THE FLORIDA STATE UNIVERSITY

COLLEGE OF ARTS AND SCIENCES

TWO ESSAYS IN FINANCE AND STATISTICS: BUILDING TRACKING PORTFOLIOS AND DESIGNING RISK MANAGEMENT PROGRAMS

By

SHAOJUN ZHANG

A Dissertation submitted to the
Department of Statistics
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Degree Awarded:
Spring Semester, 2001

UMI Number: 3005637

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

The members of the Committee approve the dissertation of Shaojun Zhang defended on April 4, 2001.

James S. Ang
Professor Co-Directing Thesis

Xufeng Niu
Professor Co-Directing Thesis

Paul Beaumont
Outside Committee Member

Jayaram Sethuraman
Committee Member

Fred W. Huffer
Committee Member

To my wife Ya Wang and daughter Cindy Zhang.

ACKNOWLEDGEMENTS

I am deeply indebted to my major professors, Dr. James Ang and Dr. Xufeng Niu, for their knowledge, guidance, and support. I would not have been able to complete this work without numerous discussions with them. I am very grateful to my committee members, Dr. Paul Beaumont, Dr. Jayaram Sethuraman, and Dr. Fred Huffer, for their insightful and constructive comments.

I owe much to the Statistics Department of the Florida State University for providing a conducive atmosphere for study and research under the leadership of Dr. Ian McKeague and Dr. Myles Hollander. Over the last five years many people inside and outside the department have shared their knowledge and lent their hands to me. A complete list of their names would stretch a long way. I sincerely thank all of them. In particular, I thank Dr. Pi-Erh Lin for clearing the path for me to pursue interdisciplinary research. Special thanks to Dr. Douglas Zahn, Dr. Lei Li, Dr. Kai-Sheng Song, Karen Kinard, Dagang Wang, Tom Jagger, Marc Loizeaux, and Bin Yu for many interesting conversations. I would like to thank the staff of the department, Pam McGhee, Virginia Grice, Betty Stokes, and James Stricherz, for their help and excellent work. I also wish to acknowledge the financial support from the Florida State University with the University Fellowship.

I thank my parents, brother and sisters. My graduate study would not have been possible without them.

Finally, I thank my wife, Ya Wang for her love, support, and encouragement throughout all the years.

TABLE OF CONTENTS

List of Tables vi

List of Figures vii

Abstract viii

1. BUILDING TRACKING PORTFOLIOS BASED ON A GENERALIZED INFORMATION CRITERION **1**

1.1 Introduction 1
1.2 Statistical Model 4
1.3 The Generalized Information Criterion 7
1.3.1 Independent Observations 8
1.3.2 Dependent Observations 10
1.4 Simulation 13
1.4.1 Data and Setup 13
1.4.2 Results 15
1.5 Application 19
1.5.1 Estimates of Abnormal Return 19
1.5.2 Empirical Assessment of Performance 22
1.6 Proofs 24
1.7 Summary and Discussion 33

2. RISK MANAGEMENT, DISPERSED INFORMATION, AND INTERNAL MARKET **35**

2.1 Introduction 35
2.2 Problem Description 37
2.3 Theoretical Model and Analysis 40
2.3.1 Ideal Centralization 41
2.3.2 Uncoordinated Decentralization 47
2.3.3 Real-World Centralization 51
2.3.4 Coordinated Decentralization 54
2.4 Summary and Discussion 59

REFERENCES **61**

BIOGRAPHICAL SKETCH **64**

LIST OF TABLES

1.1 Simulation with observed returns 17

1.2 Simulation with simulated returns 18

1.3 A sample of 200 randomly selected firms. 24

LIST OF FIGURES

1.1 Time series plots and sample autocorrelation functions 14

ABSTRACT

This dissertation contains two essays. The first essay concerns how to build a tracking portfolio of stocks whose return of investment mimics that of a chosen investment target. Statistically, this task can be accomplished by selecting an optimal model from constrained linear models. To develop an automatic procedure for building an optimal tracking portfolio, we extend the Generalized Information Criterion (GIC) to constrained linear models either with independently and identically distributed random errors or with dependent errors that follow a stationary Gaussian process. The asymptotic validity of the extended GIC is established. Simulation results show that the relative frequency of selecting the optimal constrained linear model by the GIC is close to one in finite samples. We apply the GIC based procedure for building an optimal tracking portfolio to the problem of measuring the long-term impact of a corporate event on stock returns and demonstrate empirically that it outperforms two other competing methods.

The second essay concerns how corporations organize their risk management program. We set up a theoretical framework to analyze the problem of designing a risk management program for multidivisional corporations. Our analysis shows that risk management programs currently existing in large corporations are not optimal. We propose a new risk management program in which the corporate headquarters organizes an internal market for divisions to trade state-contingent claims among themselves. We show that this new program is better than existing ones.

CHAPTER 1

BUILDING TRACKING PORTFOLIOS BASED ON A GENERALIZED INFORMATION CRITERION

1.1 Introduction

In a number of financial research studies, the essential task boils down to building a tracking portfolio of stocks whose return on investment mimics that of a chosen investment target. The following are two typical examples of such studies.

Example 1.1 Index Fund. The sole business of an index fund is to maintain a portfolio of individual stocks such that percentage changes in the value of the portfolio are approximately equal to those of the chosen index. Success of an index fund depends on how closely its portfolio mimics the index. An easy way to track the index is to buy and hold all constituent stocks of the index in the same proportions as they compose the index. It is the easy approach that every index fund manager would like to take. However, managers of open-end index funds are frequently forced to buy or sell stocks as investors deposit or withdraw their money. They simply can not constantly hold the same stocks in the same proportion. When they have to adjust their portfolio, they would like to make their new portfolio be one that can closely track the index.

Example 1.2 Long-Term Impact of a Corporate Event on Stock Returns. Such studies focus on the effect of a specific corporate event on return of investment in the event firm's stock over a time span of several years after the event has happened. A great interest in learning the long-term impact of corporate actions has recently arisen among finance researchers and generated a considerable literature that is still growing. See Fama (1998) for a summary of the literature and references therein. Since the observed post-event return of the event firm's stock has been influenced by the event, we do not know what the *status quo* return would be if the event had not happened. But in order to learn the event's impact, we have to compare the observed post-event return against the unobservable *status quo* return. One way to estimate the *status quo* return is to build a portfolio of other stocks whose return has moved in the same way as the event firm's return before the event happened and to use the observed post-event return of the portfolio as an estimate. Once we have a pre event tracking portfolio, the difference between the observed post-event return of the event firm and that of the tracking portfolio is a measure of the event's effect.

In each of the two examples, a tracking portfolio is to be built given a desired target and a group of other stocks. In fact, every nonempty subset of these stocks can form a portfolio that may track the target well. There are as many possible tracking portfolios as the number of nonempty subsets of these stocks. Among all possible portfolios, an ideal tracking portfolio would be such that its return is equal to the target's return in every month. (We use month as the time unit for measuring investment return in this paper.) In reality, any portfolio will have returns different from those of the target. An optimal, though not ideal, tracking portfolio will be the one whose returns are on average closest to the target's returns.

Since the return on a tracking portfolio is a weighted sum of returns on all stocks in the portfolio, building a tracking portfolio for one nonempty subset of stocks is equivalent to fitting a constrained linear model with the target's return as the response variable and returns on stocks in the subset as the covariates. The linear model is constrained in that all coefficients in the model sum up to one. This is because the coefficient of each covariate is the proportion of investment on the corresponding stock to the total investment on the tracking portfolio and the sum of all coefficients accounts for 100 percent of the total investment.

Because of the correspondence of each possible portfolio and a constrained linear model, the task of finding the optimal tracking portfolio can be accomplished by selecting an optimal constrained linear model. In this paper, we develop an automatic procedure to find an optimal tracking portfolio, based on a model selection criterion known as the Generalized Information Criterion (GIC). There is considerable literature on the problem of selecting variables in the context of unconstrained linear models; see review papers by Hocking (1976) and Thompson (1978a, b) for early contributions. Miller (1990) gives an excellent and comprehensive treatment of variable selection methods prior to 1990, and George (2000) reviews the key developments in the last decade. The Generalized Information Criterion (GIC) we use is proposed by Rao and Wu (1989) and is a generalization of the well known Akaike's Information Criterion (AIC, Akaike (1973)) and the Bayesian Information Criterion (BIC, Schwartz (1978)). Nishii (1984) studies asymptotic properties of several selection criteria, one of which is asymptotically equivalent to the one by Rao and Wu (1989). Nishii (1984), Rao and Wu (1989), and Pötscher (1989) prove the consistency of the GIC or its asymptotic equivalents for unconstrained linear models under the assumption that there exists a finite-dimensional true model. However, they made different assumptions on the random errors in the linear models. Nishii (1984) and Rao and Wu (1989) assume i.i.d. random errors while Pötscher (1989) assumes that the errors follow a martingale difference sequence. Shao (1997) proposes using two criteria to evaluate asymptotic validity of a model selection procedure: consistency and asymptotic loss efficiency. (See Section 1.3 for definitions of asymptotical loss efficiency and consistency.) He shows that the GIC is asymptotically loss efficient regardless the existence of a true model and consistent if a true model exists. He deals with only unconstrained linear models and assume that the random errors are i.i.d.

In this paper, we extend the Generalized Information Criterion (GIC) first to constrained linear models with i.i.d. random errors and then to constrained linear models with errors following a stationary Gaussian process. Since there is no guarantee that returns on the target are completely determined by returns on any subset of stocks, we do not take the existence of a true model for granted. Following Shao (1997), we study both asymptotic loss efficiency and consistency of the extended GIC. We prove that, under certain conditions, the extended GIC is asymptotically loss efficient regardless the existence of a true model and consistent if a true model exists.

As an application, we apply the GIC to build an optimal tracking portfolio for the purpose of measuring the long-term impact of a corporate event on stock returns. We compare performance of the GIC based procedure against two other competing methods empirically and find that the GIC based procedure gives the best results.

The rest of the paper is organized as follows. In Section 1.2, we formalize a statistical model for building an optimal tracking portfolio. In Section 1.3, we introduce both the Generalized Information Criterion (GIC) and the extended Generalized Information Criterion (EGIC) and study their asymptotic properties. Section 1.4 reports results from a simulation study. We then apply GIC to solve the problem of measuring long-term post-event abnormal return in Section 1.5. We conclude the paper with summary and discussion in Section 1.7. Proofs of theorems in this paper are given in the Appendix.

1.2 Statistical Model

Let y_t be the return of investing in a chosen target during time period t, that is,

$$y_t = \frac{\text{Target's price at the end of } t - \text{Target's price at the end of } t-1}{\text{Target's price at the end of } t-1}.$$

Let $\boldsymbol{y} = (y_1, \cdots, y_\tau)'$ be a vector of returns for $t = 1, 2, \cdots, \tau$, and $\boldsymbol{y}$ has the following representation:

$$\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{e} \ , \tag{1.1}$$

where $\boldsymbol{\mu} = E(\boldsymbol{y})$ is the mean of $\boldsymbol{y}$ and $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{\mu}$ is a vector of random variables with mean zero. Note that we allow the expected return of the target to change with time in any way.

Suppose that m other stocks are available for building a tracking portfolio of the target. Let $\boldsymbol{X} = (\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m)$ be a $\tau \times m$ matrix of rank m, where column $\boldsymbol{x}_j$ is a vector of returns on the jth stock for $t = 1, 2, \ldots, \tau$. Let V be the collection of all nonempty subsets of $\{1, 2, \cdots, m\}$. Each subset $v \in V$ indexes a group of stocks. Let $\boldsymbol{X}(v)$ be the submatrix of $\boldsymbol{X}$ whose columns are returns on stocks in the subset v. To build a tracking portfolio consisting of all stocks in the subset v, we fit the following linear model of $\boldsymbol{y}$ against the matrix $\boldsymbol{X}(v)$,

$$\boldsymbol{y} = \boldsymbol{X}(v)\boldsymbol{\beta}(v) + \boldsymbol{e}(v) \ , \tag{1.2}$$

where the dimension of $\boldsymbol{\beta}(v)$ is equal to the size of the subset v. Note that the error component $\boldsymbol{e}(v)$ depends on v and differs from the random vector $\boldsymbol{e}$ in (1.1). More specifically, the error component $\boldsymbol{e}(v)$ is the sum of the random vector $\boldsymbol{e}$ in (1.1) and the model misspecification error.

Let $\hat{\boldsymbol{\beta}}(v)$ denote an estimate of $\boldsymbol{\beta}(v)$. Then an estimate of the expected value $\boldsymbol{\mu} = E(\boldsymbol{y})$ is given by $\hat{\boldsymbol{\mu}}(v) = \boldsymbol{X}(v)\hat{\boldsymbol{\beta}}(v)$. The goodness of the estimate is measured by the ***average squared error loss***

$$L(v) = \frac{||\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}(v)||^2}{\tau} \ , \tag{1.3}$$

where $||\cdot||$ is the Euclidean norm. The objective of model selection is to find the subset v whose associated estimate $\hat{\mu}(v)$ minimizes the average squared error loss. Once the minimizing subset v is found, the portfolio that consists of all stocks in the subset v with $\hat{\boldsymbol{\beta}}(v)$ as the portfolio weights will be the optimal tracking portfolio.

In the context of building a tracking portfolio, the coefficients $\boldsymbol{\beta}(v)$ in the linear model (1.2) have to sum up to one. The left hand side of (1.2) is the return of investing one dollar in the target. The right hand side of (1.2) is the return of investing one dollar in the portfolio of stocks in the subset v, plus random noise. Each coefficient in $\boldsymbol{\beta}(v)$ is the proportion of a dollar invested in the corresponding stock, and the sum of all coefficients accounts for 100 percent of the dollar.

Give a subset v, to estimate the linearly constrained coefficients, we take the model reduction approach given in Hocking (1985, Chapter 3). We briefly describe the approach in the remainder of this section for the sake of the self-completeness of this paper. To simplify notation, we drop the subset index v temporarily.

A linear model with general linear constraints is characterized by

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}\ , \qquad (1.4)$$

$$\text{subject to} \quad \boldsymbol{G}\boldsymbol{\beta} = \boldsymbol{g}\ ,$$

where $\boldsymbol{y}$ is a vector of dimension τ, $\boldsymbol{X}$ is a $\tau \times m$ matrix of rank m, $\boldsymbol{\beta}$ is a vector of m coefficients, $\boldsymbol{G}$ is a $q \times m$ matrix of rank q, and $\boldsymbol{\epsilon}$ is a random vector.

An estimate of $\boldsymbol{\beta}$ can be obtained by the model reduction approach as follows. The coefficient vector $\boldsymbol{\beta}$ and the constraint matrix $\boldsymbol{G}$ are partitioned in the way such that the constraints are written as

$$\boldsymbol{G}_1\boldsymbol{\beta}_1 + \boldsymbol{G}_2\boldsymbol{\beta}_2 = \boldsymbol{g}\ ,$$

where $\boldsymbol{G}_1$ is a $q \times q$ matrix of rank q. Solving for $\boldsymbol{\beta}_1$ yields

$$\boldsymbol{\beta}_1 = \boldsymbol{G}_1^{-1}\boldsymbol{g} - \boldsymbol{G}_1^{-1}\boldsymbol{G}_2\boldsymbol{\beta}_2\ . \qquad (1.5)$$

Corresponding to the partition of β, we partition $\boldsymbol{X}$ as $\boldsymbol{X} = (\boldsymbol{X}_1\ \boldsymbol{X}_2)$, where $\boldsymbol{X}_1$ is a $\tau \times q$ matrix. Substituting the partition into the constrained model, we obtain the following unconstrained model

$$\boldsymbol{y}_R = \boldsymbol{X}_R\boldsymbol{\beta}_2 + \boldsymbol{\epsilon}\ .$$

where $\boldsymbol{y}_R = \boldsymbol{y} - \boldsymbol{X}_1\boldsymbol{G}_1^{-1}g$ and $\boldsymbol{X}_R = \boldsymbol{X}_2 - \boldsymbol{X}_1\boldsymbol{G}_1^{-1}\boldsymbol{G}_2$. The least square estimates for $\boldsymbol{\beta}_2$ is then given by

$$\hat{\boldsymbol{\beta}}_2 = (\boldsymbol{X'}_R\boldsymbol{X}_R)^{-1}\boldsymbol{X'}_R\boldsymbol{y}_R \ .$$

Substituting $\hat{\boldsymbol{\beta}}_2$ in (1.5), we get $\hat{\boldsymbol{\beta}}_1 = \boldsymbol{G}_1^{-1}\boldsymbol{g} - \boldsymbol{G}_1^{-1}\boldsymbol{G}_2\hat{\boldsymbol{\beta}}_2$. We can write the estimator for $\boldsymbol{\beta}$ together as

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{G}_1^{-1}\boldsymbol{g} \\ 0 \end{bmatrix} + \begin{bmatrix} -\boldsymbol{G}_1^{-1}\boldsymbol{G}_2 \\ \boldsymbol{I} \end{bmatrix} \hat{\boldsymbol{\beta}}_2 \ . \tag{1.6}$$

Then the estimate of $\boldsymbol{\mu} = E(\boldsymbol{y})$ is given by

$$\hat{\boldsymbol{\mu}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{\eta} + \boldsymbol{H}\boldsymbol{y} \ . \tag{1.7}$$

where $\boldsymbol{H} = \boldsymbol{X}_R(\boldsymbol{X'}_R\boldsymbol{X}_R)^{-1}\boldsymbol{X'}_R$ and $\boldsymbol{\eta} = (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{X}_1\boldsymbol{G}_1^{-1}\boldsymbol{g}$.

1.3 The Generalized Information Criterion

In the context of unconstrained linear models, numerous criteria have been proposed to select variables, see, e.g., Hocking (1976), Thompson (1978a, b), Miller (1990), George (2000), and references therein. In Section 1.3.1, we discuss the Generalized Information Criterion (GIC) for selecting variables in constrained linear models when observations are independent, and prove the asymptotic loss efficiency and consistency of the GIC. In Section 1.3.2, we extend the GIC to constrained linear models with dependent observations, and give conditions under which the extended GIC is still asymptotically loss efficient and consistent.

We consider constrained linear models of the following form:

$$\boldsymbol{y}_\tau = \boldsymbol{X}_\tau(v)\boldsymbol{\beta}(v) + \boldsymbol{e}_\tau(v) \ , \tag{1.8}$$

$$\text{subject to} \qquad \boldsymbol{l}'\boldsymbol{\beta}(v) = 1 \ ,$$

where v belongs to V, the collection of all possible nonempty subsets of the m covariates, $\boldsymbol{l}$ is a vector of ones with the same dimension as $\boldsymbol{\beta}(v)$, and τ is the number of observed time periods. Note that the coefficients in $\boldsymbol{\beta}(v)$ vary with the subset v but not with τ. In other words, we assume that the linear relation between $\boldsymbol{y}_\tau$ and $\boldsymbol{X}_\tau(v)$ is fixed through time. Let $\boldsymbol{\mu}_\tau = E(\boldsymbol{y}_\tau)$. A candidate model $v \in V$ is said to be ***correct*** if there exists $\boldsymbol{\beta}(v)$ such that $\boldsymbol{\mu}_\tau = \boldsymbol{X}_\tau(v)\boldsymbol{\beta}(v)$ for all τ. Let V^c be the collection of all candidate models that are correct.

In the remainder of this paper, the subset index $v \in V$ is attached to quantities that depend on choice of the subset v, and the subscript τ is used to indicate quantities that vary with τ.

1.3.1 Independent Observations

Recall the following identity representation of $\boldsymbol{y}_\tau$, first introduced in equation (1.1),

$$\boldsymbol{y}_\tau = \boldsymbol{\mu}_\tau + \boldsymbol{e}_\tau .$$

In this subsection, we assume that the elements of $\boldsymbol{e}_\tau$ are independently and identically distributed with a normal distribution of mean 0 and variance σ^2.

The ***Generalized Information Criterion (GIC)*** selects a model in the form (1.8) that minimizes

$$\Gamma_\tau(v) = \frac{||\boldsymbol{y}_\tau - \hat{\boldsymbol{\mu}}_\tau(v)||^2}{\tau} + \frac{\lambda_\tau \hat{\sigma}_\tau^2 tr(\boldsymbol{H}_\tau(v))}{\tau} \tag{1.9}$$

over $v \in V$, where $\hat{\sigma}_\tau^2$ is an estimator of σ^2, $tr(\boldsymbol{H}_\tau(v))$ is the trace of the matrix $\boldsymbol{H}_\tau(v)$ introduced in equation (1.7), and λ_τ is a sequence of non-random positive numbers. Note that $\hat{\sigma}_\tau^2$ does not depend on the model v and can be obtained by fitting the linear model with all m covariates included. Note also that $tr(\boldsymbol{H}_\tau(v))$ is equal to the number of unconstrained coefficients in the linear model (1.8).

Let $\hat{v}_\tau$ denote the subset that minimizes the Generalized Information Criterion (GIC), $\Gamma_\tau(v)$, over $v \in V$. Let v_τ^L be the subset that minimizes the average squared error loss, $L_\tau(v)$, over $v \in V$. Shao (1997) studies the asymptotic validity of a model selection procedure in terms of two criteria: consistency and asymptotic loss efficiency. The GIC selection procedure is said to be ***consistent*** if

$$P\{\hat{v}_\tau = v_\tau^L\} \to 1 \quad \text{as} \quad \tau \to \infty \ ,$$

and to be ***asymptotically loss efficient*** if

$$L_\tau(\hat{v}_\tau)/L_\tau(v_\tau^L) \xrightarrow{p} 1 \quad \text{as} \quad \tau \to \infty \ ,$$

where $\xrightarrow{p}$ denotes convergence in probability. Throughout this paper, all limiting processes are taken as $\tau \to \infty$.

The following lemma gives explicit expressions for the average squared error loss $L_\tau(v)$ and the expected average squared error loss $R_\tau(v) \equiv E(L_\tau(v))$.

Lemma 1.1 *Assume that the elements of $\boldsymbol{e}_\tau$ are i.i.d. with a normal distribution of mean 0 and variance σ^2. The average squared error loss defined in (1.3) is equal to*

$$L_\tau(v) = \Delta_\tau(v) + (\boldsymbol{e}_\tau' \boldsymbol{H}_\tau(v) \boldsymbol{e}_\tau)/\tau \ ,$$

where $\Delta_\tau(v) = (||\boldsymbol{\mu}_\tau - \boldsymbol{\eta}_\tau(v) - \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau||^2)/\tau$ and $\boldsymbol{e}_\tau \sim N(0, \sigma^2 I_{\tau\times\tau})$ is the vector of random variables in (1.1). Furthermore, $\Delta_\tau(v) = 0$ for $v \in V^c$. In addition, the expected average squared error loss, is

$$R_\tau(v) \equiv E(L_\tau(v)) = \Delta_\tau(v) + \sigma^2 tr(\boldsymbol{H}_\tau(v))/\tau \ .$$

Proof of Lemma 1.1 is given in the Appendix. Lemma 1.1 points out that the average squared error loss has two components: one is the model misspecification error $\Delta_\tau(v)$, and the other is estimation error $(\boldsymbol{e}_\tau' \boldsymbol{H}_\tau(v) \boldsymbol{e}_\tau))/\tau$ due to randomness in observed data. When the model v is correct, the model misspecification error $\Delta_\tau(v)$ is zero.

The following theorem shows that $\hat{v}_\tau$ is consistent and asymptotically loss efficient under certain conditions. The proof of the theorem is given in the Appendix.

Theorem 1.1 *Assume that the elements of $\mathbf{e}_\tau$ are i.i.d. with a normal distribution of mean 0 and variance σ^2 and the estimator $\hat{\sigma}_\tau^2$ used in computing the GIC, $\Gamma_\tau(v)$, is bounded. Under the following conditions*

$$\lambda_\tau \to \infty, \qquad \lambda_\tau/\tau \to 0 \ , \tag{1.10}$$

$$\text{and} \qquad \frac{\lambda_\tau}{\tau R_\tau(v)} \to 0 \quad \text{for all} \ \ v \in V - V^c \ , \tag{1.11}$$

the GIC minimizer $\hat{v}_\tau$ is asymptotically loss efficient. In addition, if V contains at least one correct model, then $\hat{v}_\tau$ is consistent.

Remark 1.1 Shao (1997) proves both the asymptotic loss efficiency and consistency for unconstrained linear regression models with i.i.d. random errors under condition (1.10) and the following condition (Theorem 2, Shao (1997))

$$\liminf_{\tau\to\infty} \min_{v\in V-V^c} \Delta_\tau(v) > 0 \ . \tag{1.12}$$

It is easy to see that conditions (1.10) and (1.12) together implies condition (1.11) because $R_\tau(v) \geq \Delta_\tau(v)$. Shao's condition (1.12) requires that the model misspecification error $\Delta_\tau(v)$ be bounded away from zero uniformally for all incorrect models. Our condition (1.11) suggests that, as long as the model misspecification error of incorrect models tends to 0 at a rate slower than $1/\tau$, the GIC minimizer $\hat{v}_\tau$ is still asymptotically valid.

1.3.2 Dependent Observations

In this subsection, we extend GIC to linearly constrained regression model with dependent observations. Specifically, we assume that $\{e_t\}_{t=-\infty}^{\infty}$ is a stationary Gaussian process with mean $E(e_t) = 0$ and $E(e_t e_{t+j}) = \gamma_j$. We further assume that the autocovariances $\{\gamma_j\}_{j=0}^{\infty}$ are absolutely summable, that is,

$$\Upsilon \equiv \gamma_0 + 2\sum_{j=1}^{\infty} |\gamma_j| < \infty \ . \tag{1.13}$$

Then the covariance matrix for $\boldsymbol{e}_\tau$, denoted as $\boldsymbol{\Psi}_\tau$, is given by

$$\boldsymbol{\Psi}_\tau = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{\tau-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{\tau-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{\tau-1} & \gamma_{\tau-2} & \cdots & \gamma_0 \end{bmatrix}. \tag{1.14}$$

The ***extended Generalized Information Criterion (EGIC)*** selects a model that minimizes

$$\Phi_\tau(v) = \frac{||\boldsymbol{y}_\tau - \hat{\boldsymbol{\mu}}_\tau(v)||^2}{\tau} + \frac{\lambda_\tau tr(\hat{\boldsymbol{\Psi}}_\tau \boldsymbol{H}_\tau(v))}{\tau} \tag{1.15}$$

over $v \in V$, where $\hat{\boldsymbol{\Psi}}_\tau$ is an estimate of $\boldsymbol{\Psi}_\tau$ and λ_τ is a sequence of non-random positive numbers. Note that $\hat{\boldsymbol{\Psi}}_\tau$ does not depend on model v and is obtained by fitting a linear model with all available explanatory variables included.

Let $\check{v}_\tau$ denote the model that minimizes the extended Generalized Information Criterion (EGIC), $\Phi_\tau(v)$, over $v \in V$. Let v_τ^L be the model that minimizes the average squared error loss, $L_\tau(v)$, over $v \in V$. The EGIC selection procedure is said to be ***consistent*** if $P\{\check{v}_\tau = v_\tau^L\} \to 1$. The EGIC selection procedure is said to be ***asymptotically loss efficient*** if $L_\tau(\check{v}_\tau)/L_\tau(v_\tau^L) \xrightarrow{p} 1$, where $\xrightarrow{p}$ denotes convergence in probability.

The following lemma gives explicit expressions for the average squared error loss $L_\tau(v)$ and the expected average squared error loss $R_\tau(v) \equiv E(L_\tau(v))$. Proof of the Lemma is given in the Appendix.

Lemma 1.2 *Assume that $\{e_t\}_{t=-\infty}^{\infty}$ is a stationary Gaussian process with $E(e_t) = 0$ and $\gamma_j = E(e_t e_{t+j})$. The average squared error loss defined in (1.3) is equal to*

$$L_\tau(v) = \Delta_\tau(v) + (\boldsymbol{e}_\tau' \boldsymbol{H}_\tau(v) \boldsymbol{e}_\tau)/\tau,$$

where $\Delta_\tau(v) = (||\boldsymbol{\mu}_\tau - \boldsymbol{\eta}_\tau(v) - \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau||^2)/\tau$. Furthermore, $\Delta_\tau(v) = 0$ when $v \in V^c$. In addition, the expected average squared error loss is given by

$$R_\tau(v) \equiv E(L_\tau(v)) = \Delta_\tau(v) + tr(\boldsymbol{\Psi}_\tau \boldsymbol{H}_\tau(v))/\tau\ ,$$

where the matrix $\boldsymbol{\Psi}_\tau$ is the covariance matrix given by (1.14).

The results on the consistency and asymptotically loss efficiency for $\check{v}_\tau$ are given in the following theorem. Proof of the theorem is in the Appendix.

Theorem 1.2 *Assume that $\{e_t\}_{t=-\infty}^{\infty}$ is a stationary Gaussian process with $E(e_t) = 0$, $\gamma_j = E(e_t e_{t+j})$, and $\Upsilon \equiv \gamma_0 + 2\sum_{j=1}^{\infty} |\gamma_j| < \infty$. We further assume that $\hat{\boldsymbol{\Psi}}_\tau$ used in computing the EGIC, $\Phi_\tau(v)$, is a consistent estimator of $\boldsymbol{\Psi}_\tau$ and that $tr(\boldsymbol{\Psi}_\tau \boldsymbol{H}_\tau(v))$ converges to a finite limit as $\tau \to \infty$ for any $v \in V^c$. Under the following conditions that*

$$\lambda_\tau \to \infty \ , \qquad \lambda_\tau/\tau \to 0 \ , \tag{1.16}$$

and

$$\frac{\lambda_\tau}{\tau R_\tau(v)} \to 0 \quad \text{for all} \ \ v \in V - V^c \ , \tag{1.17}$$

the EGIC minimizer $\check{v}_\tau$ is asymptotically loss efficient. In addition, if V contains at least one correct model, then $\check{v}_\tau$ is consistent.

Remark 1.2 Theorem 1.2 does not include Theorem 1.1 as a special case. Even though Theorem 1.1 deals with a special case $\boldsymbol{\Psi}_\tau = \sigma^2 \boldsymbol{I}_\tau$, it puts less restriction on the estimator of σ^2, requiring the estimator of σ^2 is bounded rather than consistent.

The following corollary shows that some common stochastic processes are included in Theorem 1.2. The proof of the corollary is simple and omitted.

Corollary 1.1 *Theorem 1.2 is valid when $\{e_t\}_{t=-\infty}^{\infty}$ is an infinite moving average Gaussian process given by $e_t = \sum_{j=0}^{\infty} \psi_j a_{t-j}$, where $a_t \sim i.i.d. N(0, \sigma_a^2)$ and $\sum_{j=0}^{\infty} |\psi_j| < \infty$.*

Remark 1.3 The Gaussian infinite moving average process specified in Corollary 1.1 includes stationary Gaussian processes AR(p), MA(q) and ARMA(p, q) as special cases.

1.4 Simulation

We carry out the following simulation study for two purposes: first, to empirically check the validity of GIC's optimal properties, and second, to understand how choice of the penalty λ_n affects GIC's performance in finite samples.

1.4.1 Data and Setup

We use historical stock returns in the simulation study. We randomly selected five stocks and extracted their monthly returns for 96 months between January 1981 and December 1988 from the database distributed by the Center of Research in Securities Prices (CRSP). The selected five stocks are Wal Mart Stores Inc. (WMT), Dayton Hudson Corp. (DH), Mac Frugals Bargains Close Outs (MFB), Service Merchandise Inc. (SM), and Family Dollar Stores Inc (FDS). Figure 1.1 shows the time series plots on the left column and the sample autocorrelation function plots on the right column of the monthly returns of the five stocks. Since the sample autocorrelation function plots show that these monthly stock returns are not autocorrelated, we use GIC instead of EGIC in this simulation study.

The monthly returns of the five stocks, denoted by $\{x_{1t}, \ldots, x_{5t}\}$, are used as independent variables in the following regression model

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + \beta_5 x_{5t} + \epsilon_t \ , \quad t = 1, \cdots, \tau$$

$$\text{subject to} \quad \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 = 1 \ .$$

Fixing the coefficients at (0.3, 0, 0, 0.4, 0.3) throughout the simulation, we simulate the response variable y_t by generating random errors ϵ_t from the normal distribution $N(0, \sigma^2)$. We choose the standard deviation σ equal to 0.0385 throughout the simulation. The number 0.0385 is the sample standard deviation of the 96 monthly returns on the CRSP value weighted market index between January 1981 and

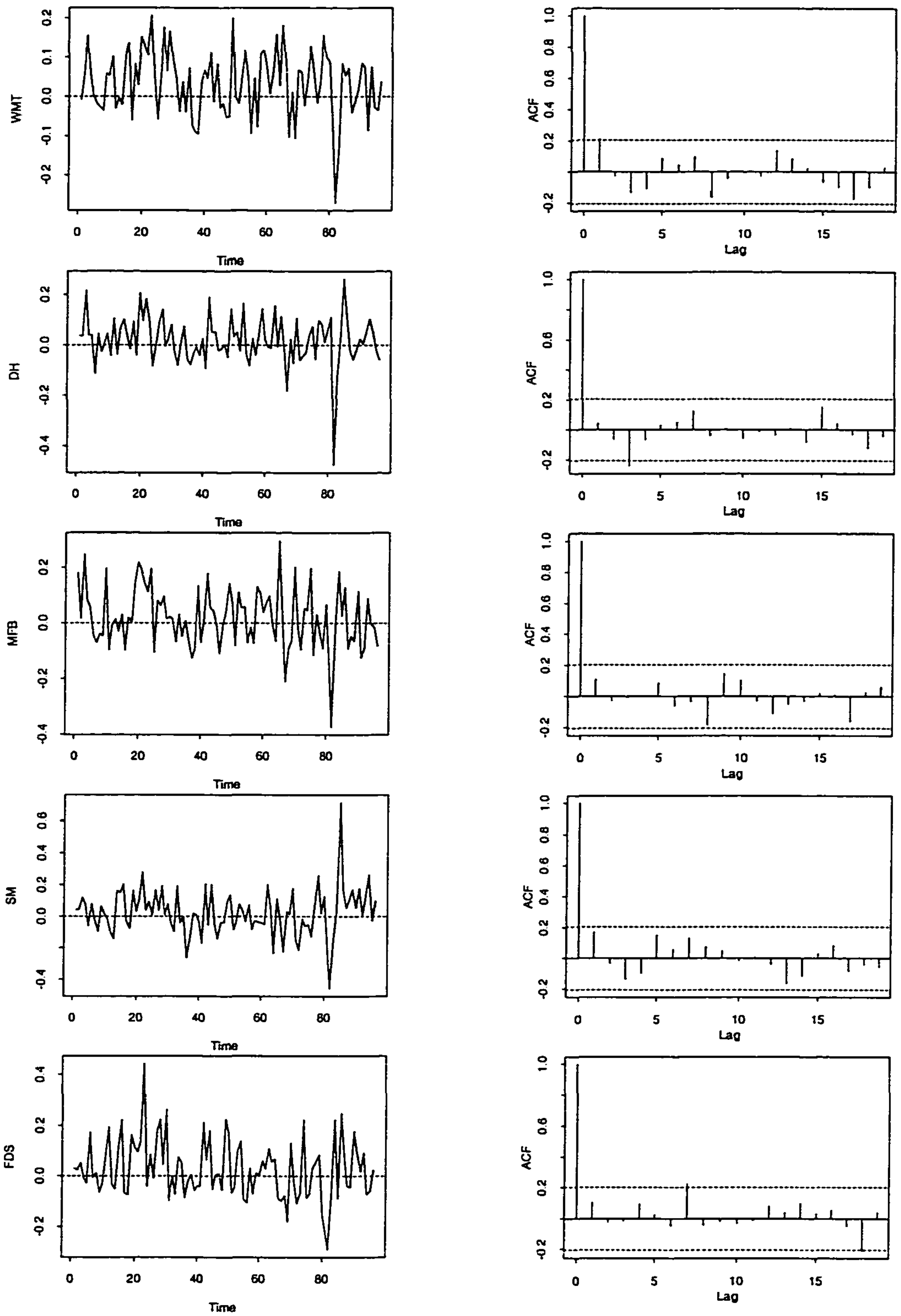

Figure 1.1. Time series plots and sample autocorrelation functions

December 1988. Returns on the value weighed market index are also extracted from the CRSP database.

Two sets of simulations are carried out in this study. For the first set, we chose the number of observations τ to be 36, 60, and 96. Two common choices of the penalty λ_τ are employed: one is the $\lambda_\tau = \log \tau$ and the other is $\lambda_\tau = \sqrt{\tau}$. Given τ and λ_τ, we generate 1,000 samples of response variables, each sample of τ observations. For each sample, we compute the Generalized Information Criterion (GIC), $\Gamma_\tau(v)$, for all $31(= 2^5 - 1)$ subsets of explanatory variables, and identify the subset minimizing GIC. We count how many times each subset minimizes GIC out of the 1,000 samples. The counts are reported in Table 1.1.

In the second set of simulation, we choose the number of observations τ to be 36, 60, 96, 120, and 240. Since we have only 96 observed monthly returns, we decide to simulate 240 values for each stock in the following way. For each stock, we compute the sample mean and standard deviation of the 96 observed monthly returns after omitting two extreme values at each tail. Sample means of the five stocks are 0.0349, 0.0234, 0.0200, 0.0218, and 0.0258, while sample standard deviations are 0.0715, 0.0724, 0.0931, 0.1131, and 0.1015. We use Kolmogorov-Smirnov Goodness-of-Fit Test to check whether returns of the five stocks are normally distributed and get the following p-values: 0.5, 0.5, 0.0622, 0.5, and 0.0288. Since monthly returns of the five stocks are approximately normally distributed, we generate 240 values for each stock from a normal distribution with mean and standard deviation respectively equal to the stock's sample mean and sample standard deviation. Other aspects of simulation are the same as in the first set of simulations. The results are reported in Table 1.2.

1.4.2 Results

Both Table 1.1 and Table 1.2 show that, as the number of observations τ increases, the probability that the GIC selection procedure picks the correct model (1, 4, 5) gets closer to 1. It confirms the validity of consistency of the GIC procedure. Both tables show that the logarithm rule tends to overestimate the model while overestimation and underestimation are equally likely when the square root rule is used, which is not surprising because the square root rule puts heavier penalty on models with many parameters than the logarithm rule.

Furthermore, both tables show that the probability of selecting the correct model under the square root penalty goes to 1 faster than under the logarithm penalty. This phenomenon can be explained by bounds on convergence rates for the error probabilities of the GIC given in both Shao (1998) and Zhang (1993). Both Shao (1998) and Zhang (1993) show that the rate at which the probability of choosing wrong models by the GIC goes to zero is an inverse function of the penalty λ_τ. Since $\sqrt{\tau}$ increases faster than $\log(\tau)$, the error probability with the penalty $\sqrt{\tau}$ goes to zero faster than that with the penalty $\log(\tau)$. Practically, this finding suggests that the square-root penalty is preferable for samples of moderate or large size, say more than 96 observations. For small samples, the square-root penalty does not seem to have an advantage over the logarithm penalty.

Contrasting Table 1.1 with Table 1.2, we notice that, under the same combination of τ and λ_τ, the probability of selecting the correct model in Table 1.2 is greater than in Table 1.1. The cause for the difference might be that explanatory variables used in the second set of simulations are generated independently from normal distributions while explanatory variables in the first set of simulations, being actual contemporaneous stock returns, are possibly correlated. Arguments in proof of Theorem 1.1 does not help to explain how structure in explanatory variables affects the probability of selecting the correct model.

In next section, we apply GIC to build a tracking portfolio so as to compute post-event long-term abnormal returns.

Table 1.1. Simulation with observed returns

Candidate Models	$\lambda_\tau = log(\tau)$			$sqrt(\tau)$		
	$\tau = 36$	60	96	$\tau = 36$	60	96
(1)	0	0	0	0	0	0
(2)	0	0	0	0	0	0
(3)	0	0	0	0	0	0
(4)	0	0	0	0	0	0
(5)	0	0	0	0	0	0
(1, 2)	0	0	0	0	0	0
(1, 3)	0	0	0	0	0	0
(1, 4)	0	1	0	7	4	0
(1, 5)	0	0	0	1	0	0
(2, 3)	0	0	0	0	0	0
(2, 4)	0	0	0	0	0	0
(2, 5)	0	0	0	0	0	0
(3, 4)	0	0	0	0	0	0
(3, 5)	0	0	0	0	0	0
(4, 5)	31	1	0	120	26	2
(1, 2, 3)	0	0	0	0	0	0
(1, 2, 4)	0	0	0	0	0	0
(1, 2, 5)	0	0	0	0	0	0
(1, 3, 4)	0	0	0	0	0	0
(1, 3, 5)	0	0	0	0	0	0
(1, 4, 5)	832	911	953	799	928	991
(2, 3, 4)	0	0	0	0	0	0
(2, 3, 5)	0	0	0	0	0	0
(2, 4, 5)	49	19	1	37	20	5
(3, 4, 5)	37	16	2	29	15	2
(1, 2, 3, 4)	0	0	0	0	0	0
(1, 2, 3, 5)	0	0	0	0	0	0
(1, 2, 4, 5)	23	34	24	5	4	0
(1, 3, 4, 5)	25	16	20	2	2	0
(2, 3, 4, 5)	2	0	0	0	1	0
(1, 2, 3, 4, 5)	1	2	0	0	0	0

Table 1.2. Simulation with simulated returns

Candidate Models	$\lambda_\tau = log(\tau)$					$sqrt(\tau)$				
	36	60	96	120	240	36	60	96	120	240
(1)	0	0	0	0	0	0	0	0	0	0
(2)	0	0	0	0	0	0	0	0	0	0
(3)	0	0	0	0	0	0	0	0	0	0
(4)	0	0	0	0	0	0	0	0	0	0
(5)	0	0	0	0	0	0	0	0	0	0
(1, 2)	0	0	0	0	0	0	0	0	0	0
(1, 3)	0	0	0	0	0	0	0	0	0	0
(1, 4)	1	0	0	0	0	4	0	0	0	0
(1, 5)	0	0	0	0	0	0	0	0	0	0
(2, 3)	0	0	0	0	0	0	0	0	0	0
(2, 4)	0	0	0	0	0	0	0	0	0	0
(2, 5)	0	0	0	0	0	0	0	0	0	0
(3, 4)	0	0	0	0	0	0	0	0	0	0
(3, 5)	0	0	0	0	0	0	0	0	0	0
(4, 5)	11	0	0	0	0	33	2	0	0	0
(1, 2, 3)	0	0	0	0	0	0	0	0	0	0
(1, 2, 4)	0	0	0	0	0	4	0	0	0	0
(1, 2, 5)	0	0	0	0	0	0	0	0	0	0
(1, 3, 4)	0	0	0	0	0	0	0	0	0	0
(1, 3, 5)	0	0	0	0	0	0	0	0	0	0
(1, 4, 5)	905	946	950	967	971	951	993	999	1000	1000
(2, 3, 4)	0	0	0	0	0	0	0	0	0	0
(2, 3, 5)	0	0	0	0	0	0	0	0	0	0
(2, 4, 5)	0	0	0	0	0	1	0	0	0	0
(3, 4, 5)	2	0	0	0	0	3	0	0	0	0
(1, 2, 3, 4)	0	0	0	0	0	1	0	0	0	0
(1, 2, 3, 5)	0	0	0	0	0	0	0	0	0	0
(1, 2, 4, 5)	43	24	22	17	15	2	1	1	0	0
(1, 3, 4, 5)	32	27	26	16	14	2	4	0	0	0
(2, 3, 4, 5)	0	0	0	0	0	0	0	0	0	0
(1, 2, 3, 4, 5)	6	3	2	0	0	0	0	0	0	0

1.5 Application

In this section, we apply GIC to build a tracking portfolio for the purpose of measuring long-term post-event abnormal stock return. A great interest in learning the long-term impact of corporate actions has recently arisen among finance researchers and generated a considerable literature that is still growing. The evidence for existence of long-term post-event abnormal stock return challenges the belief that the U.S. stock market is efficient, and motivates research in behavioral finance. See Fama (1998) for a summary of the literature and references therein. In these studies, the most important job is to precisely estimate what the event firm's return would have been if the event had not happened. In the following, we compare the performance of three estimation methods.

1.5.1 Estimates of Abnormal Return

The three-year buy-and-hold abnormal return of firm i, denoted as AR_i is measured as follows

$$AR_i = R_i - BR_i \ , \tag{1.18}$$

where R_i is the buy-and-hold return of firm i over the same three years, and BR_i is a specific benchmark-hold return over the same three years. The benchmark return is an estimate of the unobservable *status quo* return that an event firm would have had over the three years following the event month if the event had not happened. The three-year buy-and-hold return of firm i is computed by compounding monthly returns, i.e., $R_i = \prod_{t=1}^{36}(1 + r_{it}) - 1$, where r_{it} is firm i's return in month t.

We use four benchmarks to estimate the unobservable *status quo* return of an event firm. The *first benchmark* is a size and book-to-market ratio matched portfolio. Similar benchmarks are widely used in existing finance literature, e.g., see Dharan and Ikenberry (1995), Desai and Jain (1997), Barber and Lyon (1997), Lyon, Barber and Tsai (1999), and Mitchell and Stafford (2000). To identify the size and book-to-market ratio matched portfolio of an event firm, we construct 70 reference portfolios on the basis of firm size and book-to-market ratio and the matched portfolio is the one that includes the event firm. The 70 reference portfolios are formed according to the following procedure of three steps.[1]

Step 1: At the end of June of year t, we calculate firm size as price per share multiplied by shares outstanding, sort all NYSE firms by firm size into ten portfolios of equal size, and then place each AMEX/Nasdaq firm in the portfolio whose range of firm sizes covers the firm's size.

Step 2: We partition the smallest size decile portfolio into five subportfolios of equal size on the basis of firm size rankings of all firms in the portfolio without regard to listing exchange, so that we have 14 firm size portfolios. [2]

Step 3: We divide each of the 14 portfolios into five subportfolios of equal size by ranking all firms in the portfolio by their book-to-market ratios at the end of year $t-1$, so that we end up with 70 reference portfolios. In the last step of the procedure, a firm's book-to-market ratio at the end of year $t-1$ is equal to the ratio of the book common equity (COMPUSTAT data item 60) at the end of the firm's fiscal year ending in year $t-1$ over the firm's market common equity at the end of December of year $t-1$. Throughout the procedure, we include only stocks with ordinary common equity shares (that is, firms with CRSP share code being 11) and exclude firms of negative book common equity whenever book equity is needed.

[1] The procedure is created in the same spirits as Fama and French (1993), and is almost identical to that in Lyon, Barber and Tsai (1999). The only difference between our procedure and that in Lyon, Barber and Tsai (1999) is that we use firms with CRSP share code being only 11 while they allow the CRSP share code to be both 10 and 11.

[2] A majority of Nasdaq firms are small and thus fall into the smallest size decile portfolio; as a result, approximately 50 percent of all firms fall in the smallest size decile. By further partitioning the portfolio, we make the 14 size portfolios have almost the same number of stocks.

The benchmark return based on a size and book-to-market matched portfolio is computed as follows:

$$BR_i^{SZBM} = \prod_{t=1}^{36} \left[1 + \frac{\sum_{j=1}^{n_t} r_{jt}}{n_t} \right] - 1 \ , \tag{1.19}$$

where r_{jt} is the monthly return of firm j in month t and n_t is the number of firms in month t. We label the first benchmark as B1:SZBM and that is how the above abnormal return gets its superscript "SZBM".

The *second benchmark* is a portfolio of the ten firms that have the largest sample correlation coefficients with the event firm among all firms in the size and book-and-market ratio matched portfolio. To identify the ten firms, we chose the size and book-and-market ratio matched portfolio for the event firm as described above, identify all firms in the portfolio that have returns for the five years before and the three years after the event month, calculate the sample correlation coefficient between each identified firm and the event firm based on the 60 monthly returns in the pre-event five years, and then choose the ten firms that have the largest sample correlation coefficients. We label the second benchmark as B2:MC10 for the most correlated ten and compute the three-year post-event benchmark return as follows

$$BR_i^{MC10} = \sum_{j=1}^{10} \frac{\left[\prod_{t=1}^{36}(1 + r_{jt})\right] - 1}{10} \ , \tag{1.20}$$

where r_{jt} is the monthly return of firm j in month t. The benchmark return is return of investing equally in the most correlated ten firms over the three years starting with the event month.

The *third benchmark* is simply the buy-and-hold return of the single most correlated firm over the post-event three years. In contrast with the second benchmark, this benchmark use only the single firm that has the largest sample correlation coefficient with the event firm in the five years before the event month. We label this benchmark as B3:MC1 for the most correlated one.

The *fourth and last benchmark* falls between the second and the third benchmark. After obtaining the most correlated ten firms as in the second benchmark, we apply the GIC selection procedure with the event firm as the response variable and the ten firms as explanatory variables using the 60 monthly returns in the pre-event five years. We use the optimal tracking portfolio resulted from the GIC selection procedure as a benchmark and label it as B4:GIC. We compute the three-year post-event return of the benchmark B4:GIC as follows

$$BR_i^{GIC} = \sum_{j=1}^{n_i} w_j \left[\prod_{t=1}^{36} (1 + r_{jt}) - 1 \right] , \tag{1.21}$$

where r_{jt} is the monthly return of firm j in month t, n_i is the number of firms in the GIC optimal tracking portfolio, and w_j is the optimal weight of the jth firm in the GIC optimal tracking portfolio. In contrast to B2:MC10 and B3:MC1, B4:GIC removes unrelated firms from B2:MC10, keeps more relevant firms than B3:MC1, and also allows different weights for different stocks in the tracking portfolio.

1.5.2 Empirical Assessment of Performance

To assess the performance of the three benchmarks, we employ a procedure which uses actual security return data to examine the characteristics of abnormal returns produced by the three benchmarks. This type of procedure has been used widely in finance literature to compare performance of various methodologies for measuring abnormal returns, see, e.g., Brown and Warner (1980), Kothari and Warner (1997), Barber and Lyon (1997), and Lyon, Barber and Tsai (1999).

In the simulation procedure, we randomly choose with replacement a sample of 200 event months between July 1984 and December 1994, inclusively. In each selected event month, we then randomly choose an event firm without replacement that has returns for the five years before and the three years after the event month. We compute the three-year post-event abnormal return for each event firm using all three benchmarks. Since the 200 event firms are randomly selected and not many of the 200 event months were supposed to experience any event, we expect that the 200 abnormal returns concentrate around zero.

Panel A of Table 1.3 reports sample mean, median, standard deviation, interquartile range, skewness coefficient and kurtosis coefficient of the 200 abnormal returns under each benchmark. Student's t test is employed to test the null hypothesis that mean of abnormal return is zero while Fisher's distribution-free sign test is used to test the null hypothesis that median of abnormal return is zero (See Hollander and Wolf (2000) for detailed description of Fisher's distribution-free sign test). The p-values from both tests are reported in the last two columns.

The t test shows that none of the four benchmarks yield mean abnormal return significantly different from zero. The sign test reports that the first two benchmarks B1:SZBM and B2:MC10 have significantly non-zero median abnormal return while the last two benchmarks B3:MC1 and B4:GIC do not. Since sample skewness coefficients and sample kurtosis coefficients under all four benchmark are far away from the theoretical values of a standard normal distribution (the theoretical skewness and kurtosis coefficients of a standard normal distribution are 0 and 3, respectively), we believe that median is more appropriate than mean in measuring central tendency of abnormal returns and that the sign test is more appropriate than the t test in telling the difference between benchmarks. Based on the sign test, the last two benchmarks B3:MC1 and B4:GIC produce abnormal returns for this sample that are on average close to zero. Since the sample of 200 firms under current study are randomly selected without actual events occurring in specified event months, abnormal returns are expected to be close to zero on average. From this point of view, the last two benchmarks do a better job for this sample than the first two benchmarks. The reason why the first two benchmarks underestimate abnormal return (median of abnormal returns under B1:SZBM and B2:MC10 are -0.2555 and -0.1466, respectively) might be that both benchmarks include many stocks in their

Table 1.3. A sample of 200 randomly selected firms.

	Descriptive Statistics						p-values	
	mean	median	std	iqr	skewness	kurtosis	t	sign
Panel A: Abnormal returns under four benchmarks								
B1:SZBM	-0.022	-0.256	1.180	0.957	1.437	6.982	0.794	0.009
B2:MC10	0.054	-0.147	1.145	0.995	1.389	6.775	0.503	0.028
B3:MC1	-0.021	0.069	2.214	1.109	-3.323	23.699	0.896	0.289
B4:GIC	0.084	-0.030	1.381	0.941	0.115	9.856	0.390	0.621
Panel B: Paired difference in abnormal return between benchmarks								
d(B4, B1)	0.106	0.163	0.884	0.652	-2.811	17.689	0.092	0.000
d(B4, B2)	0.030	0.063	0.674	0.550	-2.494	14.443	0.533	0.019
d(B4, B3)	0.105	-0.075	1.322	0.536	5.037	38.007	0.264	0.056
d(B1, B2)	-0.076	-0.050	0.416	0.328	0.047	8.765	0.010	0.009
d(B1, B3)	-0.001	-0.213	1.889	0.847	5.366	38.348	0.993	0.000
d(B2, B3)	0.075	-0.185	1.712	0.833	5.369	39.139	0.537	0.000

benchmark portfolios (at least 10 stocks) and the average return of these many stocks is closer to the market return than an event firm's return.

Panel B of Table 1.3 reports paired difference in abnormal return between the four benchmarks. Paired difference in abnormal return between any two benchmarks is the difference between the abnormal returns under the two benchmarks for each firm in the sample. For example, the paired difference between B1:SZBM and B2:MC10, d(B1, B2), is a vector of 200 values, each value for one firm being the difference between abnormal return under B1:SZBM and that under B2:MC10. The paired difference gives a direct and precise comparison between benchmarks. Again, the sign test is more appropriate than the t test in testing the difference between benchmarks. Based on the sign test, B4:GIC is significantly different from B1:SZBM and B2:MC10 while B4:GIC is different from B3:MC1 with a p-value of 0.0560.

1.6 Proofs

Proof of Lemma 1.1:

The loss function $L_\tau(v)$ can be decomposed as follows

$$\begin{aligned}
\tau L_\tau(v) &= ||\boldsymbol{\mu}_\tau - \hat{\boldsymbol{\mu}}_\tau(v)||^2 \\
&= ||\boldsymbol{\mu}_\tau - (\boldsymbol{\eta}_\tau(v) + \boldsymbol{H}_\tau(v)\boldsymbol{y}_\tau)||^2 \\
&= ||\boldsymbol{\mu}_\tau - \boldsymbol{\eta}_\tau(v) - \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau - \boldsymbol{H}_\tau(v)\boldsymbol{e}_\tau||^2 \\
&= ||\boldsymbol{\mu}_\tau - \boldsymbol{\eta}_\tau(v) - \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau||^2 + \boldsymbol{e}'_\tau\boldsymbol{H}_\tau(v)\boldsymbol{e}_\tau \\
&\quad - 2\boldsymbol{e}'_\tau\boldsymbol{H}'_\tau(v)[\boldsymbol{\mu}_\tau - \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau - \boldsymbol{\eta}_\tau(v)] \\
&= ||\boldsymbol{\mu}_\tau - \boldsymbol{\eta}_\tau(v) - \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau||^2 + \boldsymbol{e}'_\tau\boldsymbol{H}_\tau(v)\boldsymbol{e}_\tau \,.
\end{aligned}$$

The second and third equality holds because of (1.7) and (1.1), respectively. The fourth equality holds because $\boldsymbol{H}'_\tau(v)\boldsymbol{H}_\tau(v) = \boldsymbol{H}_\tau(v)$. The last equality holds because $\boldsymbol{H}'_\tau(v)[\boldsymbol{\mu}_\tau - \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau - \boldsymbol{\eta}_\tau(v)] = \boldsymbol{H}'_\tau(v)(I - \boldsymbol{H}_\tau(v))(\mu_\tau - \boldsymbol{X}_1(v)\boldsymbol{G}_1^{-1}\boldsymbol{g}(v)) = 0$.

For $v \in V^c$, we know $\boldsymbol{\mu}_\tau = \boldsymbol{X}_\tau(v)\boldsymbol{\beta}_\tau(v)$ and that the random vector $\boldsymbol{\epsilon}$ in the constrained linear model (1.4) has mean zero. Then the least square estimate $\hat{\beta}$ in equation (1.7) is unbiased. By taking expectation on both sides of equation (1.7), we obtain $\boldsymbol{X}_\tau(v)\boldsymbol{\beta}_\tau(v) = \boldsymbol{\eta}_\tau(v) + \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau$. Therefore, we obtain $\Delta_\tau(v) = ||\boldsymbol{\mu}_\tau - \boldsymbol{\eta}_\tau(v) - \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau||^2/\tau = 0$.

The expression for the expected average squared error is obtained as follows

$$\begin{aligned}
R_\tau(v) &= E(L_\tau(v)) \\
&= \Delta_\tau(v) + E(\boldsymbol{e}'_\tau\boldsymbol{H}_\tau(v)\boldsymbol{e}_\tau)/\tau \\
&= \Delta_\tau(v) + tr(\boldsymbol{H}_\tau(v)\mathrm{Var}(\boldsymbol{e}_\tau))/\tau \\
&= \Delta_\tau(v) + \sigma^2 tr(\boldsymbol{H}_\tau(v))/\tau
\end{aligned}$$

□

Proof of Theorem 1.1:

In the proof, we use arguments in spirit similar to those in Li (1987) and Shao (1997).

First, note that the GIC procedure is to minimize

$$\begin{aligned}\Gamma_\tau(v) &= \frac{||\boldsymbol{y}_\tau - \hat{\boldsymbol{\mu}}_\tau(v)||^2}{\tau} + \frac{\lambda_\tau \hat{\sigma}_\tau^2 tr(\boldsymbol{H}_\tau(v))}{\tau} \\ &= \frac{||\boldsymbol{e}_\tau||^2 + \tau L_\tau(v) + 2\boldsymbol{e}_\tau'(\boldsymbol{\mu}_\tau(v) - \hat{\boldsymbol{\mu}}_\tau(v))}{\tau} + \frac{\lambda_\tau \hat{\sigma}_\tau^2 tr(\boldsymbol{H}_\tau(v))}{\tau} .\end{aligned}$$

For $v \in V^c$, since $\boldsymbol{\mu}_\tau(v) - \boldsymbol{\eta}_\tau(v) - \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau(v) = 0$ and $L_\tau(v) = (\boldsymbol{e}_\tau'\boldsymbol{H}_\tau(v)\boldsymbol{e}_\tau)/\tau$, we get

$$\Gamma_\tau(v) = \frac{||\boldsymbol{e}_\tau||^2}{\tau} + \frac{\lambda_\tau \hat{\sigma}_\tau^2 tr(\boldsymbol{H}_\tau(v))}{\tau} - \frac{\boldsymbol{e}_\tau'\boldsymbol{H}_\tau(v)\boldsymbol{e}_\tau}{\tau} \tag{1.22}$$

For $v \in V - V^c$, we have

$$\begin{aligned}\Gamma_\tau(v) &= \frac{||\boldsymbol{e}_\tau||^2}{\tau} + \frac{||\boldsymbol{\mu}_\tau(v) - \hat{\boldsymbol{\mu}}_\tau(v)||^2}{\tau} + \frac{(\lambda_\tau \hat{\sigma}_\tau^2 - 2\sigma^2) tr(\boldsymbol{H}_\tau(v))}{\tau} \\ &\quad + \frac{2[\sigma^2 tr(\boldsymbol{H}_\tau(v)) - \boldsymbol{e}_\tau'\boldsymbol{H}_\tau(v)\boldsymbol{e}_\tau]}{\tau} + \frac{2\boldsymbol{e}_\tau'[\boldsymbol{\mu}_\tau(v) - (\boldsymbol{\eta}_\tau(v) + \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau(v))]}{\tau} \\ &= \frac{||\boldsymbol{e}_\tau||^2}{\tau} + L_\tau(v) + o_p(L_\tau(v))\end{aligned} \tag{1.23}$$

where the last equality holds uniformly in $v \in V - V^c$. To establish the last equality, it suffices to show that in probability,

$$\max_{v \in V - V^c} \frac{\boldsymbol{e}_\tau'[\boldsymbol{\mu}_\tau(v) - (\boldsymbol{\eta}_\tau(v) + \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau(v))]}{\tau R_\tau(v)} \xrightarrow{P} 0 , \tag{1.24}$$

$$\max_{v \in V - V^c} \frac{\sigma^2 tr(\boldsymbol{H}_\tau(v)) - \boldsymbol{e}_\tau'\boldsymbol{H}_\tau(v)\boldsymbol{e}_\tau}{\tau R_\tau(v)} \xrightarrow{P} 0 , \tag{1.25}$$

$$\max_{v \in V - V^c} \frac{(\lambda_\tau \hat{\sigma}_\tau^2 - 2\sigma^2) tr(\boldsymbol{H}_\tau(v))}{\tau R_\tau(v)} \xrightarrow{P} 0 , \tag{1.26}$$

and

$$\max_{v \in V - V^c} \left| \frac{L_\tau(v)}{R_\tau(v)} - 1 \right| \xrightarrow{P} 0 . \tag{1.27}$$

We shall prove (1.24) first. Given any $\varepsilon > 0$, by Chebyshev's inequality we have

$$\begin{aligned}& P\left\{\max_{v \in V - V^c} \left| \frac{\boldsymbol{e}_\tau'[\boldsymbol{\mu}_\tau(v) - \boldsymbol{\eta}_\tau(v) - \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau(v)]}{\tau R_\tau(v)} \right| > \varepsilon\right\} \\ \leq \quad & \sum_{v \in V - V^c} \frac{E[\boldsymbol{e}_\tau'(\boldsymbol{\mu}_\tau(v) - \boldsymbol{\eta}_\tau(v) - \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau(v))]^2}{[\tau R_\tau(v)\varepsilon]^2} .\end{aligned} \tag{1.28}$$

Since $E(\mathbf{e}) = 0$, we know

$$E[\boldsymbol{e}_\tau'(\boldsymbol{\mu}_\tau(v) - \boldsymbol{\eta}_\tau(v) - \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau(v))]^2$$

$$
\begin{aligned}
&= \operatorname{Var}(\boldsymbol{e}_\tau'(\boldsymbol{\mu}_\tau(v) - \boldsymbol{\eta}_\tau(v) - \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau(v))) \\
&= (\boldsymbol{\mu}_\tau(v) - \boldsymbol{\eta}_\tau(v) - \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau(v))' \, \sigma^2 \boldsymbol{I}_\tau \, (\boldsymbol{\mu}_\tau(v) - \boldsymbol{\eta}_\tau(v) - \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau(v)) \\
&= \sigma^2 ||\boldsymbol{\mu}_\tau(v) - \boldsymbol{\eta}_\tau(v) - \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau(v)||^2 \,,
\end{aligned}
$$

where Var(·) gives the covariance matrix of its argument. Since $\tau R_\tau(v) \geq ||\boldsymbol{\mu}_\tau(v) - \boldsymbol{\eta}_\tau(v) - \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau(v)||^2$, the right hand side of (1.28) does not exceed

$$
\frac{\sigma^2}{\varepsilon^2} \sum_{v \in V - V^c} \frac{1}{\tau R_\tau(v)} \,,
$$

which tends to 0 by condition (1.11). We thus obtain (1.24).

Next, we shall prove (1.25). Since $\boldsymbol{e}_\tau$ is normally distributed with $E(\boldsymbol{e}_\tau) = 0$ and $\operatorname{Var}(\boldsymbol{e}_\tau) = \sigma^2 \boldsymbol{I}_\tau$, it is well known that $E(\boldsymbol{e}_\tau' \boldsymbol{H}_\tau(v) \boldsymbol{e}_\tau) = \sigma^2 tr(\boldsymbol{H}_\tau(v))$ and $\operatorname{Var}(\boldsymbol{e}_\tau' \boldsymbol{H}_\tau(v) \boldsymbol{e}_\tau) = 2\sigma^4 tr(\boldsymbol{H}_\tau(v)\boldsymbol{H}_\tau(v))$. Given any $\varepsilon > 0$, by Chebyshev's inequality we have,

$$
\begin{aligned}
& P\left\{ \max_{v \in V - V^c} \left| \frac{tr(\boldsymbol{\Psi}_\tau \boldsymbol{H}_\tau(v)) - \boldsymbol{e}_\tau' \boldsymbol{H}_\tau(v) \boldsymbol{e}_\tau}{\tau R_\tau(v)} \right| > \varepsilon \right\} \\
\leq & \sum_{v \in V - V^c} \frac{E[tr(\boldsymbol{\Psi}_\tau \boldsymbol{H}_\tau(v)) - \boldsymbol{e}_\tau' \boldsymbol{H}_\tau(v) \boldsymbol{e}_\tau]^2}{[\tau R_\tau(v)\varepsilon]^2} \\
= & \sum_{v \in V - V^c} \frac{\operatorname{Var}(\boldsymbol{e}_\tau' \boldsymbol{H}_\tau(v) \boldsymbol{e}_\tau)}{[\tau R_\tau(v)\varepsilon]^2} \\
\leq & \frac{m\sigma^4}{\varepsilon^2} \sum_{v \in V - V^c} \frac{1}{(\tau R_\tau(v))^2} \,.
\end{aligned}
$$

Since $R_\tau(v) > 0$, the last term goes to zero under conditions (1.11). We thus obtain (1.25).

To prove (1.26), we note that both $\hat{\sigma}^2$ and $tr(\boldsymbol{H}_\tau(v))$ are bounded. Then (1.26) holds under condition (1.11).

Finally, (1.27) is equivalent to (1.25) since

$$
\left| \frac{L_\tau(v)}{R_\tau(v)} - 1 \right| = \frac{|L_\tau(v) - R_\tau(v)|}{R_\tau(v)} = \frac{|\boldsymbol{e}_\tau' \boldsymbol{H}_\tau(v) \boldsymbol{e}_\tau - \sigma^2 tr(\boldsymbol{H}_\tau(v))|}{\tau R_\tau(v)} \,.
$$

We thus conclude the proof of equation (1.23).

Next, we show the asymptotic loss efficiency and consistency of the GIC minimizer $\hat{v}_\tau$, using equations (1.22) and (1.23). When V^c is empty, we know from (1.23) that $\hat{v}_\tau$ is asymptotically equal to the minimizer of $L_\tau(v)$, that is, $\hat{v}_\tau$ is asymptotically loss efficient.

When V^c is not empty, we can show that for any $v^c \in V^c$,

$$\Gamma_\tau(v^c) - \frac{||e_\tau||^2}{\tau} = o_p(L_\tau(v)) \tag{1.29}$$

uniformly in $v \in V - V^c$, using similar arguments in proof of (1.26) and (1.25). Equation (1.29) together with equation (1.23) implies that $\hat{v}_\tau$ will always belong to V^c asymptotically if V^c is not empty.

We can further prove, using similar arguments in proof of (1.25),

$v\in V^c$ $\lambda_\tau \hat{\sigma}_\tau^2 tr(\boldsymbol{H}_\tau(v))$

$$\max_{v\in V^c} \frac{e_\tau' \boldsymbol{H}_\tau(v) e_\tau}{\lambda_\tau \hat{\sigma}_\tau^2 tr(\boldsymbol{H}_\tau(v))} \xrightarrow{p} 0 \ .$$

Then we know from (1.22), for $v^c \in V^c$, $\Gamma_\tau(v) - ||e_\tau||^2/\tau$ is asymptotically dominated by the term $\lambda_\tau \hat{\sigma}_\tau^2 tr(\boldsymbol{H}_\tau(v))/\tau$. Because $\hat{\sigma}_\tau$ is bounded and does not depend on the index v, the dominating term $\lambda_\tau \hat{\sigma}_\tau^2 tr(\boldsymbol{H}_\tau(v))/\tau$ has the same minimizer as $L_\tau(v) = e_\tau' \boldsymbol{H}_\tau(v) e_\tau/\tau$ asymptotically. Therefore we obtain

$$P\{\hat{v}_\tau \in V^c \text{ but } \hat{v}_\tau \neq v_\tau^L\} \to 0 \ , \tag{1.30}$$

which means that $\hat{v}_\tau$ is asymptotically loss efficient when V^c is not empty.

Equation (1.30) also implies that $P\{\hat{v}_\tau = v_\tau^L\} \to 1$ when V^c is not empty. We thus conclude that $\hat{v}_\tau$ is consistent. $\square$

Proof of Lemma 1.2:

Proof of the first two parts is identical to proof of Lemma 1.1. So only the last part is proved here.

$$\begin{aligned} R_\tau(v) &= E(L_\tau(v)) \\ &= \Delta_\tau(v) + E((e_\tau' \boldsymbol{H}_\tau(v) e_\tau))/\tau \end{aligned}$$

$$
\begin{aligned}
&= \Delta_\tau(v) + tr(E(\boldsymbol{e}_\tau \boldsymbol{e}_\tau')\boldsymbol{H}_\tau(v))/\tau \\
&= \Delta_\tau(v) + tr(\boldsymbol{\Psi}_\tau \boldsymbol{H}_\tau(v))/\tau \ .
\end{aligned}
$$

□

The following two lemmas are necessary for the proof of Theorem 1.2.

Lemma 1.3 *Let $\boldsymbol{\Psi}_\tau$ be the covariance matrix given by (1.14). Let $\boldsymbol{a} = (a_1, a_2, \cdots, a_\tau)'$ and $\boldsymbol{b} = (b_1, b_2, \cdots, b_\tau)'$ be any two vectors. The following inequality holds*

$$
|\boldsymbol{a}'\boldsymbol{\Psi}_\tau \boldsymbol{b}| \le \frac{||\boldsymbol{a}||^2 + ||\boldsymbol{b}||^2}{2}\Upsilon \ ,
$$

where Υ is the absolute sum of autocovariances given in (1.13). In particular, when $||\boldsymbol{a}||^2 = ||\boldsymbol{b}||^2 = 1$, we have $|\boldsymbol{a}'\boldsymbol{\Psi}_\tau \boldsymbol{b}| \le \Upsilon$.

Proof of Lemma 1.3:

Because of the special structure $\boldsymbol{\Psi}_\tau$, we have

$$
\begin{aligned}
|\boldsymbol{a}'\boldsymbol{\Psi}_\tau \boldsymbol{b}| &= \left|\sum_{k=1}^{\tau}\sum_{l=1}^{\tau} a_k b_l \gamma_{k-l}\right| \\
&= \left|\gamma_0 \sum_{l=1}^{\tau} a_l b_l + \sum_{i=1}^{\tau-1}\left(\gamma_i \sum_{l=1}^{\tau-i}(a_{l+i}b_l + a_l b_{l+i})\right)\right| \\
&\le \gamma_0 \sum_{l=1}^{\tau} |a_l b_l| + \sum_{i=1}^{\tau-1}\left(|\gamma_i| \sum_{l=1}^{\tau-i}(|a_{l+i}b_l| + |a_l b_{l+i}|)\right) \ .
\end{aligned}
$$

Notice that

$$
\sum_{l=1}^{\tau} a_l b_l \le \sum_{l=1}^{\tau} \frac{a_l^2 + b_l^2}{2} = \frac{||\boldsymbol{a}||^2 + ||\boldsymbol{b}||^2}{2} \ ,
$$

and for any $i \in \{1, 2, \cdots, \tau - 1\}$,

$$
\sum_{l=1}^{\tau-i} a_{l+i} b_l \le \sum_{l=1}^{\tau-i} \frac{a_{l+i}^2 + b_l^2}{2} \le \frac{1}{2}\left(\sum_{l=1}^{\tau} a_l^2 + \sum_{l=1}^{\tau} b_l^2\right) = \frac{||\boldsymbol{a}||^2 + ||\boldsymbol{b}||^2}{2} \ ,
$$

and similarly

$$
\sum_{l=1}^{\tau-i} a_l b_{l+i} \le \sum_{l=1}^{\tau-i} \frac{a_l^2 + b_{l+i}^2}{2} \le \frac{1}{2}\left(\sum_{l=1}^{\tau} a_l^2 + \sum_{l=1}^{\tau} b_l^2\right) = \frac{||\boldsymbol{a}||^2 + ||\boldsymbol{b}||^2}{2} \ .
$$

We thus obtain

$$|\boldsymbol{a}'\boldsymbol{\Psi}_\tau\boldsymbol{b}| \le \frac{||\boldsymbol{a}||^2+||\boldsymbol{b}||^2}{2}(\gamma_0 + 2\sum_{j=0}^{\infty}|\gamma_j|) = \frac{||\boldsymbol{a}||^2+||\boldsymbol{b}||^2}{2}\Upsilon\ .$$

□

Lemma 1.4 *Suppose that $\boldsymbol{H}$ is an idempotent matrix of rank r. Then*

$$tr(\boldsymbol{\Psi}_\tau\boldsymbol{H}) \le r\Upsilon \quad \textit{and} \quad tr(\boldsymbol{\Psi}_\tau\boldsymbol{H}\boldsymbol{\Psi}_\tau\boldsymbol{H}) \le (r\Upsilon)^2\ .$$

Proof of Lemma 1.4:

Let the $\tau\times\tau$ matrix $\boldsymbol{\Lambda}$ be $\begin{bmatrix} \boldsymbol{I}_r & 0 \\ 0 & 0 \end{bmatrix}$, be a $\tau\times\tau$ matrix where $\boldsymbol{I}_r$ represents the identity matrix of dimension r. Since $\boldsymbol{H}$ is an idempotent matrix of rank r, there exists a $\tau\times\tau$ orthogonal matrix $\boldsymbol{C}$ such that $\boldsymbol{C}'\boldsymbol{H}\boldsymbol{C}=\boldsymbol{\Lambda}$. Then we have

$$tr(\boldsymbol{\Psi}_\tau\boldsymbol{H}) = tr(\boldsymbol{\Psi}_\tau\boldsymbol{C}\boldsymbol{\Lambda}\boldsymbol{C}') = tr(\boldsymbol{C}'\boldsymbol{\Psi}_\tau\boldsymbol{C}\boldsymbol{\Lambda}) = \sum_{k=1}^{r}\boldsymbol{c}_k'\boldsymbol{\Psi}_\tau\boldsymbol{c}_k\ ,$$

where $\boldsymbol{c}_k$ is the kth column vector of the matrix C. Since C is orthogonal, $\boldsymbol{c}_k'\boldsymbol{c}_k = 1$ for $k = 1, 2, \cdots, r$, and thus we know $tr(\boldsymbol{\Psi}_\tau\boldsymbol{H}) \le r\Upsilon$ by Lemma 1.3.

Notice that $tr(\boldsymbol{\Psi}_\tau\boldsymbol{H}) = tr(\boldsymbol{\Lambda}\boldsymbol{C}'\boldsymbol{\Psi}_\tau\boldsymbol{C}\boldsymbol{\Lambda})$ and $\boldsymbol{\Lambda}\boldsymbol{C}'\boldsymbol{\Psi}_\tau\boldsymbol{C}\boldsymbol{\Lambda}$ is symmetric and non-negative definite. Therefore we have $tr(\boldsymbol{\Psi}_\tau\boldsymbol{H}\boldsymbol{\Psi}_\tau\boldsymbol{H}) \le [tr(\boldsymbol{\Psi}_\tau\boldsymbol{H})]^2 \le (r\Upsilon)^2$. □

Proof of Theorem 1.2:

First, note that the EGIC procedure is to minimize

$$\begin{aligned}\Phi_\tau(v) &= \frac{S_\tau(v)}{\tau} + \frac{\lambda_\tau tr(\hat{\boldsymbol{\Psi}}_\tau\boldsymbol{H}_\tau(v))}{\tau} \\ &= \frac{||\boldsymbol{e}_\tau||^2 + ||\boldsymbol{\mu}_\tau(v) - \hat{\boldsymbol{\mu}}_\tau(v)||^2 + 2\boldsymbol{e}_\tau'(\boldsymbol{\mu}_\tau(v) - \hat{\boldsymbol{\mu}}_\tau(v))}{\tau} + \frac{\lambda_\tau tr(\hat{\boldsymbol{\Psi}}_\tau\boldsymbol{H}_\tau(v))}{\tau}\ .\end{aligned}$$

For $v\in V^c$, since $\boldsymbol{\mu}_\tau(v) - \boldsymbol{\eta}_\tau(v) - \boldsymbol{H}_\tau(v)\boldsymbol{\mu}_\tau(v) = 0$, we have $L_\tau(v) = (\boldsymbol{e}_\tau'\boldsymbol{H}_\tau(v)\boldsymbol{e}_\tau)/\tau$ and thus

$$\Phi_\tau(v) = \frac{||\boldsymbol{e}_\tau||^2}{\tau} + \frac{\lambda_\tau tr(\hat{\boldsymbol{\Psi}}_\tau\boldsymbol{H}_\tau(v))}{\tau} - \frac{\boldsymbol{e}_\tau'\boldsymbol{H}_\tau(v)\boldsymbol{e}_\tau}{\tau} \qquad (1.31)$$

For $v\in V - V^c$, we have

$$\Phi_\tau(v) = \frac{||\boldsymbol{e}_\tau||^2}{\tau} + \frac{||\boldsymbol{\mu}_\tau(v) - \hat{\boldsymbol{\mu}}_\tau(v)||^2}{\tau} + \frac{\lambda_\tau tr(\hat{\boldsymbol{\Psi}}_\tau\boldsymbol{H}_\tau(v)) - 2tr(\boldsymbol{\Psi}_\tau\boldsymbol{H}_\tau(v))}{\tau}$$

$$+\frac{2[tr(\Psi_\tau H_\tau(v)) - e'_\tau H_\tau(v)e_\tau]}{\tau} + \frac{2e'_\tau[\mu_\tau(v) - \eta_\tau(v) - H_\tau(v)\mu_\tau(v)]}{\tau}$$

$$= \frac{||e_\tau||^2}{\tau} + L_\tau(v) + o_p(L_\tau(v)) \qquad (1.32)$$

where the last equality holds uniformly in $v \in V - V^c$. To establish the last equality, it suffices to show that in probability,

$$\max_{v\in V-V^c} \frac{e'_\tau[\mu_\tau(v) - \eta_\tau(v) - H_\tau(v)\mu_\tau(v)]}{\tau R_\tau(v)} \xrightarrow{P} 0 , \qquad (1.33)$$

$$\max_{v\in V-V^c} \frac{tr(\Psi_\tau H_\tau(v)) - e'_\tau H_\tau(v)e_\tau}{\tau R_\tau(v)} \xrightarrow{P} 0 , \qquad (1.34)$$

$$\max_{v\in V-V^c} \frac{\lambda_\tau tr(\hat{\Psi}_\tau H_\tau(v)) - 2tr(\Psi_\tau H_\tau(v))}{\tau R_\tau(v)} \xrightarrow{P} 0 , \qquad (1.35)$$

and

$$\max_{v\in V-V^c} \left| \frac{L_\tau(v)}{R_\tau(v)} - 1 \right| \xrightarrow{P} 0 . \qquad (1.36)$$

We shall prove (1.33) first. Given any $\varepsilon > 0$, by Chebyshev's inequality we have

$$P\left\{\max_{v\in V-V^c}\left|\frac{e'_\tau[\mu_\tau(v)-\eta_\tau(v)-H_\tau(v)\mu_\tau(v)]}{\tau R_\tau(v)}\right| > \varepsilon\right\}$$

$$\leq \quad \sum_{v\in V-V^c} \frac{E[e'_\tau(\mu_\tau(v)-\eta_\tau(v)-H_\tau(v)\mu_\tau(v))]^2}{[\tau R_\tau(v)\varepsilon]^2} . \qquad (1.37)$$

Since $E(e) = 0$, we have

$$\begin{aligned} & E[e'_\tau(\mu_\tau(v) - \eta_\tau(v) - H_\tau(v)\mu_\tau(v))]^2 \\ = \ & \mathrm{Var}(e'_\tau(\mu_\tau(v) - \eta_\tau(v) - H_\tau(v)\mu_\tau(v))) \\ = \ & (\mu_\tau(v) - \eta_\tau(v) - H_\tau(v)\mu_\tau(v))'\Psi_\tau(\mu_\tau(v) - \eta_\tau(v) - H_\tau(v)\mu_\tau(v)) \\ \leq \ & \Upsilon||\mu_\tau(v) - \eta_\tau(v) - H_\tau(v)\mu_\tau(v)||^2 , \end{aligned}$$

where Var$(\cdot)$ gives the covariance matrix of its argument and the last inequality holds because of Lemma 1.3. Since $\tau R_\tau(v) \geq ||\mu_\tau(v) - \eta_\tau(v) - H_\tau(v)\mu_\tau(v)||^2$, the right hand side of (1.37) does not exceed

$$\frac{\Upsilon}{\varepsilon^2} \sum_{v\in V-V^c} \frac{1}{\tau R_\tau(v)} ,$$

which tends to 0 by condition (1.17). We thus obtain (1.33).

Next, we shall prove (1.34). Since $\boldsymbol{e}_\tau$ is normally distributed with $E(\boldsymbol{e}_\tau) = 0$ and $\mathrm{Var}(\boldsymbol{e}_\tau) = \boldsymbol{\Psi}_\tau$, it is well known that $E(\boldsymbol{e}_\tau'\boldsymbol{H}_\tau(v)\boldsymbol{e}_\tau) = tr(\boldsymbol{\Psi}_\tau\boldsymbol{H}_\tau(v))$ and $\mathrm{Var}(\boldsymbol{e}_\tau'\boldsymbol{H}_\tau(v)\boldsymbol{e}_\tau) = 2\ tr(\boldsymbol{\Psi}_\tau\boldsymbol{H}_\tau(v)\boldsymbol{\Psi}_\tau\boldsymbol{H}_\tau(v))$. By Lemma 1.4, we know $\mathrm{Var}(\boldsymbol{e}_\tau'\boldsymbol{H}_\tau(v)\boldsymbol{e}_\tau) \le (r\Upsilon)^2$. Given any $\varepsilon > 0$, by Chebyshev's inequality we have,

$$
\begin{aligned}
& P\left\{\max_{v\in V-V^c}\left|\frac{tr(\boldsymbol{\Psi}_\tau\boldsymbol{H}_\tau(v)) - \boldsymbol{e}_\tau'\boldsymbol{H}_\tau(v)\boldsymbol{e}_\tau}{\tau R_\tau(v)}\right| > \varepsilon\right\} \\
\le & \sum_{v\in V-V^c}\frac{E[tr(\boldsymbol{\Psi}_\tau\boldsymbol{H}_\tau(v)) - \boldsymbol{e}_\tau'\boldsymbol{H}_\tau(v)\boldsymbol{e}_\tau]^2}{[\tau R_\tau(v)\varepsilon]^2} \\
= & \sum_{v\in V-V^c}\frac{\mathrm{Var}(\boldsymbol{e}_\tau'\boldsymbol{H}_\tau(v)\boldsymbol{e}_\tau)}{[\tau R_\tau(v)\varepsilon]^2} \\
\le & \left(\frac{m\Upsilon}{\varepsilon}\right)^2\sum_{v\in V-V^c}\frac{1}{(\tau R_\tau(v))^2}\ .
\end{aligned}
$$

Since $R_\tau(v) > 0$, the last term goes to zero under conditions (1.17). We thus obtain (1.34).

To prove (1.35), we note that both $tr(\boldsymbol{\Psi}_\tau\boldsymbol{H}_\tau(v))$ and $tr(\hat{\boldsymbol{\Psi}}_\tau\boldsymbol{H}_\tau(v))$ are bounded. Then since $R_\tau(v) \ge \Delta_\tau(v)$, we know (1.35) holds under condition (1.17).

Finally, (1.36) is equivalent to (1.34) since

$$
\left|\frac{L_\tau(v)}{R_\tau(v)} - 1\right| = \frac{|L_\tau(v) - R_\tau(v)|}{R_\tau(v)} = \frac{|\boldsymbol{e}_\tau'\boldsymbol{H}_\tau(v)\boldsymbol{e}_\tau - tr(\boldsymbol{\Psi}_\tau\boldsymbol{H}_\tau(v))|}{\tau R_\tau(v)}\ .
$$

We thus conclude the proof of equation (1.32).

Next, we show the asymptotic loss efficiency and consistency of the EGIC minimizer $\check{v}_\tau$, using (1.31) and (1.32). When V^c is empty, we know from (1.32) that the minimizer of $\Phi_\tau(v)$, $\check{v}_\tau$, is asymptotically equal to the minimizer of $L_\tau(v)$, that is, $\check{v}_\tau$ is asymptotic loss efficiency.

When V^c is not empty, we can show that for any $v^c \in V^c$,

$$
\Phi_\tau(v^c) - \frac{||\boldsymbol{e}_\tau||^2}{\tau} = o_p(L_\tau(v)) \tag{1.38}
$$

uniformly in $v \in V - V^c$, using similar arguments in proof of (1.35) and (1.34). Equation (1.38) together with equation (1.32) implies that $\check{v}_\tau$ will always belong to V^c asymptotically if V^c is not empty.

We can further prove, using similar arguments in proof of (1.34),

$$\max_{v \in V^c} \frac{e_\tau' H_\tau(v) e_\tau}{\lambda_\tau tr(\hat{\Psi}_\tau H_\tau(v))} \to_p 0 \ .$$

Then we know from (1.31), for $v^c \in V^c$, $\Phi_\tau(v) - ||e_\tau||^2/\tau$ is asymptotically dominated by the term $\lambda_\tau tr(\hat{\Psi}_\tau H_\tau(v))/\tau$. Under the assumption that $\hat{\Psi}_\tau$ is a consistent estimator of Ψ_τ and that $tr(\Psi_\tau H_\tau(v))$ converges to a finite limit as $\tau \to \infty$ for any $v \in V^c$, the dominating term $\lambda_\tau tr(\hat{\Psi}_\tau H_\tau(v))/\tau$ has the same minimizer as $L_\tau(v) = e_\tau' H_\tau(v) e_\tau/\tau$ asymptotically. Therefore we obtain

$$P\{\check{v}_\tau \in V^c \text{ but } \check{v}_\tau \neq v_\tau^L\} \to 0 \ , \tag{1.39}$$

which means that $\check{v}_\tau$ is asymptotically loss efficient when V^c is not empty.

Equation (1.39) also implies that $P\{\check{v}_\tau = v_\tau^L\} \to 1$ when V^c is not empty. We thus conclude that $\check{v}_\tau$ is consistent. □

1.7 Summary and Discussion

In this paper, we point out that building a tracking portfolio for a target stock is equivalent to selecting variables in linear regression models with linearly constrained coefficients. We develop a procedure to build an optimal tracking portfolio by extending the Generalized Information Criterion (GIC) to constrained linear regression models with independent observations. We also extend the GIC to constrained linear regression models with errors following a stationary Gaussian process. Under mild conditions, the extended GIC is proved to be asymptotically loss efficient with respect to the average squared error loss, and furthermore, consistent when a true model exists.

A simulation study is carried out to evaluate the performance of the GIC procedure in finite samples. The simulation shows that the percentage of selecting the correct model is increasing and close to one when the sample size increases.

The results also indicate that the square-root penalty rule performs better than the logarithm penalty rule for moderate and large sample sizes.

The GIC selection criterion is applied to building an optimal tracking portfolio for measuring long-term post-event abnormal stock return. We compare the GIC method with two other methods in measuring abnormal returns of 200 randomly selected firms that are expected to have zero abnormal return. Our results show that the GIC method outperforms the other two methods.

In this paper, the extended Generalized Information Criterion (EGIC) for dependent observations has not been applied to monthly returns of individual stocks in both the simulation and empirical analyses. In fact, it is well documented in literature that monthly returns of individual stocks have insignificant autocorrelation while daily returns appear to be negatively autocorrelated, see, e.g., Campbell, Lo and MacKinlay (1997, Chapter 2) and references therein. In studies on building optimal tracking portfolios to track daily movements in a chosen financial index, EGIC shall be employed. In the context of tracking financial indices, choice of penalties in the selection criterion can be empirically investigated. For instance, we may construct two index funds based on the logarithm penalty rule and the square-root penalty rule. The performance of the two index funds can be compared according to how closely each fund mimics the target index in a given time period. Another direction of future research is to apply the GIC procedure to measure the long-term post-event abnormal returns of a sample of firms that actually experienced a specific event.

CHAPTER 2

RISK MANAGEMENT, DISPERSED INFORMATION, AND INTERNAL MARKET

2.1 Introduction

Risk management has become an indispensable task in large corporations that demands extensive firm-wide effort. According to Brown (2000), HDG Inc., pseudonym of a large multinational corporation, devotes approximately 11 full-time employees to foreign exchange risk management: four in US-based foreign exchange group, two regional treasury managers, one senior management, two in treasury accounting, and two in support group. The Foreign Exchange Management Committee in HDG is composed of high rank officers, including the Chief Financial Officer, Corporate Controller, Treasurer, regional Vice-Presidents (America, Asia-Pacific, Europe, Japan), and the Manager of Foreign Exchange. As so many parties across the organizational chart in a corporation are involved in risk management, a risk management program that optimally organizes relevant parties' activities is in demand.

An unorganized or ill-organized risk management program may result in two undesirable consequences: underhedging or overhedging. In the case of underhedging, the corporation remains exposed to the down side of certain risk factors, which will cause shareholders to suffer serious loss in value if the exposed risk factors happen to take their down sides in future. In the case of overhedging, the corporation buys unnecessary exposure to the up side of certain risk factors, which reduces shareholders' value due to transaction costs in financial markets.[1] An optimal risk management program should be effective in controlling both underhedging and overhedging.

A risk management program faces two primary issues. First, it needs to identify who makes hedging decisions. A multidivisional corporation either lets divisions manage risk at the division level or lets the corporate headquarters make decisions at the corporate level. The choice between decentralization and centralization determines roles and responsibilities of each party involved in risk management. Second, a risk management program should be effective in helping the corporation fully use its internal fund. Since external fund is costly, a corporation maximizes its value by making best use of internal fund. The difficulty a multidivisional corporation faces in using internal fund is that no one in the corporation knows precisely the sum of internal fund the corporation as a whole will have. Performance of a risk management program crucially depends on how well it overcomes the difficulty of gathering information on internal fund.

Existing risk management programs in large corporations fall into two categories according to a survey by the Financial Executives Research Foundation, which is documented in Davis and Militello (1995). Companies in the first category, including General Electric, Mobil, Union Carbide, etc., delegate hedging decisions to divisions. Companies in the second category, including Eli Lilly, Applied Material, etc., make hedging decisions at the corporate level. Even though the survey does not reveal how corporations gather information on internal fund, managers being surveyed consider it an important issue. This is evident in the following principal concerns they have expressed: how to gather exposure data, how to ensure accuracy of exposure data,

[1] The total transaction cost a corporation spends on hedging nowadays is a noticeable sum. Brown (2000) estimates that HDG Inc. spends roughly \$2.3 million on transaction costs annually for managing its foreign exchange risk. Since HDG has exposure to other risks as well, including interest rate risk, operational risk, counterparty credit risk, etc., the total transaction costs HDG spends on managing risk is certainly greater than \$2.3 million.

and how to establish controls and accountability (Page 7 in Davis and Militello (1995)).

In this paper, we set up a theoretical framework to analyze the problem of designing a risk management program. We treat risk management in a more general sense that corporations use hedging not only to avoid financial distress but also to secure financing for future investments. We demonstrate that neither of the two existing categories of risk management program is optimal; they either underhedge or overhedge. We propose a third kind of risk management program and show that it is better than existing ones.

The rest of this paper is organized as follows. In Section 2.2, we state the problem of designing a risk management program at a multidivisional corporation. In Section 2.3, we formulate a theoretical framework under which performance of four risk management programs are studied. Section 2.4 concludes the paper with summary and discussion.

2.2 Problem Description

In this section, we state the problem of designing a risk management program. Our model concerns corporate decisions at two time spots, labeled as time 0 and time 1 respectively. Time 0 is present while time 1 is future. Throughout the paper, we assume the discount rate between time 0 and time 1 is zero. At time 1, the world is in one and only one of N possible states, which are indexed by $\{1, 2, \cdots, N\}$. We assume that there exists a risk neutral probability distribution $\boldsymbol{p} = (p_1, p_2, \cdots, p_N)'$ of these states, where p_k is the probability that the world is in the kth state at time 1.

We consider a multidivisional corporation with M divisions. Let $\boldsymbol{v}_j$ be division j's operating income at time 1 for $j = 1, 2, \cdots, M$. The distribution of $\boldsymbol{v}_j$ is given by

$$\boldsymbol{v}_j = \begin{cases} v_{j1} & \text{if the world is in state 1;} \\ v_{j2} & \text{if the world is in state 2;} \\ \vdots & \qquad\qquad \vdots \\ v_{jN} & \text{if the world is in state N,} \end{cases}$$

We assume that the optimal investment of division j at time 1 is η_{jk} if the world is in state k.[2] Let $\boldsymbol{\eta}_j = (\eta_{j1}, \eta_{j2}, \cdots, \eta_{jN})'$ be division j's optimal investment for every state of time 1. A division has two sources of funds to finance its investment at time 1: operating income from existing production projects and funds acquired from external financial markets. The former is called internal fund and the latter external fund. For example, if the world is in state k at time 1, then division j's operating income is equal to v_{jk} and its optimal investment is in the amount of η_{jk}. If the operating income is greater than optimal investment, i.e., $v_{jk} > \eta_{jk}$, the division is able to finance its investment via internal fund only; otherwise, the division has to acquire external fund. Since external fund is costly, a division always prefers to have sufficient internal fund in every state. Unfortunately, the distribution of operating income does not always match that of optimal investment. While a division has greater operating income than optimal investment in some states of the world, it has less in other states. To finance investment using as much of its internal fund as possible, a division needs to shift operating income across states so as to match investment needs.

We assume that a complete financial market exists that enables divisions to shift operating income across states of the world.[3] In a complete financial markets, there are N primitive Arrow securities, one for each state of the world (Arrow (1964)). The

[2]The optimal investment is exogenously determined. It captures essence of existing theories on why firms hedge. Firms that hedge to reduce the volatility of its taxable income (Smith and Stulz (1985)) will prefer $\eta_{j1} = \eta_{j2} = \cdots = \eta_{jN}$. Firms that hedge to reduce the probability of bankruptcy or financial distress (Stulz (1996), Ross (1997) and Leland (1998)) will prefer $\eta_{jk} \geq d_{jk}$ for $k = 1, 2, \cdots, N$, where d_{jk} equals a firm's anticipated financial obligation in state k. Firms that hedge to avoid raising funds in external capital markets (Froot, Scharfstein and Stein (1993)) will prefer large η_{jk} in state k when attractive investment opportunities are likely to exist but external funds are hard to get. Firms that hedge to prove their superior investment projects or management ability to the public (DeMarzo and Duffie (1991), Breeden and Viswananthan (1996)) will prefer large η_{jk} in state k when most of their competitors do poorly.

[3]The assumption of a complete financial market is more plausible now than 20 years ago because of recent innovations in financial products and advances in financial engineering. See Merton (1992) for a review of how financial innovations improve economic performance as well as complete the financial markets.

Arrow security of state k, also called the ***kth Arrow security***, is a contract between buyer and seller, which entitles the buyer the right to receive one unit of money from the seller if the world is in state k at time 1 and nothing if in any other state. The seller then receives a fixed price at time 0 for being willing to give one unit of money to the buyer should the world be in state k at time 1. Under the risk neutral probability distribution $\boldsymbol{p}$, the risk neutral equilibrium price of a kth Arrow security is equal to the probability of the kth state occurring, p_k. However, because there exist transaction costs in real-world financial markets, actual transaction prices are different from the theoretical equilibrium price. Let b_k be the buy price a buyer pays and s_k the sell price a seller receives for one kth Arrow security in the financial market for $k = 1, 2, \cdots, N$. Because of transaction costs, the following relationship generally holds in real-world financial markets

$$b_k > p_k > s_k > 0, \quad \text{for} \quad k = 1, 2, \cdots, N \ .$$

Sources of transaction cost include cost of maintaining intermediacy by financial intermediaries, cost of credit risk, to name a few.

The following example illustrates how operating income of time 1 is shifted across states via Arrow security. Suppose a division buys one kth Arrow security at the price of b_k at time 0. Then the division will receive one unit of money if the world is in state k at time 1 and zero if in any other state. Consequently, the division's disposable internal fund increases by the amount of $1 - b_k$ in state k and decreases by the amount of b_k in any other state. The example suggests that a division whose operating income is less than its optimal investment in state k at time 1 can raise its disposable internal fund of state k by buying a sufficient number of the kth Arrow security. The division acquires external fund at higher cost only when it faces shortage of internal fund in some states no matter how it shifts operating income.

In a multidivisional corporation, since divisions are generally in different lines of business, involving different processes, products, customers, geographical locations, etc., their operating incomes are not perfectly correlated with each other. In any state of the world, some divisions may be short of internal fund while other divisions may have more internal fund than they need. Divisions of the former type are called deficit divisions and those of the latter type are surplus divisions. The potential that one division's surplus offsets other divisions' deficit makes it attractive to let the corporate headquarters make hedging and financing decisions based on consolidated operating income and optimal investment of the whole corporation. Let $\boldsymbol{u} = (u_1, u_2, \cdots, u_N)$ be the corporate operating income of time 1, which is the sum of all divisional operating incomes, i.e., $u_k = v_{1k} + v_{2k} + \cdots + v_{Mk}$ for $k = 1, 2, \cdots, N$. Let $\boldsymbol{\xi} = (\xi_1, \xi_2, \cdots, \xi_N)'$ be the corporate optimal investment of time 1, which is the sum of all divisional optimal investments, i.e., $\xi_k = \eta_{1k} + \eta_{2k} + \cdots + \eta_{Mk}$ for $k = 1, 2, \cdots, N$.

To make central decisions, headquarters needs to know corporate operating income and corporate optimal investment. Unfortunately, exact figures of $\boldsymbol{u}$ and $\boldsymbol{\xi}$ are not readily available to headquarters. In a multidivisional corporation, since only divisions know their operating income and optimal investment, headquarters has to rely on divisions' reports to estimate $\boldsymbol{u}$ and $\boldsymbol{\xi}$. The estimated figures are generally imprecise due to information loss in the reporting process and divisions' incentives to report biased figures in their favor. We will show that bias in the estimated figures reduces benefits of centralization in managing risk in Section 2.3.3.

In next section, we present a theoretical formulation of the problem of making optimal hedging and financing decisions and examine performance of four organizational programs that a multidivisional corporation can adopt to solve the problem.

2.3 Theoretical Model and Analysis

In the first subsection, we study a centralization program for risk management under the ideal assumption that the corporate headquarters has perfect information regarding corporate operating income $\boldsymbol{u}$ and corporate optimal investment $\boldsymbol{\xi}$. We then examine three other practical risk management programs in following subsections and compare them with the ideal centralization program.

2.3.1 Ideal Centralization

In this subsection, we make the ideal assumption that the corporate headquarters knows corporate operating income $\boldsymbol{u}$ and corporate optimal investment $\boldsymbol{\xi}$. The corporation's objective is to maximize its expected net present value.

The corporation's value at time 1 has four components: corporate operating income, cash flow resulting from hedging transactions, expenses of acquiring external fund, and expected gain of investment made at time 1. Suppose the world is in state k at time 1. The *first component* is operating income, equal to u_k. To explicitly write down the *second component*, let q_k^b and q_k^s be quantities of the kth Arrow security the corporation buys and sells in the financial market at time 0, respectively. Both q_k^s and q_k^b take only nonnegative values. By definition of Arrow security, the corporation's buying and selling positions in the kth Arrow security generates cash flow of $q_k^b - q_k^s$ in state k at time 1. In addition, the second component also includes $\sum_{i=1}^N(-b_i q_i^b + s_i q_i^s)$, where $\sum_{i=1}^N -b_i q_i^b$ is what the corporation pays at time 0 for buying Arrow securities and $\sum_{i=1}^N s_i q_i^s$ is what the corporation receives at time 0 for selling Arrow securities. The *second component* of the corporation's net value is thus equal to $-q_k^s + \sum_{i=1}^N s_i q_i^s + q_k^b - \sum_{i=1}^N b_i q_i^b$. Let z_k be the amount of external fund the corporation acquires in state k and c_k be the unit cost of external fund in state k. The *third component* is then equal to $c_k z_k$. Suppose that the corporation makes optimal investment ξ_k in state k. Let r_k be the expected discounted cash flow from the future beyond time 1 that is a result of one unit investment made in state k at time 1. Then the *fourth component*, the expected gain of investment made at time 1, is equal to $r_k \xi_k$. Putting the four components together, we know that the corporation's value in state k is equal to

$$u_k + q_k^b - \sum_{i=1}^{N} b_k q_k^b - q_k^s + \sum_{i=1}^{N} s_k q_k^s - c_k z_k + r_k \xi_k \ . \tag{2.1}$$

Let $\boldsymbol{q}^b = (q_1^b, q_2^b, \cdots, q_N^b)'$ be the vector containing quantities of Arrow securities the corporation buys for every state, $\boldsymbol{q}^s = (q_1^s, q_2^s, \cdots, q_N^s)'$ be the vector containing quantities of Arrow securities the corporation sells for every state, and $\boldsymbol{z} = (z_1, z_2, \cdots, z_N)'$ be the vector containing external fund the corporation acquires for every state. The corporation's expected net present value at time 0 is a function of $\boldsymbol{q}^b$, $\boldsymbol{q}^s$, and $\boldsymbol{z}$. Let $f(\boldsymbol{q}^b, \boldsymbol{q}^s, \boldsymbol{z})$ be the corporation's expected net present value at time 0. By taking expectation of (2.1) under the risk neutral probability distribution, we obtain $f(\boldsymbol{q}^b, \boldsymbol{q}^s, \boldsymbol{z})$ as follows

$$\begin{aligned}
f(\boldsymbol{q}^b, \boldsymbol{q}^s, \boldsymbol{z}) &\equiv \sum_{k=1}^{N} p_k \left[u_k + q_k^b - \sum_{i=1}^{N} b_i q_i^b - q_k^s + \sum_{i=1}^{N} s_i q_i^s - c_k z_k + r_k \xi_k \right] \\
&= \sum_{k=1}^{N} p_k u_k + \sum_{k=1}^{N} p_k q_k^b - \left(\sum_{k=1}^{N} p_k \right) \left(\sum_{i=1}^{N} b_i q_i^b \right) \\
&\quad - \sum_{k=1}^{N} p_k q_k^s + \left(\sum_{k=1}^{N} p_k \right) \left(\sum_{i=1}^{N} s_i q_i^s \right) - \sum_{k=1}^{N} p_k c_k z_k + \sum_{k=1}^{N} p_k r_k \xi_k \\
&= \mathrm{E}(\boldsymbol{u}) + \sum_{k=1}^{N} p_k r_k \xi_k - \sum_{k=1}^{N} \left[(b_k - p_k) q_k^b + (p_k - s_k) q_k^s + p_k c_k z_k \right] \quad (2.2)
\end{aligned}$$

To maximize its expected net present value, the corporation solves the following optimization problem,

$$\text{maximize} \quad f(\boldsymbol{q}^b, \boldsymbol{q}^s, \boldsymbol{z}) \tag{2.3a}$$

subject to

$$\begin{aligned}
u_1 + q_1^b - \sum_{i=1}^{N} b_i q_i^b - q_1^s + \sum_{i=1}^{N} s_i q_i^s + z_1 &\geq \xi_1 \\
u_2 + q_2^b - \sum_{i=1}^{N} b_i q_i^b - q_2^s + \sum_{i=1}^{N} s_i q_i^s + z_2 &\geq \xi_2 \\
&\vdots
\end{aligned} \tag{2.3b}$$

$$u_N + q_N^b - \sum_{i=1}^{N} b_i q_i^b - q_N^s + \sum_{i=1}^{N} s_i q_i^s + z_N \geq \xi_N$$

$$f(\boldsymbol{q}^b, \boldsymbol{q}^s, \boldsymbol{z}) \geq \mathrm{E}(\boldsymbol{u}) \tag{2.3c}$$

$$\boldsymbol{q}^b \geq \mathbf{0}, \quad \boldsymbol{q}^s \geq \mathbf{0}, \quad \boldsymbol{z} \geq \mathbf{0}\ . \tag{2.3d}$$

Constraint (2.3b) is a set of budget constraints, one for each state of the world at time 1. Take the budget constraint of state 1 as an example. On the left hand side is the total amount of disposable fund the corporation has in state 1, which is the sum of operating income, cash flow from hedging positions, and external fund. On the right hand side is the corporate optimal investment. The constraint requires that the left hand side be greater than or equal to the right hand side so that the corporation has enough disposable fund to make optimal investment in state 1. Constraint (2.3c) is a rational constraint, where $\mathrm{E}(\boldsymbol{u})$ is the expected value of the corporate operating income. If the corporation does not make any investment at time 1, its expected value will be $\mathrm{E}(\boldsymbol{u})$.[4] We call $\mathrm{E}(\boldsymbol{u})$ the reservation value of the corporation. The rational constraint (2.3c) emphasizes that a rational corporation does not take extra effort to reduce its value. Constraint (2.3d) requires all choice variables to be nonnegative.

Since the objective function and all constraints are linear functions of choice variables, the optimization problem (2.3) is a linear programming problem. Let $\Omega = \{(\boldsymbol{q}^b, \boldsymbol{q}^s, \boldsymbol{z}) : \boldsymbol{q}^b, \boldsymbol{q}^s,$ and $\boldsymbol{z}$ satisfy constraints (2.3b), (2.3c) and (2.3d).$\}$ be the collection of feasible solutions. It is easy to verify that constraint (2.3c) is equivalent to the following inequality

$$\sum_{k=1}^{N} \Big[(b_k - p_k)q_k^b + (p_k - s_k)q_k^s + p_k c_k z_k\Big] \leq \sum_{k=1}^{N} p_k r_k \xi_k\ . \tag{2.4}$$

Because $b_k > p_k > s_k > 0$ and $c_k > 0$, choice variables $\boldsymbol{q}^b$, $\boldsymbol{q}^s$, and $\boldsymbol{z}$ are bounded from above according to inequality (2.4). In addition, since $\boldsymbol{q}^b$, $\boldsymbol{q}^s$, and $\boldsymbol{z}$ are nonnegative the feasible solution space Ω is thus bounded. Furthermore, equation (2.2) shows

[4] If the corporation does not make any investment at time 1, it will not take any hedging position at time 0 nor acquire any external fund at time 1.

that the objective function is decreasing in choice variables $(\boldsymbol{q}^b, \boldsymbol{q}^s, \boldsymbol{z})$, and that the maximum value of the objective function is equal to $\mathrm{E}(\boldsymbol{u}) + \sum_{i=1}^{N} p_k r_k \xi_k$ when $\boldsymbol{q}^b = \boldsymbol{q}^s = \boldsymbol{z} = \mathbf{0}$. But the solution $\boldsymbol{q}^b = \boldsymbol{q}^s = \boldsymbol{z} = \mathbf{0}$ is not necessarily a feasible solution because it may not satisfy constraint (2.3b).

In this paper, we assume that Ω is not empty. The assumption means that the corporation can do better with investing at time 1 than without. It is well known that when the feasible solution space is bounded and not empty, there exists an optimal feasible solution to a linear programming problem. The following proposition gives certain properties of an optimal feasible solution to the linear programming problem (2.3).

Proposition 2.1 *Let* $(\boldsymbol{q}^{*b}, \boldsymbol{q}^{*s}, \boldsymbol{z}^*) = (q_1^{*b}, \cdots, q_N^{*b}; q_1^{*s}, \cdots, q_N^{*s}; z_1^*, \cdots z_N^*)$ *be an optimal feasible solution. The optimal feasible solution has the following properties.*

1. *The optimal hedging positions* q_k^{*b} *and* q_k^{*s} *can not be both positive in any state* k.
2. *If* $q_k^{*b} > 0$ *or* $z_k^* > 0$, *then the budget constraint of state* k *in (2.3b) is binding.*
3. *If* $q_k^{*s} > 0$ *for a certain state* k, *then there must exist another state* l *whose budget constraint is binding.*
4. *If* $z_k^* > 0$ *for a certain state* k *and the inequality,* $p_k c_k > (p_l - s_l)/s_l$, *holds for any state* $l = 1, 2, \cdots, N$, *then all budget constraints is binding.*
5. *If none of the* N *budget constraints is binding, then* $q_k^{*s} = q_k^{*b} = z_k^* = 0$ *for all* $k = 1, 2, \cdots, N$.

Remark 2.1 The *first property* suggests that, to maximize its expected net present value, the corporation either buys or sells Arrow security of a particular state, but does not engage in both at the same time. Buying one kth Arrow security effectively shifts operating income from all other states to state k, while selling one kth Arrow security effectively distributes operating income of state k evenly to all states. Since both buying and selling incur transaction cost, it is not optimal for the corporation to increase its disposable fund of state k by buying the kth Arrow security and then shift fund from state k to some deficit states by selling the kth Arrow security. The *second property* suggests that the corporation buys the kth Arrow security or acquires external fund in state k only when it could not meet the optimal investment of state k otherwise. The *third property* suggests that the corporation sells the kth Arrow security only when it could not meet the optimal investment in some other states otherwise. In the condition of the *fourth property*, $p_k c_k$ is the cost of one unit external fund acquired in state k, and $(p_l - s_l)/s_l$ measures the proportion of transaction cost incurred in selling one lth Arrow security to the sell price. Since the cost of external fund is usually much higher than transaction cost, the condition is satisfied for any state k and l. The *fourth property* implies that, under the condition that the cost of external fund is higher than the cost of shifting operating income across states via Arrow securities, the corporation will acquire external fund only when its operating income is not enough to meet its optimal investments no matter how the corporation shifts operating income.

Proof of Proposition 2.1:

Let w_t^* be the corporation's disposable fund in state t under the optimal feasible solution $(\boldsymbol{q}^{*b}, \boldsymbol{q}^{*s}, \boldsymbol{z}^*)$, i.e.,

$$w_t^* \equiv u_t + q_t^{*b} - \sum_{i=1}^{N} b_i q_i^{*b} - q_t^{*s} + \sum_{i=1}^{N} s_i q_i^{*s} + z_t^* \ .$$

Let $f(\boldsymbol{q}^{*b}, \boldsymbol{q}^{*s}, \boldsymbol{z}^*)$ be the corporation's optimal expected net present value.

To prove the first property, suppose $q_k^{*s} > 0$ for state k. We will show that if $q_k^{*b} > 0$ then the solution is not optimal. Suppose $q_k^{*b} > 0$. Let $\Delta = \min\{q_k^{*s}, q_k^{*b}\}$, then $\Delta > 0$. If the corporation decreases both q_k^{*s} and q_k^{*b} by Δ then the corporation's disposable fund in any state t becomes $w_t^* + (b_k - s_k)\Delta$ and the corporation expected net present value is equal to $f(\boldsymbol{q}^{*b}, \boldsymbol{q}^{*s}, \boldsymbol{z}^*) + (b_k - s_k)\Delta$. Since $b_k > s_k$, we thus obtain a better solution by substituting $q_k^{*s} - \Delta$ for q_k^{*s} and $q_k^{*b} - \Delta$ for q_k^{*b}. This contradicts to the assumption that q_k^{*s} and q_k^{*b} are optimal. Therefore $q_k^{*b} = 0$ when $q_k^{*s} > 0$. It also implies that $q_k^{*s} = 0$ when $q_k^{*b} > 0$. We thus prove that q_k^{*b} and q_k^{*s} can not both be positive in any state k.

Next, we shall prove the second property. Suppose $q_k^{*b} > 0$ for any state k. Suppose the budget constraint of state k is not binding, i.e., $w_k^* > \xi_k$. Let $\Delta = (\xi_k - w_k^*)/2$, then $\Delta > 0$. By substituting $q_k^{*b} - \Delta$ for q_k^{*b} in the optimal solution, we reduce the disposable fund of state k from w_k^* to $w_k^* - (1 - b_k)\Delta$ and increase the disposable fund of any other state t from w_t^* to $w_t^* + b_k\Delta$. Consequently, budget constraints of all states remain satisfied. More importantly, the substitution increases the objective function by $b_k\Delta$. It means that the substitution results in a better solution, which contradicts to the assumption that q_k^{*b} is optimal. Therefore the budget constraint of state k is binding. Similarly, we can prove that if $z_k^* > 0$ then the budget constraint of state k is binding.

Next, we shall prove the third property. Suppose $q_k^{*s} > 0$ for any state k. We want to show that there must be another state whose budget constraint is binding. Suppose budget constraints of all states other than state k are not binding, i.e., $w_l^* > \xi_l$ for any state l where $l \neq k$. Let $\Delta = \min\{w_l^* - \xi_l : l = 1, 2, \cdots, N, \text{ and } l \neq k\}$, then $\Delta > 0$. By substituting $q_k^{*s} - \Delta$ for q_k^{*s} in the optimal solution, we increase the disposable fund of state k from w_k^* to $w_k^* + (1 - s_k)\Delta$ and decrease the disposable fund of any other state t from w_t^* to $w_t^* + s_k\Delta$. By definition of Δ, budget constraints of all states remain satisfied. More importantly, the substitution also increases the objective function by $s_k\Delta$. It means that the substitution results in a better solution, which contradicts to the assumption that q_k^{*s} is optimal. Therefore there must exist a state whose budget constraint is binding.

Next, we shall prove the fourth property. Suppose $z_k^* > 0$ for state k. According to the second property, the budget constraint of state k is binding. Suppose the budget constraint of another state l is not binding, i.e., $w_l^* > \xi_l$. Let $\Delta = (w_l^* - \xi_l)/2$, then $\Delta > 0$. We then substitute $q_l^{*s} + \Delta$ for q_l^{*s} and $z_k^* - s_l\Delta$ for z_k^* in the optimal solution. With the substitution, the disposable fund of state k does not change, the disposable fund of state l is still greater than ξ_l, and the disposable fund of any other state increases by $s_l\Delta$. Consequently, budget constraints are still satisfied for all states. The value of the objective function after the substitution is equal to $f(\boldsymbol{q}^{*b}, \boldsymbol{q}^{*s}, \boldsymbol{z}^*) + p_k c_k s_l \Delta - (p_l - s_l)\Delta$, where $p_k c_k s_l \Delta$ is due to change in z_k^* and $-(p_l - s_l)\Delta$ is due to change in q_l^{*s}. Under the assumption that $p_k c_k > (p_k - s_k)/s_l$, the substitution results in a better solution, which contradicts to the assumption that $(\boldsymbol{q}^{*b}, \boldsymbol{q}^{*s}, \boldsymbol{z}^*)$ is optimal. Therefore the budget constraint of state l is binding.

At last, the fifth property can be derived from the second and the third properties. If none of the budget constraint is binding, then we know that, for any state k, $q_k^{*b} = z_k^* = 0$ according to the third property and $q_k^{*s} = 0$ according to the second property. □

In summary of this section, we set up a theoretical model in which the corporate headquarters solves the linear programming problem (2.3) to obtain optimal hedging and financing decisions, assuming that headquarters knows corporate operating income $\boldsymbol{u}$ and corporate optimal investment $\boldsymbol{\xi}$. Properties of optimal decisions are studied and summarized in Proposition (2.1). In the following subsections, we discuss three risk management programs that operate under the reality that each division knows its own operating income and optimal investment while no one in the corporation has perfect information about corporate operating income and corporate optimal investment.

2.3.2 Uncoordinated Decentralization

In this subsection, we study an uncoordinated decentralization program for risk management. Under such a program, divisions make hedging and financing decisions individually. Each division solves its own optimization problem as if it is a stand-alone corporation.

Take division j as an example. Let $\boldsymbol{q}_j^b = (q_{j1}^b, q_{j2}^b, \cdots, q_{jN}^b)'$ be the vector containing quantities of Arrow securities division j buys for every state, $\boldsymbol{q}_j^s = (q_{j1}^s, q_{j2}^s, \cdots, q_{jN}^s)'$ be the vector containing quantities of Arrow securities division j sells for every state, and $\boldsymbol{z}_j = (z_{j1}, z_{j2}, \cdots, z_{jN})'$ be the vector containing external fund division j acquires for every state. Then the expected net present value of division j at time 0, $f_j(\boldsymbol{q}_j^b, \boldsymbol{q}_j^s, \boldsymbol{z}_j)$, has the following expression, similar to equation (2.2),

$$f_j(\boldsymbol{q}_j^b, \boldsymbol{q}_j^s, \boldsymbol{z}_j) = \mathrm{E}(\boldsymbol{v}_j) + \sum_{k=1}^{N} p_k r_k \eta_{jk} - \sum_{k=1}^{N} \left[(b_k - p_k) q_{jk}^b + (p_k - s_k) q_{jk}^s + p_k c_k z_{jk}\right] \quad (2.5)$$

where $\boldsymbol{v}_j = (v_{j1}, v_{j2}, \cdots, v_{jN})'$ is division j's operating income at time 1, and $\boldsymbol{\eta}_j = (\eta_{j1}, \eta_{j2}, \cdots, \eta_{jN})'$ is division j's optimal investment at time 1. Note that we assume the buy price b_k and the sell price s_k of Arrow securities, the unit cost of external fund c_k, and the unit gain of investment r_k are the same to all divisions as well as the corporation.

Division j solves the following linear programming problem to maximize its expected net present value,

$$\text{maximize} \quad f_j(\boldsymbol{q}_j^b, \boldsymbol{q}_j^s, \boldsymbol{z}_j) \quad (2.6a)$$

subject to

$$\begin{aligned}
v_{j1} + q_{j1}^b - \sum_{i=1}^{N} b_i q_{ji}^b - q_{j1}^s + \sum_{i=1}^{N} s_i q_{ji}^s + z_{j1} &\geq \eta_{j1} \\
v_{j2} + q_{j2}^b - \sum_{i=1}^{N} b_i q_{ji}^b - q_{j2}^s + \sum_{i=1}^{N} s_i q_{ji}^s + z_{j2} &\geq \eta_{j2} \\
&\vdots
\end{aligned} \quad (2.6b)$$

$$v_{jN} + q^b_{jN} - \sum_{i=1}^{N} b_i q^b_{ji} - q^s_{jN} + \sum_{i=1}^{N} s_i q^s_{ji} + z_{jN} \geq \eta_{jN}$$

$$f_j(\boldsymbol{q}^b_j, \boldsymbol{q}^s_j, \boldsymbol{z}_j) \geq \mathrm{E}(\boldsymbol{v}_j) \tag{2.6c}$$

$$\boldsymbol{q}^b_j \geq \mathbf{0}, \quad \boldsymbol{q}^s_j \geq \mathbf{0}, \quad \boldsymbol{z}_j \geq \mathbf{0} \ . \tag{2.6d}$$

Let $\Omega_j = \{(\boldsymbol{q}^b_j, \boldsymbol{q}^s_j, \boldsymbol{z}_j) : \boldsymbol{q}^b_j, \boldsymbol{q}^s_j$, and $\boldsymbol{z}_j$ satisfy constraints (2.6b), (2.6c) and (2.6d)$\}$ be the feasible solution space to the optimization problem of division j. We assume that the feasible solution space Ω_j is not empty for any division j with $j = 1, 2, \cdots, M$. The assumption means that all divisions have profitable future investment opportunities on their own. Nonprofitable divisions have been cut out from the corporation.

The following proposition compares the uncoordinated decentralization program with the ideal centralization program.

Proposition 2.2 *Let $(\boldsymbol{q}^{*b}_j, \boldsymbol{q}^{*s}_j, \boldsymbol{z}^*_j)$ be an optimal solution to the optimization problem (2.6) of division j for $j = 1, 2, \cdots, M$, and $(\boldsymbol{q}^{*b}, \boldsymbol{q}^{*s}, \boldsymbol{z}^*)$ be an optimal solution to the optimization problem (2.3) of the corporation under the ideal centralization program. We then have*

$$\sum_{j=1}^{M} f_j(\boldsymbol{q}^{*b}_j, \boldsymbol{q}^{*s}_j, \boldsymbol{z}^*_j) \leq f(\boldsymbol{q}^{*b}, \boldsymbol{q}^{*s}, \boldsymbol{z}^*) \ .$$

Proof of Proposition 2.2:

Note that corporate operating income is the sum of divisional operating income, i.e., $\boldsymbol{u} = \sum_{j=1}^{M} \boldsymbol{v}_j$, and corporate investment is the sum of divisional investment, i.e., $\boldsymbol{\xi} = \sum_{j=1}^{M} \boldsymbol{\eta}_j$. We then have the following results

$$\begin{aligned} & \sum_{j=1}^{M} f_j(\boldsymbol{q}^{*b}_j, \boldsymbol{q}^{*s}_j, \boldsymbol{z}^*_j) \\ = \ & \sum_{j=1}^{M} \left\{ \mathrm{E}(\boldsymbol{v}_j) + \sum_{k=1}^{N} p_k r_k \eta_{jk} - \sum_{k=1}^{N} \left[(b_k - p_k) q^{*b}_{jk} + (p_k - s_k) q^{*s}_{jk} + p_k c_k z^*_{jk} \right] \right\} \end{aligned}$$

$$
\begin{aligned}
&= \mathrm{E}(\sum_{j=1}^{M} \boldsymbol{v}_j) + \sum_{k=1}^{N} (p_k r_k \sum_{j=1}^{M} \eta_{jk}) - \sum_{k=1}^{N} \sum_{j=1}^{M} \left[(b_k - p_k) q_{jk}^{*b} + (p_k - s_k) q_{jk}^{*s} + p_k c_k z_{jk}^{*} \right] \\
&= \mathrm{E}(\boldsymbol{u}) + \sum_{k=1}^{N} p_k r_k \xi_k - \sum_{k=1}^{N} \left[(b_k - p_k) \sum_{j=1}^{M} q_{jk}^{*b} + (p_k - s_k) \sum_{j=1}^{M} q_{jk}^{*s} + p_k c_k \sum_{j=1}^{M} z_{jk}^{*} \right] \\
&= f(\sum_{j=1}^{M} \boldsymbol{q}_j^{*b}, \sum_{j=1}^{M} \boldsymbol{q}_j^{*s}, \sum_{j=1}^{M} \boldsymbol{z}_j^{*}) \ .
\end{aligned}
$$

It is easy to verify that $(\sum_{j=1}^{M} \boldsymbol{q}_j^{*b}, \sum_{j=1}^{M} \boldsymbol{q}_j^{*s}, \sum_{j=1}^{M} \boldsymbol{z}_j^{*})$ satisfies constraints (2.3b), (2.3c), and (2.3d). Therefore $(\sum_{j=1}^{M} \boldsymbol{q}_j^{*b}, \sum_{j=1}^{M} \boldsymbol{q}_j^{*s}, \sum_{j=1}^{M} \boldsymbol{z}_j^{*})$ is a feasible solution to the optimization problem (2.3). Since $(\boldsymbol{q}^{*b}, \boldsymbol{q}^{*s}, \boldsymbol{z}^{*})$ is the optimal solution to problem (2.3), we obtain

$$
\sum_{j=1}^{M} f_j(\boldsymbol{q}_j^{*b}, \boldsymbol{q}_j^{*s}, \boldsymbol{z}_j^{*}) = f(\sum_{j=1}^{M} \boldsymbol{q}_j^{*b}, \sum_{j=1}^{M} \boldsymbol{q}_j^{*s}, \sum_{j=1}^{M} \boldsymbol{z}_j^{*}) \leq f(\boldsymbol{q}^{*b}, \boldsymbol{q}^{*s}, \boldsymbol{z}^{*}) \ .
$$

$\square$

Proposition 2.2 shows that the optimal corporate expected net present value under the uncoordinated decentralization program is no more than that under the ideal centralization program. From the proof of the proposition, $(\sum_{j=1}^{M} \boldsymbol{q}_j^{*b}, \sum_{j=1}^{M} \boldsymbol{q}_j^{*s}, \sum_{j=1}^{M} \boldsymbol{z}_j^{*})$ is the sum of all divisions' optimal hedging and financing positions under the uncoordinated decentralization program. A necessary condition for the uncoordinated decentralization program to achieve the same maximum value as the ideal centralization program is that $(\sum_{j=1}^{M} \boldsymbol{q}_j^{*b}, \sum_{j=1}^{M} \boldsymbol{q}_j^{*s}, \sum_{j=1}^{M} \boldsymbol{z}_j^{*})$ satisfies all properties in Proposition 2.1. Take Property 1 in Proposition 2.1 as an example. Property 1 demands that for any state k, if $\sum_{j=1}^{M} q_{jk}^{*b} > 0$, then $\sum_{j=1}^{M} q_{jk}^{*s} = 0$, and *vice versa*. However, since divisions' operating incomes are not correlated with each other in general, it is usually the case that, in any state of the world, some divisions are in surplus while others are in deficit. As a result, $\sum_{j=1}^{M} q_{jk}^{*b}$ and $\sum_{j=1}^{M} q_{jk}^{*s}$ are usually both positive for many states of the world, which means that the corporation as a whole buys more Arrow securities than necessary. That is, the corporation overhedges. Since the corporation pays transaction cost for each Arrow security it buys or sells, overhedging reduces the corporation's value due to excessive transaction costs.

2.3.3 Real-World Centralization

In Section 2.3.1, we study a centralized risk management program under the ideal assumption that the corporate headquarters knows exact figures of corporate operating income and corporate optimal investment. In reality, headquarters does not have first-hand knowledge of these figures and relies on divisions' reports to estimate them. The estimated figures can be imprecise for three reasons.

First, since headquarters is able to handle only a limited amount of information, it generally asks each division to report aggregate figures of its operating income and optimal investment. The restriction on the amount of information being transfered leads to loss of information, which then results in bias in estimated figures at headquarters.

Secondly, under central decision making, divisions have disincentives to collect information. When there is no direct reward for reporting correct information under centralization, it is not beneficial for divisions to engage in costly information collection. Instead, divisions tend to report handy figures, which may not be accurate at all.

Thirdly, even if divisions have accurate information, they have incentives to report biased figures. Divisions are managed by economic agents who maximize their own benefits. Since the reported figures determine how much contribution each division makes to corporate operating income and how much fund each division gets for future investment, divisions have incentives to report biased figures in their favor. In the following, we analyze how imprecise estimates affect performance of the centralization program.

Let $\bar{\boldsymbol{u}} = (\bar{u}_1, \bar{u}_2, \cdots, \bar{u}_N)'$ be headquarters' estimates of corporate operating income for every state of the world, and $\bar{\boldsymbol{\xi}} = (\bar{\xi}_1, \bar{\xi}_2, \cdots, \bar{\xi}_N)'$ be headquarters' estimates of optimal investment for every state of the world. The corporation's expected net present value with imprecise estimates, $\bar{f}(\bar{\boldsymbol{q}}^b, \bar{\boldsymbol{q}}^s, \bar{z})$, is then equal to

$$\bar{f}(\bar{\boldsymbol{q}}^b, \bar{\boldsymbol{q}}^s, \bar{\boldsymbol{z}}) = \mathrm{E}(\bar{\boldsymbol{u}}) + \sum_{k=1}^{N} p_k r_k \bar{\xi}_k - \sum_{k=1}^{N} \left[(b_k - p_k)\bar{q}_k^b + (p_k - s_k)\bar{q}_k^s + p_k c_k \bar{z}_k \right] , \quad (2.7)$$

where $\bar{\boldsymbol{q}}^b = (\bar{q}_1^b, \bar{q}_2^b, \cdots, \bar{q}_N^b)'$ is the vector containing quantities of state-contingent Arrow securities the corporation buys at time 0, $\bar{\boldsymbol{q}}^s = (\bar{q}_1^s, \bar{q}_2^s, \cdots, \bar{q}_N^s)'$ is the vector containing quantities of state-contingent Arrow securities the corporation sells at time 0, and $\bar{\boldsymbol{z}} = (\bar{z}_1, \bar{z}_2, \cdots, \bar{z}_N)'$ is the vector containing state-contingent external fund the corporation acquires at time 1.

To maximize its expected net present value, the corporation solves the following optimization problem,

$$\text{maximize} \quad \bar{f}(\bar{\boldsymbol{q}}^b, \bar{\boldsymbol{q}}^s, \bar{\boldsymbol{z}}) \quad (2.8a)$$

subject to

$$\begin{aligned}
\bar{u}_1 + \bar{q}_1^b - \sum_{i=1}^{N} b_i \bar{q}_i^b - \bar{q}_1^s + \sum_{i=1}^{N} s_i \bar{q}_i^s + \bar{z}_1 &\geq \bar{\xi}_1 \\
\bar{u}_2 + \bar{q}_2^b - \sum_{i=1}^{N} b_i \bar{q}_i^b - \bar{q}_2^s + \sum_{i=1}^{N} s_i \bar{q}_i^s + \bar{z}_2 &\geq \bar{\xi}_2 \\
&\vdots \qquad\qquad (2.8b) \\
\bar{u}_N + \bar{q}_N^b - \sum_{i=1}^{N} b_i \bar{q}_i^b - \bar{q}_N^s + \sum_{i=1}^{N} s_i \bar{q}_i^s + \bar{z}_N &\geq \bar{\xi}_N
\end{aligned}$$

$$\bar{f}(\bar{\boldsymbol{q}}^b, \bar{\boldsymbol{q}}^s, \bar{\boldsymbol{z}}) \geq \mathrm{E}(\bar{\boldsymbol{u}}) \quad (2.8c)$$

$$\bar{\boldsymbol{q}}^b \geq \mathbf{0}, \quad \bar{\boldsymbol{q}}^s \geq \mathbf{0}, \quad \bar{\boldsymbol{z}} \geq \mathbf{0} . \quad (2.8d)$$

The above optimization problem (2.8) is the same as the linear programming problem (2.3) except that the parameters $\bar{\boldsymbol{u}}$ and $\bar{\boldsymbol{\xi}}$ used in problem (2.8) are estimates of the parameters $\boldsymbol{u}$ and $\boldsymbol{\xi}$ used in problem (2.3). Let $(\bar{\boldsymbol{q}}^{*b}, \bar{\boldsymbol{q}}^{*s}, \bar{\boldsymbol{z}}^*)$ be an optimal solution to problem (2.8) and $(\boldsymbol{q}^{*b}, \boldsymbol{q}^{*s}, \boldsymbol{z}^*)$ be an optimal solution to problem (2.3). Below we discuss to what extent bias in the estimated parameters, $\bar{\boldsymbol{u}}$ and $\bar{\boldsymbol{\xi}}$, affects the corporation's hedging and financing decisions.

When the corporation has sufficient operating income to meet its optimal investment in every state of the world, small bias in estimates $\bar{\boldsymbol{u}}$ and $\bar{\boldsymbol{\xi}}$ does not have any impact. This is because when operating income is sufficient in every state, i.e. $u_k \geq \xi_k$ for all $k = 1, 2, \cdots, N$, the optimal solution to problem (2.3) is $\boldsymbol{q}^{*b} = \boldsymbol{q}^{*s} = \boldsymbol{z}^* = \boldsymbol{0}$. As long as the estimates $\bar{\boldsymbol{u}}$ and $\bar{\boldsymbol{\xi}}$ satisfy $\bar{u}_k \geq \bar{\xi}_k$ for all $k = 1, 2, \cdots, N$, the optimal solution to problem (2.8) will be the same $\bar{\boldsymbol{q}}^{*b} = \bar{\boldsymbol{q}}^{*s} = \bar{\boldsymbol{z}}^* = \boldsymbol{0}$.

However, when $(\bar{\boldsymbol{q}}^{*b}, \bar{\boldsymbol{q}}^{*s}, \bar{\boldsymbol{z}}^*)$ is not zero, estimation error in $\bar{\boldsymbol{u}}$ and $\bar{\boldsymbol{\xi}}$ reduces the corporation's value. According to Property 5 in Proposition 2.1, when $(\bar{\boldsymbol{q}}^{*b}, \bar{\boldsymbol{q}}^{*s}, \bar{\boldsymbol{z}}^*)$ is not zero, some inequality constraints in (2.8b) are binding. Suppose the constraint of state l is binding, i.e.,

$$\bar{u}_l + \bar{q}_l^{*b} - \sum_{i=1}^{N} b_i \bar{q}_i^{*b} - \bar{q}_l^{s} + \sum_{i=1}^{N} s_i \bar{q}_i^{*s} + \bar{z}_l^* = \bar{\xi}_l \ .$$

If $\bar{\xi}_l - \bar{u}_l < \xi_l - u_l$, then the corporation underhedges in state l. This is because when state l occurs at time 1, the corporation's disposable fund is equal to $u_l + \bar{q}_l^{*b} - \sum_{i=1}^{N} b_i \bar{q}_i^{*b} - \bar{q}_l^{s} + \sum_{i=1}^{N} s_i \bar{q}_i^{*s} + \bar{z}_l^*$, which is less than the optimal investment ξ_l. This forces the corporation to acquire additional external fund at time 1. It becomes even worse if $\bar{q}_l^{*s} > 0$, because the corporation then has to acquire external fund to pay holders of the lth Arrow securities it has sold at time 0. Underhedging reduces the corporation's value because of the high cost of obtaining external fund at time 1.

On the other hand, if $\bar{\xi}_l - \bar{u}_l \geq \xi_l - u_l$, then the corporation overhedges in state l, which means that the corporation has more disposable fund than it needs for investment. The corporation suffers loss in value if $\bar{q}_l^{*b} > 0$, because the corporation has bought more Arrow securities for state l than necessary. Since the corporation takes a charge in the amount of transaction cost for each Arrow security it buys, overhedging reduces the corporation's value by excessive transaction cost.

Although loss in the corporation's value due to bias in estimated operating income and optimal investment can not be written out in an explicit formula, it can be demonstrated numerically. Empirical evidence presented in Berger and Ofek (1994) shows that the market value of a multidivisional corporation appears to be approximately 13-15% less than the sum of its divisions valued separately.

They argue that this diversification discount results from misallocation of capital and inefficient cross-subsidies between divisions in the multidivisional corporation. Our analysis suggests that the existence of a diversification discount may be due to bias in information that the corporate headquarters uses in making centralized decisions. Even though headquarters has no intention for cross-subsidization of poorly-performing divisions by better-performing divisions, inefficient cross-subsidies take place due to inaccurate inputs to headquarters' decision-making process.

2.3.4 Coordinated Decentralization

In Section 2.3.2, an uncoordinated decentralization is discussed, under which each division acts as a stand-alone firm who deals only with external financial institutions for hedging and financing transactions, and the corporate headquarters offers no help in divisions' decision-making. As a result, divisions do not benefit from potential offsetting cash flows in other peer divisions, and the corporation as a whole overhedges. In this subsection, we propose a new kind of decentralized risk management program. In the new program, divisions make their own hedging and financing decisions individually just like in the uncoordinated decentralization program, but headquarters organizes an internal market that helps divisions exploit offsetting cash flows in peer divisions.

The reason why an internal market enables divisions to exploit offsetting cash flows in peer divisions is that there is no transaction cost for transactions on the internal market. A major component of transaction costs on external financial markets is the cost of credit risk. External financial institutions are in general suspicious to a firm's report on its financial strength and are afraid that the firm defaults financial contracts they take part in. To protect themselves from default risk, they usually charge a premium over fair price of any financial contract. The premium is the cost of credit risk. Within a multidivisional corporation, divisions are bound to be honest with each other by their desire to stay under the same roof and the corporate headquarters serves as a clearinghouse in the internal market, which effectively eliminates default risk of internal transactions.

To facilitate internal transactions, the corporate headquarters sets internal prices after investigating prices on external financial markets. Let δ_k be the internal price of the kth Arrow security for $k = 1, 2, \cdots, N$. Then the internal price ought to be chosen such that $b_k > \delta_k \geq p_k > s_k > 0$, where b_k is the external buy price, s_k is the external sell price, and p_k is the risk neutral equilibrium price. Divisions who wants to buy Arrow securities pay lower internal price δ_k than the external buy price b_k, and divisions who wants to sell receives higher internal price δ_k than the external sell price s_k. Therefore, both buying divisions and selling divisions will look for counterparties in internal market before in external financial markets.

The sequence of actions that divisions take to make hedging and financing decisions under the coordinated decentralization program is as follows. Take division j as an example. At time 0, division j solves the following linear programming problem,

$$\text{maximize} \quad f_j(\boldsymbol{q}_j^b, \boldsymbol{q}_j^s, \boldsymbol{z}_j) \tag{2.9a}$$

subject to

$$\begin{aligned}
v_{j1} + q_{j1}^b - \sum_{i=1}^{N} b_i q_{ji}^b - q_{j1}^s + \sum_{i=1}^{N} s_i q_{ji}^s + z_{j1} &\geq \eta_{j1} \\
v_{j2} + q_{j2}^b - \sum_{i=1}^{N} b_i q_{ji}^b - q_{j2}^s + \sum_{i=1}^{N} s_i q_{ji}^s + z_{j2} &\geq \eta_{j2} \\
&\vdots \\
v_{jN} + q_{jN}^b - \sum_{i=1}^{N} b_i q_{ji}^b - q_{jN}^s + \sum_{i=1}^{N} s_i q_{ji}^s + z_{jN} &\geq \eta_{jN}
\end{aligned} \tag{2.9b}$$

$$f_j(\boldsymbol{q}_j^b, \boldsymbol{q}_j^s, \boldsymbol{z}_j) \geq \mathrm{E}(\boldsymbol{v}_j) \tag{2.9c}$$

$$\boldsymbol{q}_j^b \geq \mathbf{0}, \quad \boldsymbol{q}_j^s \geq \mathbf{0}, \quad \boldsymbol{z}_j \geq \mathbf{0}\ . \tag{2.9d}$$

This optimization problem is the same as the optimization problem (2.6) division j would solve under the uncoordinated decentralization program. Let $(\boldsymbol{q}_j^{*b}, \boldsymbol{q}_j^{*s}, \boldsymbol{z}_j^*)$

be an optimal solution division j obtains by solving the above linear programming problem. Division j then posts a request in the internal market to buy Arrow securities in quantities of $\boldsymbol{q}_j^{*b}$ and sell Arrow securities in quantities of $\boldsymbol{q}_j^{*s}$. Once all divisions post their desired quantities, a schedule of supply and demand of each Arrow security becomes to exist in the internal market. Since there are N Arrow securities, there will be N schedules of supply and demand. Divisions clear the internal market according to these schedules.

In case that supply and demand do not match exactly on a schedule, divisions follow a *pro rata* rule to clear the internal market. Take the schedule of the kth Arrow security as an example. Suppose that supply and demand are equal to $\sum_{j=1}^{M} q_{jk}^{*s}$ and $\sum_{j=1}^{M} q_{jk}^{*b}$, respectively, where q_{jk}^{*s} and q_{jk}^{*b} are division j's optimal selling and buying positions in the kth Arrow security, respectively. Let ρ be the ratio of supply to demand. When supply is less than demand, i.e., $\rho < 1$, the *pro rata* rule specifies that each buying division gets only a proportion of what it desires. For example, if division j is in demand of the kth Arrow security, i.e., $q_{jk}^{*b} > 0$, then division j gets only ρq_{jk}^{*b} from the internal market.[5] When supply is more than demand, i.e., $rho > 1$, the *pro rata* rule specifies that each selling division sells only a proportion of what it can provide. For example, if division j wants to sell $q_{jk}^{*s} > 0$, then division j sells only q_{jk}^{*s}/ρ in the internal market.

Once the internal market clears, let $\hat{\boldsymbol{q}}_j = \{\hat{q}_{j1}, \hat{q}_{j2}, \cdots, \hat{q}_{jN}\}$ be the position division j takes in each Arrow security in the internal market. If $\hat{q}_{jk} > 0$, division j buys the kth Arrow security in quantity of $\hat{q}_{jk}$. If $\hat{q}_{jk} < 0$, division j sells the kth Arrow security in quantity of $-\hat{q}_{jk}$. Since cash flows from these internal Arrow securities changes division j's expected net present value and the distribution of internal fund at time 1, division j needs to solve a second optimization problem to compute additional hedging and financing it has to obtain from external financial markets.

[5] According to Property 1 in Proposition 2.1, $q_{jk}^{*s} = 0$ when $q_{jk}^{*b} > 0$.

Let $(\boldsymbol{q}_j^b, \boldsymbol{q}_j^s, \boldsymbol{z}_j)$ be additional hedging and financing position division j takes in external financial markets. Then division j's disposable fund in state k is equal to

$$v_{jk} + \hat{q}_{jk} - \sum_{i=1}^{N} \delta_i \hat{q}_{ji} + q_{jk}^b - \sum_{i=1}^{N} b_i q_{ji}^b - q_{jk}^s + \sum_{i=1}^{N} s_i q_{ji}^s + z_{jk} \geq \eta_{jk} \ ,$$

and division j's value in state k is equal to

$$v_{jk} + \hat{q}_{jk} - \sum_{i=1}^{N} \delta_i \hat{q}_{ji} + q_{jk}^b - \sum_{i=1}^{N} b_i q_{ji}^b - q_{jk}^s + \sum_{i=1}^{N} s_i q_{ji}^s - c_k z_{jk} + r_k \eta_{jk} \ ,$$

where v_{jk} is division j's operating income, $\hat{q}_{jk} - \sum_{i=1}^{N} \delta_i \hat{q}_{ji}$ represents cash flow from division j's positions in internal Arrow securities, $q_{jk}^b - \sum_{i=1}^{N} b_i q_{ji}^b - q_{jk}^s + \sum_{i=1}^{N} s_i q_{ji}^s$ represents cash flow from division j's positions in external Arrow securities, $c_k z_{jk}$ is expenses of acquiring external fund, and $r_k \eta_{jk}$ is expected gain of investment made in state k at time 1. By taking expectation of division j's value at time 1 under the risk neutral probability distribution, we can write division j's expected net present value as follows

$$\hat{f}_j(\boldsymbol{q}_j^b, \boldsymbol{q}_j^s, \boldsymbol{z}_j) = \tag{2.10}$$
$$\mathrm{E}(\boldsymbol{v}_j) - \sum_{k=1}^{N} (\delta_k - p_k) \hat{q}_{jk} + \sum_{k=1}^{N} p_k r_k \eta_{jk} - \sum_{k=1}^{N} \left[(b_k - p_k) q_{jk}^b + (p_k - s_k) q_{jk}^s + p_k c_k z_{jk} \right] \ .$$

The second optimization problem is then formulated as follows

$$\text{maximize} \quad \hat{f}_j(\boldsymbol{q}_j^b, \boldsymbol{q}_j^s, \boldsymbol{z}_j) \tag{2.11a}$$

subject to

$$\begin{aligned}
v_{j1} + \hat{q}_{j1} - \sum_{i=1}^{N} \delta_i \hat{q}_{ji} + q_{j1}^b - \sum_{i=1}^{N} b_i q_{ji}^b - q_{j1}^s + \sum_{i=1}^{N} s_i q_{ji}^s + z_{j1} \quad &\geq \quad \eta_{j1} \\
v_{j2} + \hat{q}_{j2} - \sum_{i=1}^{N} \delta_i \hat{q}_{ji} + q_{j2}^b - \sum_{i=1}^{N} b_i q_{ji}^b - q_{j2}^s + \sum_{i=1}^{N} s_i q_{ji}^s + z_{j2} \quad &\geq \quad \eta_{j2} \\
&\vdots
\end{aligned} \tag{2.11b}$$

$$v_{jN} + \hat{q}_{jN} - \sum_{i=1}^{N} \delta_i \hat{q}_{ji} + q^b_{jN} - \sum_{i=1}^{N} b_i q^b_{ji} - q^s_{jN} + \sum_{i=1}^{N} s_i q^s_{ji} + z_{jN} \geq \eta_{jN}$$

$$\hat{f}_j(\boldsymbol{q}^b_j, \boldsymbol{q}^s_j, \boldsymbol{z}_j) \geq \mathrm{E}(\boldsymbol{v}_j) \tag{2.11c}$$

$$\boldsymbol{q}^b_j \geq \boldsymbol{0}, \quad \boldsymbol{q}^s_j \geq \boldsymbol{0}, \quad \boldsymbol{z}_j \geq \boldsymbol{0} \,. \tag{2.11d}$$

As the last step of the decision-making process, divisions take optimal hedging and financing positions given by the second optimization with external financial institutions.[6] Even though divisions might still be able to find internal offsetting supply or demand with a second round of internal trading, the potential gain would be small because most offsetting has occurred in the first round of internal trading.

In the following, we compare the coordinated decentralization program with the uncoordinated decentralization program and the real-world centralization program. The coordinated decentralization program is evidently better than the uncoordinated decentralization program. Under the coordinated decentralization program, divisions have opportunity to trade Arrow securities in the internal market. For divisions who wants to sell Arrow security, internal trades increase their value because they receive higher internal price than external price. For divisions who wants to buy Arrow security, internal trades also increase their value because they pay lower internal price than external price. The internal market helps all divisions achieve higher values than under the uncoordinated decentralization program.

The coordinated decentralization program is better than the real-world centralization program for two reasons. First, the coordinated decentralization program avoids value reduction caused by imprecise information. Under the coordinated decentralization program, divisions make their hedging and financing decisions based on first-hand knowledge. There is no loss of information during transfer and no incentives to use biased estimates in solving the optimization problem. In

[6]The corporation can take advantage of the scale of economy in transaction costs on external financial markets by asking divisions to report their external hedging and financing positions to headquarters and letting headquarters to make transactions with external financial institutions on a larger scale.

stead, divisions are motivated to use their best knowledge in making decisions. Secondly, the coordinated decentralization program prevents cross-subsidization of poorly-performing divisions by better-performing divisions. Under the real-world centralization program, the corporate headquarters pools operating income from all divisions, makes firm-wide hedging and financing decisions, and then allocates capital to divisions for investment purpose. During the process, inefficient cross-subsidization occurs as an inevitable consequence of imprecise information or other agency issues. On the contrary, divisions remain to be independent profit centers under the coordinated decentralization program. Transfer of internal fund between divisions is accomplished through a market mechanism, which benefits givers of internal fund and costs receivers. While better-performing divisions afford to acquire more capital in both internal and external markets, poorly-performing divisions are prevented from getting excessive capital by cost in both markets. The coordinated decentralization program creates value for the corporation by the amount that the inefficient cross-subsidization destroys.

2.4 Summary and Discussion

We set up a theoretical framework to analyze the problem of designing a risk management program for a multidivisional corporation. Under the framework, the corporation uses risk management not only to avoid financial distress but also to secure financing for future investments. It solves a linear programming problem to make optimal hedging and external financing decisions.

Existing risk management programs in large corporations fall into two categories. Firms in the first category let divisions make decisions at the division level individually while firms in the second category ask divisions to report information to the corporate headquarters and let headquarters make decisions at the corporate level.

Our analysis shows that firms in the first category tends to overhedge while firms in the second category either underhedge or overhedge.

We propose a third kind of risk management program. We suggest that corporations organize an internal market for divisions to trade state-contingent claims among themselves. We show that such a coordinated decentralization program with internal market does better than existing ones in maximizing the expected net present value of the whole corporation.

The concept of internal risk management market is practically implementable. Recently, the oil giant BP Amoco PLC in Naperville, Ill. has created an internal market for their divisions to trade permits to emit the greenhouse gases (Ginsburg (2000)). Since many large corporations have already built or are building internal computer systems to enable direct communication among divisions, internal risk management market can be implemented without much extra complications or cost.

The theoretical model we develop in this paper has other applications. It is realistic enough that management can follow it to make real-world hedging and financing decisions. It also provides a framework under which management can compare organizational programs based on simulations.

REFERENCES

Arrow, K. (1964). The role of securities in the optimal allocation of risk bearing. *Review of Economic Studies*, 31, 91-95.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd International symposium on Information Theory* (Edited by B.N. Petrov and F. Csáki), 267-281. Akadémiai Kiado, Budapest.

Barber, B. M. and Lyon, J. D. (1997). Detecting long-run abnormal stock returns: The empirical power and specification of test statistics. *Journal of Financial Economics*, 43, 341-372.

Berger, P. and Ofek, E. (1994). Diversification's effect on firm value. *Journal of Financial Economics*, 37, 39-65.

Breeden, D. and Viswanathan, S. (1996). Why do firms hedge? An asymmetric information model. Working paper, Fuqua School of Business, Duke University.

Brown, G. (1999), Managing foreign exchange risk with derivatives. Unpublished working paper, Kenan-Flagler School of Business, University of North Carolina.

Campbell, J. Y. Lo, A. W. and MacKinlay, A. C. (1997). *The econometrics of financial markets.* Princeton University Press, Princeton, New Jersey.

Davis, H. A. and Militello, F. C. Jr. (1995). *Foreign exchange risk management: A survey of corporate practices.* Financial Executives Research Foundation.

Desai, H. and Jain, P. C. (1997). Long-run common stock returns following splits and reverse splits. *Journal of Business*, 70, 409-433.

DeMarzo, P. and Duffie, D. (1991). Corporate financial hedging with proprietary information. *Journal of Economic Theory*, 53, 261-286.

Dharan, B. G. and Ikenberry, D. L. (1995). The long-run negative drift of post-listing stock returns. *Journal of Finance*, 50, 1547-1574.

Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, 49, 283-306.

Fama, E. and French K. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33, 3-56.

Froot, K. A. Scharfstein, D. S. and Stein, J. C. (1993). Risk management: coordinating corporate investment and financing policies. *Journal of Finance*, 48, 1629-1658.

Froot, K. A. and Stein, J. C. (1998). Risk management, capital budgeting, and capital structure policy for financial institutions: An integrated approach. *Journal of Financial Economics*, 47, 55-82.

George, E. I. (2000). The variable selection problem. *Journal of the American Statistical Association*, 95, 1304-1308.

Ginsburg, J. (2000). Letting the free market clear the air. Business Week, November 6.

Hocking, R. R. (1976). The analysis and selection of variables in linear regression. *Biometric*, 32, 1-49.

Hocking, R. R. (1985). *The analysis of linear models.* Brooks/Cole Publishing Company.

Hollander, M. and Wolfe D. A. (2000). *Nonparametric statistical methods.* 2nd Edition. John Wiley & sons.

Kothari, S. P. and Warner, J. B. (1997). *Measuring long-horizon security price performance.* *Journal of Financial Economics*, 43, 301-339.

Leland, H. E. (1998). Agency costs, risk management, and capital structure. *Journal of Finance*, 53, 1213-1243.

Li, K.-C. (1987). Asymptotic optimality for C_p, C_L, cross-validation and generalized cross-validation: Discrete index set *Annals of Statistics*, 15, 958-975.

Lyon, J. D. Barber, B. M. and Tsai, C.-L. (1999). Improved methods for tests of long-run abnormal stock returns. *Journal of Finance*, 54, 165-201.

Merton, R. C. (1992). Financial innovation and economic performance. *Journal of Applied Corporate Finance*, 4, 12-22.

Miller, A. (1990). *Subset selection in regression.* London: Chapman and Hall.

Mitchell, M. L. and Stafford, E. (2000). Managerial decisions and long-term stock price performance. *Journal of Business*, 73, 287-329.

Nishii, R. (1984). Asymptotic properties of criteria for selection of variables in multiple regression. *Annals of Statistics*, 12, 758-765.

Pötscher, B. M. (1989). Model selection under nonstationary: autoregressive models and stochastic linear regression models. *Annals of Statistics*, 17, 1257-1274.

Rao, C. R. and Wu, Y. (1989). A strongly consistent procedure for model selection in a regression problem. *Biometrika*, 76, 369-374.

Ross, M.P. (1997). Corporate hedging: What, why and how? Unpublished working paper, Walter A. Haas School of Business, University of California, Berkeley.

Schwarz, G. (1978). Estimating the dimensions of a model. *Annals of Statistics*, 6, 461-464.

Shao, J. (1997). An asymptotic theory for linear model selection. *Statistica Sinica*, 7, 221-264.

Shao, J. (1998). Convergence rates of the generalized information criterion. *Journal of Nonparametric Statistics*, 9, 217-225.

Smith, C. W. Jr. and Stulz, R.M. (1985). The determinants of firms' hedging policies. *Journal of Financial and Quantitative Analysis*, 20, 391-405.

Stein, J. (1997). Internal capital markets and the competition for corporate resources. *Journal of Finance*, 52, 111-133.

Stulz, R. M. (1996). Rethinking risk management. *Journal of Applied Corporate Finance*, 9, 8-24.

Thompson, M. L. (1978a). Selection of variables in multiple regression: Part I, A review and evaluation. *International Statistical Review*, 46, 1-19.

Thompson, M. L. (1978b). Selection of variables in multiple regression: Part II, Chosen procedures, computations and examples. *International Statistical Review*, 46, 126-146.

Zhang, P. (1993). On the convergence rate of model selection criteria. *Communications in Statistics A*, 22, 2765-2775.

BIOGRAPHICAL SKETCH

Shaojun Zhang

EDUCATION

- Florida State University, Tallahassee, Florida, 1998 – 2001
  Ph.D. in Statistics with an Emphasis in Finance, April 2001
- Florida State University, Tallahassee, Florida, 1996 – 1998
  M.S. in Statistics, May 1998
- Tsinghua University, Beijing, P.R. China, 1991 – 1996
  B.S. in Applied Mathematics, July 1996
  B.E. in Computer Science and Technology, July 1996

AWARDS AND HONORS

- University Fellowship, Florida State University, Spring 2001
- Student Research Presentation Award, the 2000 Florida Chapter of the American Statistics Association Annual Meeting, February 18, 2000
- University Fellowship, Florida State University, 1999-2000
- University Fellowship, Florida State University, 1997-1998
- Best First Year Student in Theoretical Statistics, Department of Statistics, Florida State University, 1996-1997
- Best First Year Student in Applied Statistics, Department of Statistics, Florida State University, 1996-1997
- Outstanding Undergraduate Student, Tsinghua University, 1996
- Xiong Qinglai Outstanding Student in Mathematics, Tsinghua University, 1994
- Undergraduate Scholarship, Tsinghua University, 1992-1995

TEACHING EXPERIENCE

Instructor — Summer 1997, 1998, 1999
Department of Statistics, Florida State University — Fall 2000
Responsible for all aspects of teaching an undergraduate business statistics course, including lectures, grading as well as individual conferences and student advising.

Teaching Assistant — 08/96 – 05/97, 08/98 – 05/99
Department of Statistics, Florida State University — Spring 2001

Assisted in courses including Business Statistics, Time Series Analysis, Nonparametric Statistical Methods, and Applied Multivariate Analysis.

COMMUNITY SERVICE

- Chairman of Student Advisory Committee, Department of Statistics, Florida State University, 2000-2001
- Member of Student Advisory Committee, Department of Statistics, Florida State University, 1998-2001
- Member of Executives Committee, Chinese Student & Scholar Association, Florida State University, 1997-1998