ	5.51	6.02	6.28	6.48
t_l	0.800	0.731	0.680	0.636
t_u	0.750	0.665	0.622	0.587

Table 3: The coefficient of skewness and kurtosis and t_l and t_u of the fitted model

The advantage of allowing "contamination" is even more profound when the data has heavier tails. By comparing the DIC values for both skew t datasets in Table 2, we can see that having the additional (fixed-value) k term actually improves the fit of the NVM model. This improvement can be explained by the increase in tail weights of models, which can be seen by the increase in the estimated ν as k increases in Table 2. Even though the value of k does not directly influence the asymptotic tail behaviour of the distribution, as ν does as seen in (8) to (10), k has an implicit effect on the tails by affecting the weight near the centre of the distribution.

4 Recommendation

We assumed k to be a fixed constant in the last section. Inevitably one would ask what would the appropriate choice for k in an uncontrolled setting. If you found the fit of the basic NVM model quite reasonable but the weight at the centre seems to be slightly overestimated, we recommend using a a small k to be approximately 5% of the estimated σ^2 of the NVM model. On the other hand, if you found the tails weight to be underestimated by the NVM model, we recommend using a larger k to be about 15 to 30% of the estimated σ^2 of the NVM model depends on how much extra weight you would like to reallocate from the centre to the tails.

To conclude, we demonstrated that we can adjust the shape of the Variance Gamma model in a favourable way by using contamination as an instrument.

Acknowledgement

This project was supported by the Macquarie University though the Macquarie University New Staff Research Grant Scheme (MQNS).

References

[1] Barndorff-Nielsen, O., Kent, J. and Sørensen, M. (1982). Normal Variance-Mean mixtures and z distributions. *Int. Stat. Rev.*, **50**, 145–159.

[2] Bibby, B. and Sørensen, M. (2003). Hyperbolic processes in finance. In: Rachev, S.T., eds. *Handbook of Heavy Tailed Distributions in Finance.* Elsevier, Amsterdam, pp. 211-248.

[3] Fung, T. and Seneta, E. (2010). Modelling and Estimation for Bivariate Financial Returns. *Int. Stat. Rev.*, **78**, 117–133.

[4] Gupta, A.K., Chang, F.C. and Huang, W.J. (2002). Some skew-symmetric models. *Random Operators and Stochastic Equations*, **2**, 133–140.

[5] Madan, D.B. and Seneta, E. (1990). The variance-gamma (V.G.) model for share market returns. *J. Business*, **63**, 511–524.

[6] Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing.* Vienna, Austria.

[7] Press, S.J. (1967). A compound events model for security prices. *J. Business*, **40**, 317–335.

[8] Rzetelski, P. (2009). *Generalised normal variance-mean mixing model for asset returns.* Honours Thesis. University of Sydney.

[9] Tjetjep, A. and Seneta, E. (2006). Skewed normal variance-mean models for asset pricing and the method of moments. *Int. Stat. Rev.*, **74**, 109–126.