Testing for a Common Volatility Process and Frictionless Hedging in Bivariate Time Series Models and other improvement in model and method

Jinghui Chen

Yokohama National University

1

Abstract

This paper contained three Parts:

Part I, I proposed a statistic test for common volatility process in bivariate time series model. We considers the problem as to whether financial returns have a common volatility process in the framework of stochastic volatility models that were suggested by Harvey et al. [1994]. We propose a stochastic volatility version of the ARCH test proposed by Engle and Susmel [1993], who investigated whether international equity markets have a common volatility process. The paper also checks the hypothesis of frictionless cross-market hedging, which implies perfectly correlated volatility changes, as suggested by Fleming et al. [1998]. The paper uses the technique of Chesher [1984] in differentiating an integral that contains a degenerate density function in deriving the Lagrange Multiplier test statistic.

Part II, I extend this model to more practical situation with stable distribution. Using normal distribution to model the assets' return is not appropriate especially considering the rare events such as financial crisis. I suggest using Stable distribution which is more persuasively in fundamental generating process of price than other skewness and fat-tailed distribution. I also use state space model with $\alpha - Stable$ distribution to dynamically measure the changing of log-price, and catch different performance during different times. To compare the performance of different distribution, I use the Bayesian information criterion(BIC).

Part III, I learned a fast algorithm called fast Gaussian transform to accelerate the calculation in filtering and smoothing process. We focus on the description of this algorithm proposed in Greengard and Strain [1991]. It is one of the fast algorithm called the fast multipole method (FMM). It use the two asymptotic expansion: the near field expansion well known as Taylor expansion, and the far field multipole expansion named as Hermite expansion. There may be potential useful in high dimension particle filter and smoother , or multi-asset option pricing. It can be significant improvement for calculating the statistic proposed in Part I.

Part I

Testing for a Common Volatility Process and Frictionless Hedging in Bivariate Time Series Models

1 Introduction

This paper considers the problem as to whether financial returns have a common volatility process in the framework of stochastic volatility models that were suggested by Harvey et al. [1994]. We propose a stochastic volatility version of the ARCH test proposed by Engle and Susmel [1993], who investigated whether international equity markets have a common volatility process using a multivariate ARCH model. They found groups of countries that showed similar time-varying volatility.

Fleming et al. [1998] used the multivariate stochastic volatility model of Harvey et al. [1994], and estimated volatility linkages across stock, bond, and money markets, and found strong correlation between the markets. Fleming et al. [1998] also suggested that cross-market hedging in frictionless markets causes perfectly correlated volatility changes, extending the model of Tauchen et al. [1983]. This linkage is stronger than the presence of a common factor in volatility changes in that it implies that the idiosyncratic part of stochastic volatility changes will disappear and have a common volatility process. They also conducted a Wald test, and rejected the null hypothesis of perfectly correlated volatility to conclude that cross-market hedging is imperfect.

Contrary to what has been presented, the use of the Wald and likelihood ratio tests, is inappropriate for the null hypothesis of perfectly correlated volatility, as the asymptotic distribution of the Wald test statistics is different from the conventional chi-squared distribution, as shown, for example, in Chernoff et al. [1954] . As the null hypothesis is on the boundary of the parameter space, the correlation estimator cannot be greater than one in absolute value, so that the distribution is asymmetric, and hence non-normal, when the true correlation coefficient is unity.

The paper proposes a new Lagrange multiplier test for the hypothesis that the volatility changes of a bivariate series are perfectly correlated. We use the framework of a multivariate stochastic volatility model proposed by Harvey et al. [1994], where the log-volatility follows vector autoregressive (VAR) process of order one with diagonal autoregressive coefficient matrix.

The Lagrange multiplier test principle is the only alternative for this problem in deriving the test statistics because it uses only the estimator of the unconstrained parameters, which are asymptotically normally distributed, and does not estimate the parameter on the boundary of

the parameter space. Then the test statistic will follow the conventional chi-squared asymptotic distribution under the null hypothesis.

To the best of our knowledge, the Lagrange multiplier test statistic for the perfectly correlated volatility changes has not been proposed in the literature. It follows that the hypothesis of frictionless cross-market hedging has also not been tested, so that a new test for perfectly correlated volatility would be useful from a practical perspective.

It is not without reason why an LM test has not been proposed to date as the conventional method to obtain a score function that is used in constructing the LM test statistic is unworkable for the multivariate stochastic volatility model.

The derivative of the transition density is intractable in this integral under the null hypothesis, as the transition disturbance has zero variance, and the transition equation density degenerates. We express the score function analytically with respect to the degenerate parameter using the ingenious method devised by Chesher [1984], which is the main technical breakthrough in tackling this problem.

The new test is a stochastic volatility version of the ARCH test proposed by Engle and Susmel [1993] to investigate whether international equity markets have a common volatility process. The test can be regarded as a test for the number of stochastic volatility factors, in line with the definition of Harvey et al. [1994] and Cipollini and Kapetanios [2008], when the number of factors is one under the null hypothesis. Cipollini and Kapetanios [2008] used a linearized model for the log of squared returns, and used the principal component methodology of Stock and Watson [2002] in deciding the number of factors. Their method has the advantage in that it is applicable when the number of variables is large, even though it is not a statistical test. The new test developed in this paper is a unique statistical test for the null hypothesis of the number of stochastic volatility factors.

Although theoretically straightforward, a generalization to multi-factor models is left to for further research, as numerical calculation of the test statistic is extremely time consuming, even in the simple case given here. The bottleneck lies in the calculation of score functions by the conventional smoothing algorithm. Maximum likelihood estimation by means of the quadrature method proposed by Watanabe [1999] is efficient when the state variable is univariate. Moreover, Monte Carlo simulation methods would work well for the estimation of multivariate state space models, with some difficulty, as the filtering algorithm, which is required for estimation, is far faster than the smoothing algorithm. It should be possible to generalize the results in the paper to the multi-factor model by improving the smoothing algorithm, as well as by hardware advancement.

The remainder of the paper is as follows. Section 2 presents the model, Section 3 develops the LM test statistic, Monte Carlo experiments are presented in Section 4, the empirical analysis

4

is given in Section 5, and some concluding remarks are in Section 6, followed by the Appendices.

2   Model

Under the alternative hypothesis, the observation vector $y_t = (y_{1t}, y_{2t})'$ can be expressed as:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \exp(\frac{h_1}{2}) & 0 \\ 0 & \exp(\frac{h_2}{2}) \end{pmatrix} \begin{pmatrix} a_1 & 0 \\ a_2 & a_3 \end{pmatrix} \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}, t = 1, \cdots, T, \tag{1}$$

where the log-volatility, $(h_{1t}, h_{2t})'$, follows a stationary bivariate autoregressive process of order one, defined by:

$$\begin{pmatrix} h_{1t} \\ h_{2t} \end{pmatrix} = \begin{pmatrix} \rho & 0 \\ 0 & \psi \end{pmatrix} \begin{pmatrix} h_{1t-1} \\ h_{2t-1} \end{pmatrix} + \begin{pmatrix} b_1 & 0 \\ b_2 & \sqrt{c} \end{pmatrix} \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}, t = 1, \cdots, T, \tag{2}$$

and

$$\begin{pmatrix} h_{11} \\ h_{21} \end{pmatrix} = \begin{pmatrix} b_1/\sqrt{(1-\rho^2)} & 0 \\ b_2/\sqrt{(1-\psi^2)} & \sqrt{c/(1-\psi^2)} \end{pmatrix} \begin{pmatrix} u_{11} \\ u_{21} \end{pmatrix},$$

$$(e_{1t}, e_{2t}, u_{1t}, u_{2t})' \sim NID\left(\mathbf{0}, \mathbf{I}_4\right).$$

The disturbance term of the measurement equation (1) is assumed to be contemporaneously uncorrelated, and the transition equation (2) is assumed to be contemporaneously correlated. This model was originally suggested by Harvey et al. [1994], and was examined in detail in Danielsson [1998] and Asai et al. [2006].

It is easy to see that, under the null hypothesis defined by:

$$\mathbf{H}_0 : \sqrt{c} \to 0, \psi = \rho, b_2 = b_1,$$

we have:

$$h_{1t} \equiv h_{2t}, \text{ for any } t.$$

This is the stochastic volatility factor model discussed in Harvey et al. [1994] and Cipollini and Kapetanios [2008] in the simple case when the number of factors is one, that is , $h_{1t} = h_{2t}$.

3   LM Test statistic

We propose the LM test for the null hypothesis for the observation series y_{1t} and y_{2t}. For the sake of computational simplicity and parameter parsimony, we set $a_1 = a_2 = 1$ by standardizing the returns variance. Define the unrestricted parameter vector as $\theta_1 = (c, \psi, b_2, \rho, b_1)$, and the restricted parameter vector as $\theta_0 = (0, \psi, b_2, \psi, b_2)$.

First, I obtain the maximum likelihood estimator of the constrained parameter, θ_0, of the state space system, (1) and (2), at time t. Denote $\mathbf{y}_1 = (y_{11}, y_{12}, \ldots, y_{1t})'$, $\mathbf{y}_2 = (y_{21}, y_{22}, \ldots, y_{2t})'$, $\mathbf{h}_1 = (h_{11}, h_{12}, \ldots, h_{1t})'$, and $\mathbf{h}_2 = (h_{21}, h_{22}, \ldots, h_{2t})'$. The likelihood function is expressed as:

$$f(\mathbf{y}) = \int f(\mathbf{h}, \mathbf{y}) d\mathbf{h} = \int f(\mathbf{y}_1, \mathbf{y}_2 | \mathbf{h}_1, \mathbf{h}_2) f(\mathbf{h}_2 | \mathbf{h}_1) f(\mathbf{h}_1) d\mathbf{h}_1 d\mathbf{h}_2,$$

where the specified densities $f(\mathbf{y}_1 | \mathbf{h}_1)$, $f(\mathbf{y}_2 | \mathbf{y}_1, \mathbf{h}_1, \mathbf{h}_2)$, $f(\mathbf{h}_2 | \mathbf{h}_1)$, and $f(\mathbf{h}_1)$ are given in the Appendix.

Second, I derive the score function under the alternative hypothesis, and evaluate it under the null hypothesis. Denote:

$$\mathbf{y}_{1:t} = (y_{11}, y_{12}, \ldots, y_{1t}, y_{21}, y_{22}, \ldots, y_{2t})', \ \mathbf{y}_t = (\mathbf{y}_1, \mathbf{y}_2), \ f_t = f(\mathbf{y}_t),$$

and the score function as:

$$\frac{\partial \log f_t}{\partial \theta_1} = \left(\frac{\partial \log f_t}{\partial c}, \frac{\partial \log f_t}{\partial \psi}, \frac{\partial \log f_t}{\partial b_2}, \frac{\partial \log f_t}{\partial \rho}, \frac{\partial \log f_t}{\partial b_1} \right).$$

Note that $\log f(\mathbf{y}_t | \mathbf{y}_{1:t-1}) = \log f(\mathbf{y}_{1:t}) - \log f(\mathbf{y}_{1:t-1})$. Define the conditional score function as:

$$Q_t = \frac{\partial \log f(\mathbf{y}_t | \mathbf{y}_{1:t-1})}{\partial \theta_1'} = \frac{\partial \log f(\mathbf{y}_{1:t})}{\partial \theta_1'} - \frac{\partial \log f(\mathbf{y}_{1:t-1})}{\partial \theta_1'}.$$

Then the Fisher information matrix can be expressed as:

$$\mathcal{I}(\theta) = \frac{1}{T} \sum_{t=1}^{T} Q_t Q_t',$$

and the full information score function is given as:

$$\mathcal{U}(\theta) = \frac{1}{T} \sum_{t=1}^{T} Q_t = \frac{1}{T} \frac{\partial \log f(\mathbf{y}_{1:T})}{\partial \theta_1'},$$

which is evatulated at $\theta_1 = \theta$.

Define the LM test statistic as:

$$LM = T \mathcal{U}'(\hat{\theta}_0) \mathcal{I}(\hat{\theta}_0)^{-1} \mathcal{U}(\hat{\theta}_0) \xrightarrow{L} \chi^2(3),$$

where $\hat{\theta}_0$ is the maximum likelihood estimator of θ_0 under the null hypothesis. The asymptotic distribution has three degrees of freedom corresponding to the three restrictions under the null hypothesis.

As $\frac{\partial \log f(y_{1:t})}{\partial \theta_1'} = \frac{1}{f(y_{1:t})} \frac{\partial f(y_{1:t})}{\partial \theta_1'}$, I focus on how to obtain $\frac{\partial f(y_{1:t})}{\partial \theta_1'}$. A problem is that the score function derived in the usual way diverges as the parameter c approaches 0, so I use

the method proposed in Chesher [1984]. Denote $y_{1:t}$ as $\mathbf{y}$. The score functions with respect to each parameter under the null hypothesis are given below:

$$
\begin{aligned}
\frac{\partial \log f(\mathbf{y})}{\partial c}|_{\mathbf{H}_0} &= tr E_{\mathbf{h}_1|\mathbf{y}}(\mathbf{J}_t), \\
\frac{\partial \log f(\mathbf{y})}{\partial \psi}|_{\mathbf{H}_0} &= \frac{1}{2}\left[\mathbf{1}_{1\times t}\mathbf{V}_\rho^{1/2}\mathbf{Z}_\rho E_{\mathbf{h}_1|\mathbf{y}}[\mathbf{h}_1]\right] - \frac{1}{2}tr\left[\mathbf{Y}_2\mathbf{V}_\rho^{1/2}\mathbf{Z}_\rho E_{\mathbf{h}_1|\mathbf{y}}[\mathbf{h}_1\exp(-\mathbf{h}_1')]\right], \\
\frac{\partial \log f(\mathbf{y})}{\partial b_2}|_{\mathbf{H}_0} &= -\frac{1}{2b_1}\mathbf{1}_{1\times t}E_{\mathbf{h}_1|\mathbf{y}}\left[\mathbf{h}_1\right] + \frac{1}{2b_1}tr\left[\mathbf{y}_2^{2'}E_{\mathbf{h}_1|\mathbf{y}}\left[\exp(-\mathbf{h}_1)\circ\mathbf{h}_1\right]\right], \\
\frac{\partial \log f(\mathbf{y})}{\partial \rho}|_{\mathbf{H}_0} &= -\frac{\partial \log f(\mathbf{y})}{\partial \psi}|_{\mathbf{H}_0} + \left[-\frac{\rho}{1-\rho^2} - \frac{1}{2}b_1^{-2}tr\left(\frac{\partial \mathbf{V}_\rho^{-1}}{\partial \rho}E_{\mathbf{h}_1|\mathbf{y}}\left(\mathbf{h}_1\mathbf{h}_1'\right)\right)\right], \\
\frac{\partial \log f(\mathbf{y})}{\partial b_1}|_{\mathbf{H}_0} &= -\frac{\partial \log f(\mathbf{y})}{\partial b_2}|_{\mathbf{H}_0} - \frac{t}{b_1} + \frac{1}{b_1^3}tr\left(\mathbf{V}_\rho^{-1}E_{\mathbf{h}_1|\mathbf{y}}\left(\mathbf{h}_1\mathbf{h}_1'\right)\right),
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{J}_t &= \frac{1}{8}\left\{-2\left(\mathbf{1}_{1\times t}\mathbf{V}_\rho\mathbf{Y}_2\exp(-\mathbf{h}_1)\right) + \left(\mathbf{1}_{t\times t}\mathbf{V}_\rho\right)\right. \\
&\quad + \left.\left(\mathbf{V}_\rho\mathbf{Y}_2\exp(-\mathbf{h}_1)\exp(-\mathbf{h}_1')\mathbf{Y}_2\right) - 2\left(\mathbf{V}_\rho\mathbf{Y}_2\mathbf{H}_1^{-1}\right)\right\}.
\end{aligned}
$$

Further details are given in the Appendix.

4   Monte Carlo Experiments

In order to confirm that the proposed new test statistic is asymptotically distributed as $\chi^2(3)$ under the null, and whether it has power to reject a false null hypothesis, I conduct two Monte Carlo experiments, as given below.

4.1   Asymptotic distribution

This experiment is to generate samples drawn under the null hypothesis, H_0, calculates the new test statistic, and obtains the empirical distribution of the test statistic.

In this test, the significance level corresponds to the probability of the rejection region for the upper-tailed distribution of $\chi^2(3)$, so the experiment calculates the rejection rate that is larger than the theoretical critical value. Using the calculated statistics, I obtain the empirical distribution of the statistic and use kernel estimation or a simple histogram to show that it follows the asymptotic $\chi^2(3)$ distribution.

First, I generate samples drawn from different null hypotheses, particularly for different values of the parameter, ψ, which is the autocorrelation coefficient of the state variable. The rejection rates correspond to different critical values, and are shown in Table 1. The parameter vector,

Table 1: Rejection rates under the null hypothesis

| | Sampling from H_0 | | | | Rejection rates | | | | | |
| | | | | | $T=100$ | | $T=200$ | | $T=500$ | |
c	ψ	b_2	ψ	b_2	5%	1%	5%	1%	5%	1%
0	0.7	1	0.7	1	10.2%	4.3%	8.3%	2.7%	7.1%	1.3%
0	0.9	1	0.9	1	21.0%	9.5%	13.1%	4.8%	6.3%	1.7%
0	0.95	0.45	0.95	0.45	22.5%	10.2%	14.7%	5.6%	7.4%	1.6%

θ, follows the same definition as in the previous section, namely $\theta_0 = (0, \psi, b_2, \psi, b_2)$. As the time length T increases, the rejection rate converges to the theoretical significance level.

From the table, I can see that the rejection rate converges sufficiently well when the time length is 500, which suggests that I should use data with at least 500 observations in practice. The histogram of the samples statistics obtained when $T = 500$ is given in the Appendix.

4.2   Statistical power

In comparison with the previous experiments, I generate data drawn from the alternative hypothesis, H_1, and calculate the rejection rates to see whether the statistic has power to reject a false null hypothesis. The parameter vector under H_1 shifts from the parameter vector under H_0. The Monte Carlo results are shown in Table 2, where the vector of parameters under the null hypothesis is given as $\theta_0 = (0, 0.7, 0.32, 0.7, 0.32)$.

Table 2: Rejection rates under the alternative hypothesis

| | Sampling from H_1 | | | | Rejection rates | |
| | | | | | $T=500$ | |
c	ψ	b_2	ρ	b_1	5%	1%
0.32	0.7	0.32	0.7	0.32	28%	13%
0.45	0.7	0.32	0.7	0.32	72%	43%
0	0.7	0.25	0.7	0.32	24%	7%
0	0.7	0.19	0.7	0.32	51%	28%
0	0.5	0.32	0.7	0.32	14%	7%
0	0.9	0.32	0.7	0.32	86%	72%

Given the accurate finite sample rejection rates, it was felt reasonable to perform the Monte

Carlo simulations only 100 times.

5 Empirical Analysis

5.1 Data adjustment

Before using the LM test statistic, it is worth recalling that the error terms (e_{1t}, e_{2t}) in the measurement equation are mutually independent. However, as the usual situation is that they are contemporaneously correlated, I need to adjust the data to eliminate the correlation between y_{1t} and y_{2t}. Instead of using the original data, y_{1t} and y_{2t}, I use a linear combination of y_{1t} and y_{2t}. A similar approach is used in Engle and Kozicki [1993]. We illustrate the reason with the linear transformations, as follows.

First, under the null hypothesis, the measurement equation can be written simply as:

$$\begin{pmatrix} y_{1t}^{''} \\ y_{2t}^{''} \end{pmatrix} = \exp(\frac{h_1}{2}) \begin{pmatrix} a_1 & 0 \\ a_2 & a_2 \end{pmatrix} \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}.$$

It is easy to see that any linear operator applied to $(y_{1t}^{''}, y_{2t}^{''})$ does not change the state part, $\exp(\frac{h_1}{2})$, of the equation. If the null hypothesis is true, with linear transformation, I only alter the structure of the measurement noise, and the state part $\exp(\frac{h_1}{2})$ remains the same after the data adjustment.

Second, if the original data are drawn under the alternative hypothesis:

$$\begin{pmatrix} y_{1t}^{''} \\ y_{2t}^{''} \end{pmatrix} = \begin{pmatrix} \exp(\frac{h_1}{2}) & 0 \\ 0 & \exp(\frac{h_2}{2}) \end{pmatrix} \begin{pmatrix} a_1 & 0 \\ a_2 & a_3 \end{pmatrix} \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}, \tag{3}$$

then any linear transformation between $(y_{1t}^{''}, y_{2t}^{''})$ will retain the two "features" $\exp(\frac{h_i}{2})$, $i = 1, 2$, where the word "feature" is used in Engle and Kozicki [1993].

Finally, any linear combination between $(y_{1t}^{''}, y_{2t}^{''})$ has its own significant meaning in empirical finance. Notice that the original data, $(y_{1t}^{''}, y_{2t}^{''})$, denote the difference in the log-price, namely the financial returns of the assets:

$$\begin{aligned} y_{1t}^{''} &= \log(p_{1t}) - \log(p_{1t-1}), \\ y_{2t}^{''} &= \log(p_{2t}) - \log(p_{2t-1}). \end{aligned}$$

For example, consider two assets, S&P 500 and Nikkei 225, for which the linear combination of the return is the returns on the portfolio which shares the weights between S&P 500 and Nikkei 225. In this situation, I can also use the model to analyse two new assets which always contain two "features", which makes our test statistic useful for empirical analysis.

We need two steps to adjust the data, which are given as:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = a_0^{-1} \begin{pmatrix} y_{1t}' \\ y_{2t}' \end{pmatrix} = a_0^{-1}\Lambda^{-\frac{1}{2}} \begin{pmatrix} y_{1t}'' \\ y_{2t}'' \end{pmatrix}, \tag{4}$$

where estimation of Λ and a_0 are given by:

$$\hat{\Lambda} = \frac{1}{T}\begin{pmatrix} \sum y_{1t}''^2 & \sum y_{1t}'' y_{2t}'' \\ \sum y_{1t}'' y_{2t}'' & \sum y_{2t}''^2 \end{pmatrix}, \tag{5}$$

$$\hat{a_0} = \begin{pmatrix} \exp\left\{\left[\frac{1}{T}\left(\sum \log y_{1t}'^2\right) + 1.27\right]/2\right\} & 0 \\ 0 & \exp\left\{\left[\frac{1}{T}\left(\sum \log y_{2t}'^2\right) + 1.27\right]/2\right\} \end{pmatrix}. \tag{6}$$

If:

$$a_0^{-1}\Lambda^{-\frac{1}{2}} \xrightarrow{p} \begin{pmatrix} a_1 & 0 \\ a_2 & a_3 \end{pmatrix}^{-1}, \tag{7}$$

I can obtain a new data set, (y_{1t}, y_{2t}), that has the same distribution compared with the model that is used in the new test statistic, with the same asymptotic null distribution. Further adjustments are shown in the Appendix.

5.2 Empirical result

Using the proposed new statistical test in the previous section, I examine the relationship between different stock markets, and also investigate the exchange rate movements over different time periods.

5.2.1 Analysis of stock markets

First, I investigate whether there exists a common factor of volatility between different stock market indices. The data I use for analysis come from Yahoo finance, and I use the Adjusted-Close price p_t. The stock market indices list is given below.

We obtained daily data from January, 2011 to December, 2014 and separated them into two sets to check the performance in different years. The test needs a combination of two indices, so there are 28 pairs. We excluded data whenever there were closed-market days in one market, such as holidays.

The data adjustment follows the two steps given in the previous section. The parameter estimates are shown in the Appendix as Tables 15,16,17,18, and the test outcomes for different pairs are shown in Tables 4 and 5.

As can be seen from the tables, even for different time periods, the group that contains the FTSE, STI and AORD stock markets share the same volatility factor. Stock markets in the

Table 3: Stock market indices

Index number	Stock Market	Symbol	Country/Region
1	Dow Jones Industrial Average	**DOW**	United State
2	FTSE Index	**FTSE**	Unite Kingdom
3	DAX Index	**DAX**	Germany
4	Shanghai Composite Index	**SSCI**	China
5	Nikkei 225 Stock Average Index	**NIKKEI**	Japan
6	Hang Seng Index	**HSI**	Hong Kong
7	Straits Times Index	**STI**	Singapore
8	All Ordinaries Index	**AORD**	Australia

Table 4: LM statistics between markets from 2011 to 2012

y_1 \ y_2	DOW	FTSE	DAX	SSCI	NIKKEI	HSI	STI	AORD
DOW		9.1*	9.19*	21.48**	24.22**	5	10.44*	4.07
FTSE	5.14		1.39	20.33**	25.33**	3.66	3.74	1.07
DAX	5.96	8.7*		17.84**	30.19**	2.52	6.22	6.82
SSCI	21.07**	7.63	15.72**		18.07**	4.7	7.92*	8.46*
NIKKEI	22.15**	8.88*	29.5**	15.23**		2.34	10.23*	5.22
HSI	4.9	2.73	5.88	4.14	40.52**		2.87	4.69
STI	7.69	2.99	7.03	3.17	34.31**	18.12**		1.98
AORD	2.46	5.56	12.88**	5.19	69.35**	15.88**	2.68	

Note: * significant at 5% level, ** significant at 1% level.

Asian region (such as China and Japan) appear to have a unique factor compared with other regions during 2011 and 2012. However, in 2013 and 2014, there appear to be more groups that share the same factor, namely:

$$Group\ 1\ :\quad FTSE,\ SSCI,\ NIKKEI$$
$$Group\ 2\ :\quad FTSE,\ HSI,\ AORD$$
$$Group\ 2\ :\quad FTSE,\ STI,\ AORD$$

Table 5: LM statistics between markets from 2013 to 2014

y_1 \ y_2	DOW	FTSE	DAX	SSCI	NIKKEI	HSI	STI	AORD
DOW		0.39	8.05*	11.83**	15.43**	9.24*	4.97	5.21
FTSE	15.79**		9.9*	2.28	4.76	4.17	1.92	2.68
DAX	4.34	7.52		10.64*	21.47**	3.31	7.83*	10.22*
SSCI	11.53**	2.14	10.72*		3.39	12.84**	3.79	6.11
NIKKEI	15.05**	5.68	25.2**	2.58		9.92*	9.78*	9.88*
HSI	7.88*	1.54	7.13	7.47	12.39**		19.45**	2.58
STI	11.42**	1.91	23.23**	7.08	7.13	8.32*		1.63
AORD	6.11	0.8	8.03*	2.31	5.71	3.99	4.13	

Note: * significant at 5% level, ** significant at 1% level.

5.2.2   Analysis of exchange rates

We also use foreign exchange rates to evaluate the performance during extreme situations, especially when volatility is higher than usual. Comparing the performance for two time periods, namely the global financial crisis and normal times, I focus on the rates that are aggregated from the table instead of a single result between only two currency pairs.

First, I define two time periods representing the financial crisis and normal times. We use the Chicago Board Options Exchange (CBOE) Volatility Index (VIX) as an indicator to detect when volatility is high. It is easy to see that volatility changed from the historical chart (see Figures 1 and 2). We choose Period 1: Oct/1/2008 ∼ Oct/31/2008 as the financial crisis, and Period 2: Oct/1/2012 ∼ Oct/31/2012 as normal times.

Figure 1: VIX during global financial crisis (2008)

Period 1:
Oct/1/2008~Oct/31/2008
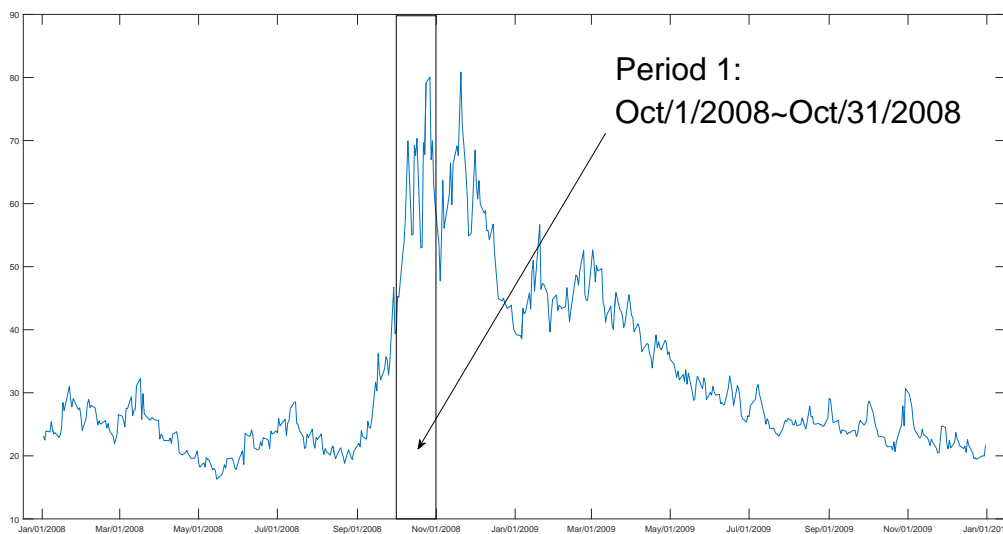
Note: VIX data are downloaded from Yahoo Finance.

Figure 2: VIX during normal times (2012)

Period 2:
Oct/1/2012~Oct/31/2012
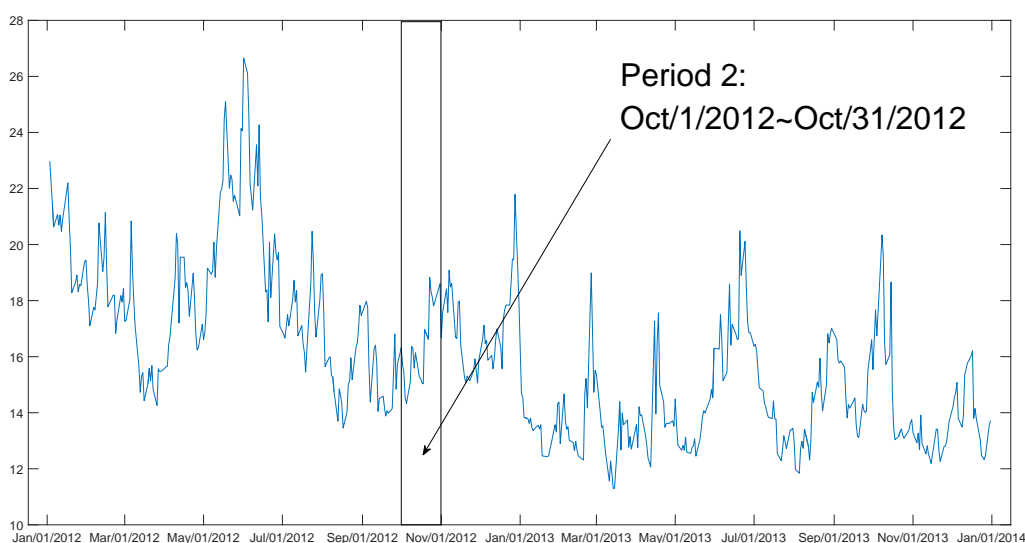
Note: VIX data are downloaded from Yahoo Finance.

Second, I use 6 major currency pairs that are traded widely. The pairs used are listed in Table 6, and all contain USD, so that all currencies are priced in USD.

We obtained hourly data for a month, which means roughly 500 data series. The estimated parameters are listed in the Appendix as Tables 19,20,21,22, and the statistical results are

Table 6: Currency Pairs list

Paris	EUR/USD	USD/JPY	GBP/USD	AUD/USD	USD/CHF	USD/CAD

shown in Tables 7 and 8.

Table 7: LM statistics for exchange rates during Period 1

y_1 \ y_2	EURUSD	USDJPY	GBPUSD	AUDUSD	USDCHF	USDCAD
EURUSD		23.73**	12.06**	34.09**	10.43*	17.89**
USDJPY	35.45**		17.28**	15.62**	28.23**	54.5**
GBPUSD	14.49**	14.34**		53.78**	18.74**	22.03**
AUDUSD	37.32**	33.16**	30.01**		25.47**	28.3**
USDCHF	24.47**	33.96**	43.9**	39.49**		23.63**
USDCAD	18.31**	37.44**	15.14**	19.02**	14.64**	

 Note: * significant at 5% level, ** significant at 1% level. The exchange rate is downloaded from FXDD's historical data.

Table 8: LM statistics for exchange rates during Period 2

y_1 \ y_2	EURUSD	USDJPY	GBPUSD	AUDUSD	USDCHF	USDCAD
EURUSD		11.27*	5.17	7.56	6.61	25.69**
USDJPY	15.88**		28.67**	12.97**	16.73**	14.86**
GBPUSD	2.15	23.74**		20.35**	5.39	8.38*
AUDUSD	28.07**	18.32**	26.51**		14.97**	22.86**
USDCHF	3.22	9.53*	4.64	6.9		17.66**
USDCAD	5.5	6.47	4.83	5.82	4.41	

 Note: * significant at 5% level, ** significant at 1% level. The exchange rate is downloaded from FXDD's historical data.

As can be seen from Tables 7 and 8, during the financial crisis volatility is larger than usual, so it is difficult to find a single asset to hedge volatility. The accepted rate is given as 0 in

Table 7. Conversely, it is easy to find currency pairs that potentially share the same volatility factor during normal times. The accepted rate is given as 43.3% in Table 8.

6   Conclusion

In this paper, I considered whether financial returns has a common volatility process in the framework of stochastic volatility models, and proposed a Lagrange Multiplier test statistic for the null hypothesis that the volatility changes of a bivariate series are perfectly correlated. It is useful in investigating the correlation between different markets, even for frictionless cross-market hedging.

In the empirical analysis of stock markets, I found some groups that potentially share common time-varying volatility, especially markets for the United Kingdom, Singapore and Australia. We also investigated the correlations between different major currencies when big events, such as financial crises, occurred. The empirical analysis suggested that, during high volatility periods, it is more difficult to find a common factor between currencies, compared with low volatility periods, so that it is harder to hedge with different currencies.

However, the approach adopted in the paper it is the simplest case of a multiple stochastic volatility model. The extension to a multi-factor model, even stochastic volatility with a fat-tailed distribution of the test, is left for further research.

Part II

Models with dynamic stable distribution

7   Introduction

Stable distribution is a consequently result of generalized central limit theory, which is the sum of some not well-define random variables (see Gnedenko et al. [1954]). It is a extension of normal distribution. Compared to normal distribution which is a special case of families, it allow the model to measure on heavy tails , Leptokurtic and skewness, which is important in financial risk measurement. In the Appendix, I list the definition and some important property of stable distribution. There is more detail description of stable distribution in Chapter 3 of Rachev [2003]. I use matlab code from Mark S. Veillette(Alpha-Stable distributions CODE).

There are many articles that using stable distribution, such as Liu and Brorsen [1995] (ML in GARCH with stable innovation) and Mittnik et al. [1998](ARMA with stable distribution). Instead of using stable distribution directly to model financial data, I use state space model with stable distribution to measure the data. In Stuck [1978], he describe such a model, and derived it with a linear filter similarly to Kalman filter. Also, there is continuous time version filter and smoother in Le Breton and Musiela [1993]. Compared to that they use symmetric distribution to establish the filter, I modify it with asymmetricity to measurement equation innovation. Unlike the normal distribution, the posterior distribution of normal distribution is conjugated to itself, the posterior of the asymmetric stable distribution is not analysis expressible. In another word, the linear filter is not the best filter compared to nonlinear Bayesian filter.

During to that, the stable distribution does not have analytical expression of density function, I should evaluate the density function in numerical method. This work is done in Nolan [1997]. We can also draw stable random variable with the method proposed in Chambers et al. [1976]

Beside the fact that stable is the result of central limit theory, there are several more reasons that I suggest to use stable distribution instead of normal distribution. First, it is safe for risk management, and useful in option pricing. We usually underestimate the probability of rare event, such as financial crisis. Second, α is also well known as tail index. The most common tail index estimator is the Hill estimator (see Hill et al. [1975]). However, they only use sub-sample of the data. According to empirical data, it is easy to find that volatility or second moment changes over time. It is nature to use fat tail distribution such as stable distribution, which does not have finite second moment instead of applying more assumption such as stochastic volatility. Third, it is easy to set up asymmetricity or skewness, leptokurtic and kurtosis with

stable distribution.

An advance that using state space model is that, I can dynamically measure the distribution, using the past information. The more detail describe of state space model can be found in Durbin and Koopman [2012]. Because it is univariate state model, I use simple numerical integration and Sequential Monte Carlo integration. The procedure can be found briefly in Doucet and Johansen [2009].

We use this model to analysis the exchange rate pair and stock markets. Compare the estimation result to normal distribution and independent stable distribution with Bayesian information Criteria (BIC). The result shows that stable distribution show far more suitable compared to normal distribution, and dynamic stable distribution is more acceptable that independent stable distribution. We also use the conditional distribution to estimate the Value at Risk (VaR). Obviously, it is much safer to use fat-tailed distribution during financial crisis.

8   Definition of Stable distribution

A stable distribution random variable has the characteristic function as

$$\psi(t) = E(e^{itx}) = \exp[it\delta - |\gamma t|^\alpha (1 - i\beta sgn(t)\Phi)]$$

where $\Phi = \tan(\pi\alpha/2)$ for $\alpha \neq 1$ or $\Phi = -\frac{2}{\pi}\log|t|$ for $\alpha = 1$.

α is the index parameter. β is the skewness parameter which is the measure of asymmetry. δ is the location parameter that play the same role compared to the mean parameter μ in normal distribution. γ is the scaling parameter similar to standard deviation parameter σ in normal distribution. The more detail describe about this distribution can be found in Rachev [2003]. Some properties are presented in Appendix.

In this article, I focus on distribution that the index parameter α range from $(1, 2)$, because in reality, asset's log-return can be considered as a martingale sequence which means expected return is 0.

9   Models

Model 1

We simply use i.i.d distribution to model the real data. This is

$$Y_t = \eta_t, \tag{8}$$

where, η_t is distributed as normal distribution or stable distribution. Y_t is obtained as

$$Y_t = \log(P_t) - \log(P_{t-1}).$$

Model 2

In finance, the phenomenon of large log-return is tend to be followed by large log-return, is called as volatility clustering in Mandelbrot [1963]. The usual way to model this feature is to use the ARCH or GARCH model. The stable version of these model can be find in **?**. In this article, I use the stochastic volatility model as:

$$Y_t = \exp(\frac{h_t}{2})\eta_t, \tag{9}$$
$$h_t = \rho h_{t-1} + \epsilon_t. \tag{10}$$

Y_t is also the log-return, and h_t is the unobserved volatility factor. η_t distribute as normal or stable distribution. The more detail describe about this model can be found in Harvey et al. [1994].

Model 3

We conduct a dynamic model using the simplest state-space system.

$$Y_t = X_t + \epsilon_t, \tag{11}$$

$$X_t = X_{t-1} + \eta_t. \tag{12}$$

Y_t is the log-price of assets initialed at price P_0:

$$Y_t = \log(P_t) - \log(P_0). \tag{13}$$

X_t and η_t are all $\alpha - Stable$ distribution random variables with skewness or normal. ϵ_t distributes normal. We have:

$$\epsilon_t \sim N(0, \sigma_\epsilon), \tag{14}$$

$$\eta_t \sim S(\alpha, \beta_\epsilon, \gamma_\eta, 0) \ or \ N(0, \sigma_\eta). \tag{15}$$

One similar linear filter is introduced in Stuck [1978] for this model with both ϵ_t and η_t symmetric stable distribution. However, when introduce the asymmetry and normal distribution, the mean of the posterior $X|Y$ is no longer linear with the observation $Y = y$. This result can be easily shown in numerical computation. If $X \sim S(1.5, \beta, 1, 0)$ and $Y = X + \epsilon$, where $\epsilon \sim N(0, 1)$, I can easily calculate the posterior mean of $X|Y$ with different skewness parameter β.

Figure 3: Posterior Mean of X given Y
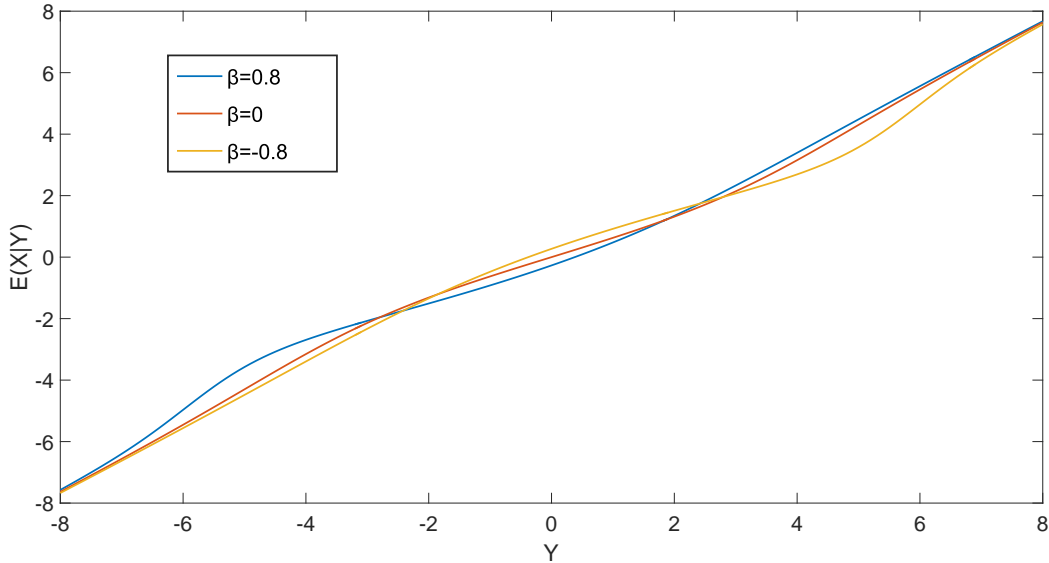
This result shows that the linear projection is not the best prediction of posterior random variable $X|Y$.

19

Model 4

Combine model 2 and model 3, I can obtain a hierarchical model that:

$$Y_t = X_t + \epsilon_t, \tag{16}$$

$$X_t = X_{t-1} + \exp(\frac{h_t}{2})\eta_t, \tag{17}$$

$$h_t = \rho h_{t-1} + \zeta_t. \tag{18}$$

Again, ϵ, ζ is distributed as normal and η is distributed as stable or normal.

The reasons I suggest using stable distribution are not only that it is fundamentally solid in central limited theorem, but also have realistic meaning in finance:

First, in finance, tail events means rare event, such as financial crisis or cooperation default. It is very important part of risk management and option pricing. To fit the data of log-return of exchange rate, I use kernel density estimation, normal distribution and stable distribution. The data is obtained from 03/04/2014 to 02/05/2016 as daily with exchange rate of euro against dollar. Comparison result is Figure 4.

The first reason is obvious that stable distribution describe the tail probability more suit than normal distribution.

Figure 4: Comparison between different density function
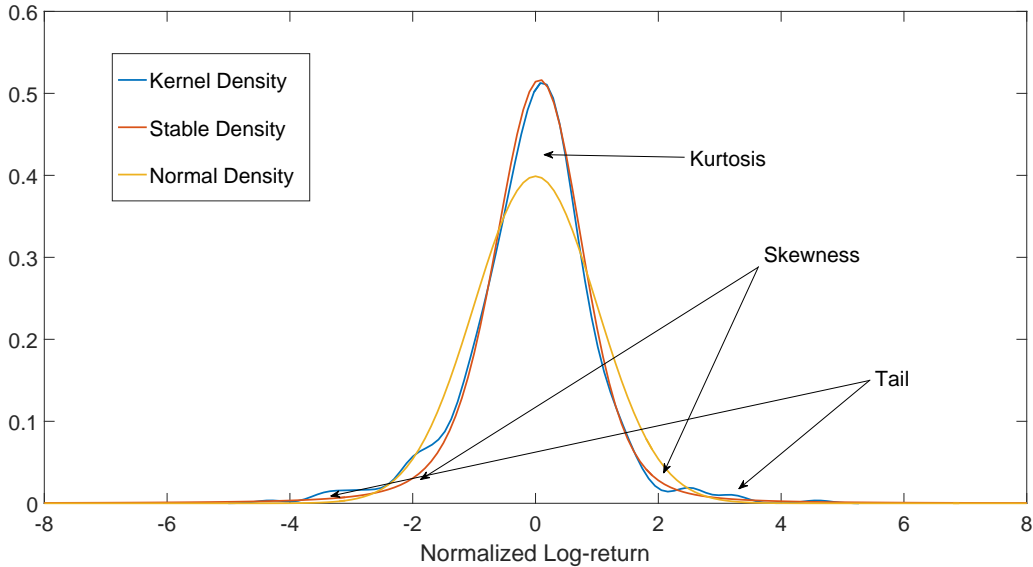
Second, the tail index α ranges from 1 to 2 so the higher moments bigger that 2 do not exist. We calculate the variance of 1 min log-return of exchange rate using rolling-over, though one day (1440 minutes). It is easy to find the implied variance of log-return changes over time. It can be explained as stochastic volatility, or simply as the variance does not exits. We also

simulate the Stable distribution and Normal distribution to calculate the rolling-over variance. The comparison between different data is listed in Appendix as Figure 5.

Third, asymmetry of the distribution becomes more and more important in finance, especially

Figure 5: Comparison with variance between different distribution
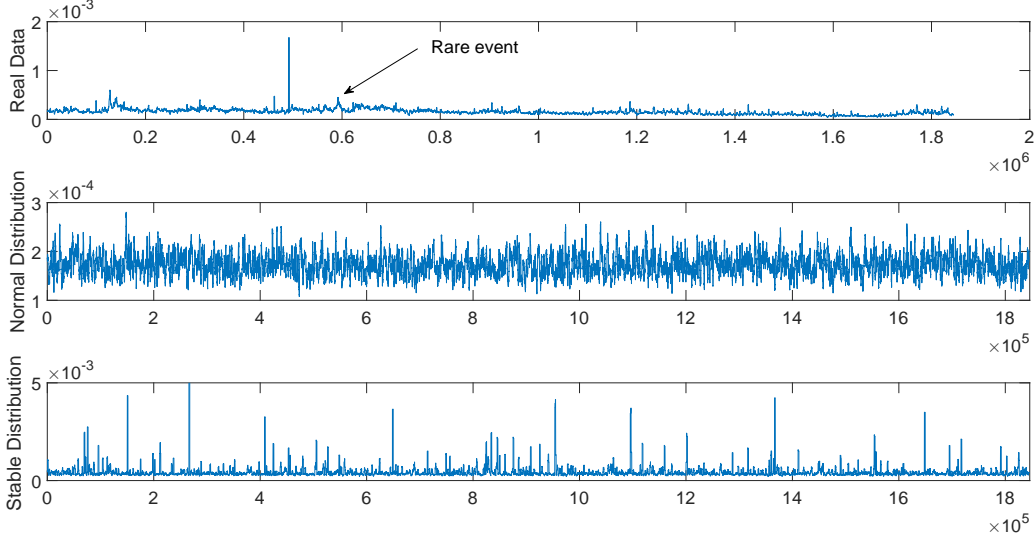
when financial crisis occurs. It is naturally contained in the parameterization in stable distribution. Also, the higher moment such as kurtosis diverse over time. These feature can also be found in Figure 4.

10   Likelihood

For Model 1, simple use the normal density or stable density as the likelihood function, and apply the maximize likelihood method to obtain the parameter estimation. It is:

$$(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = \arg\max \sum_{i=1}^{n} \log S(y_i|\alpha, \beta_y, \gamma_y).$$ (19)

For Model 2, 3 and 4, because they are general state-space model, I can obtain the likelihood in filtering step. The likelihood can be expressed as:

$$f(\mathbf{y}) = \prod_{t=1}^{T} f(y_i|y_{i-1})$$ (20)

where

$$f(y_i|y_{i-1}) = \int f(y_i|x_i)f(x_i|y_{i-1})dx_i$$ (21)

21

where x_i is the unobserved variable. In model 2, it is h_t. In model 3, it is X_t. In model 4, both X_t and h_t is unobserved.

the model detail about filtering process can be found in Durbin and Koopman [2012] and the contents about Sequential Monte Carlo can be found in Doucet and Johansen [2009].

About the initial point, I can first use the model 1 to estimate the tail index. Set that estimator value as the initial point for model 2, 3 and 4 for α. Other parameter can be also applied as the same process.

11 Empirical Analysis

We apply the different models with both normal distribution model and stable distribution model. We analysis the major currencies exchange rate. It is convenience to set mean of the variable to 0 because of the Efficient Market Hypothesis. The data I use is obtained from 03/04/2014 to 02/05/2016, the daily exchange rate of euro against dollar. We normalize data using the standard deviation of the log-return for computation convenience. We estimate model 1 with Maximize Likelihood method, and model 2,3 and 4 with both Maximize Likelihood method and MCMC Bayesian method. We also calculate the Bayesian information criterion of each model to compare the performance. Use the table in Kass and Raftery [1995], I decide which model is better.

Table 9: Evidence against higher BIC

ΔBIC	Evidence against higher BIC
0 to 2	Not worth more than a bare mention
2 to 6	Positive
6 to 10	Strong
> 10	Very Strong

Model 1

$$y_i \sim \eta_t$$

where η_t is distributed as normal or stable distribution. The estimated distribution density is shown in Figure 4 and the parameter result is:

Model 2

Table 10: Comparison with Normal and Stable distribution in Model 1

η_t	Stable	Normal
Estimated Parameters	$\alpha = 1.6431, \beta = -0.1655, \gamma = 0.5507$	$\sigma = 1$
BIC	1383	1425

This model is also targeted for log-return, so it is comparable with Model 1.

$$Y_t = \exp(\frac{h_t}{2})\eta_t,$$
$$h_t = \rho h_{t-1} + \epsilon_t.$$

The result is shown as:

Table 11: Comparison with Normal and Stable distribution in Model 2

η_t	Stable	Normal
Estimated Parameters	$\alpha = 1.7222, \beta = -0.1821, \gamma = 0.5700$	$\sigma_\eta = 0.9305$
(ML method)	$\sigma_\epsilon = 0.0747, \rho = 0.9820$	$\sigma_\epsilon = 0.2188, \rho = 0.9311$
BIC	1384	1396

Based on this two models, I can see that the stable distribution has always better performance that normal distribution.

Model 3

This model is target for price tracing.

$$Y_t = X_t + \epsilon_t,$$
$$X_t = X_{t-1} + \eta_t.$$

Y_t is the log-price of assets initialed at price P_0:

$$Y_t = \log(P_t) - \log(P_0).$$

Again, η can be chose between stable and normal. The estimate result is below:

Model 4

Add stochastic volatility to model, I have hierarchical model that:

$$Y_t = X_t + \epsilon_t,$$
$$X_t = X_{t-1} + \exp(\frac{h_t}{2})\eta_t,$$
$$h_t = \rho h_{t-1} + \zeta_t.$$

Table 12: Comparison with Normal and Stable distribution in Model 3

η_t	Stable	Normal
Estimated Parameters (ML method)	$\alpha = 1.6008, \beta = -0.1936, \gamma = 0.5093$ $\sigma_\epsilon = 0.1779$	$\sigma_\epsilon = 0.8718$ $\sigma_\eta = 0.3464$
BIC for ML	1383	1424
Estimated Parameters (MCMC method)	$\alpha = 1.6201, \beta = -0.1647, \gamma = 0.5024$ $\sigma_\epsilon = 0.2200$	analytical result using ML
BIC for MCMC	1384	1424

Again, η is distributed as stable or normal.

The estimate result is in Table 13

Based on the estimation I obtain, I also predict the Value at Risk using conditional distribution.

Table 13: Comparison with Normal and Stable distribution in Model 4

η_t	Stable	Normal
Estimated Parameters (ML method)	$\alpha = 1.7164, \beta = -0.1925, \gamma = 0.5012$ $\sigma_\epsilon = 0.1512, \rho = 0.9732, \sigma_\zeta = 0.0814$	$\sigma_\eta = 0.8904, \sigma_\epsilon = 0.2390$ $\sigma_\zeta = 0.2037, \rho = 0.9620$
BIC for ML	1384	1405
Estimated Parameters (MCMC method)	$\alpha = 1.7088, \beta = -0.2072, \gamma = 0.5454$ $\sigma_\epsilon = 0.1995, \rho = 0.9730, \sigma_\zeta = 0.1059$	$\sigma_\eta = 0.9511, \sigma_\epsilon = 0.2284$ $\sigma_\epsilon = 0.1814, \rho = 0.9638$
BIC for MCMC	1390	1404

12 Conclusion

We use stable distribution to fit both i.i.d models and dynamic models. We find that stable distribution is always outperform normal distribution. However, the dynamic structure in volatility change remain inflexible, so does the tail index α. We will focus on flexible parameter changes and adaptive process to improve the performance of model 4.

Figure 6: Value at Risk for Model 4 with Stable distribution
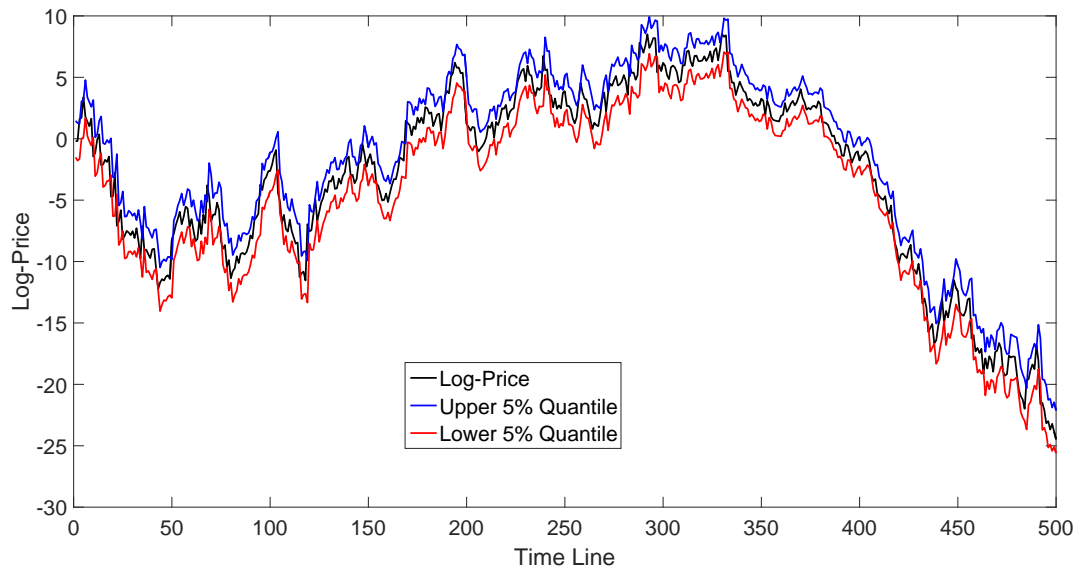
Part III

Fast algorithm in filtering and smoothing with Stochastic Volatility

13 Introduction

This part focus on the description of fast smoothing algorithm using Fast Gaussian transform (FGT) proposed in Greengard and Strain [1991]. It is one of the fast algorithm called the fast multipole method (FMM). It use the two asymptotic expansion: the near field expansion well known as Taylor expansion, and the far field multipole expansion named as Hermite expansion. There may be potential useful in high dimension particle filter and smoother , or multi-asset option pricing. Thanks for the C code provided by Sebastien PARIS LSIS/DYNI, Aix-Marseille University (AMU), UMR CNRS 7296.

13.1 Forward Filtering

Forward Filtering algorithm is widely used in State-Space model in many fields. Usually, I want to obtain likelihood of non-linear model, I will attempt apply the Sequential Monte Carlo method to calculate the likelihood. This method is also usually called as particle filter. The model detail can be found in Doucet and Johansen [2009].

Assuming that I get density of $f(x_t|y_t)$ and transition kernel $f(x_{t+1}|x_t)$, the filtering problem can be simplified as an integral such as:

$$f(x_{t+1}|y_t) = \int f(x_{t+1}|x_t)f(x_t|y_t)dx_t. \tag{22}$$

This integral can not be calculated efficiently under usual direct discretization method due to cross term $f(x_{t+1}|x_t)$.

13.2 Backward Smoothing

Backward smoothing algorithm is also widely used in State-Space model. Especially when come to statistic inference problem, I usually need to calculate the smoothed probability to obtain statistic. The more detail can also be found in Doucet and Johansen [2009].

In smoothing process, I want to obtain the probability or density function $f(x_{t-1}|y_{1:T})$ as:

$$f(x_{t-1}|y_{1:T}) = \int f(x_{t-1}|x_t, y_{1:T})f(x_t|y_{1:T})dx_t. \tag{23}$$

Assuming that I already obtain the density function $f(x_t|y_{1:T})$. Using Markovian equation $f(x_{t-1}|x_t, y_{1:T}) = f(x_{t-1}|x_t, y_{t-1})$ We can further write the equation as:

$$f(x_{t-1}|y_{1:T}) = \int f(x_{t-1}|x_t, y_{t-1})f(x_t|y_{1:T})dx_t \tag{24}$$

$$= \int f(x_{t-1}|x_t, y_{t-1})f(x_t|y_{1:T})dx_t \tag{25}$$

$$= \int \frac{f(x_t|x_{t-1}, y_{t-1})f(x_{t-1}|y_{t-1})}{f(x_t|y_{t-1})}f(x_t|y_{1:T})dx_t \tag{26}$$

$$= f(x_{t-1}|y_{t-1})\int \frac{f(x_t|x_{t-1})}{f(x_t|y_{t-1})}f(x_t|y_{1:T})dx_t \tag{27}$$

Assuming that I have already obtained the density function $f(x_t|y_{t-1})$ in filtering process. Then the core calculate problem can also be an integral as:

$$\int \frac{f(x_t|y_{1:T})}{f(x_t|y_{t-1})}f(x_t|x_{t-1})dx_t. \tag{28}$$

Again, the calculation problem is occurred by the cross term $f(x_t|x_{t-1})$.

13.3   Summary of Problem

We use the discretization method to rewrite the integral as summation of particles as:

$$\int q_t f(x_t|x_{t-1})dx_t = \sum_{i}^{n} q_t^{(i)} f(x_t^{(i)}|x_{t-1}^{(i)}) \tag{29}$$

where q_t is $f(x_t|y_t)$ in filtering problem and $\frac{f(x_t|y_T)}{f(x_t|y_{1:T})}$ in smoothing problem.
Furthermore, if the transition kernel is Gaussian kernel, the I can rewrite the integration as:

$$\int q_t e^{-\frac{(x_{t-1}-x_t)^2}{2\sigma}}dx_t = \sum_{i}^{n} q_t^{(i)} e^{-\frac{(x_{t-1}-x_t^{(i)})^2}{2\sigma}}, \tag{30}$$

the problem is I need to calculate the transition density $e^{-\frac{(x_{t-1}-x_t)^2}{2\sigma}}$ many times , in briefly the whole field to obtain the integration for each X_{t-1}. So if I got N x_t's and M x_{t-1}'s, I need calculate the density $M*N$ times.

14   Solution

14.1   The Common solution: Fast Fourier Transform (FFT)

This integral is also known as convolution of two function. The common way to calculate this integral is using Fast Fourier Transform (FFT) to reduce the calculation to $O(N\log N)$ if

$M = N$. The formal explanation of FFT is introduced in Cooley and Tukey [1965]. If I denote two function as $f_1(y-x)$, $f_2(x)$, and I want to obtain $f(y)$ as the convolution of these functions:

$$f(y) = \int f_1(y-x)f_2(x)dx. \tag{31}$$

Apply Fourier transform to function $f(y)$:

$$\begin{align}
\mathcal{F}(f(y)) &= \int e^{-ity} \int f_1(y-x)f_2(x)dxdy \tag{32}\\
&= \int e^{-it(y-x+x)} \int f_1(y-x)f_2(x)dxd(y-x) \tag{33}\\
&= \int e^{-it(y-x)}f_1(y-x)d(y-x) \int e^{-itx}f_2(x)dx \tag{34}\\
&= \mathcal{F}(f_1(y^*))\mathcal{F}(f_2(x)). \tag{35}
\end{align}$$

This means I only have to obtain the Fourier transform to each function. Then take the inverse Fourier transform to obtain the function $f(y)$:

$$f(y) = \mathcal{F}^{-1}(\mathcal{F}(f(y))) \tag{36}$$

In practice, for filtering or smoothing, assuming I have points $(y-x)^i$ and x^i for $i = 1 \cdots N$ with same constant interval, I only have to calculate the vector $f_1((y-x)^i)$ and $f_2(x^i)$, and add $N-1$ zeros to the vector to make the length equals to $2N-1$. Then transform them into Fourier vector with dimension $2N-1$, take Hadamard product(element by element) with each other to get the new Fourier vector $\mathcal{F}(f(y))$. Finally, I take the inverse Fourier transform to obtain the original probability vector $f(y)$. Notice that I should normalize every probability vector.

Beside the FFT, I will learn to use the other method called Fast Gaussian Transform (FGT).

14.2   the description of Fast Gaussian Transform

The problem I introduced above can be concluded as the integration with the Gaussian field as:

$$G(x) = \sum_{j=1}^{N} q_j e^{-\frac{|x-s_j|^2}{\delta}}, \tag{37}$$

which the notation is cited from Greengard and Strain [1991]. We call the x as target and s_j as source. Suppose I want to evaluate Gaussian kernel in Equation.37 at target t_i with source s_j,

$$G_{ij} = e^{-\frac{|t_i-s_j|^2}{\delta}}, \tag{38}$$

and I have M targets and N sources, then the direct evaluation will need $M * N$ calculation of Gaussian function. However, use the FGT, only need $O(M + N)$ calculation.

14.2.1 Hermite expansions and Taylor expansions

This section introduce Hermite expansions of the Gaussian kernel which is used in Fast Gaussian Transform.

First is Hermite polynominals is defined as:

$$H_n(t) = (-1)^n e^{t^2} D^n e^{-t^2}, \tag{39}$$

where $D = d/dt$. The Hermite polynomials then can write as:

$$e^{2ts-s^2} = \sum_{n=0}^{\infty} \frac{s^n}{n!} H_n(t). \tag{40}$$

Multiply both side with e^{-t^2} and I have:

$$e^{-(t-s)^2} = \sum_{n=0}^{\infty} \frac{s^n}{n!} h_n(t), \tag{41}$$

where Hermite functions $h_n(t)$ are defined as:

$$h_n(t) = e^{-t^2} H_n(t). \tag{42}$$

The shifted and scaled version of Hermite expansion at s_0 with scaling δ for Gaussian function is:

$$
\begin{aligned}
e^{-(t-s)^2/\delta} &= e^{-((t-s_0)-(s-s_0))^2/\delta} \\
&= \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{s-s_0}{\sqrt{\delta}}\right)^n h_n\left(\frac{s-s_0}{\sqrt{\delta}}\right) \\
&= e^{-(t-s_0)^2} \sum_{n=0}^{\infty} \frac{1}{n!} \left(\frac{s-s_0}{\sqrt{\delta}}\right)^n H_n\left(\frac{s-s_0}{\sqrt{\delta}}\right).
\end{aligned}
\tag{43}
$$

This formula allows us to evaluate the Gaussian field $e^{-(t-s)^2/\delta}$ as Hermite expansion centered at s_0, which is very important in this fast algorithm. Also, I can rewrite the Gaussian field as:

$$e^{-(t-s)^2/\delta} = \sum_{n=0}^{\infty} \frac{1}{n!} h_n\left(\frac{s-t_0}{\sqrt{\delta}}\right) \left(\frac{t-t_0}{\sqrt{\delta}}\right)^n. \tag{44}$$

This formula is usually called Taylor expansion at t_0 near the target t. Thus I call Formula.44 near-field expansion. The formula.43 is call far-field expansion because of expansion centered point s_0 is far from target point t. Also, there are many article call the Hermite expansion as Singular expansion and the Taylor expansion as Regular expansion. These are the names in the Fast Multipole Method (FMM) algorithm. We also need recurrence relation between Hermite function:

$$h_{n+1}(t) = 2t h_n(t) - 2n h_{n-1}(t). \tag{45}$$

In the original paper, there are bound condition about the error from both expansion, and also multi-dimension expansion with Gaussian function. However. I omit these statements and focus on the explanation of algorithm.

14.3   Taylor expansion of Hermite expansion

Follow the original paper's notation. Denote Target point t lies in Box C with center t_C and side length $r\sqrt{2\delta}$. N_B sources as s_j lies in Box B with center s_B and side length $r\sqrt{2\delta}$. Then the Gaussian field due to sources in B is:

$$G(t) = \sum_{j=1}^{N_B} q_j e^{-|t-s_j|^2/\delta}. \tag{46}$$

Then take Hermite expansion about s_B:

$$G(t) = \sum_{\alpha \geq 0} A_\alpha h_\alpha \left(\frac{t - s_B}{\sqrt{\delta}} \right), \tag{47}$$

where the coefficients A_α is given by:

$$A_\alpha = \frac{1}{\alpha!} \sum_{j=1}^{N_B} q_j \left(\frac{s_j - s_B}{\sqrt{\delta}} \right)^\alpha \tag{48}$$

Furthermore, expand Hermite expansion with Taylor expansion at point t_C, I will get:

$$G(t) = \sum_{\beta \geq 0} B_\alpha \left(\frac{t - t_C}{\sqrt{\delta}} \right)^\beta, \tag{49}$$

where the coefficient B_β is given by:

$$B_\beta = \frac{(-1)^{|\beta|}}{\beta!} \sum_{\alpha \geq 0} A_\alpha h_{\alpha+\beta} \left(\frac{s_B - t_C}{\sqrt{\delta}} \right). \tag{50}$$

Here I use the Taylor series of h_α:

$$h_\alpha(t) = \sum_{\beta \geq 0} \frac{(t - t_0)^\beta}{\beta!} (-1)^{|\beta|} h_{\alpha+\beta}(t_0) \tag{51}$$

Combine the Taylor expansion of $G(t)$ at near point t_C:

$$G(t) = \sum_{\beta \geq 0} C_\beta \left(\frac{t - t_C}{\sqrt{\delta}} \right)^\beta \tag{52}$$

where

$$C_\beta = \frac{(-1)^{|\beta|}}{\beta!} \sum_{j=1}^{N_B} q_j h_\beta \left(\frac{s_j - t_C}{\sqrt{\delta}} \right) \tag{53}$$

I have three asymptotic expansion if I truncate each expansion series properly.

14.4 the image of Fast Gaussian Transform

Include direct evaluation of Gaussian, I have 4 methods. We will use scatter to explain how FGT work more clearly. We first will rethink about Gaussian function, it also means how source s_j pass impact to target t. In Figures, they can be shown as:

Figure 7: Direct Evaluation's Image
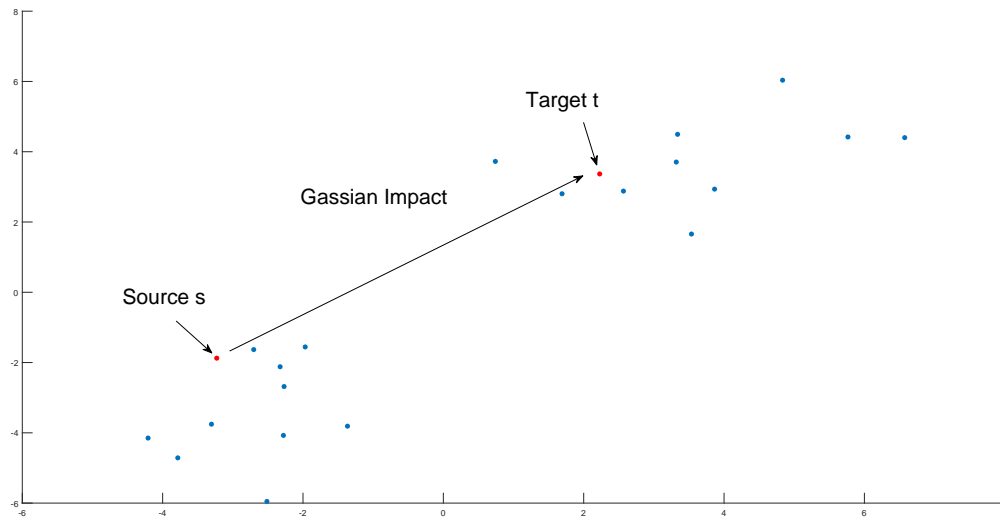
The Taylor expansion is also called as near-field expansion or Regular expansion. It means that instead of receive the impact direct from source, I can receive it through local box center.
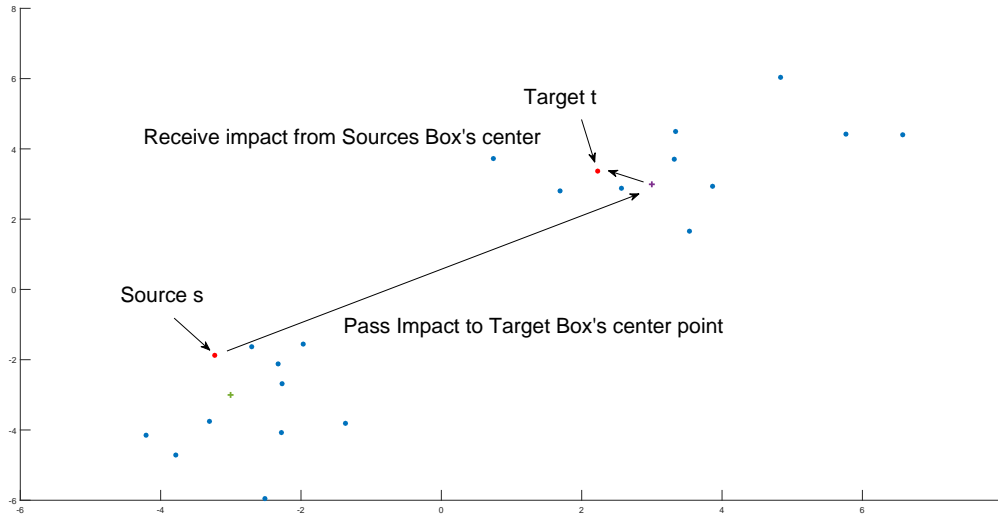
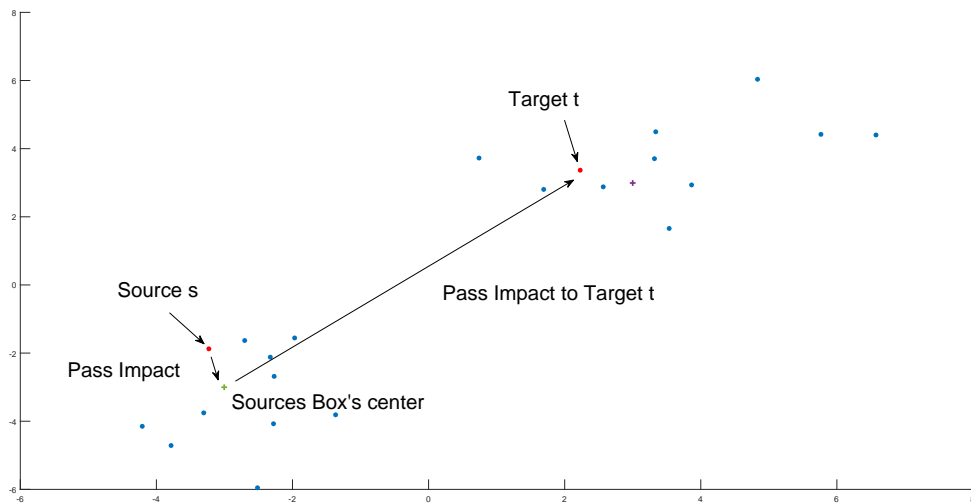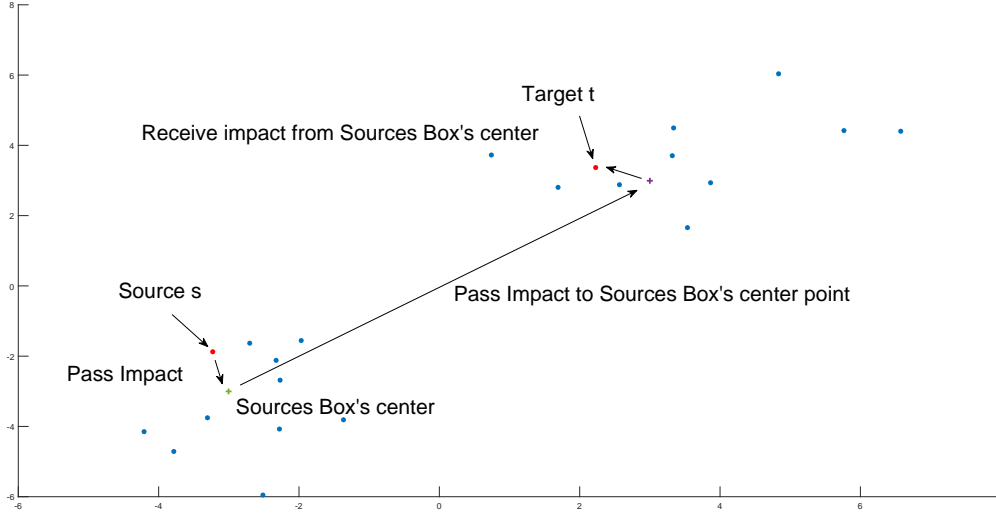Figure 8: Taylor expansion's Image

Figure 9: Hermite expansion's Image

The Hermite expansion is also called as far-field expansion or Singular expansion. It means that I gather source box's impact at source box's center, then pass it to target t.

Figure 10: Hermite expansion with Taylor expansion's Image

Hermite expansion with Taylor expansion means that with Hermite expansion, source box'c center does not pass impact to target directly, but pass to target box's center.

14.5 Numeric experiment

The model I use is stochastic volatility model as:

$$
\begin{aligned}
Y_t &= \exp(\frac{h_t}{2})\eta_t, \\
h_t &= \rho h_{t-1} + \epsilon_t.
\end{aligned}
$$

We use the particle filter to obtain the prediction density $f(x_{t+1}|y_t)$. Simulate state variable h_t as AR process with auto-correlation coefficient $\rho = 0.7$ and marginal distribution is standard normal distribution. Measurement error η_t is distributed as standard normal distribution. We compare methods between directly evaluation and FGT by calculating the prediction probability of this model. The result of different particle numbers, is shown in Table.(14.5).

15 Conclusion

By learning this Fast Gaussian transform, I know that the Gaussian convolution can be computed very fast. This is can be an important improvement of our computation of particle filter

Table 14: Elapsed time comparison between Direct evaluation and Fast Gaussion Transform

#Particles	Direct evaluation(second)	FGT(second)
100	0.4502	0.0946
200	1.7584	0.2292
300	3.9491	0.3979
400	6.9924	0.6087
500	10.8800	0.8303

and smoother.

Appendices

A   Likelihood Function

In order to express the transition equation (2) in matrix form, I express the log volatilities and disturbance terms used in (1) and (2) in vector form, as follows:

$$\mathbf{h}_1 = (h_{11}, \ldots, h_{1t})', \ \mathbf{h}_2 = (h_{21}, \ldots, h_{2t})', \tag{54}$$

$$\mathbf{u}_1 = (u_{11}, \ldots, u_{1t})', \ \mathbf{u}_2 = (u_{21}, \ldots, u_{2t})', \tag{55}$$

$$\mathbf{e}_1 = (e_{11}, \ldots, e_{1t})', \ \mathbf{e}_2 = (e_{21}, \ldots, e_{2t})'. \tag{56}$$

Then the transition equation (2) is expressed as:

$$\mathbf{h}_1 = \mathbf{V}_\rho^{1/2}(b_1\mathbf{u}_1), \ \mathbf{h}_2 = \mathbf{V}_\psi^{1/2}(b_2\mathbf{u}_1 + \sqrt{c}\mathbf{u}_2) = \mathbf{V}_\psi^{1/2}(\mathbf{V}_\rho^{-1/2}\mathbf{h}_1 b_2/b_1 + \sqrt{c}\mathbf{u}_2), \tag{57}$$

where $\mathbf{V}_\rho$ and $\mathbf{V}_\psi$ are the covariance matrices of the autoregressive processes of order one, $\mathbf{h}_1$ and $\mathbf{h}_2$, respectively, and $\mathbf{V}_\rho^{1/2}$ and $\mathbf{V}_\psi^{1/2}$ are defined by their Cholesky decomposition:

$$\mathbf{V}_\rho = (\mathbf{V}_\rho^{1/2})(\mathbf{V}_\rho^{1/2})', \ \mathbf{V}_\psi = (\mathbf{V}_\psi^{1/2})(\mathbf{V}_\psi^{1/2})',$$

where

$$\mathbf{V}_\psi^{1/2} = \begin{pmatrix} 1/\sqrt{1-\psi^2} & 0 & \ldots & 0 & 0 \\ \psi/\sqrt{1-\psi^2} & 1 & \ldots & 0 & 0 \\ \psi^2/\sqrt{1-\psi^2} & \psi & \ldots & 0 & 0 \\ \vdots & & & & \\ \psi^{n-1}/\sqrt{1-\psi^2} & \psi^{n-2} & \ldots & \psi & 1 \end{pmatrix}, \ \mathbf{V}_\rho^{1/2} = \begin{pmatrix} 1/\sqrt{1-\rho^2} & 0 & \ldots & 0 & 0 \\ \rho/\sqrt{1-\rho^2} & 1 & \ldots & 0 & 0 \\ \rho^2/\sqrt{1-\rho^2} & \rho & \ldots & 0 & 0 \\ \vdots & & & & \\ \rho^{n-1}/\sqrt{1-\rho^2} & \rho^{n-2} & \ldots & \rho & 1 \end{pmatrix}, \tag{58}$$

It is easy to see that their inverses are decomposed as $\mathbf{V}_\psi^{-1} = (\mathbf{V}_\psi^{-\frac{1}{2}})'\mathbf{V}_\psi^{-\frac{1}{2}}, \ \mathbf{V}_\rho^{-1} = (\mathbf{V}_\rho^{-\frac{1}{2}})'\mathbf{V}_\rho^{-\frac{1}{2}},$ where:

$$\mathbf{V}_\psi^{-\frac{1}{2}} = \begin{pmatrix} \sqrt{1-\psi^2} & 0 & \ldots & 0 & 0 \\ -\psi & 1 & \ldots & 0 & 0 \\ 0 & -\psi & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 0 & 0 \\ \vdots & & & & \\ 0 & 0 & \ldots & -\psi & 1 \end{pmatrix}, \ \mathbf{V}_\rho^{-\frac{1}{2}} = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & \ldots & 0 & 0 \\ -\rho & 1 & \ldots & 0 & 0 \\ 0 & -\rho & \ldots & 0 & 0 \\ 0 & 0 & \ldots & 0 & 0 \\ \vdots & & & & \\ 0 & 0 & \ldots & -\rho & 1 \end{pmatrix}. \tag{59}$$

Then the density functions of the transition and measurement equations of the model can be expressed as:

$$f(\mathbf{h}_1) = \frac{1}{(2\pi)^{\frac{t}{2}} \left| \mathbf{V}_\rho^{1/2} \right| b_1^t} \exp\left\{ -\frac{1}{2} b_1^{-2} \mathbf{h}_1' \mathbf{V}_\rho^{-1} \mathbf{h}_1 \right\}, \tag{60}$$

$$f(\mathbf{h}_2|\mathbf{h}_1) = \frac{1}{(2\pi)^{\frac{t}{2}} \left| \mathbf{V}_\psi^{1/2} \right| (\sqrt{c})^t} \exp\left\{ -\frac{1}{2} \mathbf{u}_2' \mathbf{u}_2 \right\}, \tag{61}$$

$$f(\mathbf{y}_1|\mathbf{h}_1) = \frac{1}{(2\pi)^{\frac{t}{2}}} \frac{1}{\left| \mathbf{H}_1^{1/2} \right|} \exp\left\{ -\frac{1}{2} \mathbf{y}_1' \mathbf{H}_1^{-1} \mathbf{y}_1 \right\}, \tag{62}$$

$$f(\mathbf{y}_2|\mathbf{h}_2) = \frac{1}{(2\pi)^{\frac{t}{2}}} \frac{1}{\left| \mathbf{H}_2^{1/2} \right|} \exp\left\{ -\frac{1}{2} \mathbf{y}_2' \mathbf{H}_2^{-1} \mathbf{y}_2 \right\}, \tag{63}$$

where

$$\mathbf{u}_2 = \left(\mathbf{V}_\psi^{-\frac{1}{2}} \mathbf{h}_2 - \mathbf{V}_\rho^{-\frac{1}{2}} \mathbf{h}_1 \frac{b_2}{b_1} \right) / \sqrt{c}, \tag{64}$$

$$\mathbf{H}_1 = diag(\exp(h_{11}), \ldots, \exp(h_{1t})), \ \mathbf{H}_2 = diag(\exp(h_{21}), \ldots, \exp(h_{2t})). \tag{65}$$

Then I can rewrite the likelihood function, given by:

$$f(\mathbf{y}_1, \mathbf{y}_2) = \int f(\mathbf{y}_2|\mathbf{h}_2) f(\mathbf{y}_1|\mathbf{h}_1) f(\mathbf{h}_2|\mathbf{h}_1) f(\mathbf{h}_1) d\mathbf{h}_2 d\mathbf{h}_1, \tag{66}$$

as

$$f(\mathbf{y}_1, \mathbf{y}_2) = \int f(\mathbf{y}_2|\mathbf{u}_2, \mathbf{h}_1) f(\mathbf{y}_1|\mathbf{h}_1) f(\mathbf{u}_2|\mathbf{h}_1) f(\mathbf{h}_1) d\mathbf{u}_2 d\mathbf{h}_1, \tag{67}$$

where

$$f(\mathbf{u}_2|\mathbf{h}_1) = \frac{1}{(2\pi)^{\frac{t}{2}}} \exp\left(-\frac{1}{2} \mathbf{u}_2' \mathbf{u}_2 \right), \tag{68}$$

in terms of $\mathbf{u}_2$, instead of $\mathbf{h}_2$, by the variable transformation given in (64).

B Score function with respect to c

We obtain the score function with respct to c from (67) as:

$$\frac{\partial f(\mathbf{y})}{\partial c} = \int \frac{\partial f(\mathbf{y}_2|\mathbf{u}_2, \mathbf{h}_1)}{\partial c} f(\mathbf{y}_1|\mathbf{h}_1) f(\mathbf{u}_2|\mathbf{h}_1) f(\mathbf{h}_1) d\mathbf{u}_2 d\mathbf{h}_1, \tag{69}$$

as the variance parameter c appears only in:

$$f(\mathbf{y}_2|\mathbf{h}_1, \mathbf{u}_2) = \frac{1}{(2\pi)^{\frac{t}{2}}} \frac{1}{\left| \mathbf{H}_2^{1/2} \right|} \exp\left\{ -\frac{1}{2} \mathbf{y}_2' \mathbf{H}_2^{-1} \mathbf{y}_2 \right\}$$

through $\mathbf{h}_2$ in $\mathbf{H}_2 = diag(\exp(\mathbf{h}_2))$ because, from (57), I have:

$$\mathbf{h}_2 = \mathbf{V}_\psi^{-\frac{1}{2}} \left(\sqrt{c} \mathbf{u}_2 + \mathbf{V}_\rho^{-\frac{1}{2}} \frac{b_2}{b_1} \mathbf{h}_1 \right). \tag{70}$$

Then I obtain the derivative of $f(\mathbf{y}_2|\mathbf{h}_1, \mathbf{u}_2,)$ with respect to c, as follows. First, noting (70), I define:

$$f(\mathbf{y}_2|\mathbf{h}_1, \mathbf{u}_2,) = E_t F_t, \tag{71}$$

where

$$\begin{aligned}
E_t &= \frac{1}{|\mathbf{H}_2|^{1/2}} = \exp\left(-\frac{1}{2}\mathbf{1}_{1\times t}\mathbf{h}_2\right) \\
&= \exp\left(-\frac{1}{2}\mathbf{1}_{1\times t}\mathbf{V}_\psi^{1/2}\left(\sqrt{c}\mathbf{u}_2 + \mathbf{V}_\psi^{-\frac{1}{2}}\frac{b_2}{b_1}\mathbf{h}_1\right)\right), \tag{72} \\
F_t &= \exp\left\{-\frac{1}{2}\mathbf{y}_2'\mathbf{H}_2^{-1}\mathbf{y}_2\right\} \\
&= \exp\left\{-\frac{1}{2}(\exp(-\mathbf{h}_2))'\mathbf{y}_2^2\right\}, \tag{73}
\end{aligned}$$

and, for notational convenience, I define:

$$\exp(-\mathbf{h}_2) = (\exp(-h_{21}),\ldots,\exp(-h_{2t}))',\ \mathbf{y}_2^2 = \left\{y_{21}^2, y_{22}^2, \ldots, y_{2t}^2\right\}',$$

and $\mathbf{h}_2$ denotes a function of $\mathbf{u}_2$ as the abbreviation of equation (70).

Then, from (69), I have:

$$\begin{aligned}
B_t &= \lim_{c\to 0}\frac{\partial f(\mathbf{y})}{\partial c} \tag{74} \\
&= \lim_{c\to 0}\int (other\ terms)(F_t\frac{\partial E_t}{\partial c} + E_t\frac{\partial F_t}{\partial c})d\mathbf{u}_2 d\mathbf{h}_1 \tag{75} \\
&= \lim_{c\to 0}\frac{\sqrt{c}\int (other\ terms)(F_t\ M_{1t} + E_t\ M_{2t})d\mathbf{u}_2 d\mathbf{h}_1}{c}, \tag{76}
\end{aligned}$$

where I define:

$$M_{1t} = \frac{\partial E_t}{\partial c}\frac{1}{\sqrt{c}},\ M_{2t} = \frac{\partial F_t}{\partial c}\frac{1}{\sqrt{c}}. \tag{77}$$

We need $\sqrt{c}$ in the denominator of (77) as:

$$\begin{aligned}
\frac{\partial E_t}{\partial c} &= -\frac{1}{2}E_t\ \frac{1}{2\sqrt{c}}\mathbf{1}_{1\times t}\mathbf{V}_\psi^{1/2}\mathbf{u}_2 \tag{78} \\
&= -\frac{1}{4\sqrt{c}}E_t\ \mathbf{1}_{1\times t}\mathbf{V}_\psi^{1/2}\mathbf{u}_2, \tag{79} \\
\frac{\partial F_t}{\partial c} &= -\frac{1}{2}F_t\ \left(\frac{\partial}{\partial c}\exp(-\mathbf{h}_2')\mathbf{y}_2^2\right) \tag{80} \\
&= -\frac{1}{2}F_t\ \left(-\frac{1}{2\sqrt{c}}\mathbf{u}_2'\mathbf{V}_\psi^{1/2'}\mathbf{H}_2^{-1}\mathbf{y}_2^2\right) \tag{81} \\
&= \frac{1}{4\sqrt{c}}F_t\ G_t, \tag{82} \\
G_t &= \mathbf{u}_2'\mathbf{V}_\psi^{1/2'}\mathbf{H}_2^{-1}\mathbf{y}_2^2, \tag{83}
\end{aligned}$$

as

$$\frac{\partial \mathbf{h}_2}{\partial c} = \frac{1}{2\sqrt{c}} \mathbf{V}_\psi^{1/2} \mathbf{u}_2, \quad \frac{\partial \exp(-\mathbf{h}_2)}{\partial c} = -\frac{1}{2\sqrt{c}} \mathbf{H}_2^{-1} \mathbf{V}_\psi^{1/2} \mathbf{u}_2.$$

Note that the denominators of the derivatives (79) and (82) contain c, which converges to zero, and hence is intractable by conventional methods. It is convenient to use the method proposed by Chesher [1984]. First, applying L'Hopital's rule to (76) with respect to c , I obtain:

$$B_t = \frac{1}{2} B_t + \lim_{c \to 0} \sqrt{c} \frac{\partial}{\partial c} \int (other\ terms)(F_t\ M_{1t} + E_t\ M_{2t}) d\mathbf{u}_2 d\mathbf{h}_1. \tag{84}$$

Comparing both sides of equation (84), I have:

$$B_t = 2 \lim_{c \to 0} \sqrt{c} \int (other\ terms)(\frac{\partial F_t}{\partial c}\ M_{1t} + \frac{\partial E_t}{\partial c}\ M_{2t} + F_t \frac{\partial M_{1t}}{\partial c} + E_t \frac{\partial M_{2t}}{\partial c}) d\mathbf{u}_2 d\mathbf{h}_1 \tag{85}$$

$$= 2 \lim_{c \to 0} \sqrt{c} \int (other\ terms)(2M_{1t}\ M_{2t} + F_t \frac{\partial M_{1t}}{\partial c} + E_t \frac{\partial M_{2t}}{\partial c}) d\mathbf{u}_2 d\mathbf{h}_1. \tag{86}$$

Defining $\mathbf{Y}_2 = diag(\mathbf{y}_2^2)$, the terms in the integrand of C_t can be expressed as follows:

$$M_{1t}\ M_{2t} = (\sqrt{c} \frac{\partial E_t}{\partial c})(\sqrt{c} \frac{\partial F_t}{\partial c})$$

$$= \left(-\frac{1}{4} E_t\ \mathbf{1}_{1 \times t} \mathbf{V}_\psi^{1/2} \mathbf{u}_2 \right) \left(\frac{1}{4} F_t\ \mathbf{u}_2' \mathbf{V}_\psi^{1/2'} \mathbf{Y}_2 \exp(-\mathbf{h}_2) \right)$$

$$= -\frac{1}{16} E_t F_t \left(\mathbf{1}_{1 \times t} \mathbf{V}_\psi^{1/2} \mathbf{u}_2 \mathbf{u}_2' \mathbf{V}_\psi^{1/2'} \mathbf{Y}_2 \exp(-\mathbf{h}_2) \right),$$

$$\frac{\partial M_{1t}}{\partial c} = -\frac{1}{4} \frac{\partial E_t}{\partial c} \mathbf{1}_{1 \times t} \mathbf{V}_\psi^{1/2} \mathbf{u}_2$$

$$= -\frac{1}{4} \left(-\frac{\sqrt{c}}{4c} \right) E_t\ \mathbf{1}_{1 \times t} \mathbf{V}_\psi^{1/2} \mathbf{u}_2 \mathbf{u}_2' \mathbf{V}_\psi^{1/2'} \mathbf{1}_{t \times 1}$$

$$= \frac{1}{16\sqrt{c}} E_t\ tr \left(\mathbf{1}_{1 \times t} \mathbf{V}_\psi^{1/2} \mathbf{u}_2 \mathbf{u}_2' \mathbf{V}_\psi^{1/2'} \mathbf{1}_{t \times 1} \right)$$

$$= \frac{1}{16\sqrt{c}} E_t\ tr(\mathbf{1}_{t \times t} \mathbf{V}_\psi^{1/2} \mathbf{u}_2 \mathbf{u}_2' \mathbf{V}_\psi^{1/2'}),$$

$$\frac{\partial M_{2t}}{\partial c} = \frac{1}{4} \frac{\partial F_t}{\partial c} G_t + \frac{1}{4} F_t \frac{\partial G_t}{\partial c}$$

$$= \frac{1}{16\sqrt{c}} F_t G_t^2 + \frac{1}{4} F_t \frac{\partial G_t}{\partial c},$$

$$\frac{\partial G_t}{\partial c} = \mathbf{u}_2' \mathbf{V}_\psi^{1/2'} \mathbf{Y}_2 \frac{\partial \exp(-\mathbf{h}_2)}{\partial c}$$

$$= -\frac{1}{2\sqrt{c}} \mathbf{u}_2' \mathbf{V}_\psi^{1/2'} \mathbf{Y}_2 \mathbf{H}_2^{-1} \mathbf{V}_\psi^{1/2} \mathbf{u}_2$$

$$= -\frac{1}{2\sqrt{c}} tr \left(\mathbf{u}_2' \mathbf{V}_\psi^{1/2'} \mathbf{Y}_2 \mathbf{H}_2^{-1} \mathbf{V}_\psi^{1/2} \mathbf{u}_2 \right)$$

$$= -\frac{1}{2\sqrt{c}} tr \left(\mathbf{V}_\psi^{1/2'} \mathbf{Y}_2 \mathbf{H}_2^{-1} \mathbf{V}_\psi^{1/2} \mathbf{u}_2 \mathbf{u}_2' \right),$$

$$
\begin{aligned}
G_t^2 \quad &= \mathbf{u}_2' \mathbf{V}_\psi^{1/2'} \mathbf{Y}_2 \exp(-\mathbf{h}_2) \exp(-\mathbf{h}_2') \mathbf{Y}_2 \mathbf{V}_\psi^{1/2} \mathbf{u}_2 \\
&= tr\left(\mathbf{u}_2' \mathbf{V}_\psi^{1/2'} \mathbf{Y}_2 \exp(-\mathbf{h}_2) \exp(-\mathbf{h}_2') \mathbf{Y}_2 \mathbf{V}_\psi^{1/2} \mathbf{u}_2 \right) \\
&= tr\left(\mathbf{V}_\psi^{1/2'} \mathbf{Y}_2 \exp(-\mathbf{h}_2) \exp(-\mathbf{h}_2') \mathbf{Y}_2 \mathbf{V}_\psi^{1/2} \mathbf{u}_2 \mathbf{u}_2' \right).
\end{aligned}
$$

Then I have:

$$
\begin{aligned}
B_t \;=\; &\frac{1}{8} \lim_{c \to 0} \int f(\mathbf{y}_1|\mathbf{h}_1) \frac{1}{(2\pi)^{\frac{t}{2}}} E_t F_t \\
&\left\{ -2tr\left(\mathbf{1}_{1 \times t} \mathbf{V}_\psi^{1/2} \mathbf{u}_2 \mathbf{u}_2' \mathbf{V}_\psi^{1/2'} \mathbf{Y}_2 \exp(-\mathbf{h}_2) \right) \right. \\
&+\; tr\left(\mathbf{1}_{t \times t} \mathbf{V}_\psi^{1/2} \mathbf{u}_2 \mathbf{u}_2' \mathbf{V}_\psi^{1/2'} \right) \\
&+\; tr\left(\mathbf{V}_\psi^{1/2'} \mathbf{Y}_2 \exp(-\mathbf{h}_2) \exp(-\mathbf{h}_2') \mathbf{Y}_2 \mathbf{V}_\psi^{1/2} \mathbf{u}_2 \mathbf{u}_2' \right) \\
&-\; \left. 2tr\left(\mathbf{V}_\psi^{1/2'} \mathbf{Y}_2 \mathbf{H}_2^{-1} \mathbf{V}_\psi^{1/2} \mathbf{u}_2 \mathbf{u}_2' \right) \right\} f(\mathbf{u}_2|\mathbf{h}_1) f(\mathbf{h}_1) d\mathbf{u}_2 d\mathbf{h}_1. \quad (87)
\end{aligned}
$$

We can perform the integration with respect to $\mathbf{u}_2$ in (87) analytically. As $\mathbf{u}_2|\mathbf{h}_1$ follows the t-dimensitonal standard normal distribution:

$$
\int \mathbf{u}_2 \mathbf{u}_2' f(\mathbf{u}_2|\mathbf{h}_1) d\mathbf{u}_2 = \mathbf{I}_t, \quad (88)
$$

under the null hypothesis $\mathbf{h}_2 = \mathbf{h}_1$ and $\psi = \rho$, equation (87) is expressed as:

$$
\begin{aligned}
B_t \;=\; &\frac{1}{8} \int f(\mathbf{y}_1|\mathbf{h}_1) \frac{1}{(2\pi)^{\frac{t}{2}}} E_t F_t \\
&\left\{ -2tr\left(\mathbf{1}_{1 \times t} \mathbf{V}_\rho^{1/2} \mathbf{V}_\rho^{1/2'} \mathbf{Y}_2 \exp(-\mathbf{h}_1) \right) \right. \\
&+\; tr\left(\mathbf{1}_{t \times t} \mathbf{V}_\rho^{1/2} \mathbf{V}_\rho^{1/2'} \right) \\
&+\; tr\left(\mathbf{V}_\rho^{1/2'} \mathbf{Y}_2 \exp(-\mathbf{h}_1) \exp(-\mathbf{h}_1') \mathbf{Y}_2 \mathbf{V}_\rho^{1/2} \right) \\
&-\; \left. 2tr\left(\mathbf{V}_\rho^{1/2'} \mathbf{Y}_2 \mathbf{H}_1^{-1} \mathbf{V}_\rho^{1/2} \right) \right\} f(\mathbf{h}_1) d\mathbf{h}_1. \quad (89)
\end{aligned}
$$

Noting that $\mathbf{V}_\rho = \mathbf{V}_\rho^{1/2} \mathbf{V}_\rho^{1/2'}$, and applying the cyclic property of the trace operator to simplify equation (89), I have:

$$
B_t = \left. \frac{\partial f(\mathbf{y})}{\partial c} \right|_{\mathbf{H}_0} \;=\; \int tr \mathbf{J}_t f(\mathbf{y}, \mathbf{h}_1) d\mathbf{h}_1, \quad (90)
$$

where

$$
\begin{aligned}
\mathbf{J}_t \;=\; &\frac{1}{8} \left\{ -2\left(\mathbf{1}_{1 \times t} \mathbf{V}_\rho \mathbf{Y}_2 \exp(-\mathbf{h}_1) \right) + \left(\mathbf{1}_{t \times t} \mathbf{V}_\rho \right) \right. \quad (91) \\
&+\; \left. \left(\mathbf{V}_\rho \mathbf{Y}_2 \exp(-\mathbf{h}_1) \exp(-\mathbf{h}_1') \mathbf{Y}_2 \right) - 2\left(\mathbf{V}_\rho \mathbf{Y}_2 \mathbf{H}_1^{-1} \right) \right\}. \quad (92)
\end{aligned}
$$

It follows that:

$$\left.\frac{\partial \log f(\mathbf{y})}{\partial c}\right|_{\mathbf{H}_0} = \lim_{c \to 0} \frac{1}{f(\mathbf{y})} \frac{\partial f(\mathbf{y})}{\partial c} \tag{93}$$

$$= \int tr\mathbf{J}_t \frac{1}{f(\mathbf{y})} f(\mathbf{h}_1, \mathbf{y}) d\mathbf{h}_1 \tag{94}$$

$$= tr \int \mathbf{J}_t f(\mathbf{h}_1|\mathbf{y}) d\mathbf{h}_1 \tag{95}$$

$$= tr E_{\mathbf{h}_1|\mathbf{y}}(\mathbf{J}_t), \tag{96}$$

as $f(\mathbf{h}_1|\mathbf{y}) = f(\mathbf{h}_1, \mathbf{y})/f(\mathbf{y})$. From (91), I have only to evaluate $E_{\mathbf{h}_1|\mathbf{y}} [\exp(-\mathbf{h}_1)]$ and

$$E_{\mathbf{h}_1|\mathbf{y}} \left[\exp(-\mathbf{h}_1) \exp(-\mathbf{h}_1)' \right]$$

to obtain the score function with respect to ψ. These expected values have no analytic expressions, so that they will need to be evaluated numerically.

C   Score function with respect to ψ

In the log-likelihood function, ψ appears only in $f(\mathbf{y}_1|\mathbf{h}_1, \mathbf{u}_2) = E_t F_t$, as shown in (67) and (71). The partial derivative of the likelihood with respect to ψ can be expressed as :

$$\frac{\partial f(\mathbf{y})}{\partial \psi} = \int \left(\frac{\partial E_t}{\partial \psi} E_t^{-1} + \frac{\partial F_t}{\partial \psi} F_t^{-1} \right) f(\mathbf{y}, \mathbf{u}_2, \mathbf{h}_1) d\mathbf{u}_2 d\mathbf{h}_1 \tag{97}$$

$$= \int \left(\frac{\partial E_t}{\partial \psi} E_t^{-1} + \frac{\partial F_t}{\partial \psi} F_t^{-1} \right) f(\mathbf{y}, \mathbf{h}_1) d\mathbf{h}_1, \tag{98}$$

since, as will be seen later, $\mathbf{u}_2$ can be integrated out in $\left(\frac{\partial E_t}{\partial \psi} E_t^{-1} + \frac{\partial F_t}{\partial \psi} F_t^{-1} \right)$. Then I have:

$$\left.\frac{\partial \log f(\mathbf{y})}{\partial \psi}\right|_{\mathbf{H}_0} = \int \left(\frac{\partial E_t}{\partial \psi} E_t^{-1} + \frac{\partial F_t}{\partial \psi} F_t^{-1} \right) f(\mathbf{h}_1|\mathbf{y}) d\mathbf{h}_1 = E_{\mathbf{h}_1|\mathbf{y}} \left(\frac{\partial E_t}{\partial \psi} E_t^{-1} + \frac{\partial F_t}{\partial \psi} F_t^{-1} \right), \tag{99}$$

as

$$f(\mathbf{h}_1|\mathbf{y}) = f(\mathbf{h}_1, \mathbf{y})/f(\mathbf{y}).$$

First, using the formula:

$$\frac{\partial \mathbf{V}_\psi^{1/2}}{\partial \psi} = -\mathbf{V}_\psi^{1/2} \mathbf{Z}_\psi \mathbf{V}_\psi^{1/2}, \tag{100}$$

where

$$\mathbf{Z}_\psi = \frac{\partial \mathbf{V}_\psi^{-1/2}}{\partial \psi},$$

note that

$$\frac{\partial E_t}{\partial \psi} = -\frac{1}{2} E_t \left[\mathbf{1}_{1 \times t} \frac{\partial \mathbf{V}_\psi^{1/2}}{\partial \psi} \left(\sqrt{c} \mathbf{u}_2 + \mathbf{V}_\rho^{-\frac{1}{2}} \frac{b_2}{b_1} \mathbf{h}_1 \right) \right] \tag{101}$$

$$= \frac{1}{2} E_t \left[\mathbf{1}_{1 \times t} \mathbf{V}_\psi^{1/2} \mathbf{Z}_\psi \mathbf{V}_\psi^{1/2} \left(\sqrt{c} \mathbf{u}_2 + \mathbf{V}_\rho^{-\frac{1}{2}} \frac{b_2}{b_1} \mathbf{h}_1 \right) \right], \tag{102}$$

$$\frac{\partial F_t}{\partial \psi} = -\frac{1}{2} F_t \frac{\partial}{\partial \psi} \left[\mathbf{y}_2^{2'} \exp(-\mathbf{h}_2) \right] \tag{103}$$

$$= \frac{1}{2} F_t \left[\mathbf{y}_2^{2'} \mathbf{H}_2^{-1} \frac{\partial \mathbf{V}_\psi^{1/2}}{\partial \psi} \left(\sqrt{c} \mathbf{u}_2 + \mathbf{V}_\rho^{-\frac{1}{2}} \frac{b_2}{b_1} \mathbf{h}_1 \right) \right] \tag{104}$$

$$= -\frac{1}{2} F_t \left[\mathbf{y}_2^{2'} \mathbf{H}_2^{-1} \mathbf{V}_\psi^{1/2} \mathbf{Z}_\psi \mathbf{V}_\psi^{1/2} \left(\sqrt{c} \mathbf{u}_2 + \mathbf{V}_\rho^{-\frac{1}{2}} \frac{b_2}{b_1} \mathbf{h}_1 \right) \right], \tag{105}$$

as I have:

$$\mathbf{h}_2 = \mathbf{V}_\psi^{1/2} \left(\sqrt{c} \mathbf{u}_2 + \mathbf{V}_\rho^{-\frac{1}{2}} \frac{b_2}{b_1} \mathbf{h}_1 \right), \tag{106}$$

so that:

$$\frac{\partial}{\partial \psi} \mathbf{h}_2 = \frac{\partial \mathbf{V}_\psi^{1/2}}{\partial \psi} \left(\sqrt{c} \mathbf{u}_2 + \mathbf{V}_\rho^{-\frac{1}{2}} \frac{b_2}{b_1} \mathbf{h}_1 \right), \tag{107}$$

$$E_t = \exp \left(-\frac{1}{2} \mathbf{1}_{1 \times t} \mathbf{V}_\psi^{1/2} \left(\sqrt{c} \mathbf{u}_2 + \mathbf{V}_\rho^{-\frac{1}{2}} \frac{b_2}{b_1} \mathbf{h}_1 \right) \right),$$

$$F_t = \exp \left\{ -\frac{1}{2} \left(\exp(-\mathbf{h}_2) \right)' \mathbf{y}_2^2 \right\}.$$

We have used $-\mathbf{V}_\psi^{1/2} \mathbf{Z}_\psi \mathbf{V}_\psi^{1/2}$ rather than $\frac{\partial \mathbf{V}_\psi^{1/2}}{\partial \psi}$ because it is easy to generate computationally. Evaluating each term under the null hypothesis $c = 0$ and $b_1 = b_2$, I have:

$$\frac{\partial E_t}{\partial \psi} |_{\mathbf{H}_0} = \frac{1}{2} E_t \left[\mathbf{1}_{1 \times t} \mathbf{V}_\psi^{1/2} \mathbf{Z}_\psi \mathbf{h}_1 \right], \tag{108}$$

$$\frac{\partial F_t}{\partial \psi} |_{\mathbf{H}_0} = -\frac{1}{2} F_t \left[\mathbf{y}_2^{2'} \mathbf{H}_1^{-1} \mathbf{V}_\psi^{1/2} \mathbf{Z}_\psi \mathbf{h}_1 \right] \tag{109}$$

$$= -\frac{1}{2} F_t \, tr \left[\exp(-\mathbf{h}_1') \mathbf{Y}_2 \mathbf{V}_\psi^{1/2} \mathbf{Z}_\psi \mathbf{h}_1 \right] \tag{110}$$

$$= -\frac{1}{2} F_t \, tr \left[\mathbf{Y}_2 \mathbf{V}_\psi^{1/2} \mathbf{Z}_\psi \mathbf{h}_1 \exp(-\mathbf{h}_1') \right], \tag{}$$

using the identity:

$$\mathbf{y}_2^{2'} \mathbf{H}_1^{-1} = \exp(-\mathbf{h}_1') \mathbf{Y}_2.$$

From (99), I have:

$$\frac{\partial \log f(\mathbf{y})}{\partial \psi} \bigg|_{\mathbf{H}_0} = \frac{1}{2} \left[\mathbf{1}_{1 \times t} \mathbf{V}_\rho^{1/2} \mathbf{Z}_\rho E_{\mathbf{h}_1|\mathbf{y}}[\mathbf{h}_1] \right] - \frac{1}{2} tr \left[\mathbf{Y}_2 \mathbf{V}_\rho^{1/2} \mathbf{Z}_\rho E_{\mathbf{h}_1|\mathbf{y}}[\mathbf{h}_1 \exp(-\mathbf{h}_1')] \right]. \tag{111}$$

Note that the matrix $\mathbf{Y}_2 \mathbf{V}_\rho^{1/2} \mathbf{Z}_\rho$ is lower triangular, and I have only to calculate the upper triangular part of the matrix $E_{\mathbf{h}_1|\mathbf{y}}[\mathbf{h}_1 \exp(-\mathbf{h}_1')]$ in evaluating the score function (111).

41

D   Score function with respect to b_2

First, note that, in the log-likelihood function, b_2 appears only in $f(\mathbf{y}_2|\mathbf{h}_1, \mathbf{u}_2) = E_t F_t$, through:

$$\mathbf{h}_2 = \mathbf{V}_\psi^{1/2} \left(\sqrt{c}\mathbf{u}_2 + \mathbf{V}_\rho^{-\frac{1}{2}} \frac{b_2}{b_1} \mathbf{h}_1 \right), \tag{112}$$

as shown in (67) and (71). Then I have the formula:

$$\frac{\partial \log f(\mathbf{y})}{\partial b_2} = E_{\mathbf{u}_2, \mathbf{h}_1|\mathbf{y}} \left(\frac{\partial E_t}{\partial b_2} E_t^{-1} + \frac{\partial F_t}{\partial b_2} F_t^{-1} \right), \tag{113}$$

using:

$$\frac{\partial f(\mathbf{y})}{\partial b_2} = \int (\mathbf{y}_1|\mathbf{h}_1) \frac{\partial f(\mathbf{y}_2|\mathbf{h}_1, \mathbf{u}_2)}{\partial b_2} f(\mathbf{h}_1) f(\mathbf{u}_2|\mathbf{h}_1) d\mathbf{u}_2 \mathbf{h}_1 = \int \left(\frac{\partial E_t}{\partial b_2} E_t^{-1} + \frac{\partial F_t}{\partial b_2} F_t^{-1} \right) f(\mathbf{y}, \mathbf{h}_1) d\mathbf{h}_1,$$

as I have:

$$\frac{\partial f(\mathbf{y}_2|\mathbf{h}_1, \mathbf{u}_2)}{\partial b_2} = \frac{\partial E_t}{\partial b_2} F_t + E_t \frac{\partial F_t}{\partial b_2} = \left(\frac{\partial E_t}{\partial b_2} E_t^{-1} + \frac{\partial F_t}{\partial b_2} F_t^{-1} \right) f(\mathbf{y}_2|\mathbf{h}_1, \mathbf{u}_2). \tag{114}$$

The partial derivatives of E_t and F_t are:

$$\frac{\partial E_t}{\partial b_2} = -\frac{1}{2} E_t \left[\mathbf{1}_{1 \times t} \mathbf{V}_\psi^{1/2} \left(\mathbf{V}_\rho^{-\frac{1}{2}} \frac{1}{b_1} \mathbf{h}_1 \right) \right], \tag{115}$$

$$\frac{\partial F_t}{\partial b_2} = -\frac{1}{2} F_t \frac{\partial}{\partial b_2} \left[\mathbf{y}_2^{2'} \exp(-\mathbf{h}_2) \right] \tag{116}$$

$$= \frac{1}{2} F_t \, tr \left[\mathbf{y}_2^{2'} \mathbf{H}_2 \mathbf{V}_\psi^{1/2} \left(\mathbf{V}_\rho^{-\frac{1}{2}} \frac{1}{b_1} \mathbf{h}_1 \right) \right]. \tag{117}$$

Note that, under the null hypothesis, $\rho = \psi, \mathbf{h}_1 = \mathbf{h}_2$, and $\mathbf{V}_\rho = \mathbf{V}_\psi$, so that I have:

$$\left. \frac{\partial E_t}{\partial b_2} \right|_{\mathbf{H}_0} = -\frac{1}{2} E_t \left[\mathbf{1}_{1 \times t} \frac{1}{b_1} \mathbf{h}_1 \right], \tag{118}$$

$$\left. \frac{\partial F_t}{\partial b_2} \right|_{\mathbf{H}_0} = -\frac{1}{2} F_t \frac{\partial}{\partial b_2} \left[\mathbf{y}_2^{2'} \exp(-\mathbf{h}_2) \right] \Big|_{\mathbf{H}_0}, \tag{119}$$

$$= \frac{1}{2} F_t \, tr \left[\mathbf{y}_2^{2'} \mathbf{H}_2 \frac{1}{b_1} \mathbf{h}_1 \right] \Big|_{\mathbf{H}_0}, \tag{120}$$

as

$$\frac{\partial}{\partial b_2} \exp(-\mathbf{h}_2) = -\mathbf{H}_2^{-1} b_2/b_1.$$

Then I have:

$$\left. \frac{\partial \log f(\mathbf{y})}{\partial b_2} \right|_{\mathbf{H}_0} = -\frac{1}{2b_1} \mathbf{1}_{1 \times t} E_{\mathbf{h}_1|\mathbf{y}} [\mathbf{h}_1] + \frac{1}{2b_1} tr \left[\mathbf{y}_2^{2'} E_{\mathbf{h}_1|\mathbf{y}} [\exp(-\mathbf{h}_1) \circ \mathbf{h}_1] \right], \tag{121}$$

where $\circ$ denotes the Hadamard (or element-by-element) product.

E   Score function with respect to ρ

In the log-likelihood function, ρ appears only in $f(y_1|\mathbf{h}_1, \mathbf{u}_2) = E_t F_t$ and $f(\mathbf{h}_1)$, as shown in (60) and (71). Then I have the derivative using the formula:

$$\frac{\partial \log f(\mathbf{y})}{\partial \rho}\Big|_{\mathbf{H}_0} = E_{\mathbf{h}_1|\mathbf{y}} \left(\frac{\partial E_t}{\partial \rho} E_t^{-1} + \frac{\partial F_t}{\partial \rho} F_t^{-1} + \frac{\partial f(\mathbf{h}_1)}{\partial \rho} f(\mathbf{h}_1)^{-1} \right), \tag{122}$$

analogously to that of (99). As:

$$E_t = \exp \left(-\frac{1}{2} \mathbf{1}_{1\times t} \mathbf{V}_\psi^{1/2} \left(\sqrt{c}\mathbf{u}_2 + \mathbf{V}_\rho^{-\frac{1}{2}} \frac{b_2}{b_1} \mathbf{h}_1 \right) \right), \tag{123}$$

$$F_t = \exp \left\{ -\frac{1}{2} \exp(-\mathbf{h}_2') \mathbf{y}_2^2 \right\}, \tag{124}$$

$$f(\mathbf{h}_1) = \frac{1}{(2\pi)^{\frac{t}{2}} \left| \mathbf{V}_\rho^{1/2} \right| b_1^t} \exp \left\{ -\frac{1}{2} b_1^{-2} \mathbf{h}_1' \mathbf{V}_\rho^{-1} \mathbf{h}_1 \right\}, \tag{125}$$

defining

$$\mathbf{Z}_\rho = \frac{\partial \mathbf{V}_\rho^{-1/2}}{\partial \rho}, \tag{126}$$

their derivatives are expressed as:

$$\frac{\partial E_t}{\partial \rho} = -\frac{1}{2} E_t \left[\mathbf{1}_{1\times t} \mathbf{V}_\psi^{1/2} \mathbf{Z}_\rho \frac{b_2}{b_1} \mathbf{h}_1 \right], \tag{127}$$

$$\frac{\partial F_t}{\partial \rho} = \frac{1}{2} F_t \left[\mathbf{y}_2^{2'} \mathbf{H}_1^{-1} \mathbf{V}_\psi^{1/2} \mathbf{Z}_\rho \frac{b_2}{b_1} \mathbf{h}_1 \right], \tag{128}$$

$$\frac{\partial f(\mathbf{h}_1)}{\partial \rho} = f(\mathbf{h}_1) \left[-\frac{\rho}{1-\rho^2} - \frac{1}{2} b_1^{-2} tr \left(\mathbf{h}_1' \frac{\partial \mathbf{V}_\rho^{-1}}{\partial \rho} \mathbf{h}_1 \right) \right]$$

$$= f(\mathbf{h}_1) \left[-\frac{\rho}{1-\rho^2} - \frac{1}{2} b_1^{-2} tr \left(\frac{\partial \mathbf{V}_\rho^{-1}}{\partial \rho} \mathbf{h}_1 \mathbf{h}_1' \right) \right]. \tag{129}$$

We have used $(\partial/\partial\rho) \left| \mathbf{V}_\rho^{1/2} \right| = 1/\sqrt{1-\rho^2}$ in deriving the first term of equation (129).

Noting that $\exp(-\mathbf{h}_1')\mathbf{Y}_2 = (\mathbf{y}_2^2)'\mathbf{H}_1^{-1}$ under the null hypothesis, the above can be expressed as:

$$\frac{\partial E_t}{\partial \rho}\Big|_{\mathbf{H}_0} = -\frac{1}{2} E_t \left[\mathbf{1}_{1\times t} \mathbf{V}_\rho^{1/2} \mathbf{Z}_\rho \frac{b_2}{b_1} \mathbf{h}_1 \right] \tag{130}$$

$$= -\frac{\partial E_t}{\partial \psi}\Big|_{\mathbf{H}_0}, \tag{131}$$

$$\frac{\partial F_t}{\partial \rho}\Big|_{\mathbf{H}_0} = -\frac{1}{2} F_t \; tr \left[\mathbf{Y}_2 \mathbf{V}_\rho^{1/2} \mathbf{Z}_\rho \frac{b_2}{b_1} \mathbf{h}_1 \exp(-\mathbf{h}_1') \right] \tag{132}$$

$$= -\frac{\partial F_t}{\partial \psi}\Big|_{\mathbf{H}_0}. \tag{133}$$

The score function with respect to ρ can be expressed as:

$$\frac{\partial \log f(\mathbf{y})}{\partial \rho}|_{\mathbf{H}_0} = -\frac{\partial \log f(\mathbf{y})}{\partial \psi}|_{\mathbf{H}_0} + \left[-\frac{\rho}{1-\rho^2} - \frac{1}{2}b_1^{-2}tr\left(\frac{\partial \mathbf{V}_\rho^{-1}}{\partial \rho} E_{\mathbf{h}_1|\mathbf{y}} \left(\mathbf{h}_1 \mathbf{h}_1' \right) \right) \right]. \qquad (134)$$

F Score function with respect to standard deviation b_1

In the likelihood function, b_1 appears only in:

$$E_t = \exp\left(-\frac{1}{2}\mathbf{1}_{1\times t}\mathbf{V}_\psi^{1/2} \left(\sqrt{c}\mathbf{u}_2 + \mathbf{V}_\rho^{-\frac{1}{2}}\frac{b_2}{b_1}\mathbf{h}_1 \right) \right), \qquad (135)$$

$$F_t = \exp\left(-\frac{1}{2}\exp(-\mathbf{h}_2')\mathbf{y}_2^2 \right), \qquad (136)$$

$$f(\mathbf{h}_1) = \frac{1}{(2\pi)^{\frac{t}{2}} \left| \mathbf{V}_\rho^{1/2} \right| b_1^t} \exp\left(-\frac{1}{2}b_1^{-2}\mathbf{h}_1'\mathbf{V}_\rho^{-1}\mathbf{h}_1 \right). \qquad (137)$$

Then I can derive the score function with respect to b_1 using the formula analogous to that of ρ given in (122), with ρ replaced by b_1. We can easily show from (72) and (73) that, under the null hypothesis $b_2 = b_1$, the derivatives of E_t and F_t with respect to b_1 are equal to the negative of the derivatives with respect to b_2 , namely:

$$\frac{\partial E_t}{\partial b_1}|_{\mathbf{H}_0} = -\frac{\partial E_t}{\partial b_2}|_{\mathbf{H}_0}, \qquad (138)$$

$$\frac{\partial F_t}{\partial b_1}|_{\mathbf{H}_0} = -\frac{\partial F_t}{\partial b_2}|_{\mathbf{H}_0}, \qquad (139)$$

so that no additional calculations are necessary. From (60), the derivative of $f(\mathbf{h}_1)$ can be expressed as:

$$\frac{\partial f(\mathbf{h}_1)}{\partial b_1} = f(\mathbf{h}_1)\left[-\frac{t}{b_1} + \frac{1}{b_1^3}tr\left(\mathbf{h}_1'\mathbf{V}_\rho^{-1}\mathbf{h}_1 \right) \right] \qquad (140)$$

$$= f(\mathbf{h}_1)\left[-\frac{t}{b_1} + \frac{1}{b_1^3}tr\left(\mathbf{V}_\rho^{-1}\mathbf{h}_1\mathbf{h}_1' \right) \right]. \qquad (141)$$

Using the formula:

$$\frac{\partial \log f(\mathbf{y})}{\partial b_1}\bigg|_{\mathbf{H}_0} = E_{\mathbf{h}_1|\mathbf{y}}\left(\frac{\partial E_t}{\partial b_1}E_t^{-1} + \frac{\partial F_t}{\partial b_1}F_t^{-1} + \frac{\partial f(\mathbf{h}_1)}{\partial b_1}f(\mathbf{h}_1)^{-1} \right), \qquad (142)$$

where the derivation is analogous to that of (122), and comparing it with the formula (113), I have:

$$\frac{\partial \log f(\mathbf{y})}{\partial b_1}|_{\mathbf{H}_0} = -\frac{\partial \log f(\mathbf{y})}{\partial b_2}|_{\mathbf{H}_0} - \frac{t}{b_1} + \frac{1}{b_1^3}tr\left(\mathbf{V}_\rho^{-1}E_{\mathbf{h}_1|\mathbf{y}}\left(\mathbf{h}_1\mathbf{h}_1' \right) \right). \qquad (143)$$

G Monte Carlo Results

Histogram of LM statistic distribution:

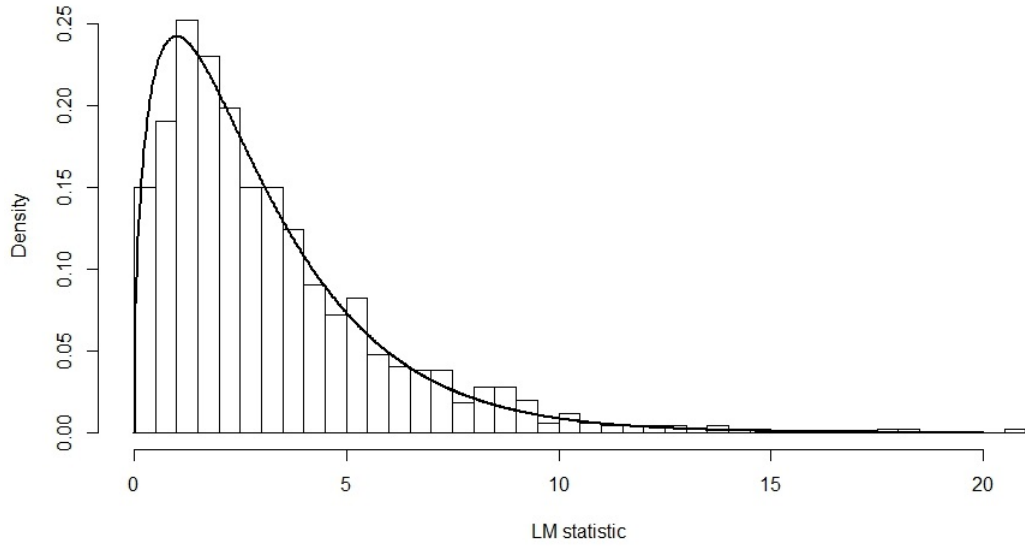Figure 11: Histogram of LM statistic for ψ=0.7

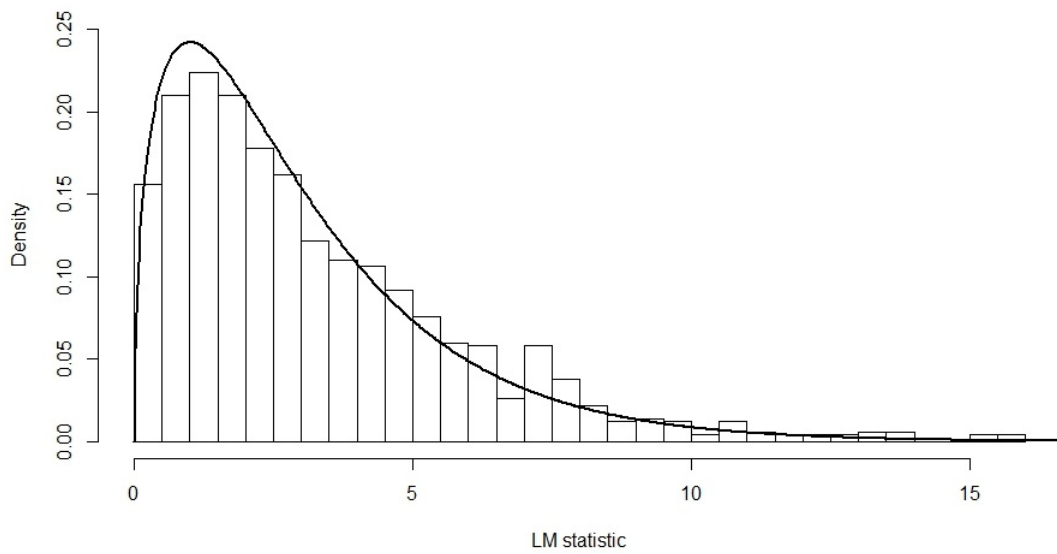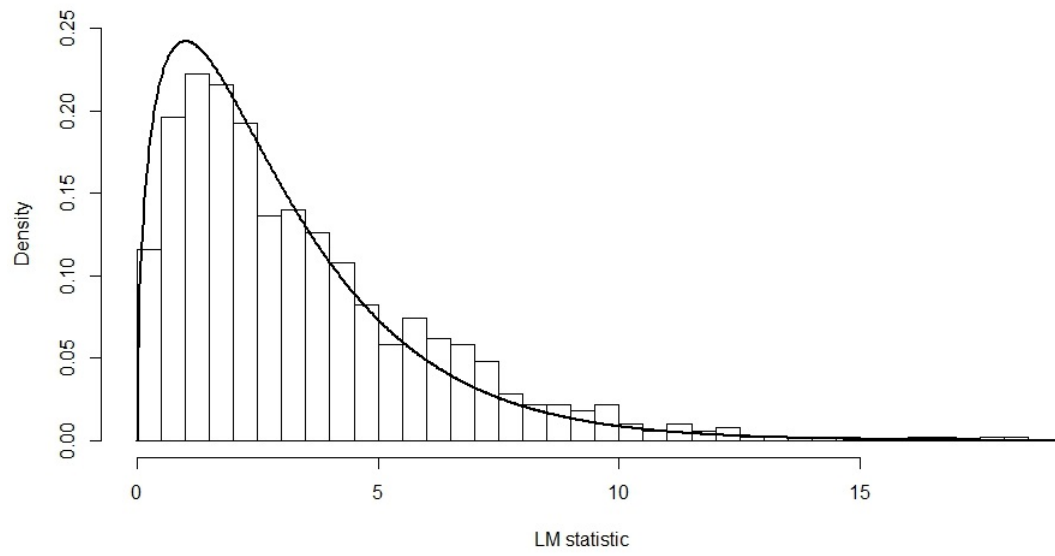Figure 12: Histogram of LM statistic for ψ=0.9

Figure 13: Histogram of LM statistic for ψ=0.95

H Data adjustment

The model to which I can apply the statistic test is:

$$\begin{pmatrix} y_{1t}'' \\ y_{2t}'' \end{pmatrix} = \begin{pmatrix} \exp(\frac{h_1}{2}) & 0 \\ 0 & \exp(\frac{h_2}{2}) \end{pmatrix} \begin{pmatrix} a_1 & 0 \\ a_2 & a_3 \end{pmatrix} \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}. \tag{144}$$

Adding to the correlation between e_{1t} and e_{2t} I have the measurement equation given as:

$$\begin{pmatrix} y_{1t}'' \\ y_{2t}'' \end{pmatrix} = \begin{pmatrix} a_1 \exp(\frac{h_1}{2}) & 0 \\ 0 & a_2 \exp(\frac{h_2}{2}) \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \lambda & 1 \end{pmatrix} \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}, \tag{145}$$

where e_{1t} and e_{2t} remain independent, and λ represents the correlation coefficient. Owing to the identification between (a_1, a_2) and the variance of (e_{1t}, e_{2t}), I set the variance of (e_{1t}, e_{2t}) equal to 1.

Recall that the null hypothesis is $h_{1t} = h_{2t}$ for any t, so that the measurement equation can be rewritten as:

$$\begin{pmatrix} y_{1t}'' \\ y_{2t}'' \end{pmatrix} = \exp(\frac{h_1}{2}) \begin{pmatrix} a_1 & 0 \\ \lambda a_2 & a_2 \end{pmatrix} \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}, \tag{146}$$

and the product and moment of (y_{1t}'', y_{2t}'') would be given as:

$$\begin{pmatrix} y_{1t}''^2 & y_{1t}'' y_{2t}'' \\ y_{1t}'' y_{2t}'' & y_{2t}''^2 \end{pmatrix} = \exp(h_1) \begin{pmatrix} a_1 & 0 \\ \lambda a_2 & a_2 \end{pmatrix} \begin{pmatrix} e_{1t}^2 & e_{1t} e_{2t} \\ e_{1t} e_{2t} & e_{2t}^2 \end{pmatrix} \begin{pmatrix} a_1 & \lambda a_2 \\ 0 & a_2 \end{pmatrix}, \tag{147}$$

$$E \begin{pmatrix} y_{1t}''^2 & y_{1t}'' y_{2t}'' \\ y_{1t}'' y_{2t}'' & y_{2t}''^2 \end{pmatrix} = E(\exp(h_1)) \begin{pmatrix} a_1 & 0 \\ \lambda a_2 & a_2 \end{pmatrix} \begin{pmatrix} a_1 & \lambda a_2 \\ 0 & a_2 \end{pmatrix} \tag{148}$$

$$= E(\exp(h_1)) \begin{pmatrix} a_1^2 & \lambda a_1 a_2 \\ \lambda a_1 a_2 & a_2^2 \end{pmatrix} \tag{149}$$

$$\equiv \Lambda, \tag{150}$$

where Λ is defined as:

$$\Lambda^{-\frac{1}{2}} = (E(\exp(h_1)))^{-\frac{1}{2}} \begin{pmatrix} a_1 & 0 \\ \lambda a_2 & a_2 \end{pmatrix}^{-1}. \tag{151}$$

It follows that:

$$\Lambda^{-\frac{1}{2}} \begin{pmatrix} y_{1t}'' \\ y_{2t}'' \end{pmatrix} = \exp(\frac{h_1}{2})(E(\exp(h_1)))^{-\frac{1}{2}} \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}. \tag{152}$$

At the first adjustment, I define:

$$\begin{pmatrix} y_{1t}' \\ y_{2t}' \end{pmatrix} = \Lambda^{-\frac{1}{2}} \begin{pmatrix} y_{1t}'' \\ y_{2t}'' \end{pmatrix} = a_0 \exp(\frac{h_1}{2}) \begin{pmatrix} e_{1t} \\ e_{2t} \end{pmatrix}, \tag{153}$$

47

where

$$a_0 = (E(\exp(h_1)))^{-\frac{1}{2}}, \tag{154}$$

and the second adjustment is given as:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = a_0^{-1} \begin{pmatrix} y_{1t}' \\ y_{2t}' \end{pmatrix}. \tag{155}$$

Note that:

$$\begin{pmatrix} \log y_{1t}'^2 \\ \log y_{2t}'^2 \end{pmatrix} = 2 \log a_0 + h_1 + \begin{pmatrix} \log e_{1t}^2 \\ \log e_{2t}^2 \end{pmatrix}, \tag{156}$$

$$E \begin{pmatrix} \log y_{1t}'^2 \\ \log y_{2t}'^2 \end{pmatrix} = 2 \log a_0 - 1.27, \tag{157}$$

according to the state variable with zero mean and $E(\log(e^2)) = -1.27$. Therefore, I reach the conclusion:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = a_0^{-1} \begin{pmatrix} y_{1t}' \\ y_{2t}' \end{pmatrix} = a_0^{-1} \Lambda^{-\frac{1}{2}} \begin{pmatrix} y_{1t}'' \\ y_{2t}'' \end{pmatrix}, \tag{158}$$

where the estimates of Λ and a_0 are given as:

$$\hat{\Lambda} = \frac{1}{T} \begin{pmatrix} \sum y_{1t}''^2 & \sum y_{1t}'' y_{2t}'' \\ \sum y_{1t}'' y_{2t}'' & \sum y_{2t}''^2 \end{pmatrix}, \tag{159}$$

$$\hat{a_0} = \begin{pmatrix} \exp\left\{\left[\frac{1}{T}\left(\sum \log y_{1t}'^2\right) + 1.27\right]/2\right\} & 0 \\ 0 & \exp\left\{\left[\frac{1}{T}\left(\sum \log y_{2t}'^2\right) + 1.27\right]/2\right\} \end{pmatrix}. \tag{160}$$

Then the adjusted data (y_{1t}, y_{2t}) can be applied to the proposed test.

I Empirical estimates

Table 15: Empirical estimates between stock markets correspond to $\hat{b}$ from 2011 to 2012

Stocks	DOW	FTSE	DAX	SSCI	NIKKEI	HSI	STI	AORD
DOW		0.1711	0.2103	0.147	0.3169	0.1998	0.1819	0.211
FTSE	0.165		0.1605	0.6314	0.2718	0.1879	0.1384	0.1495
DAX	0.199	0.1632		0.1089	0.3065	0.1676	0.1374	0.157
SSCI	0.1464	0.1016	0.1043		0.6901	0.7408	0.7272	0.7892
NIKKEI	0.318	0.2688	0.3078	0.685		0.3443	0.2865	0.3301
HSI	0.2	0.1904	0.1695	0.7045	0.343		0.1839	0.1576
STI	0.1858	0.1359	0.1374	0.7086	0.2871	0.1968		0.1137
AORD	0.2093	0.1469	0.1577	0.7487	0.3617	0.1768	0.1129	

Table 16: Empirical estimates between stock markets correspond to ψ from 2011 to 2012

Stocks	DOW	FTSE	DAX	SSCI	NIKKEI	HSI	STI	AORD
DOW		0.9717	0.9634	0.9717	0.9178	0.9699	0.9732	0.9674
FTSE	0.9714		0.9793	0.4727	0.9164	0.9732	0.9825	0.982
DAX	0.9657	0.9795		0.9811	0.9141	0.9821	0.9858	0.9833
SSCI	0.9711	0.9817	0.984		0.2652	0.0901	0.09	0.09
NIKKEI	0.9157	0.9147	0.915	0.2833		0.8388	0.8938	0.8648
HSI	0.9694	0.9725	0.9815	0.0902	0.8438		0.9662	0.9724
STI	0.9713	0.9829	0.9856	0.0901	0.8961	0.9609		0.9888
AORD	0.9674	0.9826	0.983	0.09	0.8449	0.9647	0.9885	

Table 17: Empirical estimates between stock markets correspond to b from 2013 to 2014

Stocks	DOW	FTSE	DAX	SSCI	NIKKEI	HSI	STI	AORD
DOW		0.3496	0.3341	0.3013	0.2941	0.2959	0.31	0.3048
FTSE	0.3665		0.289	0.2017	0.2479	0.2441	0.2416	0.2448
DAX	0.3328	0.2922		0.2458	0.1955	0.8115	0.3015	0.2804
SSCI	0.2966	0.1943	0.2476		0.1187	0.7535	0.2666	0.3395
NIKKEI	0.2957	0.2493	0.2046	0.1205		0.116	0.1465	0.1975
HSI	0.2861	0.2459	0.8285	0.1746	0.6946		0.7223	0.4359
STI	0.3108	0.2475	0.3074	0.6258	0.1478	0.6895		0.1905
AORD	0.3053	0.2433	0.2765	0.3251	0.2031	0.4082	0.1878	

Table 18: Empirical estimates between stock markets correspond to ψ from 2013 to 2014

Stocks	DOW	FTSE	DAX	SSCI	NIKKEI	HSI	STI	AORD
DOW		0.8696	0.8769	0.8857	0.9065	0.8611	0.8792	0.8857
FTSE	0.8633		0.8633	0.9502	0.9337	0.8932	0.918	0.9179
DAX	0.8776	0.8672		0.9294	0.9542	0.1669	0.8881	0.8981
SSCI	0.8908	0.9574	0.9269		0.9841	0.24	0.9123	0.8596
NIKKEI	0.9041	0.9342	0.9467	0.983		0.972	0.968	0.9451
HSI	0.8816	0.8928	0.1741	0.9523	0.2281		0.09	0.7186
STI	0.8751	0.9151	0.8847	0.6001	0.9662	0.09		0.9391
AORD	0.8856	0.9182	0.8984	0.8594	0.9419	0.7373	0.9395	

Table 19: Empirical estimates for exchange rates correspond to b in financial crisis

Pairs	EURUSD	USDJPY	GBPUSD	AUDUSD	USDCHF	USDCAD
EURUSD		0.322	0.6375	0.4357	0.5225	0.4753
USDJPY	0.3004		0.4268	0.4564	0.3658	0.5701
GBPUSD	0.5962	0.4318		0.4854	0.5239	0.6284
AUDUSD	0.4413	0.4552	0.4797		0.4342	0.535
USDCHF	0.5745	0.349	0.5816	0.4315		0.5294
USDCAD	0.4799	0.5578	0.6279	0.5191	0.5133	

Table 20: Empirical estimates for exchange rates correspond to ψ in financial crisis

Pairs	EURUSD	USDJPY	GBPUSD	AUDUSD	USDCHF	USDCAD
EURUSD		0.9014	0.7209	0.8183	0.7963	0.8463
USDJPY	0.9101		0.872	0.8523	0.8703	0.8261
GBPUSD	0.7437	0.8703		0.8449	0.8033	0.8038
AUDUSD	0.8089	0.8527	0.8475		0.8239	0.8616
USDCHF	0.768	0.8854	0.7953	0.8368		0.8243
USDCAD	0.8445	0.8281	0.8029	0.8661	0.8259	

Table 21: Empirical estimates for exchange rates correspond to b in normal times

Pairs	EURUSD	USDJPY	GBPUSD	AUDUSD	USDCHF	USDCAD
EURUSD		0.8462	0.8073	0.7858	0.918	0.8892
USDJPY	0.8599		0.8206	0.6888	0.85	0.8186
GBPUSD	0.8037	0.8333		0.7449	0.7302	0.7736
AUDUSD	0.8515	0.7424	0.8307		0.8041	0.8936
USDCHF	0.9163	0.8165	0.7198	0.6969		0.8686
USDCAD	0.8833	0.8113	0.818	0.8534	0.9306	

Table 22: Empirical estimates for exchange rates correspond to ψ in normal times

Pairs	EURUSD	USDJPY	GBPUSD	AUDUSD	USDCHF	USDCAD
EURUSD		0.404	0.644	0.4239	0.4179	0.6104
USDJPY	0.4325		0.5258	0.467	0.4317	0.5272
GBPUSD	0.6537	0.4644		0.5188	0.6825	0.677
AUDUSD	0.4202	0.4001	0.4774		0.4384	0.3643
USDCHF	0.4613	0.4794	0.7036	0.5476		0.6441
USDCAD	0.6384	0.5173	0.6559	0.3455	0.5727	

J   Basic Describe of Stable Distribution

There are certain definitions of stable random variables, here gives a mostly common one.

Definition 1 *Non-degenerate X is stable if and only if for all $n > 1$, there exist constants $c_n > 0$ and $d_n \in R$ such that $X_1 + \cdots + X_n \overset{d}{=} c_n X + d_n$, where $X_1, ..., X_n$ are independent, identical copies of X. X is strictly stable if and only if $d_n = 0$ for all n*

Remember that the sum of random variables usually converge to a normal distribution which is the statement of central limit theorem. A generalization of the theorem is completed in Gnedenko et al. [1954]. Here gives a briefly statement.

Theorem 1 *The sum of a number of random variables with a power-law tail (Paretian tail) distribution decreasing as $|x|^{-\alpha-1}$ where $0 < \alpha < 2$ (and therefore having infinite variance) will*

tend to a stable distribution $f(x; \alpha, 0, c, 0)$ as the number of summands grows.If $\alpha > 2$ then the sum converges to a stable distribution with stability parameter equal to 2, i.e. a Gaussian distribution.

According to this theorem, I model the log-price process in short term time frame. Considering that in market, there are individual and institution investors at the same time. Every transaction they made can be considered as an independent random variable. At short term time frame, I can also consider them as identically random variables. However, because of exchange volume is huge which institution investor made, so the variance random variable should be consider as larger enough to measure the tail probability. So, this does not generality to express variance as infinity. There is another important property with stable distribution random variables. They are closure under convolution.

Property 1 *If $X_1 \sim S(x_1; \alpha, \beta_1, \gamma_1, \delta_1)$ and $X_2 \sim S(x_2; \alpha, \beta_2, \gamma_2, \delta_2)$, then the convolution $X = X_1 + X_2$ follow the distribution of $S(x; \alpha, \beta, \gamma, \delta)$, where*

$$
\begin{aligned}
\delta &= \delta_1 + \delta_2 \\
\gamma &= (|\gamma_1|^\alpha + |\gamma_2|^\alpha)^{1/\alpha} \\
\beta &= \frac{\beta_1 |\gamma_1|^\alpha + \beta_2 |\gamma_2|^\alpha}{\gamma^\alpha}
\end{aligned}
$$

References

M. Asai, M. McAleer, and J. Yu. Multivariate stochastic volatility: A review. *Econometric Reviews*, 25(2-3):145–175, 2006.

J. M. Chambers, C. L. Mallows, and B. Stuck. A method for simulating stable random variables. *Journal of the american statistical association*, 71(354):340–344, 1976.

H. Chernoff et al. On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics*, 25(3):573–578, 1954.

A. Chesher. Testing for neglected heterogeneity. *Econometrica*, 52(4):865–872, 1984.

A. Cipollini and G. Kapetanios. A stochastic variance factor model for large datasets and an application to s&p data. *Economics Letters*, 100(1):130–134, 2008.

J. W. Cooley and J. W. Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of computation*, 19(90):297–301, 1965.

J. Danıelsson. Multivariate stochastic volatility models: estimation and a comparison with vgarch models. *Journal of Empirical Finance*, 5(2):155–173, 1998.

A. Doucet and A. M. Johansen. A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of Nonlinear Filtering*, 12(656-704):3, 2009.

J. Durbin and S. J. Koopman. *Time series analysis by state space methods*. Number 38. Oxford University Press, 2012.

R. F. Engle and S. Kozicki. Testing for common features. *Journal of Business & Economic Statistics*, 11(4):369–380, 1993.

R. F. Engle and R. Susmel. Common volatility in international equity markets. *Journal of Business & Economic Statistics*, 11(2):167–176, 1993.

J. Fleming, C. Kirby, and B. Ostdiek. Information and volatility linkages in the stock, bond, and money markets. *Journal of Financial Economics*, 49(1):111–137, 1998.

B.-V. Gnedenko, A.-N. Kolmogorov, K.-L. Chung, and J.-L. Doob. Limit distributions for sums of independent random variables. 1954.

L. Greengard and J. Strain. The fast gauss transform. *SIAM Journal on Scientific and Statistical Computing*, 12(1):79–94, 1991.

A. Harvey, E. Ruiz, and N. Shephard. Multivariate stochastic variance models. *The Review of Economic Studies*, 61(2):247–264, 1994.

B. M. Hill et al. A simple general approach to inference about the tail of a distribution. *The annals of statistics*, 3(5):1163–1174, 1975.

R. E. Kass and A. E. Raftery. Bayes factors. *Journal of the american statistical association*, 90(430):773–795, 1995.

A. Le Breton and M. Musiela. A generalization of the kalman filter to models with infinite variance. *Stochastic processes and their applications*, 47(1):75–94, 1993.

S.-M. Liu and B. W. Brorsen. Maximum likelihood estimation of a garch-stable model. *Journal of Applied Econometrics*, 10(3):273–285, 1995.

B. Mandelbrot. The variation of certain speculative prices. *The Journal of Business*, 36(4): 394–419, 1963.

S. Mittnik, S. T. Rachev, and M. S. Paolella. Stable paretian modeling in finance: Some empirical and theoretical aspects. 1998.

J. P. Nolan. Numerical calculation of stable densities and distribution functions. *Communications in statistics. Stochastic models*, 13(4):759–774, 1997.

S. T. Rachev. *Handbook of Heavy Tailed Distributions in Finance: Handbooks in Finance*, volume 1. Elsevier, 2003.

J. H. Stock and M. W. Watson. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460):1167, 2002.

B. Stuck. Minimum error dispersion linear filtering of scalar symmetric stable processes. *Automatic Control, IEEE Transactions on*, 23(3):507–509, 1978.

G. E. Tauchen, M. Pitts, et al. The price variability-volume relationship on speculative markets. *Econometrica*, 51(2):485–505, 1983.

T. Watanabe. A non-linear filtering approach to stochastic volatility models with an application to daily stock returns. *Journal of Applied Econometrics*, 14(2):101–121, 1999.