ECONOMETRIC METHODS FOR IMPROVED MEASURES OF FINANCIAL RISK

A Dissertation

Presented to

The Faculty of the Graduate School

At the University of Missouri-Columbia

In Partial Fulfillment

Of the Requirements for the Degree

Doctor of Philosophy

By

MOOHWAN KIM

Dr. Douglas J. Miller, Dissertation Supervisor

July 2011

The undersigned, appointed by the Dean of the Graduate School,

have examined the dissertation entitled

ECONOMETRIC METHODS FOR IMPROVED MEASURES OF FINANCIAL RISK

Presented by Moohwan Kim

A candidate for the degree of

Doctor of Philosophy

And hereby certify that, in their opinion, it is worthy of acceptance.

Professor Douglas J. Miller

Professor Peter Mueser

Professor Saku Aura

Professor J. Isaac Miller

Professor Nicholas Kalaitzandonakes

I would like to thank my mother for giving me encouragement. My wife, Miji and our two children, Connie and Sean were with me all the time during the long years I have been studying. I would like to share this achievement with them.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Douglas J. Miller for his advice through my entire dissertation work. I am very thankful to Mr. Ray Bacon for his hands-on help with data handling. I would like to acknowledge the committee members, Dr. Mueser, Dr. Aura, Dr. J. Issac Miller and Dr. Kalaitzandonakes for careful reading of the draft and valuable suggestions which lead to improvement in my work.

I have completed my dissertation under Dr. Douglas J. Miller's supervision. I truly appreciate his advising. He has taught me what I needed to know largely in two ways: one is a direct way. Every time I asked some question about a particular issue, he answered them without any hesitation. In the midst of doing research, whenever I faced some obstacle, he suggested an alternative way to overcome the difficulty.

The other is an indirect way. I think he intentionally designed it in a way that I must realize for myself what I had to know. Sometimes he paused and waited until I made some progress. In the meantime, I had to find a clue and apply it to the analysis. By doing that, the research process came to reasonable outcomes through tons of trials and errors so that I amounted to accumulate much basic knowledge. He had me exposed to such a situation where I wandered in the woods to find something by myself. Although this course needed much time, it helped me being an independent researcher.

Based on both explicit and implicit learning processes, I have learned that there is something that cannot be transformed by words during our two-way communication process. That is, sometimes I had to interpret what he implied (usually not expressed in

words) to make a little progress. Repeating the process continuously, I have realized that conducting research in some respect is to find some hidden principle that underlies visible leading economic indicators or phenomena through sharp sense or mature intuition. In other words, I have to read between the lines to understand what the observed economic facts are implying. Although the task is not easy at all, I will keep it in mind for the rest of my research experience (or career) and make best efforts to get such insights realized. I would like to thank my advisor for these teachings and guidance with a deep bow.

Table of Contents

List of Figures

List of Tables

Abstract

Chapter 1. Improved Measures of Financial Risk for Hedge Funds

During the current financial crisis, several US and foreign banks and investment firms have failed due to excessive losses in some of their investments. Many of these financial institutions relied on a widely-used risk model known as the Value-at-Risk (VaR) to gauge the risks taken by their businesses, and several authors have pointed to some key flaws in the VaR measure that tend to understate the risks that these firms actually faced. In particular, VaR does not properly account for the joint risks of investments based on more than one asset (i.e., the risk measure is not subadditive). Also, VaR computations are commonly based on the assumption that the probability distribution of asset returns is normal (Gaussian), which understates the probability of encountering large losses for some investments.

To overcome the subadditivity flaw of VaR, researchers in financial economics have proposed an alternative measure, the Conditional Value-at-Risk (CVaR), which is defined as the expected losses that are strictly larger than the VaR. The CVaR measure may more appropriately compute the potential losses associated with holding two or more assets.

The purpose of this paper is to evaluate two recent innovations in the financial economics literature that may help banks and investment firms to properly assess the risks they face. First, we employ Extreme Value Theory (EVT) to estimate non-normal models of the return distribution tails. In particular, we use the peaks-over-threshold (POT) method in which extremes are defined as excesses over a threshold, and we

estimate the marginal (univariate) return distributions. The POT observations are used to estimate the Generalized Pareto (GP) model of the upper and lower tail areas of the return distributions. Second, we use the estimated GP models to compare the relative performance of the VaR and CVaR for assessing forward-looking risk in observed hedge fund returns. The main objective of this analysis is to evaluate competing claims from the financial economics literature about the relative importance of the VaR flaws (e.g., subadditivity) and probability model specification errors in risk measurement.

Chapter 2. Optimal Hedging under Copula Models with High Frequency Data

The current financial disturbance is partly due to the failure of risk management tools to warn of rapidly-evolving market events. One important lesson gained from this experience is that we must improve our ability to manage such financial risks by developing a better understanding of the microstructure of financial markets. Accordingly, we may be able to use models based on high frequency (e.g., intra-day) financial data to better assess the current risk of financial positions and to improve our predictions of future price movements. However, there are special problems associated with modeling high frequency data -- for example, previous research shows that high frequency returns might be correlated in a nonlinear fashion. To handle these problems, we may use copula-based probability models, which represent the dependence structure and the univariate marginal properties of the risky asset returns. The methodology turns out to be useful in such a way that multivariate non-normality is readily modeled and the associated correlation parameter is easily updated on the basis of time varying structure. We estimate these models in order to determine optimal hedge ratios for currency futures

positions used to manage price return risks in spot exchange rates. A Dynamic hedging strategy with futures contracts is considered to allow the hedge position to be adjusted over time. With the dynamic strategy, hedgers should be able to improve hedging performance over a static hedging strategy, which keeps a constant hedge ratio over time. Various GARCH models are used to capture the volatility of the value of short futures positions coupled with foreign exchange rate fluctuations. For the purpose of measuring the conditional dependence between the two asset returns in a GARCH context, we use the Copula-based GARCH models. We compare the performance of the alternative dynamic hedging models with the hedging effectiveness of the static model.

Chapter 3. Copula Model Selection Based on Non-Nested Testing

The copula approach adopted in Essay 2 may be used to model any multivariate probability distribution by separately estimating the marginal distributions and the dependence structure. In practice, one needs to choose an appropriate copula model (from the many candidates) that provides the "best" fit to the observed data. In the statistics and econometrics literature, researchers have proposed various model selection procedures, and examples include L^2-norm distances, graphical selection procedures based on Kendall's process, adjusted R^2 statistics, and likelihood-based procedures such as the pseudo-likelihood ratio tests. If the candidate models are not members of the same copula family, we must use model selection procedures that are specifically designed to handle non-nested models. We propose a non-nested test procedure for copula model selection that is based on the Cox test statistic, which is a centered version of the standard LR statistic. That is, the test statistic combines the standard LR statistic with the expected

log-likelihood ratio statistic, which is the Kullback-Liebler information criterion (KLIC) measure of closeness between two non-nested models. The Cox test and related non-nested testing methods hold conceptual advantages over the alternative tools mentioned above, but these methods are not widely used in practice due to computational difficulties. To resolve some of these practical challenges, we could use Monte Carlo sampling methods for computing the Cox test statistic and evaluating its distributional properties. The objective of this research is to propose a model selection procedure that is computationally feasible and statistically reliable in order to facilitate applications of these improved risk models in practice.

1 Improved Measures of Financial Risk for Hedge Funds

1.1 Introduction

This study of risk management in the hedge fund industry is concerned with how new methodologies to manage risks compare to traditional methods such as Value-at-Risk (VaR). Improved techniques are potentially important from a risk management standpoint because they provide an opportunity for financial institutions and investment firms to better manage various risks.

Historical experience has shown abundant evidence of the potential risks, and these may be illustrated with the list of large-scale financial crises. For example, extensive U.S. bank failures occurred later in the Great Depression of 1929 and did harm the entire economy. In the summer of 1997, the Asian financial crisis broke out. In August 1998, the Russian government defaulted on their debt, and Russian ruble faced difficulties. The dot-com bubble burst in 2000 and the trouble extended into 2001. In the summer of 2007, big losses on U.S. subprime mortgage loans sparked financial stress, which was widespread across other sectors.

Faced with the possibility of large financial panics and smaller adverse events, every market participant is exposed to persistent and potential risks of loss associated with unexpected changes in asset values. From the standpoint of risk management, better measures of risk are strongly required to avoid substantial losses. The most widely used risk model is VaR. In words, VaR is defined as the maximum value that may be lost over a certain period (e.g., one month) with a given probability or confidence level (a technical

definition is provided later). The measure is commonly used to determine how much capital should be set aside to control the risk facing financial institutions or investment firms at the stated confidence level. The VaR measure has three main drawbacks. First, VaR is generally not subadditive in the sense that the risk of a combined portfolio may be larger than the sum of the individual risk portfolios. Non-subadditivity of VaR may lead to bias when assessing the value of diversification of a portfolio. Second, VaR is not convex, which may hamper the search for risk optimized portfolios. Third, the most common VaR applications are based on the normality assumption of asset returns, which is not appropriate for most assets and can lead to understated risk estimates.

Due to the presence of the drawbacks contained in VaR, researchers have developed Conditional Value-at-Risk (CVaR), which is defined as the expected losses that are strictly larger than VaR (a formal definition is provided later). The CVaR measure may more appropriately represent the potential losses associated with holding two or more underlying assets because it is subadditive. The improved risk measure may be useful to enhance financial institutions' risk management operations. Here the meaning of "improved" is based on an axiomatic approach that characterizes a coherent risk measure. Unlike VaR, CVaR satisfies the axioms and is a coherent risk measure. It provides potentially more valid estimates of the capital requirement for corporate and financial institutions relative to VaR.

Efforts to build probability models that capture the behaviors of extreme risk factors have been made over a long period of time. The extreme events such as financial panics are rare and thus it is hard to draw reliable inferences based only on historical

data. Conceptually, the extreme events are represented by the tail areas in the return. For example, the fat-tailed character of the distribution implies that the extremely rare events occur with higher probability than under the normal distribution. As a result, the normal distribution model of risk might understate the true risk.

To deal with those problems, researchers developed Extreme Value Theory (EVT) as described by Embrechts, Klüppelberg and Mikosch (1997). EVT is used to model the extreme quantiles or tails of the probability distribution for some random variable. As an application of the proposed methods in this paper, hedge fund returns are used to compare the relative performance of VaR and CVaR for assessing forward-looking risk. The estimated risk measures are based on the Generalized Pareto (GP) model, which is a prominent EVT tool. For instance, given an upper threshold, we fit the observed extreme values of asset returns over the threshold to the distribution, which is called peak-over-threshold (POT) approach. Then VaR is estimated from the fitted GP model, and CVaR is calculated by the expected value of losses over VaR.

There are several reasons that we focus on the hedge fund industry for this application. Since the 1990s, the hedge fund industry has grown rapidly and this is partly due to financial globalization. The primary motivation is to find an opportunity for higher profits by moving beyond traditional investment vehicles. Hedge funds are limited to wealthy individuals and institutional investors who meet a minimum standard of wealth (e.g., 5 million dollars). Due to the impressive profits, the performance fee charged by hedge fund managers is relatively high and may be more than 20 percent, which is in addition to a management fee of 2 percent of the net asset value. The outstanding

performance (i.e., commonly double-digit annual returns) of hedge funds stems from their use of long/short [1] trading strategies and investment in illiquid assets. While traditional investment funds gain value from asset appreciation, hedge fund investments may be based on long or short positions and may gain value from either rises or declines in asset prices.

Traditionally, a long investor will profit if the price of the financial instruments increases and a short investor will profit from a decrease in price. To take advantage of an expected lower price in the future, hedge fund managers can borrow shares from their brokerage firms or banks and short sell them at the current price. At some point in the future, the fund managers must return the shares and the interest to the brokerage firms or banks. If the price declines, the managers will take a profit. However, they will take a loss if the price rises. Thus, hedge funds can generate higher returns than traditional long-only investment funds by taking advantage of both asset value declines and increases. At this point, it is important to note that the use of the term "hedge" to describe these funds may be distinct from the practice of hedging risks with forward or futures contracts. Although hedge funds can use traditional hedging strategies to reduce risk, this activity is not their main objective.

More broadly, the difference between a traditional long-only investing strategy and a hedge fund strategy is that hedge funds can use flexible tools such as short-selling and

[1] An investor is referred to as taking a short position if he or she sells futures contracts and a long position if he or she purchases futures contracts.

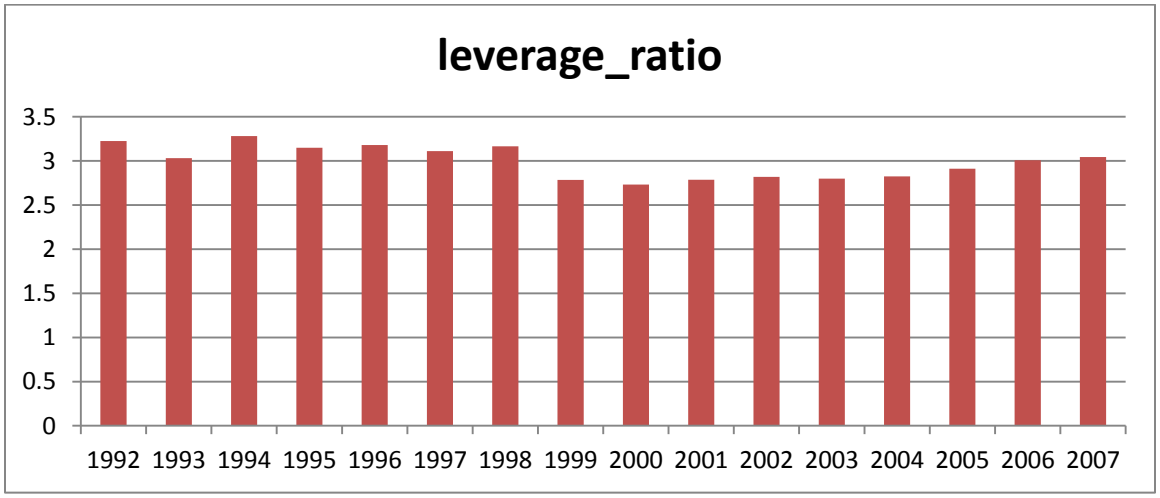
derivatives to reduce risk and generate returns. For example, the equity long/short strategy engages in short positions to hedge market exposure. This means that hedge funds have more diverse tools relative to long-only investing. However, hedge fund managers may take on more risk by using leverage, which may result in the expansion of risk and return.

Hedge funds borrow in order to increase the size of the investment portfolio and gain better performance when the underlying asset value rises. The practice of borrowing is called leverage. For example, borrowing to buy a stock is called a margin loan. In general hedge fund leverage is known as being above investor capital. Unlike mutual funds that operate under the Securities and Exchange Commission (SEC) regulations, hedge funds have great exemptions from the usual regulations. Although short selling is theoretically useful because it increases market efficiency (Renshaw 1977), it can cause financial distress. Some hedge fund strategies simultaneously buy and sell a security or other financial instruments in two different markets to profit from the price anomalies in the two different markets, which is called arbitrage. For example, in bond arbitrage, hedge funds buy illiquid bonds that may be mispriced and capture arbitrage profits by short selling derivatives based on more liquid bonds that are accurately priced.

Figure 1-1 shows the leverage ratios of hedge funds over time. The leverage ratios are calculated as the ratio of average gross leverage plus average assets under management (AUM) to average assets under management. According to this result, the weighted average hedge fund has constant leverage ratios of 2.5 times over the average assets under management. The ratio of three to one leverage means that a hedge fund

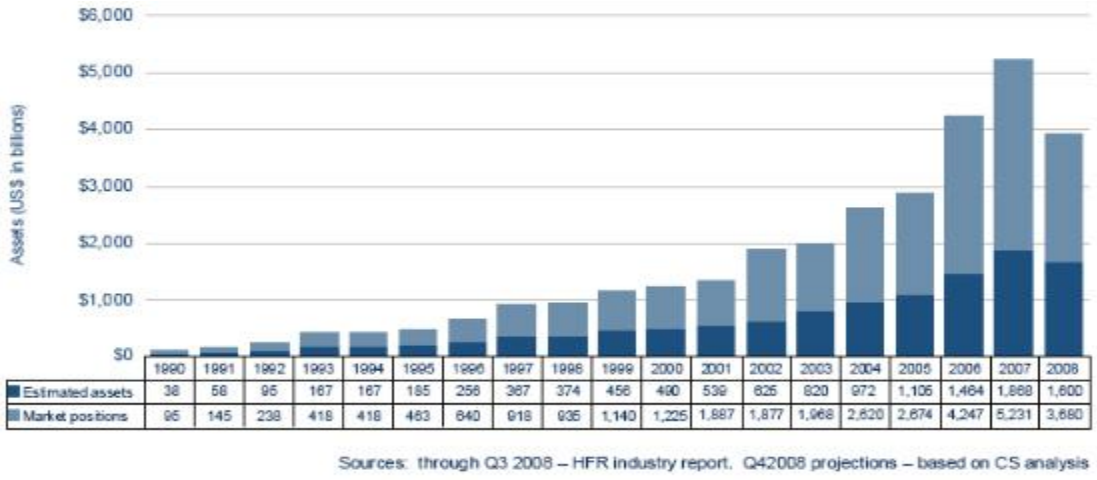
manager controls $3 million in the investment portfolio with $1 million in capital. As the ratio increases, an increase in asset values leads to an expanded gain. However, in the opposite case, the losses would be severely magnified. We can see the gradual increasing pattern since 1999, and the leverage ratio reached three times over the AUM in 2007.

Figure 1-1 Hedge Funds' Leverage Ratio with CISDM database



Along with Figure 1-1, it is important to take a look at the growth of assets and hedge fund leverage from 1990 and 2008, as cited from Lo (p.15, 2009). Figure 1-2 shows the nearly monotone increase in leverage since 2002. The vertical bars in light color denote market positions containing estimated assets in hedge funds. There is a gradual rise in assets and leverage up until 2007. Note the quick decline in 2008 after the financial crisis started in mid-2007. Facing large investment losses and sizable redemptions by investors, hedge funds were forced to deleverage the relevant assets.

Figure 1-2 Growth of assets and hedge fund leverage with HFR dataset



	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Estimated assets	38	58	95	167	167	185	256	367	374	456	490	539	825	820	972	1,105	1,464	1,868	1,600
Market positions	95	145	238	418	418	463	640	918	935	1,140	1,225	1,887	1,877	1,968	2,620	2,674	4,247	5,231	3,660

Note: The graph above is cited from p.15 in Lo (2009).

The aggressive investment practices of hedge funds bring up big concerns because of their impacts on the global financial system which may offset their crucial role in liquidity supply and price discovery. There are debates about sensible regulations for hedge funds due to the spillover effects to other sectors. Some people insist that hedge funds be regulated due to concerns about their opaqueness. On the other hand, advocates assert that hedge funds make financial markets more efficient (Ackerman, McEnally and Ravenscraft, 1999; Danielsson and Zigrand, 2007). At this time when the U.S. financial crisis of 2007-2009 recedes, it seems that there is no general agreement on regulations of hedge funds.

It is essential to understand the source of the returns embedded in each hedge fund strategy to improve the risk management of the hedge funds. Hedge fund returns have two main components. The first part can come from diversified risk exposures such as credit risk, liquidity risk, exchange rate risk, or event risk. Hedge funds that have

different risk profiles are compensated for the risks in the form of risk premiums. For example, a merger arbitrage manager can obtain an event (e.g., merger and acquisition) risk premium since there is a possibility that a merger would not occur after the merger's announcement. The second component is the manager's ability to exploit the excess market return caused by market inefficiencies such as price anomalies which lead to arbitrage opportunities. For example, currency based macro strategies would borrow a currency with a low interest rate and then purchase currencies with a high interest rate (i.e., carry trade). The strategy works fine with good market conditions but it causes trouble when exchange rates shift.

Hedge fund investors or managers consider risk management seriously. As shown in Lo (2008, p.238), considering risk management with a financial gain through an investment strategy at the same time might be profitable. For a simple illustration, he assumes that there is a fund that provides an annual expected return $E[R]$ of 10% and an annual volatility of 75%. As a risk management practice, suppose a fund manager can avoid losses below -20%. His new return is

$$R^* = \max[R, -20\%] \tag{1-1}$$

Based on the log-normality assumption of returns, the expected value of R^* is 20.9% by truncating the left tail of the return distribution (see Table 9.1, p.240). The resulting expected return doubles. It is very crucial to have the ability to manage risks in a hedge fund operation.

There are common risks embedded in hedge fund strategies as described by the U.S. Securities and Exchange Commission (2003). According to the classification, the

risk can be divided into five branches: market, credit, leverage, liquidity, and operational risk. Here we focus on some common risks to several hedge fund strategies. First, hedge fund managers are exposed to leverage risk. They need to borrow some money from their partners or banks to increase their performance. When investments go bad, big losses may occur. Furthermore, those who use leverage will take on the risk that one has to pay back when the lender requests the loan. From Figure 1-2, we can observe a sharp decline in assets and leverage in 2008 right after the subprime crisis occurred in the mid-2007.

Second, there is liquidity risk. When an economic disturbance occurs, there may be many investors who wish to exit a hedge fund. Faced with this situation, hedge funds may be forced to sell positions at prices lower than their real value. Subprime mortgages and related mortgage-backed securities suffered large losses during 2007. As a result, one of the five largest investment bank in the United States, Lehman Brothers, announced bankruptcy on September 14, 2008. This triggered a liquidity risk in the hedge fund sector because their prime brokers were reluctant to fund these positions.

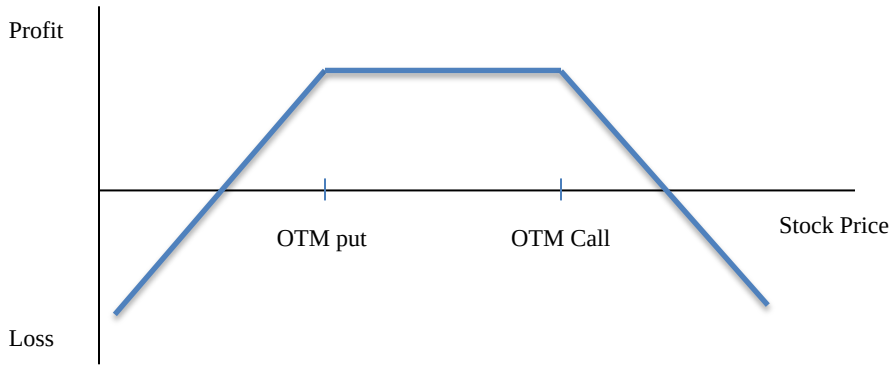
Third, the failure of Long Term Capital Management (LTCM) in 1998 is often described as an example of credit risk. A default by Russia on its debt cast serious doubt on the quality of the fund. LTCM responded by going long in illiquid assets (i.e., bonds) and short in the corresponding liquid bonds. After the Russian default, the spread between the two underlying assets widened severely. LTCM was highly leveraged and suffered from huge losses after receiving margin calls on its positions. Unfortunately, LTCM was not able to meet those calls. Note that even though the liquidity and the credit risk are separate exposures, they are interconnected as in the case of the LTCM collapse.

Fourth, hedge fund managers may be exposed to counter party risk. In the financial affair of LTCM, Bear Sterns, LTCM's prime brokerage firm, faced a large margin call as the value of the portfolios deteriorated. It was at risk if LTCM did not meet further margin calls.

Finally, specific event risks are the possibility that individual events that have not happened before would occur. The use of derivative securities such as options makes it possible to trade volatility. For example, when the hedge fund managers do not know how the underlying stock price changes up or down, the short strangle strategy may be pursued by some hedge fund managers. The strangle strategy in options trading is the simultaneous selling of an out-of-the-money (OTM) put and an out-of-the-money call with the same expiration but with different strike prices. It can be used when we expect the market volatility to stay in a bounded range before expiration. In Figure 1-3, the blue solid line denotes the profits of a short strangle strategy. The put option is said to be out-of-the-money (OTM) when its strike price is below the current stock price and a call option is said to be OTM when its strike price is above the current stock price. Writing (or selling) an OTM call option can be profitable when the underlying stock at expiration is below a call's strike price. Similarly writing an OTM put option can be profitable when the underlying stock at expiration is above a put's strike price. Hence, only if the underlying stock stays in between the two strike prices, we can profit. In the case of high volatility, the stock can rise or decline quite far from the option strike prices, and we need to close the short strangle position by buying the in-the-money option. However, it may be very expensive to buy back, and we may be exposed to unlimited risk. The hedge fund

managers can use this short strangle strategy, but when the market is volatile, large loss can occur.

Figure 1-3 Short Strangle



Many studies such as Brooks and Kat (2002) and Geman and Kharoubi (2003) find that hedge fund returns are not normally distributed. Most hedge fund indices show not only high negative skewness, which is a measure for the asymmetry around the mean, but also high kurtosis, which is a measure of tail thickness. The latter implies that there are more observations on the tails than a normal distribution. Thus, both statistics indicate that those returns are not normally distributed.

The purpose of this paper is to evaluate two recent innovations in the financial economics literature that may help hedge fund investors and managers to properly assess the risks they might confront. We will use CVaR risk measure and EVT to estimate the risk of hedge fund returns under the constituent strategies. The study will directly examine competing claims about the relative importance of VaR flaws and probability model specification errors in risk measurement.

The paper is organized as follows: The next section provides a detailed literature review. The subsequent three sections briefly discuss risk measures and then those risk measures apply to two different parametric specifications on extreme values. The sixth section describes the data and presents empirical results from various methods. Finally, concluding remarks are presented.

1.2 Literature Review

Value-at-Risk (VaR) was developed by JP Morgan. In October 1994, J.P. Morgan and Reuters (1996) made its RiskMetrics freely available on the internet, which contributes to the wide use of VaR. There are a few other significant merits. VaR can be applied to any asset class, including equities, bonds and derivatives. Also, VaR provides a means to aggregate the component risks of a given trading desk. Furthermore, VaR is expressed by potential lost money, which is easy to interpret.

Despite these advantages, there are criticisms as well. Taleb (1997) claims that the method may not accurately measure tail probabilities. While arguing that forecasts based only on past observations is naïve, he writes "Nothing predictable can be truly harmful and nothing truly harmful can be predictable." Ju and Pearson (1999) show that the use of VaR in controlling risks at the level of individual traders or trading desks leads to large biases.

Due to the vulnerability of the financial system stemming from potential financial crises, the international banking forum (i.e., Bank for International Settlements (BIS)) has built guidelines and standards for banking supervisory issues. The committee strengthens

the 2001 Basel II capital framework, which is based mainly on VaR. However, there are some arguments by Danielsson, Embrechts, et al. (2001) against contents suggested by the Basel II proposal. They point out that VaR has limitations as a risk measure and the statistical model can generate inconsistent and biased forecasts of risk. These criticisms highlight the significance of a better understanding of extreme risks in order to reduce them.

In terms of risk measurement, a seminal paper by Artzner, et al. (1999) is a turning point. Since this work, the concept of risk measures becomes refined and stands on more solid ground. There are four axioms that risk measures must satisfy to be called coherent measures. The axioms of coherence are translation invariance, subadditivity, positive homogeneity, and monotonicity. According to the proposed standards, the widely used VaR turns out not to be a coherent risk measure because it violates subadditivity. In contrast, CVaR satisfies all properties required for coherency.

The application of EVT in the finance and insurance literature begins from Embrechts, Klüppelberg and Mikosch (1997) and Reiss and Thomas (1997). The two books explain the theoretical background and show how the EVT methods can be used to estimate the risk measures in applications. For example, Longin (1996) uses the maxima of daily returns for U.S. stock indices, which is modeled with the Fréchet distribution. The Fréchet distribution is one of the standard generalized extreme value (GEV) distributions.

Danielsson and Vries (2000) analyze the extreme returns of the S&P 500 index using unconditional extreme value theory along with the Generalized Autoregressive

Conditional Heteroskedastic (GARCH) model. Use of the unconditional[2] distribution of asset returns is criticized by McNeil and Frey (2000), who conclude that the method based on the conditional return distribution provides better estimates of VaR and CVaR than unconditional EVT. However, in cases where the conditional distributions taking into account all values such as the GARCH model is employed, there is still the possibility for large unexpected events to be ignored as in Longin (2000). Danielsson and Vries (2000) argue that for the portfolio constructed from a number of assets, the huge conditional variance matrices can make the conditional approach infeasible. They suggest that there are cases where an unconditional EVT method is suited for some financial low-frequency (i.e., daily or weekly etc.) data. In contrast, they propose that the conditional volatility models may perform better over short horizons, such as intra-day data. Christoffersen and Diebold (2000) present findings that agree with Danielsson and Vries (2000).

Regarding applications to hedge fund returns, VaR has been used with EVT to control the market risk associated with hedge funds. Gupta and Liang (2005) use VaR to determine the equity capital requirement for hedge funds based on EVT. Lhabitant (2003) estimates VaR and CVaR using two funds of hedge funds on a monthly basis with an

[2] The probability that an event will occur can be characterized without any information obtained from past and related events. On the other hand, the conditional approach utilizes past events to model the distribution of future outcomes.

EVT model. Blum, Dacorogna and Jaeger (2004) calculate VaR to capture the tail risk measure from a Generalized Pareto (GP) model.

In this chapter, we rely on the unconditional volatility approach to compute the standard deviation required for VaR and CVaR. The data for this application is low-frequency (i.e., weekly or monthly) hedge fund returns. With regard to applications to risk management, McNeil and Frey (2000) show that the GP distribution of EVT works better for CVaR than the Gaussian model. In the same vein, Fernandez (2003) verifies that EVT outperforms a GARCH model with normal innovations using the Chilean daily stock index conditional on restriction of the shape parameter (e.g., $\xi > -0.25$). However, these findings depend on the data under study and may not be generalized.

Hedge fund operations are differentiated by the investment vehicles that they can provide to investors. Boudt, Peterson and Croux (2008/2009) document that modified expected shortfall (a variant of CVaR) based on portfolios that outperform the fund-of-fund index. Fung and Hsieh (2001) show that hedge funds differ from mutual funds, which usually keep long-only buy-and-hold strategy. Agarwal and Naik (2000) report that the hedge fund strategies outperform the traditionally diversified investment method by more than 6 percent. It suggests that hedge funds provide better opportunities for diversification by their low correlation with different indices and even different asset returns, such as stocks and bonds.

Some studies deal with hedge fund characteristics and management styles. Compared to the performance of mutual funds, hedge fund returns generate less correlation with those of standard asset classes (Fung and Hsieh 1997). Specifically while

hedge fund returns have low and occasionally negative correlation with other asset returns (e.g., stock market index), mutual fund returns are highly and positively correlated with asset class returns. In relation to the impact on crises, Fung and Hsieh (2000) conclude that hedge funds seem not to have had a major role.

Attempts to estimate historical returns for hedge funds are recognized in the literature to have several biases such as survivorship bias or backfill bias. Hedge fund managers do not disclose their performance to the public. They have an incentive to report performance to the data vendors only when they have relatively good results. Thus, persistently successful funds tend to be contained, causing survivorship bias. There is also instant history bias. In the case where a new fund manager starts reporting, it is more likely for recent and good performance record to be included in the database. Accordingly, they tend to highly exaggerate the returns. This observation is backed by several studies including Ackerman, McEnally and Ravenscraft (1999), Brown, Goetzmann and Ibbotson (1999), and Schneeweis, Sputgin and McCarthy (1996).

Regarding the distributional characteristics of hedge fund returns, several studies (Brooks and Kat 2002, Bacmann and Gawron 2005, Agarwal and Naik 2004) find that hedge fund index returns are often not normally distributed as skewness and kurtosis are significantly identified and are significantly correlated with each other. Amin and Kat (2003) find that skewness decreases and kurtosis rises with portfolios containing hedge funds. Due to the non-normality of hedge funds returns, Gupta and Liang (2005) cast doubt on the validity of the normal VaR, which is usually used to evaluate the validity of the capital adequacy of hedge fund operations (Jorion 2000). An interesting point is that

despite the claim that hedge funds are uncorrelated with market indices, some studies (Brooks and Kat 2002, and Agarwal and Naik 2004) show the existence of high correlation with the equity indices.

1.3 Risk Measures

In general, risk is explained by its exposure and uncertainty (Holton 2004). In financial management, exposure is characterized by financial loss and uncertainty that occurs because we are not aware of future events. So uncertainty can be described by a probability distribution of future values (i.e., prices or returns) of the financial assets, which are represented by random variables. Risk, which is subjective, is related to uncertainty[3] but they are not identical (Rachev, Stoyanov and Fabozzi 2008). A classical uncertainty measure is the standard deviation which is the square root of the variance. Despite such difficulties we can attempt to formalize it by relying on probabilistic models denoted by the portfolio loss distribution. To quantify a risk of a financial position that is regarded as a random variable, the risk measure is defined as a mapping of portfolio losses into real numbers. Let X be a random variable on the probability space $(S, \mathcal{F}, P)$. In other words, risk measures are real-valued functions

$$\rho(X): S \to \mathbb{R} \tag{1-2}$$

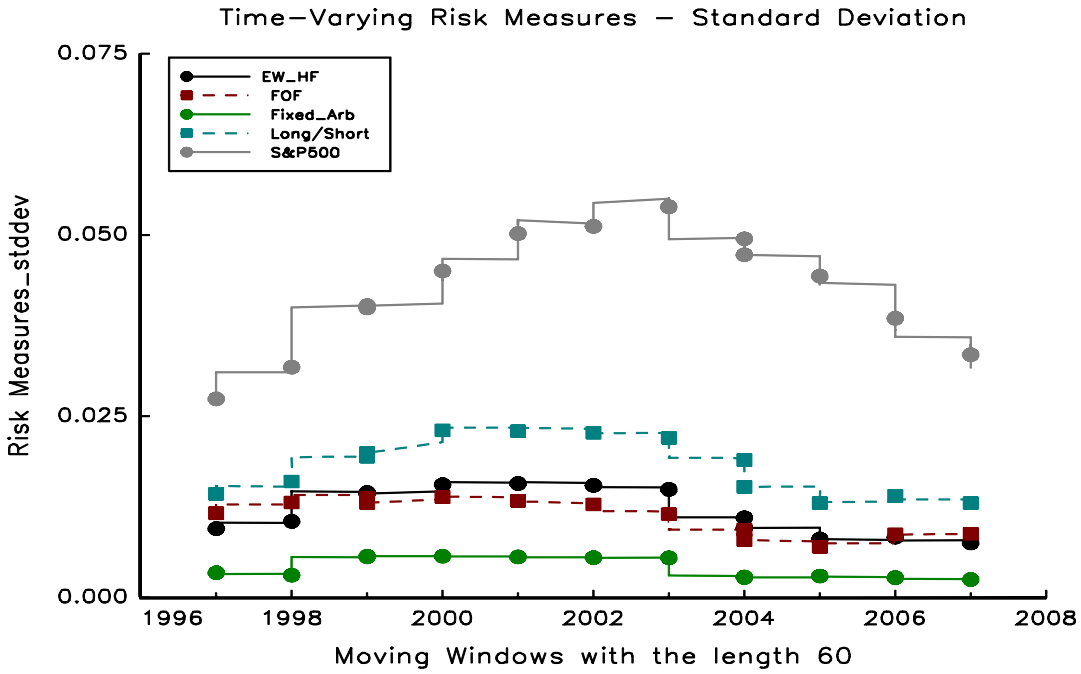
defined on S, the random loss from financial positions.

[3] Uncertainty stems from the inability to predict the future values of assets of interest. Therefore it can be formed by a probability distribution of future prices or returns.

1.3.1 Standard Deviation

The standard deviation is used to quantify the amount of volatility. In a statistical term, the standard deviation is a measure of dispersion around the mean. As the standard deviation gets higher, the volatility becomes larger, and vice versa. The standard deviation can be used to measure the risk of hedge fund returns over a specified time period. However, it is very sensitive to extreme values, which may be more likely when hedge fund returns are not normally distributed. When we incorrectly use the normal distribution, we can underestimate the possibility of large losses.

Figure 1-4 Volatility with Selected Hedge Fund and S&P500 Returns



In Figure 1-4, we compute the time-varying volatility. The lowest curve is obtained from the returns of fixed arbitrage strategy. The next two lines correspond to fund of fund (FOF) index and equal weighted hedge fund portfolio returns. The fourth curve from the

bottom represents the standard deviation of an equity long/short returns. In contrast to the hedge fund returns, the volatility of S&P 500 returns show a different pattern. It stays above the contrasting hedge funds indices' volatilities. It increases steadily over the next eleven years to peak in August, 2003. From then until July, 2007 the values fall consistently. The implications are that through the diversification the hedge fund managers seem to achieve superior returns and relatively lower volatility.

1.3.2 Value at Risk

To define VaR we need three components: a cumulative distribution function, a fixed time horizon and a given confidence level. Let X be the random variable (the loss) of risky asset returns. By convention, X is a random variable corresponding to loss and negative values of X corresponds to profits. We assume that F_X represents the cumulative distribution function for the loss random variable. Given a confidence level α where $\alpha \in (0,1)$, VaR at the specified horizon time τ is defined as

$$VaR_\alpha(X) = inf\{x|F_X(x) \geq \alpha\}. \quad (1\text{-}3)$$

In words, it means that VaR is the smallest number x such that the probability that the loss X exceeds x is no larger than $1-\alpha$. In probabilistic terms, VaR is simply an upper α −quantile of the loss distribution. For example, the VaR for $\alpha = 0.99$ indicates that there is a one percent probability that losses exceed VaR in the given time interval.

1.3.3 Conditional Value at Risk

To cope with the well-known drawbacks of VaR such as non-subadditivity and non-convexity, CVaR was proposed by Artzner, et al. (1999) and satisfies the four

coherence properties: translation invariance, subadditivity, positive homogeneity, and monotonicity. Now we formally introduce those axioms in more detail based on Equation 1-2. Here X denotes the loss of a position.

Property 1 (translation invariance) states that for every possible loss,

$$\rho(X + l) = \rho(X) + l \tag{1-4}$$

The additional loss l (i.e., a constant) should be taken into account by the same quantity when we update the existing risk measure. Property 2 (subadditivity) is summarized by the following equation.

$$\rho(X_1 + X_2) \leq \rho(X_1) + \rho(X_2) \tag{1-5}$$

This tells us that the risk of the combined portfolio is less than or equal to the sum of the individual risks. However, the idea behind this property is that risk may be reduced by diversification. When this property is violated, for example, a financial organization would have an incentive to have various small-sized subsidiaries to reduce regulatory capital requirements. Property 3 (positive homogeneity) is

$$\rho(hX) = h\rho(X), \text{ for } h > 0 \tag{1-6}$$

This implies that the risk of a financial position is proportional to its size. Note that both subadditivity and positive homogeneity imply that the risk measure is convex. Property 4 (monotonicity) means that positions exposed to higher losses need more capital requirements. For $X < Y$,

$$\rho(X) < \rho(Y) \tag{1-7}$$

where X and Y stand for potential losses.

CVaR is defined as follows:

$$CVaR_{\alpha}(X) = E[X|X \geq VaR_{\alpha}(X)] \tag{1-8}$$

The CVaR risk measure is the expected size of a loss that exceeds the VaR level. For a given loss distribution the CVaR is worse than or equal to VaR at the specified quantile. By the first and fourth properties, the standard deviation is ruled out as a coherent measure.

1.4 Normal Distribution and Risk Measures

If the mean (μ) and the standard deviation (σ) in the normal loss distribution are known, the VaR at a given confidence level, α, (say $\alpha = 0.99$) is simply obtained by

$$\mu + \sigma\Phi^{-1}(\alpha) \tag{1-9}$$

where Φ denotes the standard normal distribution function and $\Phi^{-1}(\alpha)$ is the α −quantile of the standard normal distribution. We assume $\mu = 0$. For CVaR,

$$CVaR = E[X|X > VaR_{\alpha}] = \mu + \sigma E\left(\frac{X-\mu}{\sigma}|\frac{X-\mu}{\sigma} \geq q_{\alpha}\left(\frac{X-\mu}{\sigma}\right)\right)$$

$$= \sigma\frac{\phi(\Phi^{-1}(\alpha))}{1-\alpha}. \tag{1-10}$$

where $q_{\alpha}(X)$ is the α −quantile of the standard normal distribution. The last equality in Equation 1-10 can be obtained by applying integration by parts to $E\left(\frac{X-\mu}{\sigma}|\frac{X-\mu}{\sigma} \geq q_{\alpha}\left(\frac{X-\mu}{\sigma}\right)\right)$.

$$E\left(\frac{X-\mu}{\sigma} \mid \frac{X-\mu}{\sigma} \geq q_\alpha\left(\frac{X-\mu}{\sigma}\right)\right) = \frac{1}{1-\alpha}\int_{\Phi^{-1}(\alpha)}^{\infty} l\phi(l)dl$$

$$= \frac{1}{1-\alpha}\left[-\phi(l)\right]_{\Phi^{-1}(\alpha)}^{\infty} = \frac{\phi\left(\Phi^{-1}(\alpha)\right)}{1-\alpha} \quad (1\text{-}11)$$

where the second equation from the right is based on the property of the normal distribution, $\int \Phi(l)dl = l\Phi(l) - \int l\phi(l)dl$. In practice, we need to calculate the mean of the truncated distribution, $X \geq VaR_\alpha$. Following Maddala (1983, p.365), the mean of the truncated distribution is given by

$$E(X \geq VaR_\alpha) = \frac{\phi(VaR_\alpha)}{1-\Phi(VaR_\alpha)} \quad (1\text{-}12)$$

1.5 Extreme Value Theory and Risk Measures

In the EVT literature, there have been efforts to link real data about extremal events to probability models. There are two parametric approaches: generalized extreme value (GEV) and generalized Pareto (GP). The former is central for analysis of maxima or minima while the latter deals with exceedances over a specified threshold. The GP model is the primary topic of this paper. The GP distribution was introduced by Pickands (1975).

To estimate the parameters of the GP model, the commonly used methods are the method of moments (MOM) and maximum likelihood (ML) techniques. Monte Carlo experiments conducted in Ashkar and Tatsambon (2007) show that the maximum likelihood method is preferable in terms of reducing the inconsistency rate, which is the percentage of times that each method produces an estimate of the GP upper bound that is

inconsistent with the simulated data. As an alternative method, Hosking and Wallis (1987) use the MOM for estimating parameters of the GP distribution. It turns out that both estimators are consistent and efficient.

The GEV model depends on the shape parameter. When the parameter is greater than zero the distribution is a Fréchet distribution. Longin (2000) and McNeil (1997) use estimation techniques based on limit theorems for block maxima. However, there is a tradeoff between the number and size of the blocks. As the size of the blocks increases, we can obtain more accurate estimation results, which mean a low bias in the parameter estimates. When many blocks are involved, the number of observations in each block given the fixed data points shrink. It leads to lower variance of the parameter estimates.

In general there are two different approaches to identify extreme values. Rather than using block maxima method based on a pre-specified period, we model the tails of the loss distribution using the peaks over the threshold (POT) method which focuses on the realizations exceeding an upper tail threshold. Let u be an upper tail threshold. Suppose that $F_u(x)$ is the distribution function of exceedances of X above a certain threshold u. Let $y = X - u$ be the excess, and the conditional excess distribution function is defined by

$$F_u(y) = P(X - u \leq y | X > u), \tag{1-13}$$

where $0 \leq y \leq x_F - u$ and $x_F \leq \infty$ is the right end point. Using the conditional probability property, F_u can be rewritten as

$$F_u(y) = \frac{F(u+y)-F(u)}{1-F(u)} = \frac{F(x)-F(u)}{1-F(u)} \tag{1-14}$$

The purpose of EVT is to model the probability of those exceedances beyond a given threshold.

Let $G_{\xi,\sigma}$ be a Generalized Pareto (GP) distribution with the shape parameter ξ and the scale parameter σ. As u gets large, the distribution of exceedances is approximated by the Generalized Pareto distribution:

$$F_u(y) \approx G_{\xi,\sigma}(y) = \begin{cases} 1 - \left(1 + \frac{\xi}{\sigma} y\right)^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - e^{-\frac{y}{\sigma}}, & \xi = 0 \end{cases} \quad \text{with}$$

$$y \in \begin{cases} [0, x_F], & \xi \geq 0 \\ \left[0, -\frac{\sigma}{\xi}\right], & \xi < 0 \end{cases} \tag{1-15}$$

The shape parameter (ξ) may take real values, whereas the scale (σ) is restricted to a positive value. The shape parameter is a good indicator of the extent of heaviness in the tail. A negative value implies very thin tailed distributions with less probability assigned to extreme outcomes than the normal distribution. A zero value means it has tail weight comparable to a normal distribution. Any positive value represents a heavy tailed character for the tail area. Typically, the shape parameter is greater than or equal to zero for financial return data.

The parameters of the GP model can be estimated in various ways. We rely on the method of moments (MOM) following Hosking and Wallis (1987). Ashkar and Tatsambon (2007) provide simulation results for estimating GP quantiles with a small sample size (e.g.,$T = 10$) and show that there are no significant differences between the various estimation methods. For monthly hedge fund returns, we take exceedances that

are in the top 20 percent of observations, which amounts to 12 data points. Furthermore, the estimated shape parameters over 122 windows have averages of 0.198 and 0.116 for the equal weighted portfolio and the FOF index, respectively. These results are consistent with the recommendations provided by Hosking and Wallis (1987) that moment estimators are preferable for $\xi > 0$ and ξ near 0. In contrast, maximum likelihood estimation can be recommended for very large samples when $\xi > 0.2$.

When $\xi > 0$, it may be shown that the moments of the GP distribution are $E(X^k) = \infty$ for $\xi \leq -1/k$ (Embrechts et al., 1997, p.568). When $\xi > -0.5$, the mean and the variance of the GP distribution is defined as

$$E(X) = \frac{\sigma}{1+\xi}, \quad \text{and } VAR(X) = \frac{\sigma^2}{(1+\xi)^2(1+2\xi)} \tag{1-16}$$

Following Hosking and Wallis (1987), if the moments of the generalized Pareto distribution exist[4], the equations for estimating parameters of the GP distribution using the method of moment are

$$\hat{\xi} = \frac{1}{2}\left(\frac{\bar{x}}{s^2} - 1\right) \tag{1-17}$$

$$\hat{\sigma} = \frac{1}{2}\bar{x}\left(\frac{\bar{x}}{s^2} + 1\right) \tag{1-18}$$

where $\bar{x}$ and s^2 are the sample mean and variance. For the hedge fund applications, those exceedances in every window have values which are larger than -0.5 as estimates of the

[4] The rth moments of X exists when $\xi > -1/r$. The mean exists if $\xi > -1$ and the variance exists if $\xi > -1/2$. The moment equation is defined as $E(1 - \xi X/\sigma)^r = 1/(1 + r\xi)$ if $1 + r\xi > 0$.

shape parameters, so that the first and second moments of the GP distribution exist. In the case where $\xi > -0.25$, the shape and scale parameter estimators are asymptotically normal. When $\xi = 0$, the MOM and MLE methods are both asymptotically efficient.

We compute the parameter estimates over a rolling window of a fixed length. In this case parameters in the model are not constant over time. The method proceeds as follows: First the historical data are split into two parts where one part is reserved for prediction. Then the model is fit under the fixed length of sub-samples and a one-step ahead prediction is made. Within the sample period, the window rolls ahead and produces the estimation results.

As pointed out by Castillo and Hadi (1997) and Dupuis (1996), although the MOM estimates exist, there could be a potential problem if the estimated parameters do not fall inside the feasible range. The range of the exceedances (e.g., x) that are consistent with the MOM estimates is $x > 0$ for $\xi \leq 0$, and $0 < x < \hat{\sigma}/\hat{\xi}$ for $\xi > 0$. Therefore, we must check whether $x_{n:n} < \hat{\sigma}/\hat{\xi}$ for $\xi > 0$, where $x_{n:n}$ is the largest order statistic or observation in a sample size of n. For the equal weighted hedge fund returns, most estimates from the rolling windows satisfy the conditions for consistency of the MOM estimates. In only 10 rolling windows, the conditions were violated in the sense that $x_{n:n} > \hat{\sigma}/\hat{\xi}$ for $\xi > 0$. In contrast, for the FOF returns, all rolling windows satisfy the associated conditions without exception.

Among the 10 rolling windows with exceedances that are not consistent with the MOM estimates for the equal weighted hedge fund returns, eight outcomes occur in untroubled times. In contrast, the two windows on November and December 1997 are in

the middle of the Asian financial crisis, and the ratio of the scale to the shape estimates (i.e., 0.0158) is less than the largest value (i.e., 0.0165) in the associated windows. In this case, the MOM estimates are not meaningful. However, it is important to note that although the MOM estimates are not consistent with the sample values in this case, the largest order statistic exceeds the estimated ratio by some small amount, which may be caused by sampling errors.

Given the estimated GP parameters, $F(x)$ is approximated by

$$F(x) = \big(1 - F(u)\big)F_u(y) + F(u), \tag{1-19}$$

where $x = u + y$ for a sufficiently high threshold u. From a practical standpoint, it is hard to determine an appropriate threshold, and several authors deal with the issue on the selection of the threshold (e.g., Embrechts, Klüppelberg and Mikosch 1997). However, there is not a clear answer for this issue, and a graphical tool may be used as in Gilli and Këllezi (2006). In our work, we used the largest 20 percent of the subsample for every rolling window as the threshold level following Harmantzis, Miao and Chien (2006).

In equation (1-19) above, $F_u(y)$ is replaced by $G_{\hat{\xi},\hat{\sigma}}(y)$ and $F(u)$ can be approximated non-parametically as $(n - N_u)/n$ where n is the total sample size and N_u is the number of exceedances over the threshold u. We get the estimate of $F(x)$ as follows:

$$\hat{F}(x) = 1 - \frac{N_u}{n}\left(1 + \frac{\hat{\xi}}{\hat{\sigma}}(x - u)\right)^{-1/\hat{\xi}} \tag{1-20}$$

where $\hat{\xi}$ and $\hat{\sigma}$ are the MOM estimates of the corresponding parameters.

To calculate estimates of VaR and CVaR, the above equation for $\hat{F}(x)$ may be inverted to get a quantile of the underlying distribution because $VaR_\alpha = F^{-1}(\alpha)$, where F^{-1} is the general inverse function of F. For $\alpha > F(u)$, solving $\hat{F}(x)$ for x yields

$$\widehat{VaR}_\alpha = u + \frac{\hat{\sigma}}{\hat{\xi}}\left(\left(\frac{n}{N_u}(1-\alpha)\right)^{-\hat{\xi}} - 1\right) \tag{1-21}$$

Let us rewrite CVaR as

$$CVaR_\alpha = E[X|X > VaR_\alpha] = VaR_\alpha + E[X - VaR_\alpha|X > VaR_\alpha], \tag{1-22}$$

where the second term is the average loss above the threshold VaR_α. Recall that when the exceedances over a threshold follows a GP distribution, the mean excess function, $e(\cdot)$, is denoted by

$$e(z) = E[X - z|X > z] = \frac{\sigma + \xi z}{1-\xi}, \tag{1-23}$$

where $z \geq u$ and $\xi < 1$. (McNeil, Frey, and Embrechts, 2005) The mean excess function is identical to the second term on the right hand side of Equation (1-22). As a result, for $z = VaR_\alpha - u$ and X representing the excesses y over u we can get the expression for estimated CVaR.

$$\widehat{CVaR}_\alpha = \widehat{VaR}_\alpha + \frac{\hat{\sigma} + \hat{\xi}(\widehat{VaR}_\alpha - u)}{1-\hat{\xi}} = \frac{\widehat{VaR}_\alpha}{1-\hat{\xi}} + \frac{\hat{\sigma} - \hat{\xi}u}{1-\hat{\xi}} \tag{1-24}$$

1.6 Application

The data used in this study are hedge fund indices provided by the Center for International Securities and Derivatives Markets (CISDM) for the monthly returns and Hedge Fund Research Inc. (HFR) for the weekly returns. The CISDM hedge fund indices

are median performance indices of the fund strategies in the database. We select 10 major hedge fund strategy indices as well as the fund-of-fund (FOF) index that reflects the performance of all hedge funds reporting to the database. We construct an equal-weighted hedge fund return that is composed of the 10 strategies. In Table 1-1, the different hedge funds indices used in this study are listed.

Table 1-1 CISDM Hedge Fund Classifications and Definitions

Hedge Fund Style	Definition
Distressed Securities	Invest in companies in financial distresses and bankruptcy
Emerging Markets	Invest in equity or fixed income in emerging markets
Equity Long/Short	Investing on both the long and short sides of the market
Equity Market Neutral	Trade long equity positions and an approximately equal dollar-amount of offsetting short positions in order to achieve an approximately zero net equity market exposure
Event Driven Strategy	Capture price movements caused by a merger, corporate restructuring, reorganization, and bankruptcy
Fixed Income MBS	Attempts to take advantage of mispricing opportunities among different types of mortgage backed fixed income securities while neutralizing exposure to interest rate and/or credit risk
Fixed Income Arbitrage	Benefit from price anomalies between related interest rate securities while neutralizing exposure to interest rate risk
Global Macro	Employs long and short strategies in anywhere a value opportunity exists. Manages use leverage and derivatives to enhance positions
Merger Arbitrage	Capture the price spread between current market prices of securities and their value upon successful completion of a takeover, mergers, spin-off or similar transaction
Sector	Specializing in securities from particular industries or economic sectors
FOF index	Reflects the median performance of all hedge fund of funds managers

Returns are constructed by the instantaneous changes in the index values, $r_t = lnP_t - lnP_{t-1}$, where P_t denotes the value at time t. This is called the continuously compounded return or log return. We include the FOF index because it is less sensitive to the various biases inherent in the individual hedge fund strategies. Fung and Hsieh (2000) argue that

the FOF returns more accurately represent overall returns on hedge funds than a hedge fund index because a collection of diversified hedge funds is more likely to survive than the individual hedge fund. Thus, the FOF may be more appropriate to reflect an investor's losses.

Table 1-2 Descriptive Statistics of Monthly hedge fund returns by strategy class

Variable	Mean	Std.dev	Skewness	Kurtosis*	ADtest**
Panel A: CISDM Hedge Fund Indices: Monthly Returns					
Distressed Securities	0.0101	0.0126	-1.2813	7.6614	1.9997 (0.0000)
Emerging Markets	0.0097	0.0348	-2.7491	19.7382	4.0667 (0.0000)
Equity Long/Short	0.0105	0.0172	-0.1528	2.5962	1.1302 (0.0058)
Equity Market Neutral	0.0067	0.0059	-0.0092	0.7495	0.4991 (0.2098)
Event Driven Strategy	0.0106	0.0133	-1.3146	6.9203	1.2989 (0.0022)
Fixed Income MBS	0.0088	0.0062	2.2472	16.1653	2.6272 (0.0000)
Fixed Income Arbitrage	0.0072	0.0043	-1.5073	8.0541	1.9002 (0.0000)
Global Macro	0.0077	0.0117	0.6037	2.0096	0.9848 (0.0134)
Merger Arbitrage	0.0074	0.0075	-1.0025	3.2003	1.2767 (0.0025)
Sector	0.0128	0.0287	-0.0553	6.3568	1.8887 (0.0000)
FOF index	0.0075	0.0111	-0.1727	2.1814	1.2379 (0.0031)
EW Portfolio	0.0092	0.0114	-1.4285	9.2145	1.2454 (0.0032)
Panel B: HFRX Global Hedge Fund Index: Weekly Returns					
HFR index	0.0512	0.7221	-2.3519	12.9135	11.3847(0.0000)

Note: Data source: CISDM for the monthly returns and HFR for the weekly returns. *Excess kurtosis relative to a normal distribution. **Anderson-Darling test for normality. P-values appear in parentheses by AD test statistics. Panel A in the table presents descriptive statistics of monthly returns of 10 hedge fund indexes, the fund of fund index and an equally weighted portfolio comprising the given 10 indexes from June 1992 to July 2007. Panel B reports descriptive statistics of weekly returns of constituent hedge fund indexes with 405 observations from 4/11/2003 to 12/31/2010.

Table 1-2 presents the descriptive statistics for the monthly and weekly returns of the indices used in the study. The results in the table show that the empirical distributions are a bit negatively skewed except for two styles, Fixed Income MBS and Global Macro, which means that data are left skewed. Each hedge fund index exhibits a fat tail since excess kurtosis is greater than zero. The positive excess Kurtosis (Leptokurtic) has a higher peak and heavier tails than the normal distribution. Ignoring kurtosis tends to

understate the risk of the variables with heavy tails. The AD test for normality evaluates departures from the normal distribution. The AD test statistics for all cases except for Equity Market Neutral yield p-values smaller than 0.05 (i.e., at the 5% significance level) which are significant. Therefore, the AD tests generally reject the null hypothesis of normality.

The preceding results implicitly require the unconditional moments of the data to exist. For our purposes, we assume those processes under consideration are ergodic for the first four moments. For example, a process is said to be ergodic (p.46-47, Hamilton 1994) for the mean if the autocovariance γ_j goes to zero as j increases

$$\sum_{j=0}^{\infty}|\gamma_j| < \infty \tag{1-25}$$

A process is said to be ergodic for the second moments if

$$\frac{1}{T-j}\sum_{t=j+1}^{T}(Y_t-\mu)(Y_{t-j}-\mu) \xrightarrow{p} \gamma_j \text{ for all } j \tag{1-26}$$

where T is a sample of size, Y_t is some random variable, and μ is a constant. As a special case, when a Gaussian process is stationary, the condition in Equation (1-26) ensures ergodicity for all moments. By assuming that the third and fourth moments are ergodic, we can use those summary statistics reported in Table 1-2. If the population moments are infinite or undefined, the sample moments still provide useful information.

A moving window of fixed length (e.g., 60 months or 200 weeks) can be used to estimate the model's stability. For each month or week, the prediction is updated by adding one observation forward and dropping the first observation. So we can keep the

number of observations in each window the same while updating the sample as new data becomes available.

For the monthly returns, we have 182 monthly data points that range from June 1992 to July 2007, which are used to estimate the rolling one-month VaR and CVaR. For the weekly returns, only Net Asset Value (NAV) is available in this case. The NAV data can be used as a proxy to determine the price at which investors enter or exit a hedge fund. The weekly data is collected from Bloomberg, and the HFRX Global hedge fund index[5] is used for the weekly returns. The data ranges from April 11, 2003 to December 31, 2010, which is 405 observations. The HFRX global hedge fund index is designed to be representative of the overall composition of the hedge fund universe. It consists of eligible hedge fund strategies, including convertible arbitrage, distressed securities, equity hedge, equity market neutral, event driven, macro, merger arbitrage, and relative value arbitrage. The fixed length for every moving window is 200 weeks, which is equivalent to about four years.

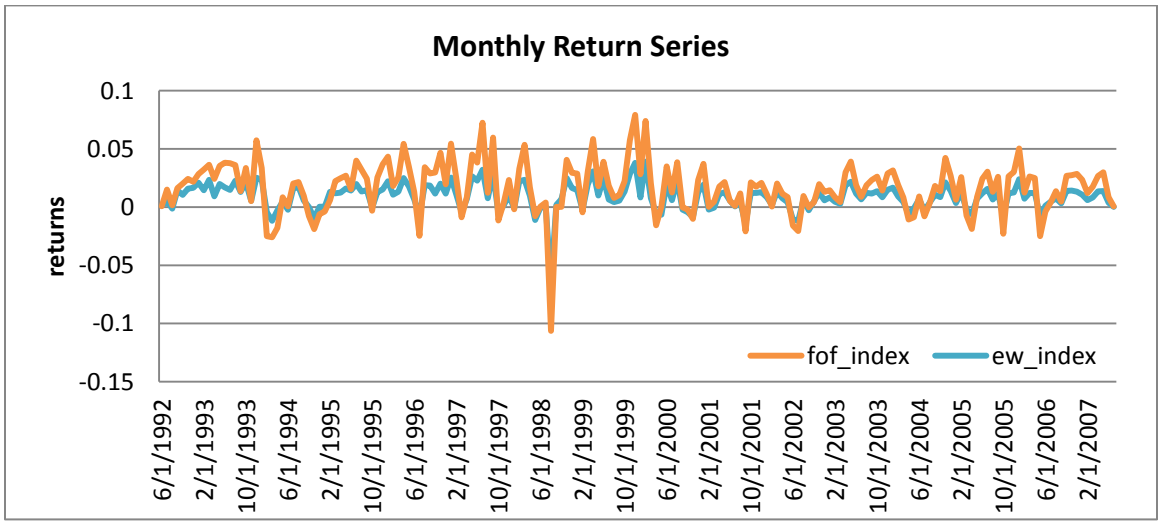
Figure 1-5 plots the observed (n=182) monthly returns of an equally-weighted hedge fund portfolio comprising 10 hedge fund strategies and a FOF index. Both returns move in a similar way and appear to exhibit dependence in the volatility. The changing environment is closely related to the hedge fund performance. The return series are affected by the financial crises in the past. For example, the negative returns around 1994

[5] The index NAV is 1000 at inception.

may be due to the collapse of the Mexican financial markets. Thereafter, there is the 1997 Asian crisis followed by the Russian crisis in August 1998.

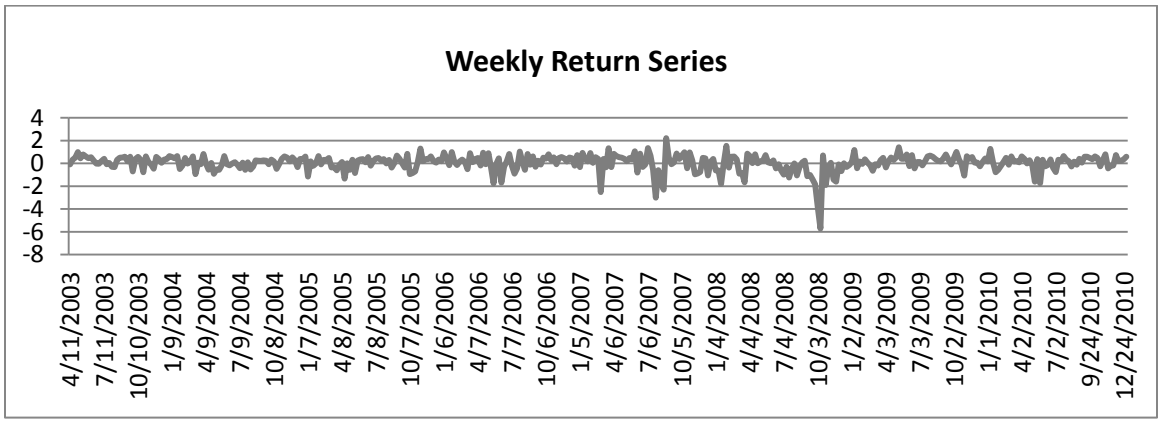
During such economic turmoils, hedge funds may suffer big losses from the unusual market events. One of the most tragic outcomes was the collapse of Long Term Capital Management (LCTM) in 1998, and the sharp drops in returns happen in June 1998. When the US stock market experienced the Dot-Com crash in January 2001, the hedge fund returns plummet, as well.

Figure 1-5 Monthly returns of equally-weighted hedge fund and Fund of Fund index



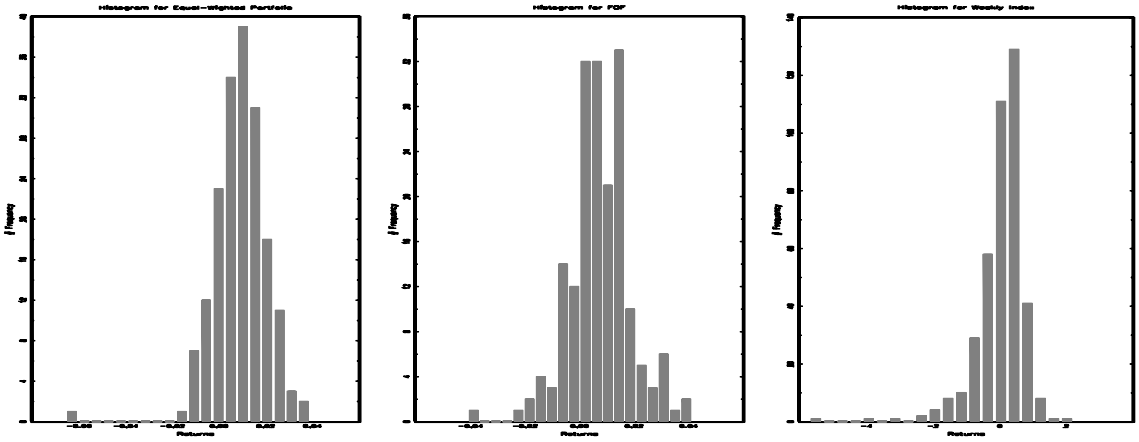
The returns for the weekly data are plotted in Figure 1-6, and several negative returns are realized in 2006 and 2007. Over the period, housing prices start falling and they may affect hedge funds' performance if the hedge funds are involved with subprime mortgages-backed securities. Also, the investment bank Lehman Brothers, which is one of the largest U.S. firms in the industry, declared bankruptcy on September 15, 2008 due to large losses on mortgage-backed securities.

Figure 1-6 Weekly returns of HFRX Global hedge fund index



In Figure 1-7, the histograms of the Equally-Weighted monthly hedge fund portfolio, the FOF index, and the weekly hedge fund returns are presented. They are useful for visualizing how the data are distributed. All three histograms indicate that the data are perhaps not normally distributed because the distributions are peaked in the center and there are a few extreme data points on the left of the distribution, which might suggest fat tails on at least the left-hand side.

Figure 1-7 Histograms of Equally-Weighted, FOF and Weekly Returns



1.6.1 Estimation Results for Monthly Returns

In this section, the time varying estimates of the risk measures are reported. For the monthly returns, 60 observations are selected as the fixed window length. For visual illustration, we take absolute values of the original risk measures and scale them. The results from both the equal- weighted hedge fund portfolio and the FOF index are shown in Figure 1-8 and Figure 1-9. The upper panel in both figures is associated with the equal-weighted hedge fund portfolio, and the lower panel is the FOF index. The estimates based on the normal distribution assumption are in Figure 1-8 and the outcome under the EVT-GP model is presented in Figure 1-9. For every graph in both figures, the CVaR measures are above the VaR estimates, which are expected since the CVaR measure is the average of exceedances above VaR.

Figure 1-8 Normal distribution based time varying risk measures

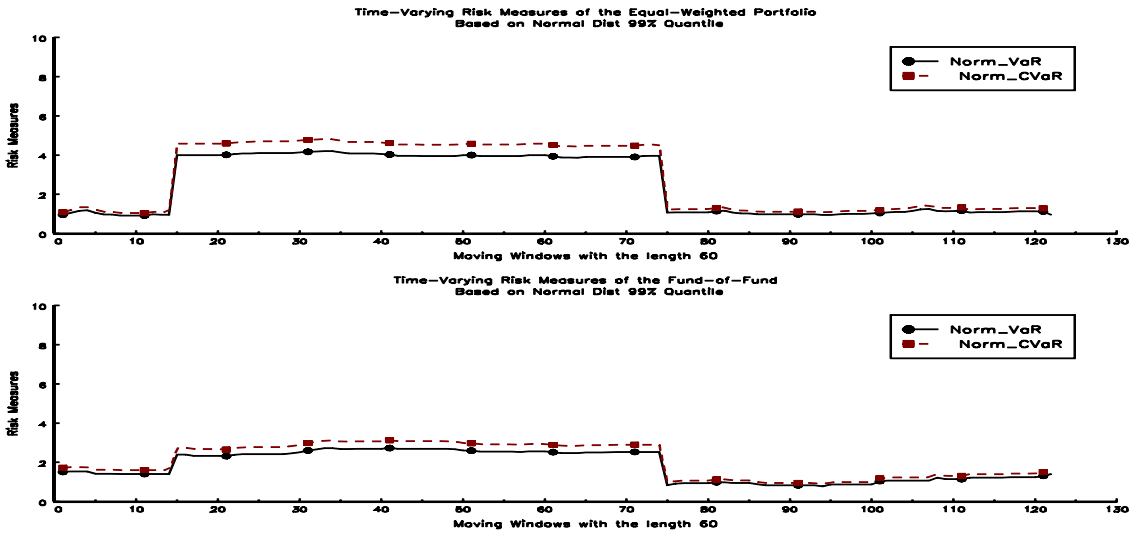
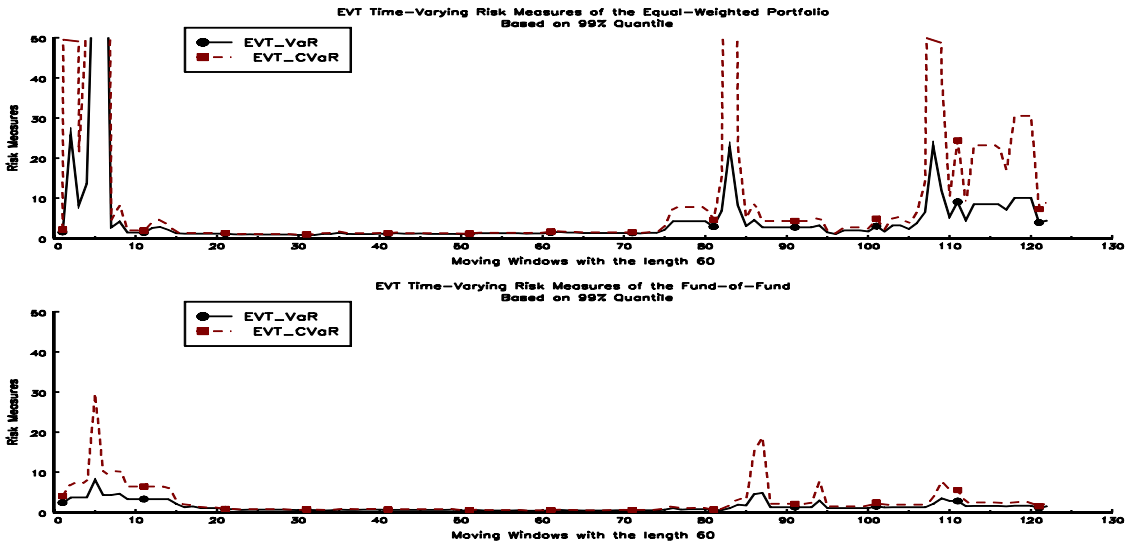


Figure 1-9 EVT-GP based time varying risk measures



Under the normality assumption about exceedances, the estimates of the risk measures demonstrate persistent risk until July 31, 1998, when the estimates suddenly increase and then remain at the same level until July 31, 2003. After that period, the estimates for risk measures return to the original level. The results for both return series under the normality assumption only capture one standout event, the collapse of Long Term Capital Management that occurred in June 1998. In other words, this approach fails to capture several economic distresses that happened at different points in time.

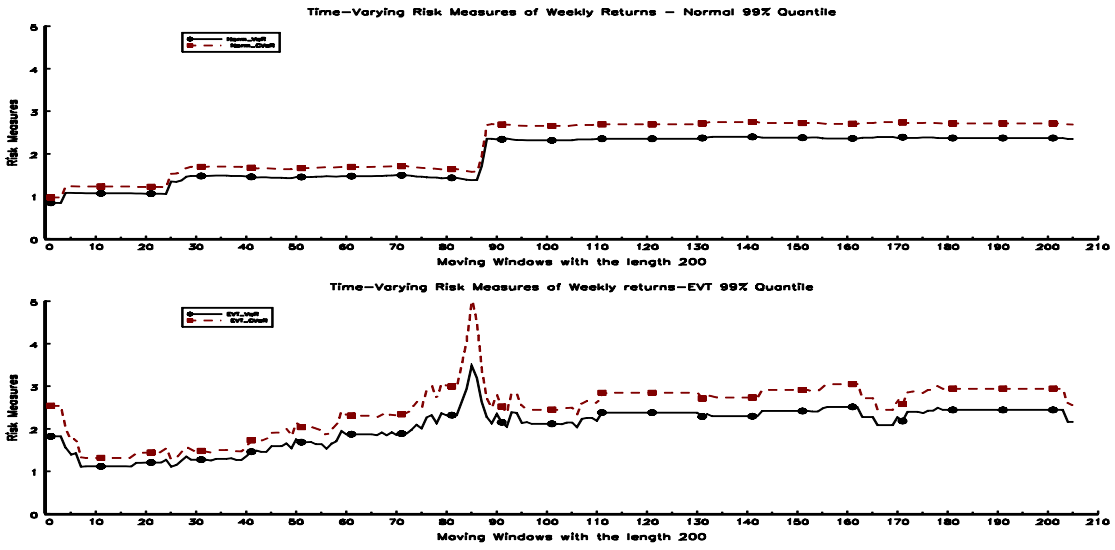
In contrast, the time varying risk measures based on the EVT-GP model capture the bad performance that scatters at several points in time. For the first 15 periods, the estimated risk measures show the Asian crisis in 1997 and the LTCM debacle and Russian default in 1998. Then there are two incidents where the risk estimates are high in June 2004 and May 2006. Whereas it is not possible to detect the sharp increase in risk

under the normality assumption, the EVT based dynamic risk measures are more sensitive to events in the observed sample.

1.6.2 Estimation Results for Weekly Returns

In Figure 1-10, the upper panel shows the estimates of VaR and CVaR based on the normal distribution. The lower panel represents the estimates based on the GP model. Both risk measures move together while the estimate of CVaR is consistently above VaR estimate, as expected. Over time those estimates are steadily increasing in both panels. However, the pattern is a little bit different. For the normal distribution case, the estimated risk measures exhibit three break points in time. For example, the third break around September 2008 reflects the collapse of Lehman Brothers. The GP based results show a slightly different time varying pattern in the estimates of VaR and CVaR. For the first few time periods, the short and quick decreases in the risk measure estimates are followed by a persistently increasing pattern over time that peaks in September 2008.

Figure 1-10 Time varying risk measures with the weekly returns



1.7 Conclusion

Financial risk management relies mainly on the ability to accurately compute the magnitudes and probabilities of large losses due to extreme events, including various economic distresses. In the literature, it is well known that traditional parametric and non-parametric methodologies have limitations to capture such extreme events. In this paper, we have illustrated the methods of extreme value theory by modeling the tail behavior of a loss distribution while considering tail related risk measures such as VaR and CVaR. The results show that the improved EVT-GP risk model more easily identifies the past global financial crises than the normal model. Although VaR is still widely used under the normal assumption, new tools such as the EVT/CVaR may become more important as investment firms look for better ways to gauge and manage the market risk.

References

Ackerman, Carl, Richard McEnally, and David Ravenscraft. 1999. The Performance of Hedge Funds. *Journal of Finance* 54, no. 3 (July): 833-874.

Agarwal, Vikas, and Narayan Y. Naik. 2000. Multi-Period Performance Persistence Analysis of Hedge Funds. *Journal of Financial and Quantitative Analysis* 35, no. 3 (September): 327-342.

Agarwal, Vikas, and Narayan Y. Naik. 2004. Risks and Portfolio Decisions Involving Hedge Funds. *Review of Financial Studies* 17, no. 1 (Spring): 63-98.

Amin, Gaurav S., and Harry M. Kat. 2003. Hedge Fund Performance 1999-2000: Do the Money Machines Really Add Value? *Journal of Financial and Quantitative Analysis* 38, no. 2 (June): 251-274.

Artzner, Philippe, Freddy Delbaen, Jean-Marc Eber, and David Heath. 1999. Coherent Measures of Risk. *Mathematical Finance* 9, no. 3 (July): 203-228.

Ashkar, F., and C. Nwentsa Tatsambon. 2007. Revisiting some estimation methods for the generalized Pareto distribution. *Journal of Hydrology* 346, no. 3-4 (November): 136-143.

Bacmann, Jean-François, and Gregor Gawron. 2005. Fat-Tail Risk in Portfolios of Hedge Funds and Traditional Investments. In *Hedge Funds: Insights in Performance Measurement, Risk Analysis, and Porfolio Allocation*, by Greg N. Fregoriou, Georges Hubner and and Fabrice Rouah Nicolas Papageorgiou. New Jersey: John Wiley & Sons.

Blum, Peter, Michel M. Dacorogna, and Lars Jaeger. 2004. Performance and Risk Measurement Challenges for Hedge Funds: Empirical Considerations. In *The New Generation of Risk Management for Hedge Funds and Provacy Equity Investments*, by Lars Jaeger. New York: Euromoney Trading Ltd.

Boudt, Kris, Brian Peterson, and Christophe Croux. 2008/2009. Estimation and Decomposition of Downside Risk for Portfolios with Non-normal Returns. *Journal of Risk* 11, no. 2 (Winter): 79-103.

Brooks, Chris, and Harry M. Kat. 2002. The Statistical Properties of Hedge Fund Index Returns and Their Implications for Investors. *Journal of Alternative Investments* 5, no. 2 (Fall): 26-44.

Brown, Stephen J., William N. Goetzmann, and Roger G. Ibbotson. 1999. Offshore Hedge Funds: Survival and Performance, 1989-95. *Journal of Business* 72, no. 1 (January): 91-117.

Castillo, Enrique, and Ali S. Hadi. 1997. Fitting the Generalized Pareto Distribution to Data. *Journal of the American Statistical Association* 92, no.440 (Dec): 1609-1620.

Christoffersen, Peter F., and Francis X. Diebold. 2000. How Relevant is Volatility Forecasting for Financial Risk Management. *Review of Economics and Statistics* 82, no. 1 (February): 1-11.

Danielsson, Jon, and Casper G. de Vries. 2000. Value-at-Risk and Extreme Returns. *Annales D'econonie et de Statistique* 60: 239-270.

Danielsson, Jon, and Jean-Pierre Zigrand. 2007. Regulating Hedge Funds. *Financial Stability Review*, no. 10 (April): 29-36.

Danielsson, Jon, Paul Embrechts, Charles Goodhart, Con Keating, Felix Muennich, Olivier Renault, and Hyun Song Shin. 2001. *An Academic Response to Basel II*. LSE Financial Markets Group.

Dupuis, D.J. 1996. Estimating the Probability of Obtaining Nonfeasible Parameter Estimates of the Generalized Pareto Distribution. *Journal of Statistical Computation and Simulation* 54: 197-209.

Embrechts, Paul, Claudia Klüppelberg, and Thomas Mikosch. 1997. *Modelling Extremal Events for Insurance and Finance*. Berlin: Springer Verlag.

Fernandez, Vivian. 2003. Extreme Value Theory and Value at Risk. *Revista de Analisis Economico* 18, no. 1 (June): 57-85.

Fung, William, and David A. Hsieh. 1997. Empirical Characteristics of Dynamic Trading Strategies: The Case of Hedge Funds. *Review of Financial Studies* 10, no. 2: 275-302.

Fung, William, and David A. Hsieh. 2000. Performance Characteristics of Hedge Funds and Commodity Funds: Natural vs. Spurious Biases. *Journal of Quantitative and Financial Analysis* 35, no. 3: 291-307.

Fung, William, and David A. Hsieh. 2001. The Risk in Hedge Fund Strategies: Theory and Evidence from Trend Followers. *Review of Financial Studies* 14, no. 2 (Summer): 313-341.

Geman, Hélyette, and Cécile Kharoubi. 2003. Hedge funds revisited: distributional characteristics, dependence structure and diversification. *Journal of Risk* 5, no. 4 (Summer): 55-73.

Gilli, Manfred, and Evis Këllezi. 2006. An Application of Extreme Value Theory for Measuring Financial Risk. *Computational Economics* 27, no. 2-3 (May): 207-228.

Gupta, Anurag, and Bing Liang. 2005. Do hedge funds have enough capital? A Value-at-Risk approach. *Journal of Financial Economics* 77, no. 1 (July): 219-253.

Hamilton, James D. 1994. *Time Series Analysis*. New Jersey: Princeton University Press.

Harmantzis, Fotios C., Linyan Miao, and Yifan Chien. 2006. Empirical Study of Value-at-Risk and Expected Shortfall Models with Heavy Tails. *Journal of Risk Finance* 7, no. 2: 117-135.

Holton, Glyn A. 2004. Defining Risk. *Financial Analysts Journal* 60, no. 6 (Nov/Dec): 19-25.

Hosking, J.R.M, and J.R. Wallis. 1987. Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics* 29, no. 3 (August): 339-349.

J.P. Morgan and Reuters. 1996. *RiskMetrics-Technical Document*. New York: Morgan Guaranty Trust Company.

Jorion, Philippe. 2000. Risk Management Lessons from Long-Term Capital Management. *European Financial Management* 6, no. 3: 277-300.

Ju, Xionwei, and Neil D. Pearson. 1999. Using Value-at-Risk to Control Risk Taking: How Wrong can you be? *Journal of Risk* 1, no. 2 (Winter): 5-36.

Lhabitant, François-Serge. 2003. Hedge Funds: A Look beyond the Sample. In *Hedge Funds: Strategies, Risk Assessment, and Returns*, by Greg N. Gregoriou and and Vassilios N. Karavas Fabrice Rouah, 209-234. Washington D.C.: Beard Books.

Liang, Bing. 2000. Hedge Funds: The Living and the Dead. *Journal of Financial and Quantitative Analysis* 35, no. 3 (September): 309-326.

Lo, Andrew W. 2009. Regulatory Reform in the Wake of the Financial Crisis of 2007-2008. *Journal of Financial Economic Policy* 1, no.1: 4-43.

Lo, Andrew W. 2008. *Hedge Funds: An Analytic Perspective*. Princeton: Princeton University Press.

Longin, François M. 2000. From value at risk to stress testing: The extreme value approach. *Journal of Banking & Finance* 24, no. 7 (July): 1097-1130.

Longin, François M. 1996. The Asymptotic Distribution of Extreme Stock Market Returns. *Journal of Business* 69, no. 3 (July): 383-408.

Maddala, G.S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

McNeil, Alexander J. 1997. Estimating the tails of loss severity distributions using extreme value theroy. *ASTIN Bulletin* 27, no. 1 (May): 117-138.

McNeil, Alexander J., and Rudiger Frey. 2000. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance* 7, no. 3-4 (November): 271-300.

McNeil, Alexander J., Rudiger Frey, and Paul Embrechts. 2005. *Quantitative Risk Management: Concepts, Techniques and Tools.* New Jersey: Princeton University Press.

Pickands, James III. 1975. Statistical Inference Using Extreme Value Order Statistics. *Annals of Statistics* 3, no. 1 (January): 119-131.

Rachev, Svetlozar T., Stoyan V. Stoyanov, and Frank J. Fabozzi. 2008. Risk and Uncertainty. In *Advanced Stochastic Models, Risk Assessment, and Portfolio Optimization*, 171-205. New Jersey: John Wiley & Sons.

Reiss, R.-D., and M. Thomas. 1997. *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields*. Basel: Birkhäuser Verlag.

Renshaw, Edward F. 1977. Short Selling and Financial Arbitrage. *Financial Analysts Journal* 33, no. 1 (Jan-Feb): 58-65.

Schneeweis, Thomas, Richard Sputgin, and David McCarthy. 1996. Survivor Bias in Commodity Trading Advisor Performance. *Journal of Futures Markets* 16, no. 7 (October): 757-772.

Taleb, Nassim. 1997. The Jorion-Taleb Debate. *Derivatives Strategy* 2, no.4 (April).

U.S. Securities and Exchange Commission. 2003. Implications of the Growth of the Hedge Funds. Washington, D.C..

2 Optimal Hedging under Copula Models with High Frequency Data

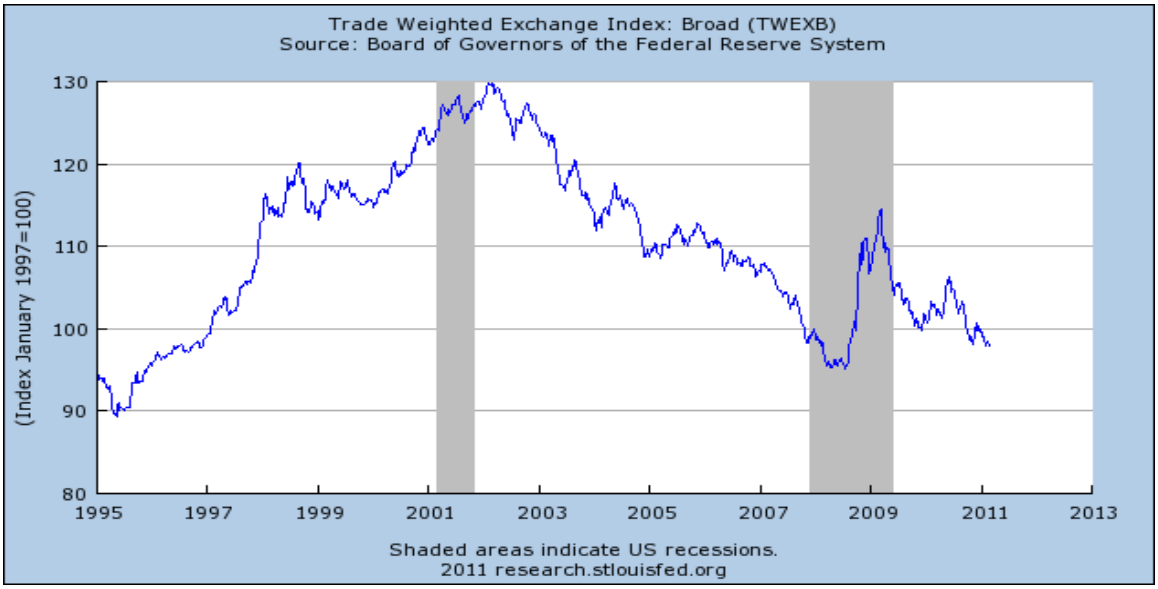
2.1 Introduction

The purpose of this chapter is to manage financial risks with the improved risk measures based on models of high frequency (e.g., intra-day) foreign exchange rates. Furthermore, we use copula-based probability models to capture the dependence structure and the univariate marginal properties of high frequency spot and futures returns. Alternative econometric models containing GARCH-based copula model are employed to determine the optimal hedge ratios for currency futures positions used to hedge against the variability of spot exchange rates. The models under consideration are compared by the extent to which the variance of the hedge portfolio return is reduced.

An exchange rate is the price of one currency in terms of another currency. As one currency depreciates, it drops in value and the other currency appreciates (i.e., rises in value). Floating exchange rates between currencies can fluctuate frequently and often substantially. From U.S.-based exporters' or investors' perspectives, foreign exchange risk is the exposure to potential losses due to appreciation of the U.S. dollar against the foreign currency. Foreign exchange fluctuations may also affect multinational firms' earnings and investors' profits. Unfavorable foreign exchange rates (i.e., a strong dollar) can affect foreign sales and investment in a negative way. As foreign currencies depreciate, U.S. exports are likely falling as U.S. firms raise prices to compensate for the

currency's decline in value. In contrast, a weaker dollar (i.e., dollar depreciation) has a positive impact on export sales and investment from overseas. The primary source of concern lies in the unpredictability of the foreign exchange rates.

Figure 2-1 Trade Weighted Exchange Index: Broad



Data Source: FRED, Federal Reserve Economic Data, Federal Reserve Bank of St. Louis: Trade Weighted Exchange Index: Broad [TWEXB]; Note: Observations range from 1995-01-04 to 2011-02-23. This figure shows a weighted average of the foreign exchange value of the U.S. dollar against the currencies of a broad group of major U.S. trading partners, which includes the Euro Area, Canada, Japan, Mexico, China, United Kingdom, Taiwan, Korea, Singapore, Hong Kong, Malaysia, Brazil, Switzerland, Thailand, Philippines, Australia, Indonesia, India, Israel, Saudi Arabia, Russia, Sweden, Argentina, Venezuela, Chile and Colombia.

Figure 2-1 shows the U.S. dollar values against a trade weighted basket of other currencies, which include most of the major trader partners over the last 16 years. From March, 1995 until the end of 2001, the dollar values have kept steadily increasing and reached a peak in early 2002. Then the dollar followed a consistently falling trend until mid-2008. When the financial collapse occurred, the period between mid-2008 and the

start of 2009 showed a rapid rise, followed by a steep fall until early 2010. Then it seemed that the U.S. dollar values fluctuated in the face of a decreasing trend.

Those agents who are interested in reducing risk must bear a lower expected return on the underlying asset due to the trade-off between risk and return (i.e., negative of loss). Due to unpredictable changes in the exchange rates, hedging foreign exchange risk is of great importance because it will influence financial performance. In a simple framework we can demonstrate the relationship between a change in the exchange rate and a change in the value of an asset.

$$\Delta V = a + b\Delta S + u \tag{2-1}$$

where ΔV is the change in the value of an asset, ΔS is unexpected change in the exchange rate, α is intercept, and u is the unexplained errors that are not captured by the variable in Equation 2-1. Assuming that a and u are on average zeros,

$$b = \frac{\Delta V}{\Delta S} \tag{2-2}$$

Here b can be regarded as foreign exchange exposure, which is the sensitivity of the domestic currency value of an asset to an unexpected change in the exchange rate.

U.S. exporters or investors can use currency futures or swap contracts to hedge against foreign exchange risk. If an investor anticipates receiving a cash flow denominated in a foreign currency, the dollar value of the foreign payment depends on the exchange rate at the time payment is made. For example, suppose a U.S. based investor will receive €125,000 on September 1. The current exchange rate implied by the futures contract is $1.4/€. The investor can lock in the exchange rate by selling €125,000

worth of futures contracts expiring in September. By doing so, the investor locks in an effective exchange rate at \$1.4/€ regardless of changes in the actual exchange rate prior to the payment date.

It is very important to see how the futures prices are determined in order to better understanding the futures hedge. The following discussion is based on a U.S. investor's point of view. Since we are dealing with futures contracts on currencies, the underlying asset is a certain number of units of the foreign currency (i.e., the payment). Let S_0 be the current spot exchange rate in dollars per foreign currency and let F_0 be the futures exchange rate in dollars per foreign currency. T is the time to maturity. We define r as the U.S. dollar risk-free interest rate and r_f as the value of the foreign risk-free interest rate. Using this notation, we can derive the price of futures contracts on currencies by equalizing two dollar denominated quantities at maturity. Suppose we have 1,000 units of foreign currency at time zero. We can invest it for T periods at the foreign risk-free interest rate, r_f, and can take a short position expecting to sell it at time T. The resulting quantity in dollar terms is $1{,}000e^{r_f T}F_0$ under interest rates with continuous compounding. The other quantity in dollars is $1{,}000S_0e^{rT}$ in a scenario where we first exchange 1,000 units of foreign currency for dollars and can invest it for T periods at interest rate r. If both values are equal at time T we end up with the following equation:

$$F_0 = S_0e^{(r-r_f)T} \tag{2-3}$$

Equation 2-3 is called interest rate parity. If the parity relationship does not hold, there is room for arbitrageurs to profit. For example, if $F_0 > S_0e^{(r-r_f)T}$, arbitrageurs can

profit by buying the underlying spot exchange rates and shorting futures contracts. As arbitrageurs sell the futures contracts, the trading drives its price down. If $F_0 < S_0 e^{(r-r_f)T}$, they can profit by selling the underlying spot exchange rates and going long in futures contracts. The long arbitrage trade in the futures market drives the futures price up. So, the spot and futures prices should satisfy the relationship during the delivery period. However, the spot price may not converge to the futures price in the absence of arbitrage.

Table 2.1 shows currency futures quotes on Friday March 4, 2011. We can observe that futures prices for the Japanese Yen and Swiss franc increase with maturity in Table 2.1. Equation 2-3 refers to spot/futures relationship over time, but the data in the table are only futures rates. March futures are close to expiry, but these are not equal to spot rates.

Table 2.1 Foreign Exchange Futures Quotes

Currency Futures

Japanese Yen (CME)-¥12,500,000; $ per 100¥						
March	1.2220	1.2237	1.2118	**1.2140**	-.0074	126,061
June	1.2232	1.2244	1.2127	**1.2149**	-.0073	8,850
Canadian Dollar (CME)-CAD 100,000; $ per CAD						
March	1.0276	1.0292	1.0251	**1.0287**	.0002	137,016
June	1.0256	1.0269	1.0230	**1.0266**	.0002	10,621
British Pound (CME)-£62,500; $ per £						
March	1.6325	1.6333	1.6251	**1.6273**	-.0055	124,708
June	1.6313	1.6314	1.6235	**1.6255**	-.0055	3,078
Swiss Franc (CME)-CHF 125,000; $ per CHF						
March	1.0827	1.0832	1.0720	**1.0731**	-.0096	59,531
June	1.0830	1.0839	1.0728	**1.0739**	-.0097	1,445
Australian Dollar (CME)-AUD 100,000; $ per AUD						
March	1.0155	1.0179	1.0115	**1.0140**	-.0012	133,306
June	1.0038	1.0063	1.0001	**1.0026**	-.0011	3,425
Mexican Peso (CME)-MXN 500,000; $ per 10MXN						
March	.82675	.83350	.82600	**.83325**	.00650	129,571
June	.82075	.82750	.82050	**.82725**	.00650	4,831
Euro (CME)-€125,000; $ per €						
March	1.3863	1.3975	1.3831	**1.3956**	.0096	213,545
June	1.3848	1.3956	1.3815	**1.3936**	.0093	11,454

Source: Thomson Reuters. This is Currency Futures quotes from the Wall Street Journal on Friday, March 4, 2011. Columns indicate month, open, high, low, settle, change, and open interest.

As Hull (2006, p.115) points out, a foreign currency can be thought of as an asset. Since the value of interest paid in a foreign currency depends on the value of the foreign currency, the positive value of the interest rate of the foreign currency generates the yield of a foreign currency.

Hedging with futures requires a futures contract that is highly correlated with the underlying asset being hedged. Since the two positions tend to move in the same directions, higher correlation between the two positions will more effectively offset each other and the hedge will perform better. For instance, if you want to hedge some UK pounds you expect to receive, you would short pound futures in the same quantity. Also, you need to choose a futures contract that has a delivery or expiry date close to or shortly after the spot pound transaction.

The ratio of the number of units in a hedging instrument to the number of units being hedged is called the hedge ratio. Under some risk measure that represents the volatility of portfolio returns, the optimal hedge ratio is easily calculated by minimizing the risk measure for the associated portfolio.

The study of the hedging process requires two practical considerations. First, we must decide which model is appropriate for construction of the hedge portfolios. Second, we must determine whether we are interested in a static hedging strategy or a dynamic method. With high-frequency foreign exchange rates and corresponding futures contracts, we attempt to see how to improve the hedging effectiveness in the various hedging models in this paper.

The normality assumption for the asset returns plays a major role in calculating the optimal hedge ratio. However, the estimated distributions of financial asset returns appear to be non-Gaussian as first noted by Mandelbrot (1963). He stressed that the normal distribution cannot reflect the movement of asset returns observing heavy tailed characteristics. To specify a model for the empirical distributions of asset returns, the following two facts are important. First, the empirical distributions tend to have high peaked and heavy tailed properties relative to the normal distribution. Those things are also observed in high-frequency data. Second, asset returns are time dependent in a nonlinear fashion. In the linear sense, there is no significant autocorrelation of the price changes, which is often used to support the efficient market hypothesis (Fama 1963) that implies the asset returns are impossible to predict from available information.

In cases where the marginal distributions of asset returns are not normally distributed, it may be hard to define a joint distribution of spot and futures returns for hedging purposes. In particular, when the associated returns are correlated in nonlinear form, it is not quite possible to derive a multivariate distribution. Faced with such difficulties, the copula method is a useful way to obtain a candidate joint distribution. The copula of a distribution is defined as the joint distribution function of fixed marginal distributions. This is supported by the Sklar theorem that proves all multivariate distribution functions have copulas and copulas can be derived from multivariate distributions with continuous marginals (Nelsen 1999). One advantage of the copula method is that a copula function can capture dependence structures without regard for the

form of the marginal distributions. Also, other correlation measures than linear correlation can be incorporated within the copula approach.

One common way to model volatility[6] of the underlying asset returns are the autoregressive conditional heteroskedastic (ARCH) models proposed by Engle (1982). The ARCH model is extended by the generalized ARCH (GARCH) model of Bollerslev (1986). These studies develop methods to deal with the univariate volatility of the asset returns. We need to extend our discussion to the multivariate case in order to capture the dynamic relationships between volatility processes of two or more asset returns. The important factor in the multivariate distribution is the dependence structure. The dependence modeling may have significant consequences in various financial applications. The constant conditional correlation (CCC) GARCH model is proposed by Bollerslev (1990) to capture the dynamic behavior of correlations. In turn models of the time varying or dynamic conditional correlations (DCC) are proposed by Engle and Shephard (2001) and Tse and Tsui (2002).

The multivariate GARCH model has several disadvantages. First, as more asset returns are involved, the number of parameters to be estimated increases. Second, it assumes that the conditional distribution of returns is Gaussian. If that is not the case, the extension to the multivariate case may be difficult or infeasible. However, copula functions can alleviate those difficulties. By combining copula functions with GARCH models, we can not only keep distributional limitations flexible but also take into account

[6] It means the conditional standard deviation of the return.

the time varying feature in the parameters of the dependence structure of the relevant marginal distributions.

In the present paper, we are interested in evaluating various competing dynamic hedging methods containing the constant conditional correlation (CCC) GARCH model, the dynamic conditional correlation (DCC) GARCH model, and the copula-based GARCH model. The GARCH process represents the dynamics of asset returns by updating past information. A variety of copulas will be employed to link the spot and futures returns. With the copula-GARCH framework, the innovations in the mean equations are described by Student-t distributions, which are the marginal distributions. Five different copula (i.e., Gaussian Student-t, Frank, Clayton, and Gumbel) functions are used to allow for the dependence structure between a pair of the spot and futures returns. Furthermore, the time varying process for the dependence structure would be modeled so that the parameters to be estimated evolve over time.

We apply such models to high frequency data on foreign exchange spot and futures returns. Earlier studies of hedging models rely heavily on less frequent data such as daily, weekly or monthly, but recent studies tend to use intra-day (e.g., a minute by minute) data for high frequency models. Among the different asset types, we examine the statistical properties of the foreign exchange rates and the corresponding futures contracts. We also evaluate the hedging effectiveness of the candidate models.

The paper is organized as follows: The next section provides a detailed literature review. The subsequent sections briefly discuss copula models and then hedging models employed in this study of high frequency foreign exchange spot and futures data. The last

section describes the data and presents empirical results from various methods. Finally, concluding remarks are presented.

2.2 Literature Review

Hedging is an activity using financial instruments such as derivative contracts (e.g., futures[7], forwards and swaps) to reduce the risk from the volatility of price changes. Whereas a forward contract agrees to buy or sell the underlying asset on a specified future date for a predetermined price, a spot contract is an agreement to buy or sell an asset today for immediate delivery. As a means of hedging owned assets against potential losses, one can use a short futures hedge. In this case, the key to this situation is how many futures contracts an investor sells to provide adequate risk protection. The optimal hedge ratio is the number of futures contracts relative to the size of the cash position.

The fundamental benchmark for the optimal hedge ratio is calculated based on the OLS method (Cecchetti, Cumby, and Figlewski 1988; Baillie and Myers 1991) and is the conditional covariance between the spot and futures returns divided by the conditional variance of the futures returns. This method is equivalent to one that minimizes the variance of the asset portfolio held by the investor, which is known as the minimum variance hedge ratio. Despite many drawbacks such as serial correlation (Herbst, Kare, and Caples 1989) and heteroskedasticity (Park and Bera 1987), the OLS-based method is commonly used in the literature. The former article tries to fix the autocorrelation

[7] A futures contract is an agreement to buy or sell an underlying asset (e.g., commodities, currencies, and securities) at predetermined prices on a specified date in the future.

disturbance problem. Park and Bera (1987) documents that an ARCH process captures the nonlinearities between the cash and futures prices better than the OLS-based method.

The optimal hedge ratio is based on two fundamental assumptions, normality of returns and a quadratic utility function over preferences. In the case of non-normality of returns, the skewness and kurtosis measure should be taken into account to measure hedge ratios. In doing so, Alexander and Baptista (2004) develop two alternative hedging methods, minimum-VaR and minimum-CVaR. Harris and Shen (2006) apply those methods to currency portfolios.

Regarding empirical estimation of the hedge ratios, the past literature has accomplished impressive developments in a variety of ways. First, the focus moves from a constant hedge ratio to a dynamic time-varying estimate of the hedge ratio (Myers 1991). Second, it is known that cointegration between the spot and futures prices affects the properties of OLS. Ghosh (1995) and Kroner and Sultan (1993) estimate the optimal hedge ratio with foreign currency futures on the daily frequency basis. Both studies state that the misspecification of the traditional method leads to underestimating the optimal hedge ratio. Their analyses are based on an error correction representation because the spot and futures prices[8] are cointegrated. Third, there are various types of generalized autoregressive conditional heteroscedasticity (GARCH) processes developed to capture

[8] It takes the form of the natural logarithms of prices.

and explain the conditional volatility of the returns[9] series in Brooks, Henry, and Persand (2002). The GARCH models proposed by Engle (1982) and Bollerslev (1986), are particularly useful to estimate the relevant parameters of the time dependent conditional variance models.

For the purpose of estimating the optimal hedge ratio, there are multiple ARCH and GARCH models that may be used. Cecchetti, Cumby, and Figlewski (1988) obtain an optimal futures hedge ratio using an ARCH model. Baillie and Myers (1991) show that the dynamic strategy based on the GARCH model performed better when describing both returns as a martingale process. Kroner and Sultan (1993) draw similar conclusion with currency spot and futures returns. Floros and Vougas (2004) document that the bivariate GARCH model provides better hedging performance relative to the alternative methods using the Greek spot and futures market data. With the Datastream benchmark BTP 10-year index and two types of futures prices (i.e., German Bund Futures and the Eurolira Futures price series), Rossi and Zucca (2002) confirm that the multivariate GARCH model-based hedging strategy can improve over the other traditional hedging strategies. Park and Switzer (1995) present comparable results based on two pairs of datasets the S&P500 spot and index futures and the Toronto 35 spot and futures.

Interestingly, some studies (e.g., Bystrom 2003 and Holms 1996) show that the constant OLS hedge ratio outperforms the GARCH hedging strategies in terms of

---

[9] The continuously compounded return can be computed by taking the first difference of the natural logarithms of prices.

variance reduction. For the former study, the Nordic electricity spot and futures contracts are used. The constant correlation bivariate GARCH and the multivariate orthogonal GARCH are employed for out-of-sample evaluation. Holms (1996) relies on a univariate GARCH(1,1) framework with the FTSE-100 stock index futures as a hedging instrument. He finds that the OLS hedge ratios vary over time and it is preferred to GARCH hedging strategy. Lien, Tse, and Tsui (2002) reach the same conclusion that the constant hedge ratio performs better than the constant conditional correlation GARCH model based strategy. The study mainly focuses on the out-of-sample forecasts based in ten different futures contracts. We need to be cautious when interpreting these counter-intuitive results. Under the situation where the model misspecification is uncertain, the superiority of the OLS- based hedging strategy may mean that OLS provides the best fitting misspecified model.

Prior to Engle (2002), most studies used the constant conditional correlation (CCC) GARCH (Bollerslev 1990), with a multivariate heteroskedastic model to capture the relationship between several exchange rates. Engle (2002) proposes the time-varying (or dynamic) conditional correlation (DCC) model with a flexible correlation structure. As he points out, the motivation is based on the optimal hedge ratio formula, which includes the estimate of the correlation between the hedge portfolio returns. Engle and Sheppard (2001) and Tse and Tsui (2002) develop time-varying models based on the GARCH framework. The use of the DCC GARCH model helps examine the short-term dynamics of the return series that may be related to a long-term pattern.

Along with the dynamic conditional correlation (DCC) GARCH, we need to place emphasis on the characteristics of the joint distribution of the hedge portfolio returns. If non-normal properties such as skewness, kurtosis, and dependence structure (e.g., between the spot and futures returns) are present, we can incorporate a copula model into the DCC GARCH process. Hsu, Tseng, and Wang (2008) shows that the copula-based GARCH model improves the performance in terms of the reduction of variance relative to other comparing models.

The copula approach is popular in finance mainly because it can describe the excess volatility and idiosyncratic behavior of financial asset values or returns. The properties of copula functions are well summarized by Nelsen (1999). Copulas can be interpreted as joint distributions represented from given or fixed marginal distributions. The copula approach conceptually separates marginal behaviors from the dependence structure. A crucial property is that the copula-based dependence measure is invariant to any increasing transformation of the original data. Therefore, the copula method may be used to model nonlinearity among multiple time series. One advantage of using copula model is to separate marginal distributions from the type of the copula model. Since Embrechts, McNeil, and Straumann (2002) and Cherubini, Luciano, and Vecchiato (2004), many applications in finance and economics are produced.

Using this method, the problem of estimating the parameters of parametric copula functions can be accomplished with several standard procedures. First, the parameters of the copula model may be estimated with the generalized method of moments (GMM). Prokhorov and Schmidt (2009) show that when the copula model is misspecified but

robust, the GMM estimator outperforms the pseudo ML estimator. Second, we can use the two-step ML method. In the first step of the inference functions for margins (IFM) approach, we choose a parametric marginal model and fit it to data. Then, we maximize the full likelihood function over the dependence parameter. This method is used by Joe (1997). Under a conditional likelihood model, the two-step method is used by Jondeau and Rockinger (2006) and Patton (2006). Third, we can use a non-parametric approach, where the marginal distribution must be derived with an empirical or kernel density estimator (Genest and Rivest 1993).

Hull and White (1998) use the normal copula to calculate the Value-at-Risk measure. Chen at al. (2004) report that normality of U.S. equity returns is not rejected under a GARCH model, and a normal copula model is not rejected for exchange rate data. Breymann et al. (2003) show the superiority of the Student's t-copula when modelling the dependence structure of high-frequency data.

The copula-GARCH approach has been employed by many researchers to evaluate hedging (Bertram, Taylor and Wang 2007, Jondeau and Rockinger 2006, Hsu, Tseng, and Wang 2008). The basic strategy in these studies is to use the GARCH framework to model the dynamics of financial returns. However, under the normality assumption for the GARCH model, its application may be restrictive in situations where the actual financial data are not likely to be normally distributed. The use of the joint distribution through copula functions may be better because the copula method can conveniently represent the dependence structure among the random variables.

Among the statistical characteristics of high frequency data, negative autocorrelation is observed in Goodhart (1989) and Goodhart and Giugale (1993). However, a different pricing algorithm yields positive first-order autocorrelation (Bollerslev and Domowitz 1993, p.1432). There are a few explanations in the literature for the causes of negative autocorrelation in minute by minute data. For example, traders' heterogeneous responses to prices may cause it because small and large firms may have different attitudes towards risk, information, and operating hours involved with more geographically dispersed markets (Bollerslev and Domowitz 1993). Goodhart and Giugale (1993) assert that the rising liquidity associated with one more market-maker improves the allocation of inventories, which may lead to negative autocorrelation. A second explanation from Bollerslev and Domowitz (1993) is that negative autocorrelation may be caused by the non-synchronous construction of the price series at the end of the interval.

As shown in Bollerslev and Domowitz (1991), it is possible that dependence in the bid-ask spread leads to serial correlation in the conditional variance of returns in high frequency data. Bollerslev and Domowitz (1993) report that such a relationship may be found through the GARCH error structure of the model. They attribute the source of the conditional variance to the number of incoming quotes, unlike the traditional proxies for the information arrival, volume or the number of transactions. Using intraday data on currency futures contracts, Laux and Ng (1993) find that when they use the expected number of price changes as a proxy for information arrival process, it has explanatory power for conditional volatility.

Intraday seasonality[10] can be related to the fat-tailedness of the returns. Ballocchi et al., (2001) provide evidence of seasonal volatility as a function of the time left to expiry for Eurofutures contracts. For the foreign exchange market, Müller et al. (1990) reveal that there is seasonal volatility in price changes during business hours. Since the existence of seasonal volatility can induce volatility model estimators to be biased, we need to filter out the periodic component of the volatility.

From the volatility estimation perspective, when the seasonal patterns of volatility processes are accounted for, it turns out that the GARCH model may not capture the heterogeneous structure of the foreign exchange market. According to Guillaume et al. (1995), the GARCH model fails to explain both geographical and temporal heterogeneity of returns. In other words, there are other unexplained components which generate long-term and consistent patterns in the time series. As Andersen and Bollerslev (1997) point out, one can get misleading inference about the intraday volatility patterns as a result of strong periodic structure of the high frequency data.

For modeling intraday seasonality, earlier studies propose either a polynomial activity function based on geographical characteristics (Dacorogna et al., 1993) or a method with a Fourier form as a nonlinear regression model (Andersen and Bollerslev 1997). Gencay et al., (2001) propose a simple method for eliminating intraday seasonality using a wavelet multi-scaling approach. Depending on the wavelet transform, it partitions the time frequency data into low and high-frequency components. The use of the filtering

[10] A time series may be described as seasonal if it exhibits a periodic pattern.

method is more beneficial in that the persistence of volatility in further lags is much smaller relative to other studies.

A key empirical question is what type of distribution is proper to describe the data. Some researchers have found that the empirical distribution of intraday foreign exchange rates is symmetric and fat-tailed (Guillaume et al., 1997). In contrast, Wasserfallen and Zimmermann (1985) reject the symmetry hypothesis, even though as the time interval extends, the size of skewness decreases. More interestingly, as time interval increases, the measure of the excess kurtosis decreases.

2.3 Wavelet Transform[11]

In the last two decades, the wavelet methodology has drawn considerable attention in the time series literature. The wavelet and other transformations are used to get additional information that is not readily available in the original series. The most popular transform to date is the Fourier transform, which is a function of time. We can analyze a signal or a series in the time domain for its frequency content. Sines and cosines are used as the basis functions, which are elements of Fourier synthesis. Namely, a given time series can be represented as a Fourier series with sine and cosine terms. In general, the frequency is related to the change in rate of a process. If it changes rapidly, it is said to be of high frequency, whereas if it changes slowly, it is of low frequency. In case of no change in the variable, we say it is zero frequency. The frequency spectrum generated by

[11] The contents in this section are based on the discussions in Chapters 1, 4 and 5 of Percival and Walden (2000) and in Chapter 4 of Gençay et al (2002).

the Fourier transform is useful for identifying the frequencies in a time series. It is well known that there is a trade-off between the time and frequency domain. This feature is revealed at the extreme cases. Given a sample in the time domain, when we expand the time series to be infinitely long by padding it with zeros (i.e., making the time period longer) the frequency domain samples are so close that they can produce a narrow spike in the frequency domain. Similarly, if the period of a sample in the time domain gets so short that it becomes an impulse, the frequency spectrum would be expressed as a flat line in the frequency domain. The Fourier analysis is attractive when the data in question are a stationary time series. The Fourier transform can be applied to spectral analysis, which is one of most widely used methods in data analysis.

In the presence of non-stationary features of time series such as drifts, trends, and sudden changes, Fourier analysis of the frequency domain may not be suitable and is replaced by the wavelet approach. It works in both the time and frequency domains whereas the Fourier transform only generates a frequency representation of a series. In a Fourier transform, there is no way to tell us when in time the frequency portions exist. The wavelet transform provides us with good time resolution and poor frequency resolution at high frequencies while at low frequencies it gives good frequency resolution and poor time resolution.

Formally, a wavelet is defined as a real valued function, $\psi(\cdot)$, over the real line, $(-\infty, \infty)$, which satisfies two basic properties: the integral of the real-valued function is zero and the square of it integrates to unity.

$$\int_{-\infty}^{\infty} \psi(u)du = 0 \qquad (2\text{-}4)$$

and

$$\int_{-\infty}^{\infty} \psi^2(u)du = 1 \tag{2-5}$$

Equation 2-4 implies that there are some negative and positive values generating waves. Equation 2-5 is called the unit energy property. If Equation 2-5 holds, there are finite intervals,$[-T, T]$, that must exist for any $\varepsilon \in (0,1)$, such that

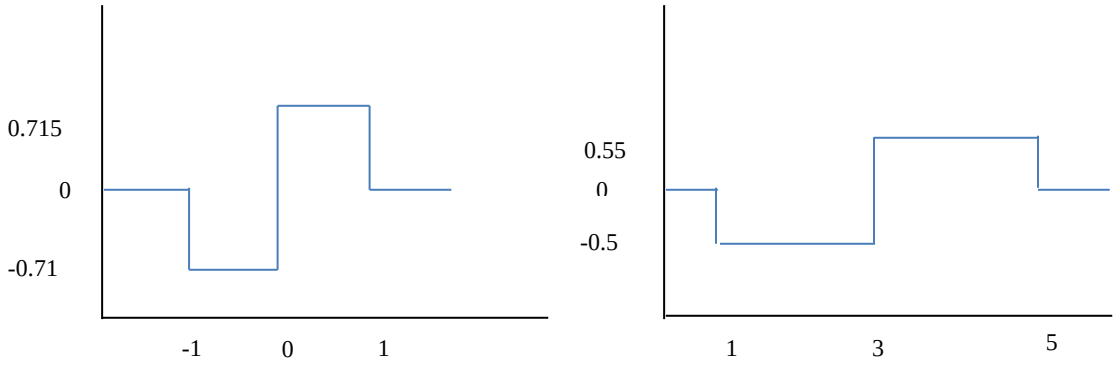
$$\int_{-T}^{T} \psi^2(u)du < 1 - \varepsilon \tag{2-6}$$

For very small ε, $\psi(\cdot)$ is negligible outside of the interval $[-T, T]$. In contrast, nonzero activity takes place in the restricted interval. Hence both Equation 2-4 and 2-5 produce a small wave, i.e., a wavelet.

For a simple illustration, we consider the Haar wavelet function, which is regarded as the first wavelet function (Haar 1910) that follows:

$$\psi^{(H)} \equiv \begin{cases} {}^{-1}/_{\sqrt{2}}, & -1 < u \leq 0; \\ {}^{1}/_{\sqrt{2}}, & 0 < u \leq 1; \\ 0, & otherwise \end{cases} \tag{2-7}$$

Figure 2-2 The Haar Wavelet Function



(a) (b)

Note: The graph (a) denotes the Haar wavelet in Equation 2-7. The Haar wavelet function with $\lambda = 2$, and $t = 3$ is plotted in (b).

The resulting Haar wavelet function is plotted in Figure 2-2. We can easily check if the above Haar wavelet function satisfies Equation 2-4 and 2-5. For example, in (a), $1/\sqrt{2} + \left(-1/\sqrt{2}\right) = 0$ and $\left(1/\sqrt{2}\right)^2 + \left(-1/\sqrt{2}\right)^2 = 1$.

It is very important to know a key concept related to wavelet analysis. Wavelets tell us how weighted local averages vary from one time period to the next. Let $x(\cdot)$ be a real valued function. The average value of $x(\cdot)$ over the interval $[a, b]$ can be given by

$$\frac{1}{b-a}\int_a^b x(u)du \tag{2-8}$$

where $a < b$ and $x(\cdot)$ is well defined. By looking into Equation 2-8 in more detail, we can set up average values at different scales and times. Let λ and t be $b - a$ and $(a + b)/2$, respectively. The former is the length of the interval, which is referred to as scale, and the latter is a midpoint of the interval, which is the center time when we consider a

time series. We can define the average value of the series $x(\cdot)$ over a scale of λ centered at time t by

$$A(\lambda, t) = \frac{1}{\lambda}\int_{t-\frac{\lambda}{2}}^{t+\frac{\lambda}{2}} x(u)du \tag{2-9}$$

Based on Equation 2-9, we can evaluate how much $A(\lambda, t)$ changes from one time period to another in an adjacent non-overlapping way. Let $D(\lambda, t)$ be the difference associated with different times between the two averages.

$$\begin{aligned} D(\lambda, t) &\equiv A\left(\lambda, t + \frac{\lambda}{2}\right) - A\left(\lambda, t - \frac{\lambda}{2}\right) \\ &= \frac{1}{\lambda}\int_{t}^{t+\lambda} x(u)du - \frac{1}{\lambda}\int_{t-\lambda}^{t} x(u)du \end{aligned} \tag{2-10}$$

Here t denotes a homogeneous time interval given a width of length λ, i.e., a scale. For example, consider a 5-minute Australian spot rate at 7:25 AM on January 2004. The value at time $t + 1$ would be recorded at 7:30 AM on January 2004. The change in scale is different from the change in time. Through temporal aggregation, we can increase the scale, λ, up to one day or even more. At this point we can discuss the effect of changes in the daily averages. Note that it is possible to take changes in averages over various scales. Thus, wavelet analysis is based on two dimensional aspects such as change in time and as well change in scale.

Since the intervals of two integrals do not overlap in Equation 2-10, we can easily combine them into a single integral

$$D(\lambda, t) = \int_{-\infty}^{\infty} \psi_{\lambda,t}(u)x(u)du \tag{2-11}$$

where

$$\psi_{\lambda,t}(u) = \begin{cases} -1/\lambda, & t-\lambda < u \le t; \\ 1/\lambda, & t < u \le t+\lambda; \\ 0, & otherwise \end{cases} \tag{2-12}$$

By changing the scale λ and time t parameters, we can extract similar information from a time series. If we apply this to the Haar wavelet function, its general form is written as

$$\psi_{\lambda,t}^{(H)} = \begin{cases} -1/\sqrt{2\lambda}, & t-\lambda < u \le t; \\ 1/\sqrt{2\lambda}, & t < u \le t+\lambda; \\ 0, & otherwise \end{cases} \tag{2-13}$$

In the case of $\lambda = 2$, and $t = 3$, the shifted and rescaled version of the Haar wavelet $\psi_{\lambda,t}^{(H)}(\cdot)$ is plotted as in (b) of Figure 2-2. According to a different level of scale λ, we can see how averages of $x(\cdot)$ are changing from one time period of length λ to the next.

There are many other basic wavelet functions. The Haar wavelet functions in Figure 2-2 are continuous wavelet transforms because the scale and location (i.e., time) are real values, not integers. In this paper, we focus on the discrete wavelet transform (DWT) based on a time series over discrete times. To enhance the understanding of maximal overlap discrete wavelet transform (MODWT), we need to have fundamental knowledge of DWT.

Let $\{X_t\}$ be a time series for $t = 0, \cdots, N-1$. So the number of elements in the series is N. The discrete wavelet transform of $\{X_t\}$ is conducted by computing

$$W = \mathcal{W}X \tag{2-14}$$

where W is a vector of length $N = 2^J$, including the j-th DWT coefficient, and $\mathcal{W}$ is an $N \times N$ real valued matrix. In the DWT the length of $\{X_t\}$ is restricted to be a power of two. In contrast, the restriction is relaxed in the MODWT. The DWT coefficients, W, can be decomposed into

$$W = \mathcal{W}X = \begin{bmatrix} \mathcal{W}_1 X \\ \mathcal{W}_2 X \\ \vdots \\ \mathcal{W}_J X \\ \mathcal{V}_J X \end{bmatrix} = \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_J \\ V_J \end{bmatrix} \tag{2-15}$$

where $\mathcal{W}_j$ has dimension $N/2^j \times N$; $\mathcal{V}_J$ is $1 \times N$; W_j is $N/2^j \times 1$, and $W_J \; and \; V_J$ are scalars. If we expand the components of the vector, W, the structure of the DWT coefficients is

$$\begin{aligned} W_1^T &= \left[\mathcal{W}_{0\bullet}^T X, \mathcal{W}_{1\bullet}^T X, \cdots, \mathcal{W}_{((N-2)/2)\bullet}^T X\right] \\ W_2^T &= \left[\mathcal{W}_{(N/2)\bullet}^T X, \mathcal{W}_{((N+2)/2)\bullet}^T X, \cdots, \mathcal{W}_{((3N-4)/4)\bullet}^T X\right] \\ &\vdots \\ W_J^T &= \left[\mathcal{W}_{(N-2)\bullet}^T X\right] \\ V_J^T &= \left[\mathcal{W}_{(N-1)\bullet}^T X\right] \end{aligned} \tag{2-16}$$

where the superscript, T, denotes transpose and $\mathcal{W}_{j\bullet}$ means a column vector containing elements of the j-th row of the $N \times N$ matrix, $\mathcal{W}$. Note that $\mathcal{W}_{0\bullet}^T$ are closely related to $\mathcal{W}_{1\bullet}^T$. The rows are circularly shifted versions of each other according to the designated rule. The $N - 1$ coefficients in W_1through W_J are called wavelet coefficients and the N-th

coefficient, V_J, in W is called a scaling coefficient. By combining all coefficients in Equation 2-16, we can consider the wavelet synthesis of X:

$$X = \mathcal{W}^T W = \sum_{j=1}^{J} \mathcal{W}_j^T W_j + \mathcal{V}_J^T V_J \tag{2-17}$$

Let us define $D_j \equiv \mathcal{W}_j^T W_j$ for $j = 1, \cdots, J$ which is the $N \times 1$ vector, and $S_J \equiv \mathcal{V}_J^T V_J$ which is the $N \times 1$vector. We can express the time series X as the sum of a constant vector S_J and other vectors $D_j, j = 1, \cdots, J$ such as

$$X = \mathcal{W}^T W = \sum_{j=1}^{J} D_j + S_J \tag{2-18}$$

where D_j is the j-th level wavelet detail and S_J is called the smooth. This is a multi-resolution analysis (MRA) of the time series, X.

A more precise definition of the DWT is given by introducing the wavelet and scaling filters in order to factor out $\mathcal{W}$ in Equation 2-14 through the pyramid algorithm. Let us define $\{h_l : l = 0, \cdots, L-1\}$ as a real valued wavelet filter, where L is the width of the filter and must be an even integer in the DWT in order to meet the orthogonality to even shifts. Here the orthogonality condition means that the inner product of the relevant components equals zero. Since h_l is a wavelet filter, it satisfies the basic properties in Equations 2-4 and 2-5. Furthermore, the orthogonal to the even shifts condition must be met.

$$\sum_{l=0}^{L-1} h_l = 0;$$

$$\sum_{l=0}^{L-1} h_l^2 = 1;$$

$$\sum_{l=0}^{L-1} h_l h_{l+2n} = 0 \tag{2-19}$$

The last two equations comprise the orthonormality condition of the wavelet filter. To illustrate the wavelet filter we can refer to Haar ($L = 2$), in which $h_0 = \frac{1}{\sqrt{2}}$ and $h_1 = -\frac{1}{\sqrt{2}}$. To check the basic properties, we can make sure that $h_0 + h_1 = 0$ and $h_0^2 + h_1^2 = 1$. And the value of $h_l h_{l-2}$ is equal to zero because the component, h_{l-2}, is zero.

Through a simple example, we can see how it works. Assume that the filter width is $L = 2$ and $N = 16 = 2^4$. The corresponding wavelet is called Haar DWT. Recalling $\mathcal{W}_j$ has the dimension of $N/2^j \times N$, $\mathcal{W}_1$ takes the form:

$$\begin{bmatrix}
-\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}
\end{bmatrix}$$

Those nonzero values usually are not unique, and here we choose $h_0 = \frac{1}{\sqrt{2}}$, and $h_1 = \frac{-1}{\sqrt{2}}$. We can obtain h_0 and h_1 by solving the following equations: $h_0 + h_1 = 0$ and $h_0^2 + h_1^2 = 1$. The rows are orthogonal to each other in the matrix,$\mathcal{W}_1$, above. For instance, pick the first two rows and then calculate their inner product. We can see the result is zero. In addition, the orthonormality can be confirmed by showing $\mathcal{W}_1 \mathcal{W}_1^T = I_{\frac{N}{2}}$, where $I_{\frac{N}{2}}$ is an $\frac{N}{2} \times \frac{N}{2}$ identity matrix. As we know the formulation of $\mathcal{W}_1$, the first level wavelet

coefficient, W_1, reduces to a $\frac{N}{2} \times 1$ vector through $\mathcal{W}_1 X$. It is called the down-sample by a factor of two.

Note that we need to circularly shift the first row in $\mathcal{W}_1$ by 2 to get the second row. In this case we have a shifting matrix, $\mathcal{T}$,

$$\begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0
\end{bmatrix}$$

Note that $\mathcal{T}\mathcal{T}^{-1} = I_N$ and $\mathcal{T}^2 = \mathcal{T}\mathcal{T}$. We can get the second row by multiplying the first row in $\mathcal{W}_1$ by $\mathcal{T}^2$. In this manner, the subsequent rows in $\mathcal{W}_1$ can be generated.

Similarly, we can construct the $\frac{N}{4} \times N$ matrix, $\mathcal{W}_2$, on a scale of two. To determine the values of the wavelet filters, we must solve the following two equations: $h_0 + h_1 + h_2 + h_3 = 0$ and $h_0^2 + h_1^2 + h_2^2 + h_3^2 = 1$. By symmetry in the sense $h_0 = -h_2$ and

$h_1 = -h_3$, we can get solutions for those two unknowns, $h_0 = h_1 = \frac{1}{2}$. In our example, $\mathcal{W}_2$ is the 4×16 matrix given by

$$\begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

As with the $\mathcal{W}_1$ matrix, the basic properties of wavelets are satisfied in the $\mathcal{W}_2$ matrix. We can get the second row in $\mathcal{W}_2$ by multiplying the first row in $\mathcal{W}_2$ by $\mathcal{T}^4$. In this fashion, for the scale of three, $\mathcal{W}_3$ consists of the 2×16 matrix and in the case of a scale of four, the last level wavelet coefficients, $\mathcal{W}_4$ is the 1×16 row vector.

The scaling filter (sometimes called the father wavelet) is closely related to the wavelet filter (or mother wavelet). In words, we can create the scaling filter $\{g_l\}$ by reversing the wavelet filter $\{h_l\}$ and then changing the sign of coefficients with even indices. For example, for the first level, the Haar scaling filter is $g_0 = g_1 = \frac{1}{\sqrt{2}}$, whereas the corresponding Haar wavelet filter is $h_0 = \frac{1}{\sqrt{2}}$, and $h_1 = -\frac{1}{\sqrt{2}}$. The scaling filters share the same properties as those in the wavelet filters. Although it is possible to discuss the scaling filter according to each of scales, as long as our interest is in the MRA, we need to focus on the scaling coefficients for the last level, $J = 4$ in this example. In our simple example, the scaling vector, V_J, is given by

$$\left[\frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}\right]$$

To summarize the DWT, it is said that the wavelet transform is localized in both time and frequency because a vector of wavelet coefficients resulting from the wavelet basis (or mother) function is associated with a particular scale and with a particular set of times. With regard to the usefulness of wavelets, they provide a localized analysis of a function in terms of changes in averages over various scales. The function describing changes in average values can be composed of two arguments, scale and time. By varying scale, we can estimate how averages of a series over various scales are changing from one period to the next. The collection of changes in the average values is called a wavelet transform.

The wavelet transform is a successive process following the pyramid algorithm. Suppose we process a time series using high-pass[12] and low-pass filters. Each of the filters can be decomposed into two parts; one is the gain function and the other is the phase[13] function, which is the phase angle of a filter. They filter out either high frequency or low frequency components. The pyramid algorithm is repeated until some frequency is removed from the series. The pyramid algorithm works as follows: we take the scaling coefficients from the first decomposition and pass it through low-pass (associated with

---

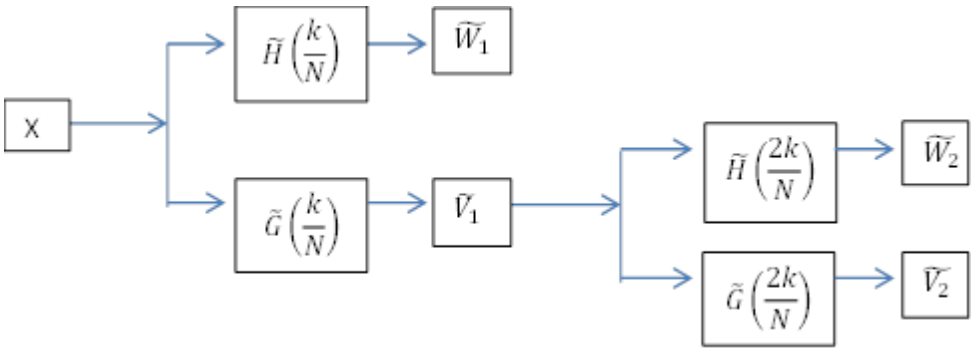
[12] When the magnitude of the gain is small at low frequencies and large at higher frequencies, the filter is said to be a high-pass filter. It is useful to capture the high-frequency dynamics of the input series.

[13] When the phase of a filter is not zero, there would be a change in the phase of the original series, which leads to misspecification of turning points in the original series.

the scaling function) and high-pass filters. We do this until the process reaches a pre-specified level.

In Figure 2-3 we describe an algorithm for computing the j-th level Maximal Overlap Discrete Wavelet Transform (MODWT) wavelet and scaling coefficients, $\widetilde{W}_j$, $\tilde{V}_j$ based on the scaling coefficients, $\tilde{V}_{j-1}$ of level $j-1$. Figure 2-3 shows a flow chart decomposing X into the MODWT wavelet and scaling coefficients $\widetilde{W}_1$ and $\tilde{V}_1$ and then decomposing $\tilde{V}_1$ into $\widetilde{W}_2$ and $\tilde{V}_2$ by the pyramid algorithm. For an arbitrary sample size N, the elements of $\widetilde{W}_j$, $\tilde{V}_j$ and $\tilde{V}_{j-1}$ are obtained by circularly filtering $\{X_t\}$ with each corresponding periodized filters $\{\tilde{h}^0_{j,l}\}$, $\{\tilde{g}^0_{j,l}\}$ and $\{\tilde{g}^0_{j-1,l}\}$, which are $\{\tilde{h}_l\}$, $\{\tilde{g}_l\}$ and $\{\tilde{g}_{l-1}\}$ periorized to length N. Here $\{\tilde{h}_l\}$ and $\{\tilde{g}_l\}$ are defined as a MODWT wavelet filter and a MODWT scaling filter, respectively. The transfer functions for $\{\tilde{h}_l\}$ and $\{\tilde{g}_l\}$ are given by $\tilde{H}(f)$ and $\tilde{G}(f)$. Each MODWT coefficient is associated with a particular portion of the original time series and hence can be associated with a time in the original series.

Figure 2-3 Flow diagram decomposing of X into wavelet and scaling coefficients



The process of figuring out the time that each coefficient should be assigned involves a study of the phase function for the filter that can be used to obtain the

coefficient from the original time series. Note that the associated squared gain functions for $\{\tilde{h}_l\}$ is real-valued and nonnegative and therefore its phase function is zero for all frequencies, which results in zero phase for the circularly wavelet filter. In order to establish the MODWT-based multi-resolution analysis (MRA) we need to define $\widetilde{D}_j$ and $\tilde{S}_j$. We can rewrite the original series X as an additive decomposition, $X = \sum_{j=1}^{J} \widetilde{D}_j + \tilde{S}_J$, in terms of the sum of a constant N-dimensional smooth vector $\tilde{S}_J$ and j wavelet details $\widetilde{D}_j, j = i, \cdots, J$. Like the case of MODWT, the filters used to generate $\widetilde{D}_j$ and $\tilde{S}_j$ hold the zero phase properties. This is very crucial in that the original series aligns with that in an additive MODWT MRA. In other words, an MRA ensures that all the characteristics in the original series are accounted for either in one of the details or in the smooth.

With this approach, high frequencies are better resolved in time and lower frequencies are better resolved in frequency. This means that a certain high frequency component can be located better in time than a lower frequency component. On the other hand, a low frequency component can be located better in frequency compared to a high frequency component. Regarding our work, one of the important technical points is to determine the value of the integer J, the level of scales. It depends on the period of oscilliation. Oscillation with a period length of about 80 occurs in our returns for spot and futures rates according to the autocorrelation function (ACF) of the absolute returns series, equivalent to almost one single day transaction. Since $2^6 = 64$, j might be 6 since the wavelet detail $\widetilde{D}_6$ captures frequencies $1/128 \leq f \leq 1/64$, where f represents

frequency. The oscillation that we are interesting in is where we want it to be (i.e., in the wavelet coefficients/details or the scaling coefficients/smooth).

Through the wavelet transform, we may remove the characteristics of the intra-day persistent components. The resulting filtered return series turn out to be stationary with long memory characteristics. However, as stated in the literature, there are gradually increasing patterns in the correlation of a pair of spot and futures returns over aggregated time intervals. While accomplishing temporal aggregation over time, we use another linear filter (i.e., exponentially weighted moving average) on the time domain. Rather than putting equal weights on every value such as a simple moving average, the exponentially weighted moving average (EWMA) can be used for our data. The latter applies the weights that decrease exponentially at distant lags. It is a kind of low pass filter that eliminates high frequency components in the series. More weights under EWMA are put on recent observations relative to a simple moving average. In the same manner, less weight should be put on the distant observations. It is more suitable because it produces a more stable and smoother frequency spectrum than the contrasting simple moving average. The value of the weight tells us the cut-off frequency to be applied. The time-stamp of the data is a short, intra-day interval, so that we need to hold the value of the parameter lower, say 0.1 or 0.2 in the application of EWMA.

2.4 Hedging Model Specifications

2.4.1 Minimum Variance Hedging Model

The model is based on a linear regression of changes in spot prices against changes in futures prices. The return on the hedge portfolio can be obtained from relations between one unit of the spot position and b units of the futures contract as follows:

$$r_p = r_s - br_f \tag{2-20}$$

where r_s and r_f represent the returns on a spot position and a futures position respectively, and r_p is the return on the hedged portfolio. The variance of the hedged portfolio return may be written as

$$h_p^2 = h_s^2 + b^2 h_f^2 - 2bh_{sf} \tag{2-21}$$

where h_p^2, h_s^2, and h_f^2 represent the variances of the hedge portfolio, the spot and futures positions respectively, and h_{sf} denotes the covariance between the spot and futures position. Then the optimal hedge ratio,

$$b^* = \frac{h_{sf}}{h_f^2} \tag{2-22}$$

is obtained from minimizing Equation 2-21. The resulting hedge ratio is the optimal number of futures contracts needed to hedge one unit of the spot position. In practice, this hedge ratio is computed as the OLS estimate in the regression of the spot return on the futures return. When the hedge ratio is above one, a hedger would be in speculation position. By contrast, when the ratio is negative, because the assets are negatively correlated, the futures position adds to the spot position, and the portfolio is not hedged.

When the associated variance and covariance are time invariant, the optimal hedge ratio is valid. However, if the variance and covariance of returns are not constant over time, the preceding hedging method is no longer appropriate. In order to improve the model specification, we can use the conditional probability of one variable given another random variable and past realized information. That is one way to update the probabilities in order to allow for new information. In the optimal hedge ratio equation, the conditional variance and covariance can be used to evaluate the time varying hedge ratios.

There is another way to compute the hedge ratio within the mean-variance framework. In this study, we use the Sharpe hedge ratio as an alternative to the minimum variance hedge ratio. Both the expected return and risk on the hedged portfolio are considered in the objective function. The Sharpe hedge ratio based on the risk-return tradeoff was proposed by Howard and D'Antonio (1984). Its objective function is constructed by the ratio of the hedged portfolio's excess return to the volatility:

$$\max_{C_f} \frac{E(r_p)-r_0}{\sigma_p} \tag{2-23}$$

where r_p represents the return on the hedged portfolio, σ_p denotes the square root of the variance of the return on the hedged portfolio, C_f is the number of units of a short futures position and r_0 is the risk-free interest rate. Note that the expected excess return can be maximized and/or the risk expressed as volatility on the denominator may be minimized in order to maximize the given objective function. Following Chen, Lee, and Shrestha (2008), the Sharpe hedge ratio can be obtained as

$$h_{sharpe} = -\frac{(\sigma_s/\sigma_f)[(\sigma_s/\sigma_f)(E(r_f)/(E(r_s)-r_0))-\rho]}{[1-(\sigma_s/\sigma_f)(E(r_f)\rho/(E(r_s)-r_0))]} \tag{2-24}$$

where r_s and r_f are log returns on the spot and futures positions, respectively, ρ is the correlation between r_s and r_f, and σ_s and σ_f are the standard deviations of r_s and r_f, respectively. Note that if $E(r_f) = 0$, the Sharpe hedge ratio reduces to the minimum variance hedge ratio.

When spot and futures prices are moving together, it is likely that there is a long run relationship. In this case an error correction term must be represented as shown later. The ignorance of the cointegration relationship among the two prices can affect the optimal hedge ratio and tend to generate a smaller hedge ratio. Ghosh (1993) shows that the consideration of the cointegration relationship produces improvements in estimation. However, Lien (2004) and Moosa (2003) document that the impact of ignoring cointegration on hedging effectiveness is not so large.

2.4.2 Constant Conditional Correlation GARCH model

We model the mean and the volatility of asset returns with extensions of the autoregressive conditional heteroscedastic (ARCH) of Engle (1982) and the generalized ARCH (GARCH) model of Bollerslev (1986). The basic idea behind volatility modeling is that the error terms in question are not serially correlated with each other over time but are dependent. Here dependence is in non-linear form because of the use of second moments. In a simple univariate GARCH(1,1) model, the current conditional volatility is explained by a constant term, which is the long-term steady state variance, the square of one lagged innovations and the past volatility term. We can describe the mean equation of a univariate time series r_t by the process

$$r_t = E(r_t|\Omega_{t-1}) + \varepsilon_t \qquad (2\text{-}25)$$

where Ω_{t-1} is the information set available at time $t-1$, and ε_t is the innovations or disturbances. The conditional variance equation of the GARCH (1,1) model is given by

$$\varepsilon_t = z_t h_t$$

$$h_t^2 = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}^2 \tag{2-26}$$

where z_t is an independent and identically distributed process with zero mean and unit variance and h_t^2 is the conditional variance. We have to impose some constraints on the parameters in order to achieve stationarity of the conditional variance such as $E(\varepsilon_t) = 0$, $var(\varepsilon_t) = \omega/(1+\alpha+\beta)$, and $\alpha + \beta < 1$.

According to the classification of multivariate GARCH models in Bauwens, Laurent, and Rombouts (2006), our focus goes to the third category, nonlinear combinations of univariate GARCH models in order to construct multivariate GARCH models. The constant and dynamic conditional correlation models and the copula-based GARCH models fall in this category. These GARCH specifications highlight the dynamics of the correlations rather than the dynamics of the covariances. For bivariate GARCH models the main component of the model specification links two individual conditional variances and the conditional correlation matrix in the CCC and DCC GARCH models and the dependence measures in the copula models.

Using the bivariate error correction model proposed by Kroner and Sultan (1993), we can construct the following model: Let s_t and f_t be the log returns of the spot and futures, respectively. S_{t-1} and F_{t-1} represent the spot and futures log prices.

$$s_t = \mu_s + a_s(S_{t-1} - \lambda F_{t-1}) + b_s(L)\varepsilon_{st}$$

$$f_t = \mu_f + a_f(S_{t-1} - \lambda F_{t-1}) + b_f(L)\varepsilon_{ft} \tag{2-27}$$

where μ_i for $i = s, f$ is the conditional mean of s_t and f_t and Lstands for lag operator used for moving average (MA) process. In our work, MA(1) or/and MA(2) error models can be used. The estimated cointegrating parameter between the individual level series is denoted by λ, which is usually a normalized outcome, and $S_{t-1} - \lambda F_{t-1}$ is the error correction term, ε_{it} for $i = s, f$ is the innovation of the respective asset return at time t. The bivariate innovation term can be formulated as follows:

$$\begin{bmatrix}\varepsilon_{s,t}\\ \varepsilon_{f,t}\end{bmatrix} |\Omega_{t-1} \sim N(0, H_t),\ H_t = \begin{bmatrix} h_{s,t}^2 & h_{sf,t} \\ h_{sf,t} & h_{f,t}^2 \end{bmatrix} \tag{2-28}$$

where it is assumed that the innovations follow the bivariate normal distribution with zero mean and the conditional covariance matrix, H_t of the innovations given the past information Ω_{t-1}. For each of the mean equations, we choose a more parsimonious model than the VAR in levels in that the number of parameters in the error correction models to facilitate forecasting. The volatility equations for both series are useful for modeling the time evolution of the conditional variances of the asset returns. Specifically, the model for the conditional variances is written as

$$\begin{aligned} h_{s,t}^2 &= \phi_s + \alpha_s \varepsilon_{s,t-1}^2 + \beta_s h_{s,t-1}^2 \\ h_{f,t}^2 &= \phi_f + \alpha_f \varepsilon_{f,t-1}^2 + \beta_f h_{f,t-1}^2 \end{aligned} \tag{2-29}$$

where ε_{it-1}^2 are the squared innovations at time $t-1$ and h_{it-1}^2 are the conditional variances for each of the log returns at time $t-1$.

The log-return data series for the GARCH models must be stationary or at most weakly defendent. Otherwise, we cannot apply the standard asymptotic properties of the parameter estimators. For instance, when we try to test the null hypothesis that a regression parameter is equal to zero in a nonstationary model, the usual t-statistics are not valid. A well-known example of the non-stationarity is the random walk model, and the simplest form is AR(1) with the coefficient of the lagged variable being unity. Unlike stationary series, the shocks in a nonstationary series would persist and never die away. One conventional way to treat non-stationarity is to use first differences of the original data.

With regard to the hedge ratio, we attempt to update it while capturing the evolution of the conditional variance. The dynamic hedge ratio at current time t can be calculated by the ratio of conditional covariance at time t between the spot and futures returns to the conditional variance of futures returns at time t (i.e., the hedge ratio is $\frac{h_{sf,t}}{h_{f,t}^2}$). As shown, both the numerator and denominator in the hedge ratio are contemporaneous. Furthermore, under the assumption about the innovations, even when there are lagged dependent variables on the right-hand sides, we can obtain consistent estimators for the parameters in the volatility equations. Alternatively, it is possible to update the Sharpe hedge ratio such as Equation 2-24 over a period of time.

Two non-stationary time series are said to be cointegrated if they are integrated of the same order and some linear combination is a stationary series. If the series are cointegrated, they have at least one long-run relationship. Even though two cointegrated series may diverge over short periods of time, they tend to move together over the long

run. For mean equations in GARCH models, we need to estimate the cointegrating parameter λ in the error correction term. The coefficient of the error correction term describes how quickly the associated variables adjust back to the long-run equilibrium when they are away from it. By using the error correction model, we can capture both short-run and long-run dynamics in the relationship.

When two random variables are $I(1)$[14] and cointegrated, the relation can be written as

$$y_t = \alpha + \beta x_t + u_t \tag{2-30}$$

where β is a cointegrating parameter and u_t has zero mean and is $I(0)$. The error term, u_t, may contain serial correlation. However, this fact does not hurt consistency of the OLS estimator of β, which is super-consistent[15] (Davidson and Mackinnon 1993, 718-719). The resulting OLS estimator of β does not follow the standard asymptotic normal distribution and the t-statistic is not a Student-t random variable. Consequently, specialized tests for cointegration are required for nonstationary data.

Although most economic variables are likely to be nonstationary, it is crucial to see if a series is $I(1)$ using a test for a unit root against $I(0)$. When a simple AR(1) model is used for the Dickey-Fuller (DF) test for a unit root, we can evaluate the null hypothesis

---

[14] $I(d)$ is said to be integrated of order d. To make a time series stationary, we must difference it d times. For example, if $\Delta p_t = p_t - p_{t-1}$ is stationary, then the series is said to be integrated of order 1, denoted by $I(1)$. $I(0)$ implies a stationary series.

[15] It follows that $\hat{\beta} - \beta = O(n^{-1})$.

by using the asymptotic critical values for the t statistic rather than the usual Student-t critical values. The reason is that the test statistic under the null hypothesis does not have an asymptotic standard normal distribution as the sample size goes to infinity. Different asymptotic distributions of the test statistic are proposed for the purpose.

To test for cointegration, there are several methods used in the literature, including the Engle-Granger method (Engle and Granger 1987) and the Johansen procedure (Johansen and Juselius 1990). Those tests enable us to see if it is possible to detect a long-run relationship between non-stationary variables. The former is a two-step method. First, we regress one variable on another and obtain the OLS residuals. Second, we check if the residuals are stationary by using an augmented Dickey-Fuller (ADF) test, which is used when autocorrelation in the errors is detected. The evaluation must be made with caution. The ADF critical values for the cointegration test are different than the usual values. When we reject the null hypothesis about the unit root, we might conclude that the associated variables are cointegrated of order of one.

Another way to conduct testing for cointegration is the Johansen approach. It can be often used to test more than one cointegrating relationship. This method has no-cointegration (i.e., a unit root) as the null hypothesis. To carry out the Johansen test, we first form the Vector Autoregressive (VAR) model with the order p of the model. Then two sets of OLS residuals based on two different dependent variables such as Δr_t and r_{t-p} are constructed. By using correlation relations we can form the associated test statistic. Finally, under the null that there are r or fewer cointegrating vectors, the test statistic based on either the trace or eigenvalue methods will be conducted.

2.4.3 Dynamic Conditional Correlation GARCH Model

In practice, the assumption of constant conditional correlations may be too restrictive. We can allow the conditional correlation matrix to be time-varying. The dynamic correlation model (Engle 2002) is

$$H_t = D_t R_t D_t \tag{2-31}$$

where $D_t = diag(h_{i,t})$ and R_t is the conditional correlation matrix. Due to the consecutive inversion in R_t, the model is computationally intensive. A dynamic correlation process can be written as

$$Q_t = (1 - \alpha - \beta)S + \alpha z_{t-1} z'_{t-1} + \beta Q_{t-1} \tag{2-32}$$

where α is a positive and β a non-negative scalar such that $\alpha + \beta < 1$, S is the unconditional correlation matrix of the standardized residuals, $z_t = u_t / h_t$. Then R_t may be obtained by rescaling Q_t as follows:

$$R_t = diag(q_{i,t})^{-1} Q_t diag(q_{i,t})^{-1} \tag{2-33}$$

where $Q_t = \begin{bmatrix} q_{s,t}^2 & q_{sf,t} \\ q_{sf,t} & q_{f,t}^2 \end{bmatrix}$ and $diag(q_{i,t}) = \begin{bmatrix} q_{s,t} & 0 \\ 0 & q_{f,t} \end{bmatrix}$.

To estimate the dynamic correlations, we first need to derive

$$\begin{bmatrix} q_{s,t}^2 & q_{sf,t} \\ q_{sf,t} & q_{f,t}^2 \end{bmatrix} = (1 - \alpha - \beta) \begin{bmatrix} 1 & s_{sf} \\ s_{sf} & 1 \end{bmatrix} + \alpha \begin{bmatrix} z_{s,t-1} \\ z_{f,t-1} \end{bmatrix} \begin{bmatrix} z_{s,t-1} \\ z_{f,t-1} \end{bmatrix}' + \beta \begin{bmatrix} q_{s,t-1}^2 & q_{sf,t-1} \\ q_{sf,t-1} & q_{f,t-1}^2 \end{bmatrix}$$

where $z_{i,t-1} = u_{i,t-1} / \sigma_{i,t-1}$ and s_{sf} is the unconditional correlation of $z_{s,t-1}$ and $z_{f,t-1}$. Then we can get the conditional correlations as

$$\begin{bmatrix} 1 & \rho_{sf,t} \\ \rho_{sf,t} & 1 \end{bmatrix} = \begin{bmatrix} 1 & q_{sf,t}/q_{s,t}q_{f,t} \\ q_{sf,t}/q_{s,t}q_{f,t} & 1 \end{bmatrix}$$

Under the structure on the conditional correlations above, its positive definiteness is met at all time periods. As Bauwens, Laurent, and Rombouts (2006) point out as a drawback, the conditional correlations would have the same dynamics because the parameters in the correlation evolution equation are scalars. The dynamic hedge ratio based on a DCC GARCH model can be obtained as the ratio of the time-varying variance components as in the CCC GARCH model, $\frac{h_{sf,t}}{h_{f,t}^2}$. Furthermore, the Sharpe hedge ratio can be updated over time with different GARCH model-dependent time-varying correlation coefficients.

2.5 Copula Models

Consider the joint cumulative distribution function $H(u,v) = Pr[U \leq u, V \leq v]$ of a pair of random variables U and V and their marginal distribution functions $F_1(u) = \Pr[U \leq u]$ and $F_2(v) = \Pr[V \leq v]$. We define the generalized inverse of a distribution function as $F^{-1}(t) = inf\{u: F(u) \geq t, 0 < t < 1\}$. A copula is defined as a function C from I^2 to $I \in [0,1]$ satisfying the following two properties;

(1) For every u, v in I,

$$C(u,0) = 0 = C(0,v) \text{ and } C(u,1) = u \text{ and } C(1,v) = v; \tag{2-34}$$

(2) For every u_1, u_2, v_1, v_2 in I such that $u_1 \leq u_2$ and $v_1 \leq v_2$,

$$C(u_2,v_2) - C(u_2,v_1) - C(u_1,v_2) + C(u_1,v_1) \geq 0. \tag{2-35}$$

For the first property, a copula is the mass of the rectangle $[0,u]\times[0,v]$ in I. The second property, namely 2-increasing[16], says that the assigned mass by C must be nonnegative (Nelsen 1999, p.8). For illustration purposes, let us introduce a product copula, which has the form

$$C(u_1,u_2)=u_1u_2, \tag{2-36}$$

where u_1 and u_2 denotes values in the unit interval. The joint distribution corresponds to the product of two independent marginals, so it is also known as the independence copula.

It is important to ask if a copula satisfies all of the properties mentioned above. Let's verify the properties with the product copula. For the first property, it obvious because $C(0,u_2)=C(u_1,0)=0$ and $C(1,u_2)=u_2$, $C(u_1,1)=u_1$. The next property we should check is that the volume (or mass) of the rectangle must be non-negative. The $C(u,v)$ can be interpreted as a mapping from a rectangle $[0,u]\times[0,v]$ to a number in I. In our illustration,

$$\begin{aligned} &C(u_2,v_2)-C(u_2,v_1)-C(u_1,v_2)+C(u_1,v_1)\\ &=u_2v_2-u_2v_1-u_1v_2+u_1v_1\\ &=u_2(v_2-v_1)-u_1(v_2-v_1)\\ &=(u_2-u_1)(v_2-v_1)\geq 0 \end{aligned} \tag{2-37}$$

[16]It refers to the non-decreasing function of rectangles, not the non-decreasing function of each argument. So some people refer to a 2-increasing function as quasi-monotone.

under the constraints $u_1 \leq u_2$ and $v_1 \leq v_2$. Thus, the product copula satisfies those properties.

For every copula C and every (u, v) in I^2, the following inequality known as the Fréchet-Hoeffding bound holds

$$\max(u + v - 1,0) \leq C(u, v) \leq \min(u, v) \tag{2-38}$$

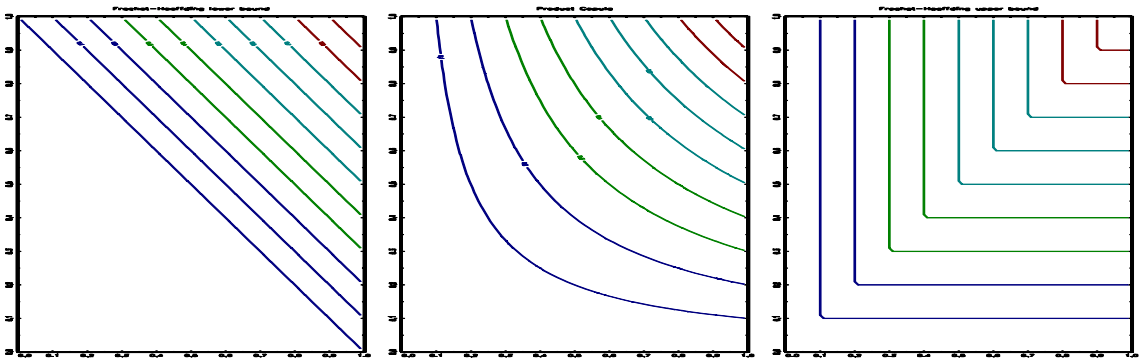
where $\max(u + v - 1,0)$ is the Fréchet-Hoeffding lower bound copula and $\min(u, v)$ is the Fréchet-Hoeffding upper bound copula. In Figure 2-4, we present the contour diagrams (i.e., level curves) of the Fréchet-Hoeffding lower bound, product, and Fréchet-Hoeffding upper bound copulas.

A link between the joint distribution functions and copulas is provided by Sklar's theorem. Joint distribution functions can be decomposed into the corresponding marginal distribution functions and a copula. Given continuous marginal distribution functions, a joint distribution function $H(u, v)$ exists such that

$$H(u, v) = C\big(F(u), F(v)\big). \tag{2-39}$$

It can be verified by the probability integral transformation that given a copula C and the associated marginal distributions, we can construct multivariate distributions with continuous margins.

Figure 2-4 Contour diagrams for the lower bound, product, upper bound copulas



There are additional properties derived from copulas that are attributable to the separation of the marginal behaviors from the dependence structure. It can be verified by Sklar's theorem and Equation 2-39 that if the univariate random variables are independent, then the dependence structure is described by Equation 2-36. If one random variable is an increasing function of another (i.e., comonotonic or perfectly positively dependent), then their copula is the Fréchet-Hoeffding upper bound copula. Similarly, if one random variable is a decreasing function of another (i.e., countermonotonic or perfectly negatively dependent), then their copula is the Fréchet-Hoeffding lower bound copula. Finally copulas are invariant with respect to increasing marginal transformations. For example, when the Gaussian copula[17] is used to link two standard normal marginals (e.g., X, Y), the same copula can be used to describe the increasing transforms, for instance, $N = \exp(X)$ and $M = \exp(Y)$.

[17] The Gaussian copula can be calculated from the bivariate normal distribution via Sklar's theorem. Given that the marginal distributions and the bivariate normal joint distribution are considered, the correlation coefficient allows us to construct the Gaussian copula. It will be defined later.

The probability density function $c(u,v)$ related to a copula $C(u,v)$ is defined as

$$c(u,v) = \frac{\partial^2 C(u,v)}{\partial u \partial v}. \tag{2-40}$$

For continuous random variables, the copula density consists of the density of the joint distribution and the marginal density distribution functions. Based on the probability integral transforms of the random variables, it can be derived as

$$c(u,v) = f\left(F_1^{-1}(u), F_2^{-1}(v)\right) \times det \begin{bmatrix} \partial X/\partial U_1 & \partial X/\partial U_2 \\ \partial Y/\partial U_1 & \partial Y/\partial U_2 \end{bmatrix}$$

$$= \frac{f\left(F_1^{-1}(u), F_2^{-1}(v)\right)}{f_1\left(F_1^{-1}(u)\right) \cdot f_2\left(F_2^{-1}(v)\right)}, \tag{2-41}$$

where f is the density of the joint distribution, f_i, are the marginal densities, $X = F_1^{-1}(U_1)$, and $Y = F_2^{-1}(U_2)$. Note that the copula density equals one when the original random variables, $X\ and\ Y$, are independent. Furthermore, by Sklar's theorem the following relationship holds

$$f\left(F_1^{-1}(u), F_2^{-1}(v)\right) = c(u,v) \cdot f_1\left(F_1^{-1}(u)\right) \cdot f_2\left(F_2^{-1}(v)\right) \tag{2-42}$$

This representation of the copula density is useful for the estimation procedures for copulas. In this paper we limit our interest to parametric classes of bivariate copulas to fit pairs of foreign exchange spot and futures returns. Every copula model is characterized by a single dependence parameter or a vector of parameters.

2.5.1 Measures of Concordance

There are various ways to measure dependence or association between random variables. The random variables X and Y are dependent if $F(x,y) \neq F_1(x)F_2(y)$ (i.e., not

independent). In the literature, various measures of dependence are closely related to copula dependence parameters. The most widely used measure is Pearson's linear correlation measure. In this section, Kendall's tau and Spearman's rho are introduced as tools to measure concordance. A pair of random variables are said to be concordant if large values of one random variable are associated with large values of the other. Conversely, the variables are said to be discordant if large values of one random variable are associated with small values of the other random variable.

Following Nelsen (1999), Kendall's tau for continuous random variables is given by the probability of concordance minus the probability of discordance such as

$$\tau_{X,Y} = Pr[(X_1 - X_2)(Y_1 - Y_2) > 0] - Pr[(X_1 - X_2)(Y_1 - Y_2) < 0] \quad (2\text{-}43)$$

If $(X_1, Y_1), (X_2, Y_2)$ and (X_3, Y_3) are independent and identically distributed, Spearman's rho is defined to be proportional to the probability of concordance minus the probability of discordance for the two vectors, $(X_1, Y_1)\ and\ (X_2, Y_3)$. Spearman's rho is given by

$$\rho_{X,Y} = 3(Pr[(X_1 - X_2)(Y_1 - Y_3) > 0] - Pr[(X_1 - X_2)(Y_1 - Y_3) < 0]) \quad (2\text{-}44)$$

Both Kendall's tau and Spearman's rho are rank based correlation measures and have the property of scale invariance with increasing transformations.

Both rank correlation measures can be represented in accordance with copulas. Given continuous random variables X and Y whose copula is C, Kendall's tau is described as

$$\tau_C = 4 \iint_{I^2} C(u, v) dC(u, v) - 1 \quad (2\text{-}45)$$

Note that the integral part in Equation 2-45 denotes the expectation of the copula function $C(u, v)$. Similarly, Spearman's rho can be expressed with copula functions

$$\rho_C = 12 \iint_{I^2} C(u, v) du dv - 3 \tag{2-46}$$

Note that Spearman's rho can be written as Pearson's correlation of the marginal distributions $F_1(x)$ and $F_2(y)$ as

$$\rho_C = \rho(F_1(x), F_2(y)) \tag{2-47}$$

Both Kendall's tau and Spearman's rho are bounded on the interval $[-1,1]$ and they are useful regardless of the functional form of the marginal distributions. For continuous random variables, those concordance measures are interchangeable with the underlying copulas.

2.5.2 Elliptical Copula Models

Elliptical distributions of the 2-dimensional random vector (X, Y) can be defined by parameters μ, Σ and ψ where μ is a vector of constants (or location vector), Σ is a 2×2 positive definite symmetric matrix (or dispersion matrix) such that the characteristic function of $X - \mu$ is expressed as a function of the quadratic form $t'\Sigma t$ i.e., $\phi_{X-\mu}(t) = \psi(t'\Sigma t)$ where ψ is the characteristic generator of the distribution. The Gaussian copula and Student-t copula are examples of elliptical copulas.

The functional form of copulas does not always exist in explicit or closed form. However, the bivariate Gaussian and Student-t copula have explicit forms. The bivariate Gaussian copula is defined as follows:

$$C^G(u.v) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v)),$$

$$C^G(u.v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} exp\left(\frac{2\rho xy - x^2 - y^2}{2(1-\rho^2)}\right) dxdy \qquad (2\text{-}48)$$

where Φ_ρ is the joint distribution function of the standard bivariate normal distribution with linear correlation coefficient ρ. For illustration, it is meaningful to check if the Gaussian copula is actually a copula. First, $C^G(0, v) = 0$ is obvious because the limiting integral over intervals of length zero is zero due to $\Phi^{-1}(0) = -\infty$. The same argument will apply to $C^G(u, 0) = 0$. In addition, since $\Phi^{-1}(1) = \infty$, $C^G(1, v)$ is determined by the other marginal distribution of a standard uniform random variable. So $C^G(1, v) = v$. By the same token, $C^G(u, 1) = u$ is satisfied. Finally, as with Equation 2-35, the last property is verified since the joint density is positive due to the exponential function. Therefore, the Gaussian copula satisfies the required properties.

The Student-t copula is defined as

$$C^t = t_{\rho,\eta}\left(t_\eta^{-1}(u), t_\eta^{-1}(v)\right),$$

$$C^t = \int_{-\infty}^{t_\eta^{-1}(u)} \int_{-\infty}^{t_\eta^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \left(1 + \frac{s^2 + w^2 - 2\rho sw}{\eta(1-\rho^2)}\right)^{-\frac{(\eta+2)}{2}} dsdw \qquad (2\text{-}49)$$

where $t_{\rho,\eta}$ is the bivariate Student-t distribution function with degrees of freedom parameter η, ρ is the linear correlation coefficient, t_η is the marginal Student-t distribution with degrees of freedom η, $s = t_\eta^{-1}(u)$, and $w = t_\eta^{-1}(v)$.

The density of the Gaussian copula is

$$c^G(u, v) = \frac{1}{\sqrt{1-\rho^2}} exp\left(\frac{X^2 + Y^2}{2} + \frac{2\rho XY - X^2 - Y^2}{2(1-\rho^2)}\right) \qquad (2\text{-}50)$$

where $X = \Phi^{-1}(u)$, $Y = \Phi^{-1}(v)$. As for the Gaussian copula, Kendall's tau and Spearman's rho can be written as a function of Pearson's correlation, ρ, as follows;

$$\tau(C^G) = \frac{2}{\pi}\arcsin(\rho) \tag{2-51}$$

$$\rho_s(C^G) = \frac{6}{\pi}\arcsin(\rho/2) \tag{2-52}$$

The density of the Student-t copula is given by

$$c^t(u,v) = \frac{1}{\sqrt{1-\rho^2}} \frac{\Gamma\left(\frac{\eta+2}{2}\right)/\Gamma\left(\frac{\eta}{2}\right)}{\eta\pi t(s;\eta)t(w;\eta)} \left(1 + \left(\frac{s^2+w^2-2\rho sw}{\eta(1-\rho^2)}\right)^{-(\eta+2)/2}\right) \tag{2-53}$$

where $s = t_\eta^{-1}(u)$, and $w = t_\eta^{-1}(v)$, where $t_\eta^{-1}(\cdot)$ denotes the inverse cumulative density function of a Student-t random variable with η degrees of freedom, and $t(\cdot;\eta)$ is the probability density function of a Student-t random variable with η degrees of freedom.

The comovement of the marginal distributions can be indicated by tail dependence. We can often observe that some relevant variables are more likely to move together. The copula functions are used for describing heavy tailed features and tail dependence. The Gaussian copula does not have tail dependence whereas the Student-t copula has more observations in the tails than the Gaussian one. We need to bear in mind that other marginal distributions may produce the different joint behaviour even for the same copula functions.

2.5.3 Archimedean Copulas

Archimedean copulas are widely used in practice due to their analytical tractability and were proposed by Schweizer and Sklar (1961). The classes of Archimedean copulas also have a convenient parametric representation. Following Nelsen (1999), given the

generator of a function C, ϕ, which is a continuous strictly decreasing convex function from [0,1] to $[0, \infty]$ such that $\phi(1) = 0$, an Archimedean copula is a function C from $[0,1]^2$ to [0,1] written as $C(u, v) = \phi^{[-1]}(\phi(u) + \phi(v))$, where $\phi^{[-1]}$ represents the pseudo-inverse[18] of ϕ. If $\phi(0) = \infty$, then the pseudo-inverse becomes an ordinary inverse function, which is said to be strict. In this case, the closed form of the copula function can be easily achieved. Archimedean copulas have three important properties. First, they are symmetric in a sense that $C(u, v) = C(v, u)$. Next, they are associative, $C(C(u, v), w) = C(u, C(v, w))$. Third, Archimedean copulas can be directly derived from Kendall's rank correlation given by

$$\tau_C = 1 + 4 \int_0^1 \frac{\phi(t)}{\phi\prime(t)} dt \tag{2-54}$$

The density function of Archimedean copulas is

$$c(u, v) = \frac{-\phi\prime\prime(C(u,v))\phi\prime(v)\phi\prime(z)}{(\phi\prime(C(u,v)))^3} \tag{2-55}$$

where a generator is assumed to be twice continuously differentiable.

2.5.3.1 Frank Copula

The bivariate Frank copula is defined by the generator,

$$\phi(t) = -ln\left(\frac{exp(-\theta t)}{exp(-\theta)}\right) \text{ with } \theta \neq 0. \tag{2-56}$$

---

[18] The pseudo-inverse of ϕ is defined as a function $\phi^{[-1]}$ such that:

$$\phi^{[-1]}(z) = \begin{cases} \phi^{-1}(z), & 0 < z \leq \phi(0) \\ 0, & \phi(0) \leq z < \infty \end{cases}.$$

The resulting Frank copula is

$$C_\theta^{Fr}(u,v) = -\frac{1}{\theta} ln\left(1 + \frac{(\exp(-\theta u)-1)(\exp(-\theta v)-1)}{\exp(-\theta)-1}\right) \quad (2\text{-}57)$$

This copula is neither lower tail nor upper tail dependent. Its copula density is

$$c_\theta^{Fr}(u,v) = \frac{\theta(1-\exp(-\theta))\exp(-\theta(u+v))}{[(1-\exp(-\theta))-91-\exp(-\theta u))(1-\exp(-\theta v))]^2} \quad (2\text{-}58)$$

Kendall's tau, $\tau\left(C_\theta^{Fr}\right)$ and Spearman's rho, $\rho_s\left(C_\theta^{Fr}\right)$ can be calculated for the Frank copula directly from the generator

$$\tau\left(C_\theta^{Fr}\right) = 1 - 4\frac{1-D_1(\theta)}{\theta} \quad (2\text{-}59)$$

$$\rho_s\left(C_\theta^{Fr}\right) = 1 - 12\frac{D_1(\theta)-D_2(\theta)}{\theta} \quad (2\text{-}60)$$

where $D_k(x) = \begin{cases} \frac{k}{x^k}\int_0^x \frac{t^k}{e^t-1}dt & x \geq 0 \\ \frac{k|x|}{1+k} + \frac{k}{|x|^k}\int_x^0 \frac{t^k}{e^t-1}dt & x < 0 \end{cases}$ is the Debye function. Furthermore, the Frank copula is comprehensive. A family of copulas which includes the Fréchet-Hoeffding lower bound, product, and Fréchet-Hoeffding upper bound copulas is called comprehensive.

2.5.3.2 Clayton Copula

When $\phi(t) = \frac{1}{\theta}\left(t^{-\theta} - 1\right)$ for $\theta \in [-1,\infty)\backslash\{0\}$ we obtain the bivariate Clayton copula as

$$C_\theta^{CL}(u,v) = max\left(\left(u^{-\theta} + v^{-\theta} - 1\right)^{-1/\theta}, 0\right) \quad (2\text{-}61)$$

In the limit as $\theta \to 0$, the copula leads to the independence copula. As $\theta \to \infty$, we obtain the comonotonicity copula. The Clayton copula produces dependence in the lower tail but not in the upper tail. The Clayton copula density is

$$c_\theta^{CL}(u,v) = (1+\theta)(uv)^{-1-\theta}\left(u^{-\theta}+v^{-\theta}-1\right)^{-2-1/\theta} \qquad (2\text{-}62)$$

Kendall's tau for the Clayton copula can be calculated by $\theta/(\theta+2)$. However, the Spearman's rho is not available in this case, but the Clayton copula is comprehensive.

2.5.3.3 Gumbel Copula

The relevant generator is $\phi(t) = \left(-\ln t\right)^\theta$ with $\theta \in [1,\infty)$, and the bivariate Gumbel copula is

$$C_\theta^{GU}(u,v) = exp\left\{-\left[(-\ln(u))^\theta + (-\ln(v))^\theta\right]^{1/\theta}\right\} \qquad (2\text{-}63)$$

If $\theta = 1$, we get the independence copula, and the copula becomes the comonotonicity copula as θ approaches infinity. The Gumbel copula density is defined by

$$c_\theta^{Gu}(u,v) = \frac{C_\theta^{Gu}(u,v)[(-lnu)(-lnv)]^{\theta-1}}{uv\left[(-lnu)^\theta+(-lnv)^\theta\right]^{2-1/\theta}}\left\{\left[(-lnu)^\theta+(-lnv)^\theta\right]^{1/\theta}+\theta-1\right\} \qquad (2\text{-}64)$$

Kendall's tau is given by $1-(\frac{1}{\theta})$. The Gumbel copula is not comprehensive.

2.5.4 Statistical Inference for Copulas

Copula methods are used extensively in finance. Patton (2006) uses copulas to model the asymmetric dependence on the foreign exchange rate markets. Jondeau and Rockinger (2006) show the dynamic dependence of the stock market returns using both the conditional and dynamic copula models. A recent survey on applications of copulas in

finance and economics can be found in Patton (2009). The conversion of Sklar's theorem allows us to construct multivariate joint distributions with a copula and arbitrary marginals. In the case of the Gaussian copula, the marginal distributions are not necessarily standard normal. For example Li (2001) uses exponential marginals to calculate marginal credit default probabilities up to time t in the future. The bivariate conditional density function of the standardized residuals from the filtered spot and futures rates is

$$f_t(x_{1,t}, x_{2,t}|\Omega_{t-1}) = c_t\left(F_1(x_{1,t}|\Omega_{t-1}), F_2(x_{2,t}|\Omega_{t-1})\right) \cdot f_1(x_{1,t}|\Omega_{t-1}) \cdot f_2(x_{2,t}|\Omega_{t-1}) \tag{2-65}$$

where $f_t(x_{1,t}, x_{2,t}|\Omega_{t-1})$ is the bivariate conditional density function of two standardized residuals, $f_{1,t}(x_{1,t}|\Omega_{t-1})$ is the conditional density of the standardized residuals of the spot returns, and $f_{2,t}(x_{2,t}|\Omega_{t-1})$ is the conditional density of standardized residuals of the futures returns.

The contribution to the log-likelihood for the observations at time t is given by

$$\sum_{t=1}^{T} lnf_t(x_{1,t}, x_{2,t}|\Omega_{t-1}, \theta) =$$

$$\sum_{t=1}^{T} lnc_t(u_t, v_t|\Omega_{t-1}, \theta_c) + \sum_{t=1}^{T} lnf_{1,t}(x_{1,t}|\Omega_{t-1}, \theta_1) + \sum_{t=1}^{T} lnf_{2,t}(x_{2,t}|\Omega_{t-1}, \theta_2) \tag{2-66}$$

where $\boldsymbol{\theta} = \{\boldsymbol{\theta_c}, \boldsymbol{\theta_1}, \boldsymbol{\theta_2}\}$ are the parameter vectors that correspond to the functions in the log-likelihood function. We can estimate the parameters in question using the maximum likelihood (ML) method, but there are several methods for estimating a set of parameters

for both the marginals and the copula. We do not use the direct maximum likelihood method in our analysis because it could be very computationally intensive. As an alternative, we focus on the inference for the marginals (or IFM) method proposed by Joe and Xu (1996). In the first stage, the parameters of the marginal distributions, $\boldsymbol{\theta}_m$ are estimated. Let $\{x_{1,t}, x_{2,t}\}_{t=1}^{T}$ be the bivariate sample data and $f_{i,t}$, $i = 1,2$ is the univariate density function. The ML estimators are obtained by maximizing the log-likelihood function of the marginals:

$$\widehat{\boldsymbol{\theta}}_m = \text{argmax}_{\theta_m} \sum_{t=1}^{T} \left[ln f_{1,t}(x_{1,t}; \theta_{1m}) + ln f_{2,t}(x_{2,t}; \theta_{2m}) \right] \qquad (2\text{-}67)$$

where $\widehat{\boldsymbol{\theta}}_m$ is a vector of estimated parameters at the first stage. In the second stage, the copula parameters are estimated by the ML method, as well. Let c and F be the copula density and the marginal cumulative distribution function, respectively. Given $\widehat{\boldsymbol{\theta}}_m$ the copula parameters are estimated by maximizing the following log-likelihood function with respect to $\widehat{\boldsymbol{\theta}}_c$,

$$\widehat{\boldsymbol{\theta}}_c = arg\max_{\theta_c} \sum_{t=1}^{T} \left(lnc\left(F_1(x_{1,t}), F_2(x_{2,t})\right); \theta_c, \hat{\theta}_m \right) \qquad (2\text{-}68)$$

where $\widehat{\boldsymbol{\theta}}_c$ is a vector of the copula parameters. Notice that the estimates of the copula parameters depend heavily on the estimates of the marginal distributions. It is very important to decide on the most appropriate parametric forms of the marginal distributions. For example, we can fit the data with a Student-t distribution which is symmetric and may capture high kurtosis. Regarding the IFM asymptotic properties, Joe (1997) shows that the IFM estimator is asymptotically normal under some regularity conditions

$$\sqrt{T}\left(\widehat{\boldsymbol{\theta}}_{IFM} - \boldsymbol{\theta}_0\right) \rightarrow N(0, G^{-1}(\boldsymbol{\theta}_0)) \quad (2\text{-}69)$$

where $\mathrm{G}^{-1}(\boldsymbol{\theta}_0)$ is the Godambe information matrix and $\boldsymbol{\theta}_0$ is the true parameter value.

2.5.5 The Copula-based GARCH Models

In this section, the marginal distributions for the spot and futures returns are specified as Student-t distributions to capture heavy tailed characteristics. Based on the error correction model for the mean equations as with Equation 2-27, the conditional volatility for the spot and futures returns is described as GARCH(1,1) with innovations that follow Student-t distributions. The specified marginals are regarded as being symmetric, fat-tailed, and non-normal.

2.5.5.1 Models for Marginal Distitributions

Most financial returns are known to be heteroskedastic and often autocorrelated, so the standardized residuals can be computed from the GARCH model with a Student t-distribution to resolve the heteroskedasticity problem. Then the standardized residuals are transformed into uniform variables by the probability integral transform, and the copula parameters are estimated in the second step. Under the previous error correction model, the bivariate innovations are assumed to follow a Student t-distribution:

$$\begin{bmatrix}\varepsilon_{s,t}\\ \varepsilon_{f,t}\end{bmatrix} |\Omega_{t-1} = \begin{bmatrix}h_{s,t}z_{s,t}\\ h_{f,t}z_{f,t}\end{bmatrix}, \text{ where } z_{i,t} \sim \text{Student-t } (z_i; v_i), i = s, f \quad (2\text{-}70)$$

Except for the parameters in the mean equations, there are four model parameters, $\theta_m = (\phi_i, \alpha_i, \beta_i, v_i)'$ in the single GARCH model with an extra parameter for the tail-fatness. As shown in the descriptive statistics, the return series in question exhibits strong

positive kurtosis. In order to capture such characteristics of the marginal behavior it is feasible to use Student-t distributions. The specification of the density for the innovations can be used to characterize the heavy tailed and non-normal features of spot and futures returns. Judging from the plots of the data, it appears that the data may be well described by a Student-t distribution in that they show symmetry about zero and a lot of data points are located in both tails (i.e., heavy tailed).

2.5.5.2 Conditional Copula function

Let us define the conditional cumulative distribution functions of the standardized residuals, $z_{s,t}$ and $z_{f,t}$ as $\mathrm{F}_{s,t}(z_{s,t}|\Omega_{t-1})$ and $\mathrm{F}_{f,t}(Z_{f,t}|\Omega_{t-1})$, respectively. Let Ψ_t be the joint conditional distribution functions of $z_{s,t}$ and $z_{f,t}$. By Sklar's theorem, we can construct the following equation:

$$\Psi_t(z_{s,t}, z_{f,t}|\Omega_{t-1}) = C_t\big(\mathrm{F}_{s,t}(z_{s,t}|\Omega_{t-1}), \mathrm{F}_{f,t}(z_{f,t}|\Omega_{t-1})\big) \tag{2-71}$$

According to Equation 2-65, the bivariate conditional density functions of $z_{s,t}$ and $z_{f,t}$ are given by

$$\begin{aligned}\psi_t(z_{s,t}, z_{f,t}|\Omega_{t-1}) = c_t\big(\mathrm{F}_{s,t}(z_{s,t}|\Omega_{t-1}), \mathrm{F}_{f,t}(z_{f,t}|\Omega_{t-1})|\Omega_{t-1}\big) \\ \times f_{s,t}(z_{s,t}|\Omega_{t-1}) \times f_{f,t}(z_{f,t}|\Omega_{t-1})\end{aligned} \tag{2-72}$$

where c_t is the copula density function, $f_{s,t}$ is the conditional density of $z_{s,t}$, and $f_{f,t}$ is the conditional density of $z_{f,t}$.

2.5.5.3 Dynamics for Dependence Structure

With an aim to capture both persistence and time variation in the conditional copula

through the dependence structure, Patton (2006) uses the logistic transformation based on a restricted ARMA(1,10)-type process. The correlation is updated by the past dependence and the previous average difference of the probability integral transforms of the two time series. However, as pointed out in Bartram et al., (2007) the logistic transformation would unnecessarily restrict the volatility of the dependence around its limiting values. To specify the persistence in dependency, Jondeau and Rockinger (2006) employ two different methods such as the time-varying correlation (TVC) model proposed by Tse and Tsui (2002) and the Markov-switching model. They attempt to compute short-run correlation over one week of data in the TVC model. In practice, a correlation measure and a pre-specified length for the time variation term in the dependence equation should be investigated with more care. In the spirit of Patton (2006) and Bartram et al. (2007), we can build the parsimonious dependence process as follows:

$$\rho_t = c + a\frac{1}{10}\sum_{j=1}^{10}|u_{t-1} - v_{t-1}| + b\rho_{t-1} \tag{2-73}$$

The dynamics in the dependence relationship depend on the previous dependence ρ_{t-1} to capture persistence, and past absolute differences (i.e., a variant of ARMA (1,10)), $|u_{t-1} - v_{t-1}|$ to capture variation in the dependence process. The past absolute value term represents the difference between the cumulative probabilities since u_t and v_t are uniform variates from the marginal cumulative distribution functions. Note that the parameter b captures persistence and a denotes the past short-term correlation between the margins. In the cases of the Normal and Student-t copulas, we use the logistic transformation of the linear correlation evolution in Equation 2-73 to keep them in the range (0, 1).

2.5.5.4 Estimation

For the various copula models, we first generate the standardized residuals from the fitted GARCH(1,1) model based on the error correction model. Second, we need to transform the standardized residuals into uniform variables using the Student-t distribution function under the probability integral transform. Third, in the case of the Gaussian copula, we use the inverse of the standard normal distribution to transform the resulting uniform variables. For the Student-t copula we use the inverse of the Student-t distribution. Fourth, following the pre-specified updating rule of the dependence structure, we maximize the log-likelihood function over the set of all possible correlation matrices. The expression for the log-likelihood is:

$$l^{C_G}(\theta) = -\frac{T}{2} ln|R| - \frac{1}{2}\sum_{t=1}^{T} S_t'(R^{-1} - I)S_t \tag{2-74}$$

where θ denotes parameters, R is the correlation matrix and $S_t = \left(\Phi^{-1}(u_t), \Phi^{-1}(v_t)\right)'$. In the Student-t copula model, the log-likelihood function for the sample data is:

$$l^{C_t}(\boldsymbol{\theta}; v) = Tln\left(\Gamma\left(\frac{v+2}{2}\right)\right) + Tln\left(\Gamma\left(\frac{v}{2}\right)\right) - 2Tln\left(\Gamma\left(\frac{v+1}{2}\right)\right) - \frac{T}{2} ln(|\Sigma|)$$

$$-\frac{v+2}{2}\sum_{t=1}^{T} ln\left(1 + \frac{\omega_t \Sigma^{-1} \omega_j}{v}\right) + \frac{v+1}{2}\sum_{t=1}^{T} ln\left(\Pi\left(1 + \frac{\omega_t^2}{v}\right)\right) \tag{2-75}$$

where ω_t is the column vector of $t - th$ row of $X = cdfti(\boldsymbol{U})$.

We need to construct the conditional variances in Equation 2-70 in order to compute the dynamic hedge ratios based on the various copula-based GARCH models after obtaining the copula parameters along with the GARCH model parameters. By

using Equation 2-65, we construct the off-diagonal element, $h_{sf,t}$, in the conditional variance matrix as

$$h_{sf,t} = h_{s,t}h_{f,t}\iint_{-\infty}^{\infty} z_{s,t}z_{f,t}\, f_t\left(z_{s,t}, z_{f,t}|\Omega_{t-1}\right)dsdr \tag{2-76}$$

Given the conditional variance matrix, we can calculate the optimal hedge ratio by the ratio of the off-diagonal element to the diagonal element for the futures returns, $\frac{h_{sf,t}}{h_{f,t}^2}$. Alternatively, we can use the Sharpe hedge ratio. In order to calculate the Sharpe hedge ratio we use Equation 2-24. We can replace the correlation coefficient with an estimated value obtained by computing the approximated double integral component in Equation 2-76.

2.6 Empirical Results

2.6.1 The data

The financial data for this study consist of the currency exchange rates and the corresponding futures contract[19] prices. Futures contracts[20] cover the major currencies, including the Australian dollar, the Swiss franc, the euro, the British pound, and the Japanese yen. The spot foreign exchange rates have round-the-clock observations, which

[19] The spot foreign exchange market is the world's largest market, with over one trillion U.S. dollars traded per day. The foreign exchange futures market is only 1/100th as large as the spot market.

[20] The contract has a quarterly expiration cycle in March, June, September, and December. Futures expire on the third Wendesday of the expiring contract month. For example, the September futures contract will expire on the 3rd Friday in September.

last throughout the entire 24 hours of the day on an electronic platform. Due to the slower trading patterns over the weekends, all raw observations from Fridays 22:00 Greenwich Mean Time (GMT) through Sunday 22:00 GMT are excluded. The corresponding futures contracts traded on the Chicago Mercantile Exchange (CME) are used. In all cases, there is no continuous series of futures prices because each individual contract in the futures market has a pre-determined life-span. As one contract expires, others with later expiration dates continue to trade, and the cycle continues. In general, the different contracts in a futures market are not homogeneous. When one contract is replaced in the series by another, the difference in price between the two will produce a gap due to different expiration dates. To construct continuous futures contracts, we resort to the back-adjusted continuous contract. We measure every gap and adjust all previous prices up or down according to whether the gap happens to be positive or negative. Working backwards, this eliminates the gaps. Back-adjusted contracts display price movements in a future market that are due to trading activity. When a futures exchange has a holiday and its trading floor is closed, the exchange does not declare settlement prices for the day in question and no data is supplied. We also account for the gaps associated with trading holidays.

The primary data are 5-minute time interval prices obtained from a private data vendor, ForexTickData.com. The entire sampling period is from January 2, 2004 to April 23, 2009. The futures data are both pit and/or electronic prices during the day trading session. Note that the further you go back in history the electronic overlap disappears and we just have pit trading. There is 24 hour GLOBEX exchange trading, but this data is

almost useless because there is limited liquidity or activity and thus the data quality is poor. We convert the spot exchange rate series to the Central Standard Time (CST) zone and take into account daylight saving time. In order to synchronize the data, we align both spot and futures series to the nearest date and time. Futures prices are always quoted as the number of U.S. dollars (USD) per unit of the foreign currency. For the spot prices, the quoted currency is per USD. For example, the EUR/USD exchange rate in a quotation is 1.4320, which means 1.4320 USD per EUR. Quotes using a country's home currency as the unit currency (e.g., EUR 1.00 = USD 1.35991 in the euro zone) are known as indirect quotation or quantity quotation and are used in British newspapers and are also common in Australia, New Zealand and the Euro zone.

Missing data points can exist in the data set, and if the missing values are not properly handled, they can generate unexpected outcomes such as incorrect sums and means. The existence of outliers[21] can cause big trouble because they may induce substantial distortions in parameter estimation and lead to model misspecification. We need to look at the distribution of the data on a graph and decide where the outliers are, and we can delete observations shown as either higher or lower than what we think is reasonable.

For the futures series, we discard some missing values, including those due to temporary closures in the exchange, which contain more than 6 missing values in a row.

---

[21] In Barnett and Lewis (1994), it is defined as "an observation (or subset of observations) which appears to be inconsistent with the remainder of that set of data", on page 7.

If there are no more than 6 missing observations, we impute the missing values through cubic spline curves[22] to the non-missing values of variables rather than changing the missing values to zero. For spot exchange rates, there are duplicate values in subsequent rows. We regard those observations as missing values and treat them in the same way we used for the data.

In the time series literature, there are three types of outliers: additive, innovation, and level shifts. An additive outlier at some time point, t_0, corresponds to a shock indicator, ξ_{t_0} such that $\xi_{t_0} = 1$ and $\xi_{t_0} = 0$ otherwise. The effect is limited to the time period t_0. Similarly an innovation outlier is about the estimated residual. Unlike the additive outliers, an innovation outlier can influence the subsequent observations in a general ARMA model context. A level shift at a specific time t_0, has a shock indicator ξ_{t_0}such that $\xi_{t_0} = 0$ for $t < t_0$ and $\xi_{t_0} = 1$ for $t \geq t_0$. In this case, the first differenced series has an additive outlier at time t_0. Due to the characteristics of outliers, we need to take into account those outliers in the modeling process combining outlier detection and residual analysis. Tukey (1977) suggests that any values greater than the 75th percentile plus 1.5 times the interquartile (IQR) distance or less than the 25th percentile minus 1.5 times the interquartile (i.e., inner fences) distance can be viewed as outliers. There are a lot of observations even beyond these outer bounds in our sample. However, original

[22] The cubic spline interpolation is used to draw smooth curves through a number of points. This spline consists of weights attached to a flat surface at the points to be connected. The weights are the coefficients on the cubic polynomials used to interpolate the data in a smooth way.

observations that pass through wavelet filter are treated and transformed so that we may not able to remove extremal data points in the whole sample.

Prices are computed as the average of the bid and ask prices. They are estimates for the transaction prices. In high-frequency data, it is very common to have transmission delays because every market maker cannot move at the same time (Goodhart et al., 1995). We need to model an effective price which is close to the transaction price. As before, we adopt the averages of bid and ask price as an effective price algorithm.

Returns on both instruments are calculated as logarithmic returns or continuous compounding returns, which can be obtained as the compounding period gets too small[23]. The return is preferred to the price because the distribution of returns is generally more symmetric over time and the return series is more likely to be stationary. Mathematically, the log return is defined as

$$r_t = 100 \times (\ln(P_t) - \ln(P_{t-1})) \tag{2-77}$$

where P_t is the price observed at time t and P_{t-1} is the price at time $t-1$. For our data, r_t represents a five-minute interval return at time t. For the spot exchange rate, the bar time stamp represents the beginning of the associated 5-minute time interval. On the other hand, the futures data bar time is stamped as the end of the 5-minute time interval. We use the modified representation for the intra-daily returns to incorporate the following idea: the disturbances (or innovations) are assumed to be serially uncorrelated, whereas

[23] A price for today would be evolved exponentially such as $P_1 = P_0 e^r$, where P_1 denotes the price for tomorrow and r is the log return.

they do not seem to be independent due to the intrinsic autocorrelation captured by the repetitive periodicity in the volatility of the return series. The observations above cause the absolute returns to be dependent as shown in the ACF plots. Algebraically, the returns can be formalized, as in Andersen and Bollerslev (1997), by

$$R_t = r_t - m_t = v_t \cdot s_t \cdot \varepsilon_t \tag{2-78}$$

where m_t is the conditional mean of the raw returns (r_t), v_t is the long-term volatility, s_t is the intra-daily seasonal volatility, and ε_t is the independent and identically distributed innovations. By squaring both sides of Equation 2-78, taking the natural logarithm, and then dividing both sides by two, we can get the following equation:

$$ln|R_t| = ln|v_t| + ln|s_t| + ln|\varepsilon_t| \tag{2-79}$$

In the literature, researchers have found that the autocorrelations of logarithmic absolute returns, $ln|R_t|$ based on 5-minute spot exchange rates decay at a hyperbolic rate rather than an exponential rate, as shown in Andersen et al., (2001) and Gençay et al., (2001). As a volatility measure, we retain the absolute value of the returns rather than the squared value because the absolute returns may better capture the autocorrelation and the seasonality of the data following Müller et al., (1997), Granger and Ding (1996), and Guillaume et al., (1997). One of most common stylized facts in high-frequency financial time series data are volatility clustering, which means that highly volatile periods are next to each other. This is associated with a long memory process in levels (i.e., returns). Another common feature of volatility includes intra-daily seasonalities (or periodicities). It is common to see intra-daily seasonalities on the daily basis, reflecting that the seasonal

pattern repeats day after day with market opening and closing (Goodhart and O'Hara 1997). The presence of seasonalities may reduce the underlying low-frequency dynamics. Such seasonalities can pull down the sample autocorrelation so that it seems not to exhibit any persistence.

There are several approaches to deal with seasonality and non-linearities. First, seasonal dummies can be used to capture seasonality. When the seasonal pattern is deterministic, the seasonality can be described by seasonal dummy variables (Baillie and Bollerslev 1990). Second, as shown in Dacorogna et al. (1993), a new time scale (called theta time) other than physical time is introduced to analyze the seasonal volatility pattern. Third, spectral analysis based on cyclical frequences can also be used to extract the seasonality in volatility such as Andersen and Bollerslev (1997). However, when a time series is not stationary due to the complicated components such as trends and volatility clustering, the Fourier analysis is not appealing.

More recently, Gençay et al., (2001) propose a wavelet multi-scaling model to filter out the underlying seasonality in volatility from the data. An important advantage of the wavelet transform over the Fourier transform is that the wavelet functions are localized in space. For example, observations in a certain interval may have a high peak. A Fourier series may not capture this feature on the range because its basis functions are the sine and cosine functions. However, the wavelet functions are good at representing such localized properties by choosing the most suitable basis functions among an infinite number of candidates.

With regard to the foreign exchange (FX) rates, their data frequency can vary from very few quotes to hundreds of quotes per hour. A dramatic decrease in correlation as data frequency enters the intra-hour level for both stocks (Epps 1979) and foreign exchange returns (Guillaume et al. 1997) is referred to as the Epps effect. The results reported in Dacorogna et al. (2001) show that a pair of currencies that are highly correlated in the long term (buy and hold) become less correlated in the intra-hour data frequency data. They find that the Epps effect disappears as the frequency approaches 10 minutes to 20 hours by relying on a particular pair of currencies among the same instruments. Table 10.5 on page 291 in chapter 10 of Dacorogna et al. (2001) includes the stabilization interval, which is the threshold after which the Epps effect vanishes. As one possible explanation, Müller et al., (1997) focus on the behavior of heterogeneous agents. For the short-term agents, they put more emphasis on an individual instrument or asset than a multivariate set of instruments. Less responsive agents will take some time to generate more stable correlation. The consideration would be applied to the case between the financial instruments and the derivatives of the instruments.

Unlike the prediction from the findings above, correlation behavior between a pair of filtered spot and futures returns at different time intervals do not stay high. The addition of successive changes of a return series is called temporal aggregation in the series over time. From the shortest time intervals (i.e., 5-minute) to the day-length intervals, the correlation coefficients between a pair of the spot and futures returns for each of the foreign exchange rates are close to about 0.1 through temporal aggregation, e.g., 1 hour and 30 minutes while there are fluctuations in narrow temporal ranges about

zero. The reason that we can observe the outcome above is that the return series employed in the calculations are the filtered returns through the wavelet transform, which is adopted to remove noises that appeared in the original return series. The process of de-noising the original returns may lead to different results for the correlation coefficients. Fortunately, in the case that we use the filtered returns based on the absolute value of the original returns, the correlation behavior is quite consistent and shows a gradual increasing trend as measured by both the Pearson and Spearman measures. On the basis of those outcomes, we can obtain an appropriate time series through aggregation such as a 2-hour or 3-hour interval series. This is reasonable in the sense that we not only take into account the Epps effects but also have meaningful correlation coefficients.

As a preliminary statistical test for the existence of an intraday price returns spike, the equality of mean returns across intraday trading intervals for each weekday was tested by applying an F-statistic. Using the NPAR1WAY procedure in SAS, we perform tests for location and scale differences across a one-way classification. It is based on the empirical distribution function and uses the Kruskal-Wallis test. In a simple case as above, a dummy variable is created that takes on the value one for the opening observations for the trading days and the value zero for the other observations. The overall F-test based on the GLM procedure in SAS is significant ($F = 8.89$, and $-value = 0.0029$), indicating strong evidence that the mean for the two levels is different at a significance level of 5%. The results indicate significant differences in mean returns at the opening of trading. On the other hand when we break the sample down into each of the weekdays, some days show insignificant differences in mean returns.

Nevertheless, it turns out that it is useful to adjust the gap between the opening of trading in a particular day and the last trading from the previous day. So every opening transaction is divided by the time gap. For the intra-daily day of the week effect, it turns out that returns on the futures contracts for the currencies on Mondays will not follow the prevailing trend from the previous Friday.

In Table 2-1, we display summary statistics for the filtered return series and present the results of applying the ADF and ARCH tests. Our data are pairs of spot and futures returns. That is, the underlying asset in the futures currency contracts is a certain number of units of the foreign currency. The number of observations reaches at least 93,226. The returns on each variable are skewed and show large excess kurtosis. Six of the total series have negative skewness, indicating that they have long tails to the left of the mean. The large positive values for excess kurtosis tell us that the returns near the mean on all cases have higher peaks and heavier tails than the normal distribution. As a result, all cases in question are found to be non-normally distributed. The test results of the Jarque-Bera statistics back up the non-normality as the statistics are large enough to reject the null hypothesis of normality.

The joint tests confirm that all returns in question have no significant serial correlations because the p-values for these test statistics are close to one. The sample Autocorrelation Functions (ACFs) are all within their two standard error limits, indicating that they are not significantly different from zero at the 5% level.

Table 2-1 Summary Statistics

Series	Nobs	Mean	Stdev	Skew	Kurt	J-B	Q(15)	ADF	ARCH(5)
AUDs	4985	-0.0033	0.049	-1.53	14.1	<0.00	<0.00	<0.01	<0.00
AUDf	4985	0.0053	0.072	-1.17	8.8	<0.00	<0.00	<0.01	<0.00
AUDs_unf	4985	-0.0031	0.052	-1.19	15.4	<0.00	<0.00	<0.01	<0.00
AUDf_unf	4985	0.0035	0.075	-1.42	10.5	<0.00	<0.00	<0.01	<0.00

The last two rows are based on the unfiltered data sets in a sense that the wavelet transforms are not used. J-B is the Jarque-Bera test for normality. The numbers in J-B represent the p-values. Q(15) is the Ljung-Box statistic for the serial correlations up to the 15th-order. P-values are reported for Q(15) column. ADF is the Augmented Dickey-Fuller unit root test. ARCH is the LM test for up to the fifth-order ARCH effects. Kurtosis refers to the excess kurtosis, the value 0 for a normal distribution.

When a shock in innovations persists over a long period of time, such economic data are regarded as having a long memory (or long range dependence) property. The presence of long memory in asset returns is contrary to the market efficiency hypothesis that future returns are hard to predict conditional on past information. The long memory process can be denoted by a fractional degree of integration which remains between 0 and 1 and exhibits hyperbolically decaying autocorrelations. With relevance to the slow decay of the correlations, it is intrinsically an asymptotic behavior. A stationary time series $\{r_t\}$ is said to be a long memory process if the autocovariance function at lag h of the process is:

$$\sum_{h=-\infty}^{\infty}|\gamma(h)| = \infty \tag{2-80}$$

where $\gamma(h) = \langle r_t, r_{t+h}\rangle$ is the autocovariance function. Another widely used definition on the spectral domain is

$$f(\lambda) \sim b|\lambda|^{-2d} \text{ as } \lambda \to 0 \tag{2-81}$$

where $f(\lambda)$ is the spectral density function for λ in a neighborhood of zero and b is a slowly varying function, but here it is assumed to be a positive constant. So, as $\lambda \to 0$, $f(\lambda)$ tends to ∞ when $d > 0$, and to zero when $d < 0$. If $d = 0$, $\{r_t\}$ is said to be short memory. In this case, $f(0)$ will be positive and finite. All stationary invertible ARMA processes are short memory. When $d = 1$, we have a model based on first differences, $(1 - B)r_t = \varepsilon_t$ which would apply the filter, $\left(1 - e^{-i\lambda}\right)\left(1 - e^{i\lambda}\right) = \left|1 - e^{-i\lambda}\right|^2$ to its spectrum density, $f(\lambda)$. Granger and Joyeux (1980) and Hosking (1981) introduce models of fractional differencing. From the theoretic perspective, the fractional $ARIMA(p, d, q)$ model is defined for the process $\{r_t\}$ as

$$\phi(B)\Delta^d r_t = \theta(B)\varepsilon_t \tag{2-82}$$

where B is the backward shift operator and $\{\varepsilon_t\}$ is zero mean white noise, $d \in (-0.5, 0.5)$ and $\phi(B)$ and $\theta(B)$ are polynomials of orders p,q in an $ARMA(p, q)$ process. If $d = 0$, the process $\{r_t\}$ is stationary, which reduces to an $ARMA(p, q)$ process. From Equation 2-82, we can see that fractional $ARIMA(p, d, q)$ in the $d - th$ difference $\Delta^d r_t$ is $ARMA(p, q)$. Fractional differencing may be obtained from the binomial expansion (e.g., Granger and Joyeux (1980) and Hosking (1981)) of the operator Δ^d:

$$\Delta^d y_t = (1 - B)^d y_t = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k y_t$$

$$= y_t - d y_{t-1} - \frac{1}{2} d(1 - d) y_{t-2} - \frac{1}{6} d(1 - d)(2 - d) y_{t-3} + \cdots \tag{2-83}$$

There are several estimation methods for long memory models. It is possible to implement the estimation in either the time domain or frequency domain. The examples

of the former are a Fractional Dickey-Fuller test (Donaldo et al., 2002) or an R/S test (Lo 1991). The R/S method is criticized by Beran (1994) because it does not yield correct statistical inference which is not efficient. As one of the examples of the latter there is Whittle estimation, which is based on the calculation of the periodogram using an approximate maximum likelihood method. Another approach is proposed by Geweke and Porter-Hudak (1983) that estimates the long memory parameter, d from the spectrum behavior using OLS. This method for estimating the long memory parameter is used extensively due to computational simplicity. The estimator is regarded as being semi-parametric because it does not assume a full parametric model for the short-term structure of the innovation (or error), and the bandwidth parameter must should be chosen. The relevant periodogram ordinates, $n^{1/2}$ is originally proposed by Geweke and Porter-Hudak (1983), where n is the number of sample observation.

In practice, researchers have found that the estimate of the long memory parameter is sensitive to the frequency of observation. In the case of implementing the GPH estimator for the long memory parameter, d, there is a controversial issue about how to determine the spectral bandwidth in the semi-parametric frequency-domain estimation. Hurvich et al., (1998) show that the GPH estimator of the long memory parameter is consistent under the Gaussian assumption and is asymptotically normal[24] as n gets big. The approach works out under the minimum mean squared error criterion. The fact

[24]The asymptotic normality is written as $\sqrt{m}(\hat{d} - d) \rightarrow N\left(0, {\pi^2}/{24}\right)$ where m is the number of Fourier frequencies, called bandwidth.

implies that since the asymptotic variance depends only on m, the bandwidth choice is critical to the asymptotic behavior of the estimator. They propose that the optimal bandwidth is $n^{4/5}$.

For our work, we set up mean equations in a simple and classical form. For instance, a mean equation for the filtered spot series has an intercept and an error correction term. In representing volatility process, we rely on a GARCH-based model. Given that volatility of various financial data is known as being persistent as in Andersen and Bolleslev (1997), Ding, Granger, and Engle (1993) and Baillie, Bollerslev, and Mikkelsen (1996), the use of the fractionally integrated GARCH (FIGARCH) model may be helpful with capturing the slow decay of volatility which is measured by absolute or squared returns. In the standard GARCH model, the volatility is supposed to decay at an exponential rate. By contrast, it decays at a slower rate in the FIGARCH model. The consideration of FIGARCH was introduced by Baillie, Bollerslev, and Mikkelsen (1996) and applied by Brunetti and Gilbert (2000) to the bivariate case. Since the data used in this study shows long-memory in volatility, we need to get fractionally integrated series on the absolute values of residuals from mean equations.

A representation of the FIGARCH process suggested by Baillie, Bollerslev, and Mikkelsen (1996) can be written as

$$\big(1-\beta(L)\big)h_t^2 = \omega + (1-\beta(L)-\phi(L)(1-L)^d)\varepsilon_t^2 \qquad (2\text{-}84)$$

where $\phi(L) = 1-\beta(L)-\alpha(L)$ and $\beta(L)$, $\alpha(L)$ are lag polynomials in the lag operator, and $(1-L)^d$ is a fractional filter. Note that when $d = 0$, Equation 2-84 reduces to a

GARCH model, and when $d = 1$, it becomes an integrated GARCH (IGARCH) model. This representation can lead to infinite ARCH form such as

$$h_t^2 = \frac{\omega}{1-\beta(1)} + \lambda(L)\varepsilon_t^2 \tag{2-85}$$

where $\lambda(L) = \left[1 - \frac{\phi(L)(1-L)^d}{1-\beta(L)}\right]$. In practice, it is important to note that the conditional variance should be positive at all times. The sufficient conditions for the non-negativity in Equation 2-85 is $\lambda(L) = \lambda_1 L + \lambda_2 L + \cdots \geq 0$. Following Bollerslev and Mikkelsen (1996), the positive definiteness in the FIGARCH(1,d,1) can be achieved under the sufficient conditions

$$\beta_i - d_i \leq \frac{1}{3}(2 - d_i)$$

$$d_i\left[\phi_i - \frac{1}{2}(1 - d_i)\right] \leq \beta_i(\phi_i - \beta_i + d_i), i = 1,2 \tag{2-86}$$

The FIGARCH process can be stationary when $0 \leq d_i \leq 1$. From the theoretic perspective, in the FIGARCH framework, the variance of the unconditional distribution of the error process, $\{\varepsilon_t\}$ is infinite. In empirical works, to overcome the problem, as suggested in Baillie, Bollerslev, and Mikkelsen (1996) a truncation of sample observations is used under some regularity conditions. They choose 1,000 observations as the pre-sample values.

2.6.2 Comparison to Hedge Effectiveness

Different approaches to estimating the optimal hedge ratio (i.e., b^*) are investigated and evaluated according to how much each method reduces the variance of the hedge portfolio return relative to the OLS-based hedge ratio. A hedge portfolio is the linear

combination of the spot and futures returns, by replacing b with b^* from Equation 2-20. The hedging performance is first evaluated for the in-sample case based on the entire set of data. Then the out-of-sample hedging effectiveness for a 120-minute time horizon may be investigated. First, the conditional variance matrix is estimated for the various model specifications. Second, these conditional variance forecasts are used to compute the optimal hedge ratios following Equation 2-22. Third, we compute the variance of the hedge portfolio return and compare it to the variance of the hedge portfolio formed by the OLS based hedge ratio which is known as the minimum variance hedge ratio. As an alternative to the minimum variance hedge ratio, the Sharpe hedge ratio based on Equation 2-24 can be calculated and compared.

With the optimal hedge ratio, b^*, the variance of the hedge portfolio return for the in-sample case is calculated from

$$r_p = r_s - b_t^* r_f \tag{2-87}$$

where b_t^* denotes the static hedge ratio. On the other hand, for the out-of-sample forecasts, the variance of the hedge portfolio return is calculated from

$$r_p = r_s - b_{t-1}^* r_f \tag{2-88}$$

where b_{t-1}^* represents the optimal dynamic hedge ratio conditional on the information set available at time $t-1$. Based on these considerations, the variance reduction relative to the OLS-based method can be computed in percentage terms by the following formula

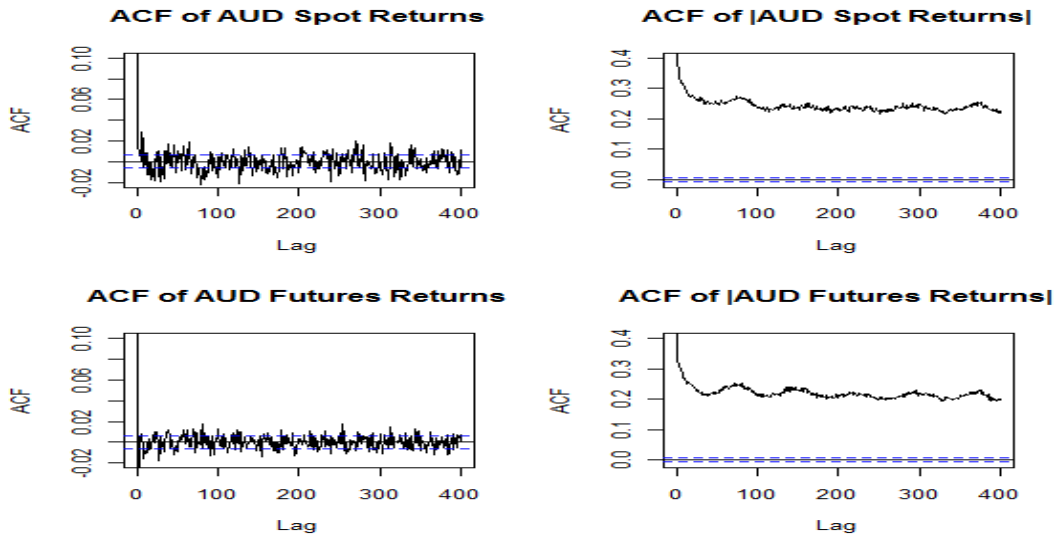
$$E = \frac{Var(r_{ols}) - Var(r_{alternative})}{Var(r_{ols})} \tag{2-89}$$

where E stands for hedging effectiveness, $Var(\cdot)$ denotes the variance of the hedge portfolio, r_{ols} is the hedge portfolio of the OLS-based optimal hedge ratio, and $r_{alternative}$ is the hedge portfolio of an alternative optimal hedge ratio. The same hedging effectiveness criterion as Equation 2-89 will be applied to the Sharpe hedge ratio framework.

2.6.3 Analysis of Foreign Exchange Returns

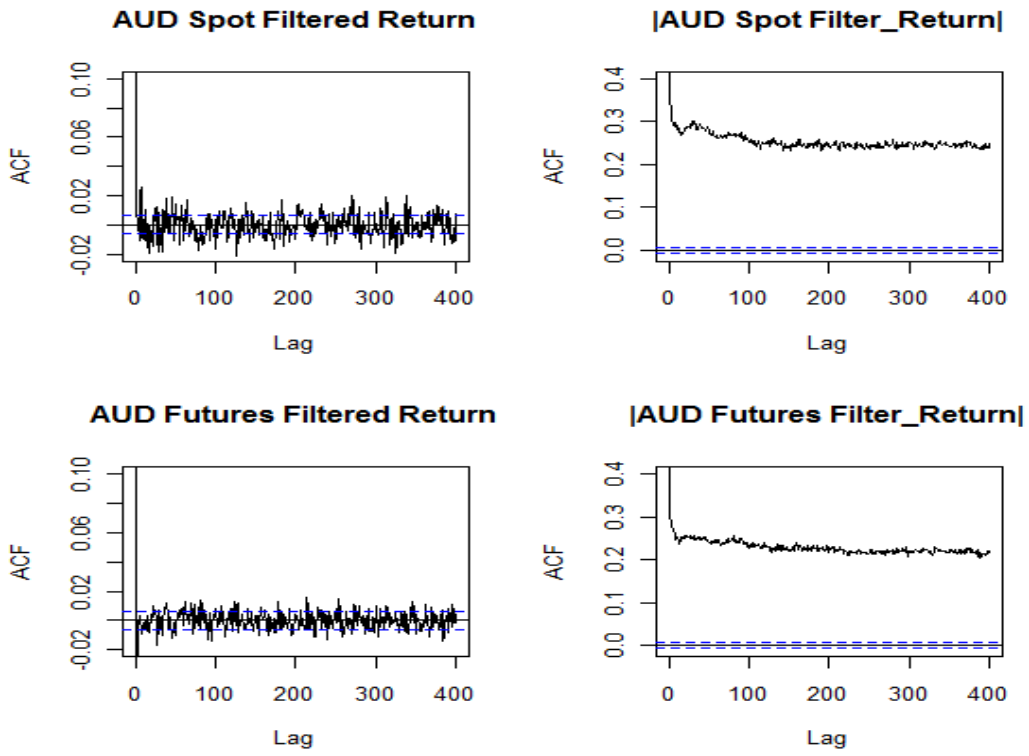
The detailed information about some summary statistics for returns over the temporally aggregated periods is displayed in Table 2-1. For the Australian dollar, the resulting time interval is 2 hours. Figure 2-5 plots the sample autocorrelations for the returns and absolute returns for two representative AUD spot and futures. It is important to note that absolute returns for both AUD spot and futures exhibit strong daily seasonality in their sample autocorrelations, where it is frequently observed in intra-day high frequency data.

Figure 2-5 Autocorrelations of AUD spot and futures returns



When intra-day seasonality is displayed in absolute returns, it is called seasonality in volatility. For both spot and futures raw returns, they are presented by the strong periodicity at the daily frequency, at the lag of 80. It is well known that the seasonality is described by the U-shape. The volatility of the opening is high and is followed by a gradual decrease, reaching the lowest level at half a trading day (i.e., 40 lags) and then an increase in volatility before closing. It is very important to get rid of the intra-day seasonal volatility. In the literature, some studies have shown that the existence of the seasonality would cause trouble in both estimating volatility models and building up models (i.e., misspecification issue) presented by Guillaume et al, (1995) and Andersen and Bollerslev (1997). In particular, Andersen and Bollerslev (1997) document that the intra-day periodicity is closely related to the conditional volatility. Without taking good care of those intraday periodicities in absolute returns, it may mislead the long-run pattern and it is likely to exaggerate the volatility and may be followed by distortion of the hedge ratios. To treat those intrinsic properties in the return volatilities, we use the wavelet transform to filter out the daily seasonality (see the section 2.3 for the details of the wavelet transform). Figure 2-6 shows the sample autocorrelations for absolute returns from the filtered return series. As shown in Figure 2-6, the wavelet filter is very effective in removing the daily seasonality preserved in absolute returns.

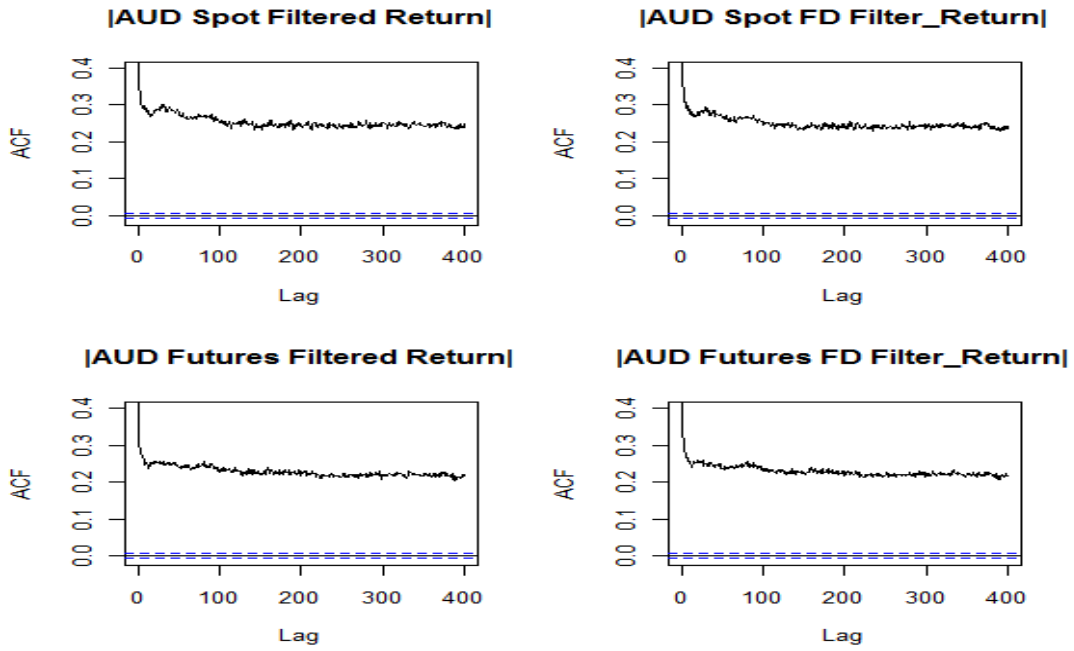
Figure 2-6 Autocorrelations of AUD spot and futures filtered returns



The sample autocorrelations for the returns are largely within the reference line although for some lagged values, the autocorrelations are occasionally out of bound. Note that the autocorrelations of the absolute returns show marked persistence, where it is not the case of the usual exponential decay. This is well known as a long memory process with hyperbolic decay. It means that the autocorrelation function decays slowly to zero at a hyperbolic rate as the lag increases. When d is the order of fractional integration, the long memory parameters of volatility measures for the AUD spot and futures series are estimated as $\hat{d}_s$ = 0.188 and $\hat{d}_f$ = 0.179 using Geweke and Porter-Hudak's semi-parametric method. Based on this result, we take the fractional differences of the two series. However, there is no difference between before- and after-treatment for visual

inspection. That is, the autocorrelations for the absolute returns are unchanged in magnitude as in the Figure 2-7.

Figure 2-7 Comparison of the autocorrelations for AUD absolute returns



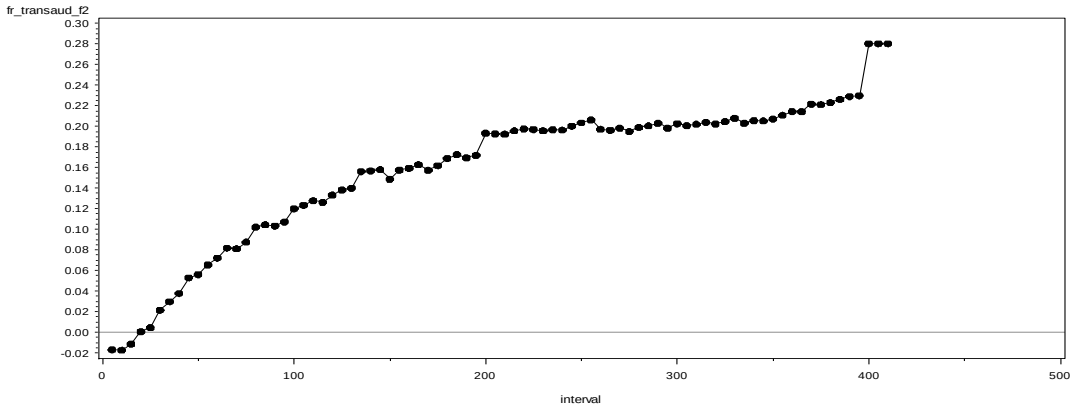
One possible explanation of this result is that a long memory process is found in the absolute returns (i.e., a volatility measure) rather than the returns. Even though the series are fractionally differenced, it does not have the effect of a long memory process for the absolute returns. The fact can be captured by a fractionally integrated GARCH (FIGARCH) model, which is known for representing long memory in volatility. In this study we do not employ a double long memory model. We focus on the long memory in volatility due to the above observation.

It is very important to note that the study of hedging effectiveness relies heavily on the correlation behavior between a pair of financial assets. Since correlation coefficients

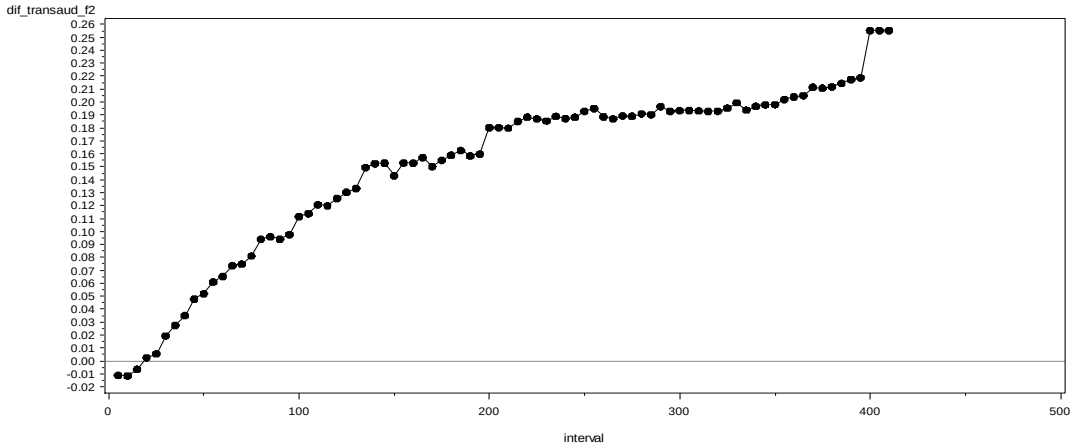
at high frequency intervals (e.g., 5-minute or 10-minute intervals) are close to zero, for the purpose of our investigation the filtered return series should be temporally aggregated. To make them smooth, we apply an exponentially weighted moving average filter. The linear correlation coefficients calculated for the AUD filtered spot and futures over the time interval of returns are displayed in Figure 2-8. For comparison purpose, the unfiltered data sets (without the use of the wavelet filter) are considered.

Figure 2-8 Pearson correlation coefficients for AUD spot and futures returns

(a) Filtered series



(b) Unfiltered series



As noted by Epps (1979) and Guillaume et al., (1997), we can see that there is decline in correlation coefficient at very high frequencies for both cases. As time interval increases, the correlations rise gradually. There is no perfect guideline about the choice of the time interval. We choose the 24^{th} point from the first time interval, which is equivalent to 2 hours. When we check for a long-run relationship between levels for the AUD spot and futures series using the Engle-Granger (1987) method, we find a long-run equilibrium relationship in the levels in both the filtered and unfiltered cases.

For the temporally aggregated data, we examine the sample autocorrelation function (ACF) of the filtered return series and the sample partial autocorrelation function (PACF) of the squared returns to determine the specification form of the mean equations. As shown in Figure 2-9, the two ACFs in the first row reflect some serial correlation. The two series mimic a variant of the AR(2) model where its characteristic roots are involved with complex numbers. There is a problem with using the AR(2) model. When we remove any linear dependence by adopting the AR(2) model, correlation between the associated residuals gets smaller so that the resulting hedge ratio turns to a negative number. To avoid this outcome, an MA-type error model is used. The two PACFs of the squared returns on the bottom show some linear dependence. It appears that there are GARCH effects using the PACFs, so the GARCH(1,1) model may be appropriate. In the case of the unfiltered series as shown Figure 2-10, all features look similar to those in Figure 2-9. Due to the long memory in volatility, the FIGARCH model can be preferred in this study.

For the specifications of the mean equations, we construct an MA model that includes the error correction term (ECT). The treatments in mean equations are considered in the same way for both the filtered and unfiltered cases. The use of the ECT can be justified by the Engle-Granger cointegration test result. For example, in the case of AUD spot and futures levels, the subsequent Augmented Dickey-Fuller (ADF) test statistic is -16.9144 (for the unfiltered case, -12.2468), which is big enough in absolute value to reject the null hypothesis of no cointegration. Since the autocorrelations over a few lags are observed, MA terms can be included. For the AUD spot filtered return series, MA(1) can be used and MA(2) is used for the AUD futures series. This means that we intentionally allow some serial correlation. The reason that we can do so is that the OLS estimator will be consistent as long as $X'u/T$ converges in probability to zero, in the OLS estimator, $\hat{\beta} = \beta + (X'X)^{-1}X'u$.

Figure 2-9 Sample ACF of returns and PACF of squared returns (filtered)

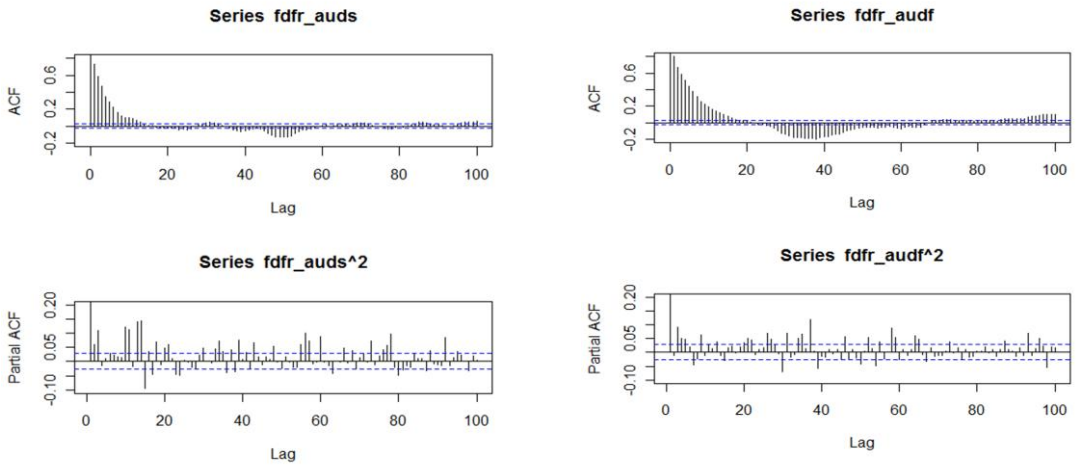


Figure 2-10 Sample ACF of returns and PACF of squared returns (unfiltered)

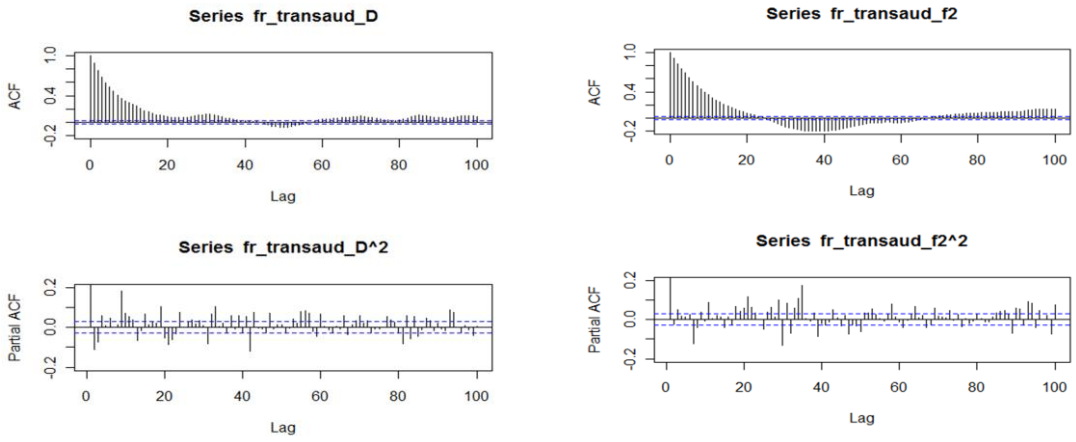


Table 2-2 reports the parameter estimates for the mean equations with the filtered and unfiltered cases. For either case, the estimation result is not very different in the magnitudes of the estimated coefficients and their significances. This outcome may occur because the filtered and unfiltered data sets contain a common linear filter, the exponentially weighted moving average. The existence of the linear filter does not significantly discriminate the difference in the filtered and unfiltered series.

Table 2-2 Estimates for mean equations with two different cases

	MA(1) (Spot)		MA(2) (Futures)	
Param	Unfiltered	Filtered	Unfiltered	Filtered
const	-0.003** (0.001)	-0.003** (0.001)	0.004** (0.001)	0.005** (0.001)
ma1	0.701** (0.008)	0.699** (0.008)	0.974** (0.013)	0.988** (0.013)
ma2	NA	NA	0.520** (0.010)	0.521** (0.010)
ect	0.003** (0.001)	0.035** (0.001)	0.001 (0.002)	0.002 (0.002)
AIC	-19108.99	-19607.77	-17825.21	-18377.07

Note: The maximum likelihood estimates are reported. The numbers in parentheses are standard errors. ** denotes significance at 5%.

Table 2-3 presents the parameter estimates from the CCC-GARCH models for the AUD spot and futures filtered and unfiltered returns. Most estimated parameters are significant at the 5% level except the correlation parameter and the constants in the unfiltered result and the correlation estimate in the filter-based model. The sum over the GARCH and ARCH parameters is almost one for both models, which indicates that the models are suited with data. Since all gradients for the estimated parameters are approaching zeros, the optimization algorithm (Newton method) seems to have attained a global maximum point. In light of the hedge ratios, although the significances of the parameters are improved in the filtered return-based model, the resultant hedge ratio (i.e., 0.0074) gets smaller compared to the hedge ratio (i.e., 0.0107) in the unfiltered case. According to the Sharpe hedge ratio[25], the filtered case (i.e., 0.0754) is larger than the unfiltered case (i.e., 0.0622). The orders in the magnitudes of the hedge ratios between the filtered and unfiltered cases are reversed compared to the minimum variance (for short, MV) hedge ratio framework. Hereafter, the mean values of the hedge ratios across time are reported.

As a benchmark, the traditional constant hedge ratio measured by the OLS estimator is 0.0925 and 0.0869 for the filtered and unfiltered returns, respectively. Under the Sharpe hedge ratio framework, the OLS-based hedge ratio based on the filtered

[25] In deriving the Sharpe hedge ratio, the risk-free interest rate is used. In this study, the 3-month Treasury bill (secondary market rate) with monthly frequency is chosen. For simplicity, the average (i.e., a constant) of the monthly values is calculated. The computed risk-free interest rate is 0.0282.

returns is 0.0866 while one for the unfiltered returns is 0.0629. The calculated hedge ratios under the Sharpe measure are usually smaller than those in the minimum variance measure. The Sharpe measure produces larger hedge ratio than the minimum variance framework. It is likely for the OLS-based hedge ratios to be larger than those in other GARCH models because the conditional variance in volatility processes crucially relies on the unexplained portion of the mean equations (i.e., residuals). As long as the mean equations are correctly specified, the hedge ratios based on the conditional variance may be smaller in magnitude than the OLS-based hedge ratios. Note that the AIC values are not directly comparable across the filtered and unfiltered data sets because the two different data sets are used.

Table 2-3 Estimates for CCC-GARCH(1,1) model with two different cases

	CCC-GARCH (unfiltered)		CCC-GARCH (filtered)	
Param	Spot	Futures	Spot	Futures
a0	0.280 (0.183)	0.240 (0.295)	0.540**(0.059)	0.392**(0.077)
a1	0.168*(0.087)	0.100**(0.044)	0.250**(0.020)	0.114**(0.012)
b1	0.815**(0.089)	0.890**(0.060)	0.732**(0.018)	0.876 **(0.013)
cor	0.013 (0.237)		0.009* (0.005)	
AIC	-25869.004		-27530.060	
Nobs	4985		4985	
h_ratio	0.0107		0.0074	
Sharpe	0.0622		0.0754	

Note: The maximum likelihood estimates are reported. The numbers in parentheses are standard errors. ** denotes significance at 5%, and * denotes significance at 10%. The AIC values are calculated as the log-likelihood value minus the number of parameters since GAUSS routine is set to minimize the negative of the log-likelihood function. Note that the AIC values are not directly comparable across the filtered and unfiltered data sets because the two different data sets are used. H_ratio denotes the minimum variance hedge ratio and Sharpe represents the Sharpe hedge ratio.

When the constant and time-varying correlation coefficients of the CCC-GARCH (or CCC-FIGARCH) models and DCC-GARCH (or DCC-FIGARCH) models that follow

are estimated, the optimal hedge ratios can be computed by either $\frac{h_{sf,t}}{h_{f,t}^2} = \rho_t \frac{h_{s,t}}{h_{f,t}}$ or the formula in Equation 2-24 attached by the time subscript for the correlation coefficient. Using the derived hedge ratios, we can use the reduction in variance of the hedged portfolio to evaluate hedging effectiveness.

Figure 2-10 display the dynamics of the hedge ratio evolution over time based on the CCC-GARCH models for two different data sets. Moreover, two different hedge ratio measures are included. The CCC-GARCH model with a constant correlation allows the hedge ratio to be time varying due to the evolution of the conditional variance matrix.

Figure 2-10 Hedge ratio evolution for CCC-GARCH Models

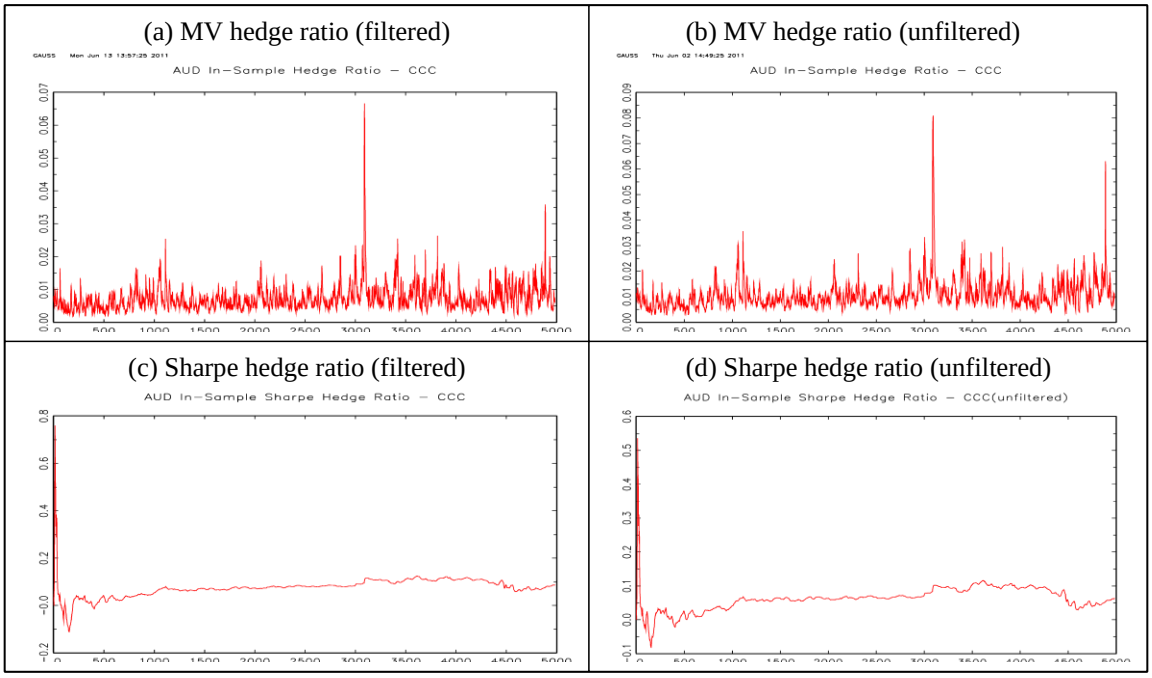


Table 2-4 report the results for the parameter estimates of the bivariate DCC-GARCH(1,1) models for the two different data sets. Regardless of the data sets, the

associated parameter estimates are statistically significant at the 5% level. Relative to the MV hedge ratios based on the CCC-GARCH model, those on the DCC-GARCH model become larger. In the case of the Sharpe measure, their magnitudes in hedge ratios reduce a little compared to those in the CCC-GARCH model. In terms of the values of the hedge ratios, the Sharpe measure produces larger hedge ratio than the minimum variance framework for the filtered and unfiltered cases.

Table 2-4 Estimates for DCC-GARCH(1,1) model with two different cases

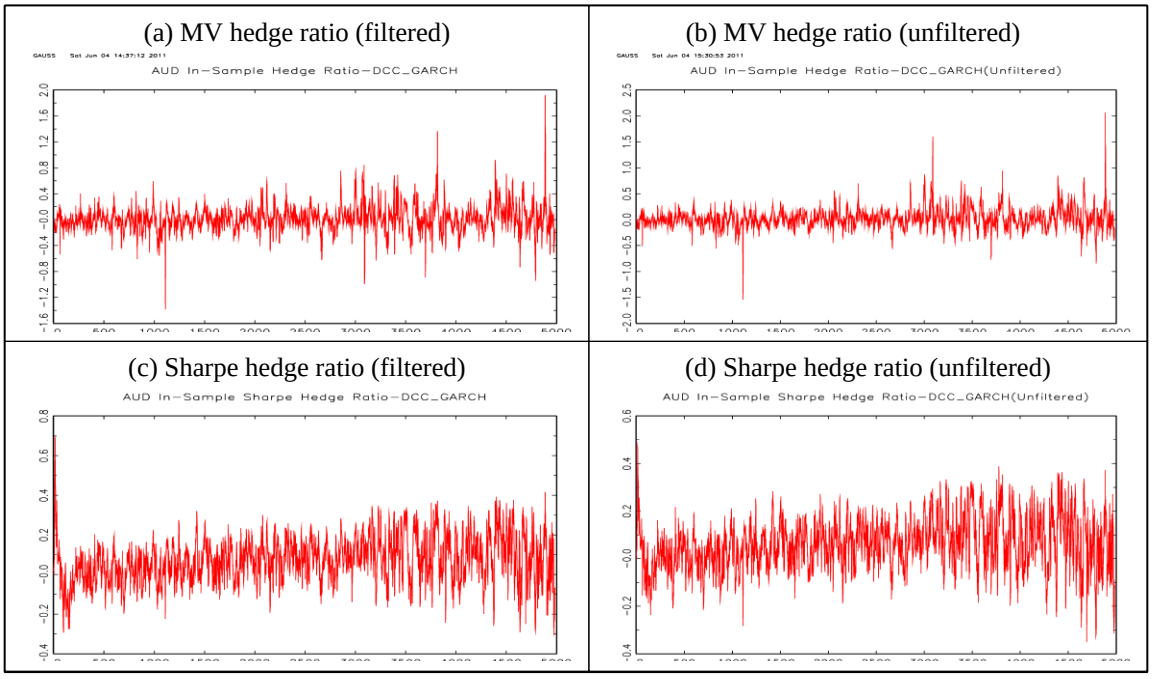
	DCC-GARCH (unfiltered)		DCC-GARCH (filtered)	
Param	Spot	Futures	Spot	Futures
a0	0.164**(0.030)	0.153**(0.037)	0.230**(0.026)	0.171**(0.042)
a1	0.148**(0.017)	0.095**(0.012)	0.236**(0.019)	0.112**(0.015)
b1	0.837**(0.019)	0.896**(0.013)	0.746**(0.018)	0.877**(0.017)
alpha	0.116**(0.013)		0.121**(0.013)	
beta	0.739**(0.034)		0.740**(0.032)	
AIC	-23956.208		-23443.032	
Nobs	4985		4985	
h_ratio	0.0197		0.0120	
Sharpe	0.0613		0.0735	

Note: The maximum likelihood estimates are reported. The numbers in parentheses are standard errors. ** denotes significance at 5%, and * denotes significance at 10%. The AIC values are calculated as the log-likelihood value minus the number of parameters since GAUSS routine is set to minimize the negative of the log-likelihood function. Note that the AIC values are not directly comparable across the filtered and unfiltered data sets because the two different data sets are used. H_ratio denotes the minimum variance hedge ratio and Sharpe represents the Sharpe hedge ratio.

Figure 2-11 presents the time varying hedge ratios for the DCC-GARCH models of the filtered and unfiltered data sets. Unlike the CCC-GARCH model, in the DCC-GARCH, the correlation coefficients are updated over time. For the AUD spot and futures series the resulting hedge ratios are 0.0120 and 0.0197 for the filtered and unfiltered data, respectively. Compared to the hedge ratios in the CCC-GARCH models,

the hedge ratios from the DCC-GARCH models become larger overall for both data sets. This result is compatible with the fact that the DCC-GARCH model is designed to describe the time varying correlation structure intrinsic to the financial data. If the correlations vary with time, the hedge ratios should be updated on the basis of more recent information. On the other hand, for the Sharpe measure, the hedge ratios (0.0735 and 0.0613) based on the DCC-GARCH model get smaller than those from the CCC-GARCH model.

Figure 2-11 Hedge ratio evolution for DCC-GARCH Models



Since long memory in the volatility process exists, a fractionally integrated GARCH (FIGARCH) model can be used for capturing the persistent volatility. In the CCC-FIGARCH and DCC-FIGARCH estimation procedures we need to set a truncation

value (i.e., 1000 as suggested by Baillie, Bollerslev, and Mikkelsen (1996)) given the infinite ARCH presentation is used in Equation 2-85.

Table 2-5 reports estimates of the parameters for the bivariate CCC-FIGARCH(1,1) models under the two different data sets. Although there are subtle differences in magnitudes for the estimated parameters, their signs and statistical significances are identical meaning that both models fit well to the data sets. The sufficient conditions for the positive definiteness of the conditional variances at every time index such as Equation 2-86 are met with the estimated parameters.

Table 2-5 Estimates for CCC-FIGARCH(1,1) model with two different cases

	CCC-FIGARCH (unfiltered)		CCC-FIGARCH (filtered)	
Param	Spot	Futures	Spot	Futures
a0	1.097** (0.162)	1.078** (0.215)	0.506** (0.081)	0.638** (0.120)
b1	-0.408** (0.032)	0.233** (0.067)	-0.338** (0.041)	0.154* (0.087)
d	0.449** (0.026)	0.384** (0.036)	0.485** (0.029)	0.407** (0.036)
phi	-0.754** (0.019)	-0.112** (0.047)	-0.704** (0.027)	-0.241** (0.065)
cor	0.013 (0.010)		0.010 (0.010)	
AIC	-26608.999		-23297.247	
Nobs	4985		4985	
h_ratio	0.0104		0.0084	
Sharpe	0.0621		0.0760	

Note: The maximum likelihood estimates are reported. The numbers in parentheses are standard errors. ** denotes significance at 5%, and * denotes significance at 10%. The AIC values are calculated as the log-likelihood value minus the number of parameters since GAUSS routine is set to minimize the negative of the log-likelihood function. Note that the AIC values are not directly comparable across the filtered and unfiltered data sets because the two different data sets are used. H_ratio denotes the minimum variance hedge ratio and Sharpe represents the Sharpe hedge ratio.

The significant estimates of the long memory parameter indicate that there exists long memory in volatility. Compared with the CCC-GARCH(1,1) models for the two data sets, the CCC-FIGARCH models produce better or similar hedge ratios. The mean values

of the hedge ratios are 0.0084 and 0.0104, which are greater than or nearly equal to those for the filtered and unfiltered data sets, respectively. For the Sharpe measure, the estimated hedge ratio increases over the CCC-GARCH model in the filtered returns. For the unfiltered case, the hedge ratio decreases over the CCC-FIGARCH. The Sharpe measure produces larger hedge ratios than the MV framework.

Table 2-6 shows the parameter estimates of the bivariate DCC-FIGARCH(1,1) models in two different data sets. Without any exceptions, all parameters are statistically significant at the 5% level for both cases.

Table 2-6 Estimates for DCC-FIGARCH(1,1) model with two different cases

	DCC-FIGARCH (unfiltered)		DCC-FIGARCH (filtered)	
Param	Spot	Futures	Spot	Futures
a0	0.569**(0.092)	0.591**(0.100)	0.467**(0.081)	0.620**(0.100)
b1	-0.434**(0.033)	0.248**(0.043)	-0.361**(0.039)	0.159**(0.051)
d	0.420**(0.025)	0.372**(0.028)	0.457**(0.027)	0.393**(0.027)
phi	-0.746**(0.020)	-0.076**(0.029)	-0.695**(0.027)	-0.214**(0.044)
alpha	0.114**(0.012)		0.112**(0.012)	
beta	0.748**(0.030)		0.761**(0.029)	
AIC	-23667.317		-23194.111	
Nobs	4985		4985	
h_ratio	0.0177		0.0118	
Sharpe	0.0608		0.0741	

Note: The maximum likelihood estimates are reported. The numbers in parentheses are standard errors. ** denotes significance at 5%. The AIC values are calculated as the log-likelihood value minus the number of parameters since GAUSS routine is set to minimize the negative of the log-likelihood function. Note that the AIC values are not directly comparable across the filtered and unfiltered data sets because the two different data sets are used. H_ratio denotes the minimum variance hedge ratio and Sharpe represents the Sharpe hedge ratio.

The value of d is an estimate of the long memory parameter in volatility. They indicate that the volatility measures (i.e., absolute returns in this study) are long memory

and stationary processes since those values are located between zero and one. As in the CCC-FIGARCH model, the conditions for the conditional variance matrix to be positive are satisfied. The hedge ratios are 0.0118 and 0.0177 for the filtered and unfiltered case, respectively. They are similar to those in the DCC-GARCH models. With the Sharpe measure, the hedge ratio for the filtered case rises while it drops in the unfiltered case. The values are high relative to the MV measure.

Figures 2-12 and 2-13 represent the time varying hedge ratios based on both the CCC- and DCC-FIGARCH(1,1) models. It is very interesting to ask whether the proposed model specifications do good jobs to describe the given data sets. The Akaike Information Criterion (AIC) may be used to select a parametric model among several fitted models. The lower the AIC, better the model. In the GARCH-based models, the CCC-GARCH(1,1) model has the lowest AIC value for the filtered data. For the unfiltered data set, the CCC-FIGARCH(1,1) would be selected according to the AIC.

Figure 2-12 MV Hedge ratio evolution for DCC-FIGARCH Models

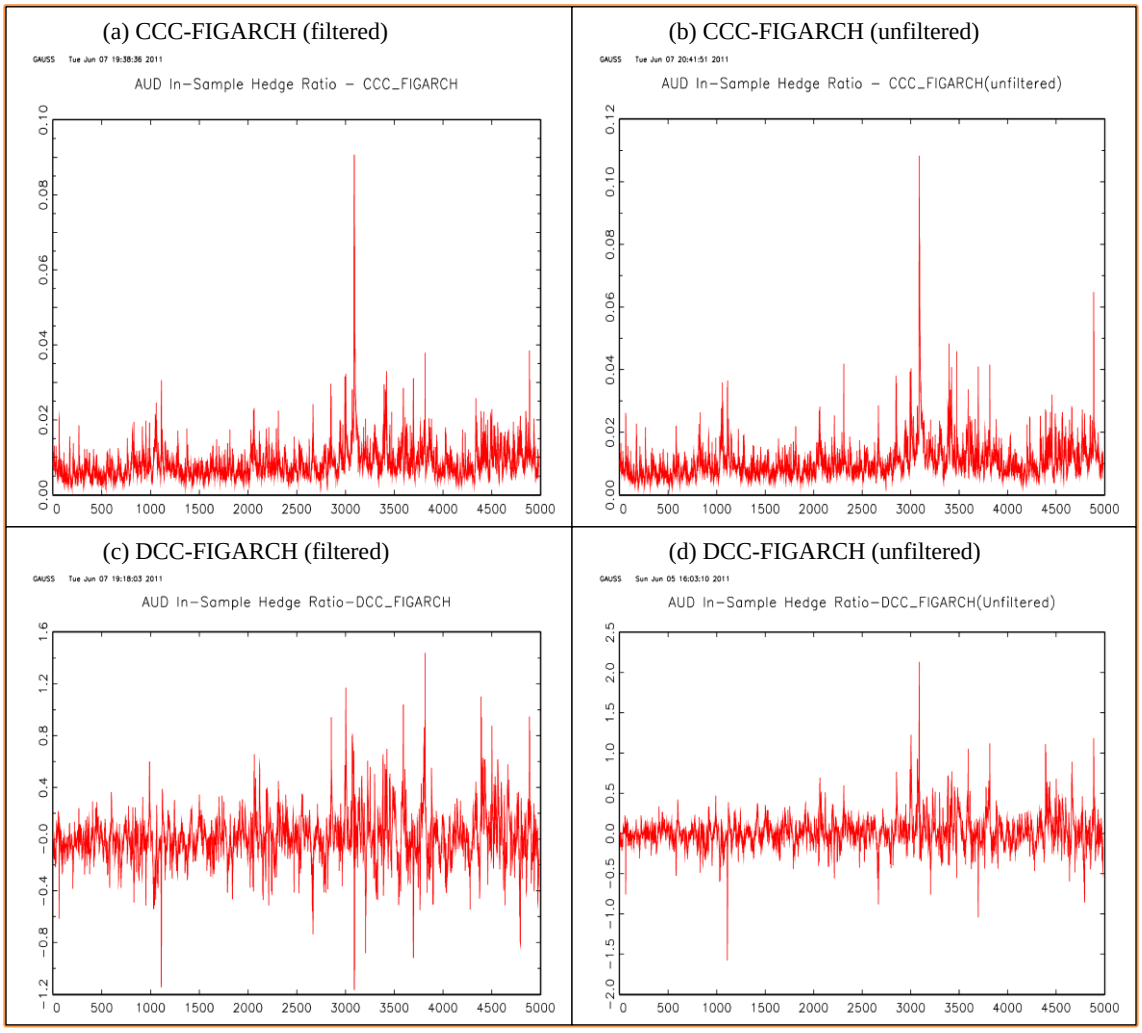
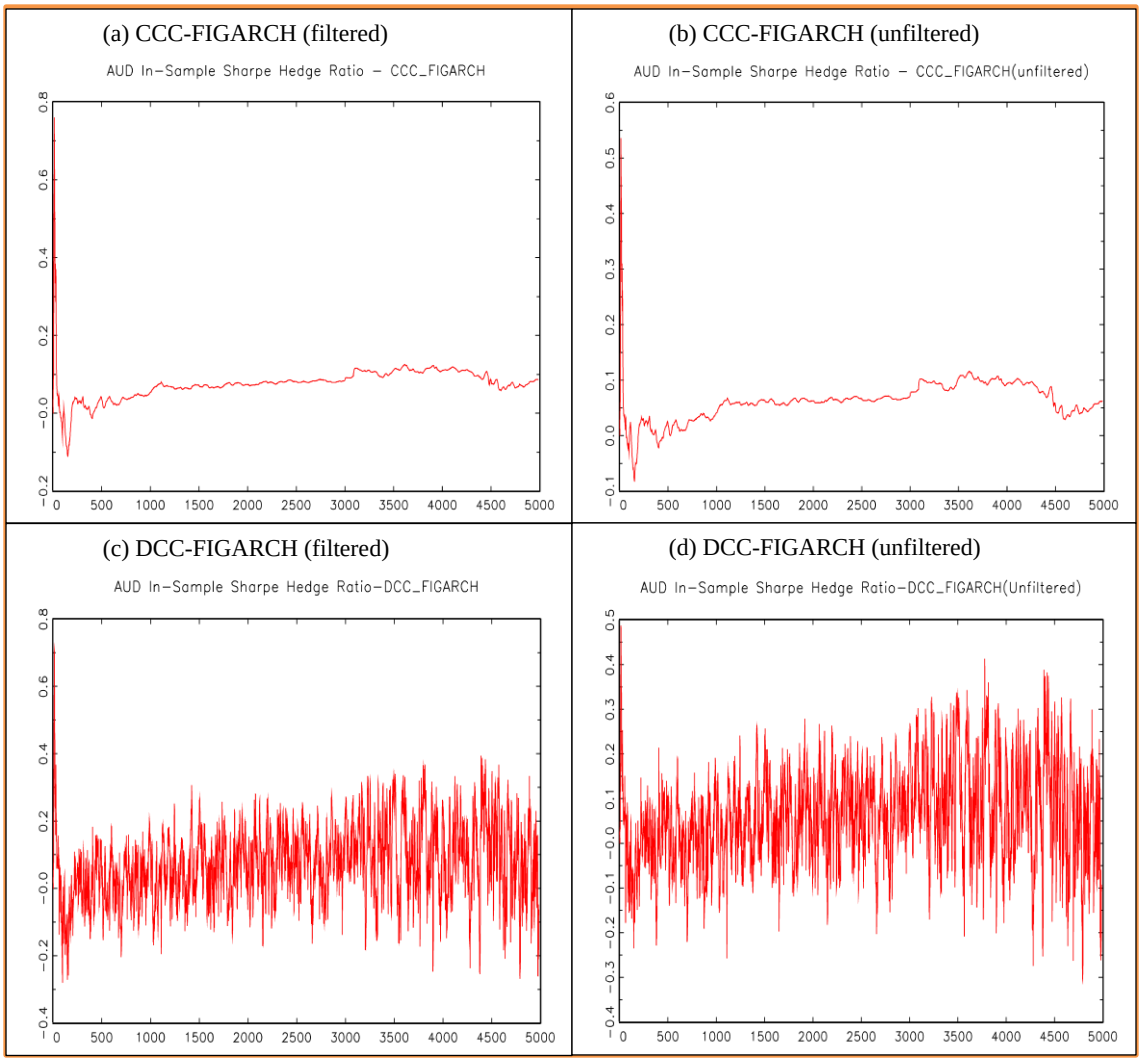


Figure 2-13 Sharpe Hedge ratio evolution for DCC-FIGARCH Models



According to the visual inspection of the plots of the filtered and unfiltered returns used in the marginal distributions, the Student-t distribution may be used for FIGARCH innovations as in Figure 2-14 and 2-15. All histograms show significant positive kurtosis (i.e., high peak) and symmetry in the distribution of data, which is also confirmed in the quantile plots (or, Q-Q plots). The two right panels of each figure display the presence of heavy-tailedness in both sides.

Figure 2-14 Histograms and Q-Q plots for AUD filtered returns

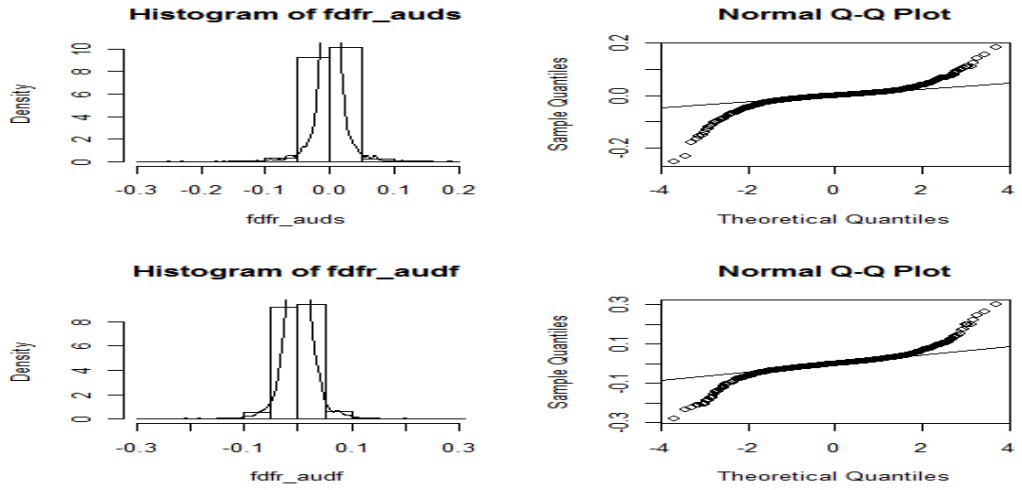
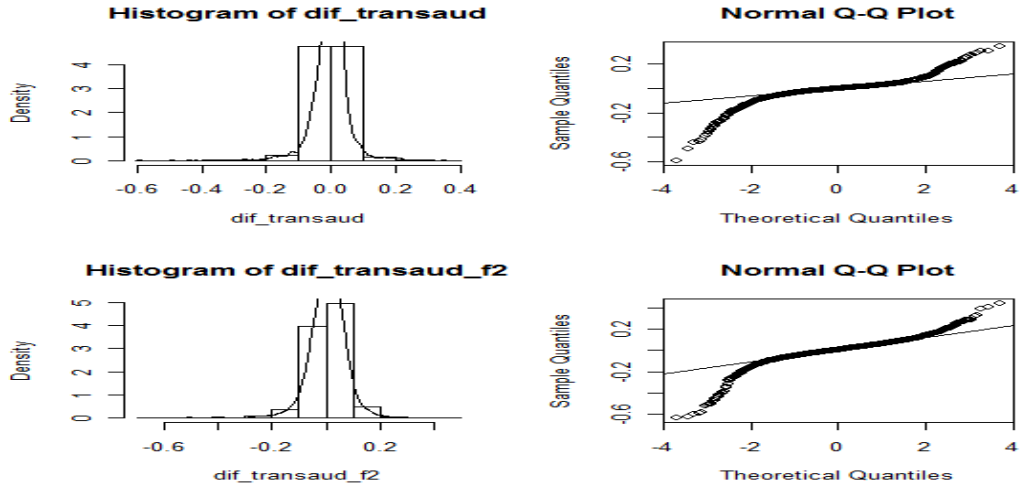


Figure 2-15 Histograms and Q-Q plots for AUD unfiltered returns



With regard to marginal distributions for copula models, Table 2-6 presents the estimated parameters for the FIGARCH(1,1) model with the Student-t innovations. The estimates for the parameters of the conditional mean equations are contained as well. All MA terms for the conditional mean equations are statistically significant, whereas the error correction terms are statistically insignificant. The error correction term is regarded as an equilibrium error that occurred in the previous period. If the spot return at the previous period is above its equilibrium level, the relevant coefficients need to be negative to restore equilibrium. Our results for the error correction coefficients are compatible with the intuition. By this process, the equilibrium error will be corrected in the next period. All parameter estimates for the FIGARCH models are statistically significant. The corresponding gradients are extremely small and appear to reflect a global maximum point on the log-likelihood function.

Regarding the statistical inference, we use Quasi Maximum Likelihood (QML) estimators, which are robust over potential model misspecification errors in the spirit of Bollerslev and Wooldridge (1992). In practice, the estimation procedure can be conducted under the normal distribution-based log-likelihood function although the errors are described by some non-normal distribution. Since we assume that the innovations that pass though the conditional mean questions follow the Student-t distributions, Gaussian QML method can be used for inference. In the context of FIGARCH, the asymptotic properties are discussed in Baillie, Bollerslev, and Mikkelsen (1996) and Bollerslev, and Mikkelsen (1996). Both papers show that the consistency and asymptotic normality of the QML estimator are satisfied within the FIGARCH(1,d,0) framework. Under the Monte

Carlo simulation studies such as Baillie, Bollerslev, and Mikkelsen (1996) and Lombardi and Gallo (2002), the resulting QML estimates for the FIGARCH(1,d,1) model exhibit consistency and asymptotic normality.

Table 2-7 Estimates for FIGARCH(1,1) with Student-t innovations

	FIGARCH (unfiltered)		FIGARCH (filtered)	
Param	Spot	Futures	Spot	Futures
const	-0.003**(0.001)	0.004**(0.001)	-0.003**(0.001)	0.005**(0.001)
ma1	0.701**(0.008)	0.974**(0.013)	0.699**(0.008)	0.988**(0.013)
ma2	NA	0.520**(0.010)	NA	0.521**(0.010)
coint	0.003**(0.001)	0.001 (0.002)	0.004**(0.001)	0.002 (0.002)
a0	0.695**(0.090)	0.637**(0.120)	0.570**(0.081)	0.594**(0.117)
b1	-0.228**(0.034)	0.279**(0.006)	-0.231**(0.039)	0.213**(0.059)
d	0.657**(0.030)	0.500**(0.043)	0.632**(0.030)	0.489**(0.043)
phi	-0.758**(0.022)	-0.185**(0.043)	-0.736**(0.025)	-0.261**(0.045)
v	6.943**(0.468)	7.733**(0.716)	8.186**(0.656)	9.447**(1.020)
AIC	-10884.965	-12581.997	-10710.106	-12385.809
Nobs	4985		4985	

Note: The maximum likelihood estimates are reported. The numbers in parentheses are standard errors. ** denotes significance at 5%, and * denotes significance at 10%. The AIC values are calculated as the log-likelihood value minus the number of parameters since GAUSS routine is set to minimize the negative of the log-likelihood function. Note that the AIC values are not directly comparable across the filtered and unfiltered data sets because the two different data sets are used.

In order to find the dependence structures between the spot and futures exchange rates, we use five alternative copula specifications. The results from these estimations are presented in Tables 2-8 (filtered case) and 2-9 (unfiltered case). For the copula models, with the conditional variances obtained from different copula-based GARCH models, we attempt to calculate the off-diagonal element in the conditional variance matrix using the formula such as Equation 2-76. The double integral component in the formula can be numerically approximated using the bivariate conditional density function in Equation 2-

72. Then the time varying hedge ratios are calculated by the ratio of the conditional covariance to the conditional variance of the futures contract.

Similarly, for the Sharpe hedge ratio, the correlation coefficient in Equation 2-24 may be replaced by an approximated double integral component used in the MV hedge ratio calculation. Note the resultant bivariate copula density is a scalar although in some dynamic copula models the time varying dependence structure is involved with the calculation of the copula density itself. Due to the variability of the spot and futures returns, we can obtain the time varying hedge ratios.

Table 2-8 Parameter Estimates for Copula Models (filtered case)

Parameters	Normal	Static Student-t	Static Frank	Clayton	Gumbel
const	2.573** (0.306)			0.510** (0.069)	1.310** (0.180)
a	-8.492** (0.985)			-1.426** (0.168)	-1.265** (0.180)
b	0.477** (0.243)			-0.042 (0.094)	0.130 (0.121)
nu		12.65** (3.699)			
rho		0.016 (0.020)	0.109** (0.041)		
AIC	205.21	4.599	-0.490	41.67	60.53
h_ratio	0.071	0.007	0.0067	0.0147	0.0153
Sharpe	0.117	0.076	0.0751	0.0803	0.0807

Note: The numbers in parentheses are standard errors. ** and * represent significance at 5% and 10%. Rho indicates static dependence parameter for the static copula functions. Other parameters are based on the time varying correlation structure. H_ratio denotes the minimum variance hedge ratio and Sharpe represents the Sharpe hedge ratio.

As shown in Tables 2-8 and 2-9, the Student-t and the Frank copulas are based on static analysis rather than the time varying structure. In the Normal copula for the filtered and unfiltered data sets, the autoregressive parameters b are statistically significant. This means that some change in the dependence measure may be persistent at the subsequent lags. The fact can affect the estimated hedge ratio in a positive way. By contrast, for Clayton and Gumbel copulas, the autoregressive parameters are not statistically significant at 5% (even 10%) for the filtered and unfiltered cases. The parameter, a, for the mean absolute difference between two uniform variables are inversely related to the dependence measures as noted by Patton (2006). This implies that recent information on returns may negatively impact the next information formation except for the Frank copula in both Table 2-8 and Table 2-9.

Table 2-9 Parameter Estimates for Copula Models (unfiltered case)

Parameters	Normal	Static Student-t	Static Frank	Clayton	Gumbel
const	2.458** (0.310)			1.197** (0.250)	1.384** (0.193)
a	-8.027** (0.995)			-2.936** (0.641)	-1.323** (0.206)
b	0.642** (0.247)			-0.305 (0.236)	0.054 (0.131)
nu		6.62** (1.063)			
rho		0.024 (0.021)	0.178** (0.074)		
AIC	250.18	18.89	0.003	76.75	55.74
h_ratio	0.057	0.005	0.004	0.011	0.010
Sharpe	0.092	0.059	0.058	0.062	0.062

Note: The numbers in parentheses are standard errors. ** represent significance at 5%. Rho indicates static dependence parameter for the static copula functions. Other parameters are based on the time varying correlation structure. H_ratio denotes the minimum variance hedge ratio and Sharpe represents the Sharpe hedge ratio.

Through the entire analysis, two different approaches for the optimal hedge ratio are chosen as reported in Table 2-8 and Table 2-9. The difference in the two approaches consists in the objective function to be optimized. The minimum variance (MV) hedge ratio minimizes the variance of the hedged portfolio. The Sharpe hedge ratio (one of the mean-variance frameworks) can be obtained by maximizing the ratio of the hedged portfolio's excess return to the volatility. It turns out that the Sharpe hedge ratios are usually larger than the MV hedge ratios. That finding is valid and predictable because by definition the Sharpe measures are based on the spot and futures returns while most MV measures except the OLS-based hedge ratio are based on the standardized innovations through various GARCH filters.

For the purpose of copula model comparison, we can choose the model with the smallest AIC over the set of copula models under consideration. We pick the Frank copula for either the filtered or unfiltered case since it provides us with the smallest AIC value in each case. For both cases, the Student-t copulas seem to fit well to the data sets too.

We evaluate the in-sample hedging effectiveness of the various copula-based GARCH models using the criterion of hedging performance presented in Equation 2-89. It tells us how the variance of the hedged portfolio based on the estimated hedge ratios reduces. The estimated hedge ratios for the relevant models are calculated by the mean value of the associated hedge ratios for each copula function. Those hedge ratios and the hedging effectiveness for the OLS, CCC, DCC, Normal copula, Clayton copula, Gumbel copula, static Frank copula, and static Student-t copulas are reported in Table 2-10.

Table 2-10 Comparisons of the Hedging Effectiveness (MV measure)

	Filtered case			Unfiltered case		
Models	H_ratio	variance	Reduction (%)	H_ratio	Variance	Reduction (%)
OLS	0.0925	0.00239		0.0869	0.00264	
CCC-G	0.0074	0.00243	-1.51	0.0107	0.00267	-1.22
CCC-FI	0.0084	0.00243	-1.47	0.0104	0.00267	-1.23
DCC-G	0.0120	0.00243	-1.35	0.0197	0.00267	-0.95
DCC-FI	0.0118	0.00243	-1.36	0.0177	0.00267	-1.01
Normal	0.0708	0.00240	-0.99	0.0567	0.00264	-0.19
Student-t	0.0073	0.00243	-1.51	0.0050	0.00268	-1.40
Clayton	0.0147	0.00242	-1.26	0.0106	0.00267	-1.22
Frank	0.0067	0.00243	-1.53	0.0042	0.00268	-1.43
Gumbel	0.0153	0.00242	-1.24	0.0103	0.00267	-1.23

Note: The benchmark method is OLS hedge ratio for comparison. The variance reduction is calculated as $(var(s-h_{ols}f) - var(s-h_{alternative}f))/ var(s-h_{ols}f)$. The negative value means that the variance of an alternative hedge ratio-based portfolio increases.

As noticed in Table 2-10, the variance reductions over alternative models compared to the OLS hedge ratio are all negatives, which mean that the variance of an alternative hedge ratio-based portfolio increases. In contrast, compared to the unhedged variance (i.e., 0.002436), for the filtered case, it turns out that all models used in this study show better performance. For the unfiltered case, all models listed in Table 2-10 show improved performance compared to the unhedged case whose variance is 0.002682. One possible explanation for this result of poor performance for some copula models over the OLS-based hedge ratios is that when we look into the reported hedge ratios, the extent to which they swing relative to the OLS-based hedge ratio is large. Such a big swing in hedge ratios may cause the sample variance of the hedged portfolio to be bigger. As a result, the reported variance reduction values are negative. Overall decreases in the hedge ratios for the different models imply that the alternative models show lower correlation coefficient values between the standardized innovations through the volatility models.

Table 2-11 Comparisons of the Hedging Effectiveness (Sharpe measure)

	Filtered case			Unfiltered case		
Models	sharpe	variance	Reduction(%)	sharpe	Variance	Reduction(%)
OLS	0.0866	0.002393		0.0629	0.002643	
CCC-G	0.0754	0.002394	-0.054	0.0622	0.002643	-0.0072
CCC-FI	0.0760	0.002394	-0.050	0.0621	0.002643	-0.0083
DCC-G	0.0735	0.002395	-0.069	0.0613	0.002643	-0.0168
DCC-FI	0.0741	0.002394	-0.064	0.0608	0.002643	-0.0223
Normal	0.1167	0.002396	-0.117	0.0921	0.002639	0.1165
Student-t	0.0755	0.002394	-0.054	0.0587	0.002644	-0.0465
Clayton	0.0803	0.002393	-0.024	0.0624	0.002643	-0.0051
Frank	0.0751	0.002394	-0.056	0.0582	0.002644	-0.0525
Gumbel	0.0807	0.002393	-0.022	0.0621	0.002643	-0.0083

Note: The benchmark method is OLS hedge ratio for comparison. The variance reduction is calculated as $(var(s-h_{ols}f) - var(s-h_{alternative}f))/ var(s-h_{ols}f)$. The negative value means that the variance of an alternative hedge ratio-based portfolio increases. The column sharpe denotes the sharpe hedge ratio.

Table 2-11 reports the estimated hedge ratios based on the Sharpe measure. It is important to note that most hedging model specifications do not outperform the OLS-based hedging strategy except the Normal copula. However, relative to the unhedged cases (those variances are 0.002436 and 0.002682, for the filtered and unfiltered case, respectively), we can benefit from any of models used in this study in light of the variance reduction criterion. Compared to the minimum variance (MV) hedge ratios, the Sharpe hedge ratios give us better performance in a sense that they generate overall lower variances over all model specifications.

Figure 2-16 and Figure 2-17 display the dynamics of the estimated MV hedge ratios from the various copulas. Each graph displays how the hedge ratios changed over a period of time. Every graph shows a similar pattern, where the hedge ratios started fluctuating at the points around the observation 4500 on the horizontal axis, equivalent of

September and October 2008. In that period of time, there was the bankruptcy of Lehman, which led to damaging financial panic in which most financial institutions and investment banks suffered large losses. In terms of hedging, when the economy is in turmoil, hedgers tend to raise the hedge ratio to protect themselves from a slack and downturn market. Every copula model in both Figure 2-16 and Figure 2-17 well explains the point that since September and October 2008, the hedgers should use higher hedge ratios to reduce the potential risk of a hedge position relative to those before September 2008.

The time varying hedge ratios based on the alternative derivation (the Sharpe measure) are presented in Figure 2-18 and Figure 2-19. The reason that all graphs in the two figures look similar is that their time-varying hedge ratios mainly depend on the observations in every time period with the dependence structure fixed. The dependence structure is obtained from the approximated double integral component in Equation 2-76.

In the literature, it is well known that there are positive relations between time horizons and hedging effectiveness. For example, as the hedging horizon increases, the hedging effectiveness increases as shown in Chen, Lee, and Shrestha (2004). Similarly, when the hedging horizon is short (e.g., 2-hour time stamp), the resulting hedging effectiveness would decrease. When this point is combined with lower correlation value within the high frequency time interval, we can expect that the hedging effectiveness that follows would be smaller than otherwise.

Based on data showing extremely low correlation in magnitude, the proposed copula and the dynamic GARCH models reveal poor performance in hedging

effectiveness relative to the OLS-based method. Both the filtered and unfiltered cases show similar results. This outcome may be expected because the correlation coefficients between the spot and futures returns is so small that it can be hard to recognize the well-known stylized facts in the hedging literature. In addition, the correlation measure is at the margin to the degree which the constituent assets do not move together. This causes the hedge ratios to be frequently turned over to negative values.

Figure 2-16 Dynamics of MV hedge ratios for copula models (filtered case)

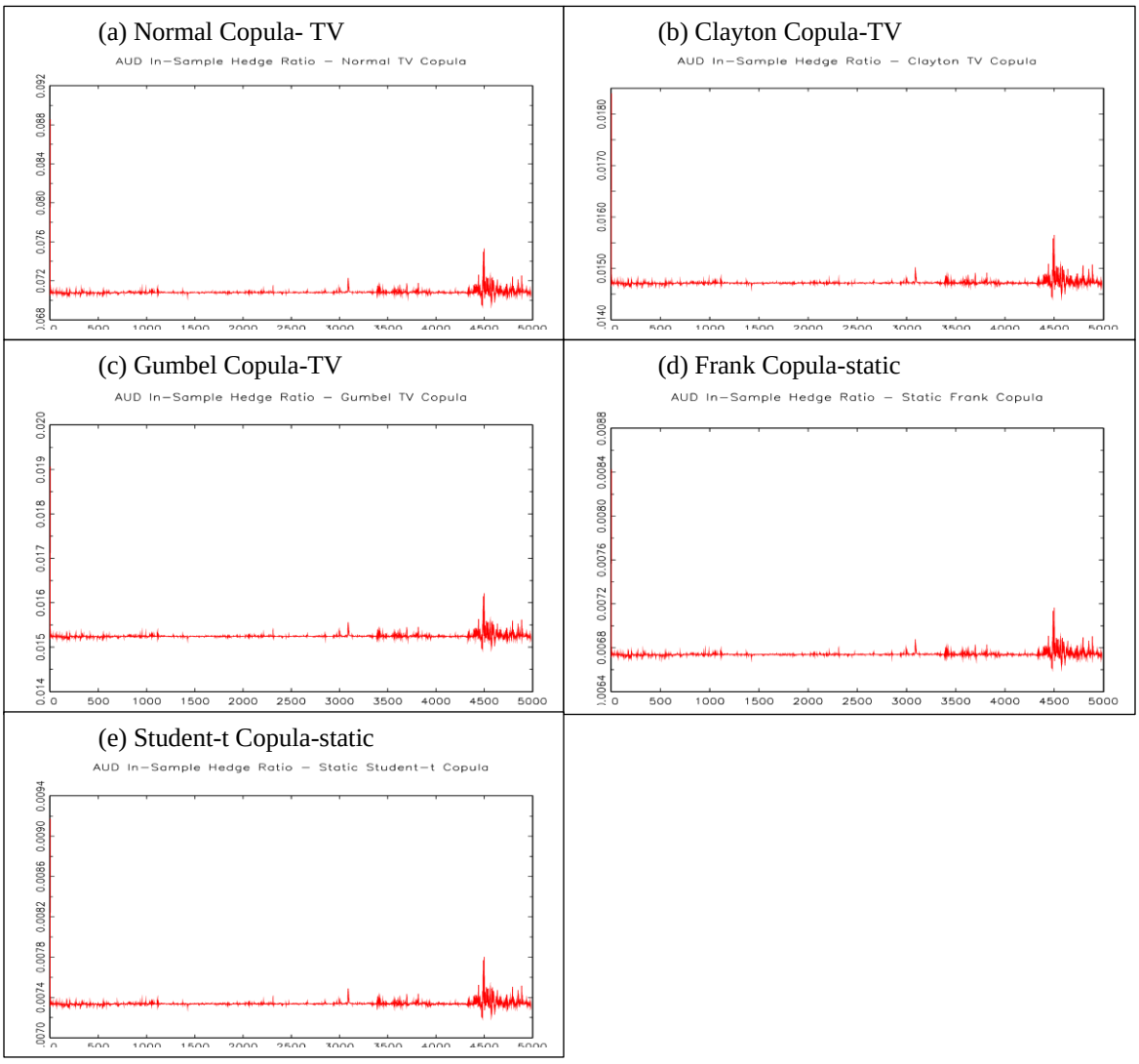
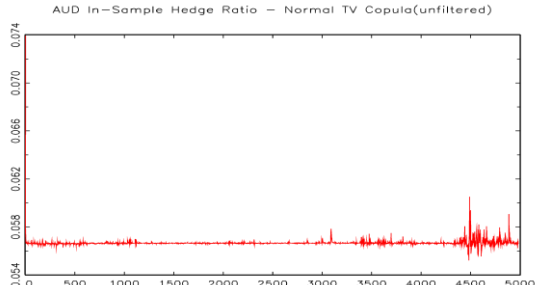
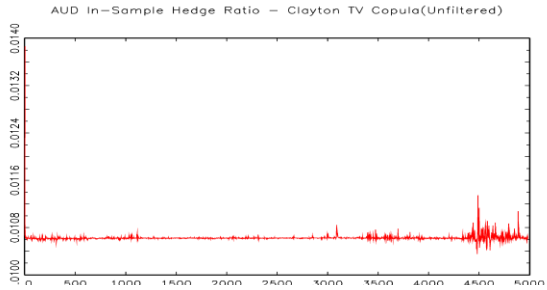


Figure 2-17 Dynamics of MV hedge ratios for copula models (unfiltered case)

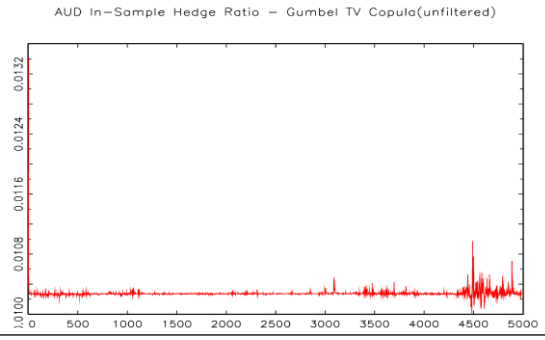
(a) Normal Copula- TV



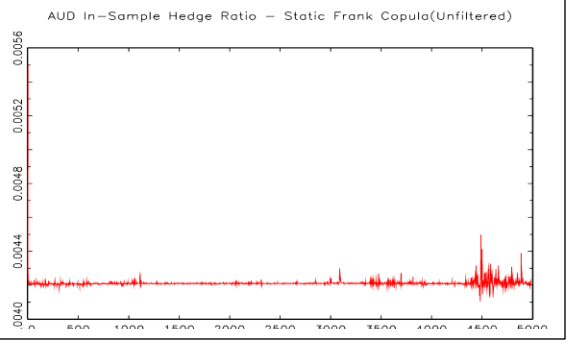
(b) Clayton Copula-TV



(c) Gumbel Copula-TV



(d) Frank Copula-static



(e) Student-t Copula-static

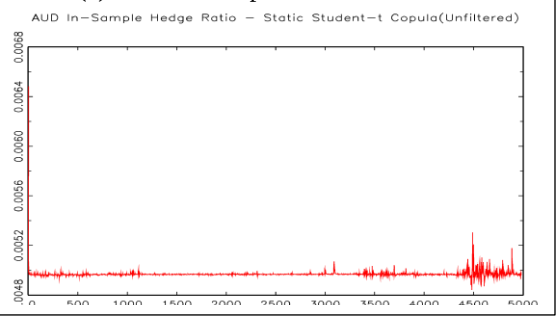


Figure 2-18 Dynamics of Sharpe hedge ratios for copula models (filtered case)

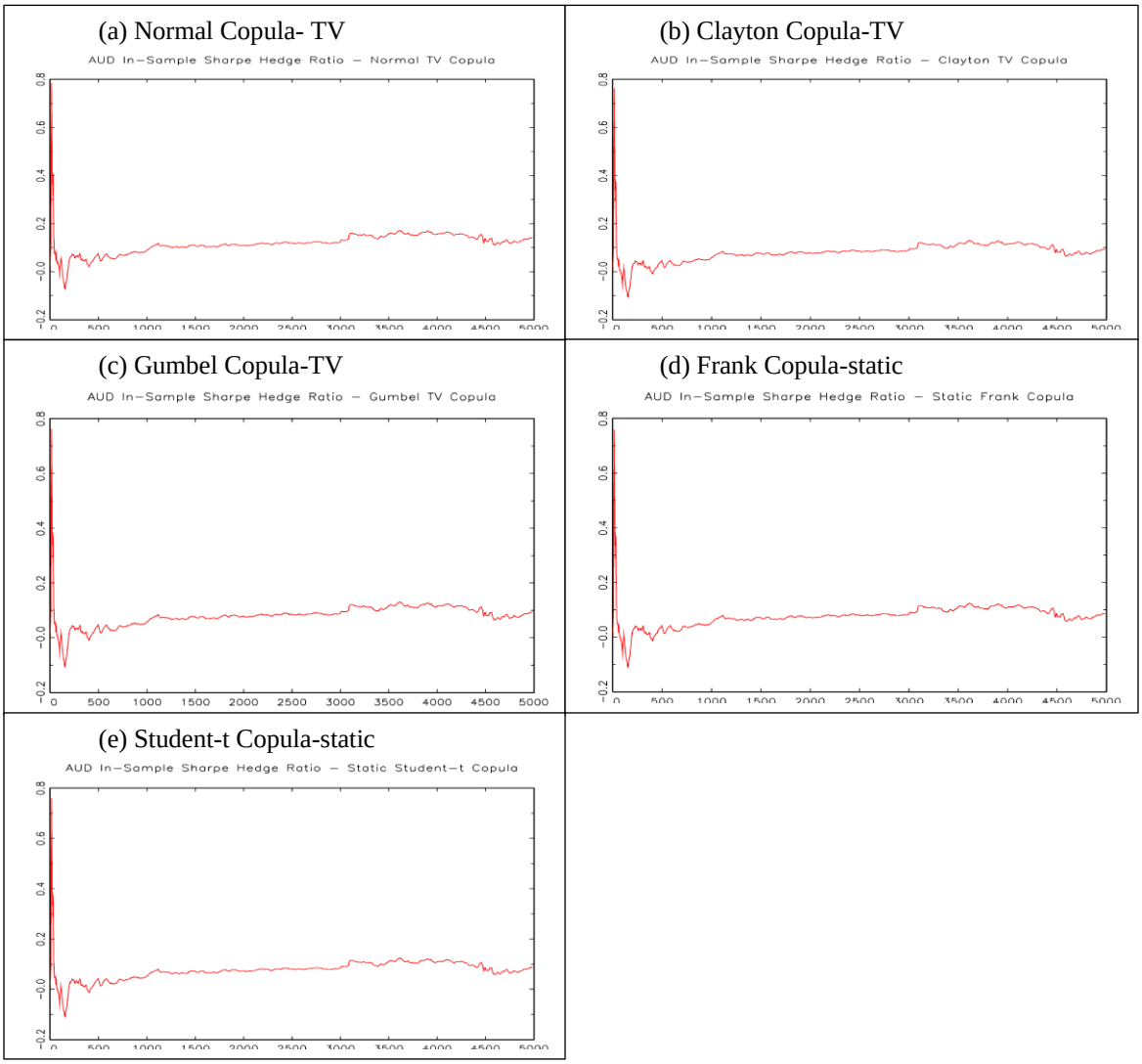
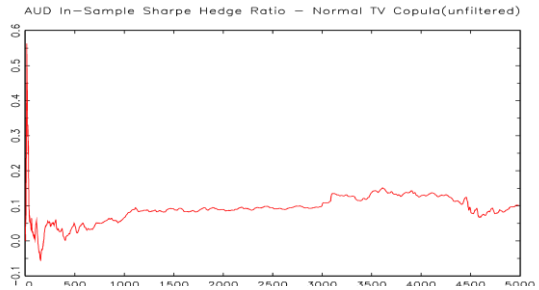
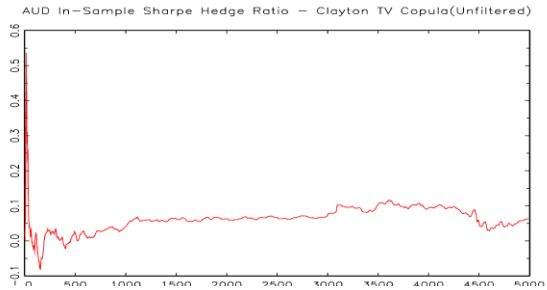


Figure 2-19 Dynamics of Sharpe hedge ratios for copula models (unfiltered case)

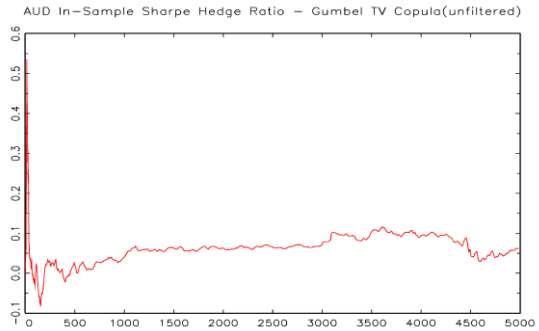
(a) Normal Copula- TV



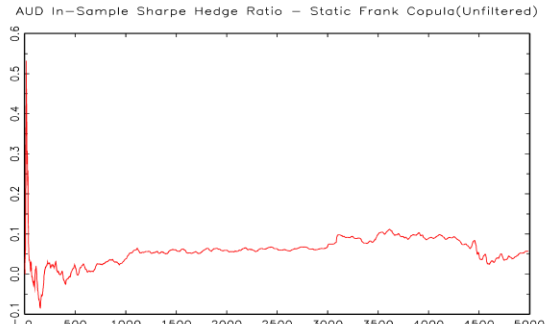
(b) Clayton Copula-TV



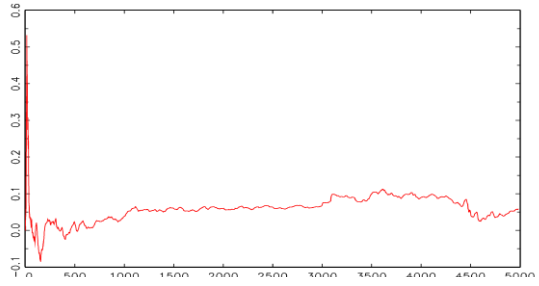
(c) Gumbel Copula-TV



(d) Frank Copula-static



(e) Student-t Copula-static



2.7 Conclusion

The purpose of the study is to evaluate the performance of copula models under different applications. In order to do that, we choose 2-hour time interval data (i.e., one of the intra-day frameworks). High frequency data give us an opportunity of a better understanding of how the relevant asset prices move over time within the daily time frame. The main objective is to investigate how effectively various statistical models can reduce the potential risk of a hedged position.

In the data handling, we use wavelet analysis to get rid of the noisy components in the original data, which are 5-minute time intervals. Due to the need for temporal aggregation, we would lose some useful information in the process. To alleviate this problem, we set the aggregation to a 2-hour time stamp, where the resulting number of observations is 4985. With various statistical models, we calculate various hedge ratios according to their models which are followed by comparing hedging effectiveness of AUD currency spot and futures contracts. Furthermore we employ two different approaches to calculating the optimal hedge ratio. The minimum variance hedge ratio are calculated and compared with the Sharpe hedge ratios.

By the variance reduction criterion for hedging effectiveness, when using some sophisticated models such as the dynamic GARCH (or FIGARCH) models, they do not produce an improvement in hedging effectiveness compared to the OLS-based MV hedge ratio. For the Sharpe measure, only the Normal copula shows some variance reduction of the hedged portfolio.

References

Alexander, Gordon J., and Alexandre M. Baptista. 2004. A Comparison of VaR and CVaR Constraints on Portfolio Selection with the Mean-Variance Model. *Management Science* 50, no.9 (September): 1261-1273.

Andersen, Torben G., and Tim Bollerslev. 1997. Intraday Periodicity and Volatility Persistence in Financial Markets. *Journal of Empirical Finance 4*, no.2-3 (June): 115-158.

Andersen, Torben G., and Tim Bollerslev. 1997. Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns. *Journal of Finance* 52, no.3 (July): 975-1005.

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys. 2001. The Distribution of Realized Exchange Rate Volatility. *Journal of the American Statistical Association* 96, no.453(March): 42-55.

Baillie, Richard T., and Tim Bollerslev. 1991. Intra-Day and Inter-Market Volatility in Foreign Exchange Rates. *Review of Economic Studies* 58, no.3 (May):565-585.

Baillie, Richard T., Tim Bollerslev, and Hans Ole Mikkelsen. 1996. Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 74, no.1 (September): 3-30.

Baillie Richard T., and Robert J. Myers. 1991. Bivariate GARCH Estimation of the Optimal Commodity Futures Hedge. *Journal of Applied Econometrics* 6, no.2 (April-June), 109-124.

Barnett, V., and T. Lewis. 1994. *Outliers in Statistical Data*. 3rd ed. Chichester, UK: John Wiley & Sons.

Bartram, Söhnke M., Stephen J. Taylor and Yaw-Huei Wang. 2007. The Euro and European Financial Market Integration. *Journal of Banking and Finance* 31, no.5:1461-1481.

Bauwens, Luc, Sebastien Laurent, and Jeroen V.K. Rombouts. 2006. Multivariate GARCH Models: A Survey, *Journal of Applied Econometrics* 21: 79-109.

Beran, Jan. 1994. *Statistics for Long-Memory Processes*. New York:Chapman & Hall.

Bollerslev, Tim, and Hans Ole Mikkelsen. 1996. Modeling and Pricing Long Memory in Stock Market Volatility. *Journal of Econometrics* 73, no.1 (July): 151-184.

Bollerslev, Tim, and Jeffrey M. Wooldridge. 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews* 11, no.2:143 - 172.

Bollerslev, Tim. 1986. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 31, no.3 (April): 307-327.

Bollerslev, Tim. 1990. Modeling the Coherence in Short-Run nominal Exchange Rates: a Multivariate Generalized ARCH Model. *Review of Economics and Statistics* 72, no.3 (Aug): 498-505.

Breymann, W., A. Dias, and Paul Embrechts. 2003. Dependence Structures for Multivariate High-Frequency Data in Finance. *Quantitative Finance* 3. 1-16.

Brooks, Chris, Olan T. Henry, and Gita Persand. 2002. The Effect of Asymmetries on Optimal Hedge Ratios. Journal of Business 75, no.2: 333-352.

Brunetti, Celso and Christopher L. Gilbert. 2000. Bivariate FIGARCH and Fractional Cointegration. *Journal of Empirical Finance* 7, no.5 (Dec): 509-530.

Byström, H.N.E. 2003. The hedging performance of electricity futures on the Nordic power exchange. *Applied Economics* 35, no.1 (January): 1-11.

Cecchetti, Stephen G., Robert E. Cumby, and Stephen Figlewski. 1988. Estimation of the Optimal Futures Hedge. *Review of Economics and Statistics* 70, no.4 (Nov): 623-630.

Chen, Sheng-Syan, Cheng-Few Lee, and Keshab Shrestha. 2004. An Empirical Analysis of the Relationship between the Hedge Ratio and Hedging Horizon: A simultaneous Estimation of the Short- and Long-run Hedge Ratios. *Journal of Futures Markets* 24, no.4 (April): 359-386.

Chen, Sheng-Syan, Cheng-Few Lee, and Keshab Shrestha. 2008. Do the Pure Martingale and Joint Normality Hypotheses Hold for Futures Contracts? Implications for the Optimal Hedge Ratios. *The Quarterly Review of Economics and Finance* 48, no.1 (February): 153-174.

Cherubini, Umberto, Elisa Luciano, and Walter Vecchiato. 2004. *Copula Methods in Finance.* West Susses, UK:John Wiley & Sons.

Dacorogna, Michael M., Ulrich A. Müller, Robert J. Nagler, Richard B. Olsena and Olivier V. Pictet. 1993. A geographical model for the daily and weekly

seasonal volatility in the foreign exchange market. *Journal of International Money and Finance* 12, no.4 (August):413-438.

Dacorogna, Michel M., Ramazan Gençay, Ulrich Müller, Richard B. Olsen, and Olivier V. Pictet. 2001. *An Introduction to High-Frequency Finance*. London: Academic Press.

Davidson, Russell and James G. Mackinnon. 1993. *Estimation and Inference in Econometrics.* New York:Oxford University Press.

Ding, Zhuanxin, Clive W.J. Granger, and Robert F. Engle. 1993. A Long Memory Property of Stock Market Returns and a New Model. *Journal of Empirical Finance* 1, no.1 (June): 83-106.

Embrechts, Paul A, Alexander J. McNeil, and D. Straumann. 2002. Correlation and Dependence Properties in Risk Management: Properties and Pitfalls, in M. Dempster, ed. *Risk Management: Value at Risk and Beyond*. Cambridge University Press.

Engle, Robert F. 1982. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* 50, no.4 (July): 987-1007.

Engle, Robert F. 2002. Dynamic Conditional Correlation-a simple class of multivariate GARCH models. *Journal of Business and Economic Statistics* 20: 339-350.

Engle, Robert F. and Sheppard, Kevin Keith. 2001. Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH. NBER Working Paper (October), no.W8554.

Engle, Robert F. and C.W.J. Granger. 1987. Cointegration and Error Correction: Representations, Estimation and Testing. *Econometrica* 55, no.2 (March): 252-276.

Engle, Robert F. and Kenneth F. Kroner. 1995. Multivariate Simultaneous Generalized Arch. *Econometric Theory* 11, no. 1(March):122-150.

Engle, Robert F. and Kevin Sheppard. 2001. Theoretical and Empirical properties of Dynamic Conditional Correlation Multivariate GARCH. *NBER Working Paper* no.8554 (October).

Epps, Thomas W. 1979. Comovements in Stock Prices in the Very Short Run. *Journal of the American Statistical Association* 74, no.366(June): 291-298.

Floros, Christos, and Dimitrios V. Vougas. 2004. Hedge ratios in Greek stock index futures market. *Applied Financial Economics* 14, no.15 (October): 1125-1136.

Gençay, Ramazan., Faruk Selçuk, and Brandon Whitcher. 2001a. Scaling properties of foreign exchange volatility. *Physica A* 289: 249-266.

Gençay, Ramazan., Faruk Selçuk, and Brandon Whitcher. 2001b. Differentiating intraday seasonalities through wavelet multi-scaling. *Physica A* 289: 543-556.

Gençay, Ramazan., Faruk Selçuk, and Brandon Whitcher. 2002. *An Introduction to Wavelets and Other Filtering Methods in Finance and Economics*. San Diego: Academic Press.

Genest Christian, and Louise-Paul Rivest. 1993. Statistical Inference Procedures for Bivariate Archimedean Copulas. *Journal of the American Statistical Association* 88, no.423 (September): 1034-1043.

Geweke, John., Susan Porter-Hudak. 1983. The Estimation and Applications of Long Memory Time Series Models. *Journal of Time Series Analysis* 4, no.4(July):221-238.

Ghosh, Asim. 1993. Hedging with stock index futures: Estimation and forecasting with error correction model. *Journal of Futures Markets*, 13, no.7 (Oct), 743–752.

Ghosh, Asim. 1995. The Hedging Effectiveness of ECU Futures Contracts: Forecasting Evidence from an Error Correction Model. *Financial Review* 30, no.3 (August): 567-581.

Goodhart, Charles A.E., and Maureen O'Hara. 1997. High frequency data in financial markets: Issues and applications. *Journal of Empirical Finance* 4, no.2-3: 73-114.

Granger, CliveW.J., and Paul Newbold.1974. Spurious regressions in econometrics. *Journal of Econometrics* 2, no.2 (July): 111-120.

Granger C. W. J., and Roselyne Joyeux. 1980. An Introduction to Long-Memory Time Series Models and Fractional Differencing. *Journal of Time Series Analysis* 1, no.1 (January): 15-29.

Guillaume, Dominique M., Olivier V. Pictet and Michel M. Dacorogna. 1995. On the intra-daily performance of GARCH processes. O&A Research Group working paper.

Guillaume, Dominique M., Michel M. Dacorogna, Rakhal R. Dave, Ulrich A.Müller, Richard B. Olsen, and Olivier V. Pictet. 1997. From the bird's eye to the microscope: A survey of new stylized facts of the intra-daily foreign exchange markets. *Finance and Stochastics* 1, no.2 (April): 95 - 129.

Haar, A. 1910. Zur Theorie der orthogonalen Funktionensysteme. *Mathematische Annalen*, 69: 331-371.

Harris, Richard D. F., and Jian Shen. 2006. Hedging and Value at Risk. *Journal of Futures Markets* 26, no.4 (April): 369-390.

Herbst, A.F, Kare, D.D, and Caples, S.C. 1989. Hedging Effectiveness and Minimum Risk Hedge Ratios in the Presence of Autocorrelation: Foreign Currency Futures. *Journal of Futures Markets* 9, no.3 (June): 185–198.

Hodgson, Allan. 1996. Information transfer, microstructures and intraday price return spikes. Accounting and Finance 36, no 2(November): 229-257.

Holmes, Phil.1996. Stock Index Futures Hedging: Hedge Ratio Estimation, Duration Effects, Expiration Effects and Hedge Ratio Stability. *Journal of Business Finance & Accounting* 23, no.1 (January): 63-77.

Hosking, J. R. M. 1981. Fractional Differencing. *Biometrika* 68, no.1(April): 165-176.

Howard, Charles T. and Louis J. D'Antonio. 1984. A Risk-Return Measure of Hedging Effectiveness. 1984. *Journal of Financial and Quantitative Analysis* 19, no.1 (March): 101-112.

Hsu, Chih-Chiang, Chih-Ping Tseng, and Yaw-Huei Wang. 2008. Dynamic Hedging with Futures: A Copula-Based GARCH Model. *Journal of Futures Markets* 28, no.11: 1095-1116.

Hull, John C. 2006. *Options, Futures, and Other Derivatives*. 6th ed. New Jersey: Pearson Prentice Hall.

Hull, J. and A. White. 1998. Value-at-Risk When Daily Changes in Market Variables are not Normally Distributed. *Journal of Derivatives* 5:9-19.

Hurvich, Clifford M, Rohit Deo and Julia Brodsky.1998. The Mean Squared Error of Geweke and Porter-Hudak's Estimator of the Memory Parameter of a Long-Memory Time Series. *Journal of Time Series Analysis* 19, no.1 (January):19-46.

Joe, H. 1997. *Multivariate Models and Dependence Concepts*. London:Chapman and Hall.

Joe, Harry., and James J. Xu. 1996. The Estimation Method of Inference Functions for Margins for Multivariate Models. Technical Report, no.166, Department of Statistics, University of British Columbia.

Jondeau, Eric., Ser-Huang Poon, and Michael Rockinger. 2007. *Financial Modeling Under Non-Gaussian Distributions*. London:Springer-Verlag.

Jondeau, Eric., Michael Rockinger. 2006. The copula-GARCH model of conditional dependencies: An international stock market application. *Journal of International Money and Finance* 25, no.5 (August):827-853.

Kroner, Kenneth F., Jahangir Sultan. 1993. Time-Varying Distributions and Dynamic Hedging with Foreign Currency Futures. *Journal of Financial and Quantitative Analysis* 28, no.4 (December): 535-551.

Li, David X. 2001. On Default Correlation: A Copula Function Approach. *Journal of Fixed Income* 9, no.4 (March): 43-54.

Lien, D.; Y.K. Tse; and A.K.C. Tsui. 2002. Evaluating the Hedging Performance of the Constant-Correlation GARCH Model. *Applied Financial Economics* 12, no.11 (Nov): 791-798.

Lien, Da-Hsiand Donald. 1996. The Effect of the Cointegrating Relationship on Futures Hedging: A Note. *Journal of Futures Markets* 16, no.7 (Oct):773-780.

Lien, Donald. 2004. Cointegration and the optimal hedge ratio: the general case. *Quarterly review of Economics and Finance* 44, no.5 (Dec): 654-658.

Lo, Andrew.W. 1991. Long-Term Memory in Stock Market Prices. *Econometrica* 59, no.5(September): 1279-1313.

Lombardi, Marco J. and Giampiero M. Gallo. 2002. Analytic Hessian Matrices and the Computation of FIGARCH Estimates. *Statistical Methods & Applications* 11, no.2 (June): 247-264.

Lundin, Mark., Michel M. Dacorogna and Ulrich A. Müller. 1999. Correlation of High-Frequency Financial Time Series. In *Financial Markets Tick by Tick*, ed. Lequeux, 91-126. New York: John Wiley & Sons Ltd.

Lütkepohl, Helmut.2005. *New Introduction to Multiple Time Series Analysis*. Berlin:Springer-Verlag.

McNeil, Alexander J., Rüdiger Frey and Paul Embrechts. 2005. *Quantitative Risk Management: Concepts, Techniques and Tools.* New Jersey: Princeton University Press.

Moosa, Imad A. 2003. The sensitivity of optimal hedge ratio to model specification. *Finance Letters* 1, no.1: 15–20.

Müller, Ulrich A., Michel M. Dacorogna, Rakhal D. Davé, Richard B. Olsen, Olivier V. Pictet, Jacob E. von Weizsäcker. 1997. Volatilities of different time resolutions - Analyzing the dynamics of market components. *Journal of Empirical Finance* 4, no. 2-3(June): 213 - 239.

Muthuswamy, Jayaram., Audipto Sarkar, Aaron Low, and Eric Terry. 2001. Time Variation in the Correlation Structure of Exchange Rates: High-Frequency Analyses. *Journal of Futures Markets* 21, no.2: 127-144.

Myers, Robert. J..1991. Estimating time-varying optimal hedge ratios on futures markets, *Journal of Futures Markets* 11, no.1 (Feb): 39-53.

Nelsen, Roger B. 1999. *An Introduction to Copulas*. New York:Springer-Verlag.

Park, Hun Y. and Bera, Anil K. 1987. Interest rate volatility, basis, and heteroscedasticity in hedging mortgages, *American Real Estate and Urban Economics Association* 15, no.2 (summer): 79-97.

Park, Tae H. and Lorne N. Switzer. 1995. Bivariate GARCH Estimation of the Optimal Hedge Ratios or Stock Index Futures: A Note. *Journal of Futures Markets* 15, no.1(Feb): 61-67.

Patton, Andrew. 2006a. Modelling Asymmetric Exchange Rate Dependence. *International Economic Review* 47, no.2 (May):527-556.

Patton, Andrew. 2006b. Estimation of Multivariate Models for Time Series of Possibly Different Lengths. *Journal of Applied Econometrics* 21, no.2 (March):147-173.

Patton, Andrew J. 2009. Copula-Based Models for Financial Time Series. In *Handbook of Financial Time Series*. Ed. Andersen, T.G., Davis, R.A., Kreiß, J.-P., Mikosch, T. Springer Verlag.

Percival, Donald B, and Andrew T. Walden. 2000. *Wavelet Methods for Time Series Analysis*. New York: Cambridge University Press.

Prokhorov, Artem and Peter Schmidt. 2009. Likelihood-based estimation in a panel setting: Robustness, redundancy and validity of copulas. *Journal of Econometrics* 153, no.1 (November): 93-104.

Rossi, Eduardo and Claudio Zucca. 2002. Hedging interest rate risk with multivariate GARCH. *Applied Financial Economics* 12, no.4 (April):241-251.

Tsay, Ruey S. and Tiao, George C.1984. Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Nonstationary ARMA Models. *Journal of the American Statistical Association* 79, no.385 (March): 84-96.

Tse, Y.K., and Albert K.C. Tsui. 2002. A Multivariate Generalized Autoregressive Conditional Heteroscedasticity Model with Time-Varying Correlations. *Journal of Business & Economic Statistics* 20, no.3 (July):351-362.

Tukey, John W. 1977. *Exploratory Data Analysis*. Reading, Massachusetts: Addison-Wesley Publishing Company.

3 Copula Model Selection Based on Non-Nested Testing

3.1 Introduction

A complete understanding of economic activity is not possible, and there are a variety of interpretations of an observed economic phenomenon. In other words, there are different models that can represent a particular set of economic data. However, given the observed data to be analyzed, we must single out the best model by the use of model selection criteria.

In general, there are nested hypothesis tests and non-nested hypothesis tests for model selection. Suppose there are two candidate models. When one model is a special case of the other model, the model structure is called nested. A specific example comes from the generalized Cobb-Douglas model given by

$$lnY = \beta_0 + \beta_1 lnL + \beta_2 lnK + \beta_3 \left(\frac{1}{2} ln^2 L\right) + \beta_4 \left(\frac{1}{2} ln^2 K\right) + \beta_5 lnL lnK + u \quad (3\text{-}1)$$

In this case by imposing the following constraints, $\beta_3 = \beta_4 = \beta_5 = 0$, we can obtain the usual Cobb-Douglas model. The constrained model can be used as the null hypothesis in the F test. The null model can be obtained from the full model by imposing some restriction on the model parameters. Examples of such tests include the standard Wald, Lagrange Multiplier, and likelihood ratio tests of significance. In contrast, there are many cases where two or more competing models are not nested. Namely, we cannot derive one model from the other model by imposing restrictions. The modern literature for non-nested hypothesis testing began with Cox (1961, 1962).

Note that there is a subtle difference between model selection and non-nested hypothesis testing. Following Gourieroux and Monfort (1993) and Pesaran and Weeks (2001), model selection would be equally treated in the sense that the true model does not belong to either the null or alternative model. In contrast, the candidate non-nested hypothesis testing is based on an asymmetric situation where one of the models is assumed to be true (say, the null hypothesis) and a testing process is conducted to test the null hypothesis against an alternative hypothesis. As a result, in the case of non-nested hypothesis testing, rejection of the null hypothesis does not mean acceptance of any one of the alternative hypotheses. Occasionally, there would be a situation where all possible models are rejected.

Suppose we conduct a test for finding an appropriate model that is compatible with a specific data generating process (i.e., the true conditional probability model that follows). For simpler exposition, we deal with linear normal regression models. The relevant hypotheses are as follows:

$$H_f: y = X\theta + u_f, \quad u_f \sim N(0, \sigma^2) \tag{3-2}$$

$$H_g: y = Z\gamma + u_g, \quad u_g \sim N(0, \omega^2) \tag{3-3}$$

where X and Z are stochastic design matrices with dimensions $T \times k_x$, and $T \times k_z$, respectively and their ranks are k_x, and k_z. Suppose we are not able to write X as a linear relationship with Z and vice versa. In this case the models are non-nested. The parameters α and β are $\alpha = (\theta', \sigma^2)'$ under H_f and $\beta = (\gamma', \omega^2)'$ under H_g. Let $P_X = X(X'X)^{-1}X'$

and $P_Z = Z(Z'Z)^{-1}Z'$ be projection matrices, and we also define $M_x = I_T - P_x$ and $M_z = I_T - P_z$.

Under some regularity conditions, a solution, b^* to the OLS minimization problem with respect to β is called a pseudo true value of β. Therefore we can estimate the relevant parameters.

$$\hat{\alpha} = \begin{pmatrix} \hat{\theta} \\ \hat{\sigma}^2 \end{pmatrix} = \begin{pmatrix} (X'X)^{-1}X'y \\ \frac{1}{T}\left(M_x X\theta + M_x u_f\right)'\left(M_x X\theta + M_x u_f\right) \end{pmatrix} = \begin{pmatrix} (X'X)^{-1}X'y \\ \frac{1}{T} e_f{}' e_f \end{pmatrix} \tag{3-4}$$

where e_f is the residual vector of model H_f. The pseudo true value of β is

$$b^*(\alpha) = \begin{pmatrix} \gamma^* \\ \omega^{*2} \end{pmatrix} = \begin{pmatrix} (Z'Z)^{-1}Z'X\theta \\ \sigma^2 + \frac{1}{T}(X\theta - P_Z X\theta)'(X\theta - P_Z X\theta) \end{pmatrix} \tag{3-5}$$

where P_Z is $Z(Z'Z)^{-1}Z'$, and γ^* and ω^{*2} are based on Equations (1.4) and (1.5) on page 15 in Gourieroux and Monfort (1995a).

The maximum likelihood estimation methods have nice properties such as consistency, asymptotic normality, and efficiency under the strong assumption that the model of interest is correctly specified. Even when the model is not well specified, the maximum likelihood properties may hold under pseudo-likelihood theory. Assume that the true conditional probability model is in the null hypothesis, H_f. That is, the true probability density is $f_0(y|x) = f(y|x;\alpha_0)$. By construction, the alternative hypothesis, H_g, can be regarded as a misspecified model. In this situation, to estimate the parameters for H_g, we need to define pseudo-true values for the parameters. If the models under consideration are parametric, the pseudo-true values can be uniquely obtained by

minimizing a discrepancy measure between the two competing parametric models, namely a null model and a misspecified model. As a discrepancy measure, the Kullback-Leibler Information criterion (KLIC) is commonly used. The KLIC evaluated conditionally at $X = x$ between two densities f and g is given by

$$K(g(y|z;\beta)|f(y|x;\alpha_0)) = E_x E_{\alpha_0}\left(log\frac{f(y|x;\alpha_0)}{g(y|z;\beta)}\right) \quad (3\text{-}6)$$

In this case the pseudo maximum likelihood (PML) estimator is based on the log-likelihood function $L_g(\beta)$ associated with the regression model, Equation (3-2);

$$L_f(\alpha) = -\frac{T}{2}ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y - x\theta)'(y - x\theta) \quad (3\text{-}7)$$

$$L_g(\beta) = -\frac{T}{2}ln(2\pi\omega^2) - \frac{1}{2\omega^2}(y - z\gamma)'(y - z\gamma) \quad (3\text{-}8)$$

where $\alpha' = (\theta', \sigma^2)$ and $\beta' = (\gamma', \omega^2)$. If the true distribution belongs to the null hypothesis, $\hat{\alpha}_T$ is the usual maximum likelihood estimator of α_0. The PML estimator of β is derived by calculating $\max_\beta L_g(\beta)$ where the objective function is the abbreviated form equivalent to Equation 3-4 and $\hat{\beta}_T$ is the PML estimator. As the sample size increases, the PML estimator converges in probability to the asymptotic pseudo-true value, $b(\alpha_0)$ following Gourieroux and Monfort (1995a).

The likelihood ratio (LR) principle can be used for testing non-nested models. The generalized likelihood ratio test based on Equations 3-7 and 3-8 is given by

$$LR_T = 2\left(L_g(\hat{\beta}_T) - L_f(\hat{\alpha}_T)\right) \quad (3\text{-}9)$$

where $\hat{\alpha}_T$ is the usual maximum likelihood estimator and $\hat{\beta}_T$ is the pseudo maximum likelihood estimator. Under the null hypothesis H_f, as $T \to \infty$ the test statistic $(1/T)LR_T$ converges to

$$2E_x E_{\alpha_0}\left[L_g\left(b^*(\alpha_0)\right) - L_f(\alpha_0)\right] \tag{3-10}$$

However, the limiting test statistic is usually different from zero. The contribution of the Cox test lies in the fact that the appropriately adjusted LR statistic, which will be shown later, is asymptotically normally distributed.

Accordingly, non-nested hypotheses for model selection may be interpreted to mean that an individual model may not be obtained from the other models by an appropriate restriction or under a limit process. For example, Cox (1961) gave an example of tests between the log-normal distribution and the exponential distribution by an asymptotic testing method. The standard LR test statistics are centered at zero when the two hypotheses are nested. In contrast, when the hypotheses are non-nested, the distribution of the test statistic does not have zero expectation. To center the test statistic for two non-nested models, the Cox statistic includes two components: the usual likelihood ratio statistic and its expected value under the alternative hypothesis over the probability distribution function of the null hypothesis. The difference between them yields the new centered log-likelihood ratio statistic.

The copula approach adopted in Chapter Two may be used to model any bivariate probability distribution by separately estimating the marginal distributions and the dependence structure. In practice, it is important to know which copula is the right one

among the competing copula models. We need to have reasonable criteria to choose an appropriate copula model fitted to the data. For a simple illustration, we demonstrate a graphical method proposed by Genest and Rivest (1993) in Figure 3-1. We use two different data series. One comes from the monthly hedge fund returns used in Chapter One. The other data set is monthly S&P 500 index returns. Both series range from June 1992 to July 2007. With the Clayton copula function, we first estimate an empirical distribution of the copula. We have to have an empirical estimate of the bivariate distribution, H given by

$$W_i \equiv \widehat{H}(X_i, Y_i) = \frac{\#\{(X_j,Y_j): X_j < X_i, Y_j < Y_i\}}{T-1} \tag{3-11}$$

where the estimate of the bivariate distribution is based on the proportion of observations that are less than or equal to (X_i, Y_i). Then the empirical copula, $K(w)$ of the random variate $C(U, V)$ can be obtained as

$$K(w) = \frac{\#\{i: 1 \le i \le T, W_i \le w\}}{T} \tag{3-12}$$

Using the relationship between the copula distribution and the generator of parametric copulas, we construct estimates of the parametric copulas (e.g., Clayton and Gumbel copulas). Every copula parameter can be related to the Kendall's tau via the identity

$$\tau = 1 + 4\int_0^1 \frac{\varphi(t)}{\varphi\prime(t)}\,dt \tag{3-13}$$

where the generator $\varphi(\cdot)$ is a function of the copula parameter θ. After obtaining the estimated copula parameter, we can get the generator involved with the corresponding

copula. Based on a unique generator, the parametric estimate of the copula distribution can be obtained as

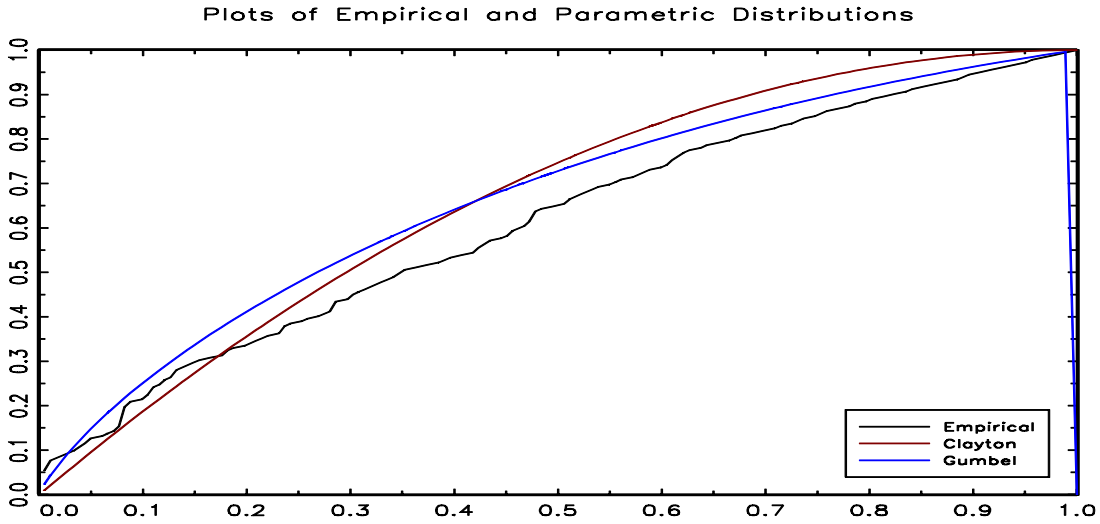
$$K_C(w) = w - \frac{\varphi(w)}{\varphi'(w^+)} \tag{3-14}$$

where w is the realized value of the bivariate pseudo random variable and $\varphi(\cdot)$ is the generator of a specific copula. For Clayton and Gumbel copulas, their generators are respectively,

$$\varphi^C(w) = \frac{1}{\theta}\left(w^{-\theta} - 1\right) \text{ and } \varphi^G(w) = (-\ln(w))^\theta \tag{3-15}$$

where φ^C denotes the generator of Clayton copula and φ^G is the generator of Gumbel copula.

Figure 3-1 Comparison of the empirical and Parametric copulas



After obtaining the nonparametric and the parametric distributions, we need to compare each parametric estimate of the copula distribution to the nonparametric distribution. By choosing a copula that is closest to the nonparametric estimate $K(w)$ in

Equation 3-12, we can select an appropriate copula which fits best to the original data. In Figure 3-1, from the upper right corner, the lowest curve represents the nonparametric (i.e., empirical) distribution, the Gumbel copula distribution is located above the nonparametric distribution, and the Clayton copula distribution is on the top. According to visual inspection, it seems like the Clayton copula distribution better resembles the nonparametric distribution.

As an alternative to graphical methods, difference measures between the competing models can be used for goodness-of-fit tests. In general, the goodness-of-fit approach tells us how well the fitted model reflects the data. Examples include the L^2-norm distances and the adjusted R^2-statistics. Here we compute Kolmogorov-Smirnov test (KS) statistics for the same monthly data. The KS test uses the maximized absolute discrepancy between the fitted and the hypothesized distribution functions

$$KS = \max_x\left\{\left|\hat{F}(x) - F(x)\right|\right\} \tag{3-16}$$

We conduct two cases; one is for the Clayton copula versus the empirical distribution, and the other is for the Gumbel copula versus the empirical distribution. The test statistic for the former case is 0.114 and the corresponding critical value at the 5 percent significance level is 0.137. Therefore, we do not reject the null hypothesis that two competing distribution functions are in the same family. In contrast, for the case of the Gumbel copula, we reject the null of the identical distributions at the 5% level since the test statistic is one and the critical value is 0.137. Therefore, it turns out that the estimate of the Clayton copula is closest to the empirical distribution. This result is consistent with the graphical comparison of the empirical and parametric copula functions.

There are a few drawbacks in applications of the Cox statistic. When the hypotheses are non-nested there is no maintained null hypothesis (i.e., union of the null and alternative hypothesis is empty). First, it is possible to form another test with a distinct outcome by reversing the null and the alternative hypotheses. However, this criticism is valid only if the main goal is to select the one correct model. Second, from the analytical and practical perspective, there is difficulty in obtaining a consistent estimator of the expected log-likelihood ratio statistic in the Cox statistic under the null hypothesis. The KLIC related term may be used in the Cox test statistic as an adjustment to the standard likelihood ratio statistic to make its mean equal to zero. It may be simple to work with and useful for separation between models. A non-nested test procedure for copula model selection that is based on the Cox test statistic is one possibility. Due to the computational difficulties of non-nested testing methods, Monte Carlo sampling methods for computing the Cox test statistic is an alternative choice.

The paper is organized as follows: The next section provides a detailed literature review. The third section briefly discusses the Cox test statistic. Finally, prior discussions are summarized.

3.2 Literature review

In the literature, there are a few approaches for the goodness-of-fit testing of copula models. One can break them down into several categories according to how the major components of the test statistics are constructed. In practice, empirical marginals can be used and an initial functional form for the copula may be presumed, but the goodness-of-

fit for copulas may be complicated since the marginal distributions are left unspecified. Unlike Breymann, Dias, and Embrechets (2003), Dorbic and Schmid (2007) show that the use of empirical marginal distributions makes a difference. However, this makes it hard to determine the asymptotic behavior of the relevant estimators (Berg 2009). Due to this limitation, the usually p-values for the test statistics must be computed by simulation.

The goodness-of-fit testing procedure is conducted by comparing the proposed empirical process to some parametric copula models. We refer to the empirical copula as the rank-based nonparametric copula. In this case, graphical selection procedures based on Kendall's process are usually used. Furthermore, by considering the discrepancy between the competing processes such as the L^2-norm divergence measure, they can be applied to goodness-of-fit testing, and examples include the Kolmogorov-Smirnov[26] statistic, the Anderson-Darling statistic, the Cramer-von Mises statistic. To select the appropriate parametric copula model, the likelihood ratio principle may be used, and Chen and Fan (2005, 2006) propose pseudo- likelihood ratio tests.

The empirical copula goodness-of-fit approach was proposed by Deheuvels (1979), and many past studies based on the approach rely on graphical selection procedures to find the appropriate model. Most results in this category follow Genest and Rivest (1993) and use the estimate of Kendall's tau to derive the best model among the class of Archimedean copulas because of the one-to-one links between Archimedean copulas and

[26] Actually the Kolmogorov-Smirnov statistic is an L^1-norm, taking the absolute difference of two distribution functions.

Kendall's tau. Other works include Genest and Favre (2007), Durrleman, Nikeghbali and Roncalli (2000), Frees and Valdez (1998), and Wang and Wells (2000). The method has a drawback in that it only applies to one-parameter models. Barbe et al. (1996) examine the limiting distribution of the empirical process[27]. Under weak regularity conditions on the empirical limiting distribution function K, they prove that the empirical process is asymptotically Gaussian. Fermanian, Radulovic, and Wegkamp (2004) and Tsukahara (2005) show that the empirical copula estimator is consistent and asymptotically normally distributed.

Another analysis based on bivariate survival (Andersen et al. 2005) conducts the goodness-of-fit test by comparing the parametric estimate of the copula and an empirical copula (nonparametric) estimate. Wang and Wells (2000) propose that one can select a copula that provides the smallest value of the given Cramer-von Mises test statistic based on the Kendall process. However, Genest et al. (2006) and Genest and Farvre (2007) point out that the limiting distribution of the proposed statistic is problematic and it leads to the miscalculation of p-values for the test statistic. In Genest et al. (2006), they verify that two variants of the statistic proposed in their study can circumvent the limitations.

For tests of goodness-of-fit for parametric families of copulas, Dorbic and Schmid (2005) use a modified version of the chi-square test. In a situation where the marginal distributions are unspecified, the proposed chi-square test is used to detect a misspecified

[27] $\alpha_n(t) = \sqrt{n}\,\{K_n(t) - K(t)\}$, where $K_n(t)$ is the empirical distribution function of the pseudo-observations.

null hypothesis. Junker and May (2005) propose the integrated Anderson-Darling statistic [28] to compare the empirical copula to the parametric copula. Other works containing the Anderson-Darling statistics include Breyman, Dias, and Embrechets (2003), Malevergne and Sornette (2003), and Berg and Bakken (2005). In the same vein, Kole et al. (2007) deal with the selection issues of copulas in risk management using several test statistics that are variants of those proposed by Malevergne and Sornette (2003).

Rosenblatt's transformation is often used to construct the test statistics as in Breymann Dias, and Embrechets (2003), and Malevergne and Sornette (2003). First, we need to obtain the uniform variables through the probability integral transformation for multivariate data. Then, using the inverse Gaussian distribution, one can get the transformed univariate observations, which follows a chi-square distribution. In the former case, they test whether the null hypothesis of ellipticality for currency returns is rejected. They regard a copula model with a minimal Akaike Information Criterion (AIC) value as the best fitting model. The proposed method is used to distinguish between a Student-t copula and a Gaussian copula. Similar research is done by Dobric and Schmid (2007), and the studies find that use of the empirical distribution may generate critical values that are misleading. As a remedy, Dobric and Schmid (2007) recommend the parametric bootstrap method for obtaining the proper critical values. In certain cases, the

[28] They use the integrated Anderson-Darling statistic since it produces a more global goodness-of-fit measure than the Anderson-Darling test statistic.

null distribution of the test statistics is unknown. The parametric bootstrap method may be used to derive the critical values by simulation. In the context of copula models, we can first estimate the model specific parameter from the original observations. Second, by adapting the estimated parameter to a particular copula model, we can generate uniform variables. Third, with the simulated uniform variables, we must estimate the same parameter so that we may compute the relevant test statistic. Eventually, we must repeat the second and third step several times to obtain desirable critical values.

The probability integral transform enables multivariate data to be mapped to a univariate function. By doing this, the Kolmogorov-Smirnov test statistic is feasible and easy to compute. Genest et al. (2009) is a variant of Rosenblatt's transform procedure and forms the Cramer-von-Mises statistic based on the discrepancy between the empirical distribution and the independence copula. Instead of using the transformed chi-square random variable, they use the transformed pseudo observations which are approximately uniformly distributed. One advantage of Rosenblatt's transform method is that both the null and alternative hypotheses are the same as in Berg (2009). Hong and Li (2005) show that the tests based on transformed variables perform better than the tests based on the original variables. However, there are drawbacks. The transformation is complicated with high dimension variables and is model specific. Thus, the testing procedure is not really distribution free.

The smoothing-based goodness-of-fit approach uses kernel methods to derive a smoothed estimate of a copula. It does not assume any particular parametric form for the

copula. Fermanian (2005) proposes a goodness-of-fit test[29] for the difference between the estimated parametric copula density and the nonparametric kernel copula density. However, he notes that the simple goodness-of-fit chi-square test based on the kernel estimator of a copula density depends on the assumed copula model and is not distribution-free. Using the smoothed copula estimator, he proves that the alternative test statistic is asymptotically chi-square under the null hypothesis. However, the kernel estimation of the copula density suffers from the "curse of dimensionality". As the dimension increases, the number of parameters to be estimated grows and the accuracy of the test is reduced. To overcome this problem, Chen et al. (2004) use the probability integral transformation to reduce the dimensionality problem, which leads to a distribution-free test. It turns out that the proposed test statistic is consistent and asymptotically standard normal under some regularity conditions.

Following Fermanian and Scaillet (2002), Scaillet (2007) constructs a test statistic based on the kernel estimator of the copula function and the kernel smoothed estimated parametric copula function. The power of the test compares well with Genest et al. (2006). Fermanian and Scaillet (2005) use the test statistic proposed by Fermanian (2005) and show how to choose between the parametric families of copulas. Panchenko (2005) develops a goodness-of-fit test for bivariate and multivariate copulas for the purpose of including larger dimensions. This test takes a kernel-based positive definite bilinear form.

[29] Unlike the simple chi-square approach, the main approach considers the convolution between the kernel and the estimated parametric copula density.

Since the asymptotic theory of the squared divergence is not known, he uses a parametric bootstrap approach to approximate the p-values for the test.

Likelihood and pseudo likelihood methods have also been proposed. Nikoloulopoulos and Karlis (2008) conduct the goodness-of-fit test with the Kullback-Leibler discrepancy measure. Chen and Fan (2005, 2006) use the likelihood ratio tests for model selection of semiparametric copula[30] models. These studies extend Vuong (1989) to form the null and alternative hypotheses. Given a benchmark model in the case of the non-nested hypothesis, each of the competing models is compared with the true copula model. When the two candidate models are non-nested, it is proved that the proposed test statistic is asymptotically standard normal. For tests of model selection, Chen and Fan (2005) rely on the sequential method in which one first tests if two models are non-nested and then conducts the test using the statistic in their Theorem 3 under the null hypothesis. However, since the test statistic in the first test is not distribution-free, the authors use a bootstrap method. In the same spirit of Chen and Fan (2005), Chen and Fan (2006) apply the pseudo likelihood ratio (PLR) test under AIC to a different situation where the marginal distributions can be described by temporal dependence. One fascinating result is that the limiting distribution of the estimators of the pseudo true copula parameters is

[30] For the estimation of a semi-parametric copula, the univariate marginal distributions should first be estimated non-parametrically. Given the marginal parameters, we can estimate the copula parameters. It is called the inference for the margins (IFM) estimation method, and Joe (1997) shows that IFM is consistent and asymptotically normal.

independent of the estimation of the dynamic parameters. So the simulation study can be easy to conduct.

Simulation studies are conducted to assess the finite-sample properties of the proposed goodness-of-fit tests for various classes of copula models. In practice, we may find that unknown parameters must be estimated to construct the associated test statistics. To circumvent these methodological issues and form approximate p-values for the test statistics, one can follow several earlier works that use a parametric bootstrap approach.

3.3 Cox Test Statistic

We are interested in conditional models in the sense that probability distribution of a vector of dependent variable, Y_t, is modeled conditional on a set of explanatory variables, X_t. In a time series model, X_t can include the lagged values of the dependent variable. As Pesaran and Weeks (2001, p.281) point out, the comparison of conditional models is valid only when the parameters of the marginal model are independent of those of the conditional model. For example, in a bivariate case,

$$f(y_t|x_t) = \frac{f(x_t,y_t)}{f_x(x_t)} \tag{3-17}$$

where $f_x(\cdot)$ is a marginal model. When we attempt to compare two competing conditional models, the respective marginal models should not be involved with the conditional models. We are going to deal with cases where conditional models have the same conditioning variables. The two competing non-nested models with a parametric form can be written by

$$H_f: \mathcal{F}_\theta = \{f(y_t|x_t, \Omega_{t-1}; \theta), \theta \in \Theta\} \tag{3-18}$$

$$H_g: \mathcal{F}_\gamma = \{g(y_t|x_t, \Omega_{t-1}; \gamma), \gamma \in \Gamma\} \tag{3-19}$$

where Ω_{t-1} denotes all past information available. For notational simplicity, we can use $f(y_t|\theta)$ and $g(y_t|\gamma)$ in place of $f(y_t|x_t, \Omega_{t-1}; \theta)$ and $g(y_t|x_t, \Omega_{t-1}; \gamma)$, respectively (see Pesaran and Weeks, 2001). In this context, we can regard the null hypothesis, H_f, as the true probability distribution with the true value, θ_0, of the parameter θ. The alternative hypothesis, H_g, is based on a mis-specified probability distribution.

Given the sample (y_t, x_t) and conditional models, $f(y_t|\theta)$ and $g(y_t|\gamma)$, we want to test H_f against H_g. In other words, among the two competing models, $\mathcal{F}_\theta$ and $\mathcal{F}_\gamma$, we want to know which model is appropriate to fit the data. The maximum likelihood estimators of θ and γ can be obtained by maximizing the log-likelihood functions under H_f and H_g:

$$\hat{\theta} = arg\max_{\theta \in \Theta} \sum_{t=1}^{T} ln\, f(y_t|\theta) \tag{3-20}$$

$$\hat{\gamma} = arg\max_{\gamma \in \Gamma} \sum_{t=1}^{T} ln\, g(y_t|\gamma) \tag{3-21}$$

We need to assume that the conditional density functions satisfy the regularity conditions to guarantee that the ML estimators have asymptotic normal distributions. In other words, $\hat{\theta}$ is the usual ML estimator and is a consistent estimator of the true parameter, θ_0. The ML estimator for H_g is a pseudo-ML estimator. As noted earlier, the data sampling process may be different from the assigned hypotheses, and the subjects of the pseudo-ML estimators are pseudo-true values. Under H_f, the standard likelihood ratio test statistic is

$$LR_{fg} = 2\left[L_g(\hat{\gamma}) - L_f(\hat{\theta})\right] \tag{3-22}$$

where $L_f(\hat{\theta})$ is the log-likelihood function evaluated at $\hat{\theta}$ for H_f and $L_g(\hat{\gamma})$ is the log-likelihood function evaluated at $\hat{\gamma}$ for H_g.

Prior to measuring the discrepancy between the two hypotheses, we need to introduce a closeness measure between two conditional densities, which is the Kullback-Leibler discrepancy (Kullback and Leibler 1951). In this regard, the usual maximum likelihood method is a variant of the Kullback-Leibler discrepancy measure in the sense that it tries to find a model that is closest to the empirical function. It would equal zero if and only if two density functions are the same. Given the hypotheses defined by Equation (3-16) and (3-17), the Kullback-Leibler information criterion (KLIC) can be denoted as

$$\begin{aligned} I_{fg}(\theta,\gamma) &= \int_y ln\left(\frac{f(y_t|\theta)}{g(y_t|\gamma)}\right) f(y_t|\theta)dy \\ &= E_f[lnf(y_t|\theta) - lng(y_t|\gamma)] \end{aligned} \tag{3-23}$$

The KLIC measure is always greater than or equal to zero such that $I_{fg} \geq 0$. Also when the dependent variable y_t is independent of each other, the KLIC measure is additive over the sample due to the property of the logarithm function.

As explained in the introduction, the Cox test statistic is based on the likelihood ratio test statistic proposed by Cox (1961, 1962). In the case of nested models, the test statistic has asymptotically zero mean value. When the restriction is truly valid for the finite sample, the estimated log-likelihood function is not significantly different from one in the unrestricted log-likelihood function. As the sample size increases, the discrepancy

between the restricted and unrestricted likelihood functions evaluated at the ML estimators will converge in probability to zero. However, when the non-nested models are present, the likelihood ratio test statistic may diverge asymptotically from zero. Furthermore, the test statistic does not follow a Chi-squared distribution. Cox proposes a new test statistic to make it converge to zero under the null hypothesis H_f as the sample size increases. Let us assume that the asymptotic true value, $\gamma_0(\cdot)$, is known. The Cox test statistic can be written as

$$S_{fg} = \frac{1}{T}\left(L_g(\hat{\gamma}) - L_f\left(\hat{\theta}\right)\right) - I_{fg}(\hat{\theta}, \hat{\gamma}_0) \tag{3-24}$$

where $I_{fg}(\theta, \gamma_0(\theta)) = E_x E_\theta[lng(y_t|x_t; \gamma_0(\theta)) - lnf(y_t|x_t; \theta)]$. By construction, the Cox statistic is defined by subtracting a consistent estimator of the asymptotic standard (or expected) log-likelihood ratio test statistic under H_f from the standard log-likelihood ratio test statistic. Gourieroux and Monfort (1994, p. 2603) demonstrate that under the null hypothesis, H_f, $\sqrt{T}S_{fg}$ is asymptotically normally distributed with zero mean and a variance ω_0^2 (see proposition 3.4.1). As a result, the asymptotic distribution of the standardized Cox statistic is:

$$\sqrt{T}\frac{S_{fg}}{\widehat{\omega}_T} \xrightarrow{d} N(0,1) \tag{3-25}$$

where $\widehat{\omega}_T$ is a consistent estimator of ω_0 under the null, H_f.

The Cox test statistic is closely related to the closeness measure. When we measure the closeness between two competing models under the null hypothesis, H_f, the following relationship is established (see Pesaran 1987):

$$C_{fg}(\theta) = I_{fg}(\theta, \gamma_0(\theta)) \tag{3-26}$$

where $\gamma_0(\theta)$ is the pseudo-true value of γ under the null, H_f. Similarly, an attempt to measure the closeness the other way around can be conducted such as

$$C_{gf}(\gamma) = I_{gf}(\gamma, \theta_0(\gamma)) \tag{3-27}$$

where $\theta_0(\gamma)$ is the pseudo-true value of θ under the null, H_g. Note that the subscripts in Equation 3-24 have the reversed order of the Equation 3-25.

For hypothesis testing of non-nested models, there is no natural null hypothesis, and it is possible to switch the role of the null and alternative hypothesis. Due to this fact, it is possible to reject all models or to fail to reject all models. However, as long as we attempt to know about the strengths or weaknesses of one model relative to other models, the use of non-nested models would be valid (Pesaran and Weeks, 2001).

An interesting variant of the Cox test statistic shows up in the model selection and hypothesis testing approach. Vuong (1989) introduces an additional model other than the null and the alternative models. In his setting, the null hypothesis is defined by the closeness between the true model and one model, C_{hf}, and the alternative hypothesis is the closeness defined by C_{hf}.

For application purposes, the Cox test statistic suffers computational difficulties. In Equation 3-24, it is hard to obtain a consistent estimator of the expected log-likelihood ratio statistic, $I_{fg}(\hat{\theta}, \hat{\gamma}_0)$. Also, it is hard to obtain a consistent estimator of the pseudo-true value $\gamma_0(\theta)$, $\hat{\gamma}_0$ in Equation 3-24. Finally, the explicit functional form of Equation 3-22 may not be available. In other words, an analytical derivation of the integral in

Equation 3-22 is generally not possible. Therefore, most of model computational problem lies with the numerator of Equation 3-23.

To resolve some of these practical challenges, we can use Monte Carlo (MC) sampling methods for computing the Cox test statistic and to evaluate its distributional properties. As in Pesaran and Pesaran (1993, particular through proposition 1 and 2), $\sqrt{T}\left(\hat{\theta}-\theta_0\right)$ and $\sqrt{T}(\hat{\gamma}-\gamma_0)$ are asymptotically normally distributed under the null hypothesis, H_f. As a practical matter, the estimator of the pseudo true value, $\hat{\gamma}_0$under the null hypothesis, H_f, requires the ML estimator $\theta=\hat{\theta}$ from the simulated observations,Y_t, according to the null, H_f. At this point, we assume a true probability density function such as $f(y|\hat{\theta})$. Then, we use the same observations Y_t to calculate the ML estimate of γ, $\hat{\gamma}_g$, under H_g. To produce the estimate of the pseudo true value, $\hat{\gamma}_0$, the MC method comes in play. The estimate is generated by

$$\hat{\gamma}_0(R)=\frac{1}{R}\sum_{i=1}^{R}\hat{\gamma}_{0,i} \tag{3-28}$$

where $R$is the number of replications. By the law of large numbers, the estimator in Equation 3-26 converges in probability to γ_0 as R increases.

Given $\hat{\gamma}_0(R)$, the consistent estimator of the Kullback-Leibler closeness measure between H_f and H_g is given by

$$C_R\left(\hat{\theta},\hat{\gamma}_0\right)=\frac{1}{R}\sum_{i=1}^{R}\left\{L_f\left(Y_i,\hat{\theta}\right)-L_g(Y_i,\hat{\gamma}_0(R))\right\} \tag{3-29}$$

The numerator of the standardized Cox test statistics can be computed as

$$S_{fg}\left(\hat{\theta},\hat{\gamma}\right)=\frac{1}{T}\left(L_g(Y,\hat{\gamma})-L_f\left(Y,\hat{\theta}\right)\right)-C_R\left(\hat{\theta},\hat{\gamma}_0\right) \tag{3-30}$$

By adding the simulation estimator of $\widehat{\omega}_T$ in Equation 3-23, we can construct the Cox test statistic based on the MC simulation method, which is asymptotically standard normal under the null. For example, when the Cox test statistic is large, we reject the null hypothesis, e.g., H_f. Similarly we need to conduct another hypothesis test with the role of each hypothesis switched. As illustrated in Pesaran and Deaton (1978), it is possible that the null hypothesis, H_g, may not be rejected. In this case, the final decision is that H_f should be rejected in favor of $H_g$on this evidence.

3.4 Copula Model Selection

Let $y_t = (y_{1t}, y_{2t})'$ be an 2×1 continuous random vector at time t. Let's define $\Omega_{1,t-1}$ and $\Omega_{2,t-1}$ as information sets generated by past observations and some pre-determined variables at time t. Given the information set $\Omega_{t-1} = \left(\Omega_{1,t-1}, \Omega_{2,t-1}\right)$, the dependence structure of y_t is fully characterized by the true conditional bivariate distribution of $y_t|\Omega_{t-1}$, denoted by $F_y^0(\cdot\,|\Omega_{t-1})$. Let $F_{yi}^0(\cdot\,|\Omega_{t-1})$ be the true conditional distribution of $y_{it}|\Omega_{t-1}$. By the conditional Sklar theorem from Patton (2006), there is a unique conditional copula $C_0(\cdot\,|\Omega_{t-1})$: $[0,1]^2 \rightarrow [0,1]$ such that

$$F_y^0(y|\Omega_{t-1}) = C_0\left(F_{y1}^0(y_1|\Omega_{t-1}), F_{y2}^0(y_2|\Omega_{t-1})\right) \tag{3-31}$$

for all $y = (y_1, y_2) \in \mathbb{R}^2$.

We specify the marginal models based on the following bivariate framework:

$$y_t = \mu_t(x_t, \alpha_1) + \sqrt{H_t(\alpha)}\varepsilon_t \tag{3-32}$$

where x_t is a Ω_{t-1} measurable random vector, $\mu_t(x_t, \alpha) = \left(\mu_{1t}(x_t, \alpha_1), \mu_{2t}(x_t, \alpha_2)\right)'$ $= E[y_t|\Omega_{t-1}]$ is the true conditional mean of y_t given Ω_{t-1} ; and $H_t(\alpha)$ $= diag\left(h_{1t}(\alpha), h_{2t}(\alpha)\right)$ where $h_{1t}(\alpha) = h_{1t}(\alpha_1, \alpha_2) = E[((y_{1t} - \mu_{1t}(x_t, \alpha_1))^2)|\Omega_{t-1}]$, and $h_{2t}(\alpha) = h_{2t}(\alpha_1, \alpha_2) = E[((y_{2t} - \mu_{2t}(x_t, \alpha_2))^2)|\Omega_{t-1}]$ are the true conditional variances of y_{it} given Ω_{t-1}, and $\alpha = (\alpha_1', \alpha_2')'$, where α_1 and α_2 do not share common elements. The standardized bivariate innovations, $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t})'$ are independent of Ω_{t-1}, and are identically independently distributed with $E(\varepsilon_{it}) = 0$ and $E(\varepsilon_{it}^2) = 1$ for $j = 1,2$. Let us define the conditional parametric distribution as $F_{\varepsilon_i}(\varepsilon_{it}|x_t; \beta_i)$ for $i = 1,2$ with a parameter vector β_i. This generates the marginal models

$$F_{y_i}(y_i|x_t; \alpha_i, \beta_i) = F_{\varepsilon_i}\left(h_{it}^{-1/2}(y_i - \mu_{it})|x_t; \beta_i\right) \tag{3-33}$$

Let $C(\cdot\,|x_t; \theta)$ be a copula model, which can be used to approximate the true conditional copula $C_0(\cdot\,|\Omega_{t-1})$. When we link the marginal models to the copula function, the joint distribution is given by

$$F_y(y|x_t; \alpha, \beta, \theta) = C\left(F_{y_1}(y_1|x_t; \alpha_1, \beta_1), F_{y_2}(y_2|x_t; \alpha_2, \beta_2)|x_t; \theta\right) \tag{3-34}$$

In this setup the underlying assumption is that the parameter vectors $(\alpha, \beta)'$ are not affected by the other equation. As a result univariate GARCH models are used to present the marginal models. This fact leads to the case where we can separately estimate the parameters of the copula models from the analysis of the marginal models. It is possible to apply the method to evaluate the bivariate Normal copula model for the dependence structure between the EURO spot exchange rate and the corresponding futures rate. The

proposed test is constructed under a null hypothesis that the Normal copula model against several alternatives such as the Student t-copula, Clayton, Frank, and Gumbel copulas.

The filtered returns can be obtained from a GARCH (1,1) model with Student-t distribution for the innovations. As Chen and Fan (2005, 2006) and Huang and Prokhorov (2010) point out, the limiting distribution of the copula parameter can be affected by the estimation results for the marginal distributions. More specifically, those studies rely on empirical marginal distributions to avoid misspecification of the marginals. As long as those marginal are not correctly specified, there is an additional adjustment term in the variance matrix of the copula parameter (see Proposition 3.1 in Chen and Fan 2006). We follow Patton (2006) and use a fully parametric model for the marginal distributions.

$$F_{y_i}(\cdot\,|x_t;\alpha_{io},\beta_{io}) = F^0_{y_i}(\cdot\,|\Omega_{t-1}), \text{ for } i = 1,2 \tag{3-35}$$

where F_{y_i}, $i = 1,2$ is assumed to be Student-t distribution.

When the modeling of the marginal distribution is complete and the relevant parameters are obtained through the maximum likelihood method, the original returns should be transformed into identically and independently distributed univariate series. Before proceeding to the second step of the copula parameter estimation, we must check whether those univariate series are independent and the density functions are correctly specified, and the LM test and the Kolmogorov-Smirnov test can be used. Then the copula parameters should be estimated by maximizing the copula density based log-likelihood function.

The model selection procedure based on non-nested hypothesis testing for this approach requires specific null and alternative hypotheses. For example:

$$H_f: C = C_N \text{ (Student t-copula)}$$

$$H_g: C = C_S \text{ (Normal copula)} \tag{3-36}$$

The Cox test process by changing the alternative copula model to the other candidate models may be implemented.

3.5 Summary

It is very important to choose the best copula model among various candidates even under the assumption that the parameters involved with different copula models are consistently and efficiently estimated. The model selection approaches rely mainly on the ranking of the outcome produced by the chosen criterion. Problems may rise because the resulting values are not supported by statistical considerations (i.e., randomness in outcome). To overcome the difficulty, Monte Carlo studies or bootstrap methods are used to compute the p-values. In other words, we need to know the distribution of the test statistics in a large sample. In this regard, we can deal with the model selection problem in the hypothesis testing framework. However, in principle, hypothesis testing may not point to a right copula model since it does not produce a deterministic outcome.

Several methods of choosing a copula model fitted to the original sample are discussed. The graphical method such as the QQ-plot method, and the norm-based distance measures including the Kolmogorov-Smirnov, Cramer-von Mises, Anderson-Darling or chi-squares test statistics, and the methods based on the likelihood ratio

principle aim to evaluate the competing models. For example, for a given sample we can select the model that maximizes (e.g., R-square) or minimizes (e.g., AIC) a chosen criterion. Based on the fact that the Clayton copula, the Gumbel copula, and the Normal copula can be regarded as non-nested models (Chen and Fan, 2006), there is the incentive to use a nonnested hypothesis testing procedure. Among the general approaches (e.g., comprehensive or encompassing procedures) to nonnested hypothesis testing, the Cox test may be used for the copula model selection. These methods above contribute to the literature by recognizing the importance of the copula model selection and developing various alternative methods. However there are no clear-cut standards for the model selection. Several limitations to various methods addressed above must be considered.

For graphical selection procedures, they can be used to estimate the degree of fit roughly and find, if any, some problem with the fits. One advantage is that the difference between competing models is easy to be identified by visual inspection. However, in general, it is hard to discriminate small differences in assessing model fits. In the copula model selection problem, some parameter space or the derivatives of a copula function to obtain the density function may not be defined (e.g., some range or point may be zero or out of bound). In that case there is no way to draw the graph. Moreover, sometimes it is possible not to discriminate between two different families of copula models on a graph (Matteis 2001). The commonly used benchmark model in the graphical method is the empirical (or nonparametric) copula estimate. The empirical copula estimator may depend on the marginal distribution estimator. As a result, the asymptotic variance of the empirical copula is closely related to the asymptotic variance of the marginal distribution

estimator. This fact leads to the inability of knowing an explicit form of the asymptotic variance of the marginal distribution estimator (Wang and Wells 2000). To overcome the difficulty, the bootstrap method can be used to estimate the asymptotic variance.

Among the norm-based divergence measures, the Cramer-Von Mises and Anderson-Darling statistics are based on a quadratic empirical distribution function (EDF). The difference between the two measures arises from the weight function. The former is a variant of the Kolmogorov-Smirnov test. The small sample distribution of the test is unknown. But the test statistic has the same distribution for the candidate (or a theoretical distribution in one sample case) distribution. In other words, it does not depend on the predetermined distribution under consideration. This is described as distribution free. It is less widely used than the Kolmogorov-Smirnov test. The problem with the Cramer-Von Mises statistic is that the test statistic makes it weaken the influence of the outliers by assigning equal weight to every observation in the sample. In contrast, the Anderson-Darling test puts more weight on the remote observations (i.e., outliers) so that it may be sensitive to the outliers. The Anderson-Darling test has the advantage that it uses a specific distribution under examination in calculating critical values. The disadvantage is that the critical values must be calculated for each distribution. However, the tables of critical values are not always available although the test can be applied to any distributions. The Kolmogorov-Smirnov statistics has an advantage in that it is a distribution free test. However, there is a serious drawback. It has low power, which means an increase in the likelihood of a Type II error. Note that the property of distribution free in EDF tests can be accounted for the fact that the original observations

are transformed into uniform ones by the probability integral transformation. The transformed observations seem likely to be ranked following the uniform distribution. As a result, the transformed uniformly distributed observations are distribution free. A more powerful test is the Chi-square test as in Junker and May (2005). The relevant test statistic is based on the absolute frequency of the data through the size of the bandwidth parameter, but the grouping criteria are ambiguous. When the bin size gets narrow, it will take a risk to lose valuable information. To resolve this problem, kernel density estimation can be used to smooth the curve of the density function. With Rosenblatt's approach (i.e., a way for generating pseudo samples (uniform variables) using transformation of the original sample observations), one advantage in a goodness-of-fit setting is that the null and alternative hypotheses are the same, so it is a distribution free test. A disadvantage is the lack of invariance, which means that the transformed uniform variable may not be independent in case of non-randomness (Berg 2009). The test statistics from the distance measures in this category have their own asymptotic distributions. In certain cases, bootstrap method can be used to calculate critical values.

The likelihood ratio principle for copula model selection can be used. As one of a kind, the Akaike information criterion can be used to rank some listed copula models as in Dias and Embrechts (2004). As pointed out in Genius and Strazzera (2002), the model selection process using the AIC does not take into account the fact that the differences in the two log-likelihood values may not be statistically significant. In the same spirit as Vuong (1989) based on non-nested hypothesis framework, Chen and Fan (2006) deal with the model selection problem based on the method that the selection rule is

statistically meaningful. In there the candidate models are discriminated by the KLIC. One advantage to this method is that one can reflect the non-Gaussian characteristics often observed in actual data. However, although a semiparametric approach such as Chen and Fan (2006) may reduce the risk of a serious misspecification, other information beyond the observed sample would be ignored.

Note the differences in forming the null and alternative hypotheses between Vuong's approach and the Cox statistic. In the Cox statistic, the two competing models are listed as a null and an alternative hypothesis, either one way or the other way around. By contrast, Vuong (1989) sets up the hypotheses by introducing an implicit true model and measuring the differences of two different models from the true model. Furthermore, the KLIC is used as the part of constructing the test statistic in the Cox procedure while it is used to rank the candidate models as a criterion in Vuong (1989).

The Cox test statistic based on the centered log-likelihood ratio has a few drawbacks in applications. First, it is possible to form another test with a distinct outcome by reversing the null and the alternative hypotheses because by design hypothesis testing does not provide a definite outcome. Second, it is hard to obtain a consistent estimator of the expected log-likelihood ratio statistic under the null hypothesis. The KLIC related term may be used in the Cox test statistic as an adjustment to the standard likelihood ratio statistic to make its mean equal to zero. A non-nested test procedure for copula model selection that is based on the Cox test statistic is one possibility. Due to the computational difficulties of non-nested testing methods, Monte Carlo sampling methods for computing

the Cox test statistic is an alternative choice. Third, the Cox test does tend to over-reject the null model under consideration as reported in several studies.

In summary, in the copula model selection literature, there is not yet clear consensus. More study must be taken. In spite of weaknesses in a particular method, the literature to the copula model selection has provided some insights to find a right copula model between the competing models.

References

Akaike, Hirotugu. 1974. A New Look at the Statistical Model Identification. *IEEE Transactopns on Automatic Control* AC-19, no.6 (December): 716-723.

Amemiya, Takeshi. 1985. *Advanced Econometrics*. Cambridge: Harvard University Press.

Andersen, P.K., C.T. Ekstrom, J.P. Klein, Y. Shu, and M.J. Zhang. 2005. A Class of Goodness of Fit Tests for a Copula Based on Bivariate Right-Censored Data. *Biometrical Journal* 47: 815-824.

Ane, T., and C. Kharoubi. 2003. Dependence Structure and Risk Measure. *Journal of Business* 76: 411-438.

Barbe, Philippe, Christian Genest, Kilani Ghoudi, and Bruno Re´millard. 1996. On Kendall's process. *Journal of Multivariate Analysis* 58, no.2 (August), 197–229.

Berg, Daniel. 2009. Copula Goodness-of-fit Testing: An Overview and Power Comparison. *European Journal of Finance* 15, no.7-8: 675-701.

Burnham, Kenneth P., and David R. Anderson. 2002. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach.* 2nd ed. New York:Springer-Verlag.

Breyman, W., A. Dias, and P. Embrechets. 2003. Dependence Structures for Multivariate High-Frequency Data in Finance. *Quantitative Finance* 3: 1-14.

Chen, Xiaohong., and Yanqin Fan. 2005. Pseudo-likelihood Ratio Tests for Semiparametric Multivariate Copula Model Selection. *Canadian Journal of Statistics* 33, no.3 (September): 389-414.

Chen, Xiaohong., and Yanqin Fan. 2006. Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspecification. *Journal of Econometrics* 135, no.1-2(Nov-Dec): 125-154.

Chrubini, U., E. Luciano, and W. Vecchiato. 2004. *Copula Methods in Finance*. John Wiley.

Cox, D.R. 1961. Tests of Separate Families of Hypotheses. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* 1: 105-123.

Cox, D.R. 1962. Further Results on Tests of Separate Families of Hypotheses. *Journal of the Royal Statistical Society. Series B (Methodological)* 24, no.2: 406-424.

Dobric, Jadran, and Friedrich Schmid. 2007. A Goodness of Fit Test for Copulas Based on Rosenblatt's Tranformation. *Computational Statistics & Data Analysis* 51: 4633-4642.

Durrleman V., A. Nikeghbali, and T. Roncalli. 2000. Which Copula is the Right One? Credit Lyonnaid *working paper*.

Fermanian, J.D., D. Radulovic, M.H. Wegkamp. 2004. Weak Convergence of Empirical Copula Processes. *Bernoulli* 10: 847-860.

Frees, Edward W., and Emiliano A. Valdez. 1998. Understanding Relationships Using Copulas. *North American Actuarial Journal* 2, no.1: 1-25.

Genest Christian, and Anne-Catherine Favre. 2007. Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. *Journal of Hydrologic Engineering* 12, no.4: 347-368.

Genest Christian, J.F. Quessy, and B. Remillard. 2006. Goodness-of-fit Procedures for Copula Models Based on the Integral Probability Transformation. *Scandinavian Journal of Statistics* 33: 337-366.

Genest Christian, and Louise-Paul Rivest. 1993. Statistical Inference Procedures for Bivariate Archimedean Copulas. *Journal of the American Statistical Association* 88, no.423 (September): 1034-1043.

Genius, Margarita and Elisabetta Strazzera. 2002. A Note about Model Selection and Tests for Non-nested Contingent Valuation Models. *Economics Letters* 74, 363-370.

Gourieroux, Christian, and Alain Monfort. 1993. Pseudo-Likelihood Methods. In *Handbook of Statistics* 11. ed. Maddala G.S., C.R.Rao, and H.D.Vinod. 335-362. Amsterdam: Elsevier.

Gourieroux, C., and A. Monfort. 1994. Testing Non-nested Hypotheses In *Handbook of Econometrics* 4. ed. R.F.Engle and D.L.McFadden. 2586-2637. Oxford: Elsevier.

Gourieroux, C., and A. Monfort. 1995a. Nonnested Tests and Model Selection Criteria In *Statistics and Econometric Models* vol.1. Translated by Quang Vuong. Cambridge: Cambridge University Press.

Gourieroux, C., and A. Monfort. 1995b. Nonnested Tests and Model Selection Criteria In *Statistics and Econometric Models* vol.2. Translated by Quang Vuong. 275-323, Cambridge: Cambridge University Press.

Godfrey, L. G., Michael McAleer, and C. R. McKenzie. 1988. Variable Addition and Lagrange Multiplier Tests for Linear and Logarithmic Regression Models. *Review of Economics and Statistics* 70, no.3(August): 492-503.

Huang, Wanling, and Artem Prokhorov. 2010. A Goodness-of-fit Test. Concordia University Dep't of Economics Working Paper Series 10-002.

Joe, H. 1997. *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.

Judge, G. George, R. Carter Hill, William E. Griffiths, Helmut Lutkepohl, and Tsoung-Chao Lee. 1988. *Introduction to the Theory and Practice of Econometrics*. 2nd ed. New York: John Wiley & Sons.

Junker, Markus and Angelika May. 2005. Measurement of Aggregate Risk with Copulas. *Econometrics Journal* 8. 428-454.

Kullback, J., and R.A.Leibler. 1951. On Information and Sufficiency. *Annals of Mathematical Statistics* 22, no.1 (March): 79-86.

Malevergne, Y., and D. Sornette. 2003. Testing the Gaussian Copula Hypothesis for Financial Assets Dependence. *Quantitative Finance* 3: 231-250.

Matteis, Roberto de. 2001. Fitting Copulas to Data. Diploma Thesis. University Zurich.

Mittelhammer, Ron C., George G. Judge, and Douglas J. Miller. 2000. *Econometric Foundations*. Cambridge: Cambridge University Press.

Nelsen, R.B. 1999. *An Introduction to Copulas*. New York: Springer-Verlag.

Nikoloulopoulos, Aristidis K. and Dimitris Karlis. 2008. Copula model evaluation based on parametric bootstrap. *Computational Statistics & Data Analysis* 52, no.7(March): 3342-3353.

Patton, Andrew. 2006. Modelling Asymmetric Exchange Rate Dependence. *International Economic Review* 47, no.2 (May): 527-556.

Pesaran, M. H. and A.S. Deaton. 1978. Testing Non-nested Nonlinear Regression Models. *Econometrica* 46, no.3 (May): 677-694.

Pesaran, M. Hashem and Bahram Pesaran. 1993. A Simulation Approach to the Problem of Computing Cox's statistic for testing Nonnested Models. *Journal of Econometrics* 57: 377-392.

Pesaran, M. Hashem. 1987. Global and Partial Non-nested Hypotheses and Asymptotic Local Power. *Econometric Theory* 3: 69-97.

Pesaran, M. Hashem, and Melvyn Weeks. 2001. Nonnested Hypothesis Testing: An overview. In *A Companion to Theoretical Econometrics*. ed. Badi H. Baltagi, 279-309. Massachusetts: Blackwell Publishers.

Tsukahara, H. 2005. Semiparametric Estimation in Copula Models. *Canadian Journal of Statistics* 33: 357-375.

Vuong, Qyang H. 1989. Likelihood Ratio Tests for Model Selection and Nonnested Hypothesis. *Econometrica* 57. no.2 (March): 307-303.

Wang, Weijing and Martin, T. Wells. 2000. Model Selection and Semiparametric Inference for Bivariate Failure-Time Data. *Journal of the American Statistical Association* 95, no.449 (March): 62-75.

White, H. 1982. Regularity Conditions for Cox's Test of Nonnested Hypothesis. *Journal of Econometrics* 19: 301-18.

Wooldridge, Jeffrey M. 2003. *Introductory Econometrics: A Modern Approach*. 2nd ed. United States: Thomson, South-Western.

Vita

Moohwan Kim attended Masan High School, South Korea. In 1988, he entered Sungkyunkwan University in Seoul, South Korea. He graduated with a Bachelor's degree in Economics in February, 1995. He received the degree of Master of Economics from Sungkyunkwan University in February, 1997. During the following years he was employed as a research associate at Korea Development Institute (KDI). Entering University of Virginia in 2001, he earned the degree of Master of Economics in 2003. He entered the Graduate school at University of Missouri at Columbia in August, 2004. In 2011, he received a Ph.D in Economics from the University of Missouri-Columbia.