

# VaR and Expected Shortfall: A Non-normal Regime Switching Framework

Robert J. Elliott

Haskayne School of Business

University of Calgary

E-mail: relliott@ucalgary.ca

Hong Miao

Haskayne School of Business

University of Calgary

E-mail: hmiao@ucalgary.ca

## Abstract

We have developed a regime switching framework to compute the Value at Risk and Expected Shortfall measures. Although Value at Risk as a risk measure has been criticized by some researchers for lack of subadditivity, it is still a central tool in banking regulations and internal risk management in the finance industry. In contrast, Expected Shortfall (ES) is coherent and convex, so it is a better measure of risk than Value at Risk. Expected Shortfall is widely used in the insurance industry and has the potential to replace Value at Risk as a standard risk measure in the near future. We have proposed regime switching models to measure value at risk and expected shortfall for a single financial asset as well as financial portfolios. Our models capture the volatility clustering phenomenon and variance independent variation in the higher moments by assuming the returns follow Student's  $t$ -distributions.

**Keywords:** Value at Risk (VaR), Expected Shortfall (ES), Regime Switching, Student- $t$  distribution, Fat tailed, Volatility Clustering

## 1 Introduction

Risk is always a concern in our lives. We face different kinds of risk everyday. Risk is particularly important in economics, and finance. However, risk is not an objective concept, and, thus, it is not at all easy to give a precise definition.

Roughly speaking, risk means a chance of injury or loss associated with a given action.

In economics, people are interested in how to measure risk. There are various attempts to look for risk measures. For a brief review of risk measures, see Miao (2006) [19]. Value at Risk (VaR), or Capital at Risk, is an industry benchmark for the disclosure of financial risk. Although Value at Risk is often criticized for lack of subadditivity, being ineffective in recognizing the dangers of concentrating credit risk, and being not able to indicate the severity of the economic consequences of exposure to the rare events, it is likely to retain its preeminent practical status since it is much easier to use than most other risk measures.

The interest in value at risk stems from the tremendous changes and expansion of financial markets. As a result of the rapid movements of financial markets, and the proliferation of derivative instruments in the past decades, many companies have complex portfolios including large numbers of cash and derivative instruments. Because of the complexity and frequent trading of various derivative instruments, the magnitudes of the risks in companies' portfolios change frequently and often are not obvious. Although many risk measures are provided by researchers, most of them are quite complicated and not understandable for senior managers. Therefore portfolio managers need a quantitative risk measure which a manager can succinctly report to the senior managers charged with the oversight of risk management and trading operations. Value at Risk (VaR) is the leading summary measure of this type.

In the late 1980s VaR was introduced and first used by major financial firms to measure the risks of their trading portfolios. Since then, the use of VaR has exploded. VaR is now widely used by other financial institutions, and non-financial corporations.

As indicated by Duffie and Pan (1997) [6], the measurement of VaR of a portfolio, including basic instruments and derivatives, relies on a model of the price change of the underlying instruments, and a model for computing the sensitivity of the prices of derivatives to the underlying prices. A variety of approaches has been developed to estimate VaR. Basically there are four groups of ways to compute VaR. They are historical simulation, fully parametric models, extreme value theory and quantile regression approaches. For details of a comparison of the four groups, see Kuuster et al. (2006) [16].

In the existing literature, two issues are key in order to calculate VaR. They are the validity of the normality assumption for the value or return

on an asset or portfolio, including the study of fat tails and the volatility clustering, and the assumption of non-linearity of the assets. The study of non-normal features of financial returns was pioneered by Mandelbrot (1963) [18] and Fama (1963) [11]. From then on, many empirical studies have proven the non-normality of financial returns. Those studies led some authors to propose other distributions with fatter tails than the normal distribution, such as Pareto's stable distribution, Student-t distribution, and the skewed Student-t distribution. Other authors have proposed distributions to describe only the behaviors of the extreme returns based on the Extreme Value Theory, see for example Neftci (2000) [20].

Another branch of research in the area of the computation of VaR is the use of regime switching models. Regime switching models are commonly used to capture the dynamics of financial returns. Guidolin and Timmermann (2003) [12] provide a four state multivariate regime switching model for the joint distribution of monthly stock and bond returns and compare it with two benchmarks. They found that the regime switching model is accurate in out-of-sample VaR predictions. Billio and Pelizzon (2000) [3] propose four different regime switching models and show that the models perform better than the GARCH and GARCHB models. Kawata and Kijima (2005) [15] introduce adjustment factors for state probabilities in regime switching models for estimate portfolio VaR, and argue that their models perform well.

All the regime switching models mentioned above assume the returns are drawn from different normal densities through time depending on states. That is, in the two state models, there are only two variance levels. This is not consistent with the empirical evidence and is in contrast with the GARCH models in which the variance is continuous. These models do not give closed formulas for parameter estimations, so they are not easy to apply in practice. To resolve this issue, we shall propose a regime switching Student-t distribution model in which we assume the returns follow a Student-t distribution. Under this assumption, the distributions can differ more than their scales across regimes. We shall derive closed form estimates for all the parameters in the model by applying filtering theory.

Although Expected Shortfall, (ES hereafter), is less commonly used in finance industry, it is widely used in the insurance industry. ES is a good measure of risk since it is coherent, convex and stable. Most of the methods used to estimate VaR can also be applied to compute ES, so we compute both of them in this paper.

The paper is constructed as follows. In section 2 we give a brief description

of the mathematical definitions of Value at Risk and Expected Shortfall, and the model set-up. In section 3 we derive the features of the model and the solutions of the parameters. In section 4 and 5 we develop closed form expressions for both VaR and ES. A conclusion summarizing the most significant findings is given in section 6.

## 2 Definitions and Model Setup

### 2.1 Definitions and Features of VaR

VaR is a single, summary statistical measure which aggregates all of the risks in a portfolio into a single number suitable for use in the boardroom, reporting to regulators, or disclosure in an annual report. It is simply a way to describe the magnitude of likely losses in a portfolio. More precisely, VaR is defined as:

**Definition 1** *VaR is a statistical estimation of a portfolio loss with the property that, with a small probability  $\alpha$ , the owner of the portfolio stands to incur that loss or more over a given (typically short) holding period.*

Typical in practice, values for the probability  $\alpha$  are 1, 2.5, and 5 percent; common holding periods are 1, 2, and 10 business days and 1 month.

Delbaen (2000) [5] gave a more precise definition which is following:

**Definition 2** *Let  $X$  be a random variable and  $\alpha \in [0, 1]$ ,  $q$  is called an  $\alpha$ -quantile if:*

$$P[X < q] \leq \alpha \leq P[X \leq q].$$

*The largest  $\alpha$ -quantile is:*

$$q_\alpha(X) = \inf \{x | P[X \leq x] > \alpha\},$$

*and the smallest  $\alpha$ -quantile is:*

$$q_\alpha(X) = \inf \{x | P[X \leq x] \geq \alpha\}.$$

**Definition 3** *Given a position  $X$  and a number  $\alpha \in [0, 1]$  we define*

$$VaR_\alpha(X) := -q_\alpha(X)$$

*and we call  $X$  VaR-acceptable if  $VaR_\alpha(X) \leq 0$  or, equivalently, if  $q_\alpha \geq 0$ .*

VaR has the following properties:

1.  $X \geq 0 \implies VaR_\alpha(X) \leq 0$ ,
2.  $X \geq Y \implies VaR_\alpha(X) \leq VaR_\alpha(Y)$ ,
3.  $VaR_\alpha(\lambda X) = \lambda VaR_\alpha(X), \forall \lambda \geq 0$ ,
4.  $VaR_\alpha(X + k) = VaR_\alpha(X) - k, \forall k \in \mathbb{R}$ .

This generalized definition of  $VaR_\alpha$  can be calculated for any random variable. However, there are good reasons for rejecting  $VaR$  as an adequate measure of risk. First of all,  $VaR_\alpha$  loses convexity properties, or is not sub-additive, so it is not coherent. (see Artzner (1999) [2] or Delbaen (2000) [5] for details on coherent risk measures.) This may result in that diversified portfolios are riskier than less diversified portfolios. Secondly,  $VaR_\alpha$  is ineffective in recognizing the dangers of concentrating credit risk. Finally,  $VaR_\alpha$  does not indicate the severity of the economic consequences of exposure to the rare events.

## 2.2 Definitions and Features of ES

As an alternative measure, ES has superior properties in many respects. ES is a coherent risk measure. In this section, we shall present the definition and properties of ES.

Although VaR is very broadly used as a standard risk measure in practice, it suffers from the three previous stated problems. Since VaR does not indicate how "bad" the bad consequences can be, or in other words, VaR does not give any information about how much the loss will be when the rare events happen, we need a measure of risk which can provide this kind of information. Therefore, ES is a natural alternative to VaR. ES is the expected loss incurred in the  $\alpha\%$  worst cases, whereas VaR indicates the minimum loss incurred in the  $\alpha\%$  worst cases. It is clearly natural for a portfolio manager or a CEO to worry about the expected loss instead of the minimum loss at the worst situation, so ES is a more natural and reasonable risk measure than VaR.

For simplification, we give simplified definition of ES. The definition below is valid only for a smooth distribution function. For a precise definition, see Rockafellar and Uryasev [21], or Elliott and Kopp [7],

**Definition 4** *ES equals the conditional expectation of  $X$ , given that  $X > VaR_\alpha(X)$ , i.e.*

$$ES_\alpha = E[X|X > VaR_\alpha(X)].$$

*Similarly, ES has the following properties:*

1.  $X \geq Y \implies ES_\alpha(X) \leq ES_\alpha(Y)$ ,
2.  $ES_\alpha(\lambda X) = \lambda * ES_\alpha(X), \forall \lambda \geq 0$ ,
3.  $ES_\alpha(X + k) = ES_\alpha(X) - k, \forall k \in \mathbb{R}$ .
4.  $ES_\alpha(X)$  is convex, i.e.

$$ES_\alpha(\lambda X + (1 - \lambda)Y) \leq \lambda ES_\alpha(X) + (1 - \lambda) ES_\alpha(Y).$$

A similar presentation of the conditional value at risk can also be found in Tasche [22], where the name *expected shortfall* is used.

## 2.3 Model Setup

We consider a regime switching model in which there is an unobserved state variable  $X$ . This variable can have different values, referred to as 'regimes' or 'states'. The reason for introducing a state variable  $X$  is that in the real world there are different events which have different influences on the economy and, therefore, on the financial markets. For example, the returns of financial instruments can be heavily affected by economic news. It is common knowledge that "important" news and "unimportant" news have different effects on the financial markets. We represent such "states of the world" by a Markov chain  $X$  with a transition matrix  $A$ . In our model, we assume the world will have 2 states. Therefore, the state space  $S$  for  $X$  is the set of 2 column vectors with unity in the  $i$ th position and zero elsewhere. That is

$$S = \{e_i, i = 1, 2\} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}.$$

Suppose  $p_{ji} = P(X_k = e_j | X_{k-1} = e_i)$ , and write  $A = (p_{ji}), 1 \leq i, j \leq 2$ , for the transition matrix of the chain  $X$ . Then, (see [10] )

$$X_k = AX_{k-1} + V_t. \tag{1}$$

where  $V_t$  is a martingale increment.

We suppose the returns  $Y_k$  ( $0 \leq k \leq T$ ) of an individual financial asset,  $S$ , follow a switching Student-t distribution with density function

$$T(x|\nu, \mu, \sigma) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2) \sqrt{\pi\nu\sigma^2}} \left(1 + \frac{1}{\nu} \left(\frac{x-\mu}{\sigma}\right)^2\right)^{-(\nu+1)/2}, \quad \text{for } \nu, \sigma > 0.$$

Thus the model can be written as

$$Y_k \sim \begin{cases} T(\nu_1, \mu_1, \theta_1) & \text{in state 1,} \\ T(\nu_2, \mu_2, \theta_2) & \text{in state 2.} \end{cases}$$

That is, under the "real world" probability space, at time  $k$ , the density function of  $Y_k$  is

$$\psi(Y_k) = \langle X_{k-1}, e_1 \rangle T_1(Y_k) + \langle X_{k-1}, e_2 \rangle T_2(Y_k).$$

Therefore, the model can describe the fat-tailed phenomenon which is very commonly noticed in high frequent financial data. The model can also capture the volatility clustering phenomenon using the different "regimes".

A Student-t distribution might arise in the following way:

Suppose:

$$Y_k = \log \left( \frac{S_k}{S_{k-1}} \right) = \mu_k(X_{k-1}) + \sigma(X_k, k) \varepsilon_{k+1}.$$

Here,  $\varepsilon_{k+1} \sim N(0, 1)$ . Therefore, the returns are normally distributed with mean  $\mu_k$  and variance  $\sigma_k^2$ . Here  $\mu_k$  depends on the state variable  $X_k$ , and  $\sigma_k^2$  depends on the state variable  $X_k$  and time  $k$ . This would be a difference between our model and the previous regime switching VaR models in which the variance,  $\sigma_k^2$ , only depends on the state variable  $X_k$ . We could then assume that the precision, the reciprocal of the variance, follows a gamma distribution, so the variance in each state can take any positive value on the real line since the support of the gamma distribution is  $\mathbb{R}^+$ . Thus, we have

$$\begin{aligned} \mu_k &= \langle \boldsymbol{\mu}, X_{k-1} \rangle, \\ \sigma_k^2 &= \langle \boldsymbol{\sigma}^2, X_{k-1} \rangle, \end{aligned}$$

where

$$\boldsymbol{\mu} = \begin{pmatrix} \mu^1 \\ \mu^2 \end{pmatrix}, \quad \boldsymbol{\sigma}^2 = \begin{pmatrix} \sigma_1^2 \\ \sigma_2^2 \end{pmatrix},$$

and

$$\frac{1}{\sigma_i^2} \sim G(\nu_i, \nu_i \theta_i), \text{ for } i = 1, 2.$$

Here  $G$  is a Gamma distribution with density function

$$G(x|\nu, \beta) = \frac{(\beta/2)^{\nu/2}}{\Gamma(\nu/2)} x^{\nu/2-1} e^{1(\beta/2)x} I_{\{x \geq 0\}}, \quad \nu, \beta > 0,$$

where

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt.$$

Then we have

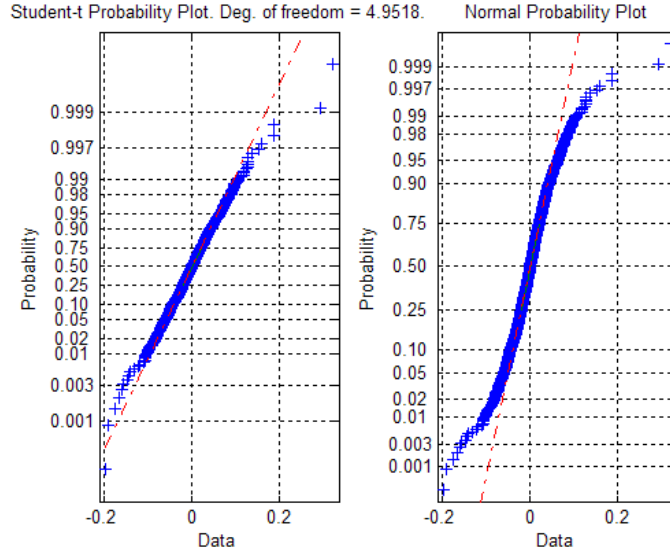
$$Y_t|e_i \sim T(\nu_i, \mu_i, \theta_i),$$

where  $\mu_i$  is a location parameter,  $\theta_i$  is a scale parameter, and  $\nu_i$  is a shape parameter or degrees of freedom.

## 2.4 Why a Student-t distribution?

The Student-t distribution itself can improve the estimation of VaR. This is shown by Lin and Shen (2006) [17]. They aimed to investigate how effectively the value at risk (VaR) estimated using the Student-t distribution captures the market risk. In their study, they considered the Student-t distribution as a fit to the empirical distribution for estimating the VaR measure, namely, VaR-t method. They found that using the student-t distribution for estimating VaR can improve the VaR estimation and offer accurate VaR estimates, particularly when the tail index technique is used to determine the degrees of freedom and the confidence level exceeds 98.5 percent. The following picture compares Student-t and Normal distributions applied to the prompt month Hub Nat. gas Nymex Futures contract.

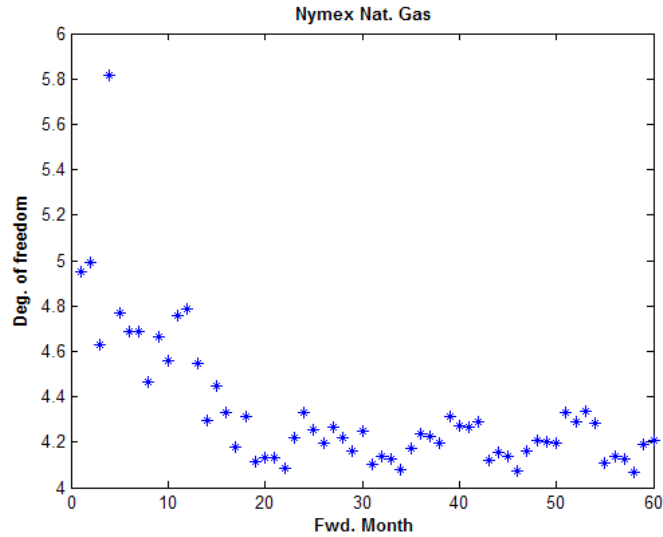




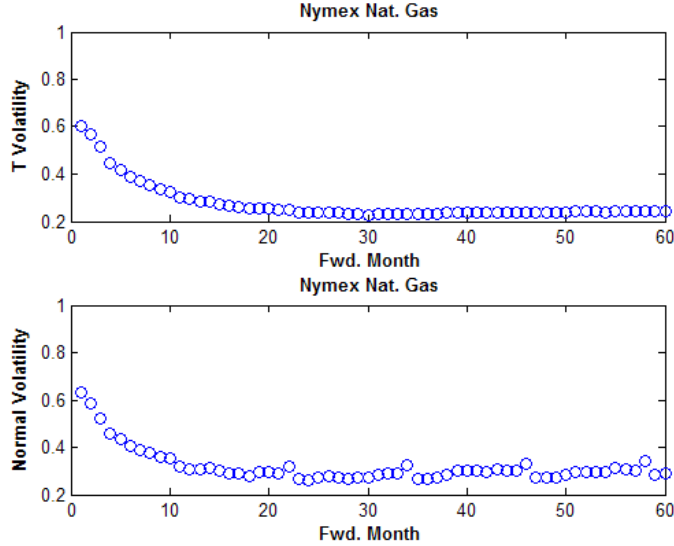
As the graphs show, the Student-t distribution is a better fit. Moreover since, see Heikkinen, and Kanto (2002) [13], and Andreev, and Kanto (2004) [1],

$$v = 4 + \frac{6}{k},$$

one can estimate the degrees of freedom by estimating the excess kurtosis  $k$ . The following graph shows this estimation for the Hub Nat. gas Nymex Futures contract.



Using these estimates for the degrees of freedom we can use Maximum Likelihood to estimate the volatility of these contracts, see Tchernister, and Rubisov (2005) [23]. The following graph shows that Student-t estimates are more stable.



### 3 Parameter Estimation

#### 3.1 Change of Measure

In this subsection, we shall develop non-linear filters for the parameters of the model. Our filters are new. These non-linear filters make the computation of VaR and ES much more easy and useful.

In practice, we only observe the return process  $Y_1, Y_2, \dots, Y_t$ . We want to estimate the transition matrix  $A$ , and the means and variances in the model based on the available information. Following Elliott et al. [8], we use the change of measure technique and initially work with a "reference" probability  $\bar{P}$ . We suppose that under the probability space  $\bar{P}$ , the  $Y_t$  are a sequence of i.i.d. random variables independent of  $X$  and each of which is  $T(1, 0, 1)$ , where  $T(1, 0, 1)$  represents the "standard" Student-t distribution with density function

$$\phi(x) = \frac{1}{\int_0^\infty t^{1/2} e^{-t} dt \sqrt{\pi}} (1+x^2)^{-1}.$$

Define:

$$\lambda_k = \frac{\psi(Y_k)}{\phi(Y_k)}, k = 1, 2, \dots,$$

and

$$\begin{aligned} \Lambda_0 &= 1, \\ \Lambda_t &= \prod_{k=1}^t \lambda_k \text{ for } k \geq 1. \end{aligned}$$

Write  $\mathcal{F}_k = \sigma\{X_0, X_1, \dots, X_k\}$  for the  $\sigma$ -field generated by  $X_0, X_1, \dots, X_k$ . Then the filtrations

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \dots \subset \mathcal{F}_k$$

and  $\{\mathcal{F}_k\}$  is a filtration which models all possible histories of  $X$ . Similarly we write

$$\mathcal{Y}_k = \sigma\{Y_1, Y_2, \dots, Y_k\},$$

$$\mathcal{G}_k = \sigma\{X_0, X_1, \dots, X_k, Y_1, Y_2, \dots, Y_k\}.$$

Then the filtrations  $\{\mathcal{F}_k\}$ ,  $\{\mathcal{Y}_k\}$  and  $\{\mathcal{G}_k\}$  represent possible histories of the state process  $X$ , the observation process  $Y$  and both process  $\{X, Y\}$ , respectively.

We define a measure  $P$  by setting  $\frac{dP}{d\bar{P}}|_{\mathcal{G}_k} = \Lambda_k$ . Then  $P$  is the "real world" probability, and we have the following lemma.

**Lemma 1** *Under  $P$  the  $Y_k$ ,  $k = 1, 2, \dots$ , are sequences of independent random variables and have the density*

$$\psi_k(Y) = \langle X_{k-1}, e_1 \rangle T_1(Y_k) + \langle X_{k-1}, e_2 \rangle T_2(Y_k).$$

**Proof.** What we must show is that for any measure function  $f(Y_k)$

$$E[f(Y_k) | \mathcal{G}_{k-1}] = \int_{-\infty}^{\infty} f(y) \psi_k(y) dy.$$

Using Bayes' Theorem we have

$$\begin{aligned} E[f(Y_k) | \mathcal{G}_{k-1}] &= \frac{\bar{E}[\Lambda_k f(Y_k) | \mathcal{G}_{k-1}]}{\bar{E}[\Lambda_k | \mathcal{G}_{k-1}]} \\ &= \frac{\bar{E}[\lambda_k f(Y_k) | \mathcal{G}_{k-1}]}{\bar{E}[\lambda_k | \mathcal{G}_{k-1}]} \end{aligned}$$

Consider the denominator

$$\begin{aligned} \bar{E}[\lambda_k | \mathcal{G}_{k-1}] &= \bar{E}\left[\frac{\langle X_{k-1}, e_1 \rangle T_1(Y_k) + \langle X_{k-1}, e_2 \rangle T_2(Y_k)}{\phi(Y_k)} \middle| \mathcal{G}_{k-1}\right] \\ &= \int_{-\infty}^{\infty} \frac{\langle X_{k-1}, e_1 \rangle T_1(y) + \langle X_{k-1}, e_2 \rangle T_2(y)}{\phi(y)} \phi(y) dy \\ &= \langle X_{k-1}, e_1 \rangle + \langle X_{k-1}, e_2 \rangle = 1. \end{aligned}$$

Also

$$\begin{aligned} &\bar{E}[\lambda_k f(Y_k) | \mathcal{G}_{k-1}] \\ &= \bar{E}\left[\frac{\langle X_{k-1}, e_1 \rangle T_1(Y_k) + \langle X_{k-1}, e_2 \rangle T_2(Y_k)}{\phi(Y_k)} f(Y_k) \middle| \mathcal{G}_{k-1}\right] \\ &= \int_{-\infty}^{\infty} \frac{\langle X_{k-1}, e_1 \rangle T_1(y) + \langle X_{k-1}, e_2 \rangle T_2(y)}{\phi(y)} f(y) \phi(y) dy \\ &= \langle X_{k-1}, e_1 \rangle \int_{-\infty}^{\infty} T_1(y) f(y) dy + \langle X_{k-1}, e_2 \rangle \int_{-\infty}^{\infty} T_2(y) f(y) dy \\ &= \int_{-\infty}^{\infty} f(y) \psi_k(y) dy. \end{aligned}$$

□

From the Lemma, we know that the probability measure is the "real world" probability measure under which we assume the sequence  $Y_k$ ,  $k = 0, 1, 2, \dots$ , are sequences of independent random variables following  $\psi_k(Y)$ .

We wish to estimate  $X$  given the observations of  $Y$  under the "real world" probability  $P$ , that is, we are interested in the quantity  $E[X_k | \mathcal{Y}_k]$ . As we have stated above,  $\bar{P}$  is a nicer measure to work with. Using the Bayes' Theorem (see [10]), we have

$$E[X_k | \mathcal{Y}_k] = \frac{\bar{E}[\Lambda_k X_k | \mathcal{Y}_k]}{\bar{E}[\Lambda_k | \mathcal{Y}_k]}. \quad (2)$$

We write  $q_k = \bar{E} [\Lambda_k X_k | \mathcal{Y}_k]$ , and notice that  $\sum_{i=1}^2 \langle X_k, e_i \rangle = 1$ , so

$$\sum_{i=1}^2 \bar{E} [\langle \Lambda_k X_k, e_i \rangle | \mathcal{Y}_k] = \bar{E} \left[ \Lambda_k \sum_{i=1}^2 \langle X_k, e_i \rangle | \mathcal{Y}_k \right] = \bar{E} [\Lambda_k | \mathcal{Y}_k] = \sum_{i=1}^2 \langle q_k, e_i \rangle.$$

Therefore

$$E [X_k | \mathcal{Y}_k] = \frac{q_k}{\sum_{i=1}^2 \langle q_k, e_i \rangle}.$$

The following lemma gives a recursive filter for  $q$ .

**Lemma 2**  *$q$  satisfies the recurrence  $q_k = AB(Y_k)q_{k-1}(h)$ , where  $B(Y_k)$  is a 2 dimension diagonal matrix with entries  $\frac{T_i(Y_k)}{\phi(Y_k)}$ .*

**Proof.**

$$\begin{aligned} q_k &= \bar{E} [\Lambda_k X_k | \mathcal{Y}_k] \\ &= \bar{E} [\Lambda_{k-1} \lambda_k (AX_{k-1} + V_k) | \mathcal{Y}_k] \\ &= \bar{E} \left[ \Lambda_{k-1} \frac{\langle X_{k-1}, e_1 \rangle T_1(Y_k) + \langle X_{k-1}, e_2 \rangle T_2(Y_k)}{\phi(Y_k)} (AX_{k-1} + V_k) | \mathcal{Y}_k \right] \\ &= \sum_{i=1}^2 \bar{E} [\Lambda_{k-1} \langle X_{k-1}, e_i \rangle | \mathcal{Y}_k] \frac{T_i(Y_k)}{\phi(Y_k)} A e_i \\ &= \sum_{i=1}^2 \langle q_{k-1}, e_i \rangle \frac{T_i(Y_k)}{\phi(Y_k)} A e_i \\ &= AB(Y_k) q_{k-1}. \end{aligned}$$

□

It is clear that to obtain an estimate of  $X$ , the parameters  $A = (p_{ji})$ ,  $\nu = (\nu_1, \nu_2)$ ,  $\mu = (\mu_1, \mu_2)$ , and  $\theta = (\theta_1, \theta_2)$  should be estimated first. To estimate these parameters we need estimates of the number of jumps from state  $r$  to state  $s$  in the time period  $k$ :

$$J_k^{rs} = \sum_{l=1}^k \langle X_{l-1}, e_r \rangle \langle X_l, e_s \rangle,$$

and the amount of time  $X$  has spent in state  $r$  up to time  $k$ :

$$O_k^r = \sum_{l=1}^k \langle X_l, e_r \rangle.$$

Also, we need an estimate of the process

$$Z_k^r = \sum_{l=1}^k \langle X_{l-1}, e_r \rangle f(Y_l),$$

where  $f(Y)$  is a function of  $Y$  for the estimate of the components of  $\mu, \theta$ , and  $\nu$ . (see [10] for details). Actually, these three processes can be constructed as one special process. (see [9])

Notice that for a  $\mathcal{G}$ -adapted process, if we write  $\hat{Z}_k = E[Z|\mathcal{Y}_k]$ , and  $\sigma(Z)_k = \bar{E}[\bar{\Lambda}_k Z_k | \mathcal{Y}_k]$  we have

$$\hat{Z}_k = \frac{\sigma(Z)_k}{\langle q_k, 1 \rangle}.$$

and

$$\sigma(Z)_k = \sum_{i=1}^2 \langle \sigma(ZX)_k, e_i \rangle.$$

The following lemmas give the recursions for  $\sigma(J^{rs}X)_k$ ,  $\sigma(O^rX)_k$ , and  $\sigma(T^rX)_k$ .

**Lemma 3** *With  $B(Y_k)$  defined as the previous Lemma, we have*

$$\sigma(J^{rs}X)_k = AB(Y_k) \sigma(J^{rs}X)_{k-1} + \langle q_{k-1}, e_r \rangle \frac{T_r(Y_k)}{\phi(Y_k)} p_{sr} e_r.$$

**Proof.** See Appendix A. □

Similarly, we have the following Lemma.

**Lemma 4**

$$\begin{aligned} \sigma(O^rX)_k &= AB(Y_k) \sigma(O^rX)_{k-1} + \langle AB(Y_k) q_{k-1}, e_r \rangle e_r, \\ \sigma(Z^rX)_k &= AB(Y_k) \sigma(Z^r(f)X)_{k-1} + \langle q_{k-1}, e_r \rangle \frac{T_r(Y_k)}{\phi(Y_k)} f(Y_k) A e_r. \end{aligned}$$

### 3.1.1 Parameter Estimation

All the regime switching papers just stated the application of Maximum Likelihood Method, or the Expectation Maximization (EM) to estimate the parameters, but no one really gave the formulas of the parameters. We shall derive the formulas for all the parameters to facility the application of the model.

In this section, we estimate the parameters of the model. Our model is determined by several parameters. They are: the transition matrix  $A = (p_{ji})$ , and the vectors  $\nu = (\nu_1, \nu_2)$ ,  $\mu = (\mu_1, \mu_2)$ , and  $\theta = (\theta_1, \theta_2)$ . Write the parameters as a set

$$\theta := \{p_{ji}, \mu_i, \nu_i, \theta_i, 1 \leq i, j \leq 2\},$$

where  $p_{ji} \geq 0$ ,  $\sum_{j=1}^2 p_{ji} = 1$ . One method of estimating parameters is the Maximum Likelihood Estimate (MLE) in which the optimal estimates maximize the log-likelihood function with respect to some fixed probability  $P_0$  :

$$L(\theta) = E \left[ \log \frac{dP_\theta}{dP_0} | \mathcal{Y} \right].$$

However the MLE is hard to compute. Also, because the variable  $X$  is not observed directly we must consider the conditional expectation of the log-likelihood. Consequently we shall use the Expectation Maximization (EM) to compute the estimates. The basic idea of the EM algorithm is:

- We start with appropriate initial values  $\hat{\theta}_0$  for

$$\theta := \{p_{ji}, \mu_i, \nu_i, \theta_i, 1 \leq i, j \leq 2\}$$

which satisfy constraints for the parameters.

- After some observations of  $Y$ , we compute the new estimates.
- Using these values, we re-estimate the parameters iteratively until some stopping criterion is satisfied.
- After more observations we repeat the process again.

Since the EM algorithm improves the estimates monotonically, the expected log-likelihood increases with each reestimation.

Consider the parameter  $p_{ji}$ , under  $P = P_\theta$ ,  $X$  is a Markov Chain with transition  $A = (p_{ji})$ . We apply the change of measure technique and introduce a new probability measure  $P_{\hat{\theta}}$  so that under the new measure,  $X$  is now a Markov Chain with transition matrix  $\hat{A} = (\hat{p}_{ji})$ , where  $P_{\hat{\theta}}(X_t = e_j | X_{t-1} = e_i) = \hat{p}_{ji}$ ), so  $\hat{p}_{ji} \geq 0$  and  $\sum_{j=1}^N \hat{p}_{ji} = 1$ .

**Lemma 5** Define  $\Lambda_0 = 1$  and  $\Lambda_k = \prod_{l=1}^k \left( \sum_{r,s=1}^N \left( \frac{\hat{p}_{sr}}{p_{sr}} \right) \langle X_l, e_s \rangle \langle X_{l-1}, e_r \rangle \right)$ . ( If  $p_{ji} = 0$ , take  $\hat{p}_{ji} = 0$  and  $\frac{\hat{p}_{ji}}{p_{ji}} = 1$ . ) Define  $P_{\hat{\theta}}$  by setting  $\frac{dP_{\hat{\theta}}}{dP_\theta} | \mathcal{F}_k = \Lambda_k$ . Then under the new measure  $P_{\hat{\theta}}$ ,  $X$  is a Markov Chain with transition  $\hat{A} = (\hat{p}_{ji})$ .

The following proposition gives the EM estimates for the entries of the transition matrix.

**Proposition 1** Given the observations up to time  $k$ ,  $\{Y_0, Y_1, Y_2, \dots, Y_k\}$ , that is, given  $\mathcal{Y}_k$ , and given the parameter set  $\theta := \{p_{ji}, \mu_i, \nu_i, \theta_i, 1 \leq i, j \leq 2\}$ , the EM estimates of  $\hat{p}_{ji}$  are given by

$$\hat{p}_{ji} = \frac{\hat{J}_k^{ij}}{\hat{O}_k^i} = \frac{\sigma(J^{ij})_k}{\sigma(O^i)_k}.$$

**Proof.** See Appendix B. □

Proposition 1 provides estimates for the elements of the transition matrix.

### 3.2 Estimate of $\mu$

**Remark 1** In the following propositions 2, 3 and 4, the new parameter estimates  $\hat{\mu}_r, \hat{\theta}_r, \hat{\nu}_r$ ,  $r = 1, 2$ , are obtained recursively. Initial estimates are chosen and the right hand sides of the expressions calculated. These give new estimates for  $\hat{\mu}_r, \hat{\theta}_r, \hat{\nu}_r$ . The right hand sides are then re-calculated using these revised estimates. This recursive procedure is continued until the differences between successive iterations are less than a pre-determined bound.



**Proposition 2** *The estimate for the parameter  $\mu = (\mu_1, \mu_2)'$ , at time  $k$ ,  $\hat{\mu} = (\hat{\mu}_1, \hat{\mu}_2)'$  is given by*

$$\hat{\mu}_r = \frac{\sum_{i=1}^2 \left\langle \sigma \left( Z^r \left( \frac{Y_i}{\nu_r \theta_r^2 + Y_i^2 - 2Y_i \mu_r + \mu_r^2} \right) X \right)_k, e_i \right\rangle}{\sum_{i=1}^2 \left\langle \sigma \left( Z^r \left( \frac{1}{\nu_r \theta_r^2 + Y_i^2 - 2Y_i \mu_r + \mu_r^2} \right) X \right)_k, e_i \right\rangle}.$$

**Proof.** See Appendix C. □

We can use this recurrence to update  $\mu_r$  at time  $k$ .

### 3.3 Estimate of $\theta$

**Proposition 3** *The estimate for the parameter  $\theta = (\theta_1, \theta_2)'$ , at time  $t$ ,  $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2)'$  is given by*

$$\hat{\theta}_r^2 = \frac{\sum_{i=1}^2 \left\langle \sigma \left( Z^r \left( \frac{Y_i^2 - 2Y_i \mu_r + \mu_r^2}{\nu_r \theta_r^2 + Y_i^2 - 2Y_i \mu_r + \mu_r^2} \right) X \right)_k, e_i \right\rangle}{\sum_{i=1}^2 \left\langle \sigma \left( Z^r \left( \frac{1}{\nu_r \theta_r^2 + Y_i^2 - 2Y_i \mu_r + \mu_r^2} \right) X \right)_k, e_i \right\rangle},$$

and an up-dated estimate for the parameter  $\nu = (\nu_1, \nu_2)'$ , at time  $t$ ,  $\hat{\nu}_t = (\hat{\nu}_1, \hat{\nu}_2)'$  is given by

$$\hat{\nu}_r = \frac{\hat{O}_k^r}{\left[ \frac{\left( \frac{\frac{1}{2} \int_0^\infty t^{\frac{\nu_r+1}{2}-1} e^{-t} (\log t) dt}{\int_0^\infty t^{\frac{\nu_r+1}{2}-1} e^{-t} dt} - \frac{\frac{1}{2} \int_0^\infty t^{\frac{\nu_r}{2}-1} e^{-t} (\log t) dt}{\int_0^\infty t^{\frac{\nu_r}{2}-1} e^{-t} dt} \right) \hat{O}_k^r}{- \frac{1}{2} \hat{Z}^r \left( \log \left( 1 + \frac{1}{\nu_r} \left( \frac{Y_i - \mu_r}{\theta_r} \right)^2 \right) \right) + \hat{Z}^r \left( \frac{(\nu_r+1)(Y_i - \mu_r)^2}{2\nu_r(\nu_r \theta_r^2 + Y_i^2 - 2Y_i \mu_r + \mu_r^2)} \right)} \right]}.$$

This is at the M stage of the estimation, after some more observations, we first use the previous estimates for the parameters to compute the value of those expectations. Then we use these parameter recurrences to update the parameters. We always use the newest estimators to compute the next one, until some criterion are reached.

## 4 Closed-form VaR and ES for a single Asset

In the previous section, we have derived closed-form recurrences for all the parameters. Now we can give a closed-form expression for VaR and ES. Since

we have the EM estimates for all the parameters, we can compute  $VaR_\alpha$  for a single asset whose returns follow the previously described regime switching Student-t distribution model.

$VaR_\alpha$  at time  $k$  is defined as the value  $V$  such that

$$\alpha = \sum_{i=1}^2 \frac{\hat{O}_i}{k} \int_{-\infty}^{-V/S_0} \frac{\Gamma((\nu_i+1)/2)}{\Gamma(\nu_i/2)\sqrt{\pi\nu_i\sigma_i^2}} \left(1 + \frac{1}{\nu_i} \left(\frac{x-\mu_i}{\sigma_i}\right)^2\right)^{-(\nu_i+1)/2} dx. \quad (3)$$

$$ES_\alpha = \frac{1}{\alpha} \sum_{i=1}^2 \frac{\hat{O}_i}{k} \int_{-\infty}^{-V/S_0} x \frac{\Gamma((\nu_i+1)/2)}{\Gamma(\nu_i/2)\sqrt{\pi\nu_i\sigma_i^2}} \left(1 + \frac{1}{\nu_i} \left(\frac{x-\mu_i}{\sigma_i}\right)^2\right)^{-(\nu_i+1)/2} dx$$

For  $\nu_i > 2$ , the variances are given by  $\frac{v_i}{\nu_i-1}\sigma_i^2$ . By solving 3 for  $V$ , we would obtain explicit values for computing  $VaR_\alpha$  and  $ES_\alpha$  during the time period 0 to  $k$  as:

$$VaR_\alpha = \frac{1}{k} \sum_{i=1}^2 \left[ \hat{O}_i \left( -\mu_i S_0 + v_i \sigma_i S_0 \sqrt{\frac{1-\lambda_i}{\lambda_i}} \right) \right],$$

and

$$ES_\alpha = \frac{1}{k} \sum_{i=1}^2 \left[ \hat{O}_i \left( -\mu_i S_0 + \frac{v_i \sigma_i S_0}{\alpha B\left(\frac{\nu_i}{2}, \frac{1}{2}\right) (\nu_i - 1)} \right) \right] \lambda_i^{\frac{\nu_i-1}{2}}. \quad (4)$$

Here  $S_{k-1}$  is the value of the position at time  $k-1$ ,  $\lambda_i = I_{[\nu_i/2, 1/2]}^{-1}(2\alpha)$  and  $I_{[\nu_i/2, 1/2]}^{-1}$  is the inverse of the incomplete beta function, which can be computed in many statistical software packages, for instance SAS. For a detailed derivation of the expressions of 3 and 4, please see Kamdem (2005) [14], and Bormetti et. al. (2006) [4].

## 5 Closed-form VaR and ES for a Portfolio

In reality, people are worried about the risk of a portfolio or the risk of the total assets held by a company, so closed form expressions of VaR and ES for portfolios are more important than for a single asset. It is not easy to obtain closed form expressions of VaR and ES for a portfolio, since the relationship between the individual assets in the portfolio might be very complicated.

However, when a portfolio's aggregate return is a linear function of the returns of its components, we can derive closed form expressions for VaR and ES. In practice, people apply the RiskMetrics methodology to estimate VaR when the portfolios satisfy the linearity property and the returns of the individual assets can be assumed to be normally jointly distributed. If the portfolios are non-linear, or the joint distribution of the returns is non-normal, Monte Carlo methods are used. Monte Carlo methods are quite flexible, but slow when the portfolio is made up of a large number of individual assets. In our model, the normality property is not satisfied, but under the linearity assumption of the portfolios, we can obtain closed form expressions for VaR and ES.

**Definition 5** *A portfolio with value  $V(k)$  at time  $0 \leq k \leq T$  is linear if its profit or loss  $\Delta V(k) = V(k) - V(0)$  is a linear function of the returns  $r_1(k), r_2(k), \dots, r_m(k)$  of its components over the same time period. i.e.*

$$\Delta V(k) = \delta_1 r_1 + \delta_2 r_2 + \dots + \delta_m r_m,$$

or

$$\Delta V(k) = \boldsymbol{\delta} \mathbf{r}^t,$$

where  $\boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_m)$  represents the initial values of the  $m$  components of the portfolio, and  $\mathbf{r} = (r_1, r_2, \dots, r_m)$ .

Similarly to the treatment in the previous section for a single asset, we assume the returns  $r_i$  ( $1 \leq i \leq m$ ) follow Student-t distributions, with two different parameter sets in the two "regimes". The multivariate Student-t distribution then takes the following form:

$$T_i(\mathbf{r}) = \frac{\Gamma((\nu_i + m)/2)}{\Gamma(\nu_i/2) \sqrt{(\pi \nu_i)^m |\boldsymbol{\Sigma}_i|}} \left( 1 + \frac{1}{\nu_i} (\mathbf{r} - \boldsymbol{\mu}_i)^t \boldsymbol{\Sigma}_i^{-1} (\mathbf{r} - \boldsymbol{\mu}_i) \right)^{-\frac{\nu_i + m}{2}}.$$

where  $\boldsymbol{\mu}_i = (\mu_1^i, \mu_2^i, \mu_3^i, \dots, \mu_m^i) \in \mathbb{R}^m$  ( $i = 1, 2$ ) is the vector of the means of the returns of the individual assets in the portfolio,  $\nu_i > 2$  represents the degree of freedom, and  $\boldsymbol{\Sigma}_i = CC^t$  is the correlation matrix in regime  $i$ .

Theoretically, by applying the previous framework, we can obtain recursive estimates for  $\nu_i, \boldsymbol{\mu}_i$ , and  $\boldsymbol{\Sigma}_i$ . Then we can obtain the closed form

expressions of VaR and ES for the portfolio as:

$$VaR_\alpha = \frac{1}{k} \sum_{i=1}^2 \left[ \hat{O}_i \left( -\delta \mu_i^t + q_{\alpha, \nu_i} \sqrt{\delta \sum_i \delta^t} \right) \right],$$

and

$$ES_\alpha = \frac{1}{k} \sum_{i=1}^2 \left[ \hat{O}_i \left( -\delta \mu_i^t + es_{\alpha, \nu_i} \sqrt{\delta \sum_i \delta^t} \right) \right],$$

where  $q_{\alpha, \nu_i}$  and  $es_{\alpha, \nu_i}$  are scalar quantities depend on the value of  $\alpha$  and  $\nu$ . See Kamdem (2005) [14] Table 1-4 for the values of  $q_{\alpha, \nu_i}$  and  $es_{\alpha, \nu_i}$ .

However it is complicated to do so, since the correlation matrix grows quadratically when the number of individual assets in the portfolio grows. For a portfolio with  $n$  individual assets, the correlation matrix has  $n^2$  entries. To deal with this situation, the interest in copula method is growing dramatically. Following this direction, we shall develop a regime switching copula model in a next paper.

## 6 Conclusion

We have developed a regime switching Student-t distribution model to estimate VaR and ES. The Student-t distribution assumption captures the real world fat-tailed fact for financial data. The regime switching setup gives our model the power to fit the volatility clustering phenomenon. We have obtained closed-form recursive estimates for all the parameters, VaR, and ES for a single asset. We shall test our result numerically by applying real financial return time series in a second paper. The framework can also be applied to portfolios, but the computation is complicated and time consuming since the number of parameters grows quickly. To handle this situation, we suggest a regime switching copula model which is our future research topic.

## References

- [1] Andreev, A., and Kanto, A., (2004), A Note on Calculation of CVAR for Student's Distribution, Helsinki School of Economics Working papers, W-369.

- [2] Artzner, P., Delbaen, F., Eber, J., and Heath, D. (1999), Coherent measures of risk, *Mathematical Finance*, 9(3), 203-228.
- [3] Billio, N., and Pelizzon, L. (2000), Value at Risk: a Multivariate switching regime approach, *Journal of Empirical Finance*, 7, 531-554.
- [4] Bormetti, G., Cisana, E., Montagna, G., and Nicrosini, O. (2007), A non-Gaussian Approach to Risk Measures, *Physica A*, 376, 532-542.
- [5] Delbaen, F. (2000), *Draft: Coherent risk measures*. Pisa-February, 28-March, 2000.
- [6] Duffie, D. and Pan, J. (1997), An overview of Value at Risk, *Journal of Derivatives*, Spring 1997, 7-49, reprinted in *Options Markets*, edited by G. Constantinides and A. G. Malliaris, London: Edward Elgar, 2001.
- [7] Elliott, R. J., and P. E. Kopp, *Mathematics of Financial Markets*, Springer-Verlag, New York, 2004.
- [8] Elliott, R. J., Aggoun, L., and Moore, J. B. (1994), *Hidden Markov Models: Estimation and Control*. Springer Verlag.
- [9] Elliott, R. J., and Miao, H. (2006), Stochastic Volatility Model with Filtering, *Stochastic Analysis and Applications*, 24(3), 661-683.
- [10] Elliott, R. J., Hunter, W. C., and Jamieson, B. M. (1998), Drift and volatility estimation in discrete time. *Journal of Economic Dynamics & Control*, 22, 209-218.
- [11] Fama, E. (1963), Mandelbrot and the Stable Paretian Hypothesis, *Journal of Business*, 36, 420-429.
- [12] Guidolin, M., and Timmermann, A. (2003), Value at Risk and expected shortfall under regime switching, University of California at San Diego working paper, Nov. 8, 2003.
- [13] Heikkinen, V., and Kanto, A., (2002), Value-at-risk estimation using non-integer degrees of freedom of Student's distribution, *Journal of Risk*, 4(4), 77-84.

- [14] Kamdem, J. S. (2005), Value-at-risk and Expected Shortfall for Linear Portfolios with Elliptically Distributed Risk Factors, *International Journal of Theoretical and Applied Finance*, 8, 537-551.
- [15] Kawata, R., and Kijima, M. (2005), Value at Risk in a market subject to regime switching, Kyoto University working paper, May 2005.
- [16] Kuester, K., Mittnik, S., and Paolella, M. (2006), Value at Risk prediction: a comparison of alternative strategies.
- [17] Lin, C., and Shen, S., (2006), Can The Student-t Distribution Provide Accurate Value at Risk? *The Journal of Risk Finance*, 7 (3), 292-300.
- [18] Mandelbrot, B. B. (1963), *The Variation in Certain Speculative Prices*, Journal of Business, 36, 394-419.
- [19] Miao, H. (2006), A review of risk measures, Haskayne School of Business PhD candidacy exam paper III.
- [20] Nefci, S. (2000), Value at Risk calculations, extreme events, and tail estimation, *The Journal of Derivatives*, Spring 2000, 1-15.
- [21] Rochafellar, R.T., and Uryasev, S. (2002), Conditional value-at-risk for general loss distributions, *Journal of Banking & Finance*, 26, 1443-1471.
- [22] Tasche, D. (2002), Expected shortfall and beyond. *Journal of Banking & Finance*, 26, 1519-1533.
- [23] Tchernister, A., Rubisov, D. H., (2005), Robust estimation of historical volatility and correlation in risk management, Bank of Montreal working paper.

## A Proof of Lemma 3

**Proof.**

$$\begin{aligned}
& \sigma(J^{rs}X)_k \\
&= \bar{E} [\Lambda_k J_k^{rs} X_k | \mathcal{Y}_k] \\
&= \bar{E} [\Lambda_{k-1} \lambda_k (J_{k-1}^{rs} + \langle X_{k-1}, e_r \rangle \langle X_k, e_s \rangle) X_k | \mathcal{Y}_k] \\
&= \bar{E} \left[ \Lambda_{k-1} \frac{\langle X_{k-1}, e_1 \rangle T_1(Y_k) + \langle X_{k-1}, e_2 \rangle T_2(Y_k)}{\phi(Y_k)} J_{k-1}^{rs} (AX_{k-1} + V_k) | \mathcal{Y}_k \right] \\
&\quad + \bar{E} \left[ \Lambda_{k-1} \frac{\langle X_{k-1}, e_1 \rangle T_1(Y_k) + \langle X_{k-1}, e_2 \rangle T_2(Y_k)}{\phi(Y_k)} \langle X_{k-1}, e_r \rangle \langle AX_{k-1} + V_k, e_s \rangle e_s | \mathcal{Y}_k \right] \\
&= \sum_{i=1}^2 \bar{E} [\Lambda_{k-1} \langle X_{k-1}, e_i \rangle J_{k-1}^{rs} | \mathcal{Y}_k] \frac{T_i(Y_k)}{\phi(Y_k)} A e_i + \bar{E} [\Lambda_{k-1} \langle X_{k-1}, e_r \rangle | \mathcal{Y}_k] \frac{T_r(Y_k)}{\phi(Y_k)} p_{sr} e_s \\
&= \sum_{i=1}^2 \langle \sigma(J^{rs}X)_{k-1}, e_i \rangle \frac{T_i(Y_k)}{\phi(Y_k)} A e_i + \langle q_{k-1}, e_r \rangle \frac{T_r(Y_k)}{\phi(Y_k)} p_{sr} e_s \\
&= AB(Y_k) \sigma(J^{rs}X)_{k-1} + \langle q_{k-1}, e_r \rangle \frac{T_r(Y_k)}{\phi(Y_k)} p_{sr} e_s,
\end{aligned}$$

where where  $B(Y_k)$  is a 2 dimension diagonal matrix with entries  $\frac{T_i(Y_k)}{\phi(Y_k)}$  on its diagonal.  $\square$

## B Proof of Proposition 1

**Proof.** As above we define  $P_{\hat{\vartheta}}$  by:

$$\frac{dP_{\hat{\vartheta}}}{dP_{\vartheta}} | \mathcal{Y}_k = \Lambda_k = \prod_{l=1}^k \left( \sum_{i,j=1}^2 \left( \frac{\hat{p}_{ji}}{p_{ji}} \right) \langle X_l, e_j \rangle \langle X_{l-1}, e_i \rangle \right).$$

Then

$$\log \frac{dP_{\hat{\vartheta}}}{dP_{\vartheta}} = \sum_{l=1}^k \sum_{i,j=1}^2 \langle X_l, e_j \rangle \langle X_{l-1}, e_i \rangle (\log \hat{p}_{ji} - \log p_{ji})$$

$$= \sum_{i,j=1}^2 J_k^{ij} \log \hat{p}_{ji} + R(a),$$

where  $R(a)$  does not depend on  $\hat{a}$ . Then

$$L(\vartheta) = E \left[ \log \frac{dP_{\hat{\vartheta}}}{dP_{\vartheta}} | \mathcal{Y}_k \right] = \sum_{i,j=1}^2 \hat{J}_k^{ij} \log \hat{p}_{ji} + R(a). \quad (5)$$

Recall that  $\sum_{j=1}^2 \hat{p}_{ji} = 1$ , and  $\sum_{j=1}^2 J_k^{ij} = O_k^i$ , so  $\sum_{j=1}^2 \hat{J}_k^{ij} = \hat{O}_k^i$ . Then, the optimal estimate of  $\hat{p}_{ji}$  is the value that maximizes the right side of (5), and subject to  $\sum_{j=1}^2 \hat{p}_{ji} = 1$ . Let  $\boldsymbol{\lambda} = (\lambda^1, \lambda^2)'$  be the Lagrange multiplier and put

$$L(\hat{P}, \boldsymbol{\lambda}) = \sum_{i,j=1}^2 \hat{J}_k^{ij} \log \hat{p}_{ji} + R(a) + \lambda^i \left( \sum_{j=1}^2 \hat{p}_{ji} - 1 \right).$$

Differentiating in  $\hat{p}_{ji}$  and  $\lambda^i$  and equating the derivatives to 0 and solving the equations, we have

$$\lambda^i = -\hat{O}_k^i,$$

and

$$\hat{p}_{ji} = \frac{\hat{J}_k^{ij}}{\hat{O}_k^i} = \frac{\sigma(J^{ij})_k}{\sigma(O^i)_k}.$$

□

## C Proof of Proposition 2

**Proof.** The density which changes  $\mu$  to  $\hat{\mu} = (\hat{\mu}_1, \hat{\mu}_2)'$  is given by

$$\frac{dP_{\hat{\mu}}}{dP_{\mu}} | \mathcal{G}_k = \prod_{l=1}^k \frac{\langle X_{l-1}, e_1 \rangle T(Y_l | \nu_1, \hat{\mu}_1, \theta_1) + \langle X_{l-1}, e_2 \rangle T(Y_l | \nu_2, \hat{\mu}_2, \theta_2)}{\langle X_{l-1}, e_1 \rangle T(Y_l | \nu_1, \mu_1, \theta_1) + \langle X_{l-1}, e_2 \rangle T(Y_l | \nu_2, \mu_2, \theta_2)}.$$



Then

$$\begin{aligned}
& E \left[ \log \frac{dP_{\hat{\mu}}}{dP_{\mu}} | \mathcal{G}_k \right] \\
&= E \left[ \sum_{l=1}^k (\langle X_{l-1}, e_1 \rangle \log T(Y_l | \nu_1, \hat{\mu}_1, \theta_1) + \langle X_{l-1}, e_2 \rangle \log (T(Y_l | \nu_2, \hat{\mu}_2, \theta_2))) | \mathcal{Y}_k \right] \\
&\quad - \hat{R}(\nu, \mu, \theta). \\
&= E \left[ \sum_{l=1}^k \langle X_{l-1}, e_1 \rangle \left( \log \frac{\Gamma(\frac{\nu_1+1}{2})}{\Gamma(\frac{\nu_1}{2}) \sqrt{\pi \nu_1 \theta_1^2}} - \frac{\nu_1+1}{2} \log \left( 1 + \frac{1}{\nu_1} \left( \frac{Y_l - \hat{\mu}_1}{\theta_1} \right)^2 \right) \right) \right. \\
&\quad \left. + \sum_{l=1}^k \langle X_{l-1}, e_2 \rangle \left( \log \frac{\Gamma(\frac{\nu_2+1}{2})}{\Gamma(\frac{\nu_2}{2}) \sqrt{\pi \nu_2 \theta_2^2}} - \frac{\nu_2+1}{2} \log \left( 1 + \frac{1}{\nu_2} \left( \frac{Y_l - \hat{\mu}_2}{\theta_2} \right)^2 \right) \right) | \mathcal{Y}_k \right] \\
&\quad - \hat{R}(\nu, \mu, \theta).
\end{aligned}$$

The best estimate of the parameter is the one which maximizes the above conditional expectation. Differentiating in  $\hat{\mu}_r$  and setting the derivative equal to 0. Then we obtain an up-dated estimate for the parameter  $\mu_r$ :

$$\begin{aligned}
\hat{\mu}_r &= \frac{E \left[ \sum_{l=1}^k \langle X_{l-1}, e_r \rangle \left( \frac{Y_l}{\nu_r \theta_r^2 + Y_l^2 - 2Y_l \mu_r + \mu_r^2} \right) | \mathcal{G}_k \right]}{E \left[ \sum_{l=1}^k \langle X_{l-1}, e_r \rangle \left( \frac{1}{\nu_r \theta_r^2 + Y_l^2 - 2Y_l \mu_r + \mu_r^2} \right) | \mathcal{G}_k \right]} \\
&= \frac{\sum_{i=1}^2 \left\langle \sigma \left( Z^r \left( \frac{Y_l}{\nu_r \theta_r^2 + Y_l^2 - 2Y_l \mu_r + \mu_r^2} \right) X \right)_k, e_i \right\rangle}{\sum_{i=1}^2 \left\langle \sigma \left( Z^r \left( \frac{1}{\nu_r \theta_r^2 + Y_l^2 - 2Y_l \mu_r + \mu_r^2} \right) X \right)_k, e_i \right\rangle}.
\end{aligned}$$

where, we use the previous estimates for  $\nu_r$ ,  $\theta_r$ , and  $\mu_r$  in the expression.  $\square$