Three Essays In Finance Economics

Author: Chuanliang Jiang

Persistent link: http://hdl.handle.net/2345/3178

This work is posted on eScholarship@BC, Boston College University Libraries.

Boston College Electronic Thesis or Dissertation, 2013

Copyright is held by the author, with all rights reserved, unless otherwise noted.
Boston College
The Graduate School of Arts and Sciences
Department of Economics

THREE ESSAYS IN FINANCE ECONOMICS

a dissertation

by

Chuanliang Jiang

submitted in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

May 2013
This dissertation contains three essays. It provides an application of quantile regression in Financial Economics.

The first essay investigates whether tail dependence makes a difference in the estimation of systemic risk. This chapter develops a common framework based on a copula model to estimate several popular return-based systemic risk measures: Delta Conditional Value at Risk ($\Delta CoVaR$) and its modification ($\Delta CoVaR^c$); and Marginal Expected Shortfall ($MES$) and its extension, systemic risk measure ($SRISK$). By eliminating the discrepancy of the marginal distribution, copula models provide the flexibility to concentrate only on the effects of dependence structure on the systemic risk measure. We estimate the systemic risk contributions of four financial industries consisting of a large number of institutions for the sample period from January 2000 to December 2010. **First**, we found that the linear quantile regression estimation of $\Delta CoVaR$, proposed by Adrian and Brunnermeier (AB hereafter) (2011), is inadequate to completely capture the non-linear contagion tail effect, which tends to underestimate systemic risk in the presence of lower tail dependence. **Second**, $\Delta CoVaR$ originally proposed by AB (2011) is in conflict with dependence measures. By comparison, the modified version of $\Delta CoVaR^c$ put forward by Girardi et al. (2011) and
MES, proposed by Acharya et al. (2010), are more consistent with dependence measures, which conforms with the widely held notion that stronger dependence strength results in higher systemic risk. **Third**, $\Delta CoVaR^\leq$ is observed to have a strong correlation with tail dependence. In contrast, $MES$ is found to have a strong empirical relationship with firms’ conditional CAPM $\beta$. **SRISK**, however, provides further connection with firms’ level characteristics by accounting for information on market capitalization and liability. This stylized fact seems to imply that $\Delta CoVaR^\leq$ is more in line with the “too interconnected to fail” paradigm, while **SRISK** is more related to the “too big to fail” paradigm. In contrast, $MES$ offers a compromise between these two paradigms.

The second essay proposes a quantile regression approach to stock return prediction. I show that incorporating distributional information together with combining model information can produce a superior forecast for the conditional mean as well as the entire distribution of future equity premium, which significantly outperforms the forecast that utilizes either source of information alone. Meanwhile, the order of combination strategies appears to make a difference in the efficiency of pooling both distributional information and model information. It turns out that aggregating distributional information in the first step, followed by combining model information in the second step is more advantageous in return forecast than the alternative combination strategies which reverse the order of combination strategy. Furthermore, the forecast based on **LASSO** model selection can be significantly improved as well if the distributional information is further incorporated. In other word, aggregating distributional information via combining multiple quantiles estimators contributes to the improvement of forecasts obtained either from model combination or model selection. This paper not only investigates the forecast of conditional mean, but also studies the forecast of the whole distribution of future stock returns. The approaches of quantile combination together with either model combination or model selection turn out to
deliver statistically and economically significant out-of-sample forecasts relative to a historical average benchmark.

The third essay proposes a quantile-based approach to efficiently estimate the conditional beta coefficient without assuming a parametric structure on the distribution of data generating process. Multiple quantiles estimates are combined in a weighting scheme to utilize distributional information across different quantile of the distribution. Monte Carlo simulation demonstrated that combining multiple quantile estimates can substantially improve the estimation efficiency for beta risk estimates in the absence of Gaussian distribution. The robustness of quantile-based beta estimates are pronounced during financial crisis when the distribution of stock returns deviates most from normality. I also explored the performance of different beta estimators in an application of portfolio management analysis and found that beta estimates from the proposed quantile combination approaches are superior to the OLS estimates in constructing Global Minimum Variance Portfolio, which generates lower variance of portfolio but does not come at the expense of persistent lower returns.
# Table of Contents

1 Does Tail Dependence Make A Difference In the Estimation of Systemic Risk? $\Delta CoVaR$ and $MES$  
1.1 Introduction ............................................. 1  
1.2 Methodology ........................................... 7  
  1.2.1 Definitions ......................................... 7  
    $\Delta CoVaR$ ........................................... 7  
    $\Delta CoVaR^\leq$ .................................... 8  
    $MES$ .................................................. 9  
    $SRISK$ ............................................... 11  
  1.2.2 Copula Based Estimation of $\Delta CoVaR$ and $MES$ ....... 12  
    $\Delta CoVaR$ Estimation ............................... 13  
    $\Delta CoVaR^\leq$ Estimation ........................ 17  
    $MES$ Estimation ....................................... 19  
    $SRISK$ Estimation ..................................... 21  
1.3 On the Dependence Consistency of Systemic Risk: $\Delta CoVaR$ and $MES$  
1.4 Estimation ............................................. 26  
  1.4.1 Modeling the Marginal Dynamics ...................... 26  
  1.4.2 Copula Models Estimation ........................... 29  
    Dynamic Copula Models ................................. 30  
  1.4.3 Backtesting CoVaR ................................ 31  
    Unconditional Coverage Testing ........................ 32
2 Model Combination or Distributional Aggregation?

A Quantile Regression Approach to Stock Return Prediction

2.1 Introduction ............................................. 78
2.2 Predictive Regression Model .............................. 83
  2.2.1 Predictive Model Selection in Quantile Regression Setting .................................. 85
2.3 Robust Point Forecasts Based on Optimally Combining Quantile Information ..................... 87
  2.3.1 Combining Quantile Forecasts of $Y(\tau)$ .............................................................. 90
    DMSFE Quantile Combination (DMSFE) ................................................................. 91
    Regression Approach of Quantile Combination (Regression) ................................. 91
  2.3.2 Combining Quantile Estimation of Coefficient $\beta(\tau)$ ........................................ 92
    Density Weight Quantile Combination (Density) ..................................................... 93
    Optimal Weight Quantile Combination (OWQ) ......................................................... 94
    Quantile Combination Illustration ................................................................. 97
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4 Model Combination</td>
<td>101</td>
</tr>
<tr>
<td>2.5 Combining Information Both Across Models and In Distribution</td>
<td>104</td>
</tr>
<tr>
<td>2.5.1 Density Combination</td>
<td>105</td>
</tr>
<tr>
<td>2.6 Forecast Evaluation</td>
<td>109</td>
</tr>
<tr>
<td>2.6.1 Statistical Evaluation of Return Predictability</td>
<td>109</td>
</tr>
<tr>
<td>Quantile Forecast Evaluation</td>
<td>109</td>
</tr>
<tr>
<td>Economically Motivated Model Restriction</td>
<td>111</td>
</tr>
<tr>
<td>Conditional Mean Forecast Evaluation</td>
<td>112</td>
</tr>
<tr>
<td>Density Forecast Evaluation</td>
<td>113</td>
</tr>
<tr>
<td>2.6.2 Economic Evaluation of Return Predictability</td>
<td>114</td>
</tr>
<tr>
<td>Optimal Portfolio Choice</td>
<td>114</td>
</tr>
<tr>
<td>2.7 Empirical Results</td>
<td>120</td>
</tr>
<tr>
<td>2.7.1 Data</td>
<td>120</td>
</tr>
<tr>
<td>2.7.2 Empirical Results for Conditional Mean Forecast</td>
<td>124</td>
</tr>
<tr>
<td>In-Sample Stock Return Prediction</td>
<td>124</td>
</tr>
<tr>
<td>Out of sample performance of predictive model</td>
<td>126</td>
</tr>
<tr>
<td>Economic Measures of Forecasting Performance Under Mean Variance Utility</td>
<td>132</td>
</tr>
<tr>
<td>2.7.3 Empirical Results for Density Forecast</td>
<td>134</td>
</tr>
<tr>
<td>Statistical Measures of Density Forecast Performance</td>
<td>134</td>
</tr>
<tr>
<td>Economic Measures of Density Forecast Performance</td>
<td>135</td>
</tr>
<tr>
<td>2.8 Conclusion</td>
<td>138</td>
</tr>
<tr>
<td>2.9 Appendix</td>
<td>139</td>
</tr>
<tr>
<td>2.9.1 Comparison of Mean Square Error(MSE) of coefficient estimates (\hat{\beta}) with MSE of point forecast (\hat{Y})</td>
<td>139</td>
</tr>
<tr>
<td>2.9.2 Choice of Quantile LASSO penalty parameter (\lambda) in Quantile Model Selection</td>
<td>141</td>
</tr>
</tbody>
</table>
I am deeply indebted to Professor Zhijie Xiao, Professor Arthur Lewbel and Professor Christopher F Baum for their guidance, support and patience throughout the development of this dissertation. Their exceptional mentorship inspires and enriches my growth as a student and a researcher. I would also like to thank my colleagues, participants on several workshops for valuable discussions and suggestions.

I dedicate my dissertation work to my family. A special feeling of gratitude to my loving parents and wife whose words of encouragement and push for tenacity ring in my ears. Without their consistent support, I would have failed to finish my PhD dissertation.
Chapter 1

Does Tail Dependence Make A Difference In the Estimation of Systemic Risk? $\Delta CoVaR$ and $MES$

1.1 Introduction

The financial crisis that occurred during 2007-2009 has spurred intensified research attempting to understand the interdependence of risk factors and the fragility of the financial system. In the past few years, the distress associated with some individual financial institutions has spread throughout the entire financial system. An individual financial institution is considered of systemic importance\(^1\) if its distress or failure increases the likelihood that other firms will go bankrupt and threatens the stability of the entire system, either due to its size or its interconnectedness with the rest of the financial industry. Basel III had proposed a capital surcharge for Systemically Important Financial Institutions (SIFIs). In December 2011, the Fed introduced such a surcharge for eight banks: Bank of America, Bank of New York Mellon, Citigroup, Goldman Sachs, JPMorgan, Morgan Stanley, State Street and Wells Fargo. As a result, the questions of how to accurately quantify the risk of spill-over effects and how to identify systemically important financial institutions

\(^1\) Federal Reserve Governor Daniel Tarullo defined it: “Financial Institutions are systemically important if the failure of the firm to meet its obligations to creditors and customers would have significant adverse consequences for the financial system and the broader economy.”
have become crucial for macro-prudential regulators and supervision authorities.

Value at Risk (VaR) and Expected Shortfall (ES), the most widely-used risk measures by financial institutions and academic researchers, have been criticized for their failure to take into account the escalated risk of spill-over effects among financial institutions during episodes of financial crisis. Ang, Chen and Xing (2006) documented that conditional correlations between assets are much higher in downturns of the financial market. More recently, Brunnermeier and Pedersen (2008) show that the negative feedback of a “loss spiral” or “margin spiral” leads to the joint depression of asset prices. Systemic risk arises because of increasing co-movement and linkages among financial institutions’ assets and liabilities during a crisis. Therefore micro-prudential regulation, which historically focused on a bank’s risk in isolation, is not adequate to contain the contagion effects of an extreme tail event.

However, the co-movement of downside risk for financial institutions during a financial crisis cannot be completely captured by Pearson’s correlation $\rho$, which only considers linear dependence but disregards nonlinear dependence between asset prices. It is widely accepted that the linear correlation $\rho$ is inadequate to characterize the full dependence structure of non-normally distributed random variables such as the daily returns of financial institutions. Figure 1.1 displays the scatter plot of real data for AIG and the Dow Jones Financial Index Return (a proxy for the financial system), and the simulated data from a bivariate normal distribution based on the first two moments of actual data. Even though their first two moments (including correlation $\rho$) are identical, their respective behaviors in the tail are quite different. The exceedance correlation and quantile dependence both lie outside the 95% confidence interval for a joint normal distribution, which implies the presence of non-

\[ 2\] Extreme tail event is defined as the days when the return of financial institutions or the market drop below a certain extreme threshold value $r_{it} < F_{r \tau}^{-1}(\tau)$.  
\[ 3\] Exceedance correlation of $X_1$ and $X_2$ (Ang & Chen 2001) is defined as $\text{corr}(X_1, X_2|X_1 \leq F_1^{-1}(u), X_2 \leq F_2^{-1}(u))$ in the left tail and $\text{corr}(X_1, X_2|X_1 \geq F_1^{-1}(u), X_2 \geq F_2^{-1}(u))$ in the right tail. Analogously, quantile dependence $P(X_2 \leq F_2^{-1}(u)|X_1 \leq F_1^{-1}(u))$ in the left tail and $P(X_2 \geq F_2^{-1}(u)|X_1 \geq F_1^{-1}(u))$ in the right tail.
linear dependence in the tail. Therefore, even though the average linear dependence strength $\rho$ between real and simulated data is controlled to be the same, their tail dependence is quite different as shown in the lower panels of Figure 1.1.

In other words, the co-movements between financial institutions and the market under conditions of distress (tail dependence) cannot be measured by their co-movements under normal times ($\rho$). As systemic risk mainly studies the extent to which extreme values tend to occur together, what really matters for the magnitude of systemic risk is the dependence between the tail risk of individual institutions and the financial system instead of the characteristics of the marginal distribution for individual firms.

There is an extensive and growing body of literature proposing alternative approaches to quantifying the systemic risk of financial institutions. Huang, Zhou and Zhu (2009) used Credit Default Swaps (CDS) and equity price correlations to construct a systemic risk indicator: Distress Insurance Premium (DIP), a measure computed under risk-neutral probability. Zhou (2010) studied systemic importance measures under the multivariate Extreme Value Theory (EVT) framework and found that the “too big to fail” argument is not always valid. Billio et al. (2010) used principal component analysis and a Granger-Causality test to measure the interconnectedness among the returns of hedge funds, banks, brokers and insurance companies. Balla, Ergen and Migueis (2012) recently used extreme value theory to investigate the extreme tail dependence among stock prices of US bank holding companies and found that it is the extreme tail dependence that makes a difference in the measurement of systemic risk. A good survey of systemic risk measures can be found in Bisias et al. (2012).

The most direct measure of systemic risk is simply the joint distribution of negative outcomes for a collection of systemically important financial institutions. Two leading metrics for return-based estimation of systemic risk are CoVaR, proposed by
Adrian and Brunnermeier (2011), and Marginal Expected Shortfall \((MES)\), proposed by Acharya et al. (2010), which estimate the contribution of a single institution to the overall systemic risk of the financial system during a tail event. More specifically, Acharya et al. (2010) defined CoVaR as the VaR of the financial system when an individual financial institution is in financial distress. They further defined an institution’s contribution to systemic risk of the financial system as the difference between CoVaR when the firm is, or is not, in financial distress. Acharya et al. (2010), Brownless and Engle (2011), among others, introduced the novel concepts of systemic risk measure (SRISK) and marginal expected shortfall \((MES)\) to estimate the systemic risk exposure of an institution. \(MES\) is the expected loss an equity investor would experience if the overall market is in the left tail. SRISK extends the \(MES\) in order to take into account both the liability and size of institutions. Over the past few years, dozens of research papers and media coverage\(^4\) have discussed, implemented and generalized these systemic risk measures. Hundreds of financial institutions in the world have been assessed for their systemic importance based on these measures\(^5\). Benoit et al. (2013) proposed a theoretical and empirical comparison of Marginal Expected Shortfall \((MES)\), Systemic Risk Measure (SRISK) and \(\Delta\)CoVaR. They assumed that the time varying correlation coefficient completely captures the dependence between the firm and market returns, and investigated under what conditions these different systemic risk measures converge in identifying systemic important financial institutions (SIFIs).

The notion of systemic risk is based on the interdependence of financial institutions and the financial system, especially in the context of extreme tail events. In this paper, we study these most widely used market-return-based systemic risk measures: \(MES\) and \(\Delta\)CoVaR in the framework of a wide range of copula models, which is very


\(^5\)For online computation of systemic risk measures, see the Stern-NYU’s V-Lab initiative at http://vlab.stern.nyu.edu/welcome/risk/.

4
straightforward in accommodating the nonlinear pattern of dependence structure. We aim to determine if these measures of systemic risk are consistent with the measures of tail dependence. Formally, the Lower Tail Dependence (LTD) is defined as

$$LTD = \lim_{u \to 0^+} P(X_2 \leq F_2^{-1}(u) | X_1 \leq F_1^{-1}(u))$$  \hspace{1cm} (1.1)$$

Where $F^{-1}(u)$ is the inverse CDF at quantile $u$. Therefore, we want to investigate if the higher value of dependence strength in the LTD between financial firms and the market indicates the higher systemic risk: $\Delta \text{CoVaR}$ or $\text{MES}$. Georg et al. (2012) investigated the dependence consistency of CoVaR mainly with respect to its statistical properties. They found that CoVaR based on a stress event $X \leq VaR_x(\alpha)$ is consistent with a dependence measure of correlation $\rho$. By contrast, the most widely used CoVaR proposed by AB (2011), which is conditional on stress event $X = VaR_x(\alpha)$, does not monotonically increase with the linear correlation $\rho$.

In contrast, this chapter provides a more thorough empirical study for a panel of 64 top US financial institutions over the period from January 2000 to December 2011. Our empirical analysis delivers the following main results: first, we found that the linear quantile regression estimation of $\Delta \text{CoVaR}$ proposed by AB (2011) is inadequate to completely take into account the non-linear contagion tail effect, which tends to underestimate systemic risk in the presence of lower tail dependence. Second, $\Delta \text{CoVaR}$, originally proposed by AB (2011), is in conflict with dependence measures in terms of correlation $\rho$, Kendall $\tau$ or lower tail dependence (LTD). Stronger dependence strength could instead lead to lower systemic risk, which contradicts the common view that higher dependence leads to higher systemic risk. In comparison, the modified version of $\Delta \text{CoVaR}^{\leq}$, based on the stress condition $X \leq VaR_x(\alpha)$, is more consistent with dependence measures. However, it tends to converge as dependence strength (correlation $\rho$ or Kendall $\tau$) is high during a financial crisis, when it is most desirable for supervision authorities to identify systemically important fi-
nancial institutions (SIFIs). MES, on the other hand, provides a moderately better response to the dependence structure if the heterogeneity of the marginal distribution can be eliminated. Third, $\Delta CoVaR^c$ is observed to have a strong correlation with tail dependence. In contrast, $MES$ is found to have a strong empirical relationship with firms’ conditional CAPM $\beta$. SRISK, however, provides further connection with firms’ level characteristics by accounting for information on market capitalization and liability. This stylized fact seems to imply that $\Delta CoVaR^c$ is more in line with the “too interconnected to fail” paradigm, and SRISK is more related to the “too big to fail” paradigm. In contrast, $MES$ is a compromise between these two paradigms. More specifically, $MES$ is a risk exposure measure. Firms’ specific marginal heterogeneity, such as volatility, and the dependence structure between firms and market jointly determine the value of $MES$. $\Delta CoVaR^c$, however, is only related to the dependence between firms and the market.

The remainder of this chapter is organized as follows: Section 2 introduces the definition of Delta Conditional Value at Risk ($\Delta CoVaR$) and its modification ($\Delta CoVaR^c$); and Marginal Expected Shortfall ($MES$) and its extension, systemic risk measure (SRISK). Then we discuss their estimation in the common framework of a copula model. Section 3 discusses the consistency of all these systemic risk measures with the dependence measures (linear correlation $\rho$, Kendall correlation and lower tail dependence). Section 4 describes the estimation strategy for both the marginal and copula models. Section 5 presents the data and empirical studies for a group of 64 top US financial institutions from January 2000 to December 2011. Section 6 concludes.
1.2 Methodology

In this section, we introduce a formal definition of systemic risk measures proposed by Adrian and Brunnermeier (2011) and Acharya et al. (2010). Let us assume a financial system composed of a large number of institutions. Denote by \( r_{mt} \) and \( r_{it} \) the daily return for the market, i.e. the financial system and firm \( i \) on time \( t \), respectively. The market return can be considered as the portfolio of all firms’ returns

\[
r_{mt} = \sum_{i=1}^{N} \omega_{it} r_{it}
\]

where \( \omega_{it} \) denotes the weight (market capitalization) of firm \( i \) in the portfolio at time \( t \).

1.2.1 Definitions

\( \Delta \) CoVaR

Paralleling the definition of Value at Risk (VaR), the conditional Value at Risk (CoVaR) is defined as the expected maximal loss of a certain portfolio at some confidence interval (\( \tau \) quantile) given another portfolio at the same time experiences expected maximal loss. Formally, the CoVaR of financial institution \( i \) corresponds to the Value at Risk of the financial system \( m \) conditioning on the occurrence of tail event \( r_{it} = VaR_{i}^{\tau} \) for the institution \( i \).

\[
Pr\left( r_{mt} \leq CoVaR_{m|i}(\tau, \tau^{*})|r_{it} = VaR_{i}^{\tau} \right) = \tau
\]

when \( \tau^{*} = \tau \), we simply denote \( CoVaR_{m|i}(\tau, \tau) = CoVaR_{\tau}^{m|r_{it}=VaR_{i}^{\tau}} \). The Value at Risk of the financial institution \( i \): \( VaR_{i}^{\tau} \) is defined as the \( \tau^{*} \) quantile of its loss
Further, AB (2011) proposed $\Delta CoVaR$ as the difference between $CoVaR_{i|\tau=VaR_{i}^\tau}$ when the financial institution $i$ is in distress and $CoVaR_{i|\tau=\text{Median}}^i$ when the financial institution $i$ is in the normal state.

$$\Delta CoVaR_{m|i} = CoVaR_{\tau|\tau=VaR_{i}^\tau} - CoVaR_{\tau|\tau=\text{Median}}^i$$

Therefore $\Delta CoVaR$ measures the increment of the VaR (difference between two conditional VaR), which aims to capture the contribution of a particular institution $i$ to the tail risk of financial system as a whole.

$$\Delta CoVaR \leq$$

Defining the financial distress being exactly at its VaR is arguably too restrictive. Girardi and Ergun (2012), among others, extended the definition of stress event to be below ($r_{it} \leq VaR_{it}(\tau)$) rather than exactly being at its VaR ($r_{it} = VaR_{it}(\tau)$), which allows for more severe distress event being further in the tail. Formally, the alternative $CoVaR$ (hereafter called $CoVaR_{\leq}$) can be defined as:

$$Pr\left(Y \leq CoVaR_{\tau} | X \leq VaR_{q}\right) = \tau$$

---

6Adrian and Brunnermeier (2011) further introduces “Exposure CoVaR”, which reverses the conditioning set to be the tail event for market return: $r_{mt} = VaR_{m}^\tau$. 

8
\[
Pr\left( Y \leq \text{CoVaR}_\tau, X \leq \text{VaR}_q \right) = \tau
\]  
(1.7)

\[
Pr\left( Y \leq \text{CoVaR}_\tau, X \leq \text{VaR}_q \right) = \tau q
\]  
(1.8)

where \( q(\tau) \) is the quantile defining the tail event of the financial institution (market). Therefore the alternative CoVaR at confident interval \( \tau \) can be solved from the following equation:

\[
\int_{-\infty}^{\text{CoVaR}_\tau} \int_{-\infty}^{\text{VaR}_q} f(x,y) \, dx \, dy = \tau q
\]

where \( f(x,y) \) is the joint density function of \( X \) and \( Y \).

By analogy, the modified version of Delta CoVaR: \( \Delta \text{CoVaR}^\leq \) can be specified as

\[
\Delta \text{CoVaR}^\leq \tau = \text{CoVaR}^\text{m|it} \leq \text{VaR}_{it}(\tau) - \text{CoVaR}^\text{i|it} \leq \text{Median}
\]  
(1.9)

Such a slight change of tail event from \( r_{it} = \text{VaR}_{it}(\tau) \) to \( r_{it} \leq \text{VaR}_{it}(\tau) \) seems to be trivial, but will prove to make a significant difference in both the estimation strategy and its statistical properties.

**MES**

Marginal Expected Shortfall (MES) measures the marginal contribution of firm \( i \) loss to systemic loss, or the average return of firm \( i \) on the \( \alpha\% \) worse days when the market as a whole is in the tail of its distribution. Formally, the coherent risk of system can be measured by its conditional Expected Shortfall (ES):

\[
ES_{m,t-1} = E_{t-1}(r_{mt} | r_{mt} < C) = \sum_{i=1}^{N} \omega_i E_{t-1}(r_{it} | r_{mt} < C)
\]  
(1.10)
where $C$ is the threshold value which defines the tail event when the market return exceeds it $r_{mt} < C$. The $MES$ of firm $i$ is defined as the partial derivative of the system’s aggregate risk (ES) with respect to the weight of firm $i$ in the market portfolio:

$$\omega_i$$

$$MES_{it} = \frac{\partial ES_{m,t-1}(C)}{\partial \omega_i} = E_{t-1}(r_{it}|r_{mt} < C)$$

(1.11)

This measures the marginal contribution of firm $i$ to the systemic risk. The higher the value of $MES$, the higher the individual contribution of the firm $i$ to the risk of the financial system as a whole. Brownless and Engle (2011) recently show that the estimation of $MES$ can be reduced into the estimations of three components such as volatility, correlation and tail expectation. The linear market model of Brownless and Engle (2011) can be summarized as:

$$r_{mt} = \sigma_{mt}\epsilon_{mt}$$

(1.12)

$$r_{it} = \sigma_{it}\rho_{it}\epsilon_{mt} + \sigma_{it}\sqrt{1 - \rho_{it}^2}\xi_{it}$$

(1.13)

$$\left(\epsilon_{mt}, \xi_{it}\right) \sim F$$

(1.14)

where the innovation process $\xi_{it}$ is uncorrelated but not independent with $\epsilon_{mt}$. Therefore the conditional $MES$ can be expressed as a function of the firm’s return volatility $\sigma_{it}$, its correlation with the market return $\rho_{it}$ and the co-movement of the tail distribution:

$$MES_{it}(C) = \sigma_{it}\rho_{it}E_{t-1}\left(\epsilon_{mt}|\epsilon_{mt} < \frac{C}{\sigma_{mt}}\right) + \sigma_{it}\sqrt{1 - \rho_{it}^2}E_{t-1}\left(\xi_{it}|\epsilon_{mt} < \frac{C}{\sigma_{mt}}\right)$$

$$= \beta_{it}E_{t-1}\left(r_{mt}|r_{mt} \leq C\right) + \sigma_{it}\sqrt{1 - \rho_{it}^2}E_{t-1}\left(\xi_{it}|\epsilon_{mt} < \frac{C}{\sigma_{mt}}\right)$$

(1.15)
where $\beta_{it} = \rho_{it} \frac{\sigma_{it}}{\sigma_{mt}}$ measures the risk exposure of financial institution $i$ to the market based on CAPM model. Suppose that the dependence structure between the firm $i$ and the market return is completely captured by the time varying conditional correlation $\rho_{it}$ (or $\xi_{it} \perp \epsilon_{mt}$), then the second term of the above equation becomes zero, and the first component completely describes the systemic risk exposure of financial institution $i$.\textsuperscript{7} Therefore, the second component is dedicated to capture the nonlinear dependence between the firms $i$ and the market. Brownless and Engle (2011) proposed a nonparametric kernel approach to estimate the tail expectation of return between firm $i$ and market.

**SRISK**

Brownless and Engle (2011) and Acharya, Engle and Richardson (2012) extended MES and proposed the systemic risk measures (SRISK) to account for the level of firms’ characteristics (size, leverage, etc), which corresponds to the expected capital shortfall of a financial firm if we have another financial crisis as a whole. More formally, Brownless and Engle (2011) defined SRISK as\textsuperscript{8}

$$SRISK_{it} = \max\left(0; \ kDebt_{it} - (1 - k)(1 - LRMES_{it}) Equity_{it}\right) \quad (1.16)$$

where $k$ is the prudential capital ratio which we take as 8%, $Debt_{it}$ is the quarterly book value of total liabilities, $Equity_{it}$ is the market value of equity (market capitalization) today, and $LRMES_{it}$ is long term Marginal Expected Shortfall (MES), which corresponds to the expected drop in equity return conditioning on the market falling by more than 40% within the next six months.\textsuperscript{9} Brownless and Engle (2011)

\textsuperscript{7}Benoit et al. (2013) compared systemic risk measures $\Delta CoVaR$ and $MES$ by assuming that the conditional correlation $\rho$ completely captures dependence between firms and market.

\textsuperscript{8}By analogy, a short-run SRISK can also be defined as $SRISK_{it} = \max(0; kDebt_{it} - (1 - k)(1 - MES_{it}) Equity_{it})$ by replacing long term $MES$ with short run $MES$.

\textsuperscript{9}For more details, see Acharya, Engle and Richardson (2012).
obtained $LRMES_u$ by implementing a simulation exercise.\textsuperscript{10} Therefore $SRISK$ can be estimated by scaling up $MES$ to account for the leverage and size of firms.

1.2.2 Copula Based Estimation of $\Delta CoVaR$ and $MES$

Since both systemic risk measures $\Delta CoVaR$ and $MES$ are concerned with the joint distribution of downside risk, in this paper, we study CoVaR and $MES$ in the framework of the copula model, which is well suited to accommodate the non-linear dependence of tail risk between financial institutions and the market. Copula based multivariate models provide the flexibility to specify the models for the dependence structure (copula) of multivariate random variables separately from their marginal distribution. The essence of copula based model for CoVaR and $MES$ is to model the local dependence between the lower quantile of two assets.

**Theorem 1 (Sklar)(1959)** Let $G$ be a joint distribution function with $n$ marginals $F_i$. Then there exists an $n$ dimension copula $C$ such that

$$G(x_1, x_2, \cdots, x_n) = C\left(F_1(x_1), F_2(x_2), \cdots, F_n(x_n)\right)$$

If all margins $F_i$ are continuous, then the copula $C$ is unique.

In contrast to the traditional dependence measure like Pearson’s correlation $\rho$ and Kendall’s $\tau$, the copula model is capable of measuring the whole dependence structure including tail dependence between random variates.

\textsuperscript{10}For the threshold value $C = 2\%$, Acharya, Engle and Richardson (2012) proposed an approximation to the LRMES without simulation exercises. Namely, $LRMES \approx 1 - exp(-18 \times MES_{it})$. 

12
Definition 1: Tail Dependence

Let \( X_1 \) and \( X_2 \) be two random variables with CDF \( F_1 \) and \( F_2 \). Then the lower tail dependence (LTD) coefficient of two random variables \((X_1, X_2)\) is defined as

\[
LTD(X_1, X_2) = \lim_{u \to 0^+} \Pr\left(X_2 < F^{-1}_2(u) \mid X_1 < F^{-1}_1(u)\right) = \lim_{u \to 0^+} \frac{C(u, u)}{u} \tag{1.18}
\]

Analogously, the upper tail dependence (UTD) coefficient of two random variables \((X_1, X_2)\) is defined as

\[
UTD(X_1, X_2) = \lim_{u \to 1^-} \Pr\left(X_2 > F^{-1}_2(u) \mid X_1 > F^{-1}_1(u)\right) = \lim_{u \to 1^-} \frac{1 - 2u + C(u, u)}{1 - u} \tag{1.19}
\]

In the following, a wide range of copula models are to be used to estimate the systemic risk measures \(\Delta CoVaR\) and \(MES\), and we will study their relationship with the tail dependence between the financial institutions and market.

\[\Delta CoVaR\] Estimation

If the joint distribution \(F(x, y)\) has a continuous marginal distribution, Sklar’s Theorem (1959) stated that there exists a unique copula function such that any bivariate distribution can be represented by the combination of their marginal distributions in a certain form of copula function:

\[
F(x, y) = C\left(F_x(x), F_y(y); \theta\right) = C(u, v; \theta) \tag{1.20}
\]

where \(u\) and \(v\) are the marginal distributions \(F_x(x)\) and \(F_y(y)\) of \(x\) and \(y\) respectively. Differentiating \(C(u, v; \theta)\) with respect to \(u\), we can obtain the conditional distribution
of $y$ given $x$ (See the proof in the Appendix):

$$F(y|x) = \frac{\partial C(u, v; \theta)}{\partial u} = C_1(u, v; \theta) = C_1(F_x(x), F_y(y); \theta)$$ (1.21)

where $F(y|x)$ denotes the conditional distribution of $y$ given $x$ and $C_1(u, v; \theta)$ represents the partial derivative of the copula with respect to the first argument.

By defining $\tau = F(y|x)$ and inverting $C_1$ with respect to the second argument can derive the marginal distribution of $y$: $F_y(y)$ as

$$F_y(y) = C_1^{-1}(F_x(x), \tau; \theta)$$ (1.22)

Let $x = VaR_x(\tau)$, and $y = CoVaR_{y|x}$. The solution for CoVaR in the context of the copula framework can be explicitly expressed as:

$$CoVaR_{y|x} = F_y^{-1}\left( C_1^{-1}\left( F_x(VaR_x(\tau)), \tau, \theta \right) \right)$$ (1.23)

The representation in Equation (1) shows the attractiveness of the copula approach for modeling the nonlinear dependence between the lower quantiles of $x$ and $y$. A wide range of CoVaR can be obtained by combining different marginal distributions $F(.)$ with different copula functional forms $C(.,.)$. Since we aim to study the nonlinear dependence of tail risk, we restrict our attention to the different specifications of copula models to see if the nonlinear dependence structure in the tail risk plays a role in the estimation of CoVaR.

**Proposition 2.1**

Suppose $X$ and $Y$ are the returns of two assets with Gaussian joint distribution (Gaussian Margins and Gaussian Copula model $C(u, v; \theta)$). Then the closed form solution of $CoVaR_{y|x}$ can be expressed as (See the proof in
Appendix 1):

\[ CoVaR^{y|x}_x = \rho \frac{\sigma_y}{\sigma_x} VaR_x(\tau) - \rho \frac{\sigma_y}{\sigma_x} \mu_x + \sigma_y \sqrt{1 - \rho^2} \Phi^{-1}(\tau) + \mu_y \] (1.24)

\[ \Delta CoVaR^{y|x} = CoVaR^{y|x} - CoVaR^{y|x=median} = \rho \frac{\sigma_y}{\sigma_x} \left( VaR_x(\tau) - VaR_x(0.5) \right) = \rho \sigma_y F^{-1}_\epsilon(\tau) \]

where \( \rho \) denotes the linear Pearson correlation between \( x \) and \( y \), \( \sigma_x \) and \( \sigma_y \) represent volatility of \( x \) and \( y \) respectively, \( \mu_x \) denotes the mean of \( x \), \( \Phi^{-1} \) denotes the inverse of standard normal distribution, and \( \epsilon \) is the standardized residual with \( \epsilon \sim N(0, 1) \).

It is very common to estimate CoVaR with a Gaussian joint distribution as normality is the workhorse distribution assumption in the financial literature. Under the Gaussian joint distribution, the calculation of CoVaR is very straightforward and become a trivial issue. The dependence structure between tail risks of two assets is linear and can be completely captured by the second moment of the joint distribution \( \rho \frac{\sigma_y}{\sigma_x} \). It is noteworthy that the linear dependence coefficient \( \rho \frac{\sigma_y}{\sigma_x} \) is not quantile specific, which means that the dependence structure in the lower tail is exactly the same as that in any other part of joint distribution. However, this statistical property is only specific to the joint Gaussian distribution. In general, there is no explicit reason to justify why the dependence of the tail risk at different quantiles should be identical. More often than not, the contribution of a financial institution to the systemic risk of financial market when it is in bankruptcy \( (x = VaR_x(\tau)) \) should be quite different from that when it is in the normal state \( (x = VaR_x(0.5)) \).

AB (2011) employed a linear quantile regression model to estimate CoVaR as
quantile regression is robust to the unknown distribution. In this paper, we consider the quantile regression of the financial market returns $r_{mt}$ on a particular institution’s return $r_{it}$ at the $\alpha$ quantile

$$Q_{r_{mt}}(\tau) = \alpha(\tau) + \beta(\tau)r_{it}$$ (1.25)

The CoVaR of financial market, conditional on the financial institution being in distress, can be defined as:

$$CoVaR_{m|VaR_{it}(\alpha)} = \alpha(\tau) + \beta(\tau)VaR_{it}(\alpha)$$ (1.26)

Therefore the systemic risk measure $\Delta CoVaR$ put forward by AB (2011) can be constructed as:

$$\Delta CoVaR(\tau) = \beta(\tau)\left(VaR_{it}(\tau) - VaR_{it}(0.5)\right)$$ (1.27)

Figure 1.2 displays the simulation result for the $\Delta CoVaR$ estimated by various copula based models and quantile regression models under different data generating processes. In each panel of the figure, the data are generated by corresponding copula model shown in the title. All marginal distributions are set to be the standard normal distribution except for student-t copula where the margin is set to student t with $df = 5$. Therefore the discrepancy of $\Delta CoVaR$ estimation can only be attributed to the discrepancy of dependence structure in tail risk. As we can see, when the data are generated from a joint normal distribution, quantile based and copula based estimation of $\Delta CoVaR$ fit quite well. However, as the data are simulated from a copula model with positive lower tail dependence in the downside risk (Clay-\footnote{AB (2011) extended the quantile regression by including some additional state variables such as VIX, liquidity spread, etc. In this paper, we consider a slightly different approach to facilitate the comparison with the copula based model.}}
ton, Rotated Gumbel), quantile based estimation of systemic risk $\Delta \text{CoVaR}$, which was the most widely used estimation strategy in the recent literature, is consistently underestimated, compared with copula based estimation. This fact is not surprising, considering that the linear quantile regression still assumes linear local dependence between downside risk, which is not adequate to estimate CoVaR when the left tail of the distribution exhibits positive and nonlinear dependence between downside risk.

$\Delta \text{CoVaR} \leqslant \text{Estimation}$

With the stress condition being $X \leqslant \text{VaR}_x(\tau)$, the modified version of $\text{CoVaR}$ proposed by Girardi and Ergun (2012) can also be analyzed under the framework of the copula model.

Starting from the above equation

$$\int_{-\infty}^{\text{CoVaR}_x} \int_{-\infty}^{\text{VaR}_q} f(x, y) dxdy = \tau q$$

(1.28)

The joint density function $f(x, y)$ can be substituted by the cross product of marginal density and copula density $f(x, y) = \frac{\partial^2 C(u, v)}{\partial u \partial v} f(x) f(y)$. By a simple manipulation, we have:

$$\int_{-\infty}^{\text{CoVaR}_x} \int_{-\infty}^{\text{VaR}_q} \frac{\partial^2 C(u, v)}{\partial u \partial v} f(x) f(y) dxdy = \tau q$$

$$\int_0^{F_y(\text{CoVaR}_x)} \int_0^{F_x(\text{VaR}_q)} \frac{\partial^2 C(u, v)}{\partial u \partial v} dudv = \tau q$$

(1.29)
Let $q = \tau$, we can solve $CoVaR_y|_{x \leq VaR_x(\tau)}$ numerically from Equation (2). Analogously, let $q = 0.5$, $CoVaR_y|_{x \leq VaR_x(0.5)}$ can be obtained as well. Therefore, $\Delta CoVaR^\leq$ can be estimated as

$$\Delta CoVaR_{\tau} = CoVaR_{\tau}|_{x \leq VaR_x} - CoVaR_{\tau}|_{x \leq VaR_{0.5}}$$

Girardi and Ergun (2012) compared the bivariate skewed t distribution (Bauwens and Laurent (2005)) with bivariate Gaussian distribution to model the density function $f(x, y)$, and they found that the former joint density outperformed the latter in the estimation of CoVaR. But they fail to explain whether this performance improvement comes from the difference in the specification of marginal distribution or the dependence structure in the tail. In our paper, we only focus on the discrepancy of dependence structure and eliminate the effect of marginal distribution as we believe that what really matters in the estimation of systemic risk is the dependence structure (especially tail dependence) rather than marginal characteristics.

Figure 1.3 illustrates the $CoVaR^\leq$ and $\Delta CoVaR^\leq$ under two copula models with different dependence structures in the tail of the distribution. The left panel present the Clayton copula joint distribution with lower tail dependence, while the right panel displays the joint distribution of the Gumbel copula with upper tail dependence. Suppose that bank $i$ financial condition moves from its normal state (at median) to its financial distress (at 1% quantile). We want to investigate what is the change of 1% quantile for the financial system. 8000 random draws are simulated from two bivariate joint distributions. The parameters of each copula model are chosen to provide a linear correlation of random variates equal to 0.8. It is not surprising that the systemic risk measure $\Delta CoVaR^\leq$ under the Clayton copula is larger than that under the Gumbel copula because the former joint distribution is characterized
by the larger lower tail dependence, which results in stronger co-movement in the tail risk of financial institutions and the market.

**MES Estimation**

Analogous to the calculation of CoVaR, the value of MES is associated with the joint distribution of downside risk between the returns of the market and financial firms.

**Proposition 2.2**

Suppose $X$ and $Y$ are the returns of two assets with the marginal distributions being $X \sim F_x$ and $Y \sim F_y$. The joint distribution of $X$ and $Y$ are defined by a parametric copula $C(u, v; \theta)$. The closed form solution of marginal expected shortfall $MES = -E(X|Y < VaR_y(\tau))$ can be expressed as (see the Proof in Appendix 2):

$$MES(\tau) = -\frac{1}{\tau} \int_0^1 F^{-1}_x(u) \frac{\partial C(u, \tau; \theta)}{\partial u} du \tag{1.30}$$

where $\tau$ is the quantile percentage for the distribution of $Y$ which defines the tail event $Y < VaR_y(\tau)$.

Figure 1.4 compares the copula-based estimation of $MES$ with the nonparametric kernel estimation following Brownless and Engle (2011).\textsuperscript{12} As we can see, Brownless–Engle’s approach provides a good approximation to the estimation of $MES$ even when the data are generated by student-t copula (symmetrical tail dependence) or Rotated Gumbel Copula characterized by lower tail dependence. This simulation exercise seems to justify the nonparametric kernel estimation in capturing

\textsuperscript{12}For the details of nonparametric kernel estimation of $MES$, see Brownless and Engle (2011)
nonlinear tail dependence.

Following the models proposed by Brownless and Engle (2011), the return of market and financial institution $i$ is specified as

$$r_{m,t} = \sigma_{m,t}\epsilon_{m,t} \quad r_{i,t} = \sigma_{i,t}\epsilon_{i,t}$$  \hspace{1cm} (1.31)

Proposition 2.2 implies that the marginal expected shortfall of financial institution $i$ can be written as:

$$MSE_{i,t-1}(K) = -E_{t-1}(r_{i,t}|r_{m,t} \leq K)$$

$$= -\sigma_{i,t}E_{t-1}\left(\epsilon_{i,t}|\epsilon_{m,t} \leq \frac{K}{\sigma_{m,t}}\right)$$

$$= -\frac{\sigma_{i,t}}{\tau_t} \int_0^1 F_{\epsilon_{i,t}^{-1}}(u) \frac{\partial C(u, \tau)}{\partial u} du$$

where the threshold value $K = r_m(\tau)$ is the unconditional quantile at $\tau\%$ of the market return $r_{mt}$, which defines the tail event when the market return exceeds $K$ (namely, $r_{m,t} \leq K$). $\tau_t = F_{\epsilon_{m,t}^{-1}}(\frac{K}{\sigma_{m,t}})$ is the CDF of the market return evaluated at the scaled threshold value $\frac{K}{\sigma_{m,t}}$. $C(u, \tau)$ is the copula function between the innovation process $\epsilon_{i,t}$ and $\epsilon_{m,t}$.

It is noteworthy that the value of $MSE$ relies not only on the dependence structure determined by copula function $C(u, v; \theta)$, but also on the marginal characteristics such as volatility $\sigma_{it}$ and $F_{\epsilon_{i,t}^{-1}}$. In other words, two firms with the same dependence structure ($C(u, v; \theta)$) with market but with different volatility ($\sigma_{it}$) would otherwise be considered as similarly systemically important, but will be treated differently by the measures of $MSE$. This has been highlighted by Archaya et al. (2012) in the case of the joint Gaussian distribution where the linear correlation $\rho$ captures the full dependence structure. Therefore, unlike $\Delta CoVaR$ or $\Delta CoVaR^{\leq}$, which solely depend on dependence structure determined by the copula model, $MSE$ is a firm-specific...
value, which is determined by the dependence structure between firms and market as well as the marginal characteristics of financial firms such as volatility $\sigma_{it}$. Thus, it is possible to observe a situation where a financial institution is more systemically risky according to the $MES$ measure (higher volatility), but could be less risky based on the ranking of $\Delta CoVaR^{c}$ (lower dependence with the market).\textsuperscript{13} Although $MES$ is associated with firms’ specific volatility, it still fails to account for some important firm level characteristics such as size and leverage which related directly to the distress condition of financial institutions.

**SRISK Estimation**

The discussion so far disregards the question of causality. Tail dependence only describes the extent of interconnectedness between the tail risks of financial firms and the market, but fails to indicate whether the crisis happens because the firms fail, or conversely, firms fail because of the crisis\textsuperscript{14}. Therefore, tail dependence alone cannot accurately measure systemic risk. It is likely that a small, unlevered firm (with higher tail dependence on the market because of the fragility of a small firm per se) can appear more “dangerous” for the financial system than a big, levered one (with lower tail dependence on the market because of robustness of the big firm per se).

Brownless and Engle (2011), Acharya, Engle and Richardson (2012) extended $MES$ and proposed the systemic risk measures (SRISK) to account for the levels of firms’ characteristics (size, leverage, etc.). We have shown in Section 2.1.4 that SRISK is

\textsuperscript{13}Benoit et al. (2013) discussed the condition when systemic risk measure $\Delta CoVaR$ and $MES$ are consistent or converge under the assumption of joint normal distribution, where $\Delta CoVaR$ and $MES$ have a clean closed form solution.

\textsuperscript{14}Acharya, Engle and Richardson (2012) argued that both of them are jointly endogenous variables.
a linear function of the long run $MES$, liability and market capitalization

$$SRISK_{it} = \max \left( 0; \ kDebt_{it} - (1 - k)(1 - LRMES_{it})Equity_{it} \right)$$ (1.32)

In the followings, we elaborate on the simulation procedure and illustrate how to obtain $LRMES_{it}$ in the framework of the copula model.

The following five step procedures are implemented for each “market-institution” pair.

Step 1. Draw $S$ sequences of length $h$ of pairs of marginal distribution $(u_{it}, v_{mt})$ from the parametric copula model $C(u, v; \theta_{it})$ for the innovation series $\epsilon_{i,t}$ and $\epsilon_{m,t}$

$$\left( u_{it}^s, v_{mt}^s \right)_{t=T+1}^{T+h}, \text{ for } s = 1, 2, \ldots, S$$ (1.33)

Step 2. Obtain the sequence $(\epsilon_{i,t}^s, \epsilon_{m,t}^s)_{t=T+1}^{T+h}$ by setting $\epsilon_{i,t}^s = F_i^{-1}(u_{it}^s)$ and $\epsilon_{m,t}^s = F_m^{-1}(v_{mt}^s)$, where $F_i$ ($F_m$) is the empirical marginal distribution for asset $i$ (market).

Step 3. Obtain the sequence of firm and market returns by setting $r_{i,t}^s = \sigma_{i,t} \epsilon_{i,t}^s$ and $r_{m,t}^s = \sigma_{m,t} \epsilon_{m,t}^s$, where $\sigma_{i,t}$ ($\sigma_{m,t}$) is the forecast standard deviation by GJR-GARCH model for asset $i$ (market).

Step 4. Calculate the simulated cumulative return of firm $i$: $R_{i,T+1:T+h}^s$ (analogously for the market return $R_{m,T+1:T+h}^s$) relying on the properties of logarithm returns

$$R_{i,T+1:T+h}^s = \exp \left( \sum_{k=1}^{h} r_{i,T+k}^s \right) - 1$$ (1.34)

Step 5. The long run $MES$ ($LRMES_{i,T+1:T+h}$) can be obtained according to Acharya et al. (2010)$^{15}$

$$MES_{i,T+1:T+h} = \frac{\sum_{s=1}^{S} R_{i,T+1:T+h}^s \mathbb{I}[R_{i,T+1:T+h}^s < C]}{\sum_{s=1}^{S} \mathbb{I}[R_{m,T+1:T+h}^s < C]}$$ (1.35)

$^{15}$Copula based estimation of $MES$ can be applied here as well
where $C$ represents the threshold value defining the systemic event.

1.3 On the Dependence Consistency of Systemic Risk: $\Delta CoVaR$ and $MES$

It still remains to be determined whether the proposed return-based systemic risk measures, $\Delta CoVaR$ and $MES$, are consistent with the dependence measures. Intuitively, a good metric of systemic risk measure should be able to reflect the fact that stronger dependence strength (higher value of $\rho$ or Kendall $\tau$) between market and firms should result in a higher value of systemic risk. In addition, given the identical value of average dependence strength ($\rho$ or Kendall $\tau$), those dependence structures with lower tail dependence (e.g., Student-t Copula or Rotated Gumbel Copula) should indicate higher value of systemic risk than those without tail dependence (Normal copula). In other words, a higher value of dependence strength in the lower tail of distribution should lead to a higher value of systemic risk.

Figure 1.5 compares the $\Delta CoVaR^=\Delta CoVaR^{\leq}$ and $MES$ estimated by a Normal Copula (without tail dependence) with those estimated by Student-t Copula (with symmetric tail dependence). The marginal distributions for all copula models are set to be Student-$t^{16}$ with $df = 5$. Having eliminated the effect of the marginal distribution, the discrepancy of systemic risk measures, therefore, can only be attributed to the effect of dependence structures.

It is surprising to find that the behavior of $\Delta CoVaR$ becomes quite strange. First, $\Delta CoVaR$ does not monotonically increase with the correlation $\rho$. It declines

---

16When I changed the degree of freedom, or control all marginal distributions to be skewed t distribution instead, the main result remain unchanged.
in the end when the strength of dependence, the correlation $\rho$, approaches a high value. Second, the lower tail dependence implied by the Student t copula does not necessarily result in a higher value of systemic risk than the Normal copula without tail dependence. When the correlation $\rho$ is sufficiently large, $\Delta CoVaR$ estimated by the Normal copula becomes significantly larger than that estimated by student t copula.

Figure 1.6 provides further evidence that $\Delta CoVaR$ is not consistent with dependence measures. As the average co-movement of data becomes stronger (Kendall $\tau$ increases), the $\Delta CoVaR$ estimated by Normal copula (without tail dependence) or Gumbel Copula (with only upper tail dependence) approaches the highest value, which contradicts the common view that stronger lower tail dependence implied by Clayton or Rotated Gumbel copula should yield a higher value of systemic risk $\Delta CoVaR$. Furthermore, the value of $\Delta CoVaR$ begins to decline when the co-movement of data becomes even stronger (Kendall $\tau$ increases), which again conflicts with the commonly-held notion that higher interconnectedness between the firm and the market should indicate a larger value of systemic risk $\Delta CoVaR$.

Figure 1.7 displays various systemic risk measures against tail dependence (either lower or upper). Again it shows that $\Delta CoVaR$ fails to provide a consistent measure of systemic risk. When tail dependence (X-axis) is sufficiently high, upper tail dependence (e.g., Gumbel Copula), on the contrary, implies higher systemic risk $\Delta CoVaR$ than lower tail dependence (e.g., Rotated Gumbel Copula). Furthermore, $\Delta CoVaR$ does not monotonically increase with tail dependence, as the left panel of Figure 1.7 illustrates. All of these odd behaviors of $\Delta CoVaR$ estimation cast into doubt its reliability in measuring systemic risk.

In contrast, $\Delta CoVaR^c$ and $MES$ are much more consistent with dependence measures. In other words, stronger lower tail dependence, implied by the Rotated
Gumbel and Student t copula, leads to higher systemic risk without exception. In addition, the values of systemic risk measure, \( \Delta CoVaR \) and \( MES \), both monotonically increase with the correlation \( \rho \), Kendall \( \tau \) or tail dependence of the data generating process. These two facts seem to provide strong support for the reliability of using \( \Delta CoVaR \) and \( MES \), instead of \( CoVaR \), to measure systemic risk. Finally, \( MES \) seems to provide a better response to the dependence measures than \( \Delta CoVaR \) in the simulation.\(^{17}\) As the correlation \( \rho \), Kendall \( \tau \) or tail dependence attains a high value, the systemic risks measured by \( \Delta CoVaR \) seem to converge and fail to detect the discrepancy of dependence structure, which implies that \( \Delta CoVaR \) would fail to identify the systemically important financial institutions (SIFIs) during financial crisis, when the correlation or tail dependence between financial institutions and market as a whole is high. By comparison, \( MES \) works much better to detect the discrepancy of dependence structures even when the dependence strength attains a high value. For instance, when Kendall \( \tau = 0.8 \), \( MES \) can still discriminate between the value estimated by normal and that estimated by Rot-Gumbel Copula, as the right panel of Figure 1.6 displays.

Figure 1.6 and 1.7 further show that the upper tail dependence also affects the estimation of systemic risk, which mainly concerns downside risk co-movement in the lower tail of distribution. This provides evidence that we cannot ignore what happens in the upper tail, even if our interest lies primarily in the lower tail\(^{18}\) (\( \Delta CoVaR \) and \( MES \) are both estimated based on the joint distribution on the lower tail). The upper tail dependence increases the market’s and firms’ propensity to flourish together, representing abnormal profit and reward in economic prosperity, which should dis-

\(^{17}\)As we discussed before, \( MES \) is not only associated with dependence structure, but also related to firm specific characteristics such as volatility. In the simulation, we eliminate the effect of marginal distribution (controlling volatility of the data generating processes to be identical) and concentrate only on the difference of dependence structure.

\(^{18}\)Distaso, Fernandes and Zikes (2012) pointed out that the variation in the upper tail dependence may affect the estimation of the conditional lower tail dependence and vice versa. Therefore they employ the symmetrized Joe-Clayton copula rather than focusing on the lower tail dependence by means of either the clayton or the rotated Gumbel copula.
count systemic risk during economic recession. In other words, when we evaluate
the systemic risk of a financial institution, we should not restrict our attention to
what is happening during financial turbulence (lower tail), but also keep an eye on
its performance during economic prosperity (upper tail).

Comparing the definition of $\Delta CoVaR^c$ with $MES$ in detail reveals that their
conditioning events are completely different. More generally, $\Delta CoVaR^c$ measures the
sensitivity of market return with respect to firms’ return, which eventually reduces
to the estimation of tail risk for market return.\footnote{Adrian and Brunnermeier (2011) further
introduced “Exposure CoVaR”, which reverse the conditioning information set to be the tail event for market return.} Therefore, the heterogeneity of
marginal distribution is not the major issue for the estimation of $\Delta CoVaR^c$. Instead
what really matters is the dependence structure between market and firms return. On
the other hand, $MES$ is a risk exposure measure, which depends on the linear projec-
tion of firm return onto market return. Therefore, firm-specific marginal distribution
characteristics such as volatility, tail thickness and skewness, all make a difference in
the estimation of $MES$. In other words, the estimation of $MES$ is determined by
the dependence structures between market and firms as well as the heterogeneity of
marginal characteristics for firms.

\section{1.4 Estimation}

\subsection{1.4.1 Modeling the Marginal Dynamics}

Even though we concentrate on the dependence structure of tail risk, this is by
no means to imply that marginal distribution is of no importance. As demonstrated
by Fermanian and Scaillet (2005), misspecification of the marginal distribution may
lead to spurious results for the dependence measure estimation. Furthermore, the
measures of systemic risk (CoVaR or MES) fundamentally depend on the estimation of marginal VaR or Expected Shortfall (ES) for either the market or firms’ return. Therefore we must first model the conditional marginal distributions.

The time series of equity data usually exhibit time varying volatility and heavy-tailedness, we model each marginal series \( i \) for simplicity by a univariate AR(1) and GJR-GARCH(1,1,1) model:

\[
Y_{i,t} = \phi_0^i + \phi_1^i Y_{i,t-1} + e_{i,t} \quad \text{where} \quad e_{i,t} = \sigma_{i,t} \epsilon_{i,t} \quad \text{and} \quad \epsilon_{i,t} \sim iid(0, 1) \quad (1.36)
\]

\[
\sigma_{i,t}^2 = \omega^i + \alpha^i e_{i,t-1}^2 + \gamma^i e_{i,t-1}^2 I(e_{i,t-1} < 0) + \beta^i \sigma_{i,t-1}^2 \quad (1.37)
\]

where \( I(e_{i,t-1} < 0) \) captures the leverage effect in which volatility tends to increase more with negative shocks than positive ones. We estimate the model using QML which guarantees the consistency of parameter estimates as long as the conditional variance is correctly specified.

After constructing the conditional mean and volatility model, the estimated standardized residuals can be specified as

\[
\hat{\epsilon}_{i,t} = \frac{Y_{i,t} - \hat{\mu}_{i,t}}{\hat{\sigma}_{i,t}} \quad (1.38)
\]

where \( \hat{\mu}_{i,t} \) is the estimated conditional mean \( \hat{\mu}_{i,t} = \hat{\phi}_0^i + \hat{\phi}_1^i Y_{i,t-1} \), and \( \hat{\sigma}_{i,t} \) is the estimated conditional volatility following the above GJR-GARCH(1,1,1) model. Given the distribution function \( F_{\epsilon}^i \) for \( \epsilon_{i,t} \), the conditional \( \tau \) quantile VaR and ES for each time series \( i \) can be computed as

\[
\hat{VaR}_{i,t}^\tau = \hat{\mu}_{i,t} + \hat{\sigma}_{i,t} F_{\epsilon}^{i-1}(\tau) \quad (1.39)
\]
\[ \hat{ES}_t^i(C) = \hat{\mu}_{i,t} + \hat{\sigma}_{i,t} E(\epsilon_{i,t} | \epsilon_{i,t} \leq \frac{C}{\hat{\sigma}_{i,t}}) \]  \hspace{1cm} (1.40)

It is generally agreed that financial time series are fat tailed and asymmetric. In order to account for both skewness and kurtosis in the estimation of the univariate distribution, the standardized innovations \( \epsilon_{i,t} \) are assumed to have a univariate skewed t distribution \(^{21}\): \( \epsilon_{i,t} \sim f(\epsilon_{i,t}; \lambda_{i,t}, \nu_{i,t}) \), where \( f \) denotes the pdf of the skewed t distribution (Hansen (1994), see also Jondeau and Rockinger (2003)), with \( \nu_{i,t} \) being the degrees of freedom and \( \lambda_{i,t} \) being the asymmetry parameter.

Figure 1.8 displays an example of the fitted parametric estimates of skewed t distribution for Market Index (the Dow Jones Financial Index) and one financial institution in our sample: AIG.\(^{22}\) As we can see, the fitted density seems to provide a reasonable fit to the empirical histogram for both market index and daily return of AIG. A QQ plot shows that most of empirical observations can be captured by the estimated skewed t distribution. Having modeled the marginal distribution of standardized residuals \( \epsilon_{i,t} \), the estimated probability integral transformations can be specified as\(^{23}\)

\[ \hat{U}_{i,t} = \hat{F}_{\text{skew},t}(\hat{\epsilon}_{i,t}; \hat{\nu}, \hat{\lambda}) \]  \hspace{1cm} (1.41)

\(^{21}\)Girardi and Ergun (2012) found that the estimation of CoVaR improves in its performance when accounting for both skewness and kurtosis by assuming a skewed t distribution for margins.

\(^{22}\)To save space, we only display the fitted estimates of the skewed distribution for AIG. The result is quite similar for the daily return of most of other financial institutions.

\(^{23}\)In the robustness check, we estimate the marginal distribution \( \hat{U}_{i,t} = \hat{F}_{i,t}(\epsilon) \) by empirical distribution function EDF, but the main results below remain unchanged.
1.4.2 Copula Models Estimation

As we have discussed before, the systemic risks are all concerned with the joint distribution of tail risk, which relies on the dependence structure of the innovation process. Therefore we need to estimate a joint distribution that allows us to capture the possible non-linear dependence across the innovation processes. A straightforward approach is to use a copula model to describe the joint distribution of innovations. With the estimation of probability integral transformations $\hat{U}_{i,t}$, the copula model for the bivariate time series of the standardized residuals can be constructed as:

$$
\epsilon_t = [\epsilon_{it}, \epsilon_{jt}]^T \sim F_{\epsilon t} = C_{t}(U, V; \theta)
$$

(1.42)

where $U = F_i(\epsilon_{it})$ and $V = F_j(\epsilon_{jt})$

(1.43)

Assuming twice differentiability of the conditional joint distribution, the copula model as well as the conditional marginal distribution yields the decomposition for the conditional joint density function:

$$
f(\epsilon_{it}, \epsilon_{jt}) = f_i(\epsilon_{it})f_j(\epsilon_{jt})c(u, v; \theta)
$$

(1.44)

where $u = F_i(\epsilon_{it})$ and $v = F_j(\epsilon_{jt})$

(1.45)

Assuming the parameters in the marginal and copula densities are independent, we can separately estimate the parameters of the copula model $\theta$ by maximizing the log likelihood of the copula density function.

$$
\hat{\theta} = \arg\max_\theta \sum_{t=1}^T \log c(u, v; \theta)
$$

(1.46)

where $c(u, v; \theta) = \frac{\partial^2 C(u, v)}{\partial u \partial v}$ is the density function for the copula model.
Dynamic Copula Models

Patton (2006) and Creal, Koopman and Lucas (2012) suggested similar observation-driven dynamic copula models for which the dependence parameter is a parametric function of lagged data and an autoregressive term. In this paper, we used the “Generalized Autoregressive Score” (GAS) model (Creal, Koopman and Lucas (2012)) to estimate time-varying parameters for a wide range of copula models in the same dynamic framework. Since the parameters of copula \( \theta_t \) are often constrained to lie in a particular range, this approach applies a strictly increasing transformation (e.g., \( \log \), \( \text{logistic} \), \( \text{arctan} \)) to the copula parameter, and then model the dynamics of transformed parameters \( f_t \) without constraints.

Let the copula model be \( C(U_t, V_t; \theta_t) \). The time varying evolution of transformed parameter \( f_t \) can be modeled as

\[
f_t = h(\theta_t) \iff \theta_t = h^{-1}(f_t)
\]

(1.47)

Where \( f_{t+1} = \omega + \beta f_t + \alpha I_t^{-1/2} s_t \)

(1.48)

\[
s_t = \frac{\partial}{\partial \theta} \log C(U_t, V_t; \theta_t)
\]

(1.49)

\[
I_t = E_{t-1}(s_t s_t')
\]

(1.50)

where \( I_t^{-1/2} s_t \) is the standardized score of the copula log-likelihood\(^{24}\), which defines a steepest ascent direction for improving the model local fit in term of likelihood. For the student-t copula, the transformation function \( \rho_t = \frac{1-\exp(-f_t)}{1+\exp(-f_t)} \) is used to ensure that the conditional correlation \( \rho_t \) takes an value inside \((-1, 1)\).

Table 2 presents the bivariate copula models studied in this paper. Since different

\(^{24}\)The score of the copula log-likelihood can be estimated numerically if there is no closed form solution.
copula models imply different dependence structures in the tail of the distribution, it is essential to be aware how well the competing specifications of copula models are capable of accurately estimating the underlying dependence process. A very simple and reliable way to select the best fitting model is to compare the value of the log likelihood function of different copula models and choose the ones with the highest likelihood.\textsuperscript{25} The metrics that will be used to test the Goodness of Fit (GoF) of copula models is Akaike’s Information Criterion (AIC) which is given as\textsuperscript{26}

\[
AIC = 2k - 2 \log \left( \hat{c}(u, v; \hat{\theta}) \right) 
\]

where \( k \) is the number of model parameters and \( \log \left( \hat{c}(u, v; \theta) \right) \) is the maximized log-likelihood at the estimate of the parameter vector \( \theta \).

\subsection*{1.4.3 Backtesting CoVaR}

Changing the distress condition from \( r_{it} = VaR_{it}(\tau) \) to the more flexible tail event \( r_{it} \leq VaR_{it}(\tau) \) facilitates the backtest of CoVaR estimates as it is very straightforward to observe the days when institution \( i \) was in financial distress. Girardi and Ergun (2012) investigated the backtest of CoVaR by modifying the financial distress from an institution being exactly at its VaR to being at most at its VaR. In this section, we briefly describe the procedure of backtest for \( CoVaR^C \).

Comparing ex-ante VaR forecasts with ex-post losses, the “hit sequence” of vio-

\textsuperscript{25}Generally speaking, the standard likelihood ratio test can’t be performed when the models are non-nested. However, Rivers and Vuong (2002) discussed how to construct non-nested likelihood ratio tests. Therefore, non-nested models can still be compared by their log-likelihood values.

\textsuperscript{26}Genest, Quessy and Remillard (2007) introduced an alternative metric for a goodness-of-fit test: Cramer-von-Mises statistic, which measures the distance between the parametric copula and the empirical copula. However, Cramer-von-Mises statistic fail to account for the number of parameters estimated, which may lead to overfitting of the data.
lation for VaR can be defined as

\[ I_{t+1}^i = \begin{cases} 
1 & \text{if } r_{it} \leq \text{VaR}_{it}(\tau) \\
0 & \text{if } r_{it} > \text{VaR}_{it}(\tau) 
\end{cases} \]  \hspace{1cm} (1.52)

Analogously, conditioning on the sub-sample \( I_{t+1}^i = 1 \) when the financial institution \( i \) being in distress, the hit variable for CoVaR can be easily constructed as

\[ I_{t+1}^{m|i} = \begin{cases} 
1 & \text{if } r_{mt} \leq \text{CoVar}_{mt}(\tau) \text{ and } r_{it} \leq \text{VaR}_{it}(\tau) \\
0 & \text{if } r_{mt} > \text{CoVar}_{mt}(\tau) \text{ and } r_{it} \leq \text{VaR}_{it}(\tau) 
\end{cases} \]  \hspace{1cm} (1.53)

Having defined the conditional hit sequence \( I_{t+1}^{m|i} \) for \( \text{CoVar}^< \), we assess the performance of \( \text{CoVar}^< \) by unconditional coverage testing and independence testing proposed by Kupiec (1995) and Christoffersen (1998).

**Unconditional Coverage Testing**

The hypothesis to test for unconditional coverage is

\[ H_o : E\left( I_{t+1}^{m|i} = \tau \right) \]  \hspace{1cm} (1.54)

which is dedicated to test whether the average violation \( (r_{mt} \leq \text{CoVar}^<) \) is equal to the coverage ratio \( \tau \). Kupiec (1995) proposed a likelihood ratio test on the difference between the observed and expected number of VaR exceedances. More formally, the likelihood ratio test statistic can be specified as

\[ LR_{uc}(\tau) = -2 \log \frac{(1-\tau)^{T-N} \tau^N}{(1-N/T)^{T-N} (N/T)^N} \sim \chi(1) \text{ under } H_o \]  \hspace{1cm} (1.55)
where $T$ is the total number of observation in the sub-sample $I_{t+1}^i = 1$ when financial institutions are in distress, and $N$ is the number of violations for CoVaR which satisfy $I_{t+1}^i \times I_{t+1}^{m|i} = 1$, that is, when both financial firms and market are in distress. Therefore $N/T$ is the empirical hit ratio for the $CoVaR \leq$ in the sub-sample $I_{t+1}^i = 1$.

**Conditional Coverage Testing**

If the forecast of CoVaR is precise, we would not expect that the violation of CoVaR is clustered. In other words, the hit sequence $I_{t+1}^{m|i}$ should be independent over time. Christoffersen (1998) suggested the following likelihood approach for the independence test:

$$LR_{cc} = 2 \log \left( \frac{(1 - \pi_{01})^{n_{00}} \pi_{01}^{n_{01}} (1 - \pi_{11})^{n_{10}} \pi_{11}^{n_{11}}}{(1 - N/T)^N (N/T)^N} \right) \sim \chi^2(1) \text{ under } H_0 \quad (1.56)$$

where $\pi_{ij}$ $i,j = 0,1$ is the transition probability. For example, conditional on today being a non-violation ($I_t^{m|i} = 0$), the probability of tomorrow being a violation ($I_{t+1}^{m|i} = 1$) is $\pi_{0,1}$. and $n_{ij} i,j = 0,1$ is the total number of days that state $j$ occurred after state $i$ occurred the previous days. Therefore, the transition probability can be calculated as $\pi_{0,1} = \frac{n_{0,1}}{n_{0,0} + n_{0,1}}$ and $\pi_{1,1} = \frac{n_{1,1}}{n_{1,0} + n_{1,1}}$.

An approximate estimates of CoVaR should satisfy at least these two backtests.
1.5 Data and Empirical Results

1.5.1 Data

The recent financial crisis provides ample evidence of a risk spill-over effect from individual financial institutions to the whole financial industry. The sample studied in this paper is common and widely used in the recent literatures. We consider daily holding period returns for almost the same panel of financial institutions studied by Acharya et al. (2010), Brownless and Engle (2011) and Benoit et al. (2013) between January 3, 2000 and December 30, 2011. The sample contains almost all U.S. financial institutions with equity market capitalization in excess of 5bln USD as of end of June 2007 (92 firms in total), among which, there are about 64 companies that have been trading continuously during the whole sample period in which each institution has 3018 observations. Following Brownless and Engle (2011), the sample of institutions can be split into four groups based on their two-digit SIC classification codes (Depositories, Insurance, Broker-Dealers and Others). Table 1 presents the full list of financial institutions studied in the paper. The return on the Dow Jones U.S. Financials Index (DJUSFN) is used as a proxy for the market return of the financial system. CoVaR and $MES$ are computed at the most common confidence interval $\tau = 5\%$. We obtain our data from Bloomberg.

Table 1.3 gives some descriptive statistics for the unfiltered return series over the whole sample. We can observe that the broker-dealers companies have been the riskiest groups: high volatility, large kurtosis, large VaR and Expected Shortfall (ES). The Jarque-Bera test confirms that all time series of daily returns are not normally distributed with significant excess kurtosis. The Ljung-Box test can be rejected for the market index (DowJones), depositories and broker-dealer companies, which
indicates the presence of autocorrelation for these groups of financial institutions. As the characteristics of the extremes of the distribution show, the asymmetry of the distribution toward right tail is mildly pronounced for all categories of financial firms.

Table 1.4 provides summary statistics on the parameter estimates (median across firms) for the institutions in various industry categories. As we can see, the difference between all categories are trivial. The individual volatility estimation displays the same persistence across all groups of financial firms. The univariate distributions have fat tails as expected. The degrees of freedom $\nu$ of skewed $t$ distribution ranges between 3.8 to 4.1. The asymmetry parameter $\lambda$ is found to be close to zero, indicating the univariate distribution being quite symmetrical. The estimation of the dynamic student-$t$ copula model shows that all financial firms are highly correlated with the market index return, with the median linear correlation $\rho$ ranging from 0.64 to 0.77, which is consistent with the fact that the dependence between financial institutions and market is strong.

Figure 1.9 displays the median across firms of conditional correlation $\rho$ for dynamic student $t$ copula estimated by 'GAS' model discussed in Section 4.2. The lower panel presents the median across firms of lower (upper) tail dependence for mixture Copula model Clayton+Gumbel+Normal. It is clear that the conditional correlation is higher during the financial crisis than that prior to crises, which is consistent with the well-documented empirical results that conditional correlations increase during an economics downturn. However, the rolling window estimation of the mixture copula model shows that tail dependences do not display the same persistence. The lower tail dependence seems to be stronger than upper tail dependence only after or in the early stage of financial crisis, which calls for further scrutiny.

Having set up the margin and copula model, we next estimate $\Delta CoVaR$ and
MES under different copula models and check if dependence structure in the tail risk between financial firm and market makes a difference in the measures of systemic risk.

1.5.2 Empirical Results

Estimation of Systemic Risk under Different Copula Models

Figure 1.10 displays the median across the firms of the time series estimates of three systemic risk measures $\Delta CoVaR$, $\Delta CoVaR^\leq$ and $MES$ estimated based on different copula models which are characterized by the different properties of tail dependence. Normal copula based estimation fails to take into account tail dependences. Rotated-Gumbel Copula based estimation, however, is supposed to account for lower tail dependences of financial data. Gumbel copula based estimation, on the contrary, is assumed to accommodate upper tail dependence. All series of estimates peak at the end of 2008 or the beginning of 2009 and then decay slowly to the lower level comparable to the one before the crisis. This pattern of time evolution is perfectly consistent with the occurrence of the financial crisis as of September 17, 2008 when Lehman Brothers filed for bankruptcy after the financial support offered by Federal Reserve Bank was stopped. The upper panel of Figure 1.10 shows clearly that the systemic risk measure for $\Delta CoVaR$ fails to reflect the fact that higher lower tail dependence tends to result in higher systemic risk as the normal copula based estimation of $\Delta CoVaR$, which disregards the tail dependence property of data, could be larger than that estimated by rotate Gumbel copula characterized by lower tail dependence.

By comparison, the systemic risk measures $\Delta CoVaR^\leq$ and $MES$, which are displayed in the middle and lower panels of Figure 1.10, seem to succeed in indicating the impor-

---

Note that all of these copula based estimations may be misspecified without model selection. It is well known that the daily return of financial assets are characterized by the tail dependence on both sides of distribution. Here we just illustrate how the outcomes could differ if tail dependence on either side is or is not accounted for.
tance of tail dependence in the estimation of systemic risk. Accounting for lower tail dependence (Rotated Gumbel) always leads to a larger value of $\Delta CoVaR^< \leq$ and $MES$.

Figure 1.11 shows the cross sectional relationship of the average systemic risk measures which is defined as:

$$\Delta CoVaR_i = \frac{1}{T} \sum_{t=1}^{T} \Delta CoVaR_{i,t} \quad \Delta CoVaR^<_{i} = \frac{1}{T} \sum_{t=1}^{T} \Delta CoVaR^<_{i,t} \text{ and } MES_i = \frac{1}{T} \sum_{t=1}^{T} MES_{i,t}$$

The x-axis displays the estimations of these systemic risks based on the normal copula model, which is well-known to disregard the tail dependence. The y-axis show the estimations which take into account either symmetric tail dependence on both sides (student-t copula) or tail dependence only on left side (Rotated Gumbel Copula). Again it indicates that all systemic risk measures except for $\Delta CoVaR$ are consistent with dependence measures in the cross sectional estimation. In other words, taking into account lower tail dependence would result in higher estimates of systemic risk measures for $\Delta CoVaR^< \leq$ and $MES$ (all values lie above the diagonal line). However, the same situation is not observed in the upper panel of Figure 1.11, as the values of the cross sectional estimation for $\Delta CoVaR$ could lie below the diagonal line, which indicates that, normal copula based estimation could, on the contrary, leads to the larger value of the systemic risk measure $\Delta CoVaR$ than those taking into account tail dependence. It is noteworthy that, accounting for lower tail dependence seems to merely scale up all values of $\Delta CoVaR^< \leq$ and $MES$, but does not significantly change the ranking of systemic importance for financial institutions. This is indicated by the strong cross sectional link in Figure 1.11, which implies that the values on the y-axis and x-axis increase or decrease concurrently.

Since different copula models result in different estimates of the systemic risk
measure, it has yet to be determined which copula model fits the data best. Table 1.5 provides a Goodness of Fit test for various copula models based on Akaike’s Information Criteria (AIC). As we can see, the dynamic copula model always outperforms its corresponding static model. This fact is not surprising as it is widely accepted that financial data exhibit time varying evolutions in distributions. In addition, the dynamic student t copula model out-performs all other parametric models with no exception, including some mixture copula models, which implies that the joint distribution of daily returns between financial firms and market exhibit tail dependence on both sides of distribution. Note that AIC or the maximal log-likelihood value only provides the average goodness of fit for the distribution as a whole. Higher AIC does not necessarily imply better goodness of fit in each area of distribution, especially in the tail of distribution, on which the estimation of systemic risk primarily concentrates. Therefore it is essential to test the performance of systemic risk estimation, especially in the tails of the distribution.

Table 1.6 presents the summary statistics of the P-value in the backtest for the $CoVaR^c$ estimated by different copula models with different marginal distributions. The testing results show that all estimates of $CoVaR^c$ satisfy the conditional coverage property (P-value for $LR_{indp}$ are all larger than 10%), which indicates that the probability of violation $r_{mt} \leq CoVaR^{m|i}(\tau)$ in the next period when financial firm $i$ is in distress (namely, $r_{it} \leq \text{VaR}^i(\tau)$) doesn’t depend on the violation today. However, the unconditional coverage testing $L_{ucp}$ do fail for some estimates. The first two columns of Table 1.6 demonstrates that the specification of margin is essential for the unconditional coverage property. Even though the dependence structure remains unchanged (normal copula), changing marginal distribution from normal to skewed t distribution alone would significantly improve unconditional coverage testing (the p-value of $LR_{ucp}$ increase significantly). This result is consistent
with most empirical results that VaR estimation based on Gaussian distribution often potentially underestimates downside risk. Girardi and Ergun (2012) found that the $CoVaR^\leq$ estimation based on the bivariate joint skewed t distribution significantly outperforms the estimation based on the bivariate Gaussian distribution. As we can see now, this improvement of performance mainly comes from the specification of marginal distribution (from normal to skewed t) rather than the change of dependence structure. Second, comparing the last four columns of Table 1.6 reveals that the specification of dependence structure does make a difference in the estimation of $CoVaR^\leq$. The last column of Table 1.6 presents the estimation of $CoVaR^\leq$ based on the Gumbel copula model which is characterized by the upper tail dependence. As shown, disregarding the property of lower tail dependence in the data would cause the estimations of $CoVaR^\leq$ for the 50% financial institutions in our sample to fail the unconditional coverage test. In contrast, taking into account tail dependence on both sides of distribution (e.g., student-t copula) would produce the most accurate estimation of $CoVaR^\leq$. As the third column displays, the $CoVaR^\leq$ estimation based on the skewed t margin and student-t copula perform best with respect to the P values for both the unconditional and conditional coverage tests. This fact is also in line with the empirical result that student-t copula in our sample data has best goodness of fit compared with other copula models (as Table 1.5 shows).

Table 1.7 reports the average values of three systemic risk measures by industry category from Jan. 2000 to Dec. 2011:

$$\Delta CoVaR = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} \Delta CoVaR^i_t$$ and $$MES = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T} \sum_{t=1}^{T} MES^i_t$$

Simulation results in section 1.3 show that as dependence strength ($\rho$ or Kendall $\tau$) attain a high value, $CoVaR^\leq$ estimated by different copula models would converge. This is the reason why normal copula (second column of Table 1.6) and rotated Gumbel (fourth column of Table 1.6) copula based estimations also perform pretty well in terms of both conditional and unconditional coverage tests. Therefore it is hard to detect the discrepancy of dependence structure based on the estimates of $CoVaR^\leq$. 

39
First, comparing the last two columns of each panel of table indicates that, when lower tail dependence implied by student-t copula is taken into account, both $\Delta CoVaR^{<}$ and $MES$ increase while $\Delta CoVaR$ decreases, which again verifies the previous discussion that only the former two systemic risk measures are consistent with dependence measures. Second, as the first two columns of Panel A show, the quantile regression estimation of $\Delta CoVaR$ proposed by AB (2011) is only comparable in magnitude to the value estimated by joint normal distribution (Norm-Norm), which tends to underestimate systemic risk in the presence of lower tail dependence. Third, the specification of marginal distribution seem to be much more influential in the estimation of $\Delta CoVaR$ and $\Delta CoVaR^{<}$ than the estimation of $MES$. Relaxing margin from normal to skewed t distribution yields a significant increase for the value of $\Delta CoVaR^{<}$ as well as $\Delta CoVaR$, while relaxing the dependence structure from Gaussian to student t only brings about a moderate increase for $\Delta CoVaR^{<}$ (comparing the second and third column of Panel B). However, the influence of margin and dependence structure seems to play equal role in the estimation of $MES$ (see Panel C). Relaxing the margin from Norm to Skewed t distribution (the first two columns of Panel C) only results in a moderate increase of value by 10.95% (from 3.38 to 3.75) for overall financial institutions. In contrast, relaxing dependence structure from norm to student-t copula (see the last two columns of Panel C) also brings about a mild increase of value by 12.80% (from 3.75 to 4.23).

Finally, Table 1.7 reveals that Depository institutions are the most systemically risky group based on the measure of $\Delta CoVaR^{<}$ or $\Delta CoVaR$, no matter what is the specification of margin and dependence structure. This finding is consistent with the results of Billio et al. (2010) and Girardi and Ergun (2012) that the commercial banks have been most systemically risky due to their illiquid assets, coupled with their structure, which is not designed to withstand rapid and large loss. Summary statistics of Table 1.4 also reveal that the dependence (correlation $\rho$) between de-
pository and market is strongest compared with other groups, which further verifies that commercial banks are more vulnerable to economics meltdown. However, based on the ranking of \( MES \), broker and dealer become the most systemic risky group instead. This discrepancy in the ranking results from the fact that the value of \( MES \) is determined not only by the dependence structure between firm and market, but also by the firms’ individual characteristics such as volatility.

**Which Firms are the SIFIs**

The discussion so far indicates that \( \Delta CoVaR^c \) and \( MES \) are all consistent with dependence measures. But it has yet to be determined if these two measures of systemic risk are convergent or divergent in identifying systemic important financial institutions (SIFIs), especially with regards to the financial crisis that started in September 2008 when the supervision authorities had to determine which financial institutions to bail out based on their systemic importance.

In this paper, we define the financial crisis as the situation in which the market return exceeds its VaR at 5%, and derive the measures of \( \Delta CoVaR^c \), lower tail dependence, \( MES \) and \( SRISK \) respectively for a list of financial institutions in our sample, which are all estimated based on the student-t copula with margins being skewed t distribution.

Table 1.8 presents the tickers of top ten most risky financial institutions on September 17, 2008 when Lehman Brothers filed for bankruptcy after financial support facility offered by Federal Reserve stopped. Comparing the concordant pattern of each pair of risk measures in the lower panel of Table 1.8 reveals some interesting facts. First, we found that \( \Delta CoVaR^c \) is observed to have a stronger correlation with tail dependence than other risk measures. Five of the ten financial institutions belong to the top 10 SIFIs with respect to the ranking of \( \Delta CoVaR^c \) and tail depen-
dence. In contrast, *MES* and firms’ conditional CAPM $\beta$ are highly concordant with each other. Nine of the ten financial institutions are commonly identified as SIFIs by the ranking of *MES* and conditional CAPM $\beta$. In addition, *SRISK* provides a strong connection with the firm level characteristics such as leverage and market capitalization.

Table 1.9 reports the value of rank similarity ratio between each pair of systemic risk measure for the date in analysis which covers the periods prior to, during and post financial crisis respectively. The similarity ratio are defined as the proportion of common financial firms in each pair of rankings at a given date. For instance, a similarity ratio equal to 0.60 for the pair of systemic risk measure $\Delta CoVaR_{\leq}$ and tail-dependence means that among the 10 firms in the top, there are 6 firms that are commonly identified as SIFIs by these two systemic risk measure. The concordant pattern of systemic risk ranking over multiple dates indicates that the converging ranking is not specific to any particular day for each pair of $\Delta CoVaR_{\leq}$ and Tail-Dependence, *MES* and CAPM $\beta$, or *SRISK* and firm characteristics such as leverage. Furthermore, the similarity ratio among each pair of systemic risk measure (last panel of Table 1.9) is pretty low, indicating that these systemic risk measures complement instead of replace each other.

Figure 1.12 provides further evidence for the concordant ranking of each pair of systemic risk measures. A stronger cross-sectional link can be found in the diagonal panels of Figure 1.12, which indicates that $\Delta CoVaR_{\leq}$ is more closely related to the measure of tail dependence. *MES*, however, shows a stronger cross sectional relation with conditional Beta. By comparison, *SRISK* seems to provide closer connection with firm level characteristics such as leverage. As the first (upper left) panel of Figure 1.12 illustrates, the time series averages of $\Delta CoVaR_{\leq}$ are highly correlated with the average values of tail dependence across firms. Analogously, the middle
panel shows the stronger cross sectional link between the average $MES$ and Beta. In addition, the cross sectional link between $MES$ and tail dependence is not as strong as that between $\Delta CoVaR^\leq$ and tail dependence. This observation is not surprising, as $\Delta CoVaR^\leq$ relies only on the dependence structure, while $MES$ is determined by the dependence structure as well as marginal characteristics like firms’ volatility $\sigma_{it}$, which is taken into account by the estimation of the conditional beta $\beta_{it} = \frac{\rho_{it} \sigma_{it}}{\sigma_{mt}}$.

The last (lower right) panel shows that $SRISK$ is more closely related to firm level information such as leverage than the other two systemic risk measures. This stylized fact seems to imply that $\Delta CoVaR^\leq$ is more in line with the “too interconnected to fail” paradigm, and $SRISK$ is more related to the “too big to fail” paradigm. In contrast, $MES$ offers a compromise between these two paradigms.

---

29 As the value of $SRISK$ can be negative for some firms, we keep only those financial firms with positive values of $SRISK$ and thus positive contributions to the systemic risk of the financial market.
1.6 Concluding Remarks

The global financial crisis initiated in 2008 has altered the regulators’ awareness of fragility of financial system. This paper develops a common framework based on copula model to estimate several widely used return-based systemic risk measures: Delta Conditional Value at Risk ($\Delta CoVaR$) and its Modified ($\Delta CoVaR^{\leq}$), Marginal Expected Shortfall($MES$) and its extension, systemic risk measure ($SRISK$). The nonlinear dependence of tail risk is straightforward to be accommodated in the copula model. By eliminating the discrepancy of the marginal distribution, the copula specification provide a flexibility to concentrate on the effect of the dependence structure in the estimation of systemic risk. Our empirical studies shows that the nonlinear dependence of tail risk does make a difference in the estimation of $\Delta CoVaR^{\leq}$ and $MES$. Simulation exercises reveal that $\Delta CoVaR$ originally proposed by AB (2011) is in conflict with dependence measures. The modified version of $\Delta CoVaR^{\leq}$ and $MES$ is more consistent with dependence measures, which is line with economic intuition that stronger dependence strength results in higher systemic risk measures. Furthermore, we found that the linear quantile regression estimation of $\Delta CoVaR$ proposed by Adrian and Brunnermeier (2011) is inadequate to completely capture the non-linear contagion tail effect, which tends to underestimate systemic risk in the presence of lower tail dependence. We estimate the systemic risk contributions of four financial industry sample consisting of a large number of institutions for the sample period from January 2000 to December 2010, and we found that $\Delta CoVaR^{\leq}$ is observed to have a strong correlation with tail dependence. In contrast, $MES$ is found to have a strong empirical relationship with firms’ conditional CAPM $\beta$. $SRISK$, however, provides further connection with firms level characteristics by accounting for information on market capitalization and liability. This stylized fact seems to imply that $\Delta CoVaR^{\leq}$ is more in line with the “too interconnected to fail” paradigm, and $SRISK$ is more related to the “too big to fail” paradigm. In contrast, $MES$ offers a
compromise between these two paradigms.
1.A Appendix

1.A.1 Proof on the relationship between conditional distribution and copula

Define $F(y|x)$ as the conditional distribution of $y$ given $x$, and let $C(u, v)$ be a copula function where $u = F_x(x)$ and $v = F_y(y)$. Therefore, the following proof shows that the conditional distribution $F(y|x)$ is equal to the first order derivative of $C(u, v)$ with respect to $u$

$$F(y|x) = Pr(Y \leq y | X = x)$$
$$= Pr\left(F_y(Y) \leq F_y(y) | F_x(X) = F_x(x)\right)$$

Let $F_y(Y) = V$ and $F_x(X) = U$

$$= Pr(V \leq v | U = u)$$
$$= \lim_{\Delta u \to 0^+} \frac{Pr(V \leq v, u \leq U \leq u + \Delta u)}{Pr(u \leq U \leq u + \Delta u)}$$
$$= \lim_{\Delta u \to 0^+} \frac{Pr(V \leq v, U \leq u + \Delta u) - Pr(V \leq v, U \leq u)}{Pr(U \leq u + \Delta u) - Pr(U < u)}$$
$$= \lim_{\Delta u \to 0^+} \frac{C(u + \Delta u, v) - C(u, v)}{\Delta u}$$
$$= \frac{\partial C(u, v)}{\partial u}$$
1.A.2 Proof of Theorem 2.1

Assume a Gaussian copula model $C(u, v; \rho)$ where $u = \Phi(x)$ and $v = \Phi(y)$ are standard Gaussian marginal distribution

\[
C(u, v; \rho) = \Phi_{\rho}(\Phi^{-1}(u), \Phi^{-1}(v)) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \exp\left\{-\frac{(s^2 - 2\rho st + t^2)}{2(1-\rho^2)}\right\} ds dt
\]

The first order derivative of $C(u, v; \rho)$ with respect to $u$ can be derived as:

\[
\frac{\partial C(u, v; \rho)}{\partial u} = \frac{\partial \Phi_{\rho}(X, Y)}{\partial X} \frac{1}{\phi(x)}
\]
where the first component $\frac{\partial \Phi (X,Y)}{\partial X}$ can be further derived as:

$$\frac{\partial \Phi (X,Y)}{\partial X} = \frac{1}{2\pi \sqrt{1-\rho^2}} \int_{-\infty}^{Y} \exp\left\{ -\frac{(x^2 - 2\rho xt + t^2)}{2(1-\rho^2)} \right\} dt$$

$$= \frac{1}{2\pi \sqrt{1-\rho^2}} \int_{-\infty}^{Y} \exp\left\{ -\frac{(x^2 - \rho^2 x^2 + \rho^2 x^2 - 2\rho xt + t^2)}{2(1-\rho^2)} \right\} dt$$

$$= \frac{1}{2\pi \sqrt{1-\rho^2}} \int_{-\infty}^{Y} \exp\left\{ -\frac{[x^2(1-\rho^2) + (t-\rho x)^2]}{2(1-\rho^2)} \right\} dt$$

$$= \frac{1}{2\pi \sqrt{1-\rho^2}} \int_{-\infty}^{Y} \exp\left\{ -\frac{x^2}{2} \right\} * \exp\left\{ -\frac{(t-\rho x)^2}{2(1-\rho^2)} \right\} dt$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{x^2}{2} \right\} \frac{1}{\sqrt{2\pi \sqrt{1-\rho^2}}} \int_{-\infty}^{Y} \exp\left\{ -\frac{(t-\rho x)^2}{2(1-\rho^2)} \right\} dt$$

$$= \phi(x) * \frac{1}{\sqrt{2\pi \sqrt{1-\rho^2}}} \int_{-\infty}^{Y} \exp\left\{ -\frac{(t-\rho x)^2}{2(1-\rho^2)} \right\} dt$$

$$= \phi(x) * \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{Y} \exp\left\{ -\frac{(t-\rho x)^2}{2(1-\rho^2)} \right\} dt \frac{d(t-\rho x)}{\sqrt{1-\rho^2}}$$

$$= \phi(x) * \phi\left( \frac{Y - \rho X}{\sqrt{1-\rho^2}} \right)$$

$$= \phi(x) * \phi\left( \frac{\Phi^{-1}(v) - \rho \Phi^{-1}(u)}{\sqrt{1-\rho^2}} \right)$$

Therefore

$$\frac{\partial C(u,v;\rho)}{\partial u} = \frac{\partial \Phi (X,Y)}{\partial X} \frac{1}{\phi(x)} = \phi\left( \frac{\Phi^{-1}(v) - \rho \Phi^{-1}(u)}{\sqrt{1-\rho^2}} \right)$$

(1.A.2.1)

Let $\tau = \frac{\partial C(u,v;\rho)}{\partial u}$, we can solve $v$ from the above equation(4) as:

$$v = \Phi \left[ \rho \Phi^{-1}(u) + \sqrt{1-\rho^2} \Phi^{-1}(\tau) \right]$$

(1.A.2.2)
Note that \( v = F_Y(y) \) and \( u = F_X(x) \).

We have assumed that both \( X \) and \( Y \) follow Gaussian marginal distribution

\[
Y \sim N(\mu_y, \sigma_y) \quad X \sim N(\mu_x, \sigma_x)
\]

Therefore

\[
v = F_Y(y) = \Phi\left(\frac{y - \mu_y}{\sigma_y}\right) \\
u = F_X(x) = \Phi\left(\frac{x - \mu_x}{\sigma_x}\right)
\]

Substituting the above two equations into the equation (5) yields

\[
\phi\left(\frac{y - \mu_y}{\sigma_y}\right) = \phi\left(\rho\phi^{-1}\left(\frac{x - \mu_x}{\sigma_x}\right) + \sqrt{1 - \rho^2}\phi^{-1}(\tau)\right) \\
\frac{y - \mu_y}{\sigma_y} = \rho\left(\frac{x - \mu_x}{\sigma_x}\right) + \sqrt{1 - \rho^2}\phi^{-1}(\tau) \\
y = \rho\frac{\sigma_y}{\sigma_x} x - \rho\frac{\sigma_y}{\sigma_x} \mu_x + \sigma_y \sqrt{1 - \rho^2}\phi^{-1}(\tau) + \mu_y
\]

Now, set both \( x \) and \( y \) to be on their tail risk, respectively. Namely, \( x = VaR_{x}^{\tau} \) and \( y = CoVaR_{y|x}^{\tau} \), we have

\[
CoVaR_{y|x}^{\tau} = \rho\frac{\sigma_y}{\sigma_x} VaR_{x}(\tau) - \rho\frac{\sigma_y}{\sigma_x} \mu_x + \sigma_y \sqrt{1 - \rho^2}\phi^{-1}(\tau) + \mu_y
\]

Therefore there is linear dependence between \( CoVaR_{y|x}^{\tau} \) and \( VaR_{x}^{\tau} \). The linear dependence coefficient \( \rho\frac{\sigma_y}{\sigma_x} \) is constant over quantiles. The time variation of this coefficient can be driven by the time variation of \( \rho_{x,y,t} \) or the ratio of their volatility \( \frac{\sigma_{y,t}}{\sigma_{x,t}} \).
Following AB (2011), we can further derive $\Delta \text{CoVaR}$ as

$$
\Delta \text{CoVaR}_y|x = \text{CoVaR}_y|x - \text{CoVaR}_y|x=\text{median} = \rho \frac{\sigma_y}{\sigma_x} (VaR_x^\tau - VaR_{0.5}^x)
$$
1.A.3 Proof of Theorem 2.2

Suppose $X$ and $Y$ are the returns of two assets with the marginal distribution being $X \sim F_x$ and $Y \sim F_y$. The marginal expected shortfall of $X$ conditioning on the tail event $Y < \text{VaR}_y(\tau)$ can be defined as $\text{MES} = -E(X|Y < \text{VaR}_y(\tau))$. Following the definition of expectation, we have

$$\text{MES}(\tau) = -E(X|Y \leq \text{VaR}_y(\tau))$$

$$= - \sum_{i=1}^{N} x_i F(x | Y \leq \text{VaR}_y(\tau))$$

$$= - \sum_{i=1}^{N} x_i \frac{F(x, Y \leq \text{VaR}_y(\tau))}{F(Y \leq \text{VaR}_y(\tau))}$$

$$= - \sum_{i=1}^{N} x_i \frac{F(Y \leq \text{VaR}_y(\tau), X = x_i)}{F(Y \leq \text{VaR}_y(\tau))} F(X = x_i)$$

$$= - \frac{1}{\tau} \sum_{i=1}^{N} x_i F(Y \leq \text{VaR}_y(\tau)|X = x_i) F(X = x_i)$$

(1.A.3.1)

Let $F(Y = y_i) = v$ and $F(X = x_i) = u$, we have already proved in the beginning of appendix that the conditional distribution $F(y_i | x_i)$ is equal to the first order derivative of copula function $C(u, v)$ with respect to $u$

$$F(Y \leq y_i|X = x_i) = \frac{\partial C(u, v)}{\partial u}$$

Therefore

$$F(Y \leq \text{VaR}_y(\tau)|X = x_i) = \frac{\partial C(u, \tau)}{\partial u}$$

(1.A.3.2)
Substituting equation (6) into the equation (5) yields

\[ MES(\tau) = -E\left( X|Y \leq VaR_y(\tau) \right) = -\frac{1}{\tau} \sum_{i=1}^{N} x_i \frac{\partial C(u, \tau)}{\partial u} Pr(X = x_i) \]

In the case of continuous random variables, the above equation can be further simplified as:

\[ MES(\tau) = -E\left( X|Y \leq VaR_y(\tau) \right) = -\frac{1}{\tau} \sum_{i=1}^{N} x_i \frac{\partial C(u, \tau)}{\partial u} Pr(X = x_i) = -\frac{1}{\tau} \int \int_{-\infty}^{+\infty} \frac{\partial C(u, \tau)}{\partial u} f(x) dx \]

\[ = -\frac{1}{\tau} \int_{-\infty}^{+\infty} x \frac{\partial C(u, \tau)}{\partial u} f(x) dx = -\frac{1}{\tau} \int_{-\infty}^{+\infty} x \frac{\partial C(u, \tau)}{\partial u} f(x) dx \]

\[ = -\frac{1}{\tau} \int_{-\infty}^{+\infty} x \frac{\partial C(u, \tau)}{\partial u} f(x) dx \]
Figure 1.1: The upper panel graphs display the scatter plot of actual data and the simulated data from a bivariate normal distribution based on the first two moments of real data. The lower panel graphs show the exceedance correlation and quantile dependence: AIG and the Dow Jones Financial Index Return (a proxy for the Financial system). The line marked with open circles shows the average exceedance correlations (probability) across 5000 simulation of series with \( T = 2646 \) from a bivariate normal distribution with the first two moments equal to those of real data. The dashed lines are 5th and 95th percentiles of the distribution for exceedance correlation (probability) estimates in this simulation. The sample data run from January 3, 2000 to December 31, 2010.
Figure 1.2: This figure displays the $\Delta CoVaR$ (y-axis) estimated by copula model and quantile regression model. The $\Delta CoVaR^{y|x}$ are defined as: $\Delta CoVaR^{y|x} = CoVaR_y^{x=VaR_x(0.05)} - CoVaR_y^{x=\text{median}}$. In each panel of the figure, the data are generated by the corresponding copula model shown in the title. All marginal distributions are set to be standard normal distribution except for the student t copula in which the margin is set to be student t with df=5. The parameter of each copula model is chosen such that the linear correlation $\rho$ or Kendall correlation $\tau$ is equal to the value displayed in the X-axis.
Figure 1.3: This figure displays 8000 random draws from two bivariate joint distributions: Clayton-copula (left) and Gumbel-copula (right). Both margins are set to be standard Gaussian $N(0, 1)$. In each panel, the parameter is chosen such that the linear correlation $\rho$ of random variates is equal to 0.8. CoVaR is estimated at 1% quantile of financial market return. Visually, $\Delta \text{CoVaR}$ in the left panel is obviously greater than that in the right panel.
Figure 1.4: This figure compares the copula-based estimation of MES with the non-parametric kernel estimation following Brownless and Engle (2011). In the left panel, data are generated by the student t copula with degree of freedom $df = 5$. In the right panel, data are generated by the Rotated Gumbel Copula. Both margins are set to be student t with $df = 5$. The parameters of both copula models are selected such that the Kendall correlation of generated data is 0.5. The X-axis denotes the unconditional cut-off value $\tau$ which determines the tail event of the market return when $r_{mt} \leq VaR_{r_{mt}}(\tau)$. The simulations are implemented for 5000 times.
Figure 1.5: This figure displays the $\Delta CoVaR^e$, $\Delta CoVaR^c$ and $MES$ ($y-axis$) estimated by Normal Copula (without tail dependence) and Student $t$ Copula (with symmetric tail dependence). The marginal distributions are set to be Student $t$ with $df = 5$. The $x-axis$ displays linear correlation $\rho$ which measures the average co-movement strength of the data. The parameters of copula models are chosen such that the linear correlation $\rho$ of generated random variates is equal to the value shown in the $x-axis$. 
Figure 1.6: This figure displays the $\Delta CoVaR^c$, $\Delta CoVaR^s$ and $MES (y – axis)$ estimated by Normal Copula (without tail dependence), Rotated Gumbel Copula (with lower tail dependence) and Gumbel Copula (with Upper tail dependence). The marginal distributions are set to be Student $t$ with $df = 5$. The x-axis displays kendall $\tau$, which measures the co-movement strength of data. The parameters of copula model are chosen such that the kendall $\tau$ of generated random variates is equal to the value shown in the $x – axis$. 
Figure 1.7: This figure displays the $\Delta CoVaR^\gamma$, $\Delta CoVaR^\leq$ and $MES$ ($y-axis$) estimated by the Copula of Student t (symmetrical tail dependence), Rotated Gumbel (lower tail dependence) and Gumbel (Upper tail dependence). All margins are Student t with $df = 5$. The x-axis displays the value of Lower tail dependence or Upper tail dependence implied by these copulas: $LTD = \lim_{u \to 0^+} P(X_2 \leq F_2^{-1}(u)|X_1 \leq F_1^{-1}(u))$ and $UTD = \lim_{u \to 1^-} P(X_2 \geq F_2^{-1}(u)|X_1 \geq F_1^{-1}(u))$. The parameters of copula model are chosen such that the tail dependence of generated random variates is equal to the value shown in the $x-axis$. 

59
Figure 1.8: This figure displays the fitted parametric estimates of a skewed \( t \) (Hansen 1994) distribution for the daily return of DowJones Financial Index (Proxy for financial market return) and AIG
Figure 1.9: The upper panel of figure displays the median across firms of conditional correlation for the dynamic student $t$ copula. The lower panel of figure presents the median across firms of lower (upper) tail dependence for mixture Copula C+G+N. The acronyms C+G+N refers to the mixture copula: Clayton+Gumbel+Normal. The parameters of mixture copula model are estimation every 1 month in the Rolling window of 24 months.
Figure 1.10: This figure displays the median across the firms of the time series estimates of $\Delta CoVaR$, $\Delta CoVaR_{\leq}$ and $MES$ under copulas models characterized by different tail dependence properties. The acronyms Rot-Gumbel refers to Rotated Gumbel Copula. The sample period runs from 2007/01/02 to 2010/12/31.
Figure 1.11: This figure displays the cross sectional relationship of systemic risk measure $\Delta CoVaR$, $\Delta CoVaR^<= \text{ and } MES$ estimated with Gaussian copula(x-axis) and Student t, Rotated Gumbel copula(y-axis) from left to right. Where $\Delta CoVaR^i = \frac{1}{T} \sum_{t=1}^{T} \Delta CoVaR^i_t$ and $MES^i = \frac{1}{T} \sum_{t=1}^{T} MES^i_t$. Each point represents an financial institution listed in the table 1. The diagonal solid line represents the equal value corresponding to x-axis and y-axis. The acronyms "St" refers to Student t copula based estimation. ”Rot-Gumbel” represents rotated Gumbel copula based estimation. The sample period covered from 2000/01/03 to 2010/12/31.
Figure 1.12: Note: The scatter plots show the cross sectional link between the time series average of the systemic risk measures displayed on the y-axis, which are all estimated by student t copula with marginal distribution being skewed t distribution. The conditional Beta is estimated as $\beta_{it} = \frac{\rho_{it}}{\sigma_{it}/\sigma_{mt}}$. The tail dependence implied by student t copula is $\tau = 2 - 2T_{1+\nu_t}(\sqrt{1+\nu_t}\sqrt{\frac{1-\rho_t}{1+\rho_t}})$. The SRISK is calculated by the simulation exercise described in the previous section 2.2.4. The solid line in each panel is the OLS regression predicted line, which indicates the strength of cross sectional link between the two variables on the axis. Each point represents a financial institution. The estimation period is from 2004/01/02 to 2010/12/30.
<table>
<thead>
<tr>
<th>Depositories</th>
<th>Insurance</th>
<th>Broker-Dealers</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAC</td>
<td>APL</td>
<td>RTFC</td>
<td>ACAS</td>
</tr>
<tr>
<td>BBT</td>
<td>AFL</td>
<td>E Trade Financial Corp</td>
<td>American Capital Ltd</td>
</tr>
<tr>
<td>BK</td>
<td>ALL</td>
<td>GS</td>
<td>AMTD</td>
</tr>
<tr>
<td>C</td>
<td>CitiGroup Inc</td>
<td>MS</td>
<td>AXP</td>
</tr>
<tr>
<td>CMA</td>
<td>BKLY</td>
<td>AIG</td>
<td>AONC</td>
</tr>
<tr>
<td>HBAN</td>
<td>BRK</td>
<td>Goldman Sachs Group Inc</td>
<td>Aon Corp</td>
</tr>
<tr>
<td>HCBK</td>
<td>CIB</td>
<td>AIG</td>
<td>SCHW</td>
</tr>
<tr>
<td>JPM</td>
<td>CI</td>
<td>AIG</td>
<td>TROW</td>
</tr>
<tr>
<td>KEY</td>
<td>CING</td>
<td>AIG</td>
<td>TROW</td>
</tr>
<tr>
<td>M</td>
<td>CNA</td>
<td>AIG</td>
<td>TROW</td>
</tr>
<tr>
<td>MTHB</td>
<td>CVH</td>
<td>AIG</td>
<td>TROW</td>
</tr>
<tr>
<td>NTRS</td>
<td>HIG</td>
<td>AIG</td>
<td>TROW</td>
</tr>
<tr>
<td>NYB</td>
<td>HNT</td>
<td>AIG</td>
<td>TROW</td>
</tr>
<tr>
<td>PBCT</td>
<td>HUM</td>
<td>AIG</td>
<td>TROW</td>
</tr>
<tr>
<td>PNC</td>
<td>LNC</td>
<td>AIG</td>
<td>TROW</td>
</tr>
<tr>
<td>RF</td>
<td>MBSI</td>
<td>AIG</td>
<td>TROW</td>
</tr>
<tr>
<td>SNV</td>
<td>MMC</td>
<td>AIG</td>
<td>TROW</td>
</tr>
<tr>
<td>STI</td>
<td>PGR</td>
<td>AIG</td>
<td>TROW</td>
</tr>
<tr>
<td>STT</td>
<td>TMK</td>
<td>AIG</td>
<td>TROW</td>
</tr>
<tr>
<td>USB</td>
<td>TVX</td>
<td>AIG</td>
<td>TROW</td>
</tr>
<tr>
<td>WFC</td>
<td>UNIH</td>
<td>AIG</td>
<td>TROW</td>
</tr>
<tr>
<td>ZION</td>
<td>USM</td>
<td>AIG</td>
<td>TROW</td>
</tr>
</tbody>
</table>
Table 1.2: Bivariate Copula Functions Studied in this paper

<table>
<thead>
<tr>
<th>Copula</th>
<th>Parametric Form</th>
<th>LTD</th>
<th>UTD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>$\Phi_{\rho}\left(\Phi^{-1}(u), \Phi^{-1}(v)\right)$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Student t</td>
<td>$t_{\nu,\rho}\left(t_{\nu}^{-1}(u), t_{\nu}^{-1}(v)\right)$</td>
<td>$2 - 2T_{1+\nu}(\sqrt{1+\nu \sqrt{\frac{1-\rho}{1+\rho}}})$</td>
<td>$2 - 2T_{1+\nu}(\sqrt{1+\nu \sqrt{\frac{1-\rho}{1+\rho}}})$</td>
</tr>
<tr>
<td>Frank</td>
<td>$-\frac{1}{\theta} \log \left(1 + \frac{(e^{-\theta u-1})(e^{-\theta v-1})}{e^{-\theta-1}}\right)$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Plackett</td>
<td>$\frac{1}{2}(\theta - 1)^{-1}(1 + (\theta - 1)(u + v) - [(1 + (\theta - 1)(u + v))^2 - 4\theta uv]^{1/2})$</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Clayton</td>
<td>$\left(u^{-\theta} + v^{-\theta} - 1\right)^{-1/\theta}$</td>
<td>$2^{-1/\theta}$</td>
<td>-</td>
</tr>
<tr>
<td>Rotated-Clayton</td>
<td>$u + v - 1 + ((1 - u)^{-\theta} + (1 - v)^{-\theta} - 1)^{-1/\theta}$</td>
<td>-</td>
<td>$2^{-1/\theta}$</td>
</tr>
<tr>
<td>Gumbel</td>
<td>$\exp\left(- \left[(- \log u)^{\theta} + (- \log v)^{\theta}\right]^{\frac{1}{\theta}}\right)$</td>
<td>-</td>
<td>$2 - 2^{1/\theta}$</td>
</tr>
<tr>
<td>Rotated-Gumbel</td>
<td>$u + v - 1 + \exp\left(- \left[(- \log u)^{\theta} + (- \log v)^{\theta}\right]^{\frac{1}{\theta}}\right)$</td>
<td>$2 - 2^{1/\theta}$</td>
<td>-</td>
</tr>
<tr>
<td>Joe-Clayton</td>
<td>$1 - \left(1 - [(1 - \bar{u})^{\gamma} + (1 - \bar{v})^{-\gamma} - 1]^{-1/\gamma}\right)^{1/\kappa}$</td>
<td>$2^{-1/\gamma}$</td>
<td>$2^{-1/\gamma}$</td>
</tr>
</tbody>
</table>
Table 1.3: **Summary Statistics for the unfiltered return series**

<table>
<thead>
<tr>
<th></th>
<th>DowJones Index</th>
<th>Depository</th>
<th>Insurance</th>
<th>Broker-Dealer</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td>0.0116</td>
<td>0.0396</td>
<td>0.0557</td>
<td>0.0472</td>
<td>0.0572</td>
</tr>
<tr>
<td><strong>Std.dev</strong></td>
<td>2.1113</td>
<td>2.8772</td>
<td>2.9194</td>
<td>4.4298</td>
<td>3.4317</td>
</tr>
<tr>
<td><strong>Skewness</strong></td>
<td>0.3022</td>
<td>0.7225</td>
<td>0.4246</td>
<td>1.3455</td>
<td>0.7839</td>
</tr>
<tr>
<td><strong>Kurtosis</strong></td>
<td>14.107</td>
<td>23.0947</td>
<td>27.9047</td>
<td>31.3129</td>
<td>25.5438</td>
</tr>
<tr>
<td>5%-VaR(left)</td>
<td>-2.7719</td>
<td>-3.9201</td>
<td>-4.0153</td>
<td>-4.8486</td>
<td>-3.8385</td>
</tr>
<tr>
<td>5%-VaR(right)</td>
<td>2.7517</td>
<td>4.6253</td>
<td>4.5193</td>
<td>5.5541</td>
<td>4.0362</td>
</tr>
<tr>
<td>5%-ES(left)</td>
<td>-0.2496</td>
<td>-0.3258</td>
<td>-0.3275</td>
<td>-0.3693</td>
<td>-0.3864</td>
</tr>
<tr>
<td>5%-ES(right)</td>
<td>0.2655</td>
<td>0.3616</td>
<td>0.3591</td>
<td>0.4283</td>
<td>0.4352</td>
</tr>
<tr>
<td><strong>P-Value</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jarque-Bera</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Ljung-Box</td>
<td>0.0001</td>
<td>0.0137</td>
<td>0.2187</td>
<td>0.0225</td>
<td>0.1909</td>
</tr>
</tbody>
</table>

This table provides summary statistics on the daily holding period return of some major US financial institutions and DowJones U.S. Financial Index for the period from 2000/01/03 until 2011/12/30. For each category, we report the average value across all institutions about the mean, the standard deviation, the skewness, the kurtosis, the 5% VaR and ES for the left and right tail of distribution. The second last row gives the median over all institutions of p-value of Jarque-Bera test for normality. and the last row reports the median of p-value for the Ljung-Box test with \( m = \log(T) \) lags.
Table 1.4: **Summary Statistics on Parameter Estimates (medians)**

<table>
<thead>
<tr>
<th></th>
<th>Dow Jones Index</th>
<th>Depository</th>
<th>Insurance</th>
<th>Broker-Dealer</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: GARCH Model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_t^2$</td>
<td>$\omega$</td>
<td>$\alpha$</td>
<td>$\gamma$</td>
<td>$\beta$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.0103</td>
<td>0.0228</td>
<td>0.0339</td>
<td>0.0413</td>
<td>0.0591</td>
</tr>
<tr>
<td></td>
<td>0.0174</td>
<td>0.0363</td>
<td>0.0344</td>
<td>0.0176</td>
<td>0.0332</td>
</tr>
<tr>
<td></td>
<td>0.1039</td>
<td>0.0866</td>
<td>0.0932</td>
<td>0.0891</td>
<td>0.0670</td>
</tr>
<tr>
<td></td>
<td>0.9291</td>
<td>0.9190</td>
<td>0.9148</td>
<td>0.9366</td>
<td>0.9224</td>
</tr>
<tr>
<td><strong>Panel B: Skewed t distribution</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\nu$</td>
<td>4.1092</td>
<td>3.9022</td>
<td>4.0798</td>
<td>3.8338</td>
<td>3.7901</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>-0.0106</td>
<td>0.0038</td>
<td>0.0135</td>
<td>0.0015</td>
<td>0.0047</td>
</tr>
<tr>
<td><strong>Panel C: 'GAS’ Dynamic Student t Copula Model (median of means)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho$</td>
<td>-</td>
<td>0.7728</td>
<td>0.6451</td>
<td>0.7634</td>
<td>0.6353</td>
</tr>
<tr>
<td>df</td>
<td>-</td>
<td>7.5087</td>
<td>8.3618</td>
<td>6.9992</td>
<td>7.6626</td>
</tr>
</tbody>
</table>

This table provides summary statistics on parameter estimates for all industry categories. Panel A reports the median across the institutions of the parameter estimates of the GJR-GARCH model for standardized residuals. Panel B reports the median across firms of the parameter estimates of the skewed t distribution for innovations. Panel C reports the median across firms of the mean over time of parameters for the "Generalized Autoregressive Score” student t copula.
Table 1.5: **AIC for the Different Copula Models**

The metric for Goodness of Fit (GoF) test of Copula model will be Akaike’s Information Criterion (AIC)

\[ AIC := 2k - 2\log(\hat{c}(u, v; \hat{\theta})) \]

**Comparing Dependence Structures Using Information Criteria (AIC)**

<table>
<thead>
<tr>
<th>Static Model</th>
<th>Q5</th>
<th>Q25</th>
<th>MEAN</th>
<th>Median</th>
<th>Q75</th>
<th>Q95</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian</td>
<td>-3159.956</td>
<td>-2555.641</td>
<td>-1862.686</td>
<td>-1830.318</td>
<td>-1189.603</td>
<td>-479.192</td>
</tr>
<tr>
<td>Clayton</td>
<td>-2501.409</td>
<td>-2118.555</td>
<td>-1534.261</td>
<td>-1539.268</td>
<td>-989.298</td>
<td>-406.144</td>
</tr>
<tr>
<td>Frank</td>
<td>-3042.834</td>
<td>-2528.719</td>
<td>-1838.441</td>
<td>-1822.108</td>
<td>-1176.598</td>
<td>-513.067</td>
</tr>
<tr>
<td>RotGumbel</td>
<td>-3070.431</td>
<td>-2590.530</td>
<td>-1849.739</td>
<td>-1837.672</td>
<td>-1181.537</td>
<td>-473.883</td>
</tr>
<tr>
<td>Student</td>
<td><strong>-3287.925</strong></td>
<td><strong>-2769.306</strong></td>
<td><strong>-1966.828</strong></td>
<td><strong>-1941.210</strong></td>
<td><strong>-1244.634</strong></td>
<td><strong>-517.695</strong></td>
</tr>
<tr>
<td>C+G+F</td>
<td>-3273.078</td>
<td>-2755.918</td>
<td>-1954.758</td>
<td>-1917.125</td>
<td>-1255.380</td>
<td>-516.206</td>
</tr>
<tr>
<td>Dynamic Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td><strong>-3424.752</strong></td>
<td><strong>-2859.831</strong></td>
<td><strong>-2084.023</strong></td>
<td><strong>-2098.774</strong></td>
<td><strong>-1348.755</strong></td>
<td><strong>-563.794</strong></td>
</tr>
</tbody>
</table>

The acronyms ”SJC” refers to symmetrized Joe-Clayton Copula, which was proposed by Andrew Patton (2006). Dynamic Copula model was estimated following the ”Generalized Autoregressive Score” (GAS) model suggested by Creal, et al (2011), which was discussed in section 3.2. The acronyms ”G+RG+N” refers to mixture copula model of Gumbel+Rotated Gumbel+Normal. ”G+RG+F”: Gumbel+Rotated Gumbel+Frank. ”C+G+F”: Clayton+Gumbel+Frank. and ”C+G+N” : Clayton+Gumbel+Normal.
Table 1.6: **Summary Statistics of P-Value For Unconditional Coverage and Independence Test for CoVaR Estimates**

<table>
<thead>
<tr>
<th></th>
<th>$r_{i,t} \leq VaR_i^t(\tau)$</th>
<th>CoVaR$^a$($\tau$)</th>
<th>(\tau = 0.05)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Margin-Copula</td>
<td>Margin-Copula</td>
<td>Margin-Copula</td>
</tr>
<tr>
<td></td>
<td>Norm - Norm</td>
<td>Skewt - Norm</td>
<td>Skewt - T</td>
</tr>
<tr>
<td>Q5</td>
<td>$LR_{ucp}$</td>
<td>0.000</td>
<td>0.072</td>
</tr>
<tr>
<td></td>
<td>$LR_{indp}$</td>
<td>0.130</td>
<td>0.179</td>
</tr>
<tr>
<td>Q10</td>
<td>$LR_{ucp}$</td>
<td>0.000</td>
<td>0.167</td>
</tr>
<tr>
<td></td>
<td>$LR_{indp}$</td>
<td>0.138</td>
<td>0.216</td>
</tr>
<tr>
<td>Q25</td>
<td>$LR_{ucp}$</td>
<td>0.000</td>
<td>0.349</td>
</tr>
<tr>
<td></td>
<td>$LR_{indp}$</td>
<td>0.273</td>
<td>0.296</td>
</tr>
<tr>
<td>Q50</td>
<td>$LR_{ucp}$</td>
<td>0.002</td>
<td>0.680</td>
</tr>
<tr>
<td></td>
<td>$LR_{indp}$</td>
<td>0.537</td>
<td>0.390</td>
</tr>
<tr>
<td>Q75</td>
<td>$LR_{ucp}$</td>
<td>0.015</td>
<td>0.842</td>
</tr>
<tr>
<td></td>
<td>$LR_{indp}$</td>
<td>0.747</td>
<td>0.479</td>
</tr>
<tr>
<td>Q95</td>
<td>$LR_{ucp}$</td>
<td>0.189</td>
<td>0.942</td>
</tr>
<tr>
<td></td>
<td>$LR_{indp}$</td>
<td>0.928</td>
<td>0.681</td>
</tr>
</tbody>
</table>

$LR_{ucp}$ refers to Kupiec’s (1995) test statistics for unconditional coverage testing. $LR_{indp}$ is Christoffersen’s (1998) test statistics for Independence testing. The acronyms Q5 denotes the 5% quantile of the summary statistics of P-Value.
Table 1.7: Comparison of $\Delta CoVaR$ and $MES$ Measures (2000/01/03 ~ 2011/12/30)

<table>
<thead>
<tr>
<th>Distress Definition</th>
<th>$r_{i,t} = VaR_{i}^{\tau}(\tau)$</th>
<th>$\Delta CoVaR^\tau(\tau)$</th>
<th>$\tau = 0.05$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Margin-Copula</td>
<td>Margin-Copula</td>
<td>Margin-Copula</td>
</tr>
<tr>
<td>Panel A</td>
<td>AB(quantile)</td>
<td>Norm - Norm</td>
<td>Skewt - Norm</td>
</tr>
<tr>
<td>Overall</td>
<td>1.73</td>
<td>1.68</td>
<td>2.28</td>
</tr>
<tr>
<td>Depository</td>
<td>1.95</td>
<td>1.92</td>
<td>2.52</td>
</tr>
<tr>
<td>Insurance</td>
<td>1.55</td>
<td>1.46</td>
<td>2.06</td>
</tr>
<tr>
<td>Broker-Dealer</td>
<td>1.68</td>
<td>1.70</td>
<td>2.33</td>
</tr>
<tr>
<td>Others</td>
<td>1.67</td>
<td>1.61</td>
<td>2.22</td>
</tr>
</tbody>
</table>

AB(quantile) refer to the estimation of $\Delta CoVaR$ by quantile regression following Adrian and Brunnermeier (2011). Norm-Norm represents bivariate normal joint distribution. As proposition 1 shows, the close form solution in this case is $\Delta CoVaR = \rho \sigma_{m} F_{\epsilon}^{-1}(\tau)$

<table>
<thead>
<tr>
<th>Distress Definition</th>
<th>$r_{i,t} \leq VaR_{i}^{\tau}(\tau)$</th>
<th>$\Delta CoVaR^\tau(\tau)$</th>
<th>$\tau = 0.05$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Margin-Copula</td>
<td>Margin-Copula</td>
<td>Margin-Copula</td>
</tr>
<tr>
<td>Panel B</td>
<td>Norm - Norm</td>
<td>Skewt - Norm</td>
<td>Skewt - T</td>
</tr>
<tr>
<td>Overall</td>
<td>1.16</td>
<td>1.81</td>
<td>1.93</td>
</tr>
<tr>
<td>Depository</td>
<td>1.27</td>
<td>2.00</td>
<td>2.02</td>
</tr>
<tr>
<td>Insurance</td>
<td>1.05</td>
<td>1.64</td>
<td>1.84</td>
</tr>
<tr>
<td>Broker-Dealer</td>
<td>1.17</td>
<td>1.85</td>
<td>1.95</td>
</tr>
<tr>
<td>Others</td>
<td>1.13</td>
<td>1.76</td>
<td>1.92</td>
</tr>
</tbody>
</table>

$MES = E(r_{it}|r_{mt} \leq q_{0.05})$

<table>
<thead>
<tr>
<th>Distress Definition</th>
<th>$r_{i,t} \leq VaR_{i}^{\tau}(\tau)$</th>
<th>$\Delta CoVaR^\tau(\tau)$</th>
<th>$\tau = 0.05$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Margin-Copula</td>
<td>Margin-Copula</td>
<td>Margin-Copula</td>
</tr>
<tr>
<td>Panel C</td>
<td>Norm - Norm</td>
<td>Skewt - Norm</td>
<td>Skewt - T</td>
</tr>
<tr>
<td>Overall</td>
<td>3.38</td>
<td>3.75</td>
<td>4.23</td>
</tr>
<tr>
<td>Depository</td>
<td>3.52</td>
<td>3.90</td>
<td>4.27</td>
</tr>
<tr>
<td>Insurance</td>
<td>2.80</td>
<td>3.13</td>
<td>3.63</td>
</tr>
<tr>
<td>Broker-Dealer</td>
<td>4.26</td>
<td>4.78</td>
<td>5.37</td>
</tr>
<tr>
<td>Others</td>
<td>3.57</td>
<td>3.91</td>
<td>4.48</td>
</tr>
</tbody>
</table>

Norm-Norm represents bivariate normal joint distribution. The close form solution in this case is $MES = E(r_{it}|r_{mt} \leq q_{0.05}) = \rho \sigma_{m} E(r_{mt}|r_{mt} \leq q_{0.05})$.  

71
Table 1.8: **Systemic Risk Rankings**: $\Delta CoVaR^\triangleleft$, Tail Dependence, $MES$, $\beta$ and SRISK

<table>
<thead>
<tr>
<th>Rank</th>
<th>$\Delta CoVaR^\triangleleft$</th>
<th>Tail-Dep</th>
<th>MES</th>
<th>$\beta$</th>
<th>SRISK</th>
<th>LEV</th>
<th>MV</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>NYX</td>
<td>JPM</td>
<td>AIG</td>
<td>AIG</td>
<td>C</td>
<td>FRE</td>
<td>BAC</td>
</tr>
<tr>
<td>2</td>
<td>BAC</td>
<td>C</td>
<td>LEH</td>
<td>LEH</td>
<td>JPM</td>
<td>FNM</td>
<td>JPM</td>
</tr>
<tr>
<td>3</td>
<td>JPM</td>
<td>BAC</td>
<td>WB</td>
<td>WB</td>
<td>BAC</td>
<td>NCC</td>
<td>WFC</td>
</tr>
<tr>
<td>4</td>
<td>C</td>
<td>TROW</td>
<td>FNM</td>
<td>FNM</td>
<td>AIG</td>
<td>MS</td>
<td>C</td>
</tr>
<tr>
<td>5</td>
<td>GNW</td>
<td>BEN</td>
<td>FRE</td>
<td>FRE</td>
<td>FNM</td>
<td>WB</td>
<td>GS</td>
</tr>
<tr>
<td>6</td>
<td>AIZ</td>
<td>LM</td>
<td>BAC</td>
<td>BAC</td>
<td>MS</td>
<td>AIG</td>
<td>MET</td>
</tr>
<tr>
<td>7</td>
<td>TROW</td>
<td>AXP</td>
<td>C</td>
<td>C</td>
<td>FRE</td>
<td>ETF</td>
<td>PRU</td>
</tr>
<tr>
<td>8</td>
<td>AXP</td>
<td>PFG</td>
<td>ABK</td>
<td>ABK</td>
<td>GS</td>
<td>SLM</td>
<td>MER</td>
</tr>
<tr>
<td>9</td>
<td>TMK</td>
<td>PRU</td>
<td>MS</td>
<td>MS</td>
<td>MER</td>
<td>MER</td>
<td>STT</td>
</tr>
<tr>
<td>10</td>
<td>FNF</td>
<td>UNM</td>
<td>MBI</td>
<td>SLM</td>
<td>WB</td>
<td>C</td>
<td>AIG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pairs</th>
<th>$\Delta CoVaR^\triangleleft$</th>
<th>Tail-Dep</th>
<th>MES</th>
<th>$\beta$</th>
<th>SRISK</th>
<th>LEV</th>
<th>MV</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta CoVaR^\triangleleft$</td>
<td>-</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Tail-Dep</td>
<td>5</td>
<td>-</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>MES</td>
<td>2</td>
<td>2</td>
<td>-</td>
<td>9</td>
<td>7</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>$\beta$</td>
<td>2</td>
<td>2</td>
<td>9</td>
<td>-</td>
<td>7</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>SRISK</td>
<td>3</td>
<td>3</td>
<td>7</td>
<td>7</td>
<td>-</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>LEV</td>
<td>1</td>
<td>1</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>-</td>
<td>3</td>
</tr>
<tr>
<td>MV</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>-</td>
</tr>
</tbody>
</table>

Notes: The upper panel displays the ranking of the top 10 financial institutions based on the estimation of $\Delta CoVaR^\triangleleft$, Tail Dependence (Tail-Dep), $MES$, conditional CAPM $\beta$ and SRISK which are all estimated by the student t copula with marginal distribution being skewed student t distribution. The last two columns display the firms’ characteristics: Leverage (LEV) and Market Capitalization (MV). In the lower panel, we report the number of concordant pair between every two systemic risk measure. The ranking is implemented on September 17, 2008 when Lehman Brothers filed for bankruptcy.
Table 1.9: Rank similarity ratio for the top 10 most risky Institutions

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta CoVaR^\infty$ vs Tail-DeP</td>
<td>0.60</td>
<td>0.60</td>
<td>0.70</td>
<td>0.50</td>
<td>0.50</td>
<td>0.70</td>
<td>0.60</td>
</tr>
<tr>
<td>MES vs Tail-DeP</td>
<td>0.30</td>
<td>0</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>SRISK vs Tail-DeP</td>
<td>0.40</td>
<td>0.30</td>
<td>0.40</td>
<td>0.40</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>$\Delta CoVaR^\infty$ vs $\beta$</td>
<td>0</td>
<td>0</td>
<td>0.40</td>
<td>0</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>MES vs $\beta$</td>
<td><strong>0.90</strong></td>
<td><strong>0.90</strong></td>
<td><strong>1.00</strong></td>
<td><strong>1.00</strong></td>
<td><strong>0.90</strong></td>
<td><strong>0.90</strong></td>
<td><strong>0.90</strong></td>
</tr>
<tr>
<td>SRISK vs $\beta$</td>
<td>0.30</td>
<td>0.30</td>
<td>0.40</td>
<td>0.30</td>
<td>0.70</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>$\Delta CoVaR^\infty$ vs LEV</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.10</td>
<td>0.10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MES vs LEV</td>
<td>0.30</td>
<td>0.40</td>
<td>0.50</td>
<td>0.40</td>
<td>0.50</td>
<td>0.30</td>
<td>0.40</td>
</tr>
<tr>
<td>SRISK vs LEV</td>
<td><strong>0.80</strong></td>
<td><strong>0.60</strong></td>
<td><strong>0.60</strong></td>
<td><strong>0.50</strong></td>
<td><strong>0.70</strong></td>
<td><strong>0.30</strong></td>
<td><strong>0.50</strong></td>
</tr>
<tr>
<td>$\Delta CoVaR^\infty$ vs MES</td>
<td>0</td>
<td>0</td>
<td>0.10</td>
<td>0</td>
<td>0.20</td>
<td>0.20</td>
<td>0.10</td>
</tr>
<tr>
<td>$\Delta CoVaR^\infty$ vs SRISK</td>
<td>0.20</td>
<td>0.30</td>
<td>0.20</td>
<td>0.40</td>
<td>0.30</td>
<td>0.30</td>
<td>0.20</td>
</tr>
<tr>
<td>MES vs SRISK</td>
<td>0.30</td>
<td>0.30</td>
<td>0.40</td>
<td>0.30</td>
<td>0.70</td>
<td>0.30</td>
<td>0.20</td>
</tr>
</tbody>
</table>

Notes: The table presents the rank similarity measure between each pair of systemic risk measure displayed in the first column. The similarity ratio measure is simply defined as the proportions of top 10 risky firms that are simultaneously identified as SIFIs by the two rankings at a given date. The acronyms “Tail-Dep” and “LEV” represent tail dependence and Leverage, respectively.
Bibliography


Chapter 2

Model Combination or Distributional Aggregation?
A Quantile Regression Approach to Stock Return Prediction

2.1 Introduction

Stock return predictability has become one of the most debated and high-profile issues in the past few decades. Forecasting the conditional mean or even the entire distribution of future equity premia is of profound importance for many aspects of finance such as portfolio choice and risk management. It is thus not surprising that academia and practitioners have devoted a vast amount of efforts attempting to understand the nature of stock return predictability and increase the forecast performance of asset pricing models. Despite being at the forefront of the current research agenda, stock return predictability at both time series and cross sectional level, is still an evergreen research area in finance. One of the most biased aspects in the current return forecast research is that most predictive models only describes the average relationship of return with a set of predictors. Yet much less is known about the predictability of other parts of the return distribution beyond the conditional mean and variance. However, from the standpoint of practitioners in risk management, the information contained in the lower tail of the distribution usually is considered to be most important in explaining the whole profile of the risk return
trade-off\footnote{In asset pricing model, higher order moments such as skewness and kurtosis have proven useful to explain variation of the risk premium (see e.g. Harvey and Siddique (2000), Dittmar (2002))}. Meanwhile, knowledge of the first two moments frequently is not adequate for portfolio management under standard preferences such as constant relative risk aversion (CRRA), which generally requires the estimation of the entire distribution of future returns. Furthermore, even the interest is exclusively focused on the first two moments (conditional mean and volatility) estimation, aggregating the multiple quantiles information across the entire distribution can potentially improve estimation efficiency for the variables concerned (see e.g. Zou and Yuan (2008), and Xiao and Zhao (2011)). In this sense, understanding the return predictability in the whole distribution is of fundamental significance in many aspects of finance economics.

Even though stock return forecasting is promising and fascinating, the results of empirical studies are quite frustrating. There is ample empirical evidence indicating that US aggregate market return is fairly predictable for the in-sample test. However, when it comes to out-of-sample forecast exercises, the predictability of a variety of popular economic variables disappears. Goyal and Welch (2008) found that most of commonly used macro-finance variables used in the literature fail to deliver a consistent superior forecast than the historical average benchmark model. This finding casts doubt on the reliability of stock return predictability. Recognizing that no simple individual model is a complete description of reality, Timmermann (2006), Rapach, Strauss and Zhou (2010), among others, proposed a variety of models combination approaches and found that pooling information across multiple models tends to produce a superior forecast to that based on a single predictive model. As a result, model combination has become a standard technique attempting to improve the forecast performance of predictive regression models in the horse race of stock return prediction. However, surprisingly very few studies recognized or acknowledged the significance of information within the model (distribution). In the absence of a Gaussian distribution, optimally integrating distributional informa-
tion turns out to produce a robust and more efficient forecast than the least square estimation. Pohlman and Ma (2010) briefly investigated how to employ the distributional information for return forecasting and portfolio construction. Even though the macro-finance variables considered in the current literature appear to have little value in forecasting the conditional mean of stock return (Goyal and Welch (2008)), their predictability ability can be strong in the other part (either left or right tail) of the stock return distribution (Cenensizoglu and Timmermann (2008)). Therefore, optimally combining the information from other parts of distribution may provide a positive contribution in forecasting either the conditional mean or the entire distribution. To explore this possibility, I consider predictive quantile regression models for equity premium forecasting.

Since the seminar work of Koenker and Bassett (1978), quantile regression approaches have been intensively used in the area of financial economics, especially in research on risk management (Chernozhukov and Umantsev (2002), Engle and Mangenelli (2004), Rubia and Sanchis-Marco (2013)), asset pricing models (Barnes and Hughes (2002), Pohlman and Ma (2010), Meligkotsidou, Vrontos and Vrontos (2009)) and stock return prediction (Cenensizoglu and Timmermann (2008), Maynard, Shimotsu and Wang (2011)). Pedersen (2010) provides detailed insights into predictability of the entire stock and bond return distribution through the use of quantile regression and found that the state variables primarily predict only location shifts in the stock return distribution, while they also predict changes in higher-order moments in the bond return distribution. Cenensizoglu and Timmermann (2012) compares statistical and economic measures of forecasting performance across a large set of stock return prediction models and found that the message conveyed from a model’s statistical RMSE performance can only be weakly associated with that emerging from an analysis of the models’ economic value. Recently, Baur, Dimpfl and Jun (2012) provides a comprehensive description of the dependence pattern of stock
returns by studying a range of quantiles of the conditional return distribution using quantile autoregression and found that lower quantile exhibit positive dependence on past returns while upper quantiles are marked by negative dependence.

This chapter is most closely related to the studies from Meligkotsidou et al. (2012) who proposed a robust point forecast by combining a set of quantile forecasts using a simple constructed time varying weighting scheme. The author found that exploiting the distributional information associated with each predictor can deliver statistically and economically significant out of sample forecasts relative to the historical benchmark average.

The main contribution of this chapter, however, is to provide a more comprehensive investigation on how to combine quantile information more efficiently in the forecast of conditional mean as well as density estimation, which is proven to be more advantageous than the simple weighting schemes used by Meligkotsidou et al. (2012). The goal of this paper is not attempting to create a “super” model to produce superior forecast in any situation, but to show that, not only the information across different models, but also the information within the model (distribution) contributes to the improvement of forecast performance. More specifically, incorporating distributional information combined with model information can produce a superior forecast for the conditional mean as well as the entire distribution of future equity premia, which significantly outperforms the forecast that utilizes either source of information alone. Meanwhile, the order of combination strategies appears to make a significant difference in the efficiency of pooling both distributional information and model information. It turns out that aggregating distributional information in the first step, followed by combining model information, is more advantageous in return forecasting than the combination strategies which revert the combination order. This chapter not only investigates the forecast of conditional mean, but also studies the forecast of the whole distribution of future stock returns, which incorporates the entire distribution-
al information provided by a find grid of quantiles forecasts for density estimation. Different test statistics are discussed to evaluate the performance of conditional mean and density forecasts.

This chapter also studies the economic evaluation of predictive performance. I attempt to investigate whether the statistical evidence of stock return predictability can be converted automatically to a significant utility gain for risk-averse investors. The investor is assumed to have either mean variance utility or power utility, which require the estimation of either the first two moments or the entire distribution. Different forecasts result in different optimal portfolio choice which in turn give rise to different realized utility gains. This chapter shows that pooling distributional information via quantile combination does add significant economic values in term of average utility gain and certainty equivalent rate. This improvement of economic value is robust to the choice of the risk aversion coefficient. However, the relative utility gain decreases as investors become more risk averse, as the accuracy of forecast in tail of distribution declines.

The rest of this chapter is organized as follows: Section 2 describes the predictive regression model which will be used in the forecast of the equity premia. In-sample test of quantile forecast performance is discussed briefly. Section 3 outlines a variety of quantile combination approaches and illustrates how to pool distributional information efficiently. Section 4 discusses a few most widely used model combination approaches. Section 5 extends to discuss how to forecast the entire density by taking advantage of the whole distributional information via quantile regression. Section 6 presents how to evaluate forecast performance in term of both statistical and economic criteria. Section 7 report the empirical studies for the proposed forecast models which pool both distributional and model information. Section 8 concludes.
2.2 Predictive Regression Model

The predictive regression model for the equity premium can be written as

\[ y_t = \mu_{t-1} + \sigma_{t-1} \epsilon_t \quad \text{with} \quad \epsilon_t \sim iid \ F_{\epsilon}(0,1) \quad (2.2.0.1) \]

where \( \mu_{t-1} \) and \( \sigma_{t-1} \) signify the location (mean) and scale (volatility) for the distribution of \( y \) conditional on information \( F_{t-1} \). \( \epsilon_t \) is a IID innovation process with mean zero and variance one.

Assume that the predictive variables \( x_{t-1} \) not only significantly influence the conditional mean (location shift effect), but also affect the conditional volatility (scale effect) of stock return \( y_t \). This is indicative of some form of heteroscedasticity.

\[ \mu_{t-1} = \beta_0 + \beta_1 x_{t-1} \quad (2.2.0.2) \]
\[ \sigma_{t-1} = \theta_0 + \theta_1 x_{t-1} \quad (2.2.0.3) \]

Therefore the data generating process (DGP) with conditional mean and variance dynamics can be defined as

\[ y_t = (\beta_0 + \beta_1 x_{t-1}) + (\theta_0 + \theta_1 x_{t-1}) \epsilon_t \quad \text{with} \quad \epsilon_t \sim iid \ F_{\epsilon}(0,1) \quad (2.2.0.4) \]

The least squares estimation (OLS) can only obtain the estimate of the central part of return distribution, but not the entire part of conditional distribution of returns, which could fail to provide insight of the whole profile of the risk-return relation.

\[ E(y_t|F_{t-1}) = \beta_0 + \beta_1 x_{t-1} \quad (2.2.0.5) \]

Suppose the predictor state variable \( x_{t-1} \) has a significantly positive influence on stock returns, Least Squares estimation is unable to tell whether the increase of stock return
$y_{t+1}$ in next period is associated with the increase of risk (scale shift on left tail) or is merely associated by the entire return distribution being shifted to the right (location shift of the entire distribution). The model prediction of volatility $\sigma_t$ can partially provide further insight for measuring the market risk. However, the concentration of study only on the first two moments of distribution is too restrictive and inadequate to conceal the underlying true return and risk relation. In other word, forecasting the entire distribution of stock return is more crucial to completely describe the whole profile of return and risk relation, which is key to understanding the price of financial assets and determining the optimal portfolio of multiple assets. In contrast, quantile regression introduced by Koenker and Bassett (1978) provides significant flexibility in the estimation of the entire return distribution without imposing restrictive distribution assumption. With the setup of data generating process in the above equation (2.4), the predictive quantile model of stock returns can be constructed as

$$Q_{y_t}(\tau|\mathcal{F}_{t-1}) = \beta_0 + \beta_1 x_{t-1} + (\theta_0 + \theta_1 x_{t-1}) Q_{\varepsilon_t}(\tau)$$

$$= \beta_{0,\tau} + \beta_{1,\tau} x_{t-1}$$

where $\beta_{0,\tau} = \beta_0 + \theta_0 Q_{\varepsilon_t}(\tau)$ and $\beta_{1,\tau} = \beta_1 + \theta_1 Q_{\varepsilon_t}(\tau)$. If the predictor variable $x_{t-1}$ is positively correlated with the volatility ($\theta_1 > 0$), the slope coefficient $\beta_\tau$ increases as $\tau$ increases from 0 to 1. By varying quantile $\tau$ from 0 to 1, a complete picture of the covariate effect on return distribution can be obtained.

Empirical results from stock return prediction literatures suggest that the conditional mean slope coefficient $\beta_1$ is insignificant from zero $\beta_1 \approx 0$. Goyal and Welch (2008) shows that all of macro-finance predictor variables appear to have little value in forecasting the mean of stock returns. Cenensizoglu and Timmermann (2008), however, provides statistical evidence that most predictor variables are useful in pre-
dicting either the left, or the right tails of stock return distribution, but not the central part of the distribution. Therefore $\beta_{1,\tau} = \beta_{1} + \theta_{1}Q_{\epsilon_{t}}(\tau) \approx \theta_{1}Q_{\epsilon_{t}}(\tau)$, which implies that the absolute value of $\beta_{1,\tau}$ increases as quantile $\tau$ converges to extreme value ($\tau \rightarrow 0$ or $\tau \rightarrow 1$)$^2$. This may explain why the location (mean or median) of stock returns tends to be much more difficult to predict than the extreme tail (lower or upper) of the distribution.

Koenker and Bassett (1978) introduced a seminal work on quantile regression estimation

$$\left[\beta_{0,\tau}, \beta_{1,\tau}\right] = \arg\max_{\beta_{0,\tau}, \beta_{1,\tau}} \sum_{t=1}^{T-1} \rho_{\tau}(y_{t} - \beta_{0,\tau} - \beta_{1,\tau}x_{t-1})$$

where $T$ is the sample size and $\rho_{\tau}(.)$ indicates the quantile check loss function

$$\rho_{\tau}(y_{t} - \beta_{0,\tau} - \beta_{1,\tau}x_{t-1}) = \left(\tau - \mathcal{I}(y_{t} - \beta_{0,\tau} - \beta_{1,\tau}x_{t-1})\right)\left(y_{t} - \beta_{0,\tau} - \beta_{1,\tau}x_{t-1}\right) \quad (2.2.0.6)$$

with $\mathcal{I}(.)$ being indicator function. In the special case where $\tau = 0.5$, the check loss function reduces to the absolute loss function which is exactly used in the median regression model. The inference of consistency and asymptotical normality for quantile estimation is well established in the literature (i.e. see Koenker (2005)).

### 2.2.1 Predictive Model Selection in Quantile Regression Setting

The stock return prediction literature proposed a large family of macro-finance variables, attempting to forecast the conditional mean or entire distribution of eq-
uity premia. In these forecasting models, the number of all possible regressors $N$ is usually very large compared with sample size. Instead of including all of these predictor variables in forecasting models (Kitchen Sink Forecast), much effort has been devoted to establish statistical model selection criterion to choose the optimal combination out of $2^N - 1$ possible outcomes. Unlike the conditional mean regression model, the commonly used model selection criteria like the Bayes Information Criterion (BIC) or the Akaike Information Criterion (AIC), are not available in a quantile setting. Belloni and Chernozhukov (2011), among others, introduced a data-driven selection of relevant covariates by employing a statistical shrinkage technique which is known as the least absolute shrinkage and selection operator (LASSO). Accordingly, the $\ell_1$ penalized regression estimator $\hat{\beta}(\tau)$ is a solution to the following optimization problem:

$$
\hat{\beta}(\tau) = \arg\max_\beta \frac{1}{T} \sum_{t=1}^{T} \hat{Q}_\tau(Y_t - X_t \beta) + \frac{\lambda \sqrt{\tau(1-\tau)}}{T} \sum_{j=1}^{N} \hat{\sigma}_j |\beta_j| \tag{2.2.1.1}
$$

where $\hat{\sigma}_j = E_T(x_{i,j}^2) = \frac{1}{T} \sum_{t=1}^{T} x_{i,j}^2$ is component-wise variation. $\hat{Q}_\tau(.)$ is the common check loss function used in the ordinary quantile regression. Therefore, the Lasso model selection is implemented by minimizing the sum of check loss function with a penalty function given by a scale $\ell_1$ norm of the parameter space, which selects the relevant covariates based on the absolute value of their shrunk coefficients (scaled by regressor variation). Increasing the penalty level will cause more and more of the parameters to be driven to zero. In the following model selection exercise, I eliminate the regressors in the forecast model if their absolute value of shrunk coefficient is below $10^{-6}$. Specifically, the support of parameter space $\hat{T}_\tau$ can be set as

$$
\hat{T}_\tau = \text{Support}(\hat{\beta}(\tau)) = \left\{ i \in \{1, 2, \cdots, N\} : |\hat{\beta}_i(\tau)| > 10^{-6} \right\} \tag{2.2.1.2}
$$
After removing all regressors that are not selected, the unrestricted quantile model is re-estimated with only including the selected relevant regressors. The post-LASSO estimator \( \widehat{\beta}(\tau) \) therefore can be derived as:

\[
\widehat{\beta}(\tau) = \arg\max_{\beta \in \hat{T}_\tau} \frac{1}{T} \sum_{t=1}^{T} \hat{Q}_\tau(Y_t - X_t \beta)
\]  

(2.2.1.3)

The perfect model selection entails that the estimated parameter space \( \hat{T}_\tau \) coincides with the true one \( \hat{T}_\tau = T_\tau \). However, this goal might be unlikely to achieve for many designs of interest. Even for this scenario, Belloni and Chernozhukov (2011) shows that the rate of converge for the post-penalized estimators perform quite well.

The success of relevant regressors selection via the LASSO procedure also depends on the precise determinant of penalty parameter \( \lambda \). The overall penalty level \( \lambda \sqrt{\tau(1-\tau)} \) is a function of the quantile of interest, in which \( \lambda \) itself is also quantile specific value. Larger value of \( \lambda(\tau) \) will result in more regressors eliminated. When \( \lambda = 0 \), the LASSO procedure is equivalent to ordinary quantile regression. Belloni and Chernozhukov (2011), Hautsch, Schaumburg and Schienle (2012) provide a detailed discussion on the selection of the data-driven penalty parameter \( \lambda(\tau) \). For more detail of empirical procedure, I refer to Appendix 2.

2.3 Robust Point Forecasts Based on Optimally Combining Quantile Information

When the data are exactly normal distributed, the most widely used approaches of least square estimation (LS) coincides with a maximum likelihood interpretation and can be considered as the most efficient estimator. However, in the absence of a Gaussian error distribution, the LS method usually fails to provide efficient esti-
mation and may have extremely poor performance when the data have a heavy-tail distribution with large amount of outlier observations. It is well known that financial data are usually characterized by the high volatility, skewness, fat tails and a significant departure from normal distribution. The standard conditional expectation estimators notoriously places too much weight on extreme observations (equal weight on the deviation from mean), rendering the LS method inadequate to forecast financial series. In the presence of outlier observation, it is well known that the Least Absolute Deviation (LAD) estimation is more robust relative to OLS. However, \( \text{LAD} \) estimation corresponds to only one specific quantile (\( \tau = 0.5 \)), which ignores other parts of the distribution that may contain more important information when it comes to the efficient quantification of risk. In addition, even in the case when median regression is more efficient than estimates at other part of quantile, combining multiple quantile estimates may further improve the efficiency of estimation. More generally, a more efficient estimator can be obtained by optimally aggregating the distributional information from multiple quantile estimators. Along this direction, Zou and Yuan (2008) proposes the composite quantile regression (\( \text{CQR} \)) for parameter estimation and variable selection in the classical linear regression models and demonstrated that combining multiple quantile information enjoys great advantages in terms of estimation efficiency. Kai, Li and Zou (2010) proposes a local composite quantile regression smoothing to further improve local polynomial regression. More recently, Xiao and Zhao (2011) develop an efficient quantile combination approach and demonstrated that the asymptotical variance of new estimator derived by optimally combining multiple quantile information approaches the Cramer-Rao lower bound under approximate conditions and thus leads to advantageous efficient estimation.

Furthermore, it is reasonable to believe that the effects of a predictive variable may differ across quantiles of distribution. In certain situations, some quantiles may
deliver better estimations than others. In other word, the sensitivity of response to factors may be quite different in the tail than the central part of distribution. Cen-nsizoglu and Timmermann (2008) found that the tail behavior of the stock market is more predictable than the central part of distribution. At the time of a Financial Crisis, all factors tend to move together in the same direction. The dependence of response and factor are usually greater and easier to be captured than that at the time of economic prosperity or normal times. Combining multiple quantiles may aggregate more useful information from the dependence structure between response and factor. From this perspective, the performance of quantile regression at a specific quantile $\tau$ can be further improved by incorporating distributional information from multiple quantiles.

Unlike the LS method which focuses on the conditional mean, quantile regression provides the flexibility to estimate the entire conditional distribution and thus naturally offers a framework to aggregate distributional information across multiple quantiles. In general, the forecast of the $\tau^{th}$ conditional quantile of the distribution of the equity premium $Y_{t+1}(\tau)$, conditional on the $i^{th}$ predictor $X_{i,t}$, can be obtained as

$$\hat{Y}_{t+1}(\tau) = \hat{\beta}_0(\tau) + \hat{\beta}_1(\tau)X_{i,t}$$

(2.3.0.4)

By choosing a fine grid of quantiles, say, $\tau = 0.01, \ldots, 0.99$, one can trace out the entire distribution of equity premium when conditioning information shifts more than just the location of the distribution. Now a question can be raised naturally: how can we take advantage of the distributional information associated with quantile regressions to produce a more efficient and robust point predictions on the conditional mean of the equity premis $E(Y_{t+1}|X_{i,t})$.

Basically, the combination of quantile information can be implemented with re-
spect to either the coefficient estimator $\beta(\tau)$ or point forecast $Y_{t+1}(\tau)$ directly$^3$. In statistics literature, much effort has been devoted to the robust estimation or prediction via quantile combination. In the following, I propose several quantile combination approaches, attempting to aggregate distributional information from multiple quantile estimation to produce a superior forecast for equity premium both in terms of statistical and economic evaluation.

2.3.1 Combining Quantile Forecasts of $Y(\tau)$

Early literature on robust point estimation from regression quantiles can be traced to Gastwirth (1966), Judge et al. (1988), which proposed a fixed weighting scheme to combine a set of quantile forecasts. In general, the quantile combination predictor $\hat{Y}_{t+1}$ can be proposed as:

$$\hat{Y}_t = \sum_{i=1}^{N} \hat{\omega}_{i,t} \hat{Y}_t(\tau_i)$$

Where $\sum_{i=1}^{N} \hat{\omega}_{i,t} = 1$ and $\hat{\omega}_{i,t} > 0$ for all $i = 1 \cdots N$ (2.3.1.1)

where $\hat{\omega}_{i,t}$ is the estimated weight for a particular $\tau_i$th quantile forecast $\hat{Y}_t(\tau_i)$. $N$ is the number of quantile to be combined. Here the weights $\omega_{i,t}$ can be considered as the probabilities assigned to different quantile forecasts, suggesting how likely the specific regression quantile $\hat{Y}_t(\tau_i)$ coincides with (or predicts) the realized return for the next period. By carefully choosing the combination weight $\omega_{i,t}$, the above robust forecast can incorporate distributional information from multiple quantile estimates, and therefore can be more accurate than the conditional expectation forecasts especially when the return distribution deviates significantly from normality.

$^3$This chapter concentrates on the forecast of equity premium $Y_{t+1}$ rather than the estimation of slope coefficient $\beta$, which is the marginal effect of predictor variable. To this end, the performance evaluation of a variety of quantile combination approaches is measured only by how well the proposed point forecast $Y_{t+1}$ after combining quantile information can predict equity premium $Y$ at next period $t + 1$. 

90
DMSFE Quantile Combination (DMSFE)

Analogous to the discounted mean square forecast error (DMSFE) approaches put forth by Stock and Watson (2004), the first class of combination method can be naturally constructed as (hereafter I call it DMSFE):

$$\omega_{i,t} = \frac{\Omega_{i,t}^{-1}}{\sum_{j=1}^{N} \Omega_{j,t}^{-1}}$$  \hspace{1cm} (2.3.1.2)

where

$$\Omega_{i,t}^{-1} = \sum_{s=P_0}^{t-1} \xi^{t-1-s} \left( Y_s - \hat{Y}_s(\tau_i) \right)^2$$  \hspace{1cm} (2.3.1.3)

where $P_0$ is the start date for out of sample forecast evaluation, $\xi$ is a discount factor which attaches higher weight to the more recent forecasting accuracy. Lower value of $\xi$ indicates less importance of forecast performance in history. This method is in the same spirit of Stock and Watson (2004), who suggested constructing a weighting scheme based on the historical performance of forecasting models. If the historical return more frequently coincided with a specific quantile forecast (the squared forecast error is smaller), it is reasonable to expect that the next period’s return is more likely as well to coincide with this specific quantile. Pohlman and Ma (2010) also claimed that the rank of the return of stock usually does not change dramatically for two consecutive periods. The location of quantile in the previous periods can predict in some degree what is the most likely quantile to be realized in next period.

Regression Approach of Quantile Combination (Regression)

The second weighting scheme of quantile combination is in the same spirit of constrained regression approach of forecast (Granger and Ramanathan (1984)), which can
be estimated straightforwardly by recursively minimizing the mean square forecast error $E(Y_t - \sum_{i=1}^{N} \hat{\omega}_{i,t} \hat{Y}_t(\tau_i))^2$ between the true value $Y_t$ and the linear combination of multiple quantile estimator $\sum_{i=1}^{N} \hat{\omega}_{i,t} \hat{Y}_t(\tau_i)$. In particular, the optimal weight vector $\omega_t$ is chosen to minimize the following optimization problem 4

$$\hat{\omega}_t = \arg\min_{\omega_t} \sum_{s=P_0+1}^{t-1} \left[ Y_s - \sum_{i=1}^{N} \hat{\omega}_{i,t} \hat{Y}_s(\tau_i) \right]^2 \quad \text{and} \quad \sum_{i=1}^{N} \hat{\omega}_{i,t} = 1 \quad \hat{\omega}_{i,t} > 0 \quad \text{for all } i$$

(2.3.1.4)

where $P_0$ is the number of observations (holdout out of sample) used to construct the first forecast. The weight $\hat{\omega}_t$ can further be constrained 5 not to exceed certain lower and upper bounds in order to reduce the volatility and stabilize forecast. Granger and Ramanathan (1984) suggested both constrained and unconstrained regression approaches to estimate weights. As the predictive quantile estimates $\hat{Y}_s(\tau_i)$ are likely to be biased predictors of $Y_s$, I include the intercept in the linear combination of quantile predictors, Therefore the bias is picked up by the intercept and quantile combination estimator $\omega_0 + \sum_{i=1}^{N} \hat{\omega}_{i,t} \hat{Y}_s(\tau_i)$ can generate unbiased predictor of $Y_s$. Since the estimation strategy is analogous to least square regression, I denote this combination scheme as Regression approach.

### 2.3.2 Combining Quantile Estimation of Coefficient $\beta(\tau)$

The above discussion concentrates on how to combine the point quantile forecast $\hat{Y}_t(\tau_i)$. It turns out that quantile combination can be conducted as well on the coefficient estimation $\beta(\tau_i)$. Zou and Yuan (2008), Kai, Li and Zou (2010) and Xiao

---

4The number of quantiles $N$ to be combined can not be too large due to the increased parameter space in the following optimization. Otherwise, the solution to optimal weight $\hat{\omega}_t$ is very likely end up with the local minimum instead of global minimum.

5In the following discussion, I will show that the weight to each quantile is determined by the shape of distribution. The quantile at the tail of distribution can be assigned a larger value than that at the center part of distribution. Without knowledge of distribution information. The constraint restriction should be carefully designed.
and Zhao (2011) all investigated how to combine quantile information to estimate the global parameter $\beta$ more efficiently. Consider a standard conditional linear quantile regression model

$$
\hat{Y}_{t+1}(\tau) = \hat{\alpha}(\tau) + \hat{\beta}(\tau) X_t
$$

(2.3.2.1)

The global coefficient $\hat{\beta}$ can be obtained by the weight average of multiple "local" quantile estimators $\hat{\beta}(\tau_j)$.

$$
\hat{\beta} = \sum_{j=1}^{N} \omega_j \hat{\beta}(\tau_j)
$$

(2.3.2.2)

Thus the forecast of $\hat{Y}_{t+1}$ can be constructed as:

$$
\hat{Y}_{t+1} = \hat{\alpha} + \hat{\beta} X_t \quad \text{where} \quad \hat{\beta} = \sum_{j=1}^{N} \omega_j \hat{\beta}(\tau_j) \quad \hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X}
$$

(2.3.2.3)

I only discuss how to combine quantile information for the robust estimation of the slope coefficient $\beta$. Xiao and Zhao (2011) suggested that the combination weight for the intercept $\alpha$ can be constructed in the same way as that for $\beta$ if the distribution is symmetric. In general, the global intercept can be derived as. $\hat{\alpha} = \sum_{j=1}^{N} \phi_j \hat{\alpha}(\tau_j)$. Without further assumptions on the distribution of $\epsilon_t$, the global parameter $\alpha$ may not be identified. Since my goal is to predict the conditional mean $Y_{t+1}$, setting $\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X}$ can ensure that the predicted value $\hat{Y}_{t+1}$ is an unbiased estimator of $Y_{t+1}$.

Density Weight Quantile Combination (Density)

Following Koenker (2005), consistency and asymptotic normality of quantile coefficients can be established under certain mild regularity conditions. The asymptotic distribution of $\hat{\beta}(\tau)$ are considerably simpler and easier to describe under the
IID model.

$$\sqrt{n} \left( \hat{\beta}(\tau) - \beta(\tau) \right) \sim N \left( 0, \frac{\tau(1-\tau)}{f^2(\hat{Q}_e(\tau))} \mathcal{C}^{-1} \right) \tag{2.3.2.4}$$

where \( \mathcal{C} \) is a positive definite matrix such that \( \mathcal{C} = \lim_{-\infty} \frac{1}{n} X^T X \).

The quantile regression estimator of \( \beta \) at \( \tau \) has an asymptotic variance matrix in the form of

$$\frac{\tau(1-\tau)}{f^2(\hat{Q}_e(\tau))} \mathcal{C}^{-1} \tag{2.3.2.5}$$

The precision of coefficient estimator \( \beta \) at each quantile \( \tau_i \), therefore, can be measured by its inverse asymptotic variance. A higher value of asymptotic variance represents a lower precision of \( \beta_{\tau_i} \). The major source of quantile heteroskedasticity is only reflected by the scalar term

$$\frac{\tau(1-\tau)}{f^2(\hat{Q}_e(\tau))} \tag{2.3.2.6}$$

Therefore, I can weight each quantile estimation of \( \beta(\tau_i) \) by the standardized value of the above scalar term:

$$\omega_i = \frac{f_i(\hat{Q}_e(\tau_i))}{\sqrt{\tau_i(1-\tau_i)}}/S_w \quad \text{where} \quad S_w = \sum_i f_i(\hat{Q}_e(\tau_i))/\sqrt{\tau_i(1-\tau_i)} \tag{2.3.2.7}$$

Note that the weight average \( \omega_i \) now is not a predetermined constant but instead depends on the shape of the distribution, which could be adjusted automatically based on the estimated density function over various quantiles. This is important when the distribution becomes skewed and the density is no longer symmetric.

**Optimal Weight Quantile Combination (OWQ)**

Xiao and Zhao (2011) constructed an estimator that combine distribution information across multiple quantiles. Considering the weighted average of multiple
quantiles estimator

\[ \hat{\beta}_{WQ} = \sum_{j=1}^{N} \omega_j \hat{\beta}(\tau) \quad \text{where} \quad \sum_{j=1}^{N} \omega_j = 1 \quad (2.3.2.8) \]

By some mild regularity conditions, Xiao and Zhao (2011) showed that the weighted average of quantile combination estimator \( \hat{\beta}_{WQ} \) has the following asymptotic distribution:

\[ \sqrt{n} \left( \hat{\beta}_{WQ} - \beta \right) \sim N \left( 0, S(\omega)C^{-1} \right) \quad (2.3.2.9) \]

where \( C \) is a positive definite matrix which is independent of quantile \( \tau \). \( S(\omega) \) is a quadratic form given by

\[ S(w) = w^T H w, \quad \text{where} \quad H = \left( \frac{\min(\tau_i, \tau_j) - \tau_i \tau_j}{f_\epsilon(Q_\epsilon(\tau_i)) f_\epsilon(Q_\epsilon(\tau_j))} \right)_{1 \leq i,j \leq N} = \left( \frac{\min(\tau_i, \tau_j) - \tau_i \tau_j}{f_\epsilon(Q_\epsilon(\tau_i)) f_\epsilon(Q_\epsilon(\tau_j))} \right)_{1 \leq i,j \leq N} \quad (2.3.2.10) \]

The optimal weight can, thus, be chosen by minimizing the asymptotic variance of \( \hat{\beta}_{WQ} \). In other word, the optimal weight \( \omega^* \) for quantile combination can be obtained by solving the following optimization problem:

\[ \omega^* = \arg\min_{\omega} \omega^T H \omega \quad \text{where} \quad H = \left( \frac{\min(\tau_i, \tau_j) - \tau_i \tau_j}{f_\epsilon(Q_\epsilon(\tau_i)) f_\epsilon(Q_\epsilon(\tau_j))} \right)_{1 \leq i,j \leq N} = \frac{H^{-1}e}{e^T H^{-1} e} \quad \text{where} \quad e = (1, 1, \ldots, 1)^T \]

Given a family of estimated conditional quantile functions, it is straightforward to estimate the conditional density of residuals \( f_\epsilon(\hat{Q}_\epsilon(\tau_i)) \) as (see Koenker (2005))

\[ f_\epsilon(\hat{Q}_\epsilon(\tau_i)) = \frac{2h}{Q_\epsilon(\tau_i + h) - Q_\epsilon(\tau_i - h)} \quad (2.3.2.11) \]
where \( h \) is the bandwidth\(^6\) which connects two neighboring points on the cumulative probability function.

The conditional densities can also be estimated by the Epanechnikov kernel, which would generate smooth densities especially when the time series sample size is short. This is going to be the case in our empirical analysis in which lower frequency data (quarterly) are employed.

\[
f_\epsilon\left(\hat{Q}_{\epsilon}(\tau_i)\right) = \frac{1}{Th_n} \sum_{t=1}^{T} K\left( \frac{\epsilon_t - \hat{Q}_{\epsilon}(\tau_i)}{h_n} \right)\quad (2.3.2.12)
\]

where \( \hat{Q}_{\epsilon}(\tau_i) \) is the estimated quantile of \( \epsilon \) at \( \tau_i \), \( K \) is the Epanechnikov kernel density and \( h_n \) is the bandwidth for the density estimation\(^7\).

A natural question can be raised on how to select quantiles \( \tau_1, \cdots, \tau_N \) when combining distributional information from multiple quantiles. There are a couple of issues that deserve highlighting. Chernozhukov (2005) demonstrated that the conventional quantile regression does not work well on the “extreme quantiles”. So the smallest and largest quantile should be chosen sensibly and carefully. Secondly, “extreme quantiles” are considered as a “rare event”, which may be trivial in the estimation of global parameter if distribution has a bell shape. In other word, the information contained in the “extreme quantiles” may be of no use and should be discarded \(^8\) when combining multiple quantiles estimators. More generally, given a small number of \( \xi \), i.e \( \xi = f_\epsilon(\hat{Q}_\epsilon(0.05)) \), we can discard information in the tail below 5% quantile or upper 95% quantile via multiplying the estimated density by the indicator function \( I(|f_\epsilon(\hat{Q}_\epsilon(\tau_i))| > \xi) \). Therefore the truncated version of the estimated density can be

---

\(^6\)bandwidth for each quantile can be estimated by Bofinger (1975) or Hall and Sheather (1988) method where \( h_n = z_{\alpha/2}^2 [1.5\phi^2(\Phi^{-1}(\tau))]/[2(\Phi^{-1}(\tau))^2 + 1]^{1/3} \) and \( z_{\alpha} \) satisfies \( \Phi(z_{\alpha}) = 1 - \alpha/2 \)

\(^7\)we follow Silverman (1986) to choose the rule-of-thumb bandwidth

\(^8\)As I will show in the following Monte Carlo simulation, when the distribution has extreme fat tail, say Cauchy distribution, the optimal weight for the extreme quantile converge to zero and even negative value, indicating negative contribution to aggregating distribution information from multiple quantiles
constructed as:
\[ \tilde{f}_t\left(\hat{Q}_t(\tau_i)\right) = f_t\left(\hat{Q}_t(\tau_i)\right) I\left(|f_t(\hat{Q}_t(\tau_i)| > \xi\right) \] (2.3.2.13)

Furthermore, since the quantile regression curves are estimated independently, the issue of “crossing” in quantile regression can be raised. Higher quantile estimator may exhibit lower value than that of the lower quantile, leading to an invalid distribution for the response. I followed Bondell et al. (2011) approach and estimated the quantiles simultaneously under the non-crossing restriction to avoid the crossing problem for the linear quantile curves.\footnote{The detail of non-crossing quantile regression proposed by Bondell et al. (2011) can be found in the followings link: http://www4.stat.ncsu.edu/ reich/Code/}

Finally, a question on what is the optimal number of quantiles to be combined, or which part of distribution contain most useful information, still remains unanswered. Xiao and Zhao (2011) demonstrated that the asymptotical variance of their proposed quantile combination estimator approaches the efficient Cramer-Rao lower bound if the number of combining quantile \( N \rightarrow \infty \). Undoubtedly, large number of quantiles \( N \) leads to computation complexity and may introduce significant measure error, while small number \( N \) may result in efficiency loss. Fortunately, Xiao and Zhao (2011) showed that the efficiency gain stabilizes quickly as \( N \) increases. Without any prior information of the shape of distribution, they suggested the set of uniformly spaced quantiles with \( N = 9 \) where \( \tau = 0.1, \ldots, \tau = 0.9 \).

**Quantile Combination Illustration**

Figure 2.1 displays the combination weight for the quantile estimator of slope coefficient \( \beta(\tau) \), under a few widely used distribution assumptions. Two combination weighting schemes are compared with each other: (1) the **density** weighting scheme which is based on the asymptotic variance of individual quantile estimator \( \beta(\tau) \); (2)
the \textbf{optimal} weighting scheme that is constructed by minimizing the asymptotic variance of combined quantile estimator $\sum_{j=1}^{N} \omega_j \hat{\beta}(\tau)$. The dispersion of combination weights, as shown in Figure 1, is determined by the shape of distribution. The \textbf{density} weights seem to be more smooth than the \textbf{optimal} weight. The \textbf{optimal} weighted average of quantile combination wins the horse race in the forecast performance relative to \textbf{density} weighted quantile combination, as will be shown in the following Monte Carlo simulation and empirical studies. Thus I only concentrate the discussion on the \textbf{optimal} weighting scheme of quantile combination suggested by Xiao and Zhao (2011).

Corresponding to the benchmark normal distribution, the \textbf{optimal} weighting scheme suggests roughly equal weights for a set of uniformly spaced quantiles from $\tau = 0.1, \cdots, \tau = 0.9$, which indicates that different part of distribution of normality play almost equal roles in aggregating distributional information. In particular, simple equally averaging multiple quantile regression estimators is asymptotically efficient and equivalent to OLS estimation under the normality assumption. In contrast, the optimal combination weight for the fat-tailed student t distribution display a symmetrical pattern with more weight assigned to the center part of distribution, and less weight to both tails of distribution. In the case of extreme fat tail distribution such as standard Cauchy (Student t with $df = 1$), the central part of distribution such as the quantiles $\tau = 0.4, 0.5, 0.6$ contribute almost all useful information. Whereas, the tail of distribution at quantiles $\tau = 0.1, 0.2, 0.8, 0.9$ are assigned even negative weight. With respect to left (right) skewed distribution, the quantile at upper (lower) tail contain almost all useful information. Higher value of weight is assigned to the right (left) tail where the larger mass of the distribution is concentrated. In the extreme case when data are generated by standard log normal distribution (right tail ‘tilt’ extremely rightwards), quantile at the tail $\tau = 0.1$ contains almost all proportional distributional information. The illustration in Figure 2.1 shows that different quan-
tiles contains different proportional distributional information, which is determined by the shape of its density function\textsuperscript{10}.

Table 2.1 presents the Monte Carlo Simulation results for the \textbf{in-sample} forecast performance of coefficient estimation $\hat{\beta}$, which is estimated using a variety of quantile combination approaches discussed above. The simulated data are generated to mimic a wide range of distributions which are characterized by heavy tails, skewness or any other features deviating from normality assumption. In the benchmark normal distribution, OLS estimation has a MLE interpretation and give rise to estimates closest to MLE. With the departure from normality assumption, the slope coefficient estimation $\hat{\beta}$ significantly improved after aggregating distributional information via multiple quantiles combination. In some distributions such as Hansen (1994) Skew-t distribution, the optimal weight quantile combination estimation achieves an efficiency gain even close to MLE estimation. Furthermore, the number of quantiles to be combined seems to have a trivial effect on the estimation efficiency gain. Increasing the number of uniform spaced quantile from 9 to 19 does not result in a significant improvement of estimation efficiency. Lastly, kernel based estimation of density perform quite well in the quantile combination estimation. There are not many differences in the efficiency gain of quantile combination even if the density is estimated by the non-parametric kernel instead of theoretic simulated distribution.

The Monte Carlo simulation in Table 2.1 is implemented in an ideal situation where both regressor and innovation process are independently simulated in the identical distribution. Now it is of interest to see if the efficiency gain from quantile combination still prevails in real data. One of characteristics of stock return forecast

\textsuperscript{10}In some distributions such as a mixture of normal distributions with different mean $u = \omega z_1 + (1 - \omega) z_2$, where $z_1 \sim N(-2, 1)$, $z_2 \sim N(2, 1)$ and $\omega = 0.5$, the quantiles in both tails can outweigh the quantiles in the central part of distribution since the larger mass of distribution concentrates on both tails.
model is the low value of goodness fit of (R2). The model simulated in Table 2 attempts to capture this feature of forecast model in stock return prediction, which is set as

\[ Y_{t+1} = \alpha + \beta X_t + \sigma u_t \]  

(2.3.2.14)

where the regressor \( X_t \) corresponds to one of widely used predictive variables in empirical studies: Dividend Yield Ratio (\( DY \)). The selected slope coefficient \( \beta = 0.0243 \) is the OLS slope coefficient for the univariate forecast with dividend yield ratio as regressor during 1926Q4 – 2010Q4. The scale value \( \sigma \) is selected as \( \sigma^2 = \frac{1-R^2}{R^2} \cdot \beta^2 \cdot Var(X) \) such that the goodness fit (\( R^2 \)) of simulated model is similar to the empirical forecast model with \( R^2 = 1.03\% \). Unlike the in-sample prediction in Table 2.1, all parameters in the simulation of Table 2 are estimated recursively in the out-of-sample forecast exercises, in which the parameter estimates are updated as the forecast moves forward by adding one additional pseudo-observation to the estimation sample at each step. As shown in Table 2.2, in the presence of a fat-tailed distribution, most of the slope coefficient \( \hat{\beta} \) estimates significantly improved after aggregating distributional information via quantile combination (except for Hansen Skew-t distribution). Meanwhile, the estimated weight average of quantile combination outperforms the uniform weighted average of quantile combination. When it turns to the evaluation of the point forecast \( \hat{Y}_{t+1} \), it is noteworthy that the efficiency gain for the slope coefficient \( \hat{\beta} \) estimation is not equivalent to that for point forecast \( \hat{Y}_{t+1} \). In other words, advantageous estimation of coefficient \( \hat{\beta} \) is not automatically converted to a superior estimation of point forecast \( \hat{Y}_{t+1} \) (See Appendix 2.9.1 for the comparison of MSE for predicted \( \hat{\beta} \) and point forecast \( \hat{Y} \). Table 2.2 shows that almost all weighted average of quantile combination estimation produces superior forecasts to OLS estimation in the out of sample forecast exercises.
2.4 Model Combination

Since the seminal work of Bates and Granger (1969), it has been understood that combining forecasts across multiple models often produces more accurate forecasts than a single selected model based on a rigorous model selection procedure. Timmermann (2006) suggested that forecast combination can be considered as a diversification strategy that can reduce uncertainty risk associated with a single predictive model and increase forecast performance in the same manner that asset diversification improves portfolio performance. It is generally believed that a particular forecasting model is unable to take into account all conditioning information during the whole sample period. Instead, the true data generating process is usually unknown and time varying. A particular individual model, more often than not, can only provide a reasonable “local” instead of a “global” approximation. In other words, no single model can dominate other models uniformly over all sample periods.

In contrast, given a collection of multiple models, combining forecasts across different models instead of placing all weight on a single model can provide flexibility to mitigate misspecification biases and measurement error in the data generating process. More specifically, even though a particular model under-perform all other models on average, it can still contribute a positive value to the optimal prediction pool as long as it occasionally although not frequently beats other models. In the study of equity premium predictability, pooling information across multiple models has been proven to be superior to a forecast based on a single model and wins the horse race in stock return forecast competition. In the past few decades, numerous forecast combination approaches have been proposed. (e.g. Timmermann (2006), Rapach, Strauss and Zhou (2010), Hsiao and Wan (2011)). Since the approaches of model combination have been discussed intensively in the current forecasting literature, I will not provide a complete survey of studies for model combination. The goal of this paper is to show that model combination joined with quantile combination
can further improve forecast performance.

In the following sections, I consider several commonly use combination approaches and examine whether model combination improve forecast performance not only for conditional mean but also for a particular quantile forecast. Since combination approaches for conditional mean have been discussed extensively, I concentrate my discussion on the model combination for quantile forecast, which is exactly in line with the combination for conditional mean (The only difference is the loss function in the estimation).

In general, the combination of multiple forecasts at \( \tau \)th quantile \( \hat{Y}_{C,t+1}^{(\tau)} \) can be derived as the weighted average of the \( N \) individual \( \tau \)th quantile forecasts \( \hat{Y}_{i,t+1}^{(\tau)} \) for \( i = 1, 2, \cdots, N \)

\[
\hat{Y}_{C,t+1}^{(\tau)} = \sum_{i=1}^{N} \omega_{i,t}^{(\tau)} \hat{Y}_{i,t+1}^{(\tau)} \tag{2.4.0.15}
\]

where \( \omega_{i,t}^{(\tau)} \) refers to the weight for model \( i \) at time \( t \) and quantile \( \tau \). \( N \) is the number of models to be combined. To prevent some abuse of notation, I omit the subscript \( t \) and \( \tau \) in the following discussion unless it is essential to highlight them.

The simplest combination scheme is **Pool Average**, which just sets \( \omega_{i} = \frac{1}{N} \) and assigns equal weight for all individual model at each quantile. This strategy is analogous to the “naive” portfolio rule that places equal weight on every asset. It turns out that this simple combination strategy performs surprisingly pretty well in the equity premium forecast (e.g. see Timmermann (2006), Rapach, Strauss and Zhou (2010)). To avoid the effect of outliers, the trimmed mean approaches also can be called for which remove outlier observations when taking average of different forecasts.\(^{11}\) The next simple combination scheme I employ is the **Median**, which is nothing but the median of all forecasts across different models: \( \left[ \hat{Y}_{i,t+1}^{(\tau)} \right]_{i=1}^{N} \)

Combination weight can also be “tilted” towards certain individual forecasts

\(^{11}\)I do not see many differences of trimmed mean approach from “Pool Average” in the following study.
which tend to have superior performance in history. In line with the idea of DMSFE proposed by Bates and Granger (1969), the weight for individual model $i$ at quantile $\tau$ also can be constructed as

$$
\omega_i = \frac{\phi_{i,t}^{-1}}{\sum_{j=1}^{N} \phi_{j,t}^{-1}} \quad (2.4.0.16)
$$

where

$$
\phi_{i,t} = \sum_{s=P_0}^{t-1} \xi^{t-1-s} \rho_{\tau} \left( Y_s - \hat{Y}_s(\tau) \right) \quad (2.4.0.17)
$$

where $\xi$ is the discount factor which assigns more weight to more recent forecasting accuracy. $\rho_{\tau}(\cdot)$ is the check loss function for quantile regression at $\tau$. When $\xi = 1$, the forecast accuracy in the whole historical period is evaluated equally. If $\xi < 1$, more recent forecasting performance is considered more important in the evaluation\textsuperscript{12}. Therefore those models who have better performance in the previous period will be assigned more weight in the next period forecast.

Another development in the large forecasting literature has been the recognition that using the first few principal components from a large set of predictor variables can avoid the curse-of-dimensionality problem which plagues many econometric models (Stock and Watson (2004)). Many studies have shown that using the first few principal components can outperform the larger multivariate model which includes too many predictive variables. Ludvigson and Ng (2007), Kelly and Pruitt (2012) and Neely et al. (2012) adopt a latent factor model structure to improve equity premium forecasting. Ando and Tsay (2011) discussed the quantile regression models with factor-augmented predictors. In particular, the quantile forecast of $Y_t$ can be constructed as

$$
\hat{Y}_t(\tau) = \hat{\beta}_0(\tau) + \hat{\beta}_1(\tau) f_{t-1} \quad (2.4.0.18)
$$

where $f_{t-1}$ are the first $r$ principal components of predictor variable $(X_{t-1})^N_1$. Ando and Tsay (2011) proposed a factor selection rule for quantile regression. In my em-

\textsuperscript{12}I set $\xi = 0.9$ and $\xi = 1$ in the following empirical studies.
pirical study, $r$ is set to 1 since the forecast accuracy is worse when $r > 1$.

### 2.5 Combining Information Both Across Models and In Distribution

In the previous sections, I have discussed how to aggregate distributional information via quantile combination and how to pool model information by weighting average multiple different forecasts. In application, these two combinations strategies can be flexibly implemented in different turns in order to take advantage of the information both across models and within distribution. More specifically, the combination strategy can be conducted in the order of model combination first followed by quantile combination afterwards.

\[
\hat{Y}_{t+1} = \frac{1}{J} \sum_{j=1}^{J} \omega_j \left( \frac{1}{I} \sum_{i=1}^{I} \varphi_i(\tau_i) \hat{Y}_j^{i+1}(\tau_i) \right)
\]  

(2.5.0.19)

where $i$ denotes the $i$th quantile at $\tau_i$ and $j$ represents the $j$th model. $\varphi_i(\tau_i)$ is the combination weight for a specific quantile at $\tau_i$, and $\omega_j$ is the averaging weight for the $j$th model.

Alternatively, we can switch the order of combination strategies. First, quantile information is aggregated in each individual model $\hat{Y}^j_{t+1} = \frac{1}{I} \sum_{i=1}^{I} \varphi_i(\tau_i) \hat{Y}^j_{t+1}(\tau_i)$.

In the second step, the point forecast $\hat{Y}_{t+1}$ can be obtained by combining different forecasts for multiple models.

\[
\hat{Y}_{t+1} = \frac{1}{J} \sum_{j=1}^{J} \omega_j \left( \frac{1}{I} \sum_{i=1}^{I} \varphi_i(\tau_i) \hat{Y}^j_{t+1}(\tau_i) \right)
\]  

(2.5.0.20)
Intuitively, these two orders of combination strategies should be equivalent if there are no measurement errors in every step of estimation and all information is completely combined. It is generally believed that the information across different models is much easier to pool (equal weights do a good job!) than the distributional information within the model, which is estimated by quantile regression. Since the goal of this paper is to forecast the conditional mean of the equity premium, model combination in the second step would be much easier to utilize information across different models. In the following empirical studies, I will show that the efficiency gain for the second combination strategy is more advantageous than the first one.

2.5.1 Density Combination

So far, I concentrate my discussion on how to improve the forecast of conditional mean by pooling information either within the distribution or across different models. However, forecasting the first moment is undoubtedly inadequate to measure the whole profile of tradeoffs between market risk and return. From the standpoint of practitioners in risk management or portfolio choice, the information contained in the tail of the distribution sometimes is more important than that in the central part of the distribution. In this sense, the forecast of entire density is of profound importance for many aspects in finance. By choosing a fine grid of quantiles, quantile regression naturally provides a flexible framework to trace out the entire distribution of stock return conditional on covariates $X_{i,t-1}$. Zhao (2013) and Gaglianone and Lima (2012) proposed a nonparametric kernel method to estimate the predictive conditional distribution of stock returns:

$$f(Y_t|X_{i,t-1}) = \frac{1}{nh_n} \sum_{\tau=0.01}^{0.99} K\left(\frac{Y_t - \hat{Q}_{Y_t}(\tau|X_{i,t-1})}{h_n}\right), \quad \text{for } i = 1, 2, \ldots, N \quad (2.5.1.1)$$
where
\[ \hat{Q}_{Y_i}(\tau|X_{i,t-1}) = \beta_{i,0}(\tau) + \beta_{i,1}(\tau)X_{i,t-1} \]  
(2.5.1.2)

Different macro-finance variables \(X_{i,t-1}\) may deliver different quantile estimates \(\hat{Q}_{Y_i}(\tau|X_{i,t-1})\), which results in different density forecasts \(f(Y_i|X_{i,t-1})\) for \(Y_t\). The benchmark model for comparison is the historical kernel density estimation which is based on historical estimation of unconditional quantile regression that disregards any predictive variables.

\[ \bar{Q}_{Y_i}(\tau) = \bar{q}_t(\tau) \]  
(2.5.1.3)

Therefore, the estimation of benchmark kernel density can be set as
\[ f^{Ben}(Y_t) = \frac{1}{nh_n} \sum_{\tau=0.01}^{0.99} K\left(\frac{Y_t - \bar{Q}_{Y_i}(\tau)}{h_n}\right) \]  
(2.5.1.4)

Before employing the distributional information provided by each quantile regression, the information across different models can be pooled firstly by combining multiple forecast models at each specific quantile \(\tau\), which had already been discussed in Section 4.

\[ \hat{Q}^C_{Y_i}(\tau|X_{i,t-1}) = \sum_{i=1}^{N} \omega_{i,t}(\tau) \hat{Q}_{Y_i}(\tau|X_{i,t-1}) \]  
(2.5.1.5)

Therefore the density forecast which first employing model information in the first step, followed by taking advantage of distributional information in the second step, can be constructed as
\[ f^C(Y_t|X_{i,t-1}) = \frac{1}{nh_n} \sum_{\tau=0.01}^{0.99} K\left(\frac{Y_t - \hat{Q}^C_{Y_i}(\tau|X_{i,t-1})}{h_n}\right), \text{ for } i = 1, 2, \cdots, N \]  
(2.5.1.6)

Alternatively, these orders of combination strategies can be reversed, which now can instead be quantile information aggregation in the first step, followed by density
model combination in the second step.

\[ f(Y_t|X_{t-1}, \omega_t) = \sum_{i=1}^{N} \omega_{i,t} f(Y_t|X_{i,t-1}) \]  

(2.5.1.7)

where

\[ f(Y_t|X_{i,t-1}) = \frac{1}{n h_n} \sum_{\tau=0.01}^{0.99} K \left( \frac{Y_t - \hat{Q}_{Y_i}(\tau|X_{i,t-1})}{h_n} \right), \quad \text{for } i = 1, 2, \ldots, N \]  

(2.5.1.8)

where \((X_{i,t-1})_{i=1}^{N}\) denote different macro-finance predictors. \((\omega_{i,t-1})_{i=1}^{N}\) is the weight for different density forecast based on different macro-finance predictors. Geweke and Amisano (2011) suggested that the predictive density \(f(Y_t|X_{t-1}, \omega_t)\) is at least weakly concave, therefore a non-negative weight will be assigned to all participating models in combination pool. Optimally weighting average of different density forecast \(\sum_{i=1}^{N} \omega_{i,t} f(Y_t|X_{i,t-1})\) may deliver more precise estimation of the whole distribution than the density forecast based on the individual source of information.

In the same spirit of model combination for conditional mean, the simplest and "naive" combination strategy is the equally weighted average of a variety of density forecasts

\[ f(Y_t|X_{t-1}, \omega_t) = \frac{1}{N} \sum_{i=1}^{N} f(Y_t|X_{i,t-1}) \]  

(2.5.1.9)

Zhao (2013) proposed an estimation of weighted average which was measured by the predictive accuracy of density forecast over the out-of-sample forecasting period. More specifically, the optimal combination weight \(\omega_{i}^{*}\) for multiple density forecasts can be estimated by maximizing the summed value of log likelihood function during forecast evaluation period (hereafter I call it **Likelihood-Weight Density Combination**).

\[ \omega_{i}^{*} = \arg\max_{\omega_t} \sum_{s=P}^{t-1} \log \left( \sum_{i=1}^{N} \omega_{i,t} f(Y_s|X_{i,s-1}) \right) \quad \text{and} \quad \sum_{i=1}^{N} \hat{\omega}_{i,t} = 1 \quad \hat{\omega}_{i,t} > 0 \quad \text{for all } i \]  

(2.5.1.10)
where \( P \) is the start date for forecast evaluation period. \( N \) is the number of models that are being combined. With the observation of data over the out-of-sample period \( \{Y_P, Y_{P+1}, \ldots, Y_{t-1}\} \), the accuracy of individual density forecast can be measured by its log predictive likelihood. A higher value of likelihood indicates greater weight to be assigned. With the estimation of optimal weight \( \omega^* \), the predictive density of stock return at time \( t \) is the weighted average of the different individual predictive density forecasts by the end of time \( t - 1 \)

\[
f(Y_t|X_{t-1}, \omega^*_t) = \sum_{i=1}^{N} \omega^*_{i,t} f(Y_t|X_{i,t-1}) \tag{2.5.1.11}
\]

Figure 2.2 displays the kernel density forecast \( f(Y_t|X_{i,t-1}) \) at 2010Q4 obtained either from some individual macro-finance variables (i.e. Dividend Price Ratio, Book to Market Ratio, Treasury Bill Rate and Inflation rate) or from density model combination estimations (right panel). The individual density forecast displayed in the left panel of Figure 2.2 looks more jagged and distinct, indicating that different macro-finance variable contain different information for the forecast of stock return density. However, after aggregating both distributional and model information, the density forecast becomes much more smooth and converge to each other, as the right panel of Figure 2 illustrates. Furthermore, the distribution of stock return at 2010Q4 (right panel) looks slightly left skewed and fat tailed compared with the normal density, which was constructed by defining the identical value of first two moments (mean \( \mu \) and variance \( \sigma \)) of the real data at 2010Q4.
2.6 Forecast Evaluation

One of the major issues in the return predictability literatures is how to conduct out-of-sample evaluation for the competing forecasting models. I next briefly describe both the statistical and economic metrics used to measure return predictability.

2.6.1 Statistical Evaluation of Return Predictability

Quantile Forecast Evaluation

The accuracy of predictive regression forecast is compared to the historical benchmark forecast. Following Campbell and Thompson (2008), I define the out-of-sample $R^2$ at quantile $\tau$ as

$$R^2(\tau) = 1 - \frac{\sum_{t=P}^{T} \rho_r(y_t - \hat{q}_{\tau,t})}{\sum_{t=P}^{T} \rho_r(y_t - \bar{q}_{\tau,t})}$$  \hspace{1cm} (2.6.1.1)

where $P$ is the start date for forecast evaluation, $\hat{q}_{\tau,t}$ is the one period ahead prediction of the $\tau$th quantile of the excess return at time $t$ using data from the previous in-sample periods; $\bar{q}_{\tau,t}$ is the simple historical unconditional quantile estimation which is used as the benchmark prediction. Since the estimation of out of sample $R^2(\tau)$ is based on a particular sample, a question naturally arises in any particular sample as to whether the positive value of $R^2(\tau)$ can truly indicate the superiority of competing model to the benchmark. The Diebold and Mariano (1995) (DM) test was designed to allow one to assess the significance of forecast superiority. More specifically, the significance of $R^2(\tau) > 0$ provides the evidence of higher accuracy of predictor model $\hat{q}_{\tau,t}$ than the historical benchmark model in forecasting $\tau$th quantile of the excess return.

In line with Diebold and Mariano (1995), I perform pairwise comparisons of
different models at each quantile $\tau$, based on the check loss function of quantile regression. The time $t$ loss differential of quantile regression between forecast $i$ and $j$ can be constructed as

$$d_{t+1,\tau} = \rho_\tau(e_{i,t+1}) - \rho_\tau(e_{j,t+1})$$

$$= \left(\tau - \mathcal{I}(y_{t+1} - \hat{q}_{i,\tau,t}^i)\right)(y_{t+1} - \hat{q}_{i,\tau,t}^i) - \left(\tau - \mathcal{I}(y_{t+1} - \hat{q}_{j,\tau,t}^j)\right)(y_{t+1} - \hat{q}_{j,\tau,t}^j)$$

(2.6.1.2)

where $\mathcal{I}(.)$ is the indicator function, and $\hat{q}_{i,\tau,t}^i$ and $\hat{q}_{j,\tau,t}^j$ are quantile forecasts for model $i$ and $j$ respectively.

When the models being compared are non-nested, Diebold and Mariano (1995) suggested that the asymptotic distribution of test statistics follows the standard normal distribution

$$t_\tau = \frac{\overline{d}_\tau}{\sqrt{\text{var}(\overline{d}_\tau)}} \sim \mathcal{N}(0,1)$$

(2.6.1.3)

where $\overline{d}_\tau$ is the average of $\{d_{t+1,\tau}\}_{t=P}^T$ over the out-of-sample evaluation period.

However, when the competing models are nested\(^{13}\), which is the exact case in this paper, the asymptotic distribution is no longer normal\(^{14}\). Therefore, a bootstrap p-value was reported for the test statistics. The procedure starts with resampling from $\{(\rho_\tau(e_{i,t+1}), \rho_\tau(e_{j,t+1}))\}_{t=P}^P$ for $N = 1000$ times and create $N$ bootstrap statistics

$$t^\text{boot}_\tau = \frac{\overline{d}^\text{boot}_\tau}{\sqrt{\text{var}(\overline{d}^\text{boot}_\tau)}}$$

Therefore, the $\alpha$ percent p-value can be obtained from the series of $N$ bootstrap statistics.

\(^{13}\)Two models are nested if one of them is the special case of the other. If neither of them can be expressed as the special case of the other model, these two models are non-nested. In this chapter, I compared the conditional forecast model (larger model) with the unconditional benchmark model (parsimonious model which is the special case of larger ones), so the pairwise comparing model are nested in the forecast evaluation exercise.

\(^{14}\)This situation is analogous to point forecast evaluation. Clark and West (2007) found that the test statistics proposed by Diebold and Mariano (1995) have a nonstandard distribution as a function of stochastic integrals of Brownian Motion.
Economically Motivated Model Restriction

Recall the bivariate predictive quantile regression model is given by

\[
\hat{Y}_{t+1}(\tau) = \hat{\alpha}_i(\tau) + \hat{\beta}_i(\tau)X_{i,t} \quad (2.6.1.4)
\]

where \(\hat{\alpha}_i(\tau)\) and \(\hat{\beta}_i(\tau)\) are out-of-sample quantile estimates of the true parameters \(\alpha_i(\tau)\) and \(\beta_i(\tau)\) respectively, based on data recursively withdrawn from the start of sample through time \(t\). Since out-of-sample forecast is based only on the information up to the time of the sample available, the parameter estimates out-of-sample are considered less efficient than the in-sample counterpart. As discussed by Campbell and Thompson (2008), and Rapach and Zhou (2012b), imposing sign restriction on both coefficient estimated \(\hat{\beta}\) and the stock return forecast \(\hat{Y}\) can significantly improve out-of-sample forecast performance. In the context of quantile forecasts, theory suggests that stock returns in the lower (upper) tail of distribution are often negative (positive). Analogously, I imposed the sign restriction on quantile forecast

\[
\begin{align*}
\hat{Y}_{t+1}(\tau) &= 0 & \text{if } \hat{Y}_{t+1}(\tau) > 0 \text{ and } \tau < 0.5 & \text{(Lower tail Restriction)} \\
\hat{Y}_{t+1}(\tau) &= 0 & \text{if } \hat{Y}_{t+1}(\tau) < 0 \text{ and } \tau > 0.5 & \text{(Upper tail Restriction)}
\end{align*}
\]

(2.6.1.5)

Meanwhile, the sign of out of sample estimated coefficient \(\hat{\beta}_i(\tau)\) are set to be consistent with the sign of in-sample counterpart estimates \(\hat{\beta}_{In-sample}(\tau)\).

\[
\begin{align*}
\hat{Y}_{t+1}(\tau) &= \bar{q}_{\tau,t} & \text{if the sign of } \hat{\beta}_i(\tau) \text{ and } \hat{\beta}_{In-sample}(\tau) \text{ are in-consistent} \\
\hat{Y}_{t+1}(\tau) &= \hat{q}_{\tau,t} & \text{if the sign of } \hat{\beta}_i(\tau) \text{ and } \hat{\beta}_{In-sample}(\tau) \text{ are consitent}
\end{align*}
\]

(2.6.1.6)

Imposing such sign restrictions turn out to reduce parameter estimation uncertainty and helps to stabilize predictive regressions.
Conditional Mean Forecast Evaluation

Campbell and Thompson (2008) suggested the out-of-sample $R^2$ to evaluate the forecast performance of conditional mean estimation, which attempts to measure the proportional reduction in MSFE for the predictive regression forecast $\hat{y}_t$ relative to the historical average $\bar{y}_t$

$$R^2 = 1 - \frac{\sum_{t=P}^{T}(y_t - \hat{y}_t)^2}{\sum_{t=P}^{T}(y_t - \bar{y}_t)}$$  \hspace{1cm} (2.6.1.7)

where $P$ is the start date of forecast, $\hat{y}_t$ is the proposed forecast model for the conditional mean, and $\bar{y}_t$ is the unconditional historical average. Undoubtedly, the significant positive value of $R^2$ indicates the superiority of the conditional forecast model $\hat{y}_t$ relative to $\bar{y}_t$.

Diebold and Mariano (1995), West (1996) (hereafter DMW) suggested a test statistic to compare forecast performance from non-nested models. Let $y_{A,t}$ and $y_{B,t}$ denote the forecast for $y_t$ obtained from models $A$ and $B$ respectively. The average forecast comparison $\bar{f}$ can be calculated as

$$\bar{f} = \frac{1}{T-P} \sum_{t=P}^{T} (y_t - y_{A,t})^2 - (y_t - y_{B,t})^2$$  \hspace{1cm} (2.6.1.8)

DMW shows that $\bar{f}$ follows a standard normal distribution asymptotically when the comparing models are non-nested

$$\frac{\bar{f}}{\text{var}(\bar{f})} \sim N(0, 1)$$  \hspace{1cm} (2.6.1.9)

However, when we compare the prediction from nested models (say, a parsimonious model $A$ and a larger model $B$), the average forecast comparison $\bar{f}$ is no longer a standard normal distribution but a Brownian motion. The estimation of additional variables in the larger model $B$ introduces noise into forecast process that will, in the finite sample, inflate its MSFE. Even in the null hypothesis of equal MSFE, the
larger model $B$ could generate greater MSFE than the parsimonious model $A$. Clark and West (2007) adjusted DMW test to produce a modified MSFE-adjusted statistic, which is very straightforward to compute by first defining the adjusted forecast comparison $\tilde{f}_t$

$$\tilde{f}_t = (y_t - y_{A,t})^2 - \left[ (y_t - y_{B,t})^2 - (y_{A,t} - y_{B,t})^2 \right]$$ (2.6.1.10)

The test statistic of Clark and West, denoted as MSFE - adjusted, corresponds to the standard t-statistic from the regression of $\tilde{f}_t$ on a constant.

Density Forecast Evaluation

As discussed in the previous sections, the mean or the downside risk of stock return is hard to predict, but the other part of distribution may be more easily predictable. To compare the whole forecast performance of any two models instead of focusing only on the mean, I used the likelihood ratio test statistics developed by Amisano and Giacomini (2007) to analyze the predictability performance of individual density relative to the benchmark model based on the historical unconditional model. The test statistics takes the form

$$AG = \frac{\Delta LR}{\hat{\sigma}/\sqrt{T-P}}$$ (2.6.1.11)

where

$$\Delta LR = \frac{1}{T-P} \sum_{t=P+1}^{T} \Delta LR_t = \frac{1}{T-P} \sum_{t=P+1}^{T} \log f(y_{t+1}|x_t) - \log g(y_{t+1}|x_t)$$ (2.6.1.12)

where $P$ is the start date for out-of-sample forecast evaluation, $\hat{\sigma}^2$ is a heteroskedasticity and autocorrelation consistent (HAC) estimator of asymptotic variance of the
\[ \Delta LR_t \]. \( f(y_{t+1}|x_t) \) and \( g(y_{t+1}|x_t) \) are the predictive densities of the candidate model and the benchmark model, respectively.

If the two competing models are non-nested, the test statistics \( AG \) asymptotically follow a normal distribution. The null hypothesis of equal performance of forecasts \( f \) and \( g \) can be rejected at level \( \alpha \) if \( |AG| > z_{\alpha/2} \), where \( z_{\alpha/2} \) is the \( (1 - \alpha/2) \) quantile of a standard normal distribution. One can choose \( f \) if \( AG > 0 \), or \( g \) if \( AG < 0 \).

### 2.6.2 Economic Evaluation of Return Predictability

The previous sections discussed how to statistically evaluate the out-of-sample performance of two competing forecast models and determine which one wins the horse race in forecast competition. This section attempts to answer the question if such statistical evidence of predictability bears any economic significance to risk averse investors who make portfolio choices strictly based on return forecasts. I perform an out-of-sample portfolio study based on an investor with Mean Variance or CRRA utility preference and show that aggregating distributional information along with model information can significantly add economic value in portfolio management. Firstly, I will explain how the forecasts of conditional mean and density estimation are employed in the portfolio selection procedures. The following portfolio choice exercise is closely in line with the studies from Cenensizoglu and Timmermann (2012).

**Optimal Portfolio Choice**

Consider a risk averse investor who constructs a dynamically balanced portfolio and has access to only one risky asset (equity) and a risk free asset (i.e. Bond). At time \( t \), the investor has to allocate \( \omega_t W_t \) wealth to the risky asset and \( (1 - \omega_t)W_t \) to the risk free asset. The budget constraint for the wealth of the investor at time \( t + 1 \)
is set as

\[
W_{t+1} = W_t \left( \omega_t R_{t+1} + (1 - \omega_t) r^f_{t+1} \right)
\]

\[
= W_t \left( r^f_{t+1} + \omega_t r_{t+1} \right)
\]

(2.6.2.1)

where $R_{t+1}$ and $r_{t+1}$ are the gross return and excess return of risky asset respectively.

Assuming the utility function at $t + 1$ is $U(W_{t+1})$, the portfolio choice question is the solution to the following optimization problem

\[
\omega^* = \arg\max_{\omega_t} E_t \left( U(W_{t+1}) \right)
\]

(2.6.2.2)

where $E_t[.]$ denotes the conditional expectation on the investor’s information at time $t$.

I consider two commonly used utility functions for investors’ preference to evaluate the forecast for the conditional mean as well as the entire density estimation. In the mean-variance framework, utility function is determined only by the first two moments of the distribution:

\[
U(W_{t+1}) = E(W_{t+1}) - \frac{\gamma}{2} Var(W_{t+1})
\]

(2.6.2.3)

where $\gamma$ is the absolute risk aversion coefficient, reflecting investor’s appetite for risk under mean variance utility preference. $E_t(.)$ and $Var_t(.)$ denote the conditional mean and variance. The optimal portfolio holdings, conditional on the predicted mean $\hat{\mu}_{t+1}$ and variance $\hat{\sigma}^2_{t+1}$, can be solved as (Campbell and Viceira (2002), Campbell and Thompson (2008) and Rapach and Zhou (2012b))

\[
\omega^* = \frac{1}{\gamma} \frac{E_t(r_{t+1})}{Var_t(r_{t+1})}
\]

(2.6.2.4)
As the forecast exercise in this paper is implemented exclusively for the continuously compounded rate\textsuperscript{15}, I reset the budget constraint for the wealth of an investor in the next period $t + 1$ as \textsuperscript{16}

$$W_{t+1} = (1 - \omega_t)\exp(r^f_{t+1}) + \omega_t\exp(r^f_{t+1} + r_{t+1})$$

(2.6.2.5)

where $r_{t+1}$ is the stock return in excess of the risk free rate, $r^f_{t+1}$, both in continuously compounded rates. Now the optimal portfolio holding under Mean Variance preference can be obtained instead as \textsuperscript{17}

$$\hat{\omega}^*_{t+1} = \frac{1}{\gamma \exp(r^f_{t+1}) \exp(\hat{\sigma}^2_{t+1}) \exp(\hat{\sigma}^2_{t+1} - 1) \exp(2\hat{\mu}_{t+1} + \hat{\sigma}^2_{t+1})}$$

(2.6.2.6)

where $\hat{\mu}_{t+1}$ and $\hat{\sigma}_{t+1}$ now are the forecast of conditional mean and volatility respectively for the continuously compounded rate. Following Campbell and Thompson (2008), I used a 5 year rolling window \textsuperscript{18} of historical return to estimate the sample variance $\hat{\sigma}^2_{t+1}$.

The second utility function I consider is relative risk aversion preference (CRRA)

$$U(W_{t+1}) = \frac{W_{t+1}^{1-\gamma}}{1 - \gamma}$$

(2.6.2.7)

where $\gamma$ is the relative risk aversion under CRRA utility preference.

Assume that the conditional distribution of excess stock return is log normal with predicted mean and variance as $\hat{\mu}_{t+1}$ and $\hat{\sigma}_{t+1}$, respectively. Campbell and Viceira

\textsuperscript{15}Note that the simple gross return $R_t = \frac{P_t + D_t - P_{t-1}}{P_{t-1}}$ can be connected with the continuous compounded return $r_t = \log(\frac{P_t + D_t}{P_{t-1}})$ by the equation $r_t = \log(1 + R_t)$.

\textsuperscript{16}I set the initial wealth $W_t = 1$

\textsuperscript{17}The optimal portfolio weight is obtained by maximizing the expected utility in equation (6.15) in term of weight $\omega_t$, but subject to the new budget constraint in equation (6.17)

\textsuperscript{18}To account for time varying property of volatility, sample variance can be estimated by GARCH type dynamic model. Andersen et al. (2006) provided an extensive survey for return volatility forecast. As matter of fact, GARCH(1,1) forecasts of future volatility fail to make noticeable difference. For the sake of brevity, these results are not reported but will be available upon request.
(2002) derived the log linearized approximation to the investor’s wealth and obtained the optimal portfolio weight approximately as

\[
\omega^* \approx \frac{1}{\gamma} \left( \hat{\mu}_{t+1} + \frac{\hat{\sigma}_{t+1}^2}{2} \right)
\]  

(2.6.2.8)

Therefore, the analysis of optimal portfolio choice only requires the knowledge of the first two moments of the return distribution, which is consistent with mean-variance analysis. However, without the restrictive assumption of log normal distribution, the above approximation of optimal portfolio weight usually does not hold. More generally, the expected utility maximization problem under CRRA utility preferences requires a forecast of the entire conditional distribution of future returns, rather than just the conditional first two moments including mean and volatility. As discussed before, different macro-finance variables \(X_{i,t-1}\) can deliver a different forecast of the entire density of future return \(f(r_{t+1}|X_{i,t})\). Combining multiple density forecasts can produce more precise forecast of the conditional distribution of future stock returns. The optimal portfolio weight \(\omega^*\) under CRRA utility preference, thus, can be obtained by solving the following optimization problem

\[
\omega^* = \arg\max_{\omega_t} \int \left( (1 - \omega_t)\exp(r_{t+1}^f) + \omega_t\exp(r_{t+1}^f + r_{t+1}) \right)^{1-\gamma} \frac{1}{1 - \gamma} f(r_{t+1}|\mathcal{F}_t) 
\]  

(2.6.2.9)

where \(r_{t+1}\) is the continuously compounded excess return. The density of \(r_{t+1}\) can be estimated by the previously discussed kernel estimation which combine a wide range of quantile information. Different density forecasts \(f(r_{t+1}|\mathcal{F}_t)\) lead to different optimal portfolio choices \(\omega^*\) over out-of-sample evaluation period. Intuitively, more precise estimates of the density forecast \(f(r_{t+1}|\mathcal{F}_t)\) should add economic value in the portfolio management.

Following Kandel and Stambaugh (1996), the optimal weight \(\omega_{t+1}^*\) under CRRA
preference was restricted in the range $\omega \in (0, 1)$ to ensure that the utility is bounded in the CRRA preference. However, under the framework of mean variance utility, the optimal weight $\omega^*_t$ was relaxed to the constraint $\omega \in (0, 1.5)$, which is in line with Campbell and Thompson (2008).

Therefore, different forecasts of the future conditional mean $\hat{\mu}_{t+1}$ and volatility $\hat{\sigma}_{t+1}$, or the entire distribution of stock returns, yield different optimal portfolio holding $\omega^*_t$, which in turn give rise to a different realized utility level in the next period.

$$U(W_{t+1}(\omega^*_t)) = U\left(W_{t}[\{1 - \omega^*_t\} \exp(r^f_t) + \omega^*_t \exp(r^f_t + r_{t+1})]\right)$$  \hspace{1cm} (2.6.2.10)

The average realized utility level \textsuperscript{19} over the out-of-sample forecast evaluation period can be constructed as

$$\overline{U}(\omega^*) = \frac{1}{T - P} \sum_{t=T-P}^T U(W_{t+1}(\omega^*_t))$$  \hspace{1cm} (2.6.2.11)

where $P$ is the start date of the forecast evaluation period.

Analogously, suppose the investors instead rely on the unconditional historical benchmark estimation $\bar{\mu}$ to predict future returns and construct the dynamic balanced optimal portfolio holding $\omega^0_t$. The average realized utility based on this “naive” model can be instead derived as

$$\overline{U}(\omega^0) = \frac{1}{T - P} \sum_{t=T-P}^T U(W_{t+1}(\omega^0_t))$$  \hspace{1cm} (2.6.2.12)

\textsuperscript{19}I set the initial wealth $W_0 = 1$.  

118
Therefore, the economic gain benefits from the conditional predictive model relative to historical average benchmark can be given as

$$\Delta U = 400 \times (U(\omega^*) - U(\omega^o))$$

(2.6.2.13)

The average utility gain $\Delta U(\%)$ can also be considered as the portfolio management fee that an investor would be willing to pay to have access to the conditional forecast model relative to the unconditional benchmark forecast. A positive value of $\Delta U(\%)$ indicates that the conditional model produce superior forecasts in the utility based metrics than the unconditional model.

Alternatively, the economic significance of the predictive model can also be evaluated by the certainty equivalent rate of return ($CER$), which is the equivalent risk free compensation that provides the same utility level as the risky portfolio. Under the mean variance preference, the $CER$

$$CER = \frac{1}{T-P} \sum_{t=T-P}^{T} U\left(W_{t+1}(\omega^*_{t+1})\right)$$

(2.6.2.14)

where $P$ is the start date for forecast evaluation.

In contrast, under CRRA utility preference, the $CER$ can be computed as

$$CER = \left\{(1-\gamma) \frac{1}{T-P} \sum_{t=T-P}^{T} U\left(W_{t+1}(\omega^*_{t+1})\right)\right\}^{1/(1-\gamma)} - 1$$

(2.6.2.15)

where $\frac{1}{T-P} \sum_{t=T-P}^{T} U\left(W_{t+1}(\omega^*_{t+1})\right)$ is the average realized utility level over the forecast evaluation period, which is obtained by deriving the optimal portfolio weight $\omega^*_{t+1}$ suggested by a particular predictive model.

\[20\] I multiply the following equation by 400 to express it to be average annualized return (basis points, bps)
2.7 Empirical Results

2.7.1 Data

The data employed in this paper are closely in line with Goyal and Welch (2008), Cenensizoglu and Timmermann (2008) and Rapach, Strauss and Zhou (2010). The equity premium is measured by the difference between the log return on the S&P 500 index returns (including dividends) and the log return on the treasury bill rate. The forecasting exercises in this paper are conducted on the quarterly frequency data, which span the sample period from 1926Q4 to 2010Q4.

In this paper, I adopt the recursive method to obtain the out-of-sample forecasts. The total number of observations $T$ is divided into the in-sample portion of $R$ observations, used to construct the first forecast and out-of-sample portion of $T - R$ observations, used for forecasting. Under the recursive scheme, the parameters estimates are updated as the forecast moves forward by adding one additional observation to the estimation sample at each step.

At each time $t = R + 1, R + 2, \cdots, T$, the parameters are estimated by utilizing the information only from $1, \cdots, t - 1$. In the following discussion, some of the combination methods are constructed based on the forecasts in the holdout out-of-sample period. Therefore, the total sample is divided into an in-sample portion of $R$ observation, $P_0$ holdout out-of-sample forecasts and $P$ out-of-sample forecasts.

More specifically, the first out-of-sample forecast is calculated for 1947Q1, which is based on the in-sample data from 1926Q4 to 1946Q4 (80 quarters). Following the recursive (expanding) scheme, such forecast experiments are repeated continuously once an additional new observation is added. Meanwhile, the initial ten years’ (40 quarters from 1947Q4 – 1956Q4) forecasts are used as the holdout out-of-sample period before the start of forecast evaluation\textsuperscript{21}. The out-of-sample forecast evaluation

\textsuperscript{21}some forecasts (i.e. DMSFE) are constructed based upon the historical forecast performance in the initial holdout out-of-sample period
starts from 1957Q4 – 2010Q4, which covers the oil price shocks of 1970s; the long
economics booms of 1960s, 1980s and 1990s; and the recent global financial crisis.
This choice of forecast evaluation period is somewhat arbitrary and exactly in line
developed the out-of-sample test of forecast ability that is robust to in-sample/out-
of-sample split.

Alternative to the recursive (expanding) window estimation, rolling window esti-
mation schemes seems to be more appealing for the data with structural breaks or time
varying dynamic evolution. However, precise estimation of optimal rolling window
is considered as a big challenge. Therefore, it is typically a bias-efficiency trade-off.
Shortening the window length, in one hand, increase the variation of forecast, but
on the other hand, introduces estimation bias and reduces estimation precision. Pe-
saran and Timmermann (2007) and Clark and McCracken (2009) demonstrated that
recursive window estimation frequently beats rolling-window estimation in term of
MSFE.

Many studies have used a similar set of predictor variables to examine stock
return predictability. Following Cenensizoglu and Timmermann (2008) and Rapach,
Strauss and Zhou (2010), the predictor regressors in this paper consists of 14 eco-
nomic variables, which can be classified into the following four categories.

- Valuation ratios capturing some measures of ‘fundamental’ value to market val-
  ue

1. Dividend-price ratio (DP): log of a 12-month moving sum of dividends
   paid on the S&P500 index minus the log of stock price (S&P500 index)

---

22Undoubtedly the choice of in-sample/out-of-sample split may affect forecast evaluation. I use
the same forecast evaluation as Rapach, Strauss and Zhou (2010) to manifest the improvement of
forecast performance for our proposed methodology.
2. **Dividend yield** (*DY*): log of a 12-month moving sum of dividends paid on the S&P500 index minus the log of lag stock price


4. **Book-to-market ratio** (*BM*): Book to market value ratio for DowJones Index

- Bond yield measures capturing the level or slope of the term structure or measures of default risk

5. **Three Month T-bill rate** (*TBL*): interest rate for three months Treasury bill (secondary market)

6. **Long term yield** (*LTY*): Long term government bond yield

7. **Term spread** (*TMS*) long term yield minus the Treasury bill rate

8. **Default Yield Spread** (*DFY*): Difference between BAA and AAA rated corporate bond yields

9. **Default Return Spread** (*DFR*): Long term corporate bond return minus the long term government bond return

- Estimates of Equity Risk
10. **Long term return** (*LTR*): Return on Long term government bonds

11. **Stock Variance** (*SVAR*): Sum of squared daily returns on the *S&P500* index

- Corporate finance variables

12. **Dividend Payout Ratio** (*DE*): log of a 12 month moving sum of dividends minus the log of a 12 month moving sum of earnings

13. **Net Equity Expansion** (*NTIS*): the ratio of 12 month net issues by NYSE-listed stocks over their end-year market capitalization

Finally, I also consider the inflation rate as a macroeconomics variable

14. **Inflation** (*INFL*): the rate of change in the consumer price index (all urban consumers). We use one time lag $X_{t-1}$ for inflation to account for the delay in CPI release

Additional details on the data sources and the construction of these variables are provided by Rapach, Strauss and Zhou (2010)$^{23}$.

$^{23}$The data are available at http://sites.slu.edu/rapachde/home/research
2.7.2 Empirical Results for Conditional Mean Forecast

In-Sample Stock Return Prediction

Table 2.3 reports the slope coefficient estimates for the in-Sample stock return predictions in the univariate quantile and OLS regression model, which cover the sample period from 1946Q4 to 2010Q4. Panel A shows that, even for the in-sample forecast exercise, many of the predictor variables suggested in the literature actually fail to predict stock return successfully. Among fourteen predictor variables, only four regressors (DP, DY, TBL and DFR) have significant slope estimates. The goodness fit measures of $R^2$ are pretty low without exception, which casts doubt on the reliability of the suggested predictive variables in forecasting stock return. Turning to the estimation results based on quantile regression. Panel B displays the slope estimates for the univariate quantile forecast. Two interesting features deserve mention. First, no predictive variables can uniformly predict stock return in every part of the distribution. For instance, the Treasury Bill Rate (TML) seems to be more successful in predicting the lower shoulder of the return distribution, whereas stock variance $SVAR$ is more predictable in the upper tail of the stock return distribution. In general, the extreme tail of return distribution tends to be more predictable than the central part of distribution. Most of the slope estimators are statistically significant from zero either in the lower or the upper tail of the distribution. Secondly, the magnitude of the slope coefficient estimates in extreme (lower or upper) tails tends to be much larger than that in the central part of distribution, which is consistent with the previous discussion in Section 2 that the magnitude of slope coefficients tends to increase when quantiles converge to the extreme tail. This observation is especially prominent in the forecast with Stock return variance $SVAR$ as regressor. An increase of stock variance $SVAR$ leads to an increase in the upper quantile of return and a decrease in the lower quantile of return. Furthermore, the median slope
coefficient ($\tau = 0.5$) is of similar magnitude as the least squares slope coefficient, indicating rough symmetry in the returns distribution. Panel C of Table 2.3 reports the wald test for slope equality across multiple quantiles. Among the fourteen quantile regression models, only three slope coefficient estimates show statistical significantly, indicating that most of predictor variables fail to capture the variation of stock returns distribution beyond the first moment.

Since the overall number of regressors proposed for stock return prediction is quite large, including all predictive variables in the forecasting model might result in inconsistent estimators. Table 2.4 reports the post-LASSO estimation of slope coefficients in the multivariate forecast model. The LASSO selection for OLS estimation in Panel B follows the procedures proposed by Belloni, Chernozhukov and Wang (2012), which suggested a pivotal method for estimating high-dimensional sparse linear regression models that are valid for both Gaussian and non-Gaussian distribution. After de-selecting regressors via LASSO procedures, the goodness of fit of the overall forecast model improves to 5.13%. Some predictive variables (e.g. $DP$ and $DY$) which originally had a significant marginal effect are eliminated by the LASSO selection method. Panel C shows the quantile LASSO selection following Belloni and Chernozhukov (2011). Surprisingly, the post-LASSO step produces the pretty identical forecast model for different quantiles of the returns distribution. The Dividend Price Ratio $DP$ and Default Yield Spread $DFY$ are most frequently selected in forecasting almost all parts of the distribution. Except for the extreme lower and upper quantile ($\tau = 0.05$ and $\tau = 0.95$), the Pseudo R2 for post-LASSO quantile regression are superior to those of ordinary quantile regression\textsuperscript{24}.

\textsuperscript{24}To save space, I hold the report for Pseudo $R^2$ for univariate quantile regression and post-LASSO regression model. The results will be available upon request.
Out of sample performance of predictive model

Table 2.5 reports the out-of-sample predictive $R^2(\%)$ for the forecast of the conditional mean of the equity premium during the sample period from 1957 Q1 - 2010 Q4. Panel A succinctly conveys the message exactly in line with Goyal and Welch (2008) and Rapach and Zhou (2012b): a univariate regression model based on individual predictive variable, more often than not, fails to outperform the “naive” simple model (unconditional historical average benchmark). Among 14 univariate predictive regression models (over the whole sample period), only one forecast based on predictive variable $DY$ significantly beats the historical average model. Most other forecast models, however, are inferior to (higher MSFE) the historical average benchmark.

Panel B presents predictive performance for the forecast models based on model selection. The Sink-Model includes all predictive variables in the regression without any selection procedures. As well known, the linear regression in the high-dimension sparse models is inconsistent, the out-of-sample performance for the Sink Model, therefore, can be expected to be extremely poor, as displayed in the first row of panel B. The model selection based on the comparison of information criteria AIC are presented in the second row of panel B. The forecast performance is still disappointing even though it improves a little bit compared with the sink model. Finally, LASSO model selection turns out to provide a successful forecast for the future equity premium, and the out of sample R2 are statistically significant positive, indicating the regressors after LASSO selection provides superior forecasts to the benchmark model.

Panel C reports the forecast performance from the model combination of different individual point forecast $\hat{Y}_{t+1}$. Four simple model combination approaches

\footnote{Up to three regressors are selected to calculate AIC and choose the model with highest value of AIC from all possible $2^{14}$ combinations.}

\footnote{If no regressors survive after LASSO selection, the post-LASSO forecast is exactly the historical benchmark model.}
are employed attempting to improve equity premium forecast by pooling information from different models. The results in Panel C demonstrate the extreme success of model combination in equity premium forecast. The out-of-sample predictive $R^2(\%)$ becomes positive at least at 10% significance level, indicating that pooling information across different models can successfully improve the equity premium forecast and beat the historical average benchmark.

Table 2.5 also conveys a interesting message that return predictability is largely a recessionary phenomenon. The values of out-of-sample $R^2(\%)$ during the recession are much higher than those during expansion or overall period. This result is consistent with the empirical finding from Henkel et al. (2011) who suggested that the out-of-sample forecast gains are highly concentrated during recessions. When a financial crisis is prevalent, all factors tend to move together. The dependence of response and factor are usually greater (correlation is higher) and is more easily to be captured than that in times of economics prosperity. This explains why stock returns are more predictable during recession period.

Now I turn to discuss what role the distributional information plays in the forecast of conditional mean of the equity premium. Table 2.6 presents the Out-of-Sample predictive $R^2(\%)$ for the conditional univariate Quantile forecast based on each individual predictor variable (Panel A), model combination for every specific quantile (Panel B), conditional mean $\hat{Y}$ forecast developed by Model Combination at each quantile joined with quantile combination afterwards (Panel C), and model selection (Quantile LASSO) in the first step followed by quantile combination at second step (Panel D). The value of $R^2(\%)$ in Panel A of table convey an interesting message: the left tail of the distribution seems to be much more difficult to predict than the right tail. The Out-of-Sample predictive $R^2(\%)$ are nearly all negative in the lower quantile of the return distribution from $\tau = 0.1$ to $\tau = 0.4$. Most of the predic-
tive variables listed in the first column of panel A fail to provide a superior forecast of the downside risk than the benchmark unconditional quantile forecast. In contrast, the upper shoulder of quantiles between $\tau = 0.6$ and $\tau = 0.8$ are more predictable, as demonstrated by the highlighted values in Panel A. This pattern of predictability becomes more prominent after model combination over each specific quantile $\tau$. As panel B of the table show, the upper tail of the distribution appears to be significantly predictable. Almost all value of the out of sample predictive $R^2(\%)$ become positive and statistically significant in the upper tail of distribution from quantile $\tau = 0.6$ to $\tau = 0.9$. Meanwhile, with respect to the downside risk forecast, principal component analysis PCA seems to provide better forecasts than other model combinations approaches because all $R^2(\%)$ obtained from PCA are positive in the lower tail of the distribution, even though some of them are still not statistically significant. In summary, combining different models in a particular quantile can significantly improve forecast performance, which is consistent with the empirical results on the conditional mean forecasts.

After pooling information from multiple models for each quantile forecast $\hat{Y}(\tau)$, I attempt to aggregate distributional information and examine whether the forecast performance for the conditional mean of equity premium can be further improved via combining distributional information from multiple quantile estimators. Corresponding to model combination on each quantile forecast, only the principal component (PCA) approach provides reasonable forecast in both tails of distribution 27(as the last row of panel B show). Therefore, when aggregating distributional information, I just consider the principal component analysis (PCA) for the model combination of quantile forecasts(the last row of panel B). Panel C presents the forecast of conditional mean $\hat{Y}$ of equity premium based on model combination for each quantile estimation in the first step joined with quantile combination in the second step. As

---

27 Other model combination approaches provide reasonable forecast only in the right tail, but not in the left tail, as shown in the panel B of table 4.
shown, by taking advantage of information both across the model (model combination) and within the distribution (quantile combination), most of forecasts for the conditional mean $\hat{Y}_t$ outperform the unconditional historical average $\bar{Y}_t$. Again, the forecast performance during recessions improved in the greatest degree.

However, a careful comparison of Table 2.5 and Table 2.6 reveals that, the improvement of forecast performance after quantile combination is very slight and even trivial. Model combination followed by quantile combination afterwards does not produce obvious superior forecasts to that derived only from model combination. This fact is not surprising as the information across different models is much easier to pool than the information within distribution\textsuperscript{28}. Directly combining multiple forecasts of the conditional mean in the second step instead of in the first step would make it easier to utilize information across different models when predicting the mean of equity premium.

Lastly, Panel D of Table 2.6 presents the performance of forecasts derived from model selection via quantile LASSO in the first step along with quantile combination in the second step. Compared with the forecast obtained only from quantile lasso model selection in table 5 ($R^2 = 2.04\%$), the improvement of forecast performance after quantile combination is very impressive and prominent which can even reach a higher value $R^2 = 3.06\%$. This result seems to provide a strong evidence that distributional information aggregated via multiple quantiles combination can result in a superior conditional mean forecast relative to the one that only utilizes the source of information alone in model selection.

In the previous discussion, a combination strategy is implemented in the order of either model combination in the first step followed by quantile combination

\textsuperscript{28}Return forecast literatures demonstrated that even a simple equal weight model combination can surprisingly achieve a extreme success in pooling model information. However, the information across different part of distribution is much difficult to aggregate.
in the second step.

\[ \hat{Y}_{t+1} = \frac{1}{I} \sum_{i=1}^{I} \varphi_i(\tau_i) \left( \frac{1}{J} \sum_{j=1}^{J} \omega_j \hat{Y}_{t+1}^j(\tau_i) \right) \]  \hspace{1cm} (2.7.2.1) \]

where \( i \) denotes the \( i \)th quantile at \( \tau_i \), \( j \) represents the \( j \)th model. \( \varphi_i(\tau_i) \) is the combination weight for a specific quantile at \( \tau_i \), and \( \omega_j \) is the combination weight for the \( j \)th model.

Alternatively, we can switch this order of combination. Firstly, multiple quantile estimators in each individual model \( j \) are combined in a weighted average scheme to utilize distribution information within the model and produce the point forecast \( \hat{Y}_{t+1}^j = \frac{1}{I} \sum_{i=1}^{I} \varphi_i(\tau_i) \hat{Y}_{t+1}^j(\tau_i) \) for the \( j \)th model. After that, the point forecast \( \hat{Y}_{t+1} \) can be obtained eventually by combining different models afterward. Generally,

\[ \hat{Y}_{t+1} = \frac{1}{J} \sum_{j=1}^{J} \omega_j \left( \frac{1}{I} \sum_{i=1}^{I} \varphi_i(\tau_i) \hat{Y}_{t+1}^j(\tau_i) \right) \]  \hspace{1cm} (2.7.2.2) \]

Now I change the order of the combination strategy and turn to examine if the distributional information within the model can be utilized in the first step and help further improve forecast performance in the second step. Table 2.7 presents the out-of-sample predictive \( R^2(\%) \) horse race comparing the forecasts derived from quantile combination firstly along with model combination afterwards. Panel A presents forecast evaluation for each univariate predictive model which combines multiple quantile estimators. Four quantile combination approaches are employed to aggregate distribution information within the model: (1) optimal weight quantile combination (OWQ) which was developed by Xiao and Zhao (2011); (2) Density weighted quantile combination (Density) which is based on the asymptotical distribution of quantile estimator for coefficient; (3) Regression approach combination which is in the same spirit of Granger and Ramanathan (1984); (4) combination quantile forecast that depend on recent historical forecast performance (DFMSE). All of these approaches have been
discussed in Section 3.

The last four rows of panel A show that the improvement (compared with \textit{OLS} estimation in table 5) of forecast performance after quantile combination is obvious. More specifically, after combining multiple quantile estimations, the out of sample Predictive $R^2(\%)$ for the individual model based on $DP$ and $DY$ all become positive and statistically significant at least 10% level. Meanwhile, the forecast performance of almost all other individual models improved in some degree compared with the OLS estimation, even though most of which are still inferior to the historical average benchmark ($R^2(\%)$ is negative, but closer to zero than OLS forecast). Among these four quantile combination approaches, it is not difficult to find that the first two quantile combination approaches (\textit{OWQ} and \textbf{Density} weight) perform somewhat better than the last two quantile combination methods (\textit{Regression} and \textbf{DFMSE}). It appears that the former two combination approaches are more successful in collecting distributional information from the predictive regression models than the last two methods. Therefore, in the following discussion, I only concentrate on the first two approaches when implementing both model and quantile combination.

Finally, Panel B of Table 2.7 presents how the forecast performance for the conditional mean $\hat{Y}$ of the equity premium can be further improved by utilizing both distributional information (quantile combination) and model information (model combination). The comparison of Table 2.7 with Table 2.5 provides insights into the extreme success of distributional information combination for stock return prediction. Relative to the only model combination in Table 2.5, combining quantile estimation followed by model combination can generate significantly higher $R^2_{\text{OOS}}$ without exception for the conditional mean forecast of the equity premium, which indicates the positive contribution of distributional information in the conditional mean forecast. It is noteworthy that the order of the combination strategy does make a difference in collecting both distributional and model information efficiently. By comparison with
Table 6, the combination strategy in Table 2.7 is much more efficient.

Economic Measures of Forecasting Performance Under Mean Variance Utility

This section is dedicated to analyze stock return forecast with profit or utility-based metrics, which attempts to investigate whether the statistical evidence of stock return predictability can be converted directly to a significant economic gain for risk-averse investors. If the different statistical and economic measures are not consistent, how closely related are these two measures of forecast performance?

Table 2.8 presents the economic evaluation for the conditional mean forecasts derived from different estimation strategies under Mean Variance utility preference with absolute risk aversion coefficient $\gamma = 5$. Average utility gain $\Delta U(\%)$ can be considered as the portfolio management fee (in annualized percent return) that an investor would be willing to pay to have access to the conditional return forecast relative to the historical average benchmark forecast. $CER$ refers to certainty equivalent rate of return that provides the same risk free utility level as the risky portfolio which is constructed based on return forecast. A higher value of $CER$ indicates the better return forecast in terms of utility gain metrics. Panel B of the table reports the utility gain for an investor under mean variance preference who makes portfolio choices following the guidance of the return forecasts that are derived only from model combination estimation strategy. Panel C presents the economic gain for the investor under the same preferences who uses the return forecasts obtained from the estimation strategy of quantile Lasso model selection in first step followed by quantile combination in the second step. Finally, Panel D reports the utility gain for the in-

\footnote{Cenensizoglu and Timmermann (2012) found that the message conveyed from a model’s RMSE performance can be quite different from that emerging from an analysis of the model’s economic value.}
vestor who uses the return forecasts based on quantile combination followed by model combination. The comparison of Panel B and Panel D provides insight that quantile combination does add additional significant economic value if an investor takes advantage of distributional information to forecast returns and makes portfolio choice accordingly. Higher values of average utility gain ($\Delta U(\%)$) and $CER$ are obtained without exception when an investor has access to return forecasts which assimilate both distributional information and model information.

Figure 2.3 compares the economic gain for the forecasts that are derived from model selection followed by quantile combination (LASSO+Quantile Combination) with the forecasts that are obtained only from model selection (LASSO). As investors become more risk-averse ($\gamma > 3$), LASSO model selection combined with the “Regression” approach of quantile combination outperforms the forecast based on only LASSO model selection in term of measures of average utility gain ($\Delta U(\%)$) and certainty equivalent return ($CER$), which provides strong evidence that aggregating distributional information can further improve the performance of forecast that only based on model selection.

Figure 2.4 and 2.5 show the contribution of distributional information when it works together with model combination. Figure 4 displays how the average utility gain (Y axis) under mean variance preference changes with absolute risk aversion coefficient (X axis). The values of average utility gain $\Delta U(\%)$ are derived based on a specific return forecasts which either pools only model information or both model and distributional information. I intend to show that the conclusion arrived in Table 2.8 is robust to the choice of absolute risk coefficient (risk appetite of an investor). As illustrated in Figure 2.4, quantile combination followed with model combination can give rise to higher utility gain than that only based on model combination, no
matter how risk averse of the investor. The relative utility gain appears to attain its largest value when the investor become mildly risk averse ($\gamma = 4$ or $5$), but decays slowly as the risk aversion coefficient $\gamma$ increase. When investors become more risk averse, the accurate forecast of conditional mean becomes less important than the tail of the distribution, which explains why the economic gain relative to benchmark forecast shrink as the risk aversion coefficient $\gamma$ increase. Figure 5 describes almost the same pattern of the relationship between the values of $CER$ derived from a variety of forecasts and the absolute risk aversion coefficient $\gamma$. In a nutshell, pooling distributional information as well as model information can further improve utility gain under mean variance preferences.

Finally, the time series of utility gain ($\Delta U\%$) is displayed in Figure 2.6, which conveys a succinct message that utilizing distributional information via quantile combination contributes to the improvement of out-of-sample performance for forecasts either obtained from model combination or derived from model selection. As time moves on, the utility gain relative to the benchmark model declines. Since all forecasts are estimated recursively, the historical benchmark forecast performs better and better as more observations added in the sample for estimation. This explain why the utility gain relative to benchmark model decays with time.

### 2.7.3 Empirical Results for Density Forecast

**Statistical Measures of Density Forecast Performance**

Table 9 reports the Amisano and Giacomini (2007) (AG) test results for the predictive density comparison of individual models and a wide range of combination models. The benchmark model is the kernel density estimation based on historical unconditional quantile regression. Panel A displays that all of individual density fore-
casts under-perform the benchmark historical estimation without exception (AG test statistics are all negative), which indicates that all macro-financial variables fail to provide superior density forecast than the benchmark model in the term of AG test statistics. This fact is line with the empirical finding that most of macro-financial variables cannot predict the conditional mean of the equity premium (Goyal and Welch (2008)).

Panel B and C of Table 9, however, show the significant improvement of density forecast performance after utilizing both model and distributional information. All values of the AG test statistics become strongly positive and some of them are even statistically significant at the 5% level. More specifically, Panel B reports the AG test results for the estimation strategy of combining models of every quantile estimation in the first step, followed by kernel density estimation employing distributional information in the second step. In contrast, Panel C reverses the combination strategy, which estimates the individual kernel density first and then combines the different models of density forecasts in the second step. Table 9 shows that both the distributional and model information are useful to make forecasts in density estimation.

**Economic Measures of Density Forecast Performance**

Table 2.10 presents the economic evaluation of a wide ranges of density forecasts under relative risk aversion preference (CRRA). Again, the economic significance of density forecast are evaluated by two metrics: (1) Average Utility Gain $\Delta U(\%)$ can be considered as the portfolio management fee that an investor under CRRA preference would be willing to pay to have access to the proposed conditional kernel density forecast relative to unconditional benchmark forecast. (2) $CER$ refers to certainty equivalent rate of return that provides the same utility level as the risky portfolio. As Panel A of Table 10 shows, most conditional density forecasts based on macro-finance
variables fail to provide economic gain relative to historical benchmark kernel density forecasts. Panel B reports the results of economic evaluation for the estimation strategy of model combination in the first step followed by kernel density estimation in the second step. As we can see, combining the information both across different models and multiple quantiles can lead to superior density forecast in term of utility metrics. All values of $\Delta U(\%)$ and $CER$ obtained from the proposed density forecast in Panel B are greater than those derived from benchmark kernel density estimation in which no predictive variables are used to make the forecast. In contrast, Panel C reverses the combination order in which the kernel density estimation is implemented for each macro-finance variable first, followed by combining different density forecasts in the second step. Again density forecasts utilizing both model and distributional information significantly outperform the benchmark kernel density estimation.

Figure 2.7 and 2.8 displays how the average utility gain $\Delta U(\%)$ and $CER$ under $CARRA$ preference changes with relative risk aversion coefficient $\gamma$. I intend to show that the conclusion arrived in table 2.10 is robust to the choice of relative risk aversion coefficient (risk appetite of an investor).

The first two panels of Figure 2.7 show the utility gain based on density forecast which pools the distributional information in the first step, followed by models combination in the second step. The other panels of Figure 2.7 derive the value of $\Delta U(\%)$ based on the density forecast in which the order of combination strategies reverse, namely, model combination in the quantile estimation followed by density kernel estimation. As we can see, pooling distributional information as well as model information can produce better density forecasts relative to the hisotrical benchmark forecast, which results in higher economic value in the measurement of average utility gain $\Delta U(\%)$ and $CER^{30}$. The first combination strategy perform slightly better than $\Delta U(\%)$ is positive in Figure 2.7, indicating a better density forecast relative to the historical benchmark forecast. In Figure 2.8, the value of $CER$ based on the forecast of model combination

\[30\]
the second combination strategy, and the likelihood weight density forecast generates a higher utility gain than the other density forecasts. Furthermore, the average utility gain $\Delta U(\%)$ attains the greatest value when investors are mildly risk averse $\gamma = 3$, but decays slowly as the risk aversion coefficient $\gamma$ increases. This fact is not surprising since the forecast of tail distribution become more and more important as investors become more and more risk averse, but the accuracy of forecasts in the tail of distribution deteriorates dramatically\(^{31}\). Therefore, the relative utility gain decreases as investors become more risk averse since the accuracy of forecast in the tails of the distribution decrease relative to historical benchmark forecast. This can explain why the value of $\Delta U(\%)$ decays gradually as the risk aversion coefficient $\gamma$ increases.

\(^{31}\)The forecast performance in the extreme tail of distribution can be really poor.
2.8 Conclusion

This paper finds that incorporating distributional information via combining multiple quantiles estimates can improve the performance of forecasts that use the single source of information either from model combination or model selection. A Monte Carlo simulation shows that by optimally weighting of multiple quantile estimators, the efficiency gain for coefficients as well as point forecast improves significantly. This paper contributes to the stock return prediction literature by providing empirical evidence that distribution information can provide a positive contribution to both conditional mean and density forecasts, which have not been recognized by many studies. A portfolio study shows that quantile combination does add additional significant economic value if an investor takes advantage of distributional information to forecast return and makes portfolio choice accordingly. Higher value of average utility $\Delta U\%$ and $CER$ are obtained without exception when an investor has access to return forecasts which assimilate both distributional information and model information.
2.9 Appendix

2.9.1 Comparison of Mean Square Error (MSE) of coefficient estimates $\hat{\beta}$ with MSE of point forecast $\hat{Y}$

Consider a linear regression model

$$Y_{t+1} = \alpha + \beta X_t + \epsilon_{t+1} \quad \epsilon_{t+1} \sim \mathcal{N}(0, 1) \quad t = 1, 2, \cdots, T$$

For illustration purposes, I assume the regressor $X_t$ as an univariate vector with dimension $T \times 1$. The predicted model is estimated as

$$\hat{Y}_{t+1} = \hat{\alpha} + \hat{\beta} X_t$$

And the conditional mean square error of point forecast $\hat{Y}$ can be obtained as

$$MSE(\hat{Y} | X) = E\left((\hat{Y} - Y)^2 | X\right)$$

$$= E(\hat{\alpha} + \hat{\beta} X - \alpha - \beta X - \epsilon)^2$$

$$= E\left((\hat{\alpha} - \alpha) + (\hat{\beta} - \beta) X - \epsilon\right)^2$$

$$= E(\hat{\alpha} - \alpha)^2 + X^2 E\left((\hat{\beta} - \beta)^2 | X\right) + E\epsilon^2 + 2XE\left((\hat{\alpha} - \alpha)(\hat{\beta} - \beta) | X\right)$$

$$- 2E((\hat{\alpha} - \alpha)\epsilon) - 2XE((\hat{\beta} - \beta)\epsilon | X)$$

$$= MSE(\hat{\alpha}) + X^2 MSE(\hat{\beta}) + \sigma^2 + 2Xcov(\hat{\alpha}, \hat{\beta}) - 2cov(\hat{\alpha}, \epsilon) - 2Xcov(\hat{\beta}, \epsilon)$$

In the out-of-sample prediction, the sample covariance between the estimated...
parameters ($\hat{\alpha}$ or $\hat{\beta}$) and the innovation process ($\epsilon$) can be derived as

$$
cov(\hat{\alpha}, \epsilon) = \frac{1}{T - P} \sum_{t=P}^{T} (\hat{\alpha}_t - \alpha)\epsilon_{t+1}
$$

$$
cov(\hat{\beta}, \epsilon) = \frac{1}{T - P} \sum_{t=P}^{T} (\hat{\beta}_t - \beta)\epsilon_{t+1}
$$

Where $P$ is the start date for out-of-sample prediction.

Assuming the innovation process $\epsilon_t$ is IID, these two terms will converge to zero

$$
cov(\hat{\alpha}, \epsilon) = 0, \quad cov(\hat{\beta}, \epsilon) = 0
$$

Therefore the $MSE(\hat{y})$ is reduced as

$$
MSE(\hat{Y}) = MSE(\hat{\alpha}) + X^2MSE(\hat{\beta}) + \sigma^2 + 2Xcov(\hat{\alpha}, \hat{\beta})
$$

In the linear regression model, the estimated intercept $\hat{\alpha}$ can be set as $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$ such that the sample mean $\bar{Y}$ can be precisely predicted $\hat{Y} = \hat{\alpha} + \hat{\beta}\bar{X}$ Therefore, the covariance between the estimated parameter can be further reduced as

$$
cov(\hat{\alpha}, \hat{\beta}) = -\bar{X}var(\hat{\beta})
$$

As a result, the MSE of predicted model $\hat{Y}$ is not only determined by the MSE of predicted parameters $\hat{\alpha}$ and $\hat{\beta}$, but also determined by the covariance of predicted parameters $cov(\hat{\alpha}, \hat{\beta})$, the sign of which could be undetermined. Therefore, a more precise estimation of the coefficient does not necessarily lead to a superior forecast of the response variable $\hat{Y}$.
2.9.2 Choice of Quantile LASSO penalty parameter $\lambda$ in Quantile Model Selection

The penalty parameter $\lambda$ in quantile LASSO procedure is determined in a bootstrap-type way following Belloni and Chernozhukov (2011), and Hautsch, Schaumburg and Schienle (2012).

- Step 1: Draw randomly a $T \times 1$ i.i.d process from uniform $(0, 1)$ distribution, denoted as $U_1, U_2, \cdots, U_T$, which are controlled to be independent with regressor, $x_1, x_2, \cdots, x_K$, and then calculate the corresponding value of the following random variable $\Lambda$

$$\Lambda = T \max_{1 \leq i \leq K} \left| E \left[ \frac{x_{t,i}(\tau - I\{U_t \leq \tau\})}{\hat{\sigma}_i \sqrt{\tau(1 - \tau)}} \right] \right|$$

(2.9.2.1)

where $\hat{\sigma}_i$ is the componentwise sample variance for each regressor $\hat{\sigma}_i = \frac{1}{T} \sum_{t=1}^{T} x_{t,i}$

- Step 2: Repeat Step 1 $R = 1000$ times, generate the bootstrap-type empirical distribution of random variables $\Lambda$ conditioning on regressors $X$ at each quantile: $\Lambda_1, \Lambda_2, \cdots, \Lambda_R$. For a selected confidence interval $\alpha = 0.10$, set the penalty parameter $\lambda$ as

$$\lambda = c \, Q(\Lambda^i, 1 - \alpha|X)$$

(2.9.2.2)

Where $Q(\Lambda^i, 1 - \alpha|X)$ denotes the $(1 - \alpha)$ quantile of empirical distribution for $\Lambda$. Following Belloni and Chernozhukov (2011), I set constant $c = 2$. Furthermore, the parameter $c$ can be set as the value which maximizes the in-sample equity premium predictability for the family of macro-finance variables.
Figure 2.1: This figure displays the combination weight (y axis) of quantile estimator $\beta(\tau)$ for a set of uniformly spaced quantiles (x axis) under different distribution assumption. The red line marked with circles denotes the optimal weights quantile (OWE) combination developed by Zhao & Xiao (2011). The blue line marked with squares display the density combination weights discussed in the section 2.3.2.
Figure 2.2: This figure displays the kernel density forecast \( f(Y_t | X_{t,t-1}) \) at 2010Q4 obtained either from some individual macro-finance variables (left panel), or from model combination estimations (right panel). The notation ‘Likelihood Weight’ refers to likelihood weighted density combination discussed in section 5.1. By analogy, ‘Pool Average’ denotes equal weight density combination. ‘DMSFE-Quantile-Density’ represent the DMSEF of model combination at each quantile firstly, followed by kernel density estimation. Normal Density is displayed as well with the identical value of first two moment at 2010Q4.
Figure 2.3: This figure displays utility gain (ΔU% and CER) for a variety of forecasts either obtained from model selection or derived from model selection jointed with quantile combination. The notation ‘LASSO’ denotes Square-Root Lasso estimation for conditional mean, which follow the method proposed by Belloni, Chernozhukov and Wang (2010). ‘LASSO + Regression’ refers to quantile lasso model selection at first step followed by Regression approach of quantile combination at second step. By analogy, ‘LASSO + DMSFE(0.9)’ represents quantile LASSO estimation firstly followed by quantile combination approach of discount mean forecast error (DMSFE) with discount factor 0.9.
Figure 2.4: This figure displays how the average utility gain (Y axis) under mean variance preference changes with absolute risk aversion coefficient (X axis). The black line marked with triangle is the value of $\Delta U(\%)$ derived from the return forecast only based on model combination; The red line marked with circle denotes that the values of $\Delta U(\%)$ are derived from the return forecast based on optimal weighted quantile combination followed by a specific model combination. By analogy, the blue line marked with diamond represents that the values of $\Delta U(\%)$ are derived from the return forecast based on density weighted quantile combination plus a specific model combination. The horizontal green line denotes zero $\Delta U(\%)$ which corresponds to the historical average benchmark forecast.
Figure 2.5: This figure displays how the CER (Y axis) under mean variance preference changes with absolute risk aversion coefficient (X axis). The black line marked with triangle is the value of CER derived from the return forecast only based on model combination; The red line marked with circle denotes that the values of CER are derived from the return forecast based on optimal weighted quantile combination followed by a specific model combination. By analogy, the blue line marked with diamond represents that the values of CER are derived from the return forecast based on density weighted quantile combination plus a specific model combination. The green line denotes the value of CER which corresponds to the historical average benchmark forecast.
Figure 2.6: This figure displays the time series of utility gain $\Delta U\%$ derived from a variety of forecasts under Mean Variance Utility with absolute risk aversion coefficient $\gamma = 5$. ‘OWQ + Pool Average’ denotes the forecast from optimal weight quantile combination in first step followed by pool average of model combination in second step. ‘Lasso + Regression’ denotes the forecast from quantile lasso model selection in first step followed by the Regression approach of quantile combination in second step.
Figure 2.7: This figure displays how the average utility gain (Y axis) under CRRA preference changes with relative risk aversion coefficient $\gamma$ (X axis). The red line marked with circle is the value of $\Delta U(\%)$ derived from the density forecast based on model combination followed by kernel density estimation; The bold horizontal green line denotes zero $\Delta U(\%)$ which corresponds to the historical benchmark density estimation. The acronyms ‘Likelihood-Weight-Density’ refers to likelihood weighted combination of density forecast. ‘Equal-Weight-Density’ refers to simple equal weighted combination of density forecast. ‘Mean-Quantile-Density’ denotes equal weight combination of quantile estimation firstly followed by density kernel forecast afterwards. DMSFE-Quantile-Density’ denotes DMSFE combination of quantile estimation firstly followed by density kernel forecast afterwards.
Figure 2.8: This figure displays how the Certain Equivalent Return \( CER \) (Y axis) under CRRA preference changes with relative risk aversion coefficient \( \gamma \) (X axis). The red line marked with circle is the values of \( CER \) derived from the density forecast based on model combination followed by kernel density estimation; The blue line marked with diamond denotes the values of \( CER \) which corresponds to the historical benchmark density forecast. The acronyms ‘Likelihood-Weight-Density’ refers to likelihood weighted combination of density forecast. ‘Equal-Weight-Density’ refers to simple equal weighted combination of density forecast. ‘Mean-Quantile-Density’ denotes equal weight combination of quantile estimation firstly followed by density kernel forecast afterwards. DMSFE-Quantile-Density’ denotes DMSFE combination of quantile estimation followed by density kernel forecast.
Table 2.1: Monte Carlo Simulation for The Forecast Performance of Quantile Combination (In-Sample)

**True Model:** \( Y_{t+1} = \alpha + \beta X_t + \epsilon_{t+1} \) and \( \alpha = 0, \beta = 1, X_t \perp \epsilon_{t+1} \)

**Predicted Model:** \( \hat{Y}_{t+1} = \hat{\alpha} + \hat{\beta} X_t \) where \( \hat{\beta} = \sum_{j=1}^{k} \omega_j \hat{\beta}(\tau_j), \hat{\alpha} = \bar{Y} - \bar{X} \hat{\beta} \)

**Evaluation:** \( RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{\beta}_i - \beta)^2} \) and \( \hat{\beta} = \sum_{j=1}^{k} \omega_j \hat{\beta}(\tau_j) \)

<table>
<thead>
<tr>
<th>Distribution</th>
<th>( \epsilon \sim N(0,1) )</th>
<th>( \epsilon \sim \text{LogNorm}(0,1) )</th>
<th>( \epsilon \sim t_1 )</th>
<th>( \epsilon \sim t_3 )</th>
<th>( \epsilon \sim \text{SkewT-Left} ) ( (\nu = 6, \lambda = -0.2) )</th>
<th>( \epsilon \sim \text{SkewT-Right} ) ( (\nu = 6, \lambda = 0.2) )</th>
<th>Mixture1</th>
<th>Mixture2</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>0.0577</td>
<td>0.0598</td>
<td>1.2895</td>
<td>0.0905</td>
<td>0.0600</td>
<td>0.0605</td>
<td>0.0587</td>
<td>0.0570</td>
</tr>
<tr>
<td>Median</td>
<td>0.0722</td>
<td>0.0366</td>
<td>0.0134</td>
<td>0.0798</td>
<td>0.0650</td>
<td>0.0651</td>
<td>0.0037</td>
<td>0.1677</td>
</tr>
<tr>
<td>MLE</td>
<td><strong>0.0551</strong></td>
<td><strong>0.0108</strong></td>
<td><strong>0.0082</strong></td>
<td><strong>0.0729</strong></td>
<td><strong>0.0552</strong></td>
<td><strong>0.0544</strong></td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Uniform-W</td>
<td>0.0592</td>
<td>0.0425</td>
<td>0.0235</td>
<td>0.0739</td>
<td>0.0567</td>
<td>0.0658</td>
<td>0.0440</td>
<td>0.0601</td>
</tr>
<tr>
<td>Density-W</td>
<td>0.0598</td>
<td>0.0193</td>
<td>0.0137</td>
<td>0.0735</td>
<td>0.0557</td>
<td>0.0557</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Optimal-W</td>
<td>0.0593</td>
<td>0.0152</td>
<td>0.0139</td>
<td>0.0740</td>
<td>0.0552</td>
<td>0.0551</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Density-W-Ker</td>
<td>0.0598</td>
<td>0.0210</td>
<td>0.0139</td>
<td>0.0735</td>
<td>0.0554</td>
<td>0.0551</td>
<td>0.0033</td>
<td>0.0476</td>
</tr>
<tr>
<td>Optimal-W-Ker</td>
<td>0.0623</td>
<td>0.0163</td>
<td>0.0152</td>
<td>0.0774</td>
<td>0.0589</td>
<td>0.0575</td>
<td>0.0028</td>
<td>0.0347</td>
</tr>
</tbody>
</table>

A uniform spaced quantile estimators are combined \( \tau = [0.1, 0.2, \ldots, 0.9] \)

<table>
<thead>
<tr>
<th>Distribution</th>
<th>( \epsilon \sim N(0,1) )</th>
<th>( \epsilon \sim \text{LogNorm}(0,1) )</th>
<th>( \epsilon \sim t_1 )</th>
<th>( \epsilon \sim t_3 )</th>
<th>( \epsilon \sim \text{SkewT-Left} ) ( (\nu = 6, \lambda = -0.2) )</th>
<th>( \epsilon \sim \text{SkewT-Right} ) ( (\nu = 6, \lambda = 0.2) )</th>
<th>Mixture1</th>
<th>Mixture2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uniform-W</td>
<td>0.0592</td>
<td>0.0402</td>
<td>0.0190</td>
<td>0.0734</td>
<td>0.0561</td>
<td>0.0564</td>
<td>0.0385</td>
<td>0.0618</td>
</tr>
<tr>
<td>Density-W</td>
<td>0.0599</td>
<td>0.0200</td>
<td>0.0139</td>
<td>0.0733</td>
<td>0.0557</td>
<td>0.0557</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Optimal-W</td>
<td>0.0592</td>
<td>0.0151</td>
<td>0.0142</td>
<td>0.0738</td>
<td>0.0559</td>
<td>0.0550</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Density-W-Ker</td>
<td>0.0600</td>
<td>0.0215</td>
<td>0.0140</td>
<td>0.0734</td>
<td>0.0554</td>
<td>0.0551</td>
<td>0.0026</td>
<td>0.0485</td>
</tr>
<tr>
<td>Optimal-W-Ker</td>
<td>0.0643</td>
<td>0.0171</td>
<td>0.0144</td>
<td>0.0778</td>
<td>0.0602</td>
<td>0.0588</td>
<td>0.0025</td>
<td>0.0355</td>
</tr>
</tbody>
</table>

This table reports the performance evaluation (RMSE) for the estimation of Beta Risk \( \beta \) from the simulated CAPM model. The Monte Carlo simulation exercises are repeated 1000 times with observation \( T = 300 \) per replication.

1. The ‘Mixture1’ Gaussian distribution is generated by \( u = \omega z + (1 - \omega) z_2 \), where \( z_1 \sim N(0, 1/9), z_2 \sim N(0, 9) \) and \( \omega = 0.5 \).
2. The ‘Mixture2’ Gaussian distribution is generated by \( u = \omega z + (1 - \omega) z_2 \), where \( z_1 \sim N(-2, 1), z_2 \sim N(2, 1) \) and \( \omega = 0.5 \).
3. ‘Uniform-W’ refers to Equal weight quantile combination estimation where \( \hat{\beta} = \frac{1}{n} \sum_{i=1}^{n} \hat{\beta}(\tau_i) \), \( \omega_i = \frac{1}{n} \).
4. ‘Density-W’: Density Weight Quantile Combination \( \omega_i = \frac{f_i(F_i^{-1}(\tau_i))}{\sqrt{\tau_i(1-\tau_i)}} / S_w \) and \( S_w = \sum_{i=1}^{n} \frac{f_i(F_i^{-1}(\tau_i))}{\sqrt{\tau_i(1-\tau_i)}} \) where the density \( f_i(F_i^{-1}(\tau_i)) \) is obtained by the theoretical simulated distribution.
5. ‘Optimal-W’: Optimal weight Quantile Combination developed by Zhao & Xiao (2011), where the density is obtained by the theoretical simulated distribution.
6. ‘Density-W-Ker’: Density Weight Quantile Combination where the density \( f_i(F_i^{-1}(\tau_i)) \) is estimated by the non-parametric kernel estimation.
7. ‘Optimal-W-Ker’: Optimal weight Quantile Combination developed by Zhao & Xiao (2011), where the density is estimated by the non-parametric kernel estimation.
Table 2.2: Monte Carlo Simulation for The Forecast Performance of Quantile Combination (Out-of-Sample)

True Model: \( Y_{t+1} = \alpha + \beta X_t + \sigma u_t \) and \( \alpha = 0.0946, \ \beta = 0.0243, \ u_t \sim \mathcal{F}(0,1) \)

Predicted Model: \( \hat{Y}_{t+1} = \hat{\alpha} + \hat{\beta} X_t \) where \( \hat{\beta} = \sum_{j=1}^{n} \omega_j \hat{\beta}(\tau_j) \), \( \hat{\alpha} = \hat{Y} - \hat{X} \hat{\beta} \)

Evaluation: \( \text{RMSE}_{\beta} = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (\beta_t - \beta)^2} \) and \( \text{RMSE}_Y = \sqrt{\frac{1}{T-R} \sum_{t=1}^{T-R} (\hat{Y}_{t+1} - Y_{t+1})^2} \)

<table>
<thead>
<tr>
<th>Distribution</th>
<th>( \epsilon \sim N(0,1) )</th>
<th>( \epsilon \sim \text{LogNorm}(0,1) )</th>
<th>( \epsilon \sim \text{t}_1 )</th>
<th>( \epsilon \sim \text{t}_3 )</th>
<th>( \epsilon \sim \text{Left Skewed St} )</th>
<th>Mixture1</th>
<th>Mixture2</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>Uniform-Weight</td>
<td>1.0303</td>
<td>0.7173</td>
<td>0.1602</td>
<td>0.8906</td>
<td>1.2177</td>
<td>0.8133</td>
<td>1.0358</td>
</tr>
<tr>
<td>Density-Weight</td>
<td>1.0486</td>
<td>0.3480</td>
<td>0.0996</td>
<td>0.8555</td>
<td>1.2753</td>
<td>0.1058</td>
<td>0.8545</td>
</tr>
<tr>
<td>Optimal-Weight</td>
<td>1.1252</td>
<td>0.2690</td>
<td>0.1004</td>
<td>0.9412</td>
<td>1.6270</td>
<td>0.1398</td>
<td>0.6601</td>
</tr>
</tbody>
</table>

Relative RMSE for \( \hat{Y}_{t+1} \): \( \frac{\text{RMSE}}{\text{RMSE(O LS)}} \) where \( \text{RMSE} = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (\hat{\beta}_t - \beta)^2} \)

<table>
<thead>
<tr>
<th>Distribution</th>
<th>( \epsilon \sim N(0,1) )</th>
<th>( \epsilon \sim \text{LogNorm}(0,1) )</th>
<th>( \epsilon \sim \text{t}_1 )</th>
<th>( \epsilon \sim \text{t}_3 )</th>
<th>( \epsilon \sim \text{Left Skewed St} )</th>
<th>Mixture1</th>
<th>Mixture2</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>Uniform-W</td>
<td>1.0010</td>
<td>0.9976</td>
<td>0.9641</td>
<td>0.9948</td>
<td>0.9671</td>
<td>0.9917</td>
<td>1.0005</td>
</tr>
<tr>
<td>Density-W-Ker</td>
<td>1.0010</td>
<td>0.9797</td>
<td>0.9549</td>
<td>0.9941</td>
<td>0.9671</td>
<td>0.9803</td>
<td>0.9976</td>
</tr>
<tr>
<td>Optimal-W-Ker</td>
<td>1.0036</td>
<td>0.9771</td>
<td>0.9533</td>
<td>0.9959</td>
<td>0.9516</td>
<td>0.9829</td>
<td>0.9948</td>
</tr>
<tr>
<td>DMSFE</td>
<td>1.0013</td>
<td>0.9509</td>
<td>0.9339</td>
<td>0.9937</td>
<td>0.9563</td>
<td>0.9831</td>
<td>1.0045</td>
</tr>
<tr>
<td>Regression</td>
<td>1.0000</td>
<td>0.9554</td>
<td>0.9534</td>
<td>0.9985</td>
<td>0.9998</td>
<td>0.9993</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

This table reports the Monte Carlo Simulation results for the forecast performance of coefficient \( \hat{\beta} \) and point forecast \( \hat{Y}_{t+1} \) which are estimated using a wide range of quantile combination approaches. All parameters are estimated recursively in the out of sample forecast exercises, in which the parameters estimation are updated as forecast moves forward by adding one additional observation to the estimation sample at each step. The total sample \( T = 337 \) is divided into an in-sample portion \( R = 100 \) and out of sample evaluation period \( P = 237 \). Monte Carlo simulation is based on 1000 replication of simulated data with \( T = 337 \) observation per replication.

The selected slope coefficient \( \beta = 0.0243 \) is the OLS slope coefficient for the univariate forecast model with dividend yield ratio (DY) as regressor during 1926Q4 ~ 2010Q4. The scale value \( \sigma \) is selected as \( \sigma^2 = \frac{1-R^2}{R^2} \cdot \beta^2 \cdot \text{Var}(X) \) such that the goodness of fit \( R^2 \) for the simulated regression model is similar to the empirical forecast model with \( R^2 = 1.03\% \).

1 The ‘Mixture’ Gaussian distribution is generated by \( u = \omega_1 + (1 - \omega)Z_2 \), where \( z_1 \sim N(0,1/9) \), \( z_2 \sim N(0,9) \) and \( \omega = 0.5 \).
2 The ‘Mixture’ Gaussian distribution is generated by \( u = \omega_1 + (1 - \omega)Z_2 \), where \( z_1 \sim N(-2,1) \), \( z_2 \sim N(2,1) \) and \( \omega = 0.5 \).
3 ‘Uniform-W’ refers to Equal weight quantile combination estimation where \( \hat{\beta} = \frac{1}{T} \sum_{i=1}^{T} \beta(\tau_i) \), \( \omega = \frac{1}{T} \).
4 ‘Density-W-Ker’: Density Weight Quantile Combination \( \omega_i = \frac{f(F^{-1}(\tau_i))}{\sqrt{\tau_i(1-\tau_i)}} / S_w \) and \( S_w = \sum_i^{N} \frac{f(F^{-1}(\tau_i))}{\sqrt{\tau_i(1-\tau_i)}} \), where the density is estimated by the non-parametric kernel estimation.
5 ‘Optimal-W-Ker’: Optimal weight Quantile Combination developed by Zhao & Xiao (2011), where the density is estimated by the non-parametric kernel estimation.
6 ‘DMSFE’: Discount mean square forecast error (DMSFE) approaches to combine quantile estimator, see the discussion in section 3.1
7 ‘Regression’: Regression Approach of Quantile Combination, see the discussion in section 3.1
Table 2.3: Slope coefficient estimates for the In-Sample Stock Return Prediction (1946Q4:2010Q4)

Panel A: Univariate OLS Regression

<table>
<thead>
<tr>
<th></th>
<th>DP</th>
<th>DY</th>
<th>EP</th>
<th>DE</th>
<th>SVAR</th>
<th>BM</th>
<th>NTIS</th>
<th>TBL</th>
<th>LTY</th>
<th>LTR</th>
<th>TMS</th>
<th>DFY</th>
<th>DFR</th>
<th>INFL</th>
</tr>
</thead>
<tbody>
<tr>
<td>β</td>
<td>0.027</td>
<td>0.030</td>
<td>0.015</td>
<td>0.022</td>
<td>-0.262</td>
<td>0.026</td>
<td>0.059</td>
<td>-0.322</td>
<td>-0.224</td>
<td>0.119</td>
<td>0.591</td>
<td>0.189</td>
<td>0.599</td>
<td>-0.712</td>
</tr>
<tr>
<td>t-stat</td>
<td>2.42</td>
<td>2.69</td>
<td>1.44</td>
<td>1.28</td>
<td>-0.48</td>
<td>1.34</td>
<td>0.22</td>
<td>-1.95</td>
<td>-1.26</td>
<td>1.20</td>
<td>1.65</td>
<td>0.17</td>
<td>2.70</td>
<td>-1.63</td>
</tr>
<tr>
<td>R2(%)</td>
<td>2.26</td>
<td>2.79</td>
<td>0.81</td>
<td>0.64</td>
<td>0.09</td>
<td>0.71</td>
<td>0.02</td>
<td>1.47</td>
<td>0.62</td>
<td>0.56</td>
<td>1.06</td>
<td>0.01</td>
<td>2.78</td>
<td>1.04</td>
</tr>
</tbody>
</table>

Panel B: Univariate Quantile Regression

<table>
<thead>
<tr>
<th>τ</th>
<th>0.05</th>
<th>0.10</th>
<th>0.20</th>
<th>0.30</th>
<th>0.40</th>
<th>0.50</th>
<th>0.60</th>
<th>0.70</th>
<th>0.80</th>
<th>0.90</th>
<th>0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>DP</td>
<td><strong>0.06</strong>*</td>
<td><strong>0.06</strong>*</td>
<td><strong>0.06</strong>*</td>
<td>0.01</td>
<td>-5.42***</td>
<td>0.05</td>
<td>1.76**</td>
<td>-0.22</td>
<td>-0.48</td>
<td>-0.28</td>
<td>-0.12</td>
</tr>
<tr>
<td>DY</td>
<td><strong>0.07</strong>*</td>
<td><strong>0.06</strong>*</td>
<td>*<em>0.04</em></td>
<td>-0.02</td>
<td>-5.19***</td>
<td>0.03</td>
<td>0.82</td>
<td>-0.69**</td>
<td>-0.63**</td>
<td>-0.16</td>
<td>0.41</td>
</tr>
<tr>
<td>EP</td>
<td><strong>0.06</strong>*</td>
<td><strong>0.06</strong>*</td>
<td>*<em>0.04</em></td>
<td>0.02</td>
<td>-0.66</td>
<td>0.00</td>
<td>0.44</td>
<td>-0.42*</td>
<td>-0.41*</td>
<td>-0.01</td>
<td>0.97*</td>
</tr>
<tr>
<td>DE</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.03</td>
<td>-0.90</td>
<td>0.00</td>
<td>0.17</td>
<td>-0.44**</td>
<td>-0.28</td>
<td>0.20</td>
<td>0.87*</td>
</tr>
<tr>
<td>SVAR</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.03</td>
<td>-1.08</td>
<td>-0.01</td>
<td>-0.08</td>
<td>-0.49**</td>
<td>-0.45**</td>
<td>0.24**</td>
<td>0.65</td>
</tr>
<tr>
<td>BM</td>
<td>0.03*</td>
<td>0.02*</td>
<td>0.01</td>
<td>0.03</td>
<td>-1.08</td>
<td>-0.01</td>
<td>-0.08</td>
<td>-0.49**</td>
<td>-0.45**</td>
<td>0.24**</td>
<td>0.65</td>
</tr>
<tr>
<td>NTIS</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.03</td>
<td>0.25</td>
<td>0.03</td>
<td>-0.43</td>
<td>-0.39*</td>
<td>-0.28</td>
<td>0.20</td>
<td>0.71</td>
</tr>
<tr>
<td>TBL</td>
<td>0.03*</td>
<td>0.02*</td>
<td>0.01</td>
<td>0.04*</td>
<td>1.54***</td>
<td>0.04**</td>
<td>-0.34</td>
<td>-0.11</td>
<td>-0.03</td>
<td>0.18</td>
<td>0.58</td>
</tr>
<tr>
<td>LTY</td>
<td>0.03**</td>
<td>0.04***</td>
<td>0.03***</td>
<td>0.02</td>
<td>1.54***</td>
<td>0.04**</td>
<td>-0.34</td>
<td>-0.11</td>
<td>-0.03</td>
<td>0.18</td>
<td>0.58</td>
</tr>
<tr>
<td>LTR</td>
<td>0.03**</td>
<td>0.04***</td>
<td>0.03***</td>
<td>0.02</td>
<td>1.54***</td>
<td>0.04**</td>
<td>-0.34</td>
<td>-0.11</td>
<td>-0.03</td>
<td>0.18</td>
<td>0.58</td>
</tr>
<tr>
<td>TMS</td>
<td>0.03**</td>
<td>0.03**</td>
<td>0.01</td>
<td>0.04*</td>
<td>1.82***</td>
<td>0.04**</td>
<td>-0.66**</td>
<td>-0.32</td>
<td>-0.11</td>
<td>0.12</td>
<td>1.30***</td>
</tr>
<tr>
<td>DFY</td>
<td>0.01</td>
<td>0.01</td>
<td>-0.01</td>
<td>0.02</td>
<td>3.49***</td>
<td>0.03</td>
<td>-0.73**</td>
<td>-0.28</td>
<td>-0.25</td>
<td>0.07</td>
<td>1.37***</td>
</tr>
<tr>
<td>DFR</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>0.01</td>
<td>4.17***</td>
<td>0.01</td>
<td>-0.39</td>
<td>0.04</td>
<td>0.28</td>
<td>0.11</td>
<td>1.86</td>
</tr>
<tr>
<td>INFL</td>
<td>0.12</td>
<td>0.36</td>
<td>0.53</td>
<td>0.10</td>
<td>0.28</td>
<td>0.05</td>
<td>0.34</td>
<td>0.83</td>
<td>0.24</td>
<td>0.16</td>
<td>0.46</td>
</tr>
</tbody>
</table>

Panel C: Wald test statistic for slope equality test

| F-val | 1.54 | 1.09 | 0.90 | **1.59*** | 1.21 | **1.85*** | 1.12 | 0.58 | 1.28 | 1.44 | 0.96 | **8.81*** | 0.62 | 0.87 |
| P-val | 0.12 | 0.36 | 0.53 | 0.10 | 0.28 | 0.05 | 0.34 | 0.83 | 0.24 | 0.16 | 0.46 | 0.00 | 0.80 | 0.55 |

This table reports the slope coefficient estimates for each of predictor variable in the univariate quantile and OLS regression model. The sample period cover from 1946Q4 to 2010Q4. The standard error of the coefficient estimates for univariate quantile regression are based on bootstrapped standard errors. The symbols *, **, *** denote significance at the 10%, 5% and 1% levels, respectively.
Table 2.4: Slope coefficient estimates for the In-Sample Stock Return Prediction (19464:20104)

**Panel A: OLS Univariate Regression**

<table>
<thead>
<tr>
<th></th>
<th>DP</th>
<th>DY</th>
<th>EP</th>
<th>DE</th>
<th>SVAR</th>
<th>BM</th>
<th>NTIS</th>
<th>TBL</th>
<th>LTY</th>
<th>LTR</th>
<th>TMS</th>
<th>DFY</th>
<th>DFR</th>
<th>INFL</th>
</tr>
</thead>
<tbody>
<tr>
<td>β</td>
<td>0.027</td>
<td>0.030</td>
<td>0.015</td>
<td>0.022</td>
<td>-0.262</td>
<td>0.026</td>
<td>0.059</td>
<td>-0.322</td>
<td>-0.224</td>
<td>0.119</td>
<td>0.591</td>
<td>0.189</td>
<td>0.599</td>
<td>-0.712</td>
</tr>
<tr>
<td>t-stat</td>
<td>2.42</td>
<td>2.69</td>
<td>1.44</td>
<td>1.28</td>
<td>-0.48</td>
<td>1.34</td>
<td>0.22</td>
<td>-1.95</td>
<td>-1.26</td>
<td>1.20</td>
<td>1.65</td>
<td>0.17</td>
<td>2.70</td>
<td>-1.63</td>
</tr>
<tr>
<td>R²(%)</td>
<td>2.26</td>
<td>2.79</td>
<td>0.81</td>
<td>0.64</td>
<td>0.09</td>
<td>0.71</td>
<td>0.02</td>
<td>1.47</td>
<td>0.62</td>
<td>0.56</td>
<td>1.06</td>
<td>0.01</td>
<td>2.78</td>
<td>1.04</td>
</tr>
</tbody>
</table>

**Panel B: OLS Lasso Regression (Belloni, Chernozhukov & Wang (2010))**

<table>
<thead>
<tr>
<th></th>
<th>DP</th>
<th>DY</th>
<th>EP</th>
<th>DE</th>
<th>SVAR</th>
<th>BM</th>
<th>NTIS</th>
<th>TBL</th>
<th>LTY</th>
<th>LTR</th>
<th>TMS</th>
<th>DFY</th>
<th>DFR</th>
<th>INFL</th>
</tr>
</thead>
<tbody>
<tr>
<td>β</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>t-stat</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>2.1125</td>
<td>-0.8356</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>2.0369</td>
<td>-1.2639</td>
<td>2.4052</td>
<td>0.0</td>
</tr>
<tr>
<td>R²(%)</td>
<td>5.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Panel C: Quantile Lasso Regression (Belloni & Chernozhukov (2011))**

<table>
<thead>
<tr>
<th>τ</th>
<th>DP</th>
<th>DY</th>
<th>EP</th>
<th>DE</th>
<th>SVAR</th>
<th>BM</th>
<th>NTIS</th>
<th>TBL</th>
<th>LTY</th>
<th>LTR</th>
<th>TMS</th>
<th>DFY</th>
<th>DFR</th>
<th>INFL</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.398</td>
<td></td>
<td>0.119</td>
<td></td>
<td>-2.603</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.193*</td>
<td></td>
<td>-5.379***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.20</td>
<td>0.063**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-2.598</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.30</td>
<td>0.031</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-2.318</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.40</td>
<td>0.014</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-1.951</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.50</td>
<td>0.006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.166</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.60</td>
<td>0.219</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.550</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.70</td>
<td>0.011</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.917**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.80</td>
<td>0.006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.862***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.90</td>
<td>-0.003</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3.968*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.95</td>
<td>-0.007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.492</td>
<td></td>
<td>0.718</td>
<td></td>
</tr>
</tbody>
</table>

This table reports the post-LASSO estimation of slope coefficient after model selection via LASSO procedures. The LASSO selection for OLS estimation in Panel B follows the procedure proposed by Belloni, Chernozhukov & Wang (2010). The Quantile Lasso Regression in Panel C follows Belloni & Chernozhukov (2011). The symbols *, **, *** denote significance at the 10%, 5% and 1% levels, respectively.
Table 2.5: Out of Sample Predictive $R^2_{OS}(\%)$ for Conditional Mean Regression (19571:20104)

<table>
<thead>
<tr>
<th>Panel A: Univariate Least Square Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predictor Variable</td>
</tr>
<tr>
<td>Overall</td>
</tr>
<tr>
<td>Recession</td>
</tr>
<tr>
<td>Expansion</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Predictive Model Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model Selection</td>
</tr>
<tr>
<td>Sink-Model</td>
</tr>
<tr>
<td>AIC</td>
</tr>
<tr>
<td>LASSO</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C: Model Combination Over the Estimated $\hat{Y}_{t+1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model Combination Over the Estimated $\hat{Y}_{t+1}$</td>
</tr>
<tr>
<td>Overall Recession Expansion</td>
</tr>
<tr>
<td>Pool Average</td>
</tr>
<tr>
<td>Pool Median</td>
</tr>
<tr>
<td>Pool DMSFE(1)</td>
</tr>
<tr>
<td>Pool DMSFE(0.75)</td>
</tr>
<tr>
<td>PCA</td>
</tr>
</tbody>
</table>

This table reports the Out of Sample predictive $R^2(\%)$ for conditional mean forecast based on each predictor variable (panel A). Multivariate forecast based some model selection procedures (Panel B) and forecasts based on multiple model combination (Panel C). The out of sample period start from 1957 Q1 to 2010 Q4. The value in Brackets report P-value for the Clark and West (2007) MSFE-adjusted statistic. ‘Recession’ denotes NBER-dated business-cycle recession period. The symbols *, **, *** denote significance at the 10%, 5% and 1% levels, respectively.
| Panel A: Individual Conditional Quantile Model Based on Each Predictor Variable |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| $\tau =$        | 0.10           | 0.20           | 0.30           | 0.40           | 0.50           | 0.60           | 0.70           | 0.80           | 0.90           |
| $DP$            | -3.27          | -2.08          | -1.79          | -1.55          | 1.12           | 1.91           | 2.83           | 2.08           | -7.16          |
| $DY$            | -1.93          | -2.44          | -1.49          | -1.25          | 1.09           | 1.83           | 3.04           | 1.97           | -2.08          |
| $EP$            | -3.71          | -4.68          | -2.93          | -2.40          | 0.62           | 0.64           | 0.70           | -1.16          | -4.92          |
| $DE$            | -5.88          | -6.81          | -7.38          | -6.14          | -2.35          | -1.11          | 0.48           | 1.61           | -0.23          |
| SVAR            | 2.30           | -1.14          | -0.17          | -1.92          | -1.77          | -0.52          | 1.04**         | 1.45**         | -2.15          |
| $BM$            | -0.92          | -1.47          | -1.36          | -1.02          | 0.15           | 0.45           | 1.86           | -0.77          | -4.34          |
| NTIS            | -1.88          | -3.12          | -1.35          | -0.77          | 0.27           | -0.20          | -0.36          | -0.34          | -0.57          |
| TBL             | -1.23          | -4.26          | -1.36          | -1.75          | -0.37          | -1.22          | -0.20          | 0.08           | -2.49          |
| LTY             | -8.23          | -8.42          | -4.27          | -1.94          | -0.77          | -2.29          | -2.39          | -1.61          | -4.93          |
| LTR             | -0.49          | -3.69          | -0.54          | 0.61           | -0.27          | 0.37           | -0.43          | -0.99          | -0.87          |
| TMS             | -4.96          | -7.34          | -2.60          | -2.37          | -1.36          | -0.56          | 0.50           | 0.10           | 0.64           |
| DFY             | 0.65           | -1.67          | -0.94          | -1.10          | -1.52          | 1.25**         | 2.93***         | 3.14*          | -1.37          |
| DFR             | -2.85          | -3.48          | 0.54           | -0.53          | -0.38          | -0.44          | -0.21          | -0.70          | -3.13          |
| $INFL - lag$    | -1.43          | -3.04          | -0.73          | 0.08           | 0.12           | 0.17           | 0.12           | -0.44          | -1.02          |
| Quantile-Lasso  | -0.21          | 1.81**         | 2.93***        | 1.73*          | 0.15           | 1.29           | 0.60           | 2.03           | 1.90           |
| P-val           | 0.53           | 0.04           | 0.00           | 0.06           | 0.47           | 0.22           | 0.39           | 0.17           | 0.24           |

| Panel B: Model Combination Over Each Quantile |
|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| $\tau =$        | 0.10           | 0.20           | 0.30           | 0.40           | 0.50           | 0.60           | 0.70           | 0.80           | 0.90           |
| Mean            | 1.13           | -0.74          | 0.13           | -0.18          | 0.29           | 1.17***        | 2.92***         | 1.99***         | 5.18***        |
| Median          | 1.61           | -1.04          | -0.73          | -0.34          | 0.24           | 0.89***        | 2.33***         | 1.29***         | 2.83***        |
| DMSFE           | 1.33           | -0.77          | 0.13           | -0.17          | 0.31           | 1.18***        | 2.95***         | 2.01***         | 5.11***        |
| PCA             | 5.95***        | 1.73*          | 0.72           | 0.43           | 1.14*          | 1.86***        | 3.93***         | 2.64**         | 1.36           |

| Panel C: Model Combination at each quantile firstly + Quantile Combination afterwards |

<table>
<thead>
<tr>
<th>Overall</th>
<th>Recession</th>
<th>Expansion</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA + DMSFE (1)</td>
<td>0.78* (0.08)</td>
<td>1.15 (0.11)</td>
</tr>
<tr>
<td>PCA + DMSFE (0.9)</td>
<td>0.58 (0.21)</td>
<td>1.94 (0.13)</td>
</tr>
<tr>
<td>PCA + Regression</td>
<td>1.07* (0.10)</td>
<td>7.81*** (0.01)</td>
</tr>
</tbody>
</table>

| Panel D: Quantile Lasso firstly + Quantile Combination afterwards |

<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th>Recession</th>
<th>Expansion</th>
</tr>
</thead>
<tbody>
<tr>
<td>LASSO + DMSFE (1)</td>
<td>3.06** (0.04)</td>
<td>5.51* (0.07)</td>
<td>0.93 (0.16)</td>
</tr>
<tr>
<td>LASSO + DMSFE (0.9)</td>
<td>2.83* (0.06)</td>
<td>6.12* (0.07)</td>
<td>-0.01 (0.23)</td>
</tr>
<tr>
<td>LASSO + Regression</td>
<td>1.88** (0.04)</td>
<td>7.76** (0.04)</td>
<td>-3.21 (0.24)</td>
</tr>
</tbody>
</table>

This table reports the Out of Sample predictive $R^2(\%)$ for conditional univariate Quantile forecast (Panel A), Quantile forecast $\hat{Y}(\tau)$ developed by combing multiple forecasts at each quantile $\tau$ (Panel B), conditional mean $\hat{Y}$ forecast developed by Model Combination at each quantile followed by quantile combination afterwards (Panel C), and model selection (Quantile LASSO) at first step followed by quantile combination at second step (Panel D).

The acronyms DMSFE denotes the combination approach of Discount Mean Square Forecast Error (DMSFE). Regression denotes the regression approach for quantile combination discussed in section 2. PCA: Principal Component Analysis. The bootstrap P-values are reported in quantile forecasts which are derived from re-sampling the sample for 1000 times. The value in Brackets of panel C report P-value for the Clark and West (2007) MSFE-adjusted statistic. The symbols *, **, *** denote significance at the 10%, 5% and 1% levels, respectively.
Table 2.7: Out of Sample Predictive $R^2$ (%) for the Forecast of Quantile Combination Followed by Model Combination (19571:20104)

### Panel A: Quantile Combination for the Predictive Model Based on Each Individual Variable

<table>
<thead>
<tr>
<th></th>
<th>DP</th>
<th>DY</th>
<th>EP</th>
<th>DE</th>
<th>SVAR</th>
<th>BM</th>
<th>NTIS</th>
<th>TBL</th>
<th>LTY</th>
<th>LTR</th>
<th>TMS</th>
<th>DFY</th>
<th>DFR</th>
<th>INFL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau = 0.1$</td>
<td>-3.27</td>
<td>-1.93</td>
<td>-3.71</td>
<td>-5.88</td>
<td>2.30</td>
<td>-0.92</td>
<td>-1.88</td>
<td>-1.23</td>
<td>-8.23</td>
<td>-0.49</td>
<td>-4.96</td>
<td>0.65</td>
<td>-2.85</td>
<td>-1.43</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$\tau = 0.5$</td>
<td>1.12</td>
<td>1.09</td>
<td>0.62</td>
<td>-2.35</td>
<td>-1.77</td>
<td>0.15</td>
<td>-0.27</td>
<td>-0.37</td>
<td>-0.77</td>
<td>-0.27</td>
<td>-1.36</td>
<td>-1.52</td>
<td>-0.38</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$\tau = 0.9$</td>
<td>-7.16</td>
<td>-2.08</td>
<td>-4.92</td>
<td>-0.23</td>
<td>-2.15</td>
<td>-4.34</td>
<td>-0.57</td>
<td>-2.49</td>
<td>-4.93</td>
<td>-0.87</td>
<td>0.64</td>
<td>-1.37</td>
<td>-3.13</td>
<td>-1.02</td>
</tr>
</tbody>
</table>

### Quantile Combine

<table>
<thead>
<tr>
<th></th>
<th>DP</th>
<th>DY</th>
<th>EP</th>
<th>DE</th>
<th>SVAR</th>
<th>BM</th>
<th>NTIS</th>
<th>TBL</th>
<th>LTY</th>
<th>LTR</th>
<th>TMS</th>
<th>DFY</th>
<th>DFR</th>
<th>INFL</th>
</tr>
</thead>
<tbody>
<tr>
<td>OWQ</td>
<td>0.63**</td>
<td>0.30**</td>
<td>-3.22</td>
<td>-1.43</td>
<td>2.00***</td>
<td>-1.74</td>
<td>-0.61</td>
<td>-0.42</td>
<td>-2.00</td>
<td>-0.15</td>
<td>0.27</td>
<td>-0.16</td>
<td>-1.02</td>
<td>0.99</td>
</tr>
<tr>
<td>Density</td>
<td>0.78**</td>
<td>1.09**</td>
<td>-2.47</td>
<td>-3.10</td>
<td>0.02</td>
<td>-0.24</td>
<td>-1.08</td>
<td>-0.64</td>
<td>-2.15</td>
<td>-0.42</td>
<td>-0.55</td>
<td>0.24</td>
<td>-0.40</td>
<td>0.76</td>
</tr>
<tr>
<td>Regression</td>
<td>0.85*</td>
<td>0.95*</td>
<td>-3.35</td>
<td>-6.81</td>
<td>-1.93</td>
<td>-0.89</td>
<td>-1.40</td>
<td>-1.31</td>
<td>-3.23</td>
<td>-0.43</td>
<td>-2.68</td>
<td>-3.57</td>
<td>-1.75</td>
<td>0.13</td>
</tr>
<tr>
<td>DFMSE</td>
<td>0.34</td>
<td>0.79*</td>
<td>-3.31</td>
<td>-8.45</td>
<td>-3.06</td>
<td>-2.16</td>
<td>-3.84</td>
<td>-1.59</td>
<td>-1.40</td>
<td>-2.29</td>
<td>-4.49</td>
<td>-3.25</td>
<td>-2.70</td>
<td>-1.09</td>
</tr>
<tr>
<td>OLS</td>
<td>-0.80*</td>
<td>0.36*</td>
<td>-6.51</td>
<td>-4.93</td>
<td>-0.47</td>
<td>-6.87</td>
<td>-5.38</td>
<td>-0.56</td>
<td>-3.43</td>
<td>-0.72</td>
<td>-0.12</td>
<td>0.15</td>
<td>-4.79</td>
<td>0.16</td>
</tr>
</tbody>
</table>

### Panel B: Quantile Combination Firstly + Model Combination Afterwards

<table>
<thead>
<tr>
<th></th>
<th>Overall</th>
<th>Recession</th>
<th>Expansion</th>
</tr>
</thead>
<tbody>
<tr>
<td>OWQ+Pool Average</td>
<td>1.86*** (0.01)</td>
<td>3.63*** (0.01)</td>
<td>0.51 (0.17)</td>
</tr>
<tr>
<td>OWQ+Pool Median</td>
<td>1.64*** (0.01)</td>
<td>2.49** (0.02)</td>
<td>0.99* (0.08)</td>
</tr>
<tr>
<td>OWQ+Pool DMSFE (1)</td>
<td>1.82*** (0.01)</td>
<td>3.58*** (0.01)</td>
<td>0.48 (0.18)</td>
</tr>
<tr>
<td>OWQ+Pool DMSFE (0.75)</td>
<td>2.25*** (0.01)</td>
<td>4.49** (0.02)</td>
<td>0.53 (0.18)</td>
</tr>
<tr>
<td>OWQ+PAC</td>
<td>-0.52 (0.38)</td>
<td>0.74 (0.32)</td>
<td>-1.49 (0.56)</td>
</tr>
<tr>
<td>Density+Pool Average</td>
<td>1.24** (0.02)</td>
<td>2.27** (0.03)</td>
<td>0.44 (0.20)</td>
</tr>
<tr>
<td>Density+Pool Median</td>
<td>0.89** (0.02)</td>
<td>1.36** (0.05)</td>
<td>0.54 (0.12)</td>
</tr>
<tr>
<td>Density+Pool DMSFE (1)</td>
<td>1.23** (0.03)</td>
<td>2.27** (0.03)</td>
<td>0.43 (0.20)</td>
</tr>
<tr>
<td>Density+Pool DMSFE (0.75)</td>
<td>1.61** (0.03)</td>
<td>3.16** (0.04)</td>
<td>0.43 (0.20)</td>
</tr>
<tr>
<td>Density+PAC</td>
<td>-2.23 (0.90)</td>
<td>-2.51 (0.79)</td>
<td>-2.03 (0.87)</td>
</tr>
</tbody>
</table>

This table reports out of sample predictive $R^2$ (%) for the forecast derived from Quantile Combination firstly followed by Model Combination afterwards. The acronyms OWQ denotes the **Optimal Weight** quantile combination approach developed by Xiao and Zhao (2011). Density denotes the density weight quantile combination discussed before. The values in Brackets of panel B report P-value for the Clark and West (2007) MSFE-adjusted statistic. “Recession” denotes NBER-dated business-cycle recession period. The symbols *, **, *** denote significance at the 10%, 5% and 1% levels, respectively.
Table 2.8: Utility Gain Under Mean Variance Utility Preference ($\gamma = 5$)

### Panel A: Individual Model

<table>
<thead>
<tr>
<th></th>
<th>DP</th>
<th>DY</th>
<th>EP</th>
<th>DE</th>
<th>SVAR</th>
<th>BM</th>
<th>NTIS</th>
<th>TBL</th>
<th>LTY</th>
<th>LTR</th>
<th>TMS</th>
<th>DFY</th>
<th>DFR</th>
<th>INFL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta U(%)$</td>
<td>1.44</td>
<td>1.73</td>
<td>0.48</td>
<td>-2.12</td>
<td>-0.29</td>
<td>-1.33</td>
<td>-1.26</td>
<td>0.89</td>
<td>1.02</td>
<td>-0.75</td>
<td>0.29</td>
<td>-0.45</td>
<td>-2.63</td>
<td>0.50</td>
</tr>
<tr>
<td>CER</td>
<td>1.0140</td>
<td>1.0148</td>
<td>1.0116</td>
<td>1.0052</td>
<td>1.0097</td>
<td>1.0071</td>
<td>1.0073</td>
<td>1.0127</td>
<td>1.0130</td>
<td>1.0086</td>
<td>1.0112</td>
<td>1.0093</td>
<td>1.0039</td>
<td>1.0117</td>
</tr>
</tbody>
</table>

### Panel B: Model Combination Across Different Estimators $\hat{Y}_{t+1}$

<table>
<thead>
<tr>
<th></th>
<th>$\Delta U(%)$</th>
<th>CER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pool Average</td>
<td>0.72</td>
<td>1.0122</td>
</tr>
<tr>
<td>Pool Median</td>
<td>0.38</td>
<td>1.0114</td>
</tr>
<tr>
<td>Pool DMSFE</td>
<td>1.51</td>
<td>1.0142</td>
</tr>
<tr>
<td>PAC</td>
<td>-1.07</td>
<td>1.0078</td>
</tr>
</tbody>
</table>

### Panel C: Quantile LASSO + Quantile Combination $\hat{Y}_{t+1}$

<table>
<thead>
<tr>
<th></th>
<th>$\Delta U(%)$</th>
<th>CER</th>
</tr>
</thead>
<tbody>
<tr>
<td>LASSO+Regression</td>
<td>2.89</td>
<td>1.0177</td>
</tr>
<tr>
<td>LASSO+DFMSE(1)</td>
<td>1.19</td>
<td>1.0134</td>
</tr>
<tr>
<td>LASSO+DFMSE(0.9)</td>
<td>1.83</td>
<td>1.0150</td>
</tr>
</tbody>
</table>

### Panel D: Quantile Combination Firstly + Model Combination Afterwards

<table>
<thead>
<tr>
<th></th>
<th>$\Delta U(%)$</th>
<th>CER</th>
</tr>
</thead>
<tbody>
<tr>
<td>OWQ+Pool Average</td>
<td>2.02</td>
<td>1.0155</td>
</tr>
<tr>
<td>OWQ+Pool Median</td>
<td>1.09</td>
<td>1.0132</td>
</tr>
<tr>
<td>OWQ+Pool DMSFE</td>
<td>2.55</td>
<td>1.0168</td>
</tr>
<tr>
<td>OWQ+PAC</td>
<td>0.77</td>
<td>1.0124</td>
</tr>
<tr>
<td>Density+Pool Average</td>
<td>1.52</td>
<td>1.0142</td>
</tr>
<tr>
<td>Density+Pool Median</td>
<td>0.54</td>
<td>1.0118</td>
</tr>
<tr>
<td>Density+Pool DMSFE</td>
<td>1.98</td>
<td>1.0154</td>
</tr>
<tr>
<td>Density+PAC</td>
<td>-0.61</td>
<td>1.0089</td>
</tr>
</tbody>
</table>

This table presents the economic evaluation of return forecasts under Mean Variance utility preference. The absolute risk aversion coefficient $\gamma = 5$. Average utility gain $\Delta U(\%)$ can be considered as the portfolio management fee (in annualized percent return) that an investor would be willing to pay to have access to the conditional return forecast relative to historical average benchmark forecast. CER refers to certainty equivalent rate of return that provides the same utility level as risky portfolio. Panel B reports the utility gain for an investor under mean variance preference who makes portfolio choice following the guidance of the return forecast based on only model combination. Panel C reports the utility gain for an investors under mean variance preference who makes portfolio choice following the guidance of the return forecast based on quantile combination followed by model combination. The out of sample evaluation period starts from 1957 Q1 to 2010 Q4. The acronyms OWQ denotes the Optimal Weight quantile combination approach developed by Xiao and Zhao (2011). Density denotes the density weight quantile combination discussed before.
Table 2.9: Predictive Density Comparison

<table>
<thead>
<tr>
<th>Panel A: Individual Density Forecast</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AG stat</th>
<th>DP</th>
<th>DY</th>
<th>EP</th>
<th>DE</th>
<th>SVAR</th>
<th>BM</th>
<th>NTIS</th>
<th>TBL</th>
<th>LTY</th>
<th>LTR</th>
<th>TMS</th>
<th>DFY</th>
<th>DFR</th>
<th>INF L</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>-3.48</td>
<td>-2.45</td>
<td>-0.99</td>
<td>-2.66</td>
<td>-0.53</td>
<td>-1.39</td>
<td>-0.08</td>
<td>-2.94</td>
<td>-1.90</td>
<td>-0.82</td>
<td>-2.59</td>
<td>-0.32</td>
<td>-0.86</td>
<td>-0.43</td>
</tr>
<tr>
<td>P-value</td>
<td>0.00</td>
<td>0.01</td>
<td>0.16</td>
<td>0.00</td>
<td>0.30</td>
<td>0.08</td>
<td>0.47</td>
<td>0.00</td>
<td>0.03</td>
<td>0.21</td>
<td>0.00</td>
<td>0.37</td>
<td>0.19</td>
<td>0.33</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel B: Model Combination + Kernel Density Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pool Average + Kernel Density</td>
</tr>
<tr>
<td>Pool Median + Kernel Density</td>
</tr>
<tr>
<td>Pool DMSFE + Kernel Density</td>
</tr>
<tr>
<td>PAC + Kernel Density</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C: Kernel Density Estimation + Density Combination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kernel Density + Likelihood Weight</td>
</tr>
<tr>
<td>Kernel Density + Pool Average</td>
</tr>
</tbody>
</table>

This table reports the Amisano and Giacomini 2007 (AG) test results for the individual models and combination models. Panel B reports the result for the estimation strategy of model combination on each quantile estimation firstly followed by kernel density estimation afterwards. Panel C reverts the combination order, which present results for kernel estimation firstly followed by density combination. Likelihood Weight refers to the weight average of different kernel density via maximizing the sum value of log likelihood function during forecast evaluation period. PCA: Principal Component Analysis. (.) is the asymptotical P-value. The symbols *, **, *** denote significance at the 10%, 5% and 1% levels, respectively.
Table 2.10: Utility Gain Under CRRA Utility Preference ($\gamma = 5$)

### Panel A: Individual Model

<table>
<thead>
<tr>
<th></th>
<th>DP</th>
<th>DY</th>
<th>EP</th>
<th>DE</th>
<th>SVAR</th>
<th>BM</th>
<th>NTIS</th>
<th>TBL</th>
<th>LTY</th>
<th>LTR</th>
<th>TMS</th>
<th>DFY</th>
<th>DFR</th>
<th>INFL</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta U(%)$</td>
<td>0.33</td>
<td>0.05</td>
<td>-0.39</td>
<td>-1.46</td>
<td>-0.07</td>
<td>0.07</td>
<td>-0.27</td>
<td>0.03</td>
<td>-0.20</td>
<td>-0.17</td>
<td>-0.18</td>
<td>-0.64</td>
<td>-0.18</td>
<td>0.12</td>
</tr>
<tr>
<td>CER</td>
<td>0.0158</td>
<td>0.0151</td>
<td>0.0139</td>
<td>0.0110</td>
<td>0.0147</td>
<td>0.0151</td>
<td>0.0142</td>
<td>0.0150</td>
<td>0.0144</td>
<td>0.0145</td>
<td>0.0144</td>
<td>0.0132</td>
<td>0.0144</td>
<td>0.0153</td>
</tr>
</tbody>
</table>

### Panel B: Model Combination + Kernel Density Estimation

<table>
<thead>
<tr>
<th></th>
<th>$\Delta U(%)$</th>
<th>CER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pool Average + Kernel Density</td>
<td>0.13</td>
<td>0.0153</td>
</tr>
<tr>
<td>Pool Median + Kernel Density</td>
<td>0.10</td>
<td>0.0152</td>
</tr>
<tr>
<td>Pool DMSFE + Kernel Density</td>
<td>0.12</td>
<td>0.0153</td>
</tr>
<tr>
<td>PAC + Kernel Density</td>
<td>0.15</td>
<td>0.0153</td>
</tr>
<tr>
<td>Benchmark Kernel Density</td>
<td>0</td>
<td>0.0149</td>
</tr>
</tbody>
</table>

### Panel C: Kernel Density Estimation + Density Combination

<table>
<thead>
<tr>
<th></th>
<th>$\Delta U(%)$</th>
<th>CER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kernel Density + Likelihood Weight</td>
<td>0.25</td>
<td>0.0156</td>
</tr>
<tr>
<td>Kernel Density + Pool Average</td>
<td>0.19</td>
<td>0.0154</td>
</tr>
<tr>
<td>Benchmark Kernel Density</td>
<td>0</td>
<td>0.0149</td>
</tr>
</tbody>
</table>

This table presents the economic evaluation of a wide range of density forecasts under relative risk aversion preference (CRRA). Average utility gain $\Delta U(\%)$ can be considered as the portfolio management fee that an investor would be willing to pay to have access to the conditional kernel density forecast relative to unconditional benchmark forecast. $CER$ refers to certainty equivalent rate of return that provides the same utility level as risky portfolio. Panel B reports the result of economic evaluation for the estimation strategy of model combination in the first step on each quantile estimation, followed by kernel density estimation in the second step. Panel C reverts the combination order, which present results for kernel estimation firstly followed by density combination in the second step. **Likelihood Weight** refers to the weight average of different kernel density via maximizing the sum value of log likelihood value during forecast evaluation period. **PCA**: Principal Component Analysis. The out of sample evaluation period starts from 1957 Q1 to 2010 Q4.
Bibliography


