New Quantitative Approaches to Asset Selection and Portfolio Construction

Irene Song

Submitted in partial fulfillment of the

requirements for the degree

of Doctor of Philosophy

in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2014

ABSTRACT

New Quantitative Approaches to Asset Selection and Portfolio Construction

Irene Song

Since the publication of Markowitz's landmark paper *Portfolio Selection* in 1952, portfolio construction has evolved into a disciplined and personalized process. In this process, security selection and portfolio optimization constitute key steps for making investment decisions across a collection of assets. The use of quantitative algorithms and models in these steps has become a widely-accepted investment practice by modern investors. This dissertation is devoted to exploring and developing those quantitative algorithms and models.

In the first part of the dissertation, we present two efficiency-based approaches to security selection: (i) a quantitative stock selection strategy based on operational efficiency and (ii) a quantitative currency selection strategy based on macroeconomic efficiency. In developing the efficiency-based stock selection strategy, we exploit a potential positive link between firm's operational efficiency and its stock performance. By means of data envelopment analysis (DEA), a non-parametric approach to productive efficiency analysis, we quantify firm's operational efficiency into a single score representing a consolidated measure of financial ratios. The financial ratios integrated into an efficiency score are selected on the basis of their predictive power for the firm's future operating performance using the LASSO (least absolute shrinkage and selection operator)-based variable selection method. The computed efficiency scores are directly used for identifying stocks worthy of investment. The basic idea behind the proposed stock selection strategy is that as efficient firms are presumed to be more profitable than inefficient firms, higher returns are expected from their stocks. This idea is tested in a contextual and empirical setting provided by the

U.S. Information Technology (IT) sector. Our empirical findings confirm that there is a strong positive relationship between firm's operational efficiency and its stock performance, and further establish that firm's operational efficiency has significant explanatory power in describing the cross-sectional variations of stock returns. We moreover offer an economic argument that posits operational efficiency as a systematic risk factor and the most likely source of excess returns of investing in efficient firms.

The efficiency-based currency selection strategy is developed in a similar way; i.e. currencies are selected based on a certain efficiency metric. An exchange rate has long been regarded as a reliable barometer of the state of the economy and the measure of international competitiveness of countries. While strong and appreciating currencies correspond to productive and efficient economies, weak and depreciating currencies correspond to slowing down and less efficient economies. This study hence develops a currency selection strategy that utilizes macroeconomic efficiency of countries measured based on a widely-accepted relationship between exchange rates and macroeconomic variables. For quantifying macroeconomic efficiency of countries, we first establish a multilateral framework using effective exchange rates and trade-weighted macroeconomic variables. This framework is used for transforming the three representative bilateral structural exchange rate models: the flexible price monetary model, the sticky price monetary model, and the sticky price asset model, into their multilateral counterparts. We then translate these multilateral models into DEA models, which yield an efficiency score representing an aggregate measure of macroeconomic variables. Consistent with the stock selection strategy, the resulting efficiency scores are used for identifying currencies worthy of investment. We evaluate our currency selection strategy against appropriate market and strategic benchmarks using historical data. Our empirical results confirm that currencies of efficient countries have stronger performance than those of inefficient countries, and further suggest that compared to the exchange rate models based on standard regression analysis, our models based on DEA improve on the predictability of the future performance of currencies.

In the first part of the dissertation, we also develop a data-driven variable selection method for DEA based on the group LASSO. This method extends the LASSO-based variable selection method used for specifying a DEA model for estimating firm's operational efficiency. In our proposed method, we derive a special constrained version of the group LASSO with the loss function suited for variable selection in DEA models and solve it by a new tailored algorithm based on the alternating direction method of multipliers (ADMM). We conduct a thorough evaluation of the proposed method against two widely-used variable selection methods: the efficiency contribution measure (ECM) method and the regression-based (RB) test, in the DEA literature using Monte Carlo simulations. The simulation results show that our method provides more favorable performance compared with its benchmarks.

In the second part of the dissertation, we propose a generalized risk budgeting (*GRB*) approach to portfolio construction. In a *GRB* portfolio, assets are grouped into possibly overlapping subsets, and each subset is allocated a risk budget that has been pre-specified by the investor. Minimum variance, risk parity and risk budgeting portfolios are all special instances of a *GRB* portfolio. The *GRB* portfolio optimization problem is to find a GRB portfolio with an optimal risk-return profile where risk is measured using any positively homogeneous risk measure. When the subsets form a partition, the assets all have identical returns and we restrict ourselves to long-only portfolios, then the *GRB* problem can in fact be solved as a convex optimization problem. In general, however, the *GRB* problem is a constrained non-convex problem, for which we propose two solution approaches. The first approach uses a semidefinite programming (SDP) relaxation to obtain an (upper) bound on the optimal objective function value. In the second approach we develop a numerical algorithm that integrates augmented Lagrangian and Markov chain Monte Carlo (MCMC) methods in order to find a point in the vicinity of a very good local optimum. This point is then supplied to a standard non-linear optimization routine with the goal of finding this local optimum. It should be emphasized that the merit of this second approach is in its generic nature: in particular, it provides a starting-point strategy for any non-linear optimization algorithms.

Contents

List of Figures

List of Tables

List of Algorithms

Acknowledgments

I am indebted to: *those who have given me a just enough prod behind so I could jump to the skies*: my advisors – professors Martin Haugh, Garud Iyengar, Soulaymane Kachani and Iraj Kani, and all the professors at the department of Industrial Engineering and Operations Research (IEOR); *those who have taken care of the IEOR family*: the IEOR staff; *those who have walked in the dark with me*: Rodrigo A. Carrasco, Song-Hee Kim, Tulia Herrera & Matthieu Plumettaz, Tony Qin, Ali Sadighian, Aya Wallwater and all my friends at Columbia; *those with whom I can afford to be stupid*: Nitin Ananda, Christine Chung, Laura Kim, Seine Kim, Hyun-Choo Lee, Sohl Lee, Lee Lui, Dongha Yang, Bomi Yoo and all my friends in the city and back home; *those who have encouraged me to take the road less traveled and make a difference*: my grandfathers in heaven, my grandmother in Pusan, Aunt Jae-Ok, Uncle Jae-Yoon, Halmae and all my relatives back home; and lastly, *those whom I love the most*: my mom, my dad and my big sister Seung-Eun. There is no way to acknowledge them all, or even any of them properly. Thank you.

To My Mom, My Dad and My Big Sister Seung-Eun

Chapter 1

Introduction

Since the publication of Markowitz's landmark paper *Portfolio Selection* in 1952, portfolio construction has evolved into a disciplined and personalized process. This process typically includes the following four basic steps:

(i) selecting asset classes to be included in the portfolio (asset class selection);

(ii) deciding weights for each asset class in the portfolio (asset allocation policy);

(iii) selecting securities within each asset class in the portfolio to achieve superior returns relative to that asset class (security selection);

(iv) deciding weights for individual securities within each asset class in the portfolio to optimize its risk-return trade-off (portfolio optimization).

The first two decisions on the types and weights of asset classes are generally addressed by individual investors as part of their investment policy. These decisions are largely driven by investors' appetite for risk and their investment goal. For example, risk averse investors with an aim of capital preservation prefer to allocate a larger portion of their investment portfolio to lower-risk securities such as fixed income and cash equivalents. On the other hand, equities may be as much as 100% of the investment portfolio of risk-taking investors with a primary objective of capital appreciation. The last two decisions lead to an investment strategy that can tactically add value (on a risk-adjusted basis) to the investment portfolio. Compared to the way investment policies are determined, these are made more objectively by quantitative methods in most cases. This dissertation is devoted to exploring and developing those quantitative security selection and portfolio optimization methods for making critical investment decisions.

In the first part of the dissertation, we present efficiency-based approaches to security selection backed by strong quantitative discipline, and yet securely grounded on fundamental analysis. There are two asset classes, namely stock and currency, and are likewise two corresponding asset selection strategies considered in this part of the study. In designing a quantitative stock selection strategy, we study the relationship between firm's operational efficiency and its stock performance. Operational efficiency of a firm measures its success in producing maximum output(s) from its given set of inputs (Farrell, 1957). Efficiently operating firms are therefore expected to be more profitable than inefficiently operating firms. Considering that the price of a stock tends to reflect firm's economic value and accounting profitability, a positive link between firm's operational efficiency and its stock performance is arguably plausible. The question that we attempt to address, though, is whether or not one can exploit such a link in building a profitable investment strategy. In our study, we quantity firm's operational efficiency into a consolidated measure of financial ratios by means of data envelopment analysis (DEA). Firm-specific information is therefore inherent in this measure, and presumably, so is its operating prospects. We form various portfolios based on such measures and evaluate their performance over different investment horizons in a contextual and empirical setting provided by the U.S. Information Technology (IT) sector. Moreover, by analyzing returns of these portfolios, we investigate the systematic nature of operational efficiency and provide a line of evidence supporting an economic argument that posits operational efficiency as a systematic risk factor.

A quantitative currency selection strategy is constructed in a similar way. An exchange rate has long been served as a useful gauge for assessing the health of the economy and measuring international competitiveness of countries. While strong and appreciating currencies correspond to productive and efficient economies, weak and depreciating currencies correspond to slowing down and less efficient economies. Hence, when constructing a currency portfolio, it is sensible to compare macroeconomic efficiency of countries. Our quantitative model for estimating macroeconomic efficiency of countries is founded on a widely-accepted relationship between exchange rates and macroeconomic variables. In our model development, we first establish a multilateral framework using effective exchange rates and trade-weighted macroeconomic variables. This framework is used for transforming the three representative bilateral structural exchange rate models: the flexible price monetary model (Bilson, 1978; Frenkel, 1976), the sticky price monetary model (Dornbusch, 1976; Frankel, 1979), and the sticky price asset model (Hooper and Morton, 1982), into their multilateral counterparts. We then translate these multilateral models into DEA models. These DEA models integrate various macroeconomic variables into a single score that can be interpreted as a measure of country's macroeconomic efficiency and thus, can be used for identifying currencies worthy of investment. Based on the rankings of the estimated efficiency scores, we select currencies to be included in the investment portfolio, and measure its performance against appropriate market and strategic benchmarks using historical data. We must emphasize that in this study, in addition to presenting a currency selection strategy, we are introducing a new way of presenting traditional exchange rate models.

The DEA method used for computing operational efficiency of firms and macroeconomic efficiency of countries is a mathematical programming approach to the estimation of frontier functions. The key advantages of the method include: (i) its non-parametrical nature, (ii) its ability to accommodate a multiplicity of inputs and outputs, and (iii) its efficiency computation based on deviation from the optimality rather than the measures of central tendency. The procedure constructs an empirically optimal production frontier consisting of the best performing entities in the sample and measures each entity's efficiency in terms of its proximity to the frontier. Performance is a relative concept, which can be measured in relation to the average or the optimum. However, one can argue that there is a general consensus among the investing public that the latter offers a more effectual performance measure. This makes our use of efficiency-based metrics in making investment decisions a sensible and defensible exercise.

In the first part of the dissertation, we also develop a joint variable selection method for DEA using the group LASSO (least absolute shrinkage and selection operator). We derive a special constrained version of the group LASSO with the loss function suited for variable selection in DEA models and solve it by a new tailored algorithm based on the alternating direction method of multipliers (ADMM). The proposed method is evaluated against a wide variety of scenarios using Monte Carlo simulations. Furthermore, two widely-used variable selection methods: the efficiency contribution measure (ECM) method and the regression-based (RB) test, in the DEA literature serve as benchmarks for performance evaluation.

In the second part of the dissertation, we propose a generalized risk budgeting (GRB) approach to portfolio construction, a risk-based portfolio optimization strategy. The financial crisis of 2008 and its aftermath have reinforced the key role of risk in asset allocation, and as a result, risk-based investment strategies have become very popular in recent years. In contrast to conventional portfolio construction approaches that concern with capital allocation, these approaches concern with risk allocation. For example, the risk parity approach equalizes the risk contribution of each asset in the portfolio. The limiting factors in most of the prevailing risk-based approaches are: (i) they just focus on minimizing the total portfolio risk disregarding the expected asset returns; (ii) they are restricted to long-only portfolios; and (iii) risk budgets are defined for individual assets. Our approach, on the other hand, provides a more generic risk allocation framework that can accommodate different needs of different investors. In our framework, investors are allowed to take short positions on assets, optimize their portfolio on the basis of its risk-return profile, and define risk budgets for possibly overlapping subsets of assets.

In the GRB approach, portfolio risk is estimated by any positively homogeneous risk measure, and portfolio optimization involves a constrained non-convex problem. When the subsets of assets pre-specified by the investor form a partition, the assets all have the same expected return and the investment portfolio is confined to long-only portfolios, then the respective GRB portfolio optimization problem can in fact be solved as a convex optimization problem. In general, however, it is a constrained non-convex problem, for which we propose two solution approaches. In the first approach, we use a semidefinite programming (SDP) relaxation to obtain an (upper) bound on the optimal objective function value. In the second approach, we develop a numerical algorithm that integrates augmented Lagrangian and Markov chain Monte Carlo (MCMC) methods in order to find a point in the proximity of a very good local optimum. This point is then supplied to a standard non-linear optimization routine with the goal of finding this local optimum. It should be emphasized that the merit of this second approach is in its generic nature: in particular, it provides a starting-point strategy for any non-linear optimization algorithms.

The remaining of the dissertation is organized as follows. In Part I, Chapter 2 provides an overview of DEA, Chapter 3 and Chapter 4 cover security selection strategies based on operational efficiency and macroeconomic efficiency respectively, and Chapter 5 details a joint variable selection algorithm for DEA. Part II presents a generalized risk-budgeting approach to portfolio construction. A general conclusion of the dissertation is given in Chapter 7.

Part I

Efficiency-Based Approaches to Asset Selection

Chapter 2

Preliminaries of Data Envelopment Analysis (DEA)

Data envelopment analysis (DEA) is a non-parametric mathematical programming approach to the estimation of production frontiers. Since its introduction in the seminal paper by Charnes *et al.* (1978), it has grown into a popular quantitative analytical tool in various fields, including management science, operations research and economics (Cooper *et al.*, 2004). A single comprehensive measure of productive efficiency estimated by this method has broadly served as a basis for making managerial decisions in practice. Over the past few decades, we have seen many successful applications of DEA in the performance evaluation of economic entities, also referred to as decision making units (DMUs), reside in diverse areas, ranging from non-profit sectors, such as hospitals, to for-profit sectors, such as banks.[1] Along with its rising popularity, DEA has certainly developed into a widely-accepted field of research in its own.

[1] To name a few, Kuntz and Vera (2007) conducted performance analysis of hospitals by means of DEA, and Yeh (1996) applied DEA in conjunction with financial ratios for evaluating the performance of banks. For more applications of DEA, interested readers can refer to *Data Envelopment Analysis and Its Applications to Management* (Charles and Kumar, 2012).

2.1 Relative Efficiency in DEA

The notion of efficiency in DEA is closely related to that of Pareto efficiency[2] in "welfare economics." The definition of Pareto efficiency, formulated by the Swiss-Italian economist Vilfredo-Pareto, is given as follows:

> "A Pareto optimum is a welfare maximum defined as a position [in an economy] from which it is impossible to improve anyone's welfare by altering production or exchange without impairing someone else's welfare (Pearce, 1986)."

This definition is extended to "production economics" by Koopmans (1951), a Dutch-American mathematician and economist. By studying the interactions between inputs and outputs of production, Koopmans introduced "efficiency prices" in his definition of efficiency to guide production and exchange to positions that are similar to Pareto efficiency. Both the Pareto-Koopmans efficiency and the relative efficiency in DEA extend this approach in their definition.

Definition 2.1.1 (Pareto-Koopmans Efficiency)**.** *A DMU is fully (100%) efficient if and only if no further improvements can be made in its performance without worsening some of its other inputs or outputs (Cooper* et al.*, 2004).*

Since the theoretically possible levels of efficiency is generally unknown in practice, (2.1.1) is replaced by the following definition, in which efficiency of a DMU is determined based solely on the empirically available information.

Definition 2.1.2 (Relative Efficiency)**.** *A DMU is rated as fully (100%) efficient if and only if comparisons with other DMUs do not convey any evidence of inefficiency in input usage and/or output production (Cooper* et al.*, 2004).*

An advantage of using (2.1.2) in the estimation of efficiency is that it avoids the need for assigning a priori measures of relative importance to any input or output. Accordingly,

[2] The terms "Pareto efficiency" and "Pareto optimality" are used interchangeably in economics.

the essence of the DEA method is that it requires neither an a priori choice of weights of inputs/outputs nor an explicit functional form for the production function.

By solving a set of linear programs (LPs), DEA constructs a piecewise linear production frontier representing the observed relation between inputs and maximal outputs (or outputs and minimal inputs) in the sample, and labels any deviation from the frontier as inefficient. For example, the originally proposed efficiency measure of a DMU is the maximum of a ratio between the weighted sum of outputs and that of inputs (see the objective function of (2.1)) and is obtained for a particular $\text{DMU}_p, p \in \{1, \ldots, n\}$, in the sample by solving the LP equivalent[3] of the following fractional program.

$$\begin{aligned} \max_{u,v} \quad & \frac{\sum_{r=1}^{s} y_{r,p} u_r}{\sum_{k=1}^{l} x_{k,p} v_k} && \\ \text{subject to} \quad & \frac{\sum_{r=1}^{s} y_{r,j} u_r}{\sum_{k=1}^{l} x_{k,j} v_k} \leq 1, \; j = 1, \ldots, n, && \\ & u_r \geq 0, \; r = 1, \ldots, s, && \\ & v_k \geq 0, \; k = 1, \ldots, l && \end{aligned} \tag{2.1}$$

where $X = x_{k,j} \in \mathbb{R}^{l \times n}$ are the input parameters, $Y = y_{r,j} \in \mathbb{R}^{s \times n}$ are the output parameters, u and v are the variables for output and input weights respectively (Charnes *et al.*, 1978). The inequality constraint is imposed to ensure that the estimated efficient frontier envelops all the sample data points. DEA basically generalizes the so-called productivity ratio of a single output to a single input to the case of multiple outputs and multiple inputs.

2.2 Basic Features of DEA Models

Numerous DEA models are available in the literature for the estimation of relative efficiency. These models differ broadly in four aspects: (i) their approach to measuring technical efficiency, (ii) their orientations in efficiency analysis, (iii) their assumptions on production

---

[3] Charnes *et al.* (1978) showed that the fractional program (2.1) can be transformed into an equivalent LP.

frontiers, and (iv) their ability to handle different data types.

First, in terms of measuring technical efficiency, DEA models take either a radial approach or a non-radial approach. In the radial approach, inputs and outputs are assumed to change proportionally. This approach is therefore prone to neglect non-radial input and output slacks. Because it does not detect input excesses and output shortfalls, radial models can only classify each DMU as weakly-efficient or inefficient. In contrast, non-radial DEA models directly deal with input excesses and output shortfalls, and thus, are capable of distinguishing efficient DMUs from inefficient ones.

Second, DEA models can be classified as output-oriented, input-oriented or base-oriented. While output-oriented DEA models focus on output augmentation to achieve efficiency (outputs are controllable), input-oriented DEA models aim to minimize the amount of inputs required for producing a certain amount of outputs (inputs are controllable). Base-oriented DEA models are concerned with determining the optimal mix of inputs and outputs (both inputs and outputs are controllable).

The third basis for variation among DEA models is returns-to-scale, which (in economics) describes what happens when the scale of production increases over the long run when all input levels are variable (chosen by the firm). There are two basic types of returns-to-scale: constant returns-to-scale (CRS) and variable returns-to-scale (VRS). Models that assume CRS production technology presume that the size of a DMU does not affect its efficiency. More precisely, a DMU operates under CRS technology if an increase in its inputs results in a proportionate increase in its outputs. If it is suspected that an increase in inputs does not result in a proportional change in outputs, models that assume VRS production technology should be considered. In terms of linear programming, the production possibility set of a VRS model is spanned by the convex hull of input and output variables. The VRS specification, in general, is a safer option if the DEA model does not include all the variables deemed to be relevant in the analysis (Galagedera and Silvapulle, 2003).

Lastly, two important properties in DEA models are the units invariant property and the translation invariant property. A DEA model is considered units invariant if it yields an efficiency score that is independent of the measurement units of the inputs and outputs. The translation invariant property allows a DEA model to handle negative data.[4] Formally, a DEA model is said to be translation invariant if translating the original input and/or output data yields a new problem with the same optimal solution as the old one. Being a VRS model is a key condition for having this property. Therefore, when dealing with negative data in DEA, an implicit assumption is that the production technology satisfies VRS. Not all VRS models, however, have the translation invariant property, and a good example of this is the basic additive model introduced in the next section.

2.3 Basic Models in DEA

The first standard DEA model in a LP form is the LP equivalent of (2.1) proposed by Charnes *et al.* (1978). This model is commonly known as the (primal) CCR model and is one of the three representative basic DEA models together with the BCC model (Banker *et al.*, 1984) and the additive model (Charnes *et al.*, 1985c). The output-oriented formulations of the primal and dual CCR models for evaluating a particular DMU_p, $p \in \{1, \ldots, n\}$, are given by (2.2) and (2.3) respectively.

$$\begin{aligned} \max_{u,v} \quad & \sum_{r=1}^{s} y_{r,p} u_r && && (2.2) \\ \text{subject to} \quad & \sum_{k=1}^{l} x_{k,p} v_k &=& \; 1, \\ & \sum_{r=1}^{s} y_{r,j} u_r &\leq& \sum_{k=1}^{l} x_{k,j} v_k, \; j = 1, \ldots, n, \\ & u_r &\geq& \; 0, \; r = 1, \ldots, s, \\ & v_k &\geq& \; 0, \; k = 1, \ldots, l; \end{aligned}$$

[4] For discussions on the negative data in DEA, refer to Pastor and Ruiz (2007).

$$
\begin{aligned}
\max_{\theta_p,\lambda} \quad & \theta_p \\
\text{subject to} \quad x_{k,p} & \geq \sum_{j=1}^{n} x_{k,j}\lambda_k,\ k = 1,\ldots,l, \\
\sum_{j=1}^{n} y_{r,j}\lambda_j & \geq y_{r,p}\theta_p,\ r = 1,\ldots,s, \\
\lambda_j & \geq 0,\ j = 1,\ldots,n.
\end{aligned} \tag{2.3}
$$

While the primal CCR model seeks to maximize efficiency by directly manipulating the weights u and v, the dual CCR model looks for a composite DMU (with input $X\lambda$ and output $Y\lambda$) that takes in at most the same input as the DMU_p, but produces a multiple $(\theta_p Y_p)$ of the output. The output-oriented BCC model is obtained when the above CCR model (2.3) is augmented by adding a convexity constraint, $\sum_{j=1}^{n} \lambda_j = 1$. This convexity constraint accounts for VRS production technology; i.e. without this constraint, the model assumes CRS production technology. The CCR and BCC models, hence, differ only in their assumption of the underlying production technology.

While both the CCR and BCC models are a radial DEA model with the units invariant property, the additive model is a non-radial DEA model without the units invariant property. The formal definition of the dual additive model with VRS technology[5] is given

[5] The corresponding formulation with CRS technology is a special instance without the convexity constraint $\sum_{j=1}^{n} \lambda_j = 1$ in (2.4) (or without the variable w in (2.5)).

by

$$
\begin{aligned}
\max_{s^-,s^+,\lambda} \quad Z_p &= \sum_{k=1}^{l} s_{k,p}^- + \sum_{r=1}^{s} s_{r,p}^+ \\
\text{subject to} \quad \sum_{j=1}^{n} \lambda_j y_{r,j} &= y_{r,p} + s_{r,p}^+,\ r = 1,\dots,s, \\
\sum_{j=1}^{n} \lambda_j x_{k,j} &= x_{k,p} - s_{k,p}^-,\ k = 1,\dots,l, \\
\sum_{j=1}^{n} \lambda_j &= 1, \\
\lambda_j &\geq 0,\ j = 1,\dots,n, \\
s_{k,p}^- &\geq 0,\ k = 1,\dots,l, \\
s_{r,p}^+ &\geq 0,\ r = 1,\dots,s
\end{aligned}
\tag{2.4}
$$

where $s_{k,p}^-$ and $s_{r,p}^+$ represent the respective input excesses and output shortfalls. This means that a DMU_p is efficient if and only if $Z_p^* = 0$ at optimality. It is worth noting that not all DEA models provide a relative efficiency score, and the additive model is an example of such models; i.e. it merely segregates efficient DMUs from inefficient DMUs. The associated primal model of (2.4) is given by

$$
\begin{aligned}
\min_{u,v,w} \quad & \sum_{k=1}^{l} x_{k,p} v_k - \sum_{r=1}^{s} y_{r,p} u_r + w, \\
\text{subject to} \quad & \sum_{k=1}^{l} x_{k,j} v_k - \sum_{r=1}^{s} y_{r,j} u_r + w \geq 0,\ j = 1,\dots,n, \\
& u_r \geq 1,\ r = 1,\dots,s, \\
& v_k \geq 1,\ k = 1,\dots,l.
\end{aligned}
\tag{2.5}
$$

Additionally, under VRS production technology, the additive model has the translation invariant property. We should note that although the BCC model likewise assumes VRS production technology, it is translation invariant only in a limited sense. Depending on the model orientation, the BCC model is invariant with respect to the translation of inputs or outputs, but not both.

2.4 Window Analysis

All three basic DEA models presented in the previous section are concerned with the cross-sectional analysis of DMUs. They implicitly assume that each DMU is observed only once. Observations for DMUs are, nevertheless, usually available over multiple time periods in practice. It is moreover often desirable to perform time-series analysis that focuses on the temporal evolution of efficiency of DMUs. In such a setting, one can apply window analysis (Charnes *et al.*, 1985a)[6] to DEA models to incorporate panel data.

Window analysis is a technique grounded on the principles of moving averages (Charnes *et al.*, 1995; Yue, 1992) and was developed in order to provide discriminatory results when the number of DMUs is small compared to the number of variables. In window analysis, each DMU in a different period is treated as if it were a different unit. In doing so, the performance of a unit in a particular period is compared to its own performance in other periods, in addition to the performance of other units. This increases the number of data points in the analysis, thus providing a higher degree of freedom (Avkiran, 2004; Reisman, 2003), and results in efficiency scores from inter-temporal analysis.

To formalize, consider n DMUs, which are observed in T periods ($t = 1, \ldots, T$) and which all use l inputs to produce s outputs. The sample, hence, has $n \times T$ observations, and an observation j in period t, DMU$_t^j$ has an l-dimensional input vector $\mathbf{x}_t^j = (x_{1,t}^j, \ldots, x_{l,t}^j)'$ and a s-dimensional output vector $\mathbf{y}_t^j = (y_{1,t}^j, \ldots, y_{s,t}^j)'$. The window starting at time t, $1 \leq t \leq T$ with a window size of w, $1 \leq w \leq T - t$ is denoted by t_w and has $(n \times w)$ observations. The matrix of inputs for this window analysis is given by

$$X_{t_w} = (\mathbf{x}_t^1, \mathbf{x}_t^2, \ldots, \mathbf{x}_t^n, \mathbf{x}_{t+1}^1, \mathbf{x}_{t+1}^2, \ldots, \mathbf{x}_{t+1}^n, \ldots, \mathbf{x}_{t+w}^1, \mathbf{x}_{t+w}^2, \ldots, \mathbf{x}_{t+w}^n)$$

[6] This name, window analysis, and the basic concept are due to G. A. Klopp (1985b) who developed these techniques in his capacity as a chief statistician for the U.S. Army Recruiting Command.

and the matrix of outputs is given by

$$Y_{t_w} = (\mathbf{y}_t^1, \mathbf{y}_t^2, \ldots, \mathbf{y}_t^n, \mathbf{y}_{t+1}^1, \mathbf{y}_{t+1}^2, \ldots, \mathbf{y}_{t+1}^n, \ldots, \mathbf{y}_{t+w}^1, \mathbf{y}_{t+w}^2, \ldots, \mathbf{y}_{t+w}^n).$$

This chapter provided an introduction to DEA that will be used as a principal tool for analyzing operational efficiency of firms and macroeconomic efficiency of countries. Readers interested in bettering their understanding of DEA can refer to Cooper *et al.* (2000, 2004). We should note that unless otherwise mentioned, the same notations for variables and parameters introduced in this chapter will be used throughout the first part of the dissertation.

Chapter 3

Quantitative Stock Selection Based on Operational Efficiency

3.1 Introduction

Operational efficiency[1] refers to firm's ability to transform its operating resources into profits. It is traditionally measured based on publicly available accounting information, which under the assumption of the efficient market hypothesis,[2] is expected to be reflected in stock prices. Considering the widespread acceptance of stock performance as the best measure of investment value of a firm (Brealey and Myers, 1991), it is natural to assume that there is a relationship between firm efficiency and stock performance. However, as trading on available information is not expected to provide any abnormal profit beyond that explained by exposure to systematic factors, it is debatable whether or not one can build a profitable investment strategy based on firm's operational efficiency. It is to this

[1]The terms "operational efficiency," "operating efficiency," "firm efficiency" and "productivity" are often used interchangeably although in certain instances in the literature there are important conceptual or technical differences among these terms.

[2]The weak form of the efficient market hypothesis posits that prices fully reflect the information implicit in the sequence of past prices. The semi-strong form of the hypothesis asserts that prices reflect all relevant information that is publicly available while the strong form asserts that information known to any participants is reflected in market prices (Dimson and Mussavian, 2000).

question that the present study is devoted.

In this study, we build a stock selection strategy based on firm's operational efficiency and evaluate its performance over time in a contextual and empirical setting provided by the U.S. Information Technology (IT) sector. Our aim in so doing is threefold. First, is to present a methodological framework for aggregating a diverse set of financial ratios into a single summary measure of firm's operational efficiency. Second, is to highlight the advantages of methodologies that are based on deviation from the optimality (measure of efficiency) rather than the average (measure of centrality). Last and most importantly, is to investigate the relationship between firm's operational efficiency and its stock price performance, and the systematic nature of operational efficiency.

In contrast to traditional approaches to security selection, our strategy does not rely upon either the estimation of the fundamental value of individual stocks (traditional fundamental analysis) or the quantification of the excess returns (traditional quantitative analysis). Instead, we evaluate the operational efficiency of a firm based on firm fundamentals. The estimation of operational efficiency generally necessitates the knowledge of a production function, which for a complex business process is difficult to specify and is often unattainable in reality due to a wider scope it allows for human subjectivity (Farrell, 1957). If the measure of operational efficiency is to be used as a basis for determining the investment worthiness of a firm, it would be sensible to compare performance with the empirically observed optimum rather than to a postulated standard of perfect efficiency. For this reason, we compute an efficiency score for a firm by employing DEA on a series of financial ratios. For the purpose of the current study, we employ only a limited set of standard financial ratios as input (output) variables to the DEA methodology. We will defer the examination of a more complete list of financial ratios and industry specific ratios as well as other quantitative, technical and macroeconomic indicators that can serve as the most suitable inputs (outputs) to the DEA methodology to a future work.

In the context of our study, the DEA method quantifies firm's operational efficiency into a single efficiency score representing a consolidated measure of financial ratios. The use of financial ratios has long been the core aspect of financial analysis for providing an essential guidepost for investment decisions (Horrigan, 1966). Yet, there are only a few prescriptions for how these ratios should be used collectively to evaluate the performance of a firm (Ou and Penman, 1989). Accordingly, in addition to building a distinct investment strategy, this study also provides a systematic approach to integrate various financial ratios into a meaningful efficiency measure,[3] which contains a broad range of firm-specific information and can serve as an effective tool for isolating and comprehending the consensus estimate of future company performance. We should point out that there are several important studies in the literature that apply the DEA methodology to financial statement data for assessing performance of various economic entities.[4]

Based on the estimated efficiency scores, we rank firms and form three types of investment portfolios for performance evaluation. The first two represent firms in the top and bottom efficiency deciles. The third is constructed as a long-short portfolio with its long positions on the most efficient firms and its short positions on the least efficient firms. In order to examine: (i) the impact of firm efficiency on stock performance and (ii) whether efficient firms significantly outperform inefficient firms, we track and measure the performance of these portfolios over different investment horizons in terms of various return, risk and risk-return trade-off indicators.

Since the magnitude of over- or under-performance of the efficiency-based portfolios depends critically on the choice of a benchmark, a residual-based portfolio is constructed using conventional relative value analysis as a reference point for comparative purposes. The average performance measure of the firms in each industry within the IT sector is estimated

[3]It should be noted that the construction of an efficiency score in this study does not rely upon any accounting identity, such as the DuPont identitities, which break down return on equity (ROE) into various elements in order to identify the sources of variations in return, e.g. ROE = Profit Margin × Asset Turnover × Equity Multiplier. We are grateful to Jeffrey Wimmer for pointing this out.

[4]For example, Smith (1990) evaluated 47 pharmaceutical firms using a DEA model with average equity and debt as inputs, and earnings, interest payments, and tax payments as outputs. Ozcan and McCue (1996) developed a DEA-based aggregate metric, which they refer to as "financial performance index" and used it in conjunction with various financial ratios to indicate performance levels of hospitals. Inevitably, DEA has been actively applied to financial statement data for reviewing performance of various economic entities, including U.S. electronic companies (Yue, 1991), U.S. computer companies (Kozmetsky *et al.*, 1994), banks (Yeh, 1996) and credit unions (Paradi and Phille, 2002).

by means of regression analysis, and each firm's cheapness (value) is quantified in terms of residuals. Firms whose performance measures are below (above) the industrial average are considered cheap (expensive) relative to their peers and are selected for investment (short selling). In addition to the residual-based portfolio, IT sector indices are included in the performance evaluation as market benchmarks.

We further study the role of firm efficiency in explaining the cross-sectional behavior of stock returns based on the supposition that firm inefficiency is not likely to be captured by traditional systematic risk factors, such as stock β, firm size, and book-to-market ratio in the original Fama-French model (1992). In order to examine the explanatory power of firm efficiency in describing the cross-section of stock returns, we conduct Fama-MacBeth regressions (1973) with efficiency score and various known predictors of stock returns as control variables. These predictors include firm size, book-to-market ratio, stock β and the measures of profitability, illiquidity, accruals, momentum and reversal.

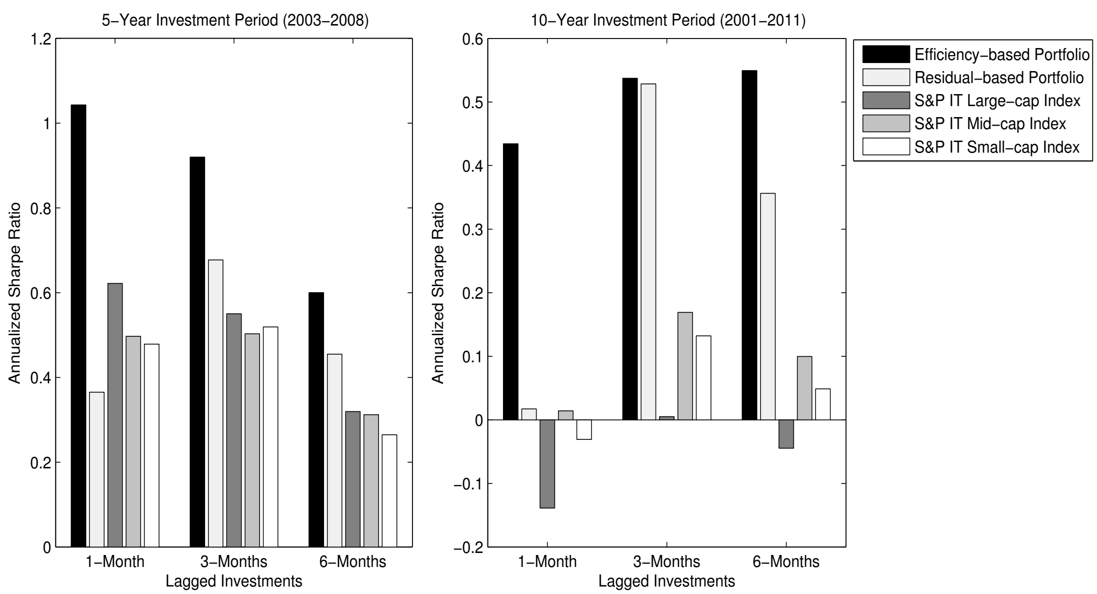


Figure 3.1: Annualized Sharpe Ratio

All the empirical findings in this study are obtained from the out-of-sample analyses and in relation to the future realized (ex-post) stock performance. With regard to the U.S.

IT sector, our results confirm the existence of a strong positive relationship between firm efficiency and stock performance. Figure 3.1 presents the annualized Sharpe ratios of the efficiency-based portfolio against its benchmarks for various investment lags (1-, 3-, and 6-month lagged investments). The bar charts are shown, side by side, for 5- and 10-year investment periods. The 5-year investment period is from 2003 to 2008, excluding the 2001 recession, its aftermath and the 2008 financial crisis. In the time of a stable economy, the efficiency-based portfolio has high values of Sharpe ratios – 1.04, 0.92 and 0.60 for 1- , 3- and 6-month lagged investments respectively. The decline we see in the Sharpe ratios is consistent with the deterioration in the world economy that began in December of 2007. The 10-year investment period is from 2001 to 2011, which includes challenging economic times, such as the burst of the dot-com bubble and the Global Financial Crisis. During this period, it is observed that the efficiency-based portfolio comprised of the most efficient firms outperforms all of its benchmarks.

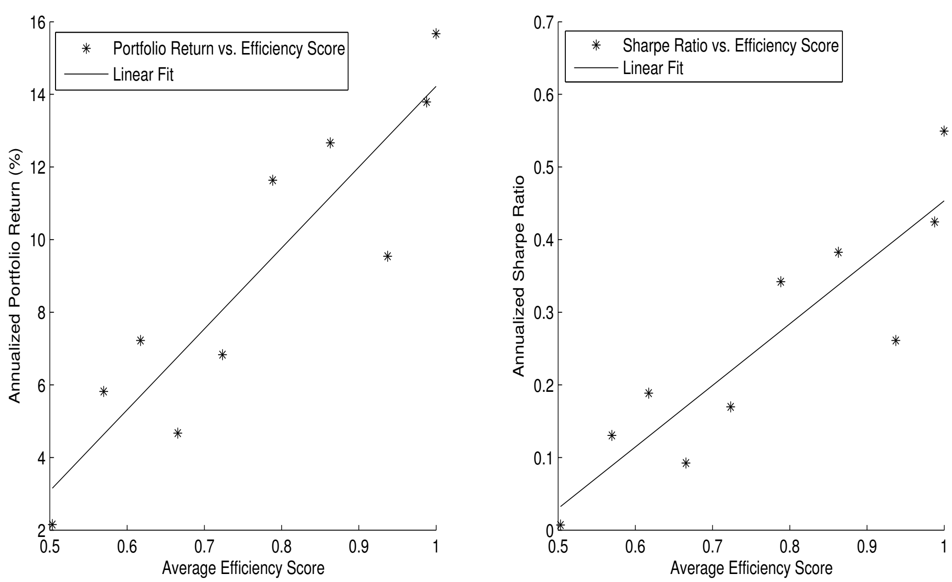


Figure 3.2: Average Efficiency Scores of Efficiency Decile Portfolios vs. Performance Metrics

The relationship between firm efficiency and stock performance is further demonstrated in Figure 3.2. Figure 3.2 presents two scatter plots of each efficiency decile portfolio's average efficiency score versus its annualized Sharpe ratio and total return. As inferred from these diagrams, the most efficient firms have significantly stronger performance than the least efficient ones, and the linear fit shown in each plot manifests a positive relationship between firm's operational efficiency and stock performance. We reach a similar conclusion on the relationship between firm efficiency and the cross-section of stock returns; i.e. firm's operational efficiency has significant explanatory power in describing the cross-sectional variation in stocks returns.

The presence of market inefficiency may be the reason behind the outperformance of our investment strategy. For instance, firm efficiency may contain firm-specific information that is not yet reflected in market prices due to delay in the price adjustment process. If firm-specific information diffuses gradually through the investing public, then the early identification of this information can provide profits. A number of empirical studies in recent years, in fact, have documented a variety of ways, in which stock returns can be predicted from publicly available information.[5] There are also possible behavioral explanations. Investors may systematically underestimate the profitability and hence the returns of the most efficient firms.[6] For example, Novy-Marx (2012) found that firm's gross profitability predicted its future returns as well as conventional value metrics like book-to-market ratios.[7] His study suggested that although firms' future stock prices eventually reflect their profitability, the market systematically underestimates this today, thus making their shares relatively undervalued. In our study, however, we will offer an economic argument that posits operational efficiency as a systematic risk factor explaining excess stock returns of efficient firms. Intuitively, one may expect to be compensated for investing in inefficient

[5]For example, Jegadeesh and Titman (1993) studied a trading strategy that buys well-performing stocks and sells poor-performing stocks and reported average annual excess returns of 12%, where excess returns were defined relative to the standard capital asset pricing model. Lakonishok *et al.* (1994) reached a similar conclusion through analyzing a strategy that buys value stocks and sells glamour stocks identified with variables, such as price earnings ratios, dividends, book-to-market values, cash flows and sales growth. Michaely *et al.* (1995) investigated profitable trading strategies involving dividend announcements and omissions. Simiarly, Chan *et al.* (1996) documented the excess returns of portfolios formed on the basis of past returns and earnings announcements. Shleifer (2000) provides a comprehensive review of these studies.

[6]We are gratetul to an anonymous referee for pointing this out.

[7]Fama and French (2006) also discovered similar results.

firms. Our empirical findings, nonetheless, bring evidence to the contrary; i.e. investors choosing to invest in efficient firms are more likely to be remunerated. Our economic reasoning for this phenomenon is that efficient firms are subject to the risk of efficiency loss over time, to which market assigns a positive premium.

Previous studies, which investigated the link between firm efficiency and stock performance, include works by Alam and Sickles (1998) and Frijns *et al.* (2012). Alam and Sickles (1998) analyzed the role of firm efficiency in the stock performance of U.S. airlines and found that the most efficient firms within this industry outperformed the most inefficient ones. Frijns *et al.* (2012) studied the role of firm efficiency in explaining the cross-section of stock returns in the U.S. market. Consistent with Alam and Sickles (1998), they documented that efficient firms outperformed inefficient firms and further concluded that firm efficiency played an important role in asset pricing in the U.S. market.[8]

A similar study of constructing an investment portfolio based on a DEA-derived metric was also conducted by Edirisinghe and Zhang (2008). In their study, the authors developed a relative financial strength (RFS) indicator by applying the DEA methodology to financial statement data and used this indicator for identifying financially sound firms for investment. They also tested their strategy on an empirical setting provided by the U.S. IT sector and reported similar results to ours that firms with a higher RFS indicator had superior share price performance relative to those with a lower RFS indicator.

The remainder of the chapter is organized as follows: Section 3.2 delineates the methodology of our study; the data is described in Section 3.3; Section 3.4 presents empirical findings; Section 3.5 provides evidence for a positive relationship between firm efficiency and the cross-section of stock returns in the U.S. IT sector; Section 3.6 offers an economic explanation for the relative outperformance of the efficient firms in the U.S. IT sector; and Section 3.7 concludes.

---

[8]Nguyen and Swanson (2009) also studied the relationship between firm efficiency and stock performance. They employed stochastic frontier analysis with firm's market value as the output measure to relate firm efficiency to stock performance in the U.S. market. In contrast to Alam and Sickles (1998) and Frijns *et al.* (2012), they reported that highly inefficient firms outperformed highly efficient firms even after adjusting for firm characteristics and risk factors.

3.2 Methodology

In this section, we first detail the estimation of firm efficiency using DEA. Next, we propose an efficiency-based stock selection strategy and describe the construction of its regression analysis-based benchmark. We then outline the methods used for evaluating the proposed strategy and explain our approach in determining whether or not an efficiency score has explanatory power for the cross-section of stock returns.

3.2.1 The Estimation of Operational Efficiency

Consider a set of firms facing the same production possibility set. Due to firm specific characteristics, each firm is operating at different efficiency levels. For any combination of profit levels and firm characteristics in a sample, we can estimate a production possibility frontier. Each point on the frontier represents a theoretically possible maximum profit level a firm can attain given its fundamentals. However, as the true value of the achievable profit levels is unknown, the estimated frontier is only an empirical benchmark consisting of the best performing firms in the sample. Each firm's proximity to the frontier is therefore a measure of firm's efficiency relative to its peers. Throughout this study, we refer to this measure as an *efficiency score*.

An efficiency score is determined based on the numerical technique of DEA. As is true in any mathematical modeling, the accuracy of the estimated efficiency score relies upon the choice of a DEA model, and its input and output variables. For this study, we select the weighted additive model (WAM) developed by Pastor (1994) for measuring firm efficiency. There are three main reasons for choosing this model: (i) it is fully translation and units invariant (Lovell and Pastor, 1995), (ii) it supports variables measured on a ratio scale,[9] and (iii) it takes a non-radial approach to estimating efficiency.

The mathematical formulation of the WAM model for evaluating the performance of a

[9] Because we are dealing with financial data, the presence of negative and interval scale variables measured in different units is unavoidable.

particular DMU_p, $p \in \{1, \ldots, n\}$, is given by

$$\begin{aligned}
\min_{u,v,w} \quad & w + \sum_{k=1}^{l} v_k x_{k,p} - \sum_{r=1}^{s} u_r y_{r,p} \\
\text{subject to} \quad & w + \sum_{k=1}^{l} v_k x_{k,j} - \sum_{r=1}^{s} u_r y_{r,j} \geq 0,\ j = 1, \ldots, n, \\
& u_r \geq \frac{1}{(l+s)R^{+}_{r,p}},\ r = 1, \ldots, s, \\
& v_k \geq \frac{1}{(l+s)R^{-}_{k,p}},\ k = 1, \ldots, l
\end{aligned} \tag{3.1}$$

where $R^{+}_{r,p} = \max_{j=1,\ldots,n}\{y_{r,j}\} - y_{r,p}$ and $R^{-}_{k,p} = x_{k,p} - \min_{j=1,\ldots,n}\{x_{k,j}\}$. The only difference between (3.1) and the basic additive model (2.5) introduced in Chapter 2 is in the lower bounds of u and v.

The dual problem of (3.1) provides a better mathematical interpretation of the model.

$$\begin{aligned}
\max_{s^{-}_{p}, s^{+}_{p}, \lambda} \quad & Z_p = \frac{1}{l+s}\left(\sum_{k=1}^{l} \frac{s^{-}_{k,p}}{R^{-}_{k,p}} + \sum_{r=1}^{s} \frac{s^{+}_{r,p}}{R^{+}_{r,p}}\right) \\
\text{subject to} \quad & \sum_{j=1}^{n} \lambda_j y_{r,j} = y_{r,p} + s^{+}_{r,p},\ r = 1, \ldots, s, \\
& \sum_{j=1}^{n} \lambda_j x_{k,j} = x_{k,p} - s^{-}_{k,p},\ k = 1, \ldots, l, \\
& \sum_{j=1}^{n} \lambda_j = 1, \\
& \lambda_j \geq 0,\ j = 1, \ldots, n, \\
& s^{-}_{k,p} \geq 0,\ k = 1, \ldots, l, \\
& s^{+}_{r,p} \geq 0,\ r = 1, \ldots, s.
\end{aligned} \tag{3.2}$$

From the first two constraints of (3.2), s^{+}_{p} and s^{-}_{p} can be viewed as output shortfalls and input excesses respectively. R^{+}_{p} and R^{-}_{p} can be seen as the ranges of possible improvements in each variable for a DMU_p. To avoid problems with zeros, the slacks corresponding to variables with $R^{+}_{r,p} = 0$ for some $r = 1, \ldots, s$ and/or $R^{-}_{k,p} = 0$ for some $k = 1, \ldots, l$ are ignored. The objective function of (3.2) is to maximize a weighted sum of input excesses and output shortfalls. Mathematically, this means to maximize a weighted l_1-distance to the efficient frontier. From the optimal solution Z_p^*, an efficiency measure can be derived as $\Gamma_p^* = 1 - Z_p^*$. This measure, initially proposed by Cooper *et al.* (1999), has some desirable properties for our analysis. For instance, $0 \leq \Gamma_p^* \leq 1$ and $\Gamma_p^* = 1$ if and only if a DMU_p achieves Pareto-Koopmans efficiency. The convexity constraint, $\sum_{j=1}^{n} \lambda_j = 1$, ensures that only interpolation between observed performance is possible in forming best practice comparison groups, and therefore avoids possibly inappropriate extrapolation of performance.

With regard to variable selection, in this study we focus on a single-output setting and choose return on equity (ROE) as the output variable representing firm profitability. ROE is arguably one of the most significant firm profitability measures, and its value has been used as a legitimate criterion for selecting stocks by many analysts. Moreover, it correlates with information other than earnings that predicts future profitability of a firm, and thus explains stock returns (Penman, 1991). In order to utilize efficiency scores in making investment decisions, the choice of an output variable should not only measure the profitability of a firm, but it should also be closely related to stock returns. In this regard, ROE is a prudent choice for an output variable.

In selecting input variables, we look for variables meeting two criteria: (i) variables whose values can be used as proxies for firm fundamentals and (ii) variables that have sufficient explanatory power for predicting the output variable. With respect to these criteria, a set of financial ratios with the most predictive power for future ROE is selected as input variables.[10,11] More specifically, from each of the commonly-used five categories of financial

[10]In this study, we impose positive weight contraints on input variables. In order to include variables with a negative relationship with ROE, one has to change their sign before using them in a DEA framework.

[11]Comparing raw financial statement items of various firms over several periods could result in misleading interpretations of firms' performance. One reason for this is that the value of the dollars reported on financial statements change over time due to inflation or deflation. Although one may overcome this complication by adjusting raw values via a price index, the issue of the comparability of firms in different sizes still remains. As a response to such problems, financial analysts have been primarily using financial ratios to compare firms among themselves and with industry benchmarks (Rees, 1995).

ratios: profitability, leverage, liquidity, efficiency and market value (Jordan *et al.*, 2010), we pick one representative ratio with the strongest power in forecasting ROE using a LASSO-based variable selection method we developed for DEA.[12] We should note that selecting one representative measure from each category is sufficient as ratios within a category are found to be highly correlated with each other.[13] In summary, for estimating firm efficiency at time t, the DEA model uses ROE at time t as an output variable and financial ratios observed at time $t-1$ as input variables.

3.2.2 Efficiency-based Portfolio Construction and Investment Strategy

By construction, firm efficiency in this study is measured relative to the frontier consisting of the best performing peers. A higher efficiency score indicates that a firm is more efficient at managing its financial positions and generating profits than its competitors while a lower efficiency score signals underperformance. Accordingly, the efficiency-based investment strategy identifies firms on or near the frontier as worthy of investment. The major difference between the proposed and traditional methods is that it is based on frontier analysis rather than central tendency.[14]

The following steps outline the proposed strategy based on a ranking of efficiency scores at time t.

Step 1 For each firm in the universe, a quarterly time series of efficiency scores is constructed up to time t.

Step 2 For each firm in the universe, an n-quarter exponentially weighted moving average of efficiency scores is computed from the time series constructed in Step 1.

Step 3 All the firms in the universe are ranked by their average efficiency score in descending order.

---

[12]Refer to Appendix A.3.

[13]For empirical evidence on this, see Horrigan (1966).

[14]Conventional analyses often rely upon standard statistical inferences, regressions and other quantitative methods that are principally grounded on the notions of central tendency.

Step 4 Firms in the top efficiency decile are selected to form an equally-weighted portfolio.

The first step involves building a time series of efficiency scores for each firm under consideration using DEA. As the efficiency score estimation at each time t does not include any fundamental data prior to time $t-1$, an efficiency score, by itself, is incapable of capturing a trend in firm efficiency. In order to incorporate temporal effects, an average efficiency score is computed using exponentially weighted moving average (EWMA) in Step 2. The smoothing factor for EWMA is chosen as per the mean absolute error (MAE) and mean squared error (MSE) analyses. We compare the MAE and the MSE of 4, 8, 10, 12, and 16 quarters and select the one with the smallest MAE and MSE as the smoothing factor. In Step 3, firms are sorted by their average efficiency score in descending order and are partitioned into efficiency deciles. In the last step, the top efficiency decile comprised of the firms closest to the frontier is selected to form a long-only equally-weighted investment portfolio. Over the investment horizon, the four steps are repeated at the end of each quarter, and the investment portfolio from Step 4 is rebalanced on a monthly basis to sustain equal dollar weights.

3.2.3 Benchmark Construction

For comparison purposes, we build a benchmark investment strategy based on fundamentally-oriented relative value analysis. Conventionally, relative value analysis emphasizes the adjustment process of firm's financial ratios to predetermined targets, such as the industrial average (Lev, 1969). Firms whose values are below the industrial average are deemed as being under-valued, and an increase in their value is anticipated. In contrast, those with values above the industrial mean are often considered as relatively expensive by investors, and their value is due for a decrease.[15] A common investment strategy, hence, is to identify cheap stocks relative to the industrial average and invest in them.

[15] Many studies have supported and empirically proved the existence of the financial ratio average adjustment process. These studies include *Generalized Financial Ratio Adjustment Processes and Their Implications* by Frecka and Lee (1983).

In this study, the industrial average is estimated using a multiple linear regression model, and the cheapness of a firm is measured in terms of residuals. There are two main reasons for choosing regression analysis for constructing a benchmark. First, regression analysis is viewed as a standard numerical technique in relative value analysis[16] and second, it is often referenced as an alternative method for DEA.[17] Accordingly, this study not only provides a comparison of central-tendency and frontier approaches to financial statement analysis, but it also provides that of two distinct numerical techniques.

As in the DEA model, a dependent variable of the regression model is ROE and independent variables are selected in the same way DEA input variables are selected. For each of the five categories of financial ratios, we perform LASSO-based variable selection[18] with ROE as a regressand and select the variable with the most explanatory power. Formally, we have

$$r_{j,t} = \alpha_t + \sum_{k=1}^{l} \beta_{k,t} x_{k,(j,t-1)} + \epsilon_{j,t},\ j = 1, \ldots, n$$

where $r_{j,t}$ is ROE of a firm j at time t and $x_{k,(j,t-1)}$ is a financial ratio k of the firm j observed at time $t-1$. $\epsilon_{j,t} = r_{j,t} - \hat{r_t}$, where $\hat{r_t}$ is the average ROE value of the n firms in the industry, is the estimated cheapness of the firm j at time t. $\epsilon_{j,t} < 0$ means that the firm j is below the industrial average, and its value is anticipated to go up and vice versa.

Mirroring the efficiency-based investment strategy, we propose the following residual-based strategy.

Step 1 For each firm in the universe, a quarterly time series of residuals is constructed

[16] Abarbanell and Bushee (1997) performed multiple linear regression on changes in prior earnings per share (EPS) and financial statements data, such as inventory, accounts receivable, capital expenditures and gross margin to predict changes in future EPS. The technique was also used in Lewellen (2004)'s work for forecasting returns with financial ratios.

[17] To name a few, Bowlin *et al.* (1985) and Thanassoulis (1993) examined regression analysis as an alternative approach to DEA for performance assessment. Furthermore, it has been shown recently that the standard (output-oriented, variable returns-to-scale) DEA model with a single output and multiple inputs can be formulated as non-parametric least squares regression subject to shape constraints (monotonicity and concavity) on the frontier and sign constraints on the regression residuals (Johnson and Kuosmanen, 2010).

[18] Refer to Appendix A.3.

up to time t.

Step 2 For each firm in the universe, an n-quarter exponentially weighted moving average of residuals is computed from the time series constructed in Step 1.

Step 3 All the firms in the universe are ranked by their average residual in ascending order.

Step 4 Firms in the top residual decile are selected to form an equally-weighted portfolio.

Since each step is analogous to that of the efficiency-based strategy, refer to Section 3.2.2 for the detailed explanation.

3.2.4 Evaluation of the Efficiency-based Stock Selection Strategy

The proposed efficiency-based stock selection strategy is assessed in terms of portfolio performance and its robustness.

3.2.4.1 Performance Metrics for Portfolio Performance Evaluation

To evaluate the portfolio performance, we compute and compare a number of widely-used performance metrics of the monthly total returns on the efficiency- and residual-based portfolios and market indices. The performance metrics can be broadly divided into three groups: (i) return indicators, (ii) risk indicators and (iii) risk-return trade-off measures.

First, for return indicators, we look at average annualized total return and the portfolio α excluding the systematic return. The portfolio α is estimated by running performance regressions. These regressions control for known systematic risk factors or factors traditionally known to affect stock returns. We perform three distinct, but related performance regressions. The first is the one-factor capital asset pricing model (CAPM), which controls for the degree of market risk of the portfolio. Formally, it is defined by the following regression model.

$$(r_{p,t} - r_{f,t}) = \alpha_p + \beta_{p,r_m - r_f}(r_{m,t} - r_{f,t}) + \epsilon_{p,t}$$

where $r_{p,t}$ is the total return of the portfolio p, $r_{m,t}$ is the market return and $r_{f,t}$ is the risk-free return at time t. The second is the Fama-French (1993) three-factor model that adds two other factors to CAPM; i.e.

$$(r_{p,t} - r_{f,t}) = \alpha_p + \beta_{p,r_m - r_f}(r_{m,t} - r_{f,t}) + \beta_{p,SMB}SMB_t + \beta_{p,HML}HML_t + \epsilon_{p,t}$$

where SMB_t is the size factor and HML_t is the value factor at time t. The final model is the six-factor model that adds momentum (Carhart, 1997), long-term reversal, and short-term reversal factors to the Fama-French three-factor model; i.e.

$$\begin{aligned}(r_{p,t} - r_{f,t}) &= \alpha_p + \beta_{p,r_m - r_f}(r_{m,t} - r_{f,t}) + \beta_{p,SMB}SMB_t + \beta_{p,HML}HML_t + \\ &\quad \beta_{p,MOM}MOM_t + \beta_{p,LTR}LTR_t + \beta_{p,STR}STR_t + \epsilon_{p,t}\end{aligned}$$

where MOM_t is the momentum factor, LTR_t is the long-term reversal factor, and STR_t is the short-term reversal factor at time t. For all three models, statistically significant positive (negative) α indicates over-performance (underperformance) of the portfolio p relative to the market after controlling for known systematic risks.

Second, for risk measures, we consider volatility, skewness, and kurtosis of the portfolio returns – the standard second, third and fourth moments of the portfolio return distribution for measuring risk respectively. Value-at-risk (VaR) and conditional value-at-risk (CVaR) (at 95% confidence level) of the portfolio returns are also computed to capture tail risk. Another measure of risk considered is the average maximum drawdown (MDD) defined as the maximum decline of a return series from a peak to a nadir over the investment period. As it measures how sustained one's losses can be, it is the risk measure of choice for many money management professionals; i.e. a reasonably low MDD is critical to the success of any investment.

Third, for risk-adjusted performance measures, the Sharpe ratio and the Sortino ratio are considered. The Sharpe ratio is the standard risk-adjusted performance measure that determines reward per unit of risk, and the Sortino ratio is a modification of the Sharpe ratio that penalizes only the downward deviations of the portfolio return from the target risk-free rate.

In addition to the described performance measures, an annual turnover ratio, which is defined as the ratio between the amount of the securities purchased (or sold) and the portfolio value at the end of the 12-month period, is used to evaluate the consistency and stability of the efficiency-based stock selection strategy over the investment period. All the performance metrics described above are computed both with and without the transaction costs of 50 basis points.

3.2.4.2 Robustness of the Efficiency-based Stock Selection Strategy

We evaluate the robustness of the efficiency-based stock selection strategy by analyzing the discriminatory power of an efficiency score. We test how well an efficiency score differentiates efficient firms from inefficient ones by comparing the performance of the top and bottom efficiency decile portfolios. We evaluate the relative performance of the top efficiency decile portfolio compared to the bottom efficiency decile portfolio by examining the return spread between the two. We refer to this spread as the "top-minus-bottom" spread throughout the study. In the analysis of the top-minus-bottom spread, we first carry out the standard difference-in-means test; i.e. we perform the t-test to determine whether the top-minus-bottom spread is positive and significant. We then look at the descriptive statistics and risk-adjusted performance metrics of the top-minus-bottom spread. We further estimate the CAPM, the Fama-French three-factor model, and the six-factor model using the top-minus-bottom spread as a regressand. The resulting intercept, $\alpha_{top-bottom}$, from each of the three models indicates the over-/underperformance of the top efficiency decile relative to the bottom efficiency decile.

Additionally, we form equally-weighted long-short portfolios with top and bottom efficiency decile portfolios and evaluate their performance. We consider two standard types of long-short portfolios: 1X0-X0, e.g. 130-30 portfolio, and leveraged neutral portfolios. An 1X0-X0 portfolio is a portfolio with 1X0% exposure to its long portfolio and X0% exposure to its short portfolio. A leveraged neutral portfolio is a portfolio that consists of a collection of long positions and short positions split equally, supported by cash account earning the risk-free return. Hence, the total return of a leveraged neutral portfolio includes the income on the collateral as well as the total earnings (including lending gains) on the long positions and the total losses (including borrowing costs) on the short positions. Assuming lending gains and borrowing costs offset each other, the total return on a leveraged neutral portfolio at time t is computed as

$$r_{p,t} = r_{f,t} + leverage \cdot (r_{long,t} - r_{short,t}) \tag{3.3}$$

where $r_{p,t}$ is the total return, *leverage* is defined as the notional size of the long and short positions divided by the collateral amount, $r_{f,t}$ is the risk-free rate, $r_{long,t}$ is the total return on the long positions and $r_{short,t}$ is the total return on the short positions.

3.2.5 The Role of Efficiency Scores in Explaining the Cross-Section of Stock Returns

We study the relationship between firm efficiency and expected stock returns using Fama-MacBeth regressions (1973). To examine the role of efficiency scores in explaining the cross-section of stock returns, we first estimate the following regression model.

$$r_{j,t+k} = \alpha_t + \beta_{ES,t} ES_{j,t} + \epsilon_{j,t} \tag{3.4}$$

where $ES_{j,t}$ is the efficiency score of a firm j at the end of quarter t and $r_{j,t+k}$ is the monthly total return for a firm j at time $t+k$, $k = 1, 3$, and 6 months. Then, we estimate the average parameter value by averaging $\beta_{ES,t}$ across time.

Similarly, we also conduct three distinct Fama-MacBeth regressions controlling for various firm attributes. The first set consists of the three original Fama and French factors (1992): firm size (SIZE), book-to-market ratio (B/M), and stock β (BETA), plus the efficiency score (ES). Consistent with Fama and French (1992), the logarithm of firm size and the logarithm of B/M ratio are used. Formally, we have

$$r_{j,t+k} = \alpha_t + \beta_{ES,t} ES_{j,t} + \beta_{SIZE,t} SIZE_{j,t} + \beta_{BM,t} BM_{j,t} + \beta_{BETA,t} BETA_{j,t} + \epsilon_{j,t}. \quad (3.5)$$

In the second set, we additionally control for profitability. Considering the direct use of ROE as an output variable in the efficiency score estimation, it is important to test whether there is any overlap between the efficiency score and the profitability measure. In the third set, we include five more control variables: accruals, leverage, illiquidity, momentum and reversal proxies. The definitions of these variables are given in Appendix A.4.5. The respective regression models for the second and third sets are given as follows.

$$\begin{aligned} r_{j,t+k} &= \alpha_t + \beta_{ES,t} ES_{j,t} + \beta_{SIZE,t} SIZE_{j,t} + \beta_{BM,t} BM_{j,t} + \beta_{BETA,t} BETA_{j,t} \quad (3.6) \\ &\quad + \beta_{PROF,t} PROF_{j,t} + \epsilon_{j,t} \end{aligned}$$

where $PROF_{j,t}$ is the profitability measure of a firm j at time t.

$$\begin{aligned} r_{j,t+k} &= \alpha_t + \beta_{ES,t} ES_{j,t} + \beta_{SIZE,t} SIZE_{j,t} + \beta_{BM,t} BM_{j,t} + \beta_{BETA,t} BETA_{j,t} \quad (3.7) \\ &\quad + \beta_{PROF,t} PROF_{j,t} + \beta_{ACCR,t} ACCR_{j,t} + \beta_{LEV,t} LEV_{j,t} + \beta_{MOM,t} MOM_{j,t} \\ &\quad + \beta_{REV,t} REV + \epsilon_{j,t} \end{aligned}$$

where $ACCR_{j,t}$, $LEV_{j,t}$, $MOM_{j,t}$, and $REV_{j,t}$ are the measures of accruals, leverage, momentum and reversal for a firm j at time t respectively. The inclusion of various control variables in the Fama-MacBeth regressions would help us identify any difference between the efficiency score and the other known return predictors.

3.3 Data

We backtest our efficiency-based stock selection strategy on the U.S. IT sector, which has the largest market capitalization in S&P.[19] Naturally, S&P IT sector indices are selected as market benchmarks, and we consider all three (large, mid, and small) market cap sizes for comparative purposes. All firms considered for our sample are identified according to the global industry classification standards (GICS).

The sample is collected from Bloomberg market data for two sets of time periods: the model estimation (in-sample) and the strategy implementation (out-of-sample) periods. The model estimation period for constructing the DEA and multiple linear regression model is from 1996 to 2000. The data from this period is used for two purposes: first, for selecting respective input and independent variables for DEA and multiple linear regression models, and second, for model calibration. We restrict the sample for this time period to those firms with the full 4 years of financial data.

The strategy implementation period is from 2001 to 2011. This period includes a full business cycle,[20] the 2001 recession caused by the burst of the IT bubble, its aftermath, and the 2008 financial crisis. It therefore provides us with a comprehensive picture of how the strategy would have worked in different economic periods, including the most stressful period for the IT sector. For the strategy implementation period, we have two different sets of stock universes: a fixed universe and a rolling universe. The fixed stock universe is confined to firms, which have the full 10 years (2001 – 2011) of financial data and stocks that have been actively traded during the past 10 years. This universe is used solely for conducting a preliminary test of the efficiency-based stock selection strategy to verify its applicability. For the rolling stock universe, we select firms that meet two criteria at the beginning of each investment quarter: (i) the full 4 years of historical financial data is available, and (ii) firms' stocks have been actively traded during the past 16 quarters. By

[19] As of December 31, 2010, the IT sector has the largest market capitalization of 18.4% (approximately 2,396.29 billion USD) in S&P. Source: S&P factsheet.

[20] As per the National Bureau of Economic Research (NBER), 2001 – 2011 includes a full business cycle with the 2001 and 2007 – 2009 recessions.

design, the rolling stock universe contains no forward looking data and thus, avoids any survivorship bias. For both universes, the liquidity of the firms is determined by reviewing average daily bid-ask ratios and stock prices. After withdrawing unqualified firms, we maintain a sample size of 150 and 300 for fixed and rolling stock universes respectively.

In our backtest, we consider 1-, 3- and 6-month lagged investments due to the reporting lags.[21] We make investment decisions for $t+1$, $t+3$, and $t+6$ months based on the estimated average efficiency scores and residuals at time t. Considering the length of the average reporting lag, which is about 96 days, 5 to 6-month lagged investment ensures that accounting information is known before we start investing. However, we consider 1- and 3-month lagged investments in our study to test the robustness of the proposed stock selection strategy.

For the performance evaluation, we use monthly total returns that are winsorized at the 1% and 99% points of their distributions (we refer to this as "1% winsorization") in order to reduce the effect of possibly spurious outliers[22] and the 1-month T-bill rate as the target risk-free rate. The data on returns for stocks are obtained from Bloomberg market data. Factors for the performance regression models are extracted from the Fama-French database.

3.4 Empirical Results

Empirical findings are reported in three separate sections. The first two sections present the results from the model estimation and strategy implementation periods. The empirical evidence for the robustness of the proposed strategy is provided in the last section.

[21] 10K (annual reports) must be filed with SEC within 60 to 90 days after the company's fiscal year end and 10Q (quarterly reports) must be filed within 40 to 45 days after each of the first three quarter ends of the company's fiscal year. Exact filing deadlines depend on the size of the filer's public float. Also, there is often a lag between the date of the receipt of a filing and the date of its public availability. The mean (median) public access lag for both NY/AMEX and NASDAQ firms is five (three) days for 10K reports. For 10Q reports, the mean (median) public access lag for NY/AMEX firms is six (five) days and for NASDAQ firms is six (three) days (Easton and Zmijewski, 1993).

[22] We are grateful to an anonymous referee for pointing this out.

3.4.1 The Model Estimation Period

Table 3.1: The Complete List of Financial Ratios

Category	Financial Ratio
Efficiency	Accounts Payable Turnover
	Accounts Receivable Turnover
	Asset Turnover
Leverage	Asset to Equity
	Common Equity to Total Asset
	Long Term Debt to Common Equity
	Long Term Debt to Total Equity
	Total Debt to Common Equity
	Total Debt to Total Asset
Liquidity	Cash Ratio
	Current Ratio
	Quick Ratio
Market Valuation	Price to Book Ratio
Profitability	Earning Before Interests and Taxes (EBIT) Margin
	Gross Margin
	Profit Margin
	Return on Asset
	Return on Equity

Table 3.2: Variable Selection Results

Category	DEA Input Variables	Regression Independent Variables
Profitability	Return on Equity	Return on Equity
Leverage	Common Equity to Total Asset	Total Debt to Total Common Equity
Liquidity	Cash Ratio	Cash Ratio
Efficiency	Accounts Receivable Turnover	Asset Turnover
Market Value	Price to Book Ratio	Price to Book Ratio

Table 3.1 contains the complete list of financial ratios considered for variable selection[23], and Table 3.2 lists the respective input and independent variables chosen for DEA and multiple linear regression models. For the formulae used for computing each financial ratio, refer to Appendix A.2. An efficiency score for each firm in the sample at quarter t is computed using a WAM with ROE at time t as an output variable and the selected financial ratios observed at time $t-1$ as input variables. Similarly, the average performance measure of the firms for each quarter t is estimated by a multiple linear regression model with ROE

[23]This complete list contains the selected set of financial ratios, for which minimum required data is available. A larger set of financial ratios can be considered pending data availability.

at time t as a dependent variable and the selected financial ratios observed at time $t-1$ as independent variables. It is worth noting that one-quarter lagged ROE is selected as a representative of the profitability category for both models and turns out to be the most powerful explanatory variable for forecasting future ROE among all the variables considered. This result is in line with previous studies, which verified historical ROE as a fairly good indicator of future ROE (Paynor and Little, 1966; Wilcox, 1984).

As per the MAE and MSE analyses, 4-quarter is chosen as the smoothing factor for computing exponentially weighted moving average (EWMA) of efficiency scores and residuals.[24]

3.4.2 The Strategy Implementation Period

All the performance metrics reported in this section are per annum figures unless otherwise mentioned.

3.4.2.1 Preliminary Results

As can be seen from Table 3.3, the efficiency-based portfolio formed from the fixed stock universe significantly outperforms the market indices in all three aspects: return, risk and risk-return trade-off, over the 10-year investment period (2001 – 2011). It has an average total return that is about 13% – 21% higher than the market returns and yet has a comparably low volatility. Accordingly, it exhibits strong risk-adjusted performance with Sharpe ratios of 0.69, 0.87 and 0.77 and Sortino ratios of 1.21, 1.56 and 1.32 for 1-, 3- and 6-month lagged investments respectively. The maximum drawdown of the efficiency-based portfolio is also about 10% – 23% lower than that of the market indices. We can make similar observations on other risk measures as well. Considering the collapse of the dot-com bubble and the downturn of the IT sector in early 2000s, such outperformance strongly supports the validity of the proposed stock selection methodology.

[24] Appendix A.5 presents the sensitivity analysis of the EWMA length in the efficiency-based stock selection.

Table 3.3: Performance of the Efficiency-based Portfolio and Market Indices

Table 3.3 presents the performance of the efficiency-based portfolio constructed from the constituents of the fixed stock universe and that of S&P IT sector indices measured over the 10-year investment period (2001 – 2011).

	Efficiency-based Portfolio			S&P IT Large Cap			S&P IT Mid Cap			S&P IT Small Cap		
	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Total Return (%)	19.62	23.71	20.02	−1.55	2.17	0.84	2.54	6.76	4.63	1.24	5.69	3.26
Volatility (%)	25.29	24.97	23.53	26.46	25.23	24.66	28.94	27.92	27.00	28.76	27.67	27.18
Kurtosis	3.05	2.98	3.13	3.34	3.36	3.42	3.29	3.41	3.32	3.23	3.28	3.38
Skewness	0.18	0.14	0.05	−0.29	−0.19	−0.25	−0.15	−0.11	−0.25	−0.41	−0.38	−0.38
VaR (%)	35.48	34.00	32.61	42.95	39.96	39.44	45.67	42.90	42.05	45.74	42.80	42.69
CVaR (%)	45.22	43.02	42.69	53.20	49.84	49.84	58.02	56.79	56.79	75.07	59.36	59.36
MDD (%)	40.94	40.69	40.95	64.10	55.21	53.40	57.94	50.94	50.94	56.92	54.60	54.60
Sharpe Ratio	0.69	0.87	0.77	−0.14	0.01	−0.04	0.01	0.17	0.10	−0.03	0.13	0.05
Sortino Ratio	1.21	1.56	1.32	−0.19	0.01	−0.06	0.02	0.25	0.14	−0.04	0.19	0.07

Table 3.4: Summary Statistics for the Efficiency Scores of Efficiency Decile Portfolios

Table 3.4 presents the summary statistics – mean, standard deviation, maximum value, median and minimum value – for efficiency scores of the decile portfolios. At the beginning of each investment quarter t, all firms are sorted into deciles based on their average efficiency scores in descending order to form 10 efficiency-based portfolios. The top efficiency decile portfolio is comprised of the most efficient firms while the bottom efficiency decile portfolio is comprised of the least efficient firms.

Efficiency-based Portfolios	Mean	Standard Deviation	Maximum	Median	Minimum
1	1.00	0.00	1.00	1.00	1.00
2	0.99	0.00	1.00	0.99	0.97
3	0.94	0.01	0.96	0.94	0.92
4	0.86	0.02	0.91	0.86	0.83
5	0.79	0.02	0.84	0.79	0.76
6	0.72	0.01	0.76	0.72	0.69
7	0.67	0.01	0.69	0.67	0.64
8	0.62	0.01	0.64	0.62	0.59
9	0.57	0.01	0.59	0.57	0.54
10	0.50	0.01	0.53	0.51	0.47
Whole Sample	0.77	0.01	1.00	0.76	0.47

Table 3.5: The Correlation Matrix for the Efficiency Score and Profitability and Valuation Measures

Table 3.5 presents the correlation matrix for the efficiency score (ES) against prior year's profitability and valuation measures, namely ROE, ROA, price to operating cash ratio (P/CF), and price to earnings ratio (P/E). *, **, *** represent significance at 10%, 5% and 1% respectively.

	ES	ROE	ROA	P/CF	P/E
ES	1.00				
ROE	−0.07	1.00			
ROA	−0.15	0.93***	1.00		
P/CF	−0.01	−0.06	−0.06	1.00	
P/E	−0.02	−0.18*	−0.19*	0.10	1.00

We now move on to discuss the empirical results obtained from the rolling stock universe. The discussion is divided into three parts corresponding to the three categories of performance metrics described in Section 3.2.4: return, risk and risk-return trade-off indicators. Throughout the discussion, we primarily focus on the performance of the top efficiency decile portfolio with an average efficiency score of 1.00 relative to the residual-based benchmark. Table 3.4 provides more details on the distribution of the efficiency scores of the efficiency decile portfolios. In Table 3.5, we also provide a correlation matrix for the efficiency score against prior year's profitability and valuation measures, namely ROE, ROA, price to operating cash ratio (P/CF), and price to earnings ratio (P/E).[25] As can be seen from the matrix, the correlation between the efficiency score and each of the profitability

[25] We are thankful to an anonymous referee for suggesting to include this correlation matrix.

and valuation measures is low and statistically insignificant. This shows how dissimilar the efficiency score is from common measures used for evaluating firm's future performance.

Table 3.6: Performance of the Top Efficiency- and Residual-based Portfolios

Table 3.6 presents the return, risk and risk-return trade-off metrics of the top efficiency- and residual-based portfolio comprised of constituents from the rolling stock universe. Panel A and B show results obtained before and after incorporating transaction costs respectively. The estimates for β coefficients for the factor models are given in Appendix A.4. The t-stats are shown in parenthesis. *, **, *** represent significance at 10%, 5% and 1% respectively.

Panel A. Before Transaction Costs

	Efficiency-based Portfolio			Residual-based Portfolio		
	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Total Return (%)	12.93	15.15	15.67	2.73	20.33	14.10
One-factor α (%)	10.97**	10.22**	11.68***	2.89	15.89**	11.22*
	(2.42)	(2.21)	(2.52)	(0.46)	(2.44)	(1.79)
Three-factor α (%)	8.99**	7.89*	10.38**	0.05	11.74**	6.94
	(2.23)	(1.92)	(2.42)	(0.01)	(2.19)	(1.33)
Six-factor α (%)	9.20**	8.35**	11.38***	0.56	12.88***	8.38*
	(2.31)	(2.06)	(2.74)	(0.11)	(2.50)	(1.67)
Turnover Ratio (%)	85.33	86.18	87.04	231.23	213.74	210.45
Volatility (%)	24.88	24.40	24.99	34.98	34.60	34.15
Kurtosis	3.07	3.11	3.25	3.09	3.04	3.65
Skewness	−0.02	−0.01	0.23	0.15	0.02	0.12
VaR (%)	36.51	35.19	36.00	55.01	49.84	50.69
CVaR (%)	46.63	44.71	42.97	63.18	62.75	70.63
MDD (%)	48.70	48.50	47.29	66.06	50.57	58.19
Sharpe Ratio	0.43	0.54	0.55	0.02	0.53	0.36
Sortino Ratio	0.69	0.88	0.94	0.03	0.87	0.57

Panel B. After Net Transaction Costs of 50 Basis Points

	Efficiency-based Portfolio			Residual-based Portfolio		
	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Total Return (%)	11.90	14.11	14.61	0.45	17.55	11.55
One-factor α (%)	9.95**	9.22**	10.64**	0.63	13.20**	8.73
	(2.21)	(2.00)	(2.32)	(0.10)	(2.07)	(1.41)
Three-factor α (%)	7.99**	6.94*	9.37**	−2.19	9.12*	4.52
	(2.00)	(1.70)	(2.20)	(−0.42)	(1.73)	(0.88)
Six-factor α (%)	8.18**	7.39*	10.35***	−1.73	10.22**	5.93
	(2.07)	(1.84)	(2.51)	(−0.33)	(2.02)	(1.20)
Volatility (%)	24.85	24.36	34.99	34.43	34.07	34.15
Kurtosis	3.07	3.10	3.24	3.09	3.00	3.68
Skewness	−0.02	−0.02	0.22	0.16	0.00	0.12
VaR (%)	36.74	35.40	36.16	55.68	50.25	51.22
CVaR (%)	46.75	44.94	43.21	63.56	63.26	71.21
MDD (%)	49.50	49.15	47.91	69.13	52.79	59.86
Sharpe Ratio	0.39	0.50	0.51	−0.05	0.45	0.28
Sortino Ratio	0.62	0.80	0.86	−0.07	0.72	0.44

3.4.2.2 Portfolio Return Measures

Examining the return indicators given in Table 3.6, we see that the efficiency-based portfolio significantly outperforms the residual-based portfolio for 1- and 6-month lagged investments.

For these lagged investments, the residual-based portfolio has average total returns of 2.73% and 14.10% while the efficiency-based portfolio has those of 12.93% and 15.67%, which are higher by 1,020 and 157 basis points per annum. Moreover, the returns on the efficiency-based portfolio are less driven by the systematic risks than those on the residual-based portfolio. For instance, the efficiency-based portfolio has notably higher factor α's than the residual-based portfolio for 1- and 6-month lagged investments. In particular, it has nearly 8.55% higher one-, three- and six-factor α's on average for the 1-month lagged investment. Such strong performance of the efficiency-based portfolio, especially after controlling for the various sources of risk, suggests that information embedded in an efficiency score is not just a proxy for known systematic risk factors.

The outperformance of the efficiency-based portfolio becomes even more apparent when transaction costs are considered. Because of the quarterly stock selection and monthly re-balancing applied to the portfolios, both efficiency- and residual-based portfolios are subject to high transaction costs. Nevertheless, transaction costs have more negative impacts on the performance of the residual-based portfolio. For example, assuming relatively large transaction costs of 50 basis points,[26] the average total returns of the residual-based portfolio decrease by approximately 2.53% while those of the efficiency-based portfolio decrease by only 1.05%. This is mainly because the residual-based portfolio has higher turnover ratios, which are about 2.5 times the turnover ratios of the efficiency-based portfolio.

Finally, in contrast to the residual-based portfolio, the efficiency-based portfolio exhibits comparable and consistent performance over all three investment lags. This shows that the efficiency-based strategy is less sensitive to the reporting lags and thus, is more robust.

3.4.2.3 Portfolio Risk Measures

Examining the risk measures reported in Table 3.6, we see that the standard deviations of the returns on the efficiency-based portfolio are 24.88%, 24.40% and 24.99% for 1-, 3- and 6-month lagged investments respectively. These are nearly 10% lower than the values of the

[26] In the market today, the transaction cost for non-retail traders is less than 10 basis points.

residual-based portfolio. Kurtosis is highest at 3.25 for the returns on the 6-month lagged efficiency-based investment. The return distribution of the efficiency- and residual-based portfolios have comparable values of kurtosis and skewness. The effect of transaction costs on volatility, kurtosis and skewness is insignificant for both efficiency- and residual-based portfolios.

In terms of extreme risk measures, VaR and CVaR are about 16% and 21% lower on average, respectively, for the efficiency-based portfolio. This implies that the residual-based portfolio has a return distribution with a heavier left tail and therefore, is more likely to experience extreme losses than the efficiency-based portfolio. When transaction costs are taken into account, VaR and CVaR increase by about 54 and 49 basis points, respectively, for the residual-based portfolio while they increase by about 20 basis points for the efficiency-based portfolio.

Lastly and perhaps most importantly, compared to the residual-based portfolio, the efficiency-based portfolio has lower and more consistent maximum drawdown values of about 48% for all three lagged investments. After transaction costs, maximum drawdowns of the efficiency-based portfolio increase by less than 1% while those of the residual-based portfolio increase by 2% on average.

3.4.2.4 Portfolio Risk-Return Trade-off Measures

The observations on the risk-adjustment performance follow directly from our analyses of the return and risk measures above. As can be seen from Table 3.6, the efficiency-based portfolio has Sharpe ratios of 0.43, 0.54 and 0.55 respectively for 1-, 3- and 6-month lagged investments. Although these values are lower than the ones obtained from the preliminary testing, they are still much higher than those of the market indices. In contrast, the residual-based portfolio has lower and less consistent values of Sharpe ratios – 0.02, 0.53 and 0.36 – for different investment lags. We can make similar observations about the values of Sortino ratios and further deduce that the efficiency-based portfolio has considerably lower downside risk than the residual-based portfolio, except for the 3-month lagged investment.

Both Sharpe and Sortino ratios of the efficiency-based portfolio are also less affected by transaction costs than the residual-based portfolio, conforming to our previous analyses of the return and risk measures.

In summary, the performance of the efficiency-based portfolio is consistently stronger than that of the residual-based portfolio and the S&P IT sector indices even in a stressful market environment. These results demonstrate convincingly the strength and stability of the efficiency-based stock selection methodology.

3.4.3 Robustness of the Efficiency-based Stock Selection Strategy

Table 3.7: Performance of the Bottom Efficiency- and Residual-based Portfolios

Table 3.7 presents the return, risk and risk-return trade-off measures of the bottom efficiency- and residual-based portfolio returns before transaction costs. The estimates for β coefficients for the factor models are given in Appendix A.4. The t-stats are shown in parenthesis. *, **, *** represent significance at 10%, 5% and 1% respectively.

	Efficiency-based Portfolio			Residual-based Portfolio		
	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Total Return (%)	−0.45	5.86	2.16	11.56	7.70	8.19
One-factor α (%)	−0.65	1.62	−0.73	10.87*	3.17	5.05
	(−0.11)	(0.29)	(−0.13)	(1.93)	(0.62)	(0.90)
Three-factor α (%)	0.14	0.83	−1.41	9.79**	0.76	2.76
	(0.03)	(0.19)	(−0.33)	(2.00)	(0.18)	(0.58)
Six-factor α (%)	0.22	1.15	−0.36	9.95**	1.07	3.61
	(0.05)	(0.26)	(−0.09)	(2.02)	(0.25)	(0.77)
Volatility (%)	32.84	31.04	30.47	31.33	30.16	30.67
Kurtosis	3.95	3.52	3.84	2.94	2.76	2.80
Skewness	−0.23	0.14	−0.11	−0.19	−0.04	0.04
VaR (%)	52.56	48.03	48.15	46.95	46.15	46.83
CVaR (%)	69.71	60.44	67.08	59.78	55.64	55.48
MDD (%)	67.80	56.62	58.94	57.64	63.35	59.52
Sharpe Ratio	−0.08	0.12	0.01	0.30	0.19	0.20
Sortino Ratio	−0.11	0.19	0.01	0.46	0.29	0.31

We study the effectiveness of efficiency scores in identifying investment worthy firms by comparing the performance of the top and bottom efficiency decile portfolios. As shown in Table 3.7, the bottom efficiency decile portfolio by far underperforms the top efficiency decile portfolio. Its average total returns for 1-, 3- and 6-month lagged investments are 13.37%, 9.29% and 13.52% lower than those of the top efficiency decile portfolio. The t-stats (p-values) obtained from the difference-in-means test for the spread between the two portfolio returns are also statistically significant with values of 1.71 (0.04), 1.20 (0.12) and

2.37 (0.01) for 1-, 3- and 6-month lagged investments respectively. With regard to the performance regression results for the bottom efficiency decile portfolio, none of the factor α's are statistically different from zero. Plus, the values of $\alpha_{top-bottom}$ reported in Table 3.8 are positive and statistically significant for all three lagged investments for all three models: the CAPM, the Fama-French three-factor model, and the six -actor model.

Table 3.8: Performance of the Top-Minus-Bottom Spread

Table 3.8 presents the descriptive statistics and Sharpe ratios of the top-minus-bottom spread for three lagged investments. The estimates for β coefficients for the factor models are given in Appendix A.4. The t-stats are shown in parenthesis. *, **, *** represent significance at 10%, 5% and 1% respectively.

	Top-minus-bottom Spread		
	1-Month	3-Month	6-Month
Total Return (%)	10.51	7.07	9.88
One-factor $\alpha_{top-bottom}$ (%)	13.17**	9.95*	11.71**
	(2.16)	(1.75)	(2.43)
Three-factor $\alpha_{top-bottom}$ (%)	11.19*	9.67*	11.62***
	(1.92)	(1.70)	(2.55)
Six-factor $\alpha_{top-bottom}$ (%)	10.74*	9.41*	11.25**
	(1.84)	(1.65)	(2.45)
Volatility (%)	18.88	17.65	14.65
Kurtosis	4.49	3.85	3.56
Skewness	0.55	0.06	0.47
Sharpe Ratio	0.44	0.28	0.54

Examining risk indicators shown in Table 3.7, we see that 1-, 3- and 6-month lagged investment returns on the bottom efficiency decile portfolio are 7.96%, 6.64% and 5.48% more volatile than those on the top efficiency decile portfolio. Similarly, on average, they also have 14% higher VaR, 21% higher CVaR and 13% higher maximum drawdown values. The return distribution of the bottom efficiency decile portfolio has slightly heavier left tail, and therefore, frequent small losses and extremely adverse scenarios are more likely compared with the top efficiency decile portfolio. This is further corroborated by the kurtosis and skewness values of the distribution of the top-minus-bottom spread shown in Table 3.8. The top-minus-bottom spread's heavy right tail together with its moderate volatility measures clearly points to the consistent outperformance of the top efficiency decile portfolio over the bottom efficiency decile portfolio.

In accordance with the return and risk measures described above, the bottom efficiency decile portfolio has considerably weak risk-adjusted performance. Its respective Sharpe ratios of −0.08, 0.12 and 0.01 and Sortino ratios of −0.11, 0.19 and 0.01 for 1-, 3- and 6-month lagged investments are actually lower than those of the market indices, namely S&P mid- and small-cap indices.

Table 3.9: Performance of the Efficiency Decile Portfolios

Table 3.9 presents the descriptive statistics and risk-return trade-off measures of the efficiency decile portfolios for 6-month lagged investment.

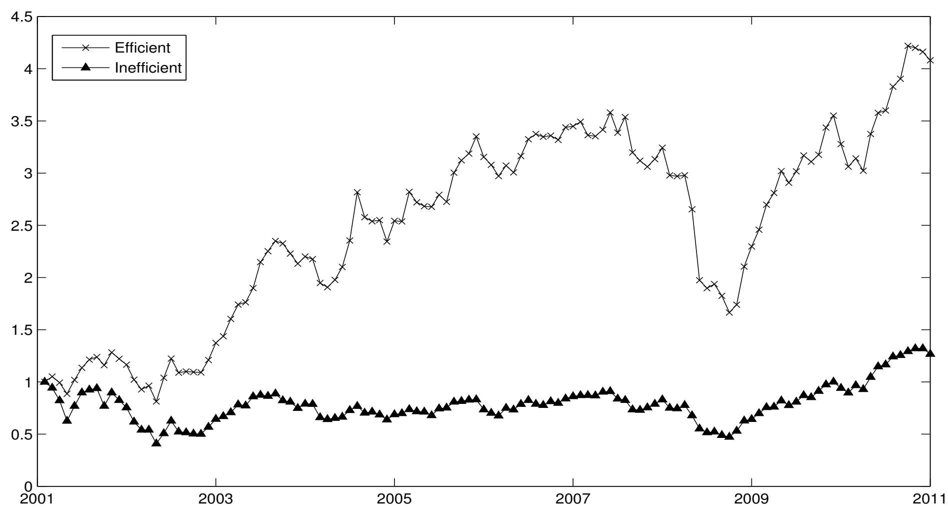
	Descriptive Statistics				Risk-Return Trade-off Measures	
Decile	Total Return (%)	Volatility (%)	Kurtosis	Skewness	Sharpe Ratio	Sortino Ratio
1	15.67	24.40	3.25	0.23	0.55	0.94
2	13.79	28.85	2.93	0.06	0.42	0.69
3	9.54	28.09	3.38	−0.05	0.26	0.40
4	12.67	30.04	3.63	0.15	0.38	0.62
5	11.63	29.22	3.16	0.04	0.34	0.54
6	6.83	28.02	3.09	−0.32	0.17	0.25
7	4.67	29.21	3.15	−0.29	0.09	0.13
8	7.22	29.09	3.31	−0.17	0.19	0.28
9	5.82	30.64	3.23	−0.13	0.13	0.19
10	2.16	31.04	3.84	−0.11	0.01	0.01

Table 3.9 presents the descriptive statistics and risk-return trade off measures of all ten efficiency decile portfolios for 6-month lagged investment.[27] As is apparent from the values shown in this table, relatively more efficient firms have stronger performance than relatively less efficient firms. For instance, the most efficient 150 firms, i.e. the firms in the top five efficiency deciles, have higher and yet, less volatile returns than the least efficient 150 firms, i.e. the firms in the bottom five efficiency deciles. The returns of the most efficient 150 firms are also more positively skewed than those of the least efficient 150 firms. Such results confirm the robustness of efficiency scores in identifying productively operating firms.

To further analyze the power of efficiency scores in differentiating efficient firms from inefficient firms, we compare the values of the top efficiency decile portfolio to those of the bottom efficiency decile portfolio over time. In Figure 3.3, we plot the 6-month lagged investment values of the top and bottom efficiency decile portfolios over a 10-year period starting in June of 2001. As can be seen from the figure, efficient firms certainly have stronger performance than inefficient ones. The difference in performance increases consis-

[27] Refer to Appendix A.7 for performance measures of all 10 decile portfolios for 1- and 3-month lagged investments.

Figure 3.3: Performance of the Most Efficient Firms vs. the Most Inefficient Firms



tently over time, though it slightly decreases during the 2008 financial crisis. These findings are in line with the previous findings of Alam and Sickles (1998) and Frijns *et al.* (2012).

Compared to the bottom residual decile portfolio, the bottom efficiency decile portfolio also conveys notably poor performance. Overall, the bottom efficiency decile portfolio has worse return, risk and risk-return trade-off measures. Such results demonstrate that efficiency scores are more robust and useful in determining investment worthiness of a firm than residuals obtained from regression analysis.

Table 3.10 reports the performance of the long-short efficiency-based portfolio comprised of the firms in the top and bottom efficiency deciles. More precisely, we present the results obtained from the 150-50 and 1.00 leveraged neutral portfolios[28] and compare them to the performance of the top efficiency decile portfolio. Examining the descriptive statistics of the 150-50 portfolio, we see that both average total returns and standard deviations of the three lagged investments are higher than those of the long-only portfolio. While the standard deviations are at most 2.01% higher than those of the long-only portfolio, the

[28] 110-10, 120-20, 130-30, 140-40 and 150-50 portfolios and leverage values of $0.50, 0.75, \ldots, 2.00$ are considered in our long-short portfolio analysis. Refer to Appendix A.6 for the complete results.

average total returns are at least 3.59% higher. Likewise, the 150-50 portfolio has stronger risk-adjusted performance with Sharpe ratios of 0.65, 0.67 and 0.83 and Sortino ratios of 1.20, 1.23 and 1.71 for 1-, 3-, and 6-month lagged investments. Such high Sortino ratios indicate that the 150-50 portfolio has lower downside semi-deviations than the long-only portfolio. In terms of the return distribution, the returns on the 150-50 portfolio exhibit heavier tails, but are more positively skewed.

Table 3.10: Performance of the Long-Short Efficiency-based Portfolios

Table 3.10 presents descriptive statistics and risk-adjusted performance measures of the long-short efficiency-based equally-weighted portfolios comprised of constituents from a rolling stock universe. The leveraged neutral portfolio uses the leverage value of 100%.

	150-50 Portfolio			Leveraged Neutral Portfolio		
	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Total Return (%)	19.33	19.30	23.22	12.79	9.21	12.20
Volatility (%)	26.37	25.93	25.56	18.93	17.70	14.68
Kurtosis	4.70	4.30	4.10	4.49	3.85	3.57
Skewness	0.62	0.57	0.77	0.56	0.06	0.48
VaR (%)	37.28	36.60	35.07	27.14	26.12	20.55
CVaR (%)	49.75	53.44	39.37	38.26	35.94	24.00
Max Drawdown (%)	49.56	48.68	47.05	33.69	31.60	28.44
Sharpe Ratio	0.65	0.67	0.83	0.56	0.40	0.60
Sortino Ratio	1.20	1.23	1.71	1.01	0.65	1.26

The returns on the leveraged neutral portfolio include risk-free earnings from the collateral and long-short returns, which are scaled down by the leverage value (see (3.3)). Compared to the long-only portfolio, the leveraged neutral portfolio has stronger risk-adjusted performance for 1- and 6-month lagged investments with Sharpe ratios of 0.56 and 0.60 and Sortino ratios of 1.01 and 1.26. The weak performance of the 3-month lagged investment could be explained by the relatively strong performance of the corresponding bottom-decile efficiency portfolio (see Table 3.7). Also, considering the fairly decent performance of the whole IT sector during this 3-month lagged investment period (see Table 3.3), our result does not seem too unreasonable.

Overall, the solid performance of the long-short portfolios empirically confirms the validity of efficiency scores in differentiating firms that are worthy of investment from those that are not. Our goal in the remainder of the chapter is to better understand the reasons behind the outperformance of the efficiency-based portfolio.

3.5 Firm Efficiency and the Cross-Section of Stock Returns

In this section, we examine whether firm efficiency has explanatory power for describing the cross-sectional behavior of stock returns in the U.S. IT sector. In Table 3.11 we report the results of the four Fama-MacBeth regressions: (3.4), (3.5), (3.6), and (3.7), introduced in Section 3.2.5. As described earlier, each of (3.5), (3.6), and (3.7) involves a set of control variables known to affect stock returns in addition to efficiency scores. These variables include the original Fama-French factors (firm size, book-to-market ratio, and stock β) and the measures of profitability, accruals, leverage, illiquidity, momentum, and reversal. For firm profitability, we use ROA[29] as our proxy.

Among the original Fama-French factors, firm size has the strongest explanatory power for the cross-section of stock returns. Its parameter estimates have p-values less than 1% for all regressions, except for the case, in which the 1-month lagged investment returns are regressed on the Fama-French factors and efficiency scores. In contrast, book-to-market ratio and stock β are statistically insignificant when efficiency score and other control variables are present in the regressions. This suggests that efficiency scores along with other return predictors embrace part of the information contained in book-to-market ratio and stock β. Among other control variables apart from the Fama-French factors and efficiency score, the momentum factor has the most significant explanatory power. The p-values for its parameter estimates are less than 1% for all three lagged investment returns.

In regard to firm efficiency, its parameter estimates β_{ES} are positive and mostly significant for the 1- and 6-month lagged investment returns despite the inclusion of other control variables (see Panel A and Panel C). The four Fama-MacBeth regressions of the 1-month lagged investment returns result in the respective parameter estimates β_{ES} of 1.55, 0.57, 1.04, and 1.05 with the corresponding p-values less than 1%, 14%, 3% and 2%. Similarly, those of the 6-month lagged investment returns result in the respective parameter estimates β_{ES} of 1.31, 0.53, 0.74, and 0.77 with the corresponding p-values less than 1%, 9%, 9%

[29] We tested both ROE and ROA as the proxy for firm profitability, and their results were very similar.

Table 3.11: Cross-Sectional Regression Analysis of Monthly Returns

Table 3.11 contains the results of the Fama-MacBeth regressions. It presents the average parameter values from the four separate cross-sectional regressions of $t+k$, $k = 1, 3$, and 6 months, monthly total returns on: (i) efficiency score (ES); (ii) efficiency score (ES), firm size (SIZE), book-to-market ratio (B/M), stock β (BETA); (iii) efficiency score (ES), firm size (SIZE), book-to-market ratio (B/M), stock β (BETA), ROA; and (iv) efficiency score (ES), firm size (SIZE), book-to-market ratio (B/M), stock β (BETA), ROA, and the measures of accruals (ACCR), leverage (LEV), illiquidity (ILLIQ), momentum (MOM) and reversal (REV) observed at t. The time-series t-stats are shown in parenthesis. *, **, *** represent significance at 10%, 5% and 1% respectively.

Panel A. $t+1$-month Monthly Stock Returns

β_{ES}									
1.55***									
(2.76)									
β_{ES}	β_{SIZE}	β_{BM}	β_{BETA}						
0.57	−0.14**	0.10	0.35						
(1.49)	(−2.38)	(0.64)	(0.62)						
β_{ES}	β_{SIZE}	β_{BM}	β_{BETA}	β_{ROA}					
1.04**	−0.25***	0.25	0.84	0.01					
(2.22)	(−3.96)	(1.42)	(1.29)	(1.07)					
β_{ES}	β_{SIZE}	β_{BM}	β_{BETA}	β_{ROA}	β_{ACCR}	β_{LEV}	β_{ILLIQ}	β_{MOM}	β_{REV}
1.05**	−0.28***	−0.02	0.56	0.01	0.16**	−0.38	0.00	−0.01***	0.00
(2.27)	(−4.05)	(−0.14)	(0.96)	(1.33)	(2.32)	(−0.19)	(−0.49)	(−2.72)	(0.26)

Panel B. $t+3$-month Monthly Stock Returns

β_{ES}									
0.84*									
(1.64)									
β_{ES}	β_{SIZE}	β_{BM}	β_{BETA}						
0.27	−0.24***	0.06	0.82						
(0.81)	(−3.64)	(0.36)	(1.30)						
β_{ES}	β_{SIZE}	β_{BM}	β_{BETA}	β_{ROA}					
0.30	−0.23***	0.09	0.74	0.00					
(0.69)	(−3.66)	(0.56)	(1.21)	(0.37)					
β_{ES}	β_{SIZE}	β_{BM}	β_{BETA}	β_{ROA}	β_{ACCR}	β_{LEV}	β_{ILLIQ}	β_{MOM}	β_{REV}
0.53	−0.25***	−0.06	1.14*	0.00	0.04	−1.75	0.00	−0.01***	−0.02**
(1.18)	(−3.91)	(−0.37)	(1.87)	(0.47)	(0.62)	(−0.81)	(0.41)	(−2.63)	(−2.39)

Panel C. $t+6$-month Monthly Stock Returns

β_{ES}									
1.31***									
(2.71)									
β_{ES}	β_{SIZE}	β_{BM}	β_{BETA}						
0.53*	−0.21***	0.09	0.41						
(1.70)	(−3.33)	(0.52)	(0.65)						
β_{ES}	β_{SIZE}	β_{BM}	β_{BETA}	β_{ROA}					
0.74*	−0.22***	0.11	0.45	0.01					
(1.71)	(−3.54)	(0.63)	(0.72)	(0.79)					
β_{ES}	β_{SIZE}	β_{BM}	β_{BETA}	β_{ROA}	β_{ACCR}	β_{LEV}	β_{ILLIQ}	β_{MOM}	β_{REV}
0.77*	−0.22***	−0.07	0.19	0.01	0.04	2.51	0.00	−0.01***	−0.02**
(1.68)	(−3.47)	(−0.39)	(0.33)	(0.90)	(0.49)	(1.15)	(−1.00)	(−2.95)	(−1.88)

and 10%. Such results imply that to a certain extent, higher firm efficiency is associated with higher expected (future) stock returns. Although β_{ES} is insignificant for the 3-month lagged investment returns, firm efficiency still has stronger impact on the cross-section of stock returns than the other control variables, excluding firm size and stock β. The fact that the parameter estimates of the efficiency scores are statistically more significant than those of the profitability measure particularly infers that firm efficiency is different from firm profitability.

From these results, it is reasonable to deduce that firm efficiency contains firm-level information that is not present in other known return predictors. In conclusion, our empirical findings confirm that firm efficiency plays an important role in explaining the cross-section of stock returns in the U.S. IT sector.

3.6 The Risk of Efficiency Loss

The empirical results thus far suggest that investing in firms with higher efficiency measures could be profitable. One possible economic argument for such implication could be that investors receive compensation for the risk of efficiency loss, which reflects the tendency for efficient firms to lose their comparative efficiency over time, for instance, due to competitive pressures. If this were the case, we should first see a gradual decline in profitability of various strategies we have tested for higher lagged investments.

In order to confirm this conjecture, we construct a long-short portfolio consisting of the most and least efficient 150 firms in the sample, and examine its various $t+k$ lagged monthly returns where $k = 1, \ldots, 36$. A total of 19 investments starting at a different period of time in history (2001 – 2005) is considered to avoid history bias as much as possible. Figure 3.4 presents the scatter plot and the exponential fit of the average monthly returns of these investments obtained at different lags. The exponential fit has a statistically significant exponent of -0.0357 (at 95% confidence level), indicating a gradual decay in profitability of the long-short strategy. This demonstrates simultaneously the decrease in profitability of the efficient firms as well as the increase in that of the inefficient firms over time. The next logical step would be to develop a systematic risk factor related to this efficiency reversion and establish empirical evidence of the associated risk premium in the stock return data.

We build an operational efficiency factor (OE) closely following the method of Fama-

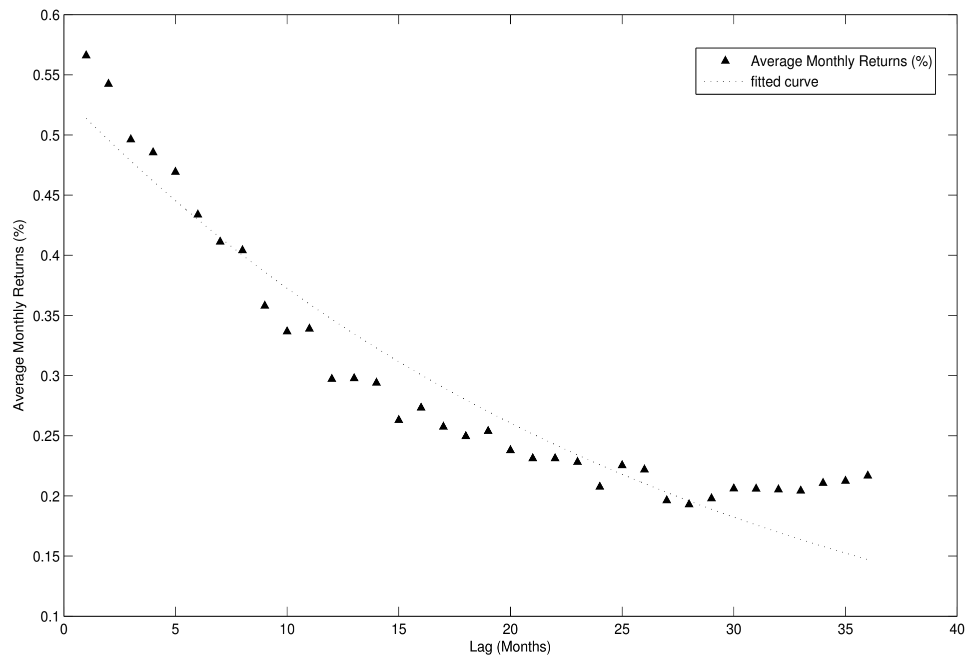


Figure 3.4: Decay of the Average Monthly Long-Short Portfolio Returns

French (1992). We partition the rolling stock universe into two groups, the efficient and the inefficient, based on the median efficiency score and form two equally-weighted portfolios, which are reformed and rebalanced annually. Then, by taking the difference between the returns of the efficient and inefficient portfolios, we derive OE; i.e.

$$OE = r_{efficient} - r_{inefficient},$$

where $r_{efficient}$ and $r_{inefficient}$ are the respective 6-month lagged monthly returns of the efficient and inefficient portfolios. To see the significance of this factor, we regress the 6-month lagged returns of the equally-weighted efficiency decile portfolios, which are also reformed and rebalanced annually, against OE and $r_m - r_f$. As can be seen from Table 3.12, β_{OE}'s are high (and positive) and statistically significant for portfolios with high efficiency scores, but on the other, they are low (and negative) and statistically significant for the ones with low efficiency scores. In addition to this regression, we run regressions of portfolio

Table 3.12: Operational Efficiency and Market Factor Regression

Table 3.12 presents the results of regressing the 6-month lagged monthly stock returns of the efficiency decile portfolios on the operational efficiency (OE) and market ($r_m - r_f$) factors. *, **, *** represent significance at 10%, 5% and 1% respectively.

Rank	α	β_{OE}	$\beta_{r_m - r_f}$	Adjusted R^2 (%)
1	0.44	0.83***	1.32***	65.21
	(1.08)	(4.27)	(15.00)	
2	0.14	0.65***	1.42***	65.74
	(0.33)	(3.09)	(15.09)	
3	0.26	0.49**	1.48***	66.58
	(0.58)	(2.29)	(15.25)	
4	0.00	0.57***	1.45***	72.96
	(0.00)	(3.19)	(17.81)	
5	0.17	0.43**	1.49***	73.69
	(0.44)	(2.32)	(17.99)	
6	0.16	−0.25	1.42***	72.33
	(0.39)	(−1.32)	(16.37)	
7	0.36	−0.14	1.56***	78.46
	(0.96)	(−0.82)	(19.58)	
8	0.34	−0.12	1.44***	75.45
	(0.91)	(−0.69)	(18.01)	
9	0.32	−0.63***	1.37***	68.37
	(0.72)	(−2.93)	(14.15)	
10	0.17	−0.74***	1.41***	72.11
	(0.41)	(−3.60)	(15.27)	

returns for all decile portfolios versus $r_m - r_f$, SMB, HML and including and then excluding OE to further look at usual inferences for model, coefficients, intercept and their significance. Table 3.13 presents the results of these regressions. When all three factors are included in the factor regression, OE is mostly more significant than HML with positive β_{OE} for the top five efficiency decile portfolios, and its inclusion improves the explanatory power of the regression model. Furthermore, β_{OE} is negative and significant for the last two efficiency decile portfolios. It appears that β_{OE} is negative and not significant for portfolios consist of firms with an efficiency score of higher than 0.60. We can also make an observation that the model intercepts are not significantly different from zero. OE therefore seems to have considerable explanatory power for explaining the return variations of efficiency decile portfolios.

In order to establish more sufficient arguments in support of that investors receive compensation for the risk of efficiency loss, we study firm efficiency through time and show that efficient firms tend to lose their efficiency and inefficient ones tend to gain it. Figure 3.5 shows two plots of the average efficiency scores of the efficient and inefficient portfolios over

Table 3.13: Operational Efficiency, Market, SMB and HML Factor Regression

Panel A shows the regressions of the monthly stock returns of the equally-weighted efficiency decile portfolios on the operational efficiency (OE), market ($r_m - r_f$), SMB, and HML factors. Panel B shows the regressions of monthly stock returns of the equally-weighted efficiency decile portfolios on the market ($r_m - r_f$), SMB and HML factors. The t-stats are shown in parenthesis. *, **, *** represent significance at 10%, 5% and 1% respectively.

Panel A

Rank	α	β_{OE}	$\beta_{r_m-r_f}$	β_{SMB}	β_{HML}	Adjusted R^2 (%)
1	0.30	1.06***	1.31***	0.58***	−0.79***	75.95
	(0.87)	(6.25)	(16.25)	(4.33)	(−5.99)	
2	−0.17	0.72***	1.29***	1.00***	−0.49***	77.41
	(−0.47)	(4.09)	(15.18)	(7.08)	(−3.55)	
3	−0.09	0.53***	1.31***	1.10***	−0.41***	78.58
	(−0.25)	(2.96)	(15.24)	(7.70)	(−2.91)	
4	−0.34	0.57***	1.27***	1.03***	−0.28***	84.50
	(−1.18)	(4.04)	(18.69)	(9.11)	(−2.52)	
5	−0.11	0.52***	1.38***	0.91***	−0.51***	83.94
	(−0.36)	(3.46)	(19.21)	(7.63)	(−4.34)	
6	−0.16	−0.17	1.28***	1.03***	−0.52***	84.53
	(−0.54)	(−1.13)	(17.84)	(8.67)	(−4.42)	
7	0.03	−0.09	1.41***	1.01***	−0.42***	88.55
	(0.13)	(−0.68)	(21.95)	(9.45)	(−4.02)	
8	−0.02	−0.13	1.26***	1.07***	−0.27***	87.08
	(−0.06)	(−0.97)	(19.56)	(10.05)	(−2.57)	
9	0.06	−0.49***	1.27***	0.88***	−0.63***	78.25
	(0.16)	(−2.64)	(14.38)	(6.02)	(−4.38)	
10	−0.03	−0.50***	1.37***	0.78***	−0.87***	83.46
	(−0.10)	(−3.05)	(17.48)	(6.00)	(−6.83)	

Panel B

Rank	α	$\beta_{r_m-r_f}$	β_{SMB}	β_{HML}	Adjusted R^2 (%)
1	0.72*	1.13***	0.64***	−0.56***	68.06
	(1.86)	(13.02)	(4.17)	(−3.85)	
2	0.12	1.16***	1.04***	−0.33**	74.35
	(0.32)	(13.79)	(6.95)	(−2.36)	
3	0.12	1.22***	1.13***	−0.29**	77.15
	(0.32)	(14.71)	(7.69)	(−2.11)	
4	−0.11	1.17***	1.06***	−0.15	82.45
	(−0.37)	(17.37)	(8.87)	(−1.37)	
5	0.10	1.29***	0.94***	−0.39***	82.42
	(0.31)	(18.41)	(7.56)	(−3.37)	
6	−0.23	1.31***	1.02***	−0.55***	84.50
	(−0.77)	(19.54)	(8.60)	(−4.92)	
7	0.00	1.43***	1.00***	−0.44***	88.60
	(−0.01)	(23.85)	(9.45)	(−4.39)	
8	−0.07	1.28***	1.06***	−0.30***	87.09
	(−0.25)	(21.35)	(10.01)	(−2.95)	
9	−0.13	1.36***	0.85***	−0.74***	77.14
	(−0.35)	(16.04)	(5.69)	(−5.19)	
10	−0.23	1.46***	0.75***	−0.98***	82.28
	(−0.69)	(19.25)	(5.60)	(−7.71)	

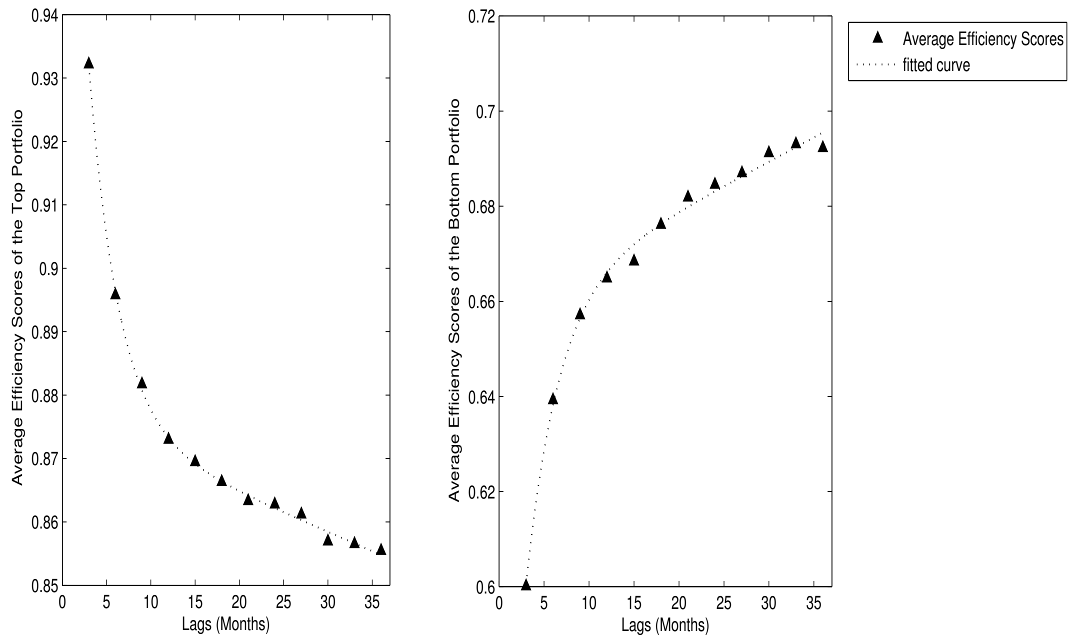


Figure 3.5: Decay (Growth) of the Average Efficiency Scores of the Efficient (Inefficient) Portfolio

time respectively. As can be seen from this figure, the average efficiency scores of the efficient (inefficient) portfolio decrease (increase) as time progresses. In Figure 3.6, we plot the logarithms of average monthly returns and those of average efficiency scores of the long-short portfolio on one chart. From this chart, we can observe that the decay rate of efficiency scores and that of monthly returns coincide. The empirical results of our analysis so far appear to be in support of the supposition that the risk of efficiency loss over time is the most likely source of excess returns of investing in efficient firms.

One may, however, argue that the loss of efficiency could be anticipated by investors. Considering that the most efficient firms on average have an efficiency score of 1/1, rational investors must expect a decline in these firms' efficiency in the future. Stock prices should therefore already reflect the loss of efficiency.[30] The question remains as to how firm efficiency could be regarded as a systematic factor. In order to construct a more convincing argument, we test whether the operational efficiency factor is correlated to (future) macroe-

[30]We are grateful to an anonymous referee for pointing this out.

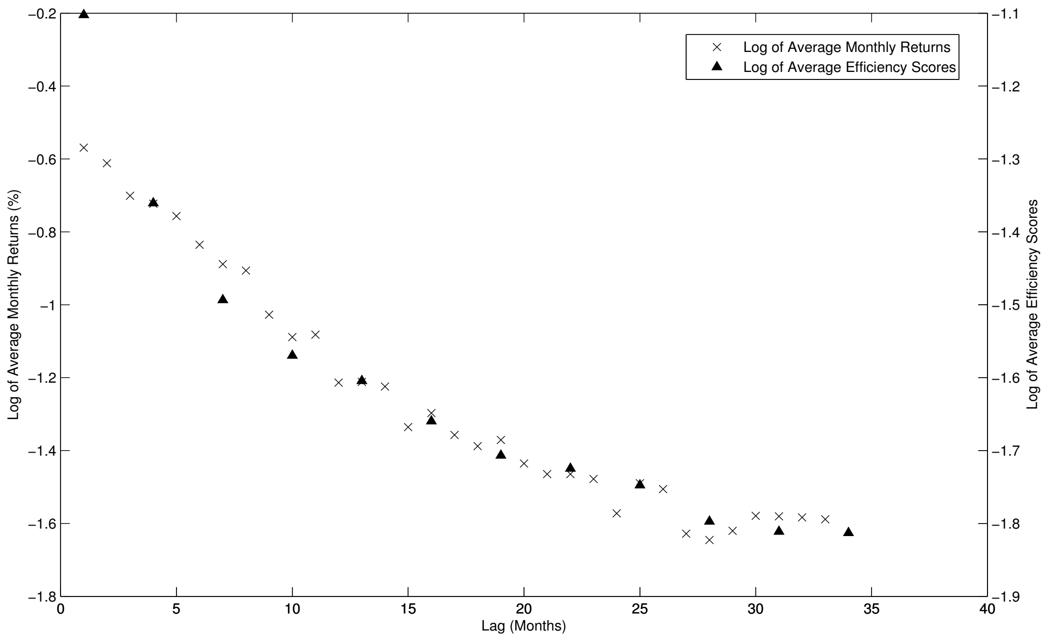


Figure 3.6: Decay of the Average Efficiency Scores of the Long-Short Portfolio vs. Decay of the Average Monthly Long-Short Portfolio Returns

conomic variables. In our analysis, we consider the operational efficiency factor constructed from 1-, 3-, and 6-month lagged investments, and the 6 to 11-quarter ahead unemployment rate and annual GDP growth.[31] Table 3.14 presents the results of the correlation analysis. We do not report on the correlation between the operational efficiency factor based on the 6-month lagged investment and the 11-quarter ahead macroeconomic variables due to the data unavailability at the time of our study.

As can be seen from the table, there is a statistically significant negative correlation between the operational efficiency factor and future unemployment. The strongest correlation of −0.41 is observed between the operational efficiency factor based on the 1-month lagged investment and the 11-quarter ahead unemployment rate. Unlike the unemployment rate, the future GDP growth is positively correlated with the operational efficiency factor.

[31] We also considered inflation and industrial production for our analysis and found that the correlations between the operational efficiency factor and these two were less significant.

Table 3.14: Correlations between the Operational Efficiency Factor and Future Macroeconomic Variables

Table 3.14 presents the correlations between the operational efficiency factor (OE) and future macroeconomic variables, namely the future unemployment rate, GDP growth and inflation. *, **, *** represent significance at 10%, 5% and 1% respectively.

Future Unemployment

Quarters ahead	1-Month	3-Month	6-Month
6	−0.30**	−0.27*	−0.26
7	−0.32**	−0.30*	−0.27*
8	−0.34**	−0.31**	−0.26*
9	−0.36**	−0.32**	−0.26
10	−0.39***	−0.35**	−0.28*
11	−0.41***	−0.37**	

Future GDP Growth

Quarters ahead	1-Month	3-Month	6-Month
6	0.29*	0.23	0.15
7	0.25	0.24	0.13
8	0.30*	0.29*	0.15
9	0.33**	0.22	0.21
10	0.29*	0.17	0.17
11	0.31*	0.31*	

From these results, we can deduce that the efficiency premium increases during the periods of economic growth, i.e. rising GDP and declining unemployment. This could be partly explained by the fact that investors tend to be less discerning during the periods of growth, and thus are more comfortable with making riskier investments in less efficient firms.

Overall, our results seem to indicate that the efficiency premium tends to vary systematically with macroeconomic conditions and business cycles. In which case, firm efficiency would still be considered as a systematic factor regardless of investors' recognition of efficiency loss.

3.7 Conclusions

This study investigated a potential link between firm efficiency and stock performance in the U.S. Information Technology (IT) sector. We analyzed this relationship by building a stock selection strategy based on firm efficiency and tracking its performance over a 10-year investment period. By means of DEA, we estimated firm's operational efficiency at utilizing its funds and producing profits. In particular, we applied DEA on a series of financial ratios to compute an efficiency score of a firm, which by design, measures the financial soundness of the firm and thus, indicates the direction of firm's future profitability. We formed various portfolios based on the estimated efficiency scores and evaluated their performance against market indices and residual-based benchmark constructed using regression analysis. We should note that our motivation for using the U.S. IT sector is on the one hand to limit the scope and data requirements of the empirical study and on the other hand to motivate a framework for further research directed at identifying sector, industry and market specific inputs (outputs) for DEA to improve its accuracy and efficacy.

In our empirical study, we found that efficient firms significantly outperform inefficient firms. The portfolio comprised of the firms in the top efficiency decile outperformed both the market indices and residual-based benchmark even after controlling for the known systematic risk factors. In contrast, the portfolio comprised of the firms in the bottom efficiency decile underperformed all the benchmarks. The long-short portfolios with long positions on the most efficient firms and short positions on the least efficient firms also exhibited strong performance over the 10-year investment period. Such results suggest the existence of a strong positive relationship between firm efficiency and stock performance. We also studied the role of firm efficiency in explaining the cross-section of stock returns. This analysis was conducted using Fama-MacBeth regressions where we regressed the cross-section of (future) stock returns on efficiency score and various factors known to affect stock returns. We found that firm efficiency has significant explanatory power in describing the cross-sectional attributes of stocks returns in the U.S. IT sector. Further analysis of the progression of efficiency scores and that of profitability supports the supposition that investing in highly efficient firms could be profitable as investors receive compensation for the risk of efficiency loss over time. In relation to the autocorrelated behavior of stock returns, our initial study confirmed that there is no clear evidence that including past stock performance would improve our strategy.

The current study can be extended both in technical details and in scope of application.

With regard to technical enhancements, first, since DEA methodology supports multiple outputs, it may be valuable to include different profitability measures, such as return on assets (ROA) and return on invested capital (ROIC), in addition to ROE as output variables. Second, one could incorporate window analysis to improve the estimate of the temporal average of firm efficiency. Third, the ranking of the firms can be further refined by applying various scale ranking methods available in the DEA literature. For instance, one could employ ranking methods based on multivariate statistical analysis, including canonical correlation analysis and discriminant analysis of ratios on the estimated efficiency scores (Friedman and Sinuany-Stern, 1998).

In terms of application, firstly, industry-level analysis can be applied to identify industry specific factors that can serve as the most suitable input (output) variables to the DEA methodology for assessing firms in each industry rather than using common variables for all the firms. One could also further catalogue or employ other quantitative, technical or macroeconomic indicators that can serve as the same. Secondly, the proposed investment strategy can be applied to other sectors and markets to evaluate its applicability and efficacy. In fact, the subsequent chapter presents an efficiency-based approach to currency selection. Lastly, and perhaps most importantly, our findings on the relationship between firm efficiency and the cross-section of stock returns as well as the risk of efficiency loss over time can be generalized by analyzing a larger sample of stocks. The results obtained from these analyses will bring more insights to the relationship between firm efficiency and share value performance.

To conclude, this study substantiated the existence of a strong positive relationship between firm efficiency and stock performance in the U.S. IT sector, and further provided empirical evidence on the practicality and robustness of the efficiency-based stock selection strategy founded on DEA models. Finally, the conservative treatment of data in our backtests should only serve to increase confidence in our findings.

Chapter 4

Quantitative Currency Selection Based on Macroeconomic Efficiency

4.1 Introduction

International economists and policymakers have long believed that the principal drivers of exchange rates are macroeconomic variables. Money supply, real output, interest rate, inflation and current account, for instance, are widely-accepted as the fundamental determinants of exchange rates. Based on these macroeconomic variables, numerous structural econometric models have been proposed to explain the observed behavior of exchange rates. Some of the well-known models include the flexible price monetary model (Bilson, 1978; Frenkel, 1976), the sticky price monetary model (Dornbusch, 1976; Frankel, 1979), and the sticky price asset model (Hooper and Morton, 1982). These three models were introduced during the period known as the "heroic age of exchange rate theory" (Isard, 1995) followed by the breakdown of the Bretton Woods system.[1] Since then they have been subject to most extensive testing in the literature. When tested in-sample, they forecasted exchange rates

[1]The Bretton Woods system of fixed exchange rates lasted from 1946 to 1973.

of major currencies exceedingly well. When tested out-of-sample, nevertheless, they all failed to outperform a simple random walk model in predicting exchange rates (Meese and Rogoff, 1983). Despite such results, macroeconomic variables are continued to be viewed as the major determinants of exchange rates by market participants. The currency selection strategy proposed in this study is also founded on this relationship between exchange rates and macroeconomic variables.

In building exchange rate models for a currency selection strategy, our aims are twofold. First, is to describe a potentially non-linear multilateral relationship between exchange rates and macroeconomic variables. Exchange rates are relative prices of national currencies. Therefore, under today's floating rate regime, they are determined by the interplay of supply and demand in foreign exchange markets to a certain extent. Supply and demand in currency markets are influenced by international trade and capital flows. Considering that trade and capital flows in the global economy are associated with various endogenous and exogenous macroeconomic variables of different countries, it is reasonable to presume a non-linear multilateral relationship between exchanges rates and macroeconomic variables.

Second, is to develop a multilateral aggregate measure of macroeconomic efficiency of countries. Exchange rates have long been considered as a reliable barometer of the state of the economy and the measure of international competitiveness of countries. While strong and appreciating currencies correspond to productive and efficient economies, weak and depreciating currencies correspond to slowing down and less efficient economies. It is, hence, sensible to compare macroeconomic efficiency of countries when evaluating performance of currencies for investment.

Exchange rate models in this study are derived from the three structural exchange rate models[2] mentioned earlier. In our model development, we first establish a multilateral framework using effective exchange rates and trade-weighted macroeconomic variables. This framework is used for transforming bilateral structural models into their multilateral

[2]These models are the flexible price monetary model, the sticky price monetary model and the sticky price asset model. Also, hereinafter, we refer to "structural exchange rate models" as "structural models."

counterparts. We then translate these multilateral models into DEA models for building an efficient frontier of countries. With respect to the estimated frontier, we quantify macroeconomic efficiency of each country into an efficiency score representing a consolidated measure of macroeconomic variables. We directly use this score for identifying currencies worthy of investment. The proposed currency selection strategy is evaluated using historical data and its successful performance would confirm the link between exchange rates and macroeconomic variables.

Although our models are obtained from the structural models, they are distinguished from conventional exchange rate models mainly in two aspects. First, traditional exchange rate models are based on standard regression analysis and hence, implicitly assume a linear relationship between exchange rates and macroeconomic variables. In contrast, the proposed exchange rate models are based on DEA, and are developed to capture a non-linear relationship between the two. Second, while traditional exchange rate models are primarily concerned with predicting future bilateral exchange rates between two countries, our models aim to measure multilateral macroeconomic efficiency of countries. Considering that macroeconomic efficiency of countries is linked to their currency performance to a certain degree, our additional objective is to provide a modeling framework for predicting future performance of currencies. To our knowledge, we are not aware of any other studies that examined structural models in a multilateral framework using DEA. In that regard, this study further provides a new way of looking at traditional exchange rate models.

The remainder of this chapter is organized as follows: in Section 4.2, we provide an overview of the three representative structural models; in Section 4.3, we detail the transformation of the structural models into their relevant DEA models; in Section 4.4, we present empirical results, and in Section 4.5, we present our conclusions.

4.2 An Overview of Structural Exchange Rate Models

The flexible price monetary model (Bilson, 1978; Frenkel, 1976), the sticky price monetary model (Dornbusch, 1976; Frankel, 1979), and the sticky price asset model (Hooper and Morton, 1982) take a monetary approach to the exchange rate determination. In the monetary approach, an exchange rate is defined as the relative price of two different currencies, and is therefore determined by the supply and demand for those currencies. The two key underlying assumptions in this approach are the goods market equilibrium and the financial market equilibrium. Accordingly, to a certain extent, the three models presume the conditions of purchasing power parity (PPP) and uncovered interest rate parity (UIP).[3] Their quasi-reduced form specifications are subsumed in the following general specification.

$$s_{i,j} = \alpha + \beta_m(m_i - m_j) + \beta_o(o_i - o_j) + \beta_r(r_i - r_j) + \beta_\pi(\pi_i - \pi_j) + \beta_z(z_i - z_j) + \epsilon \quad (4.1)$$

where:

- $s_{i,j}$ is the logarithm of the price of a currency j in a currency i[4];
- $m_i - m_j$ is the logarithm of the ratio of the money supplies of countries i and j;
- $o_i - o_j$ is the logarithm of the ratio of the real outputs of countries i and j;
- $r_i - r_j$ is the short-term interest rate differential of countries i and j;
- $\pi_i - \pi_j$ is the expected long-run inflation differential of countries i and j;
- $z_i - z_j$ is the current account differential of countries i and j;
- ϵ is the disturbance term.

We use the same notations for parameters and variables throughout the chapter unless otherwise mentioned.

[3] For details on the PPP and UIP conditions, refer to Appendix B.1.

[4] $s_{i,j}$ is the log level of the nominal bilateral exchange rate between countries i and j.

The flexible price monetary model sets β_π and β_z to zero in (4.1) under the assumption that the PPP condition holds continuously. This assumption is relaxed in the sticky price monetary model, which allows for slow domestic price adjustment and consequent deviations from PPP. More specifically, the sticky price monetary model allows for sustained inflation differentials across countries, and thus, sets only β_z to zero in (4.1). The sticky price asset model extends the sticky price monetary model. It allows for large and sustained changes in the long-run real exchange rate that are assumed to be correlated with the unanticipated shocks in the current account. Therefore, none of the coefficients in (4.1) is constrained to be zero. For further details on the flexible price monetary model, the sticky price monetary model, and the sticky price asset model, refer to Appendix B.1.

It is worth noting that imposing the constraint that the variables of two countries enter in a differential form implicitly assumes that the elasticities of two countries' macroeconomic variables are equal. Although making such a parsimonious assumption is conventional in empirical research, it has been pointed out that this could be a source of model mis-specification (Meese and Rogoff, 1983).

4.3 Methodology

4.3.1 The Estimation of Macroeconomic Efficiency

4.3.1.1 A Multilateral Framework

Suppose we have a universe of n countries where each country is a trade partner of one another. For building a currency portfolio from this universe, it is desirable to measure the relative performance of one currency against all others and select the best performers. In this regard, the usefulness of a bilateral exchange rate, which measures the relative strength of one currency to one other currency only, is limited, and so are the bilateral exchange rate models introduced in Section 4.2. An effective exchange rate,[5] on the other hand, is a more

[5] An effective exchange rate is also commonly known as a multilateral or a trade-weighted exchange rate.

general, multilateral measure that extends the bilateral exchange rate in the following way.

$$E_i = \sum_{j=1}^{n} w_{i,j} S_{i,j}$$

where E_i is the effective exchange rate of a country i, $w_{i,j}$ is the trade weight between countries i and j that satisfy

$$\sum_{j=1}^{n} w_{i,j} = 1 \text{ and } w_{i,i} = 0 \text{ for all } i,$$

and $S_{i,j}$ is the bilateral exchange rate of a country i with respect to a country j. Ergo the multilateral exchange rate of a country i is a weighted average of its bilateral exchange rates with respect to its trading partners.

We can employ the same weighting scheme to extend the bilateral exchange rate model into a multilateral one. For simplicity, let us assume that there is only one fundamental factor that determines exchange rates. Then, the bilateral exchange rate of a country i with respect to a country j can be estimated by the following (time-series) regression model. Note that henceforth we drop the subscript i from both α and β for notational simplicity.

$$s_{(i,j),t} = \alpha + \beta(a_{i,t} - a_{j,t}) + \epsilon_{i,t},\ t = 1, \ldots, T \tag{4.2}$$

where $s_{(i,j),t}$ is the logarithm of $S_{(i,j),t}$, and $a_{i,t}$ and $a_{j,t}$ are the macroeconomic variables of countries i and j respectively. By aggregating each of $a_{i,t}$ and $a_{j,t}$ by the trade weights between a country i and its trading partners and substituting $s_{(i,j),t}$ by the logarithm of $E_{i,t}$, we obtain the following multilateral exchange rate model.

$$\mathcal{E}_{i,t} = \alpha + \beta_a x^a_{i,t} + \epsilon_{i,t},\ t = 1, \ldots, T$$

where $\mathcal{E}_{i,t}$ is the logarithm of the effective exchange rate of the country i and $x^a_{i,t} = a_{i,t} - \sum_{j=1}^{n} w_{(i,j),t} a_{j,t}$. The same transformation can be applied to each of the macroeconomic variables in (4.1) for deriving the respective multilateral versions of the flexible price monetary model, the sticky price monetary model, and the sticky price asset model. We will use the term "trade-weighted" to refer to the transformed macroeconomic variables. The following general (time-series) specification subsumes the multilateral specifications of the three structural models.

$$\mathcal{E}_{i,t} = \alpha + \beta_m x^m_{i,t} + \beta_o x^o_{i,t} + \beta_r x^r_{i,t} + \beta_\pi x^\pi_{i,t} + \beta_z x^z_{i,t} + \epsilon_{i,t},\ t = 1, \ldots, T \tag{4.3}$$

where:

- $x^m_{i,t} = m_{i,t} - \sum_{j=1}^{n} w_{(i,j),t} m_{j,t}$;
- $x^o_{i,t} = o_{i,t} - \sum_{j=1}^{n} w_{(i,j),t} o_{j,t}$;
- $x^r_{i,t} = r_{i,t} - \sum_{j=1}^{n} w_{(i,j),t} r_{j,t}$;
- $x^\pi_{i,t} = \pi_{i,t} - \sum_{j=1}^{n} w_{(i,j),t} \pi_{j,t}$;
- $x^z_{i,t} = z_{i,t} - \sum_{j=1}^{n} w_{(i,j),t} z_{j,t}$.

As the same as before, β_π and β_z are set to zero in the flexible price monetary model, β_z is set to zero in the sticky price monetary model, and none of the coefficients is set to zero in the sticky price asset model.

4.3.1.2 A Motivation for the Development of the DEA Structural Exchange Rate Models

We could use the multilateral exchange rate model (4.3) for computing the expected values of n currencies in the universe and measure the performance of each currency in terms of its deviation from its temporal mean. This approach, however, neglects cross-sectional effects that currencies may have on each other. As alternatives, we could conduct either cross-sectional analysis of n currencies at one specific point t in time or panel analysis that concerns with two dimensional (cross-sectional $\times$ time-series) panel data. The relevant cross-sectional and panel (data) regression models are given by

$$\mathcal{E}_{i,t} = \alpha + \beta_m x_{i,t}^m + \beta_o x_{i,t}^o + \beta_r x_{i,t}^r + \beta_\pi x_{i,t}^\pi + \beta_z x_{i,t}^z + \epsilon_{i,t},\ i = 1, \ldots, n; \tag{4.4}$$

and

$$\mathcal{E}_{i,t} = \alpha + \beta_m x_{i,t}^m + \beta_o x_{i,t}^o + \beta_r x_{i,t}^r + \beta_\pi x_{i,t}^\pi + \beta_z x_{i,t}^z + \epsilon_{i,t},\ i = 1, \ldots, n; t = 1, \ldots, T \tag{4.5}$$

respectively. Note that in (4.4) the subscript t is dropped from both α and β for notational simplicity, and the coefficients in (4.5) do not require any subscripts.

Being a cross-sectional model, these two models are more intuitive and thus, are more suitable for analyzing cross-sectional performance of currencies. However, there are three main problems. First, both models have an homogeneous specification, which assumes that the elasticities of macroeconomic variables are equal across all n countries. Second, the sample size of the cross-sectional currency study is typically small, and this is problematic for regression models, which generally require a large sample size to make reliable estimations. Lastly, as true in any linear regression models, these models presume a linear relationship between exchange rates and macroeconomic variables. Such a relationship, nonetheless, may not even exist as many of the macroeconomic variables included in the structural models are considered as endogenous variables.

The relative performance evaluation of currencies, hence, calls for a cross-sectional model that can handle the following three things: (i) a non-linear relationship among multiple endogenous and exogenous variables, (ii) a small sample size, and (iii) the elasticities of macroeconomic variables that are specific to each country. This has led us to choose DEA as our numerical tool for building a currency portfolio.

Unlike regression models based on central tendencies, DEA is a methodology directed to extremal processes and is known to be particularly adept at uncovering relationships that may remain hidden for other methodologies (Smith, 1990). There are largely three key advantages of using DEA over standard regression methods for assessing relative strength of currencies:

(i) DEA permits the use of multiple outputs, thus enabling us to incorporate multiple endogenous variables as output variables;

(ii) it works fine with a relatively small sample size (Bauer *et al.*, 1998; Evanoff and Israilevich, 1991; Grifell-Tatje and Lovell, 1997); and

(iii) it allows for input and output variable weights that are specific to each observation in the sample.

As an illustration, let us consider the basic additive DEA model (2.5) introduced in Chapter 2 for computing a relative efficiency measure of a particular $\text{DMU}_p, p \in \{1, \ldots, n\}$ in the sample. For this chapter, we will replace variables u, v, and w in (2.5) by γ, β and α respectively; i.e. the additive model is now given by

$$\begin{aligned} \min_{\alpha,\beta,\gamma} \quad & \alpha + \sum_{k=1}^{l} \beta_k x_{k,p} - \sum_{r=1}^{s} \gamma_r y_{r,p} \\ \text{subject to} \quad & \alpha + \sum_{k=1}^{l} \beta_k x_{k,i} - \sum_{r=1}^{s} \gamma_r y_{r,i} \geq 0, \ i = 1, \ldots, n, \\ & \gamma_r \geq 1, \ r = 1, \ldots, s, \\ & \beta_k \geq 1, \ k = 1, \ldots, l \end{aligned} \tag{4.6}$$

where $X = x_{k,j} \in \mathbb{R}^{l \times n}$ are the input parameters, $Y = y_{r,j} \in \mathbb{R}^{s \times n}$ are the output parameters, β and γ are the variables for input and output weights respectively. One can solve (4.6) for all n DMUs simultaneously by solving the following LP.

$$\begin{aligned} \min_{\alpha,\beta,\gamma} \quad & \sum_{i=1}^{n} \left(\alpha_i + \sum_{k=1}^{l} \beta_{k,i} x_{k,i} - \sum_{r=1}^{s} \gamma_{r,i} y_{r,i} \right) \\ \text{subject to} \quad & \alpha_i + \sum_{k=1}^{l} \beta_{k,i} x_{k,j} - \sum_{r=1}^{s} \gamma_{r,i} y_{r,j} \geq 0, \ i, j = 1, \ldots, n, \\ & \gamma_{r,i} \geq 1, \ r = 1, \ldots, s; i = 1, \ldots, n, \\ & \beta_{k,i} \geq 1, \ k = 1, \ldots, l; i = 1, \ldots, n. \end{aligned} \tag{4.7}$$

As is apparent from (4.7), α, β, and γ are specific to each DMU in the sample.

Suppose there is only a single output variable ($s = 1$). Let us introduce a supplementary variable ϵ_i for each i and impose constraints,

$$y_i = \alpha_i + \sum_{k=1}^{l} \beta_{k,i} x_{k,i} + \epsilon_i, \; i = 1, \ldots, n.$$

Then, by means of Afriat's inequalities in Afriat's Theorem (Afriat, 1967, 1972), (4.7) can be equivalently expressed in the following non-parametric least-squares form (Johnson and Kuosmanen, 2010).

$$\begin{aligned} \min_{\alpha,\beta,\epsilon} \quad & \sum_{i=1}^{n} \epsilon_i^2 \\ \text{subject to} \quad & y_i = \alpha_i + \sum_{k=1}^{l} \beta_{k,i} x_{k,i} + \epsilon_i, \; i = 1, \ldots, n, \\ & \alpha_i + \sum_{k=1}^{l} \beta_{k,i} x_{k,i} \leq \alpha_j + \sum_{k=1}^{l} \beta_{k,j} x_{k,i}, \; i, j = 1, \ldots, n, \\ & \epsilon_i \leq 0, \; i = 1, \ldots, n, \\ & \beta_{k,i} \geq 0, \; i = 1, \ldots, n; k = 1, \ldots, l. \end{aligned} \tag{4.8}$$

Suppose we have an effective exchange rate and trade-weighted macroeconomic variables in (4.3) as an output variable and input variables respectively. The first constraint in (4.8) then becomes the general multilateral specification of the three structural models with α_i and $\beta_{k,i}$ that are specific to each country i. In this respect, (4.8) can be viewed as a non-parametric, cross-sectional currency model with country-specific elasticities of macroeconomic variables. Furthermore, as (4.8) deals with macroeconomic variables, we can interpret nonzero ϵ_i resulting from (4.8) for each country i as its measure of macroeconomic inefficiency relative to its trading partners. Consequently, DEA enables us to analyze relative strength of currencies in terms of their corresponding countries' relative macroeconomic efficiency. This presumably is a more fundamental way of evaluating performance of currencies.

4.3.1.3 DEA Structural Exchange Rate Models

Consistent with our choice of a DEA model in Chapter 3, we select the weighted additive model (WAM) for measuring relative macroeconomic efficiency of countries. Similar to our previous reasoning, we choose this DEA model because it supports negative and interval scale variables, and provides an efficiency score (Lovell and Pastor, 1995).[6] Refer to Section 3.2.1 for technical details of the model.

We consider three different DEA model specifications for computing relative macroeconomic efficiency of countries. The first specification can be viewed as the direct DEA counterpart of the three multilateral structural models: the flexible price monetary model, the sticky price monetary model, and the sticky price asset model, described in Section 4.3.1.1. In this specification, an effective exchange rate is used as a single output variable and the rest of the trade-weighted macroeconomic variables are used as input variables. Henceforth, we refer to this version of the DEA structural models as the single-output DEA model.

In the second specification, we leverage DEA's ability to handle multiple outputs. In theoretical exchange rate models, it is customary to regard money supply and real output as exogenous variables, and interest rate, inflation and current account as endogenous variables. These endogenous variables are, nonetheless, treated as legitimate regressors in ordinary or generalized least-squares regressions of the structural models (Meese and Rogoff, 1983). In contrast, we classify endogenous variables as output variables in our second DEA model specification. We refer to this specification as the multiple-output DEA model.

In the last specification, we augment the multiple-output DEA model by including lagged exchange rates as an additional input variable. A number of studies found that the addition of lags allowed the traditional exchange rate models to outperform a random walk model. Examples of these studies include works by Woo (1985), Boughton (1987), and Schinasi and Swamy (1989).[7] We refer to the last specification as the extended-multiple-

---

[6] Because we are dealing with macroecomic variables, the presence of negative and interval scale variables measured in different units is unavoidable.

[7] Schinasi and Swamy (1989) also included time-varying parameters.

output DEA model. Table 4.1 summarizes the three DEA model specifications for each of the three structural models. We should note that all variables included in the DEA models are consistent with those included in the multilateral structural models. Additionally, we apply window analysis to incorporate panel data in the DEA models.

Table 4.1: DEA Structural Exchange Rate Model Specifications

Table 4.1 presents the three DEA model specifications for each of the three structural models: the flexible price monetary model, the sticky price monetary model, and the sticky price asset model. All output and input variables are trade-weighted variables. See Section 4.3.1.1 for more details.

I. Single-Output DEA Model

Model	*Output Variables*	*Input Variables*
Flexible Price Monetary Model	Exchange rate	Money supply, interest rate real output
Sticky Price Monetary Model	Exchange rate	Money supply, interest rate, real output, inflation
Sticky Price Asset Model	Exchange rate	Money supply, interest rate, real output, inflation, current account

II. Multiple-Output DEA Model

Model	*Output Variables*	*Input Variables*
Flexible Price Monetary Model	Exchange rate, interest rate	Money supply, real output
Sticky Price Monetary Model	Exchange rate, interest rate, inflation	Money supply, real output
Sticky Price Asset Model	Exchange rate, interest rate, inflation, current account	Money supply, real output

III. Extended-Multiple-Output DEA Model

Model	*Output Variables*	*Input Variables*
Flexible Price Monetary Model	Exchange rate, interest rate,	Money supply, real output, lagged exchange rate
Sticky Price Monetary Model	Exchange rate, interest rate, inflation	Money supply, real output lagged exchange rate
Sticky Price Asset Model	Exchange rate, interest rate, inflation, current account	Money supply, real output, lagged exchange rate

4.3.2 Efficiency-based Portfolio Construction and Investment Strategy

The currency selection strategy we propose here closely follows the efficiency-based stock selection strategy described in Chapter 3. By design, macroeconomic efficiency of a country is measured relative to the frontier consisting of the best-performing countries. A country with a higher efficiency score is more efficient at managing its monetary policy compared to its peers, and therefore, its currency is expected to be stronger than foreign currencies. Accordingly, the efficiency-based investment strategy identifies currencies on or near the frontier as worthy of investment.

The following steps outline our currency investment strategy at time t.

Step 1 For each currency in the universe, a quarterly time series of efficiency scores is constructed up to time t.

Step 2 For each currency in the universe, an n-quarter exponentially weighted moving average of efficiency scores is computed from the time series constructed in Step 1.

Step 3 All the currencies in the universe are ranked by their average efficiency score in descending order.

Step 4 Currencies ranked in the top 25% are selected to form an equally-weighted portfolio.

The first step involves the computation of relative macroeconomic efficiency of countries. For each of the three multilateral structural models, its best performing DEA counterpart among the ones described in Table 4.1 is used for estimating countries' efficiency scores. In order to derive a more stable efficiency score, an average efficiency score is calculated using exponentially weighted moving average (EWMA) in Step 2. The smoothing factor for EWMA is chosen in the same way it was chosen for the efficiency-based stock selection strategy. In Step 3, currencies are sorted by their corresponding country's average efficiency score in descending order and are partitioned into four groups. In the last step, the top efficiency group comprised of the currencies closest to the frontier is selected to form a long-only equally-weighted investment portfolio. Over the investment horizon, the four steps are repeated at the end of each quarter, and the investment portfolio from Step 4 is rebalanced on a monthly basis to sustain equal dollar weights.

4.3.3 Benchmark Construction

As the structural models are originally presented in the form of regression models, our natural choice of a benchmark is regression analysis. In particular, we use a fixed effects model in panel data analysis. A fixed effects model can be viewed as a counterpart of window analysis in DEA and has a number of advantages over the simply pooled ordinary least-squares (OLS) procedure (Hsiao, 2003). For example, a simply pooled OLS model cannot adjust for currency specific effects, which if correlated with other explanatory variables, would produce omitted variables bias and thus, mis-specified models. This problem is serious as it results in flawed estimates. In contrast, a fixed effects model uses the currency specific intercepts to capture the unobserved and/or unmeasurable currency specific characteristics.

The formal representation of the model is given by

$$\mathcal{E}_{i,t} = \alpha_i + \beta' \mathbf{x}_{i,t} + \epsilon_{i,t},\ i = 1, ..., n; t = 1, \ldots, T \tag{4.9}$$

where $\mathcal{E}_{i,t}$ is the logarithm of the effective exchange rate of a currency i at time t, and $\mathbf{x}_{i,t}$ is a vector of trade-weighted macroeconomic variables included in each of the three multilateral structural models for a currency i at time t. Since intercept terms α vary across currencies, they are indexed by individual currency. Under- and over-valuation of currencies is determined in terms of residuals $\epsilon_{i,t}$ resulting from (4.9). We should note that since any regression model is data demanding, for estimating (4.9), we include as much history as needed, but as little as possible considering the unstable currency market during the 1990s.

Mirroring the efficiency-based investment strategy, we propose the following residual-based strategy.

Step 1 For each currency in the universe, a quarterly time series of residuals is constructed up to time t.

Step 2 For each currency in the universe, an n-quarter exponentially weighted moving average of residuals is computed from the time series constructed in Step 1.

Step 3 All the currencies in the universe are ranked by their average residuals in ascending order.

Step 4 Currencies ranked in the top 25% are selected to form an equally-weighted portfolio.

Since each step is analogous to that of the efficiency-based strategy, refer to Section 4.3.2 for a detailed explanation.

4.4 Empirical Results

Using historical data, we assess the appropriateness of each DEA structural model listed in Table 4.1 for measuring relative macroeconomic efficiency of countries. Two different sets of time periods are used for our testing: a model estimation (in-sample) period and a strategy implementation (out-of-sample) period. In the in-sample period, which is from 2000 to 2005, we compare the performance of investment strategies based on the three DEA structural model specifications: single-output, multiple output, and extended-multiple-output DEA models, with and without window analysis. Then, we select the best-performing DEA model specification for each structural model for out-of-sample testing. For window analysis, we try window sizes of 2, 4, 8, and 16 quarters. The investment strategy based on the finalized model is evaluated against market and strategic benchmarks in the out-of-sample period, which is from 2006 to 2011. Since the distribution of exchange rate regimes changed significantly between 1991 and the end of 1999 (Fischer, 2001),[8] we exclude the periods prior to 2000 from our testing.

The metrics used for performance evaluation include the first four moments of the cur-

[8] In 1991 almost 65% of emerging market countries had intermediate pegged exchange rate regimes. 5% had hard pegs and 30% floated. By the end of the decade, the numbers were very different. The proportion with intermediate regimes had dropped to 40% while the number with hard pegs and floats had risen to 10% and 50% respectively (Fischer, 2001).

rency return distribution: average total return,[9] volatility, skewness and kurtosis, and Sharpe ratio. All metrics reported in this study are annualized figures unless otherwise mentioned. Also, the 1% winsorization is applied to returns. For both testing periods, the sample currency universe consists of 20 – 22 currencies on average, and 1-, 3- and 6-month lagged investments are considered due to the reporting lag in the release of the macroeconomic data.

Table 4.2: Proxies for Macroeconomic Variables

Table 4.2 presents proxies used for macroeconomic variables included in the structural models.

Macroeconomic Variable	Proxy
Exchange rate	REER
Money supply	Broad money
Real output	Real GDP
Interest rate	3-month short-term interest rate
Inflation	Year-on-year changes in CPI
Current account	Current account

Table 4.2 lists proxies used for macroeconomic variables included in the structural models. As a proxy for effective exchange rates, we use real effective exchange rates (REERs) commonly used by economists and policymakers. For money supply, we use broad money as a proxy,[10] and for real output, we use real GDP, a macroeconomic measure of the value of economic output adjusted for price changes, as a proxy. Consistent with previous studies, 3-month short-term interest rate and year-on-year changes in consumer price index (CPI) are selected as the respective proxies for interest rate and inflation. As current account is directly available, we do not need a separate proxy for this variable. Refer to Appendix B.2 for the detailed data sources.

[9] Following the industry standard, total return of a currency is defined as the price return of the currency plus the accrued interest from holding it. The 3-month depository rate is used for computing accrued interest.

[10] We tried both narrow and broad money in our in-sample study and concluded that broad money is a more appropriate proxy for money supply.

4.4.1 The Model Estimation Period

As per the MAE and MSE analyses, a smoothing factor of 2-quarters is used for computing an exponentially weighted moving average of efficiency scores. Also, we use a window size of 2-quarters for all three DEA model specifications as it had the best in-sample performance among different window sizes that we tested.[11] The efficiency scores in our empirical study therefore represent how each country's economy is performing from one quarter to another against the economies of other countries. It is worth mentioning that the underlying assumption of window analysis is that there are no technical changes within each of the windows. However, this may not be plausible in practice due to the changes in regulation, policy, economic conditions or competitive situation, and is likely to make comparisons of units in different periods unfair and unrealistic. Considering that there is no theory or justification that underpins the definition of the window size (Tulkens and Eeckaut, 1995), and also considering the rapid economic growth of emerging markets as well as the number of economic crises in the last decade, a rather small window size of 2-quarters seems reasonable.

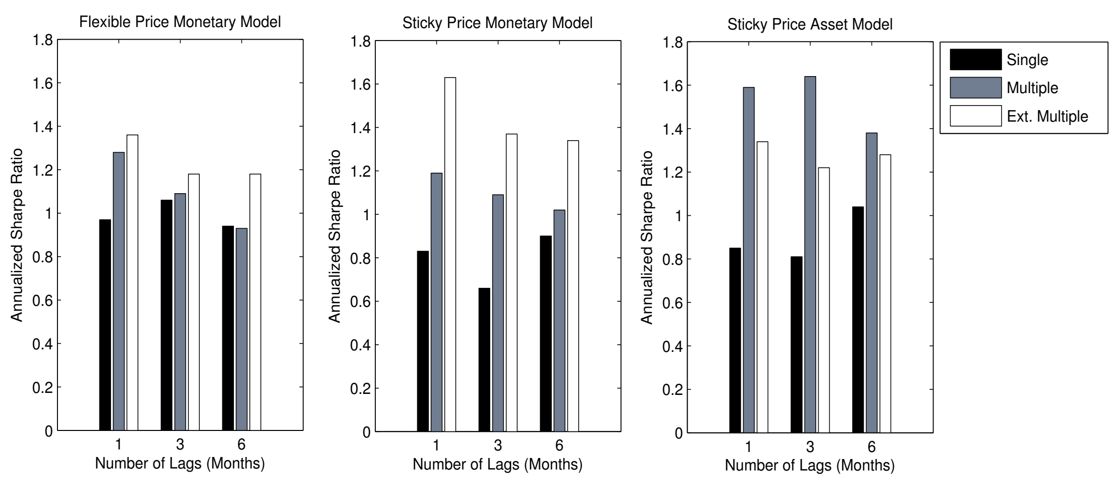


Figure 4.1: In-Sample Risk-Adjusted Performance

Figure 4.1 presents the summary results of the in-sample risk-adjusted performance of

[11] Window sizes of 1 (no window analysis), 2, 4, 8, and 16 quarters were tested in-sample, and the respective results are available in Appendix B.3.

the three DEA model specifications of the three multilateral structural models. As the bar charts illustrate, the extended-multiple-output DEA model exhibits the most stable and strongest performance for both the flexible price monetary model and the sticky price monetary model. For the sticky price asset model, the multiple-output DEA model shows significant outperformance over the other two DEA model specifications. The single-output DEA model has the weakest performance across all three structural models. Such results highlight the advantage of DEA's ability to accommodate a multiplicity of outputs. More detailed results are presented in Table 4.3. Consistent with the results shown on Figure 4.1, the average total returns of the multiple-output DEA model and the extended-multiple-output DEA model are higher than those of the single-output DEA model. All three DEA model specifications have comparable performance in terms of volatility and kurtosis. Based on the in-sample performance results, we choose the extended-multiple-output DEA model for both the flexible price monetary model and the sticky price monetary model, and the

Table 4.3: In-Sample Performance of the Efficiency-based Portfolios

Table 4.3 compares performance of the efficiency-based portfolios constructed from the single-output, multiple output, and extended-multiple-output DEA structural models.

	Flexible Price Monetary			Sticky Price Monetary			Sticky Price Asset		
Total Return (%)	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Single	8.79	9.41	8.46	7.92	6.56	8.17	6.94	6.74	7.98
Multiple	12.40	10.96	9.47	11.11	10.69	10.78	10.42	11.05	10.22
Ext. Multiple	12.31	11.47	11.47	12.91	11.68	12.14	10.23	9.55	9.88
Volatility (%)	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Single	6.91	6.91	6.72	7.02	6.85	6.75	5.71	5.77	5.64
Multiple	8.08	8.19	7.87	7.62	7.89	8.47	5.23	5.47	5.87
Ext. Multiple	7.52	8.00	7.94	6.63	7.03	7.46	6.07	6.14	6.07
Kurtosis	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Single	2.35	2.37	2.33	2.53	2.66	2.78	2.61	2.62	2.88
Multiple	2.39	2.03	2.19	2.49	2.35	2.39	2.31	2.10	2.38
Ext. Multiple	2.28	2.14	2.44	2.26	2.42	2.45	2.56	2.71	2.48
Skewness	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Single	0.09	0.10	0.17	0.15	0.02	0.30	0.05	0.09	0.36
Multiple	−0.01	0.02	−0.01	−0.14	0.07	0.07	−0.21	0.02	0.25
Ext. Multiple	−0.04	0.09	0.33	0.07	0.19	0.30	0.37	0.45	0.35
Sharpe Ratio	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Single	0.97	1.06	0.94	0.83	0.66	0.90	0.85	0.81	1.04
Multiple	1.28	1.09	0.93	1.19	1.09	1.02	1.59	1.64	1.38
Ext. Multiple	1.36	1.18	1.18	1.63	1.37	1.34	1.34	1.22	1.28

multiple-output DEA model for the sticky price asset model for out-of-sample testing.

4.4.2 The Strategy Implementation Period

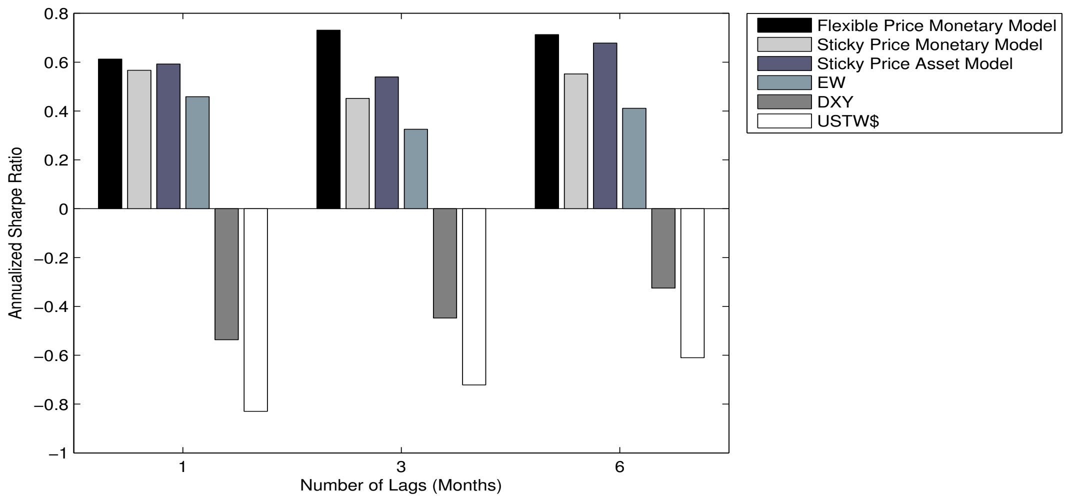


Figure 4.2: Risk-Adjusted Performance of the Efficiency-based Portfolios and Market Benchmarks

Figure 4.2 charts the Sharpe ratios of the efficiency-based portfolios based on the three structural models, U. S. dollar indices, and an equally-weighted (EW) portfolio consisting of all the currencies in the sample. Two market indices considered here are: the U. S. Dollar Index (DXY), which is a weighted average of the dollar's value compared with the basket of six other major currencies (Euro, Japanese yen, Pound sterling, Canadian dollar, Swedish krona and Swiss franc) and the USTW$ index, a related, broader-based dollar index, which uses a much larger basket of currencies including many of the developing market currencies. The outperformance of the efficiency-based portfolios, especially over the market indices, is apparent from the bar chart. Among the three structural models, the flexible price monetary model shows the strongest risk-adjusted performance while the sticky price monetary model shows the weakest.

We can make similar remarks about the outperformance of the efficiency-based portfolios in Figure 4.3. This figure compares: (i) performance of the top efficiency- and residual-based portfolios that consist of currencies ranked in the top 25% and (ii) performance of

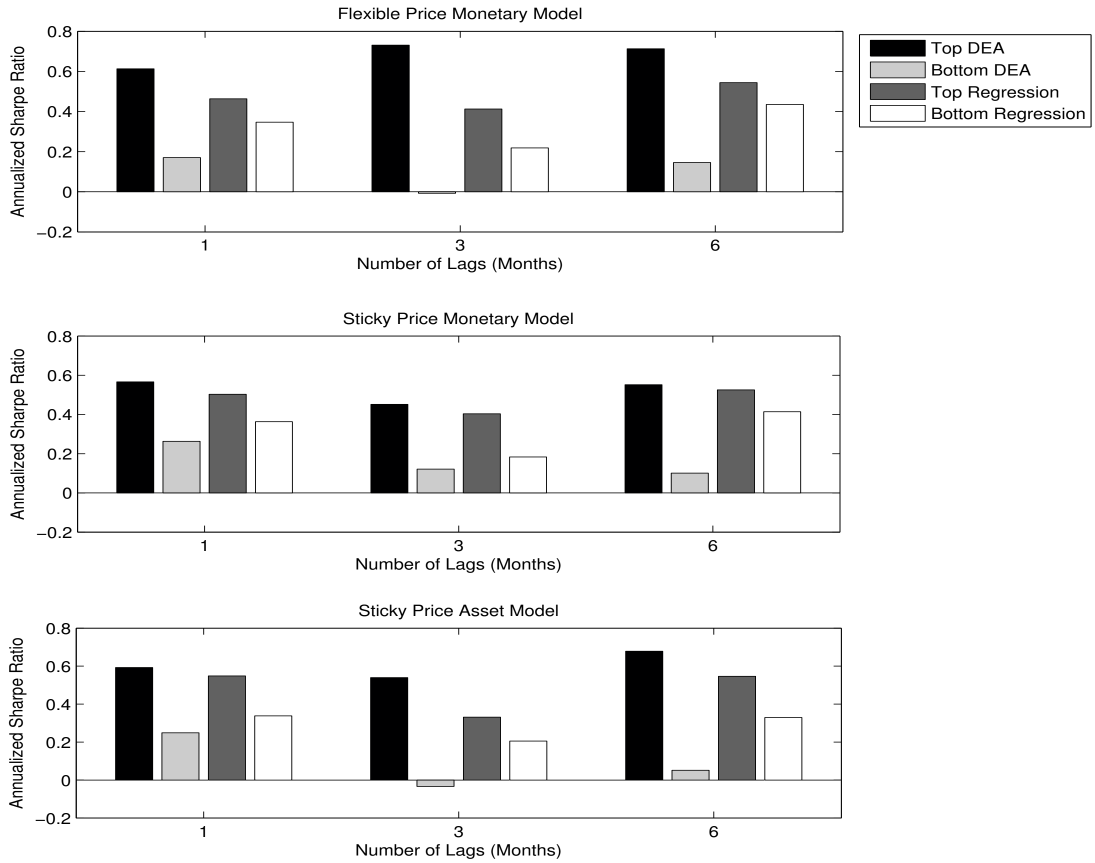


Figure 4.3: Risk-Adjusted Performance of the Efficiency- and Residual-based Portfolios

the bottom efficiency- and residual-based portfolios that consist of currencies ranked in the bottom 25%. We study the performance of the bottom efficiency- and residual-based portfolios in order to further examine the discriminatory power of macroeconomic efficiency scores and regression residuals. As the three bar charts demonstrate, for each structural model, its corresponding top and bottom efficiency-based portfolios have the highest and lowest Sharpe ratios respectively. Although the performance of the top efficiency-based portfolios is somewhat comparable with that of the top residual-based portfolios, the bottom efficiency-based portfolios significantly underperform their residual-based counterparts. In fact, the performance distinction between the top and bottom efficiency-based portfolios seems sharper than that between the top and bottom residual-based portfolios. Figure 4.4

further confirms this observation.

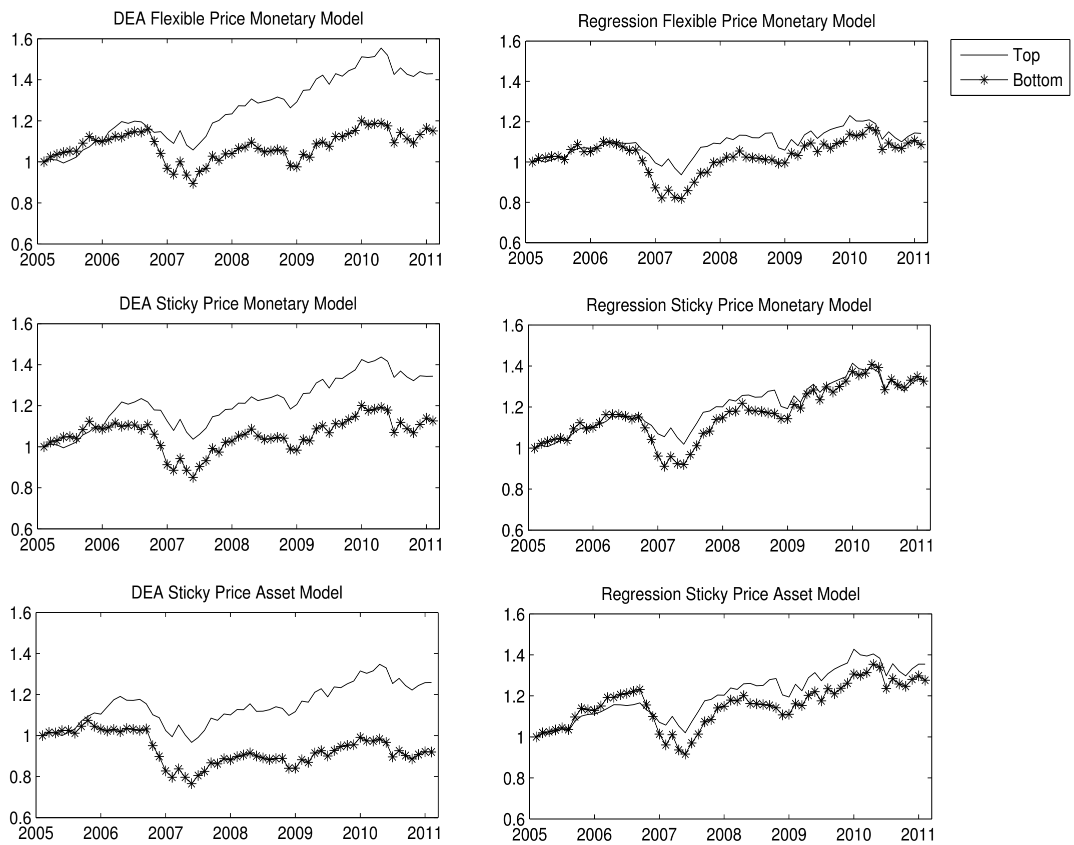


Figure 4.4: Performance of the Top vs. Bottom Efficiency- and Residual-based Portfolios

Figure 4.4 plots the values of the 6-month lagged investments of the top and bottom efficiency- and residual-based portfolios from the end of 2005 to the beginning of 2011. As can be seen from the figure, the top efficiency-based portfolios show considerably stronger performance than the bottom efficiency-based portfolios, and the performance gap between them widens as time progresses. In contrast, it is difficult to draw a clear distinction between the values of the top and bottom residual-based portfolios. For example, the values of the top and bottom residual-based portfolios actually seem to overlap with each other for the sticky price monetary model. This suggests that macroeconomic efficiency scores are a more accurate gauge of relative strength of currencies than regression residuals. Moreover, as the portfolio values shown on Figure 4.4 do not include depository rates, we can deduce that the top (bottom) efficiency-based portfolios' favorable (unfavorable) performance is largely owed to the strong (weak) performance of their constituent currencies.

Table 4.4 provides more details on the total and price return performance of the top and bottom efficiency-based portfolios. In terms of average return, the top efficiency-based portfolios have notably higher total and price returns than the bottom efficiency-based portfolios. The return spreads between the top and bottom efficiency-based portfolios were in fact statistically significant for the flexible price monetary model and the sticky price asset model according to the standard difference-in-means test (t-test). For example, for the 3-month lagged investment of the flexible price monetary model, the t-stats (p-values) for the total and price return spreads were 2.04 (0.02) and 2.23 (0.01) respectively. Similarly, for that of the sticky price asset model, the respective t-stats (p-values) for the total and price return spreads were 2.27 (0.01) and 2.59 (0.01).[12] In addition, consistent with the observations we made from Figure 4.4, both the total and price return spreads between the top and bottom residual-based portfolios were not statistically significant for any of the three structural models.

In terms of portfolio risk, as can be seen from Table 4.4, returns of the bottom efficiency-based portfolios are more volatile than those of the top efficiency-based portfolios. According to the standard difference-in-variances test (F-test), the spreads between the volatilities of these portfolio returns were statistically significant at 95% confidence level for the flexible price monetary model and the sticky price monetary model.[13] Plus, the return distribution of the bottom efficiency-based portfolios seems to have a heavier left tail than that of the top efficiency-based portfolios.

In summary, for the flexible price monetary model, the top efficiency-based portfolios have statistically higher returns and lower volatility than the bottom efficiency-based port-

[12]These results were obtained from the right tail t-test. For the sticky price monetary model, the results were less significant.

[13]These results were obtained from the left tail F-test. For the sticky price asset model, the results were less significant.

Table 4.4: Out-of-Sample Performance of the Top and Bottom Efficiency-based Portfolios

Table 4.4 presents performance metrics of the top and bottom efficiency-based portfolios for the flexible price monetary model, the sticky price monetary model and the sticky price asset model for 1-, 3-, and 6-month lagged investments. Panel A and B present performance metrics computed from total and price returns of the efficiency-based portfolios respectively.

Panel A: Total Return

	Flexible Price Monetary			Sticky Price Monetary			Sticky Price Asset		
Return (%)	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Top	6.94	7.68	7.40	6.54	5.36	6.08	7.00	6.51	7.46
Bottom	3.45	1.35	2.85	4.78	2.97	2.40	4.11	1.09	1.73
Volatility (%)	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Top	8.70	8.53	8.73	8.71	8.69	8.87	9.11	9.39	9.25
Bottom	10.85	11.47	11.45	12.05	12.59	12.10	10.07	10.33	10.55
Kurtosis	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Top	2.83	2.83	3.29	3.03	2.72	2.84	2.71	2.56	2.67
Bottom	4.86	4.29	3.01	4.00	3.62	3.30	3.03	3.09	3.26
Skewness	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Top	−0.27	−0.24	−0.38	−0.48	−0.48	−0.48	−0.34	−0.28	−0.29
Bottom	−0.96	−0.67	−0.27	−0.60	−0.33	−0.47	−0.39	−0.40	−0.56
Sharpe Ratio	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Top	0.61	0.73	0.71	0.57	0.45	0.55	0.59	0.54	0.68
Bottom	0.17	−0.01	0.15	0.26	0.12	0.10	0.25	−0.03	0.05

Panel B: Price Return

	Flexible Price Monetary			Sticky Price Monetary			Sticky Price Asset		
Return (%)	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Top	3.62	4.44	4.30	3.32	2.24	3.03	4.08	3.66	4.71
Bottom	−0.30	−2.24	−0.63	0.75	−1.06	−1.45	0.53	−2.42	−1.66
Volatility (%)	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Top	8.69	8.50	8.75	8.71	8.70	8.91	9.14	9.43	9.30
Bottom	10.90	11.54	11.49	12.08	12.62	12.19	10.06	10.41	10.68
Kurtosis	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Top	2.81	2.82	3.31	3.04	2.75	2.84	2.76	2.60	2.71
Bottom	4.81	4.28	3.01	3.99	3.59	3.27	2.99	3.13	3.32
Skewness	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Top	−0.28	−0.25	−0.37	−0.50	−0.50	−0.48	−0.38	−0.30	−0.31
Bottom	−0.91	−0.62	−0.23	−0.56	−0.32	−0.43	−0.34	−0.38	−0.55
Sharpe Ratio	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Top	0.23	0.35	0.36	0.20	0.09	0.21	0.27	0.24	0.38
Bottom	−0.17	−0.32	−0.16	−0.07	−0.20	−0.22	−0.11	−0.37	−0.27

folio. For the sticky price monetary model, although the top and bottom efficiency-based portfolios have comparable returns, the top efficiency-based portfolios have statistically lower volatility. For the sticky price asset model, although the top and bottom efficiency-based portfolios have comparable risk, the top efficiency-based portfolios have statistically higher returns.

Conforming to the observations we made on the portfolio return and risk measures, the top efficiency-based portfolios have substantially higher Sharpe ratios than the bottom efficiency-based portfolios. For instance, in terms of price returns, Sharpe ratios of the top efficiency-based portfolios are on average greater than 0.20 whereas those of the bottom efficiency-based portfolios are mostly below zero. Such strong price return performance of the top efficiency-based portfolios confirms that countries with higher macroeconomic efficiency scores have stronger currencies than those with lower scores. The outperformance of the top efficiency-based portfolios over the bottom efficiency-based portfolios moreover illustrates the power of efficiency scores in differentiating relatively strong currencies from relatively weak currencies.

Additionally, Table 4.5 presents performance metrics of the top and bottom efficiency-based portfolios constructed from the unselected DEA structural models with multiple outputs. Compared to the extended-multiple-output DEA models, the return spreads between the top and bottom efficiency-based portfolios are more evident for the multiple-output DEA models. For all three structural models, the top efficiency-based portfolios of the extended-multiple-output DEA models have statistically lower volatility than their bottom efficiency-based portfolios. On the other hand, the top efficiency-based portfolios of the multiple-output DEA models have statistically higher returns than their bottom efficiency-based portfolios. From these observations, we can infer that an inclusion of lagged exchange rates appears to help lower the variance of the portfolio returns while it does not seem to help increase the levels of the portfolio returns.

Among the three structural models, the flexible price monetary model and the sticky price asset model seem to have the strongest out-of-sample performance for both DEA and

Table 4.5: Out-of-Sample Performance of the Unselected DEA Models with Multiple Outputs

Table 4.5 presents performance metrics of the top and bottom efficiency-based portfolios for the unselected DEA structural models with multiple outputs. Panel A and B present performance metrics computed from total and price returns of the efficiency-based portfolios respectively.

Panel A: Total Return

	Multiple-Output DEA Model						Ext.-Multiple-Output DEA Model		
	Flexible Price Monetary			Sticky Price Monetary			Sticky Price Asset		
Return (%)	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Top	7.60	6.64	8.12	7.11	5.79	6.40	6.17	5.82	6.39
Bottom	3.34	1.54	1.98	2.21	1.47	1.25	4.36	2.16	3.79
Volatility (%)	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Top	9.77	9.99	10.00	9.20	9.69	9.66	8.73	8.88	8.80
Bottom	8.82	9.02	9.34	10.48	9.42	10.77	12.59	12.81	13.29
Kurtosis	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Top	3.26	3.02	2.82	2.78	2.52	2.68	3.45	3.16	3.25
Bottom	3.06	3.03	2.88	4.63	2.76	3.29	3.84	3.65	3.47
Skewness	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Top	−0.64	−0.40	−0.43	−0.49	−0.36	−0.46	−0.76	−0.62	−0.69
Bottom	−0.57	−0.50	−0.49	−0.97	−0.30	−0.53	−0.79	−0.58	−0.55
Sharpe Ratio	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Top	0.61	0.52	0.69	0.60	0.45	0.54	0.52	0.49	0.59
Bottom	0.20	0.01	0.09	0.06	0.00	0.01	0.22	0.06	0.20

Panel B: Price Return

	Multiple-Output DEA Model						Ext.-Multiple-Output DEA Model		
	Flexible Price Monetary			Sticky Price Monetary			Sticky Price Asset		
Return (%)	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Top	4.15	3.27	4.95	3.74	2.50	3.28	2.69	2.40	3.06
Bottom	−0.38	−2.04	−1.45	−1.63	−2.26	−2.30	0.47	−1.60	0.02
Volatility (%)	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Top	9.86	10.02	9.93	9.28	9.74	9.68	8.71	8.86	8.80
Bottom	8.83	9.08	9.41	10.5	9.44	10.86	12.63	12.86	13.30
Kurtosis	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Top	3.36	3.12	2.78	2.82	2.55	2.69	3.40	3.14	3.24
Bottom	3.00	3.00	2.85	4.64	2.74	3.24	3.83	3.65	3.47
Skewness	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Top	−0.68	−0.44	−0.42	−0.52	−0.38	−0.47	−0.78	−0.63	−0.70
Bottom	−0.49	−0.45	−0.46	−0.95	−0.28	−0.49	−0.75	−0.55	−0.55
Sharpe Ratio	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Top	0.26	0.18	0.38	0.23	0.11	0.22	0.12	0.11	0.21
Bottom	−0.22	−0.38	−0.28	−0.31	−0.39	−0.32	−0.09	−0.24	−0.09

regression models. The addition of an inflation variable in the sticky price monetary model and the sticky price asset model does not seem to improve the results by much. This may be because the REER is already an inflation adjusted figure. Regardless, all three DEA structural models perform relatively well both in- and out-of-sample, and their performance compares favorably to that of the U. S. dollar indices and the residual-based portfolios.[14]

Based on the empirical results, we can conclude that the structural models constructed by means of DEA yield satisfactory results in terms of identifying currencies that are worthy of investment.

4.5 Conclusions

In this study, we developed a currency selection strategy that utilizes macroeconomic efficiency of countries. With an aim of validating the link between exchange rates and macroeconomic variables, we built our models from the three representative bilateral structural exchange rate models: the flexible price monetary model, the sticky price monetary model, and the sticky price asset model. We constructed their multilateral DEA counterparts using trade-weighted macroeconomic variables and quantified macroeconomic efficiency of countries into a single score. This score served as a guidepost for making investment decisions in our strategy.

The investment portfolios formed on the basis of macroeconomic efficiency performed better than the market indices and the residual-based benchmarks in our testing. The risk-return profile of the efficiency-based portfolios was superior, especially to that of the U.S. dollar indices. The efficiency scores obtained from the DEA methodology also had stronger discriminatory power than the regression residuals. The relatively weak discriminatory power of regression residuals could be partly due to the homogeneous model specification and/or the small sample size. In this regard, two of the distinguishing advantages of the DEA methodology highlighted in this study are: first, it allows for the elasticities of macroe-

[14]For detailed performance measures of the market and residual-based benchmarks, refer to Appendix B.3.

conomic variables to vary across different countries and second, it is known to work fine with small sample sizes. The first feature seems to be beneficial, particularly for comparing developed economies (economies of G10) with emerging economies (economies of emerging markets). Therefore, the DEA methodology may be more apt for conducting cross-sectional evaluation of currencies than any other statistical methods.

For future research, one can reconstruct other existing exchange rate models within a multilateral framework presented in this study and analyze their performance. Furthermore, as DEA is known to work better with homogenous economic entities, one can investigate the effectiveness of the proposed DEA models when they are applied to a smaller set of currencies such as G10 currencies.

The goal of the DEA structural exchange rate models developed in this study has been to assess the economic performance of countries as opposed to projecting their future exchange rates. By testing these models, we found that countries with higher macroeconomic efficiency scores tend to have stronger and thus, appreciating currencies compared to those with lower scores. Furthermore, the contrasting performance of the top and bottom efficiency-based portfolios observed in our out-of-sample test confirms that our efficiency scores are valid performance metrics of countries' macroeconomic conditions, and hence, of currencies. From our results, it seems reasonable to conclude that macroeconomic variables have substantial influence on exchange rate dynamics as presumed by conventional exchange rate models.

Chapter 5

Joint Variable Selection for DEA via Group Sparsity

5.1 Introduction

Despite the large number of papers published on DEA,[1] surprisingly little attention has been paid to variable selection in the literature. Variable selection approaches in DEA are often based on experts' opinions, past experience or economic theories, as a matter of fact. One major concern about these approaches is that they are prone to include (omit) irrelevant (relevant) variables, thus leading to a model mis-specification.

Several studies demonstrated significant negative impact that a model mis-specification has on the accuracy of DEA efficiency estimates. For instance, Sexton et al. (1986) investigated the effect of including an irrelevant variable in a DEA model and reported that any variable included in the analysis, in fact, can change the shape and position of the production frontier, which in turn alters the ranking of efficiency estimates. Similarly, Smith (1990) documented the danger of a model mis-specification when a relevant variable is omitted from a DEA model.

[1] According to the literature survey by Liu et al. (2013) the DEA field has accumulated over 4,500 papers in the ISI Web of Science database in the last three decades.

Variable selection plays a pivotal role in DEA also because the greater the number of variables included in a DEA model, the higher the dimensionality of the production space and the less discerning the analysis (Jenkinson and Anderson, 2003). An increase in the number of variables included in a DEA model, for instance, tends to shift the compared DMUs towards the efficient frontier. This results in a decline in DEA's discriminatory power (Fried *et al.*, 2008; Golany and Roll, 1989). It is therefore essential to limit the number of variables included in the analysis. Still, there is no consensus on how best to do this.

In this chapter, we propose a data-driven joint variable selection method for DEA. In particular, we extend the group LASSO designed for variable selection on (often predefined) groups of variables in linear regression models to DEA models. We derive a special constrained version of the group LASSO with the loss function suited for variable selection in DEA models and solve it by a new tailored algorithm based on the alternating direction method of multipliers (ADMM). We conduct a thorough performance evaluation of the proposed method against two of the most widely-used variable selection approaches in the DEA literature: the efficiency contribution measure (ECM) method and the regression-based (RB) test, by means of Monte Carlo simulations.

The remainder of the chapter is organized as follows: in Section 5.2, we detail the development of the joint variable selection method for DEA; in Section 5.3, we introduce the ECM method and the RB test; in Section 5.4, we describe our simulation study; in Section 5.5, we present the numerical results of the simulation study; in Section 5.6, we provide a real-world example of the application of the proposed variable selection method; and in Section 5.7, we present our conclusions.

5.2 Joint Variable Selection

In DEA, we are often in a situation where we want to select a small number of most relevant input variables across the DMUs. One popular variable selection approach through convex optimization is the LASSO (Tibshirani, 1996). It is a simple regularization technique, which adds an l_1-norm (sum of the absolute values) of the variables to the original objective function. Due to the special geometric properties of the l_1-ball, the solution to the LASSO problem is sparse; i.e. only a small number of entries are nonzero, and these correspond to the selected variables. Although the LASSO was originally designed for variable selection in linear regression models, it can be easily extended to DEA models. For instance, let us recall the basic additive model presented in previous chapters.

$$\begin{aligned}
\min_{u,v,w} \quad & \sum_{i=1}^{n}\left(\sum_{k=1}^{l} x_{k,i}v_{k,i} - \sum_{r=1}^{s} y_{r,i}u_{r,i} + w_i\right) \\
\text{subject to} \quad & \sum_{k=1}^{l} x_{k,j}v_{k,i} - \sum_{r=1}^{s} y_{r,j}u_{r,i} + w_i \geq 0,\ i,j = 1,\ldots,n, \\
& u_{r,i} \geq 1,\ r = 1,\ldots,s; i = 1,\ldots,n, \\
& v_{k,i} \geq 1,\ k = 1,\ldots,l; i = 1,\ldots,n.
\end{aligned} \tag{5.1}$$

Then, its respective LASSO formulation is given by

$$\begin{aligned}
\min_{u,v,w} \quad & \sum_{i=1}^{n}\left(\sum_{k=1}^{l} x_{k,i}v_{k,i} - \sum_{r=1}^{s} y_{r,i}u_{r,i} + w_i + \lambda\sum_{k=1}^{l} v_{k,i}\right) \\
\text{subject to} \quad & \sum_{k=1}^{l} x_{k,j}v_{k,i} - \sum_{r=1}^{s} y_{r,j}u_{r,i} + w_i \geq 0,\ i,j = 1,\ldots,n, \\
& u_{r,i} \geq 1,\ r = 1,\ldots,s; i = 1,\ldots,n, \\
& v_{k,i} \geq 1,\ k = 1,\ldots,l; i = 1,\ldots,n
\end{aligned}$$

for some $\lambda > 0$, the regularization parameter that controls the level of sparsity in the solution. We do not need absolute values of v in the objective function since they are constrained to be positive. Note that in this case, the solution is, in fact, sparse only after a "shift," by subtracting 1 from each entry. We should also note that one can readily incorporate output variables into variable selection by adding $\lambda\sum_{r=1}^{s} u_{r,i}$ to the objective function. For an example of the LASSO application on DEA models, readers can refer to Chapter 3.

The results obtained from the LASSO formulation may be hard to interpret because the selection of the variables is not guaranteed to be consistent across all the DMUs. For instance, for a given variable k, $v_{k,i}$ may not be selected for all i's, i.e. across all DMUs. In fact, it has been shown that the LASSO tends to select only one variable from a group of highly correlated variables and does not care which one is selected (Zou and Hastie, 2005). Hence, an approach that enforces selection consistency across the DMUs is called for. Note that if we stack the column vectors v_i's as a matrix V, then our goal is to select a small number of rows from V. In the next section, we develop a joint variable selection method for this additive model (5.1). It should, however, be noted that the proposed variable selection method can be readily adapted for various DEA models.

5.2.1 Group Sparsity-inducing Regularization

Before we discuss our approach to joint variable selection for DEA, we need to introduce a more general regularization technique, group LASSO (Yuan and Lin, 2006), which is an extension to LASSO and tends to induce variable sparsity at group level, i.e. to select a small number of groups of correlated variables. It achieves this goal via $l_{2,1}$-regularization,

$$\min_{\beta} \quad F(\beta) + \lambda \sum_{j=1}^{J} \|\beta_j\|_2,$$

which is the sum of the group l_2-norms with a pre-defined grouping of the variables $\{\beta_j\}_{j=1}^{J}$. The original group LASSO problem considered in Yuan and Lin (2006) has $F(\beta) := \frac{1}{2}\|X\beta - y\|_2^2$, i.e. least-squares regression. The same regularization technique has also been applied to logistic regression (Meier *et al.*, 2008).

The requirement of a pre-defined grouping of the variables is often a limiting factor for applications of group LASSO. However, in the case of DEA joint variable selection, the grouping structure is readily available – we simply group the variables in $V = [v_{k,i}]$ by rows; i.e. by each variable $k = 1, \ldots, l$ across n DMUs. We can then solve a special constrained version of the group LASSO with the loss function,

$$F(u, v, w) := \sum_{i=1}^{n} \left(\sum_{k=1}^{l} x_{k,i} v_{k,i} - \sum_{r=1}^{s} y_{r,i} u_{r,i} + w_i \right).$$

The nonzero entries in the group-sparse solution that we obtain (after shifting) are then guaranteed to be consistent across all the DMUs.

The standard group LASSO problem has been studied extensively in the machine learning and optimization literature, and a number of convex optimization algorithms, e.g. Liu *et al.* (2009); Meier *et al.* (2008); Qin *et al.* (2013); van den Berg *et al.* (2008); Wright *et al.* (2009); Yuan and Lin (2006), have been proposed to solve it. However, most of these algorithms are designed to solve the unconstrained group LASSO problem. In the subsequent sections, we propose a new tailored algorithm based on the alternating direction method of multipliers (ADMM) to solve our special constrained group LASSO problem.

5.2.2 Problem Formulation

Using a change of variables $\tilde{u} = u - e$ and $\tilde{v} = v - e$, we can transform the original additive model (5.1) into the following model with zero lower bounds on $\tilde{u}$ and $\tilde{v}$:

$$\begin{aligned}
\min_{\tilde{u},\tilde{v},w} \quad & \sum_{i=1}^{n} \left(\sum_{k=1}^{l} x_{k,i} \tilde{v}_{k,i} - \sum_{r=1}^{s} y_{r,i} \tilde{u}_{r,i} + w_i \right) \\
\text{subject to} \quad & \sum_{k=1}^{l} x_{k,j} (\tilde{v}_{k,i} + 1) - \sum_{r=1}^{s} y_{r,j} (\tilde{u}_{r,i} + 1) + w_i \geq 0, \; i, j = 1, \ldots, n, \\
& \tilde{u}_{r,i} \geq 0, \; r = 1, \ldots, s; i = 1, \ldots, n, \\
& \tilde{v}_{k,i} \geq 0, \; k = 1, \ldots, l; i = 1, \ldots, n.
\end{aligned}$$

For joint variable selection on v, we propose to apply the group LASSO regularization on $\tilde{v}$; i.e.

$$
\begin{aligned}
\min_{\tilde{u},\tilde{v},w} \quad & \sum_{i=1}^{n}\left(\sum_{k=1}^{l} x_{k,i}\tilde{v}_{k,i} - \sum_{r=1}^{s} y_{r,i}\tilde{u}_{r,i} + w_i\right) + \lambda \sum_{k=1}^{l} \|\tilde{v}^k\|_2 \\
\text{subject to} \quad & \sum_{k=1}^{l} x_{k,j}(\tilde{v}_{k,i}+1) - \sum_{r=1}^{s} y_{r,j}(\tilde{u}_{r,i}+1) + w_i \geq 0, \; i,j = 1,\ldots,n, \\
& \tilde{u}_{r,i} \geq 0, \; r = 1,\ldots,s; i = 1,\ldots,n, \\
& \tilde{v}_{k,i} \geq 0, \; k = 1,\ldots,l; i = 1,\ldots,n
\end{aligned}
$$

where $\tilde{v}^k$ is the vector of $\tilde{v}_{k,i}$ for $i = 1,\ldots,n$. This formulation can be interpreted as selecting the elements of v which are sufficiently large, i.e. larger than the lower bounds. If joint variable selection on u is also needed, we can apply the group LASSO regularization on u in a similar way. For notational simplicity, we drop the $\sim$ signs for $\tilde{u}, \tilde{v}, \tilde{X}$, and $\tilde{Y}$ from now on.

Writing in matrix form, the group LASSO formulation for the additive model of DEA is

$$
\begin{aligned}
\min_{U,V,W} \quad & tr(X^\top V - Y^\top U + W) + \lambda \sum_{k=1}^{l} \|v^k\|_2 \\
\text{subject to} \quad & X^\top V - Y^\top U + W + B \geq 0, \\
& U \geq 0, \\
& V \geq 0
\end{aligned} \tag{5.2}
$$

where v^k is k-th row of V, $W = \begin{pmatrix} ew_1 & \cdots & ew_n \end{pmatrix}$ and tr denotes the trace of a matrix.

Introducing non-negative slack variables, we can transform (5.2) into

$$\begin{aligned} \min_{U,V,\bar{V},W,S} \quad & tr(X^\top V - Y^\top U + W) + \lambda R(\bar{V}) \\ \text{subject to} \quad & X^\top V - Y^\top U + W + B = S_x,\ S_x \geq 0, \\ & U = S_u,\ S_u \geq 0, \\ & V = \bar{V} = S_v,\ S_v \geq 0 \end{aligned} \tag{5.3}$$

where $R(\bar{V}) = \sum_{k=1}^{l} \|v^k\|_2$. For the ease of visualization, we write the problem in terms of vectorized decision variables, i.e. stacking columns on top of each other, which is the same as the (:) operator in `Matlab`. We use the lower-case letter to denote the vectorized version of the matrix counterpart, which is denoted by upper-case. The following are some elementary transformations:

$$X^\top V \ \rightarrow \begin{pmatrix} X^\top & & \\ & \ddots & \\ & & X^\top \end{pmatrix} v = \bar{X} v\ ,\ tr(X^\top V) \ \rightarrow \ x^\top v\ ,\ W \ \rightarrow \begin{pmatrix} e & & \\ & \ddots & \\ & & e \end{pmatrix} w = \bar{I} w.$$

Problem (5.3) then becomes

$$\begin{aligned} \min_{s \geq 0, v, u, w, \bar{v}} \quad & x^\top v - y^\top u + e^\top w + \lambda R(\bar{v}) \\ \text{subject to} \quad & \bar{X} v - \bar{Y} u + \bar{I} w + b = s_x, \\ & v = s_v, \\ & u = s_u, \\ & v = \bar{v}. \end{aligned}$$

By writing the above problem in a more compact form, we obtain

$$\begin{aligned} \min_{s\geq 0, z, \bar{v}} \quad & c^\top z + \lambda R(\bar{v}) \\ \text{subject to} \quad & \begin{pmatrix} A_s \\ A_v \end{pmatrix} z + \begin{pmatrix} b \\ 0 \end{pmatrix} = \begin{pmatrix} s \\ \bar{v} \end{pmatrix} \end{aligned} \tag{5.4}$$

where $z = \begin{pmatrix} v \\ u \\ w \end{pmatrix}$, $c = \begin{pmatrix} x \\ -y \\ e \end{pmatrix}$, $A_s = \begin{pmatrix} \bar{X} & -\bar{Y} & \bar{I} \\ I & 0 & 0 \\ 0 & I & 0 \end{pmatrix}$, $A_v = \begin{pmatrix} I & 0 & 0 \end{pmatrix}$, $s = \begin{pmatrix} s_x \\ s_v \\ s_u \end{pmatrix}$.

We note that the BCC model (for all n DMUs) with the group LASSO regularization is

$$\begin{aligned} \min_{u,v,w} \quad & \sum_{i=1}^{n} \left(x_i^\top v_i + w_i \right) + \lambda \sum_{k=1}^{l} \|v^k\|_2 \\ \text{subject to} \quad & y_i^\top u_i = 1,\ i = 1, \dots, n, \\ & X^\top v_i - Y^\top u_i + w_i e^\top + b \geq 0,\ i = 1, \dots, n, \\ & u_i \geq 0,\ i = 1, \dots, n, \\ & v_i \geq 0,\ i = 1, \dots, n \end{aligned} \tag{5.5}$$

where v_i is the vector of $v_{k,i}, k = 1, \dots, l$, and similar definitions apply to u_i, x_i, and y_i. X and Y are matrices whose columns are x_i's and y_i's respectively. The corresponding problem for the CCR model is the same except that there is no variable w. A similar transformation involving non-negative slacks applied to the additive model can also be applied to the above model (5.5). Accordingly, the optimization algorithm described in the next section can be used to solve this problem as well.

5.2.3 Optimization Algorithm

The alternating direction method of multipliers (ADMM) was first proposed in the 1970s (Gabay and Mercier, 1976; Glowinski and Marroco, 1975). It belongs to the family of the classical augmented Lagrangian method (Hestenes, 1969; Powell, 1972; Rockafellar, 1973), which iteratively solves the linearly constrained problem

$$\begin{aligned} \min_{x} \quad & F(x) \\ \text{subject to} \quad & Ax = b. \end{aligned} \tag{5.6}$$

The augmented Lagrangian of problem (5.6) is $\mathcal{L}(x,\gamma) = F(x) + \gamma^T(b - Ax) + \frac{1}{2\mu}\|Ax - b\|^2$, where γ is the Lagrange multiplier and μ is the penalty parameter for the quadratic infeasibility term. The augmented Lagrangian method minimizes $\mathcal{L}(x,\gamma)$ followed by an update to γ in each iteration.

For a structured unconstrained problem

$$\min_{x} F(x) \equiv f(x) + g(Ax) \tag{5.7}$$

where both functions $f(\cdot)$ and $g(\cdot)$ are convex, we can decouple the two functions by introducing an auxiliary variable y and transform problem (5.7) into an equivalent linearly constrained problem

$$\begin{aligned} \min_{x,y} \quad & f(x) + g(y) \\ \text{subject to} \quad & Ax = y. \end{aligned} \tag{5.8}$$

The augmented Lagrangian of this problem is

$$\mathcal{L}(x,y,\gamma) = f(x) + g(y) + \gamma^T(y - Ax) + \frac{1}{2\mu}\|Ax - y\|^2.$$

ADMM (Algorithm 1) finds the approximate minimizer of $\mathcal{L}(x,y,\gamma)$ by alternatively optimizing with respect to x and y once. This is often desirable because joint minimization of $\mathcal{L}(x, y, \gamma)$ even approximately could be hard.

Algorithm 1 ADMM

1: Choose $\gamma^{(0)}$.
2: **for** $k = 0, 1, \ldots, K$ **do**
3: $\quad x^{k+1} \leftarrow \arg\min_x \mathcal{L}(x, y^{(k)}, \gamma^{(k)})$
4: $\quad y^{k+1} \leftarrow \arg\min_y \mathcal{L}(x^{k+1}, y, \gamma^{(k)})$
5: $\quad \gamma^{(k+1)} \leftarrow \gamma^{(k)} - \frac{1}{\mu}(Ax^{k+1} - y^{k+1})$
6: **end for**
7: **return** $y^{(K)}$

Our strategy is to apply Algorithm 1 to solve problem (5.4). First, we write down the augmented Lagrangian of the problem,

$$\mathcal{L}(z, s, \bar{v}, \gamma_s, \gamma_v) := c^\top z + \lambda R(\bar{v}) - \gamma_s^\top (A_s z + b - s) + \frac{1}{2\mu}\|A_s z + b - s\|^2 - \gamma_v^\top (A_v z - \bar{v}) + \frac{1}{2\mu}\|A_v z - \bar{v}\|^2.$$

Next, we minimize with respect to $z, s, \bar{v}$ sequentially. The subproblem with respect to z can be simplified to solving a sparse linear system with a fixed left-hand-side,

$$\frac{1}{\mu}(A_s^\top A_s + A_v^\top A_v)z = A_s^\top \gamma_s + A_v^\top \gamma_v - c + \frac{1}{\mu}(A_s^\top s + A_v^\top \bar{v}).$$

As long as we keep μ constant, we can compute the Cholesky factor of the left-hand-side for once and cache it for subsequent iterations, where the computation for this step is almost as cheap as a gradient step (via forward/backward substitution). The subproblem with respect to s is a projection problem onto the non-negative orthant,

$$\min_{s \geq 0} \frac{1}{2}\|A_s z + b - s - \mu\gamma_s\|^2.$$

We can obtain the solution easily by $(A_s z + b - \mu\gamma_s)^+$, where $(\cdot)^+$ is an element-wise truncation operation at 0. The subproblem with respect to $\bar{v}$ is the proximal problem associated with the group LASSO penalty,

$$\min_{\bar{v}} \frac{1}{2}\|A_v z - \bar{v} - \mu\gamma_v\|^2 + \mu\lambda R(\bar{v}).$$

The optimal solution can be computed in closed-form: $\bar{v}^* = \mathcal{T}_{\mu\lambda}(A_v z - \mu\gamma_v)$, where $\mathcal{T}$ is the block soft-thresholding operator such that the k-th block of $\bar{v}^*$, $[\bar{v}^*]_k = \frac{A_v z - \mu\gamma_v}{\|A_v z - \mu\gamma_v\|_2} \max(0, \|A_v z - \mu\gamma_v\|_2 - \mu\lambda)$, for $k = 1, \ldots, l$.

5.2.4 Convergence

The convergence of ADMM has been established for the case of two-way splitting as above. We restate the results from Eckstein and Bertsekas (1992) in the following theorem.

Theorem 5.2.1. *Consider problem* (5.8)*, where both f and g are proper, closed, convex functions, and $A \in \mathbb{R}^{n\times l}$ has full column rank. Then, starting with an arbitrary $\mu > 0$ and $x^0, y^0 \in \mathbb{R}^l$, the sequence $\{x^k, y^k, \gamma^k\}$ generated by Algorithm 1 converges to a Kuhn-Tucker pair $\big((x^*, y^*), \gamma^*\big)$ of problem* (5.8)*, if* (5.8) *has one. If* (5.8) *does not have an optimal solution, then at least one of the sequences $\{(x^k, y^k)\}$ and $\{\gamma^k\}$ diverges.*

It is known that μ does not have to be decreased to a very small value (or can simply stay constant) in order for the method to converge to the optimal solution of problem (5.8) (Bertsekas, 1999; Nocedal and Wright, 1999).

We observe that problem (5.4) can be treated as a two-way splitting, with variables z and $\begin{pmatrix} s \\ \bar{v} \end{pmatrix}$, and obviously the matrix $\begin{pmatrix} A_s \\ A_v \end{pmatrix}$ has full column rank. Hence, Theorem 5.2.1 applies to our ADMM algorithm.

In the next section, we introduce benchmark variable selection methods, against which we evaluate the performance of the proposed joint variable selection method. Hereinafter, we refer to our group LASSO-based variable selection method as the GL method.

5.3 Benchmarks

The four most widely-used approaches in the DEA literature[2] are: (i) the regression-based (RB) test (Ruggiero, 2005), (ii) the efficiency contribution measure (ECM) method (Pastor *et al.*, 2002), (iii) the principal component analysis (PCA-DEA) (Adler and Golany, 2002; Ueda and Hoshiai, 1997), and (iv) the bootstrapping method (Simar and Wilson., 2001). Nataraja and Johnson (2011) evaluated these four approaches and reported their performance. According to their results, the RB test and the ECM method are best suited for various sample sizes provided that there is low correlation among variables while the PCA-DEA and bootstrapping methods show some limitations. The two major limitations of the PCA-DEA method are: first, as it replaces the original variables with principal components (PCs), the original data set is not retained, and therefore it is impossible to recover true efficiency levels, and second, it is vulnerable to the curse of dimensionality. The main issue with the bootstrapping method is that it involves the heavy computational burden, and yet, has the weakest performance among the four. Consequently, we select the ECM method and the RB test to serve as our performance evaluation benchmarks.

5.3.1 The Efficiency Contribution Measure (ECM) Method

The efficiency contribution measure (ECM) method evaluates the effect of a candidate variable x_{cand} on the efficiency computation by comparing two DEA formulations: one with the candidate variable and one without it. The ECM of x_{cand} for a particular DMU_0, denoted by γ_0, is a single scalar measure that quantifies the marginal impact of x_{cand} on the measurement of efficiency. In essence, the ECM method performs a statistical test to determine the statistical significance of x_{cand}'s contribution when measured by means of ECMs. It should be noted that the ECM method consists of two procedures for the progressive selection of variables: a forward selection (addition of variables) and backward elimination (removal of variables), and only supports radial DEA models, such as the CCR

[2]Readers can refer to Nataraja and Johnson (2011) for detailed reviews of prevailing variable selection methods in the DEA literature.

and BCC models.

To provide further technical details, suppose $\gamma = (\gamma_1, ..., \gamma_n)$ are the observed ECMs of a random sample, $\Gamma = (\Gamma_1, ..., \Gamma_n)$ drawn from a population (Γ, F) where Γ being a random variable distributed according to F, a cumulative density function on $[1, \infty)$. The underlying idea of the ECM method is that if x_{cand} is an irrelevant variable, then the impact it has on the efficiency evaluation should be negligible, and high values of Γ associated with x_{cand} are unlikely to be observed. For the statistical test, two additional parameters, $\bar{\gamma}$ and p_0, are introduced. $\bar{\gamma}$ represents the tolerance level for the degree of efficiency score change caused by x_{cand}, and p_0 represents the tolerance level for the proportion of DMUs whose associated efficiency score change exceeds $\bar{\gamma}$. x_{cand} is considered relevant to the production process if more than p_0% of DMUs have associated efficiency score change greater than $\bar{\gamma}$. More formally, a hypothesis test with a binomial test statistic is performed to see if the marginal impact of this candidate variable on the efficiency estimation is significant. For technical details and applications of the ECM method, readers can refer to Pastor *et al.* (2002) and Chen and Johnson (2010) respectively.

5.3.2 The Regression-based (RB) Test

In the regression-based (RB) test, initial efficiency estimates obtained from the set of known production variables are regressed against the set of candidate variables. The formal representation of the regression model is given by

$$E = \alpha + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_l x_l + \epsilon$$

where E is the efficiency score obtained from the DEA model including only an output variable y and without loss of generality, an input variable x_1, and x_2 through x_l are the candidate variables. If the coefficient β_k in the regression is statistically significant at a given level of significance and has the proper sign, i.e. $\beta > 0$ for input variables and $\beta < 0$ for output variables, the candidate variable x_k is considered relevant to the production process and is added to the DEA model. A new efficiency score E is then computed with this updated DEA model, and the new candidate set is tested. This process is repeated until all candidate variables are either found irrelevant or included in the model, and there are no more remaining variables to be tested. For technical details of the RB test, readers can refer to Ruggiero (2005).

5.4 Experimental Design and Data Generation

In our simulation study, we focus on output-oriented radial DEA models, namely the CCR (Charnes *et al.*, 1978) and BCC (Banker *et al.*, 1984) models. The reason behind this particular choice of models is that one of the benchmark methods, the ECM method, is not compatible with non-radial models. A presentation based on input-oriented formulations or non-radial models can be similarly developed.

In practical applications of DEA, the true form of a production process is mostly unknown, and the observed data used for estimating this unknown production function are often limited and contain measurement errors. These are the major setbacks in evaluating the practical importance of theoretical results in DEA, as a matter of fact. In order to overcome this problem, this study uses Monte Carlo simulations to generate a large number of observations for a plausible production process, the form of which is known.

The production process we consider is the linearly homogeneous Cobb-Douglas function, in which a number of inputs, represented by vector x, are used to produce a single output y; i.e.

$$y_{1,i} = \beta \prod_{k=1}^{l} x_{k,i}^{\alpha_k}, \; i = 1, \ldots, n \tag{5.9}$$

where α_k and β are assumed to be known parameters. The parameter α_k here plays an important role in the production model. First, it defines the returns-to-scale (RTS) specification for the production process; i.e. $\sum_{k=1}^{l} \alpha_k = 1$ indicates a CRS production process while $\sum_{k=1}^{l} \alpha_k < 1$ indicates a VRS production process. Second, mathematically, α_k indicates the flexibility of production with respect to the input x_k under efficient production;

i.e. it represents the importance of the input x_k in the production process other things being equal. For our simulation study, we set $\alpha_k = 1/m$ where m is a predetermined value as the base case scenario and $\beta = 1$ throughout.

By incorporating a technical inefficiency of the i-th DMU denoted by $\epsilon_i \in [0, 1]$ into (5.9), we obtain

$$y_{1,i} = \beta \prod_{k=1}^{l} x_{k,i}^{\alpha_k} \epsilon_i, \; i = 1, \ldots, n. \tag{5.10}$$

An alternative additive representation of (5.10) is

$$Y_{1,i} = B + \sum_{k=1}^{l} \alpha_k X_{k,i} - u_i, \; i = 1, \ldots, n$$

where $Y_{1,i} = \ln y_{1,i}, B = \ln \beta, X_{k,i} = \ln x_{k,i}$, and $u_i = \ln \epsilon_i$. The efficiency component $u_i \geq 0$ represents the shortfall of output from the production frontier.

With respect to statistical distributions of variables u and X, consistent with previous studies (Nataraja and Johnson, 2011; Smith, 1990), u is drawn from a half-normal distribution with mean zero and variance σ^2, which we vary to obtain a sample average efficiency score of 85% and is assumed to be uncorrelated with any X in order to generate a realistic range of inefficiency values. The values of X are generated from a uniform distribution on an interval $[10, 20]$ and are exponentiated and used in (5.10) to yield the values of y and x to be used in DEA models.

Smith's (1990) simulation study involving a Cobb-Douglas production function has shown that the performance of DEA in estimating true efficiencies diminishes as the number of inputs in the production process increases. We therefore assume that the true production process is determined by three inputs, x_1, x_2, and x_3 only. In the base case scenario, we also independently generate an irrelevant random variable x_4 from a uniform distribution on the same interval $[10, 20]$ to maintain symmetry with the three other relevant inputs. We test the basic variable set consisting of y, x_1, x_2, x_3, and x_4 using the three variable selection methods: the GL method, the ECM method, and the RB test, to determine the model specification. We should note that in the remaining of the chapter, "candidate variables"

Table 5.1: Outline of the Experimental Scenarios

Table 5.1 delineates the experiments used for evaluating the performance of the three variable selection methods: the GL method, the ECM method and the RB test. A total of 12 experimental scenarios are considered for (i) a CRS production process and (ii) a VRS production process. The respective values of input contribution parameter α for CRS and VRS production processes are shown on the third column separated by a semicolon.

Experiment	Correlation Between Inputs	Input Contribution to Output	Description
1	Independently generated	$\alpha_k = 1/3$; $\alpha_k = 1/4, k = 1, 2, 3$	Base case
2	$\rho_{1,2} = 0.8, \rho_{1,3} = 0.2$	$\alpha_k = 1/3$; $\alpha_k = 1/4, k = 1, 2, 3$	Correlated inputs
3	$\rho_{1,2} = \rho_{1,3} = 0.8$	$\alpha_k = 1/3$; $\alpha_k = 1/4, k = 1, 2, 3$	Highly correlated inputs
4	Independently generated	$\alpha_1 = 1/3, \alpha_2 = 4/9, \alpha_3 = 2/9$; $\alpha_1 = 1/4, \alpha_2 = 1/3, \alpha_3 = 1/6$	Input contribution to output varied
5	$\rho_{1,2} = 0.8, \rho_{1,3} = 0.2$	$\alpha_1 = 1/3, \alpha_2 = 4/9, \alpha_3 = 2/9$; $\alpha_1 = 1/4, \alpha_2 = 1/3, \alpha_3 = 1/6$	Correlated inputs and input contribution to output varied
6	$\rho_{1,2} = 0.8, \rho_{1,3} = 0.2$	$\alpha_1 = 1/3, \alpha_2 = 2/9, \alpha_3 = 4/9$; $\alpha_1 = 1/4, \alpha_2 = 1/6, \alpha_3 = 1/3$	Correlated inputs and input contribution to output varied
7	$\rho_{1,4} = 0.8$	$\alpha_k = 1/3$; $\alpha_k = 1/4, k = 1, 2, 3$	Correlated input and a random variable
8	Independently generated	$\alpha_k = 1/3$; $\alpha_k = 1/4, k = 1, 2, 3$	Small sample size, $n = 25$
9	Independently generated	$\alpha_k = 1/3$; $\alpha_k = 1/4, k = 1, 2, 3$	Large sample size, $n = 300$
10	Independently generated	$\alpha_k = 1/4$; $\alpha_k = 1/5, k = 1, 2, 3, 4$	Base case with one more relevant input x_4
11	Independently generated	$\alpha_k = 1/2$; $\alpha_k = 1/3, k = 1, 2$	Base case without a relevant input x_3
12	Independently generated	$\alpha_k = 1/3$; $\alpha_k = 1/4, k = 1, 2, 3$	Base case with three irrelevant inputs x_4, x_5 and x_6

refer to x_1, x_2, x_3, and x_4, "model efficiency estimates" refer to efficiency estimates obtained from a DEA model with a set of input variables identified by each variable selection method, and "true efficiency estimates" refer to efficiency estimates obtained from a DEA model with true input variables (x_1, x_2, and x_3).

In addition to the base case, diversified experimental scenarios are considered. These include varying: the covariance structure of inputs, sample size, contribution of each input to output and the dimensionality of the production process. For those experiments concerned with correlated input variables, we adopt the following equation from Wang and Schmidt (2002) to establish the desired covariance structure of inputs.

$$x_k = \rho_{k,j} x_j + w\sqrt{1 - \rho_{k,j}^2}, \quad k = 2, 3, 4, \quad j = 1, 2, 3, \quad k \neq j.$$

Here, $\rho_{k,j}$ is the correlation between x_k and x_j, and w is a random variable generated from a uniform distribution on the interval $[10, 20]$. Table 5.1 delineates the experimental scenarios considered in the simulation study. We should note that these scenarios are in line with the experiments used in Nataraja and Johnson's study (2011). Each experiment is tried 100 times, and the simulation results averaged over 100 trials are presented in the next section. It should be noted that we tuned the parameters of the algorithms used in the GL method on the training data (10% of the full data), and the reported results are out-of-sample results.

Table 5.2: Parameter Specification for the Simulation Study

Algorithm	Parameter value
ECM	$p_0 = 0.15$, $\bar{\rho} = 1.10$, $\alpha = 0.05$
RB	$\alpha = 0.90$

Table 5.2 presents the parameter specification for the benchmark methods, the ECM method and the RB test. We keep these parameter values the same throughout the Monte Carlo simulations. For the ECM method, following the recommendations of Pastor et al. (2002), we set $p_0 = 15\%$, $\bar{\gamma} = 10\%$ and the significance level α to 5% for the hypothesis test. Also, for comparative purposes, we use the backward procedure, which begins with the full model and then eliminates one variable that has the least impact on the efficiency calculation at each successive step. For the RB test, we set the significance level α to 90% following Ruggiero's (2005) suggestion, and without loss of generality, we choose x_1 as the first variable to be included in the initial efficiency estimation assuming no prior knowledge of production input variables.

5.5 Numerical Results

Performance criteria used for evaluating the three variable selection methods: the GL method, the ECM method and the RB test, can be broadly divided into three sets. The first and second measurement criteria we consider are the mean squared error (MSE) and the correlations between the true and model efficiency estimates. For correlation metrics, we use Pearson's and Spearman's rank correlation coefficients. The last set consists of two measures: (i) the percentage of all DMUs correctly identified as efficient or inefficient and (ii) the percentage of efficient DMUs correctly identified as efficient. All the methods under evaluation are implemented in `Matlab`, and all our experiments were performed on an Intel Core i5-680 (3.60GHz), 64-bit operating system.

The results for the first and second performance criteria for CRS and VRS production processes are presented in Table 5.3 and 5.4 respectively. Table 5.5 provides the results for the last set of performance criteria. We will discuss these results in relation to variations in the covariance structure of inputs, sample size, importance of inputs in the production process, and the dimensionality of the production space.

Table 5.3: Performance of the Variable Selection Methods for a CRS Production Process

Table 5.3 presents MSE, Pearson's correlation coefficient and Spearman's rank correlation coefficient between the true and model efficiency estimates. True efficiency estimates are obtained using a CCR model with true input variables. Model efficiency estimates are obtained using a CCR model with the three sets of input variables selected by the GL method, the ECM method and the RB test respectively.

Metrics	MSE			Correlation Coefficient			Rank Correlation Coefficient		
Experiments	GL	ECM	RB	GL	ECM	RB	GL	ECM	RB
1	0.0000	0.0000	0.0118	1.0000	1.0000	0.9302	0.9998	1.0000	0.9242
2	0.0001	0.0013	0.0022	0.9985	0.9763	0.9789	0.9976	0.9662	0.9744
3	0.0001	0.0026	0.0069	0.9981	0.9599	0.9398	0.9976	0.9486	0.9288
4	0.0005	0.0001	0.0054	0.9908	0.9980	0.9631	0.9884	0.9975	0.9583
5	0.0007	0.0023	0.0027	0.9900	0.9629	0.9712	0.9857	0.9496	0.9641
6	0.0003	0.0008	0.0019	0.9936	0.9852	0.9874	0.9908	0.9775	0.9856
7	0.0003	0.0000	0.0573	0.9944	0.9988	0.6613	0.9918	0.9984	0.6356
8	0.0014	0.0010	0.0176	0.9660	0.9673	0.8432	0.9428	0.9445	0.8092
9	0.0000	0.0000	0.0100	1.0000	1.0000	0.9494	0.9999	1.0000	0.9466
10	0.0001	0.0005	0.0096	0.9988	0.9883	0.9369	0.9976	0.9833	0.9278
11	0.0000	0.0000	0.0057	1.0000	1.0000	0.9625	1.0000	1.0000	0.9583
12	0.0000	0.0000	0.0003	1.0000	1.0000	0.9921	0.9999	1.0000	0.9880

Table 5.4: Performance of the Variable Selection Methods for a VRS Production Process

Table 5.4 presents MSE, Pearson's correlation coefficient and Spearman's rank correlation coefficient between the true and model efficiency estimates. True efficiency estimates are obtained using a BCC model with true input variables. Model efficiency estimates are obtained using a BCC model with the three sets of input variables selected by the GL method, the ECM method and the RB test respectively.

Metrics	MSE			Correlation Coefficient			Rank Correlation Coefficient		
Experiments	GL	ECM	RB	GL	ECM	RB	GL	ECM	RB
1	0.0002	0.0010	0.0022	0.9947	0.9768	0.9662	0.9885	0.9681	0.9565
2	0.0002	0.0024	0.0020	0.9943	0.9459	0.9529	0.9895	0.9242	0.9338
3	0.0006	0.0032	0.0028	0.9830	0.9345	0.9437	0.9757	0.9118	0.9223
4	0.0012	0.0030	0.0013	0.9687	0.9284	0.9687	0.9540	0.8989	0.9533
5	0.0009	0.0042	0.0037	0.9764	0.9153	0.9197	0.9656	0.8836	0.8879
6	0.0005	0.0013	0.0012	0.9862	0.9672	0.9660	0.9757	0.9471	0.9472
7	0.0002	0.0019	0.0056	0.9945	0.9557	0.9193	0.9833	0.9355	0.8986
8	0.0038	0.0046	0.0091	0.8845	0.8444	0.7842	0.7848	0.7636	0.6881
9	0.0001	0.0159	0.0009	0.9980	0.8562	0.9892	0.9964	0.8273	0.9860
10	0.0004	0.0085	0.0039	0.9871	0.8161	0.9153	0.9695	0.7492	0.8735
11	0.0000	0.0001	0.0008	1.0000	0.9974	0.9863	0.9994	0.9954	0.9825
12	0.0002	0.0014	0.0020	0.9941	0.9672	0.9527	0.9879	0.9564	0.9238

Table 5.5: The Identification of Efficient and Inefficient DMUs

Panel A and B of Table 5.5 present: (i) the percentage of DMUs correctly identified as efficient or inefficient and (ii) the percentage of the efficient DMUs correctly identified as efficient by the three methods: the GL method, the ECM method, and the RB test, for both CRS and VRS production frontiers respectively.

Panel A. % of DMUs Correctly Identified as Efficient/Inefficient

RTS:	CRS			VRS		
Experiments	GL	ECM	RB	GL	ECM	RB
1	100%	100%	98%	99%	97%	97%
2	100%	95%	99%	99%	92%	94%
3	100%	95%	94%	98%	91%	91%
4	99%	100%	98%	96%	91%	96%
5	99%	95%	98%	97%	88%	89%
6	98%	95%	99%	97%	94%	95%
7	99%	100%	89%	99%	94%	93%
8	96%	96%	88%	90%	85%	78%
9	100%	100%	99%	100%	92%	99%
10	100%	98%	96%	99%	81%	90%
11	100%	100%	99%	100%	100%	99%
12	100%	100%	98%	99%	97%	95%

Panel B. % of Efficient DMUs Correctly Identified as Efficient

RTS:	CRS			VRS		
Experiments	GL	ECM	RB	GL	ECM	RB
1	100%	100%	82%	98%	89%	89%
2	97%	46%	88%	96%	57%	67%
3	96%	30%	28%	84%	39%	41%
4	92%	98%	85%	87%	65%	86%
5	86%	41%	74%	84%	36%	47%
6	81%	50%	95%	86%	67%	75%
7	98%	98%	9%	99%	80%	72%
8	89%	88%	57%	82%	75%	58%
9	100%	100%	87%	100%	46%	95%
10	99%	91%	79%	99%	47%	74%
11	100%	100%	89%	100%	100%	95%
12	100%	100%	100%	100%	87%	89%

5.5.1 The Impact of Variations in the Covariance Structure of Inputs

In practice, input variables are often highly correlated with each other as they are all related to the scale and types of operations of DMUs being evaluated. It is, hence, important to test the robustness of each variable selection method with respect to variations in the covariance structure of inputs. Figure 5.1 illustrates the impact of varying correlation among input variables on the performance of the three variable selection methods for a VRS production process. As $\rho_{1,2}$ and $\rho_{1,3}$ are varied from 0.45 to 0.90, the GL method exhibits consistently strong performance with low MSE and high correlation coefficient between the true and model efficiency estimates. In contrast, both the ECM method and the RB test show fluctuations in their performance.

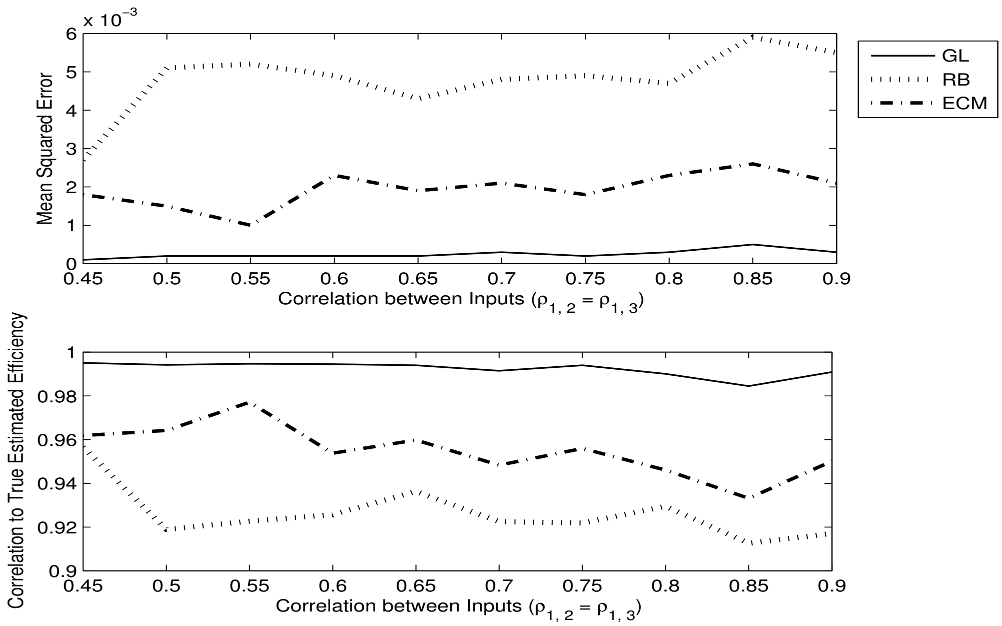


Figure 5.1: The Impact of Variations in the Covariance Structure of Inputs (for a VRS Production Process)

Similarly, for a CRS production process with highly correlated input variables, i.e. Experiment 3 where $\rho_{1,2} = \rho_{1,3} = 0.80$, the GL method outperforms both of its benchmarks. For instance, the GL method identifies 96% of efficient DMUs correctly whereas the ECM method and the RB test identify less than 30% of them correctly. The respective MSEs of the ECM method and the RB test are also 69 and 25 times higher than that of the GL method. When the relevant input variable x_1 is highly correlated with the irrelevant input variable x_4 in Experiment 7, the GL and ECM methods show comparably strong performance for a CRS production process. The RB test, on the other hand, tends to choose x_4 as a relevant variable and has contrastingly weak performance. For a VRS production process, the GL method outperforms both the ECM method and the RB test. Overall, the GL method is most robust to variations in the covariance structure of inputs. Consistent with previous findings, all three methods generally perform better when there is low correlation among input variables.

5.5.2 The Impact of Variations in Sample Size (n)

As different applications involve different sample sizes, we use Experiment 8 and 9 to investigate the impact of small and large sample sizes on the performance of the three variable selection methods. As can be seen from Table 5.3, 5.4 and 5.5, the GL method outperforms both the ECM method and the RB test regardless of the sample size. The outperformance of the GL method is more evident for a VRS production process. For instance, while the ECM method and the RB test identify 75% and 58% of efficient DMUs as efficient respectively, the GL method correctly identifies 85% of them. These observations can be summarized in Figure 5.2. From this figure, it is clear that the RB test gets most affected by the small sample size. It is surprising to see the considerably weak performance of the ECM method when the sample size is increased for a VRS production process. In general, consistent with Nataraja and Johnson's (2011) results, the performance of the three methods improves as the sample size increases.

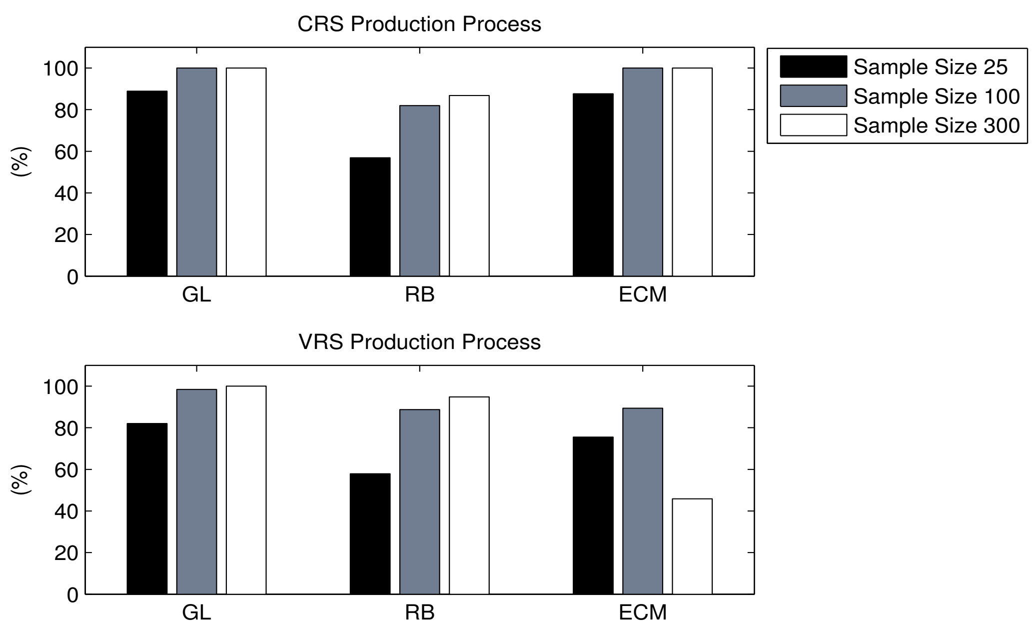


Figure 5.2: The Impact of Variations in Sample Size on the Correct Identification of Efficient DMUs

An increase in sample size, however, negatively influences the running time of each variable selection method. In Experiment 9 with a large sample size of 300, while the execution time only takes 17.09 (25.08) seconds for the GL method for a CRS (VRS) production process, it takes 904.45 (314.51) and 114.30 (97.47) seconds for the ECM method and the RB test respectively (see Table 5.6). Across all 12 experiments, the GL method has the shortest execution time while the ECM method, which uses the backward selection algorithm, has the longest execution time.

Table 5.6: The Execution Time (seconds)

Table 5.6 presents the amount of CPU time each method took on average for one trial of each of the 12 experiments.

	CRS Production Process			VRS Production Process		
Experiments	GL	ECM	RB	GL	ECM	RB
1	1.62	27.02	4.14	2.12	27.51	4.14
2	1.54	26.97	4.38	2.17	33.23	5.11
3	1.39	29.57	2.78	1.85	33.73	3.12
4	1.94	24.20	4.69	1.85	30.12	5.35
5	1.65	28.30	4.17	2.18	32.88	3.40
6	1.51	28.41	3.65	1.86	32.66	4.50
7	1.66	24.77	1.49	2.19	27.40	3.69
8	0.36	2.42	0.34	0.35	2.68	0.34
9	17.09	904.45	114.30	25.08	314.51	97.47
10	1.95	33.80	4.53	2.96	44.91	4.47
11	1.77	16.83	3.27	1.98	18.19	4.14
12	2.58	60.98	4.41	2.84	62.32	4.50

5.5.3 The Impact of Variations in the Importance of Inputs in the Production Process

Since it is reasonable to assume considerable variations in the relative importance of inputs in the production process, it is essential to test the robustness of the results with respect to the variations in input contribution to output. When input contribution to output is varied, i.e. Experiments 4, 5, and 6, all three variable selection methods, in general, have better performance under a CRS production process. Figure 5.3 plots the percentage of correctly identified efficient DMUs for Experiments 1, 4, 5, and 6 for each of the three variable selection methods. As can be seen from this figure, the GL method exhibits the most stable and strongest performance across the four experiments. Although the RB test outperforms the GL method in Experiment 6 under a CRS production process, it underperforms under a VRS production process.

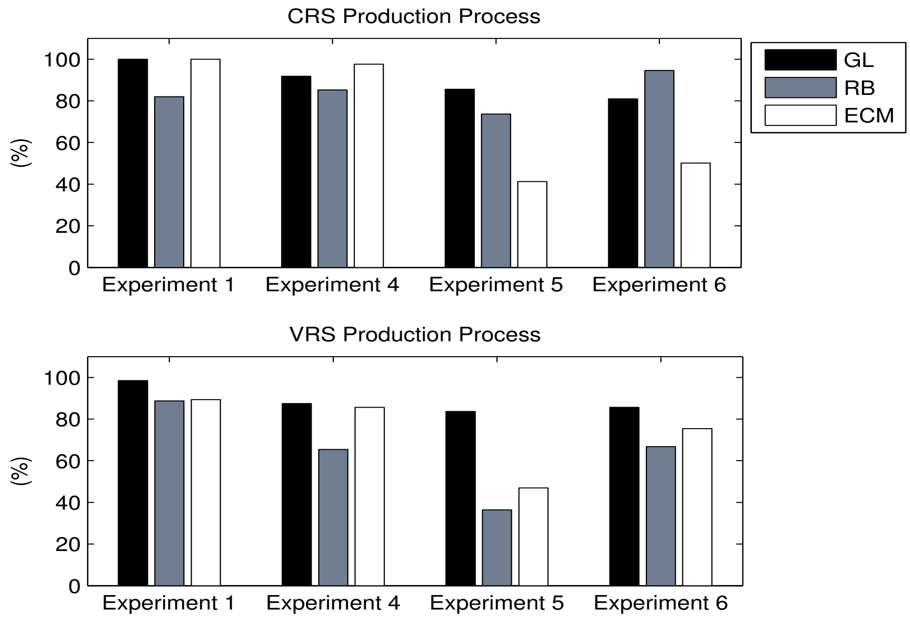


Figure 5.3: The Impact of Variations in Input Contribution on the Correct Identification of Efficient DMUs

The performance of both the ECM method and the RB test gets negatively impacted by the variations in input contribution to output when inputs are correlated, especially under a VRS production process. For example, in terms identifying efficient and inefficient DMUs, when input contribution is varied for correlated inputs, i.e. Experiment 5, the ECM method and the RB test identify 36% and 47% of efficient DMUs correctly under a VRS production process. These values are 5% and 27% lower than the corresponding values obtained for a CRS production process. Similar observations can be made in terms of MSEs between the true and model efficiency estimates. Consistent with the results obtained so far, the GL method shows comparable or better performance compared to the other two benchmark methods.

5.5.4 The Impact of Variations in the Dimensionality of the Production Process

Experiments 10 and 11 consider variations in the dimensionality of the production process. When the dimensionality of the production function is increased in Experiment 10, the GL method outperforms both the ECM method and the RB test. For a CRS production process, the RB test has the weakest performance in terms of MSE and correlations between the true and model efficiency estimates. Under a VRS production process, both the ECM method and the RB test show considerably weaker performance. For instance, the MSEs and the correlation between the true and model efficiency estimates are at least 10 times higher and 8% lower than those obtained for the GL method respectively. As the dimensionality of the production function decreases in Experiment 11, the performance of all three methods improves.

From the results obtained from various experiments in our simulation study, we can conclude that the GL method significantly outperforms both the ECM method and the RB test and is most robust to variations in the covariance structure of inputs, sample size, importance of inputs, and the dimensionality of the production space. Also, the GL method is the fastest algorithm and is least vulnerable to the choice of underlying production technology among the three variable selection methods presented.

5.6 The Application of the Joint Variable Selection Method for DEA

As a real-world example, we apply the proposed variable selection method to determine a DEA model specification for estimating firm efficiency. For comparative purposes, consistent with Chapter 3, we use the U.S. IT sector as our empirical setting.[3] For computing operational efficiency of the U.S. IT firms, we use return on equity (ROE) as an output variable. As for the input variables, we employ the GL method (GL-DEA) to select a repre-

[3] We used the same test data used in Chapter 3.

sentative measure from each of the five categories of financial ratios: profitability, leverage, liquidity, efficiency, and market value. The LASSO-based variable selection method for DEA (LASSO-DEA) and the original LASSO variable selection method for linear regression models (original LASSO) used in Chapter 3 are considered as reference points.[4]

Table 5.7: Input Variable Selection Results

Category	GL-DEA	LASSO-DEA	Original LASSO
Profitability	Return on Equity	Return on Equity	Return on Equity
Leverage	Total Debt to Total Common Equity	Common Equity to Total Asset	Total Debt to Total Common Equity
Liquidity	Current Ratio	Cash Ratio	Cash Ratio
Efficiency	Accounts Receivable Turnover	Accounts Receivable Turnover	Asset Turnover
Market Value	Price to Book Ratio	Price to Book Ratio	Price to Book Ratio

Table 5.8: Performance of the Top-Minus-Bottom Spread

Table 5.8 presents the descriptive statistics and Sharpe ratios of the top-minus-bottom spread for three lagged investments. It also presents the t-stats obtained from the standard difference-in-means test, the right tail t-test in particular.

	GL-DEA			LASSO-DEA			Original LASSO		
	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Total Return (%)	17.42	13.70	5.00	10.51	7.07	9.88	4.98	0.21	−3.70
Volatility (%)	15.02	12.93	13.85	18.88	17.65	14.65	12.80	12.26	14.86
Kurtosis	2.88	3.70	3.37	4.49	3.85	3.56	4.06	3.69	3.83
Skewness	0.45	0.61	−0.07	0.55	0.06	0.47	−0.02	−0.13	0.40
Sharpe Ratio	1.02	0.90	0.22	0.44	0.28	0.54	0.22	−0.15	−0.38
t-test	3.23***	3.29***	1.68**	1.71**	1.20	2.37***	1.63**	0.47	0.01

Table 5.7 lists the three sets of financial ratios selected using: (i) the GL-DEA, (ii) the LASSO-DEA, and (iii) the original LASSO, respectively. Based on these three DEA model specifications, we compute three sets of efficiency scores for each firm in the sample, and form three sets of top and bottom efficiency decile portfolios following the efficiency-based investment strategy proposed in Chapter 3. For each set, we consider the return spread between the top and bottom efficiency decile portfolios (top-minus-bottom return spreads) for the performance evaluation. Table 5.8 presents the descriptive statistics and Sharpe ratios of these top-minus-bottom return spreads over the 10-year investment horizon (2001

[4]We should note that the use of the group LASSO variable selection method for linear regression models is not necessary in this particular application. For the details on the LASSO-based variable selection method for DEA, see Appendix A.3.

– 2011). In addition, the t-stats obtained from the standard difference-in-means test are also reported in the table.

It is apparent from Table 5.8 that the financial ratios selected by the GL-DEA and the LASSO-DEA are more relevant to measuring firm's operational efficiency than those selected by the original LASSO method. Such results highlight the importance as well as the advantage of using a variable selection method tailored for DEA when building a DEA model. With regard to the GL-DEA and the LASSO-DEA, both show somewhat comparable performance although the t-test results suggest that the efficiency scores estimated from the DEA model specification determined by the GL-DEA have a stronger discriminatory power for differentiating efficient firms from inefficient ones. Moreover, the advantage of the GL-DEA over the LASSO-DEA is that its results are easier to interpret as the selection is done jointly across all DMUs. Through this real-life example, we were able to see the effectiveness of the proposed method in constructing a DEA model in practice.

5.7 Conclusions

As Golany and Roll (1989) noted in their study, surprisingly a few number of studies give an overview of DEA as an application procedure that must focus on the choice of variables. Even though no functional form is specified in DEA models, DEA results heavily rely on the selection of input and output variables. Wrong choices of variables are likely to compromise the accuracy of the analysis. In this regard, a model specification must be a central concern in DEA.

In this study, we developed a data-driven joint variable selection method based on the group LASSO and reported its significant outperformance over the prevailing variable selection methods, namely the efficiency contribution measure (ECM) method and the regression-based (RB) test. We evaluated the performance of our proposed method and its benchmarks by means of Monte Carlo simulations and examined the sensitivity of the results to the variations in the covariance structure of inputs, sample size, importance of inputs, and the dimensionality of the production space. Based on the results obtained from a diversified set of simulation experiments, we can conclude that the GL method is more robust and effective than its benchmarks.

The proposed GL method is a more sophisticated and quantitative variable selection method that will help finding a parsimonious DEA model, which uses as many variables as needed, but as few as possible.

Part II

A Generalized Risk Budgeting Approach to Portfolio Construction

Chapter 6

A Generalized Risk Budgeting Approach to Portfolio Construction

6.1 Introduction

Risk-based asset allocation models have received considerable attention in recent years. Some of this attention has been motivated by the difficulty in estimating expected returns. Mean-variance optimization, for example, is very sensitive to expected asset returns and if applied naively, generally results in portfolios with extreme portfolio weights that are unstable over time. While there are now many methods for addressing these problems, e.g. Black and Litterman (1992), there has been a trend of late to focus on approaches that are more robust to any assumptions on expected returns. The "$1/N$" approach of DeMiguel *et al.* (2009) is notable in this regard as are the recent developments in risk-based asset allocation models which are the focus of this chapter.

As the term "risk-based" suggests, risk generally plays a more important role in risk-based portfolio construction models. Examples of these models include the classic minimum variance approach of Markowitz and the more contemporary risk parity and risk budgeting approaches. In this study, we propose a generalized risk budgeting (GRB) approach to portfolio construction.

The concept of risk parity goes back to 1996 when Bridgewater Associates launched a risk parity fund called the All Weather fund. Although the risk parity product was originally introduced to market by Bridgewater, the term "risk parity" was first coined by Qian (2006) who formalized the definition of risk parity in terms of a risk budget where weights of assets are determined in such a way that they all contribute equally to the overall portfolio risk. Maillard *et al.* (2010) referred to such a portfolio as an equal risk contribution (ERC) portfolio. They analyzed properties of an unconstrained long-only ERC portfolio and showed that its volatility lies between the volatilities of the long-only minimum variance and equally-weighted portfolios. We note here that the terms "risk parity" and "equal risk contribution" are used interchangeably in the literature, but hereafter we will use the former.

A risk parity portfolio, however, is not always desirable. An investor may prefer to allocate different risk budgets to each asset, and this preference would require a more general risk budgeting portfolio. Theoretical properties of risk budgeting portfolios were analyzed by Bruder and Roncalli (2012). Extending the result of Maillard *et al.* (2010), they showed that the volatility of a long-only risk budgeting portfolio lies between the volatilities of a long-only minimum variance portfolio and a long-only weighted portfolio whose weights are proportional to their risk budgets. They further demonstrated that when the portfolio risk is computed using a convex risk measure and risk budgets are defined to be strictly positive, then a long-only risk budgeting portfolio exists and is unique.

Since the introduction of this approach, there have been many additional studies on risk parity and risk budgeting approaches. Most of them, however, have focussed on seeking a long-only minimal risk portfolio that satisfies (pre-defined) risk budgeting constraints. The majority of these methods, therefore, lack flexibility. For example, by disregarding the expected asset returns in their problem formulations, many of these methods make the implicit assumption that all asset returns are identical in expectation. Whether or not the disregarding of expected returns results in a better-performing portfolio, such an assumption does not hold in practice. In addition, it is clearly desirable that investors be able to freely express their views on expected returns when constructing a portfolio.

In this chapter we propose a generalized risk budgeting (*GRB*) problem formulation that leads in general to a non-convex optimization problem. We refer to this problem as the *GRB* portfolio optimization problem. We then develop solution approaches for this *GRB* problem.[1] The key advantage of our formulation over the prevailing risk parity or risk budgeting approaches is that it offers a much greater degree of flexibility in the way risk-based portfolios are constructed. It allows for short sales of assets, the use of risk factors to model asset returns, and most importantly, it allows investors to define risk budgets for overlapping subsets of assets.

When the subsets form a partition, the assets all have the same expected return and we restrict ourselves to long-only portfolios, we show that the problem can be formulated as a convex optimization problem and is therefore easily solved. This result generalizes Bruder and Roncalli (2012)'s approach for constructing a long-only risk budgeting portfolio with minimum variance. For the more general *GRB* problem, we propose two solution approaches. The first approach is a semidefinite programming (SDP) relaxation to the problem that allows us to obtain an (upper) bound on the optimal objective function of the *GRB* problem. Moreover, the solution to this SDP often yields a very good starting point for a generic non-linear optimization solver. To our knowledge, we are not aware of any other studies that apply an SDP relaxation to the risk parity or risk budgeting problems. Our second approach develops a numerical algorithm that combines augmented Lagrangian and Markov chain Monte-Carlo (MCMC) methods with the goal of finding a point in the vicinity of a very good local optimum. This point is then supplied to a non-linear optimization routine to compute this local optimum. The merit of this second approach is in its generic nature: in particular, it provides a starting-point strategy for any non-linear optimization algorithm.

---

[1] Henceforth, we refer to the *GRB* portfolio optimization problem as the *GRB* problem.

The remainder of the chapter is organized as follows. In Section 6.2 we formally define the *GRB* problem and describe our two solution approaches as well as the special case that can be solved as a convex optimization problem. We provide numerical results for the SDP relaxation and the augmented Lagrangian-MCMC approach in Section 6.3. We then conclude in Section 6.4.

6.2 The Generalized Risk Budgeting (*GRB*) Problem

In portfolio construction and analysis it is often preferable to group assets according to attributes such as asset class, country, sector and industry. In the case of an investment portfolio with a broad coverage of asset classes, for example, it may be more insightful and therefore preferable to look at the marginal risk contribution of each asset class rather than each individual asset in the portfolio. The generalized risk budgeting (*GRB*) strategy is based on this very idea of managing the marginal risk contributions of subsets of assets to the total portfolio risk. In a *GRB* portfolio, the risk contribution from each (pre-specified) subset of assets is set equal to some pre-specified risk budget. Note that we are using the term "subset" rather than"partition" since depending on the attributes used for the asset classification, assets may belong to more than one group. We will see later that minimum variance, risk parity and risk budgeting portfolios are all special instances of a *GRB* portfolio.

The objective of the *GRB* problem is to find a *GRB* portfolio that is optimal on the basis of its risk-return profile. Portfolio risk in the *GRB* problem is computed via a positively homogeneous risk measure for which we can use Euler's theorem to provide a risk decomposition. Examples of positively homogeneous risk measures include portfolio volatility, value-at-risk (VaR) and any coherent risk measures such as conditional value-at-risk (CVaR) (Artzner *et al.*, 1999).

Towards this end, let $\mathcal{R}(x) : \mathbb{R}^d \to \mathbb{R}$ denote a generic risk measure that is a positively homogeneous function of degree one in the portfolio weight vector, x. Euler's theorem then provides the following additive risk decomposition:

$$\mathcal{R}(x) = \sum_{i=1}^{d} x_i \frac{\partial \mathcal{R}(x)}{\partial x_i} \tag{6.1}$$

where the marginal risk contribution of the i-th asset is

$$RC_i(x) = x_i \frac{\partial \mathcal{R}(x)}{\partial x_i}.$$

If $\mathcal{M}_1, \ldots, \mathcal{M}_s \subseteq \{1, \ldots, d\}$ denote s subsets of portfolio assets, then the marginal risk contribution of the k-th subset is

$$RC_{\mathcal{M}_k}(x) := \sum_{i \in \mathcal{M}_k} RC_i(x).$$

Let $\beta_1, \ldots, \beta_s$ now denote the risk budgets for $\mathcal{M}_1, \ldots, \mathcal{M}_s$, respectively. We can then formulate the GRB problem as:

$$\begin{aligned} (GRB) \qquad\qquad & \max_{x \in \mathcal{X}} \quad \mu' x - \lambda \mathcal{R}(x), \\ & \text{subject to} \quad \textstyle\sum_{i \in \mathcal{M}_k} RC_i(x) = \beta_k \mathcal{R}(x), \quad k = 1, \ldots, s \end{aligned}$$

where $\mu \in \mathbb{R}^d$ is a vector of expected returns, λ is a risk aversion parameter and $\mathcal{X} := \{x \in \mathbb{R}^d : \mathbf{1}' x = 1\}$. Note that the constraint $\sum_{i \in \mathcal{M}_k} RC_i(x) = \beta_k \mathcal{R}(x)$ implies that $\sum_{k=1}^{s} \beta_k = 1$ when the $\mathcal{M}_k$'s form a partition. We note that the GRB problem becomes a minimum variance problem when $\mu = \mu_0 \mathbf{1}$, there is only one subset $\mathcal{M}_1$ which is equal to the universe of assets, and the risk measure is portfolio volatility. It is a risk parity problem when $\mu = \mu_0 \mathbf{1}$, the $\mathcal{M}_k$'s are all singletons and all risk budgets, β_k, are equal. Finally it is a risk budgeting problem when $\mu = \mu_0 \mathbf{1}$ and the $\mathcal{M}_k$'s are again all singletons.

The GRB problem is a constrained non-convex optimization problem for which efficient solution algorithms are unavailable. Although there are numerous methods available for computing risk parity portfolio weights, e.g. Spinu (2013) and Bai *et al.* (2013), these methods are in general not applicable to the *GRB* problem. In Section 6.2.1 below we consider a special case of the *GRB* problem which can be solved as a convex optimization problem. We then proceed to discuss our solution approaches for the general non-convex case. Note that the parameter and variable notations introduced in this section will be used throughout the chapter unless otherwise stated.

6.2.1 A Special Case of the *GRB* Problem

We now consider a special case of the *GRB* problem in which all assets have the same expected return, i.e. $\mu = \mu_0 \mathbf{1}$, each asset belongs to one and only one subset, i.e. the $\mathcal{M}_k$'s form a partition of the asset space, and non-negativity constraints are imposed at the partition level. We can then reformulate the *GRB* problem as follows:

$$\begin{aligned} \min_{x \in \mathcal{X}} \quad & \mathcal{R}(x), \\ \text{subject to} \quad & \textstyle\sum_{i \in \mathcal{M}_k} RC_i(x) = \beta_k \mathcal{R}(x), \quad k = 1, \ldots, s, \\ & \textstyle\sum_{i \in \mathcal{M}_k} x_i \geq 0, \quad k = 1, \ldots, s. \end{aligned} \tag{6.2}$$

Assuming also that each $\beta_k > 0$, we then have the following result that extends Bruder and Roncalli (2012).

Theorem 6.2.1. *Assuming $\mathcal{R}(y) \neq 0$ for nonzero y, then problem (6.2) is equivalent to the convex optimization problem (6.3):*

$$\begin{aligned} \mathit{min}_{y} \quad & \mathcal{R}(y) \\ \textit{subject to} \quad & \textstyle\sum_{k=1}^{s} \beta_k \ln\left(\sum_{i \in \mathcal{M}_k} y_i\right) \geq c \end{aligned} \tag{6.3}$$

where c is an arbitrary constant. In particular the normalized optimal solution $\tilde{y}^*$ to (6.3) is also the optimal solution to (6.2). (A normalized solution is one where $\sum_{i=1}^{d} \tilde{y}_i^* = 1$. See the discussion after the proof.)

Proof. Let $\mathcal{L}(y, \gamma)$ denote the Lagrangian of the optimization problem (6.3) so that

$$\mathcal{L}(y, \gamma) = \mathcal{R}(y) - \gamma \left(\sum_{k=1}^{s} \beta_k \ln \left(\sum_{i \in \mathcal{M}_k} y_i \right) - c \right).$$

At optimality, the solution y^* satisfies the KKT conditions. That is, y^* satisfies (i) the first-order conditions

$$\frac{\partial \mathcal{L}(y, \gamma)}{\partial y_i} = \frac{\partial \mathcal{R}(y)}{\partial y_i} - \gamma \left(\frac{\beta_k}{\sum_{j \in \mathcal{M}_k} y_j} \right) = 0, \tag{6.4}$$

for $i = 1, \ldots, d$ and where k is the index of the subset $\mathcal{M}_k$ containing i, and (ii) the complementary slackness conditions

$$\gamma \left(\sum_{k=1}^{s} \beta_k \ln \left(\sum_{i \in \mathcal{M}_k} y_i \right) - c \right) = 0. \tag{6.5}$$

Note that as $\ln : \mathbb{R}^+ \to \mathbb{R}$, we must have $\sum_{i \in \mathcal{M}_k} y_i > 0$ for $k = 1, \ldots, s$, and hence y cannot be $\mathbf{0}$. Then since $\mathcal{R}(y) \neq 0$ for nonzero y, at least one of $\dfrac{\partial \mathcal{R}(y)}{\partial y_i}$ must be nonzero by (6.1). The strict positivity of the β_k's and (6.4) then imply $\gamma > 0$. We therefore have

$$\frac{\partial \mathcal{R}(y)}{\partial y_i} = \gamma \left(\frac{\beta_k}{\sum_{j \in \mathcal{M}_k} y_j} \right) \tag{6.6}$$

for $i = 1, \ldots, d$ and where $\gamma > 0$. Multiplying both sides of (6.6) by y_i and then summing over $i \in \mathcal{M}_k$ yields

$$\begin{aligned} \sum_{i \in \mathcal{M}_k} y_i \frac{\partial \mathcal{R}(y)}{\partial y_i} &= \gamma \left(\frac{\beta_k}{\sum_{j \in \mathcal{M}_k} y_j} \right) \sum_{i \in \mathcal{M}_k} y_i \\ &= \gamma \beta_k \end{aligned}$$

for $k = 1, \ldots, s$. We therefore see that the risk contribution of each $\mathcal{M}_k$ is proportional to its risk budget, β_k. The normalized optimal solution $\tilde{y}^*$ is then the optimal solution x^* to (6.2) as claimed. □

Note that as $\sum_{k=1}^{s} \beta_k \ln \left(\sum_{i \in \mathcal{M}_k} y_i \right) = c$ by (6.5), we could directly obtain x^* from (6.3) if we used $c^* = c - \ln \left(\sum_{i=1}^{d} y_i \right)$ in (6.3), rather than the original c which led to the solution y. Note also that we recover the results of Bruder and Roncalli (2012) if the $\mathcal{M}_k$ are all singletons.

6.2.2 An SDP Relaxation for the General *GRB* Problem

Our first approach for the *GRB* problem uses a semidefinite programming (SDP) relaxation to obtain an upper bound on the optimal objective function value. There are two advantages of the SDP approach: (i) the solution to the SDP problem (which is generally infeasible for the *GRB* problem) can be used as a (hopefully very good) starting point for a standard non-linear optimization routine, and (ii) the SDP solution can often provide a "certificate" of near-optimality when the SDP solution has an objective function that is close to the objective function of the best local optimal solution that we have found.

In our development of the SDP approach we will assume initially that our risk measure is portfolio volatility so that

$$\mathcal{R}(x) \quad := \quad \sqrt{x'\Sigma x}$$

where $\Sigma \in \mathbb{R}^{d \times d}$ is the covariance matrix of asset returns. The marginal risk contributions of the individual assets then satisfy

$$RC_i(x) = x_i \frac{(\Sigma x)_i}{\sqrt{x'\Sigma x}}, \; i = 1, \ldots, d.$$

With this measure of risk we can rewrite the *GRB* problem in the following equivalent form:

$$\begin{aligned} \max_{x,X} \quad & \mu' x - \lambda \mathcal{R}(x), \\ \text{subject to} \quad & \textstyle\sum_{i \in \mathcal{M}_k} \text{tr}(\Gamma_i X) = \beta_k \, \text{tr}(\Sigma X), \; k = 1, \ldots, s, \\ & X = xx', \\ & \mathbf{1}' x = 1 \end{aligned} \tag{6.7}$$

where $\Gamma_i = e_i e_i' \Sigma$, e_i denotes the i-th column of the identity matrix $I \in \mathbb{R}^{d \times d}$, $x \in \mathbb{R}^d$, and $\mathrm{tr}(\cdot)$ denotes the trace of a matrix. Since $X = xx'$ is the only non-convex constraint in (6.7), we obtain a convex relaxation of the *GRB* problem by relaxing this constraint to $X \succeq xx'$. We then obtain the following SDP relaxation of our *GRB* problem:

$$\max_{x,X} \mu' x - \lambda \mathcal{R}(x), \tag{6.8}$$

$$\text{subject to} \sum_{i \in \mathcal{M}_k} \mathrm{tr}(\Gamma_i X) = \beta_k \, \mathrm{tr}(\Sigma X), \quad k = 1, \ldots, s$$

$$\begin{bmatrix} X & x \\ x' & 1 \end{bmatrix} \succeq 0,$$

$$\mathbf{1}' x = 1$$

where the linear matrix inequality (LMI) is equivalent to $X \succeq xx'$. Note that we can recover (6.7) from (6.8) by imposing an additional (non-convex) constraint that the left-hand-side of the LMI in (6.8) to be a rank one matrix. The SDP relaxation can be solved efficiently and the SDP solution provides an upper bound on the optimal objective function of the *GRB* problem. For example, one can easily implement and solve (6.8) using[2] `CVX` (Boyd and Grant, 2008, 2014).

In addition to the portfolio volatility risk measure, we can also derive SDP relaxations of the *GRB* problem for other risk measures under certain assumptions. For example, consider the value-at-risk (VaR) and conditional value-at-risk (CVaR) of a portfolio. If $F_r(z) := P\{r \leq z\}$ is the CDF of the portfolio return r, then the VaR and CVaR at the confidence level $\alpha \in (0, 1)$ are defined as

$$VaR_\alpha := \min\{z | F_r(z) \geq \alpha\}$$

[2]A special SDP mode in `CVX` allows positive (negative) semidefinite constraints $\succeq$ ($\preceq$) to be imposed using `Matlab`'s standard inequality operators >= (<=).

and

$$CVaR_\alpha := E[r|r \geq VaR_\alpha(r)].$$

Suppose the asset returns are normally distributed with mean vector μ and as before, covariance matrix Σ. Then for a portfolio with the weight vector x it easily follows that

$$VaR_\alpha(x) = \mu' x + \Psi_1(\alpha)\sqrt{x'\Sigma x}$$

and

$$CVaR_\alpha(x) = \mu' x + \Psi_2(\alpha)\sqrt{x'\Sigma x}$$

where (see McNeil *et al.* (2005))

$$\begin{aligned}
\Psi_1(\alpha) &= \sqrt{2}\,\mathrm{erf}^{-1}(2\alpha - 1), \\
\Psi_2(\alpha) &= (\sqrt{2\pi}\exp(\mathrm{erf}^{-1}(2\alpha - 1))^2(1-\alpha)^{-1}, \\
\mathrm{erf}(z) &= (2/\sqrt{\pi})\textstyle\int_0^z e^{-t^2}dt.
\end{aligned}$$

Likewise, the respective marginal VaR and CVaR contributions of the i-th asset are given by

$$RC_{i,VaR_\alpha}(x) = \mu_i + x_i \frac{(\Sigma x)_i}{\sqrt{x'\Sigma x}} \Psi_1(\alpha)$$

and

$$RC_{i,CVaR_\alpha}(x) = \mu_i + x_i \frac{(\Sigma x)_i}{\sqrt{x'\Sigma x}} \Psi_2(\alpha).$$

Without loss of generality, let us consider the risk budgeting constraints that arise when we use VaR as our risk measure. These constraints take the form

$$\sum_{i \in \mathcal{M}_k} \left(\mu_i + x_i \frac{(\Sigma x)_i}{\sqrt{x'\Sigma x}} \Psi_1(\alpha) \right) = \beta_k \left(\mu' x + \Psi_1(\alpha)\sqrt{x'\Sigma x} \right),\ k = 1, \ldots, s. \tag{6.9}$$

The VaR version of the *GRB* problem is given by:

$$
\begin{aligned}
&\max_{x,X} \;\; \mu' x - \lambda VaR_\alpha(x) && (6.10)\\
\text{subject to} \quad & \sum_{i \in \mathcal{M}_k} \left(\mu_i \sqrt{x'\Sigma x} + x_i (\Sigma x)_i \Psi_1(\alpha)\right) = \beta_k \left(\mu' x \sqrt{x'\Sigma x} + \Psi_1(\alpha) x' \Sigma x\right), \; k = 1, \ldots, s \\
& X = xx' \\
& \mathbf{1}' x = 1.
\end{aligned}
$$

Note that the risk budgeting constraints of (6.10) are obtained by multiplying both sides of (6.9) by $\sqrt{x'\Sigma x}$. Let us introduce a new variable

$$
Z := \begin{bmatrix} 1 & y & x' \\ y & y^2 & yx' \\ x & yx & xx' \end{bmatrix} \in \mathbb{R}^{(d+2)\times(d+2)}
$$

where $y = \sqrt{x'\Sigma x}$ and $x \in \mathbb{R}^d$. We can then rewrite (6.10) as:

$$\begin{aligned}
\max_{x,t,\delta,w,y,X,Z} \quad & \mu' x - \lambda VaR_\alpha(x), \qquad (6.11)\\
\text{subject to} \sum_{i\in\mathcal{M}_k} & (\mu_i t + \mathrm{tr}(\Gamma_i X)\Psi_1(\alpha)) = \beta_k\left(\mu'\delta + \Psi_1(\alpha) w\right),\ k = 1,\ldots,s,\\
& Z = \begin{bmatrix} 1 & y & x' \\ y & y^2 & yx' \\ x & yx & xx' \end{bmatrix},\\
& rank(Z) = 1,\\
& X = \begin{bmatrix} \mathbf{0} & \mathbf{0} & I \end{bmatrix} Z \begin{bmatrix} \mathbf{0} & \mathbf{0} & I \end{bmatrix}^\top,\\
& \delta = \begin{bmatrix} \mathbf{0} & \mathbf{0} & I \end{bmatrix} Z \begin{bmatrix} \mathbf{0} & I & \mathbf{0} \end{bmatrix}^\top,\\
& w = \begin{bmatrix} 0 & 1 & \mathbf{0}' \end{bmatrix} Z \begin{bmatrix} 0 & 1 & \mathbf{0}' \end{bmatrix}^\top,\\
& t = \begin{bmatrix} 0 & 1 & \mathbf{0}' \end{bmatrix} Z \begin{bmatrix} 1 & 0 & \mathbf{0} \end{bmatrix}^\top,\\
& \mathbf{1}' x = 1,\\
& t \geq 0
\end{aligned}$$

where I is the identity matrix in $\mathbb{R}^{d\times d}$, $\Gamma_i = e_i e_i' \Sigma$ with e_i being the i-th column of I, and $\mathbf{0}$ is the vector of d zeros in $\mathbb{R}^d$. Since the constraint $rank(Z) = 1$ is the only non-convex constraint in (6.11), we can obtain a SDP relaxation of the VaR version of the GRB problem by relaxing this constraint to $Z \succeq 0$. Because $VaR_\alpha(x)$ is not a convex function of x, it would also be necessary to replace the objective function in (6.11) with a concave function that dominates it in order to obtain a valid relaxation. This last step would not be necessary if we used CVaR as our risk measure because $CVaR(x)$ is already a convex function of x. We also note that while we have assumed normally distributed asset returns here, other distributions such as the t distribution could be used instead. See Boyd and Vandenberghe (1997) for SDP relaxations of non-convex problems, more generally.

6.2.3 An Augmented Lagrangian-MCMC Approach

Our second approach to solving the *GRB* problem involves combining the augmented Lagrangian approach with MCMC sampling to generate a point in the proximity of the *global* optimum of the *GRB* problem. This point can then used as a starting point for a non-linear optimization routine to converge to a globally optimal *GRB* portfolio. The underlying idea of the algorithm is to effectively sample points with a higher objective function value and simultaneously drive the sample path in the direction of the feasible region using the augmented Lagrangian terms.

Let Ω be the state space and $p(x) = C^{-1}p^*(x)$ denote some target probability distribution on Ω where $C := \int_\Omega p^*(x)dx$ is the normalization constant. The MCMC method is an approach to sample from $p(x)$ when the normalizing constant is hard compute. In the MCMC approach, one constructs a Markov chain on Ω using a "proposal" distribution $q(x_{t+1}|x_t)$ in such a way that $p(x)$ is the unique stationary distribution for the Markov chain. Modulo some technical conditions[3], the main requirement of MCMC is that the unnormalized distribution, $p^*(x)$, should be easy to compute. Given a current sample x_t at time t the proposal distribution, $q(\cdot|x_t)$, is used to generate a candidate sample, x_{t+1}, which is then accepted with probability

$$\alpha(x_t, x_{t+1}) := \min\left\{1, \frac{q(x_t|x_{t+1})p^*(x_{t+1})}{q(x_{t+1}|x_t)p^*(x_t)}\right\}. \tag{6.12}$$

If the candidate point x_{t+1} is rejected we then set $x_{t+1} = x_t$ and continue sampling in this manner.

Since our goal is to solve the *GRB* problem, one possibility would be to set

$$p^*(x) = \exp(\gamma F(x))\mathbb{I}_{\mathcal{F}}(x)$$

[3]See Robert and Casella (2004) for further technical details on MCMC algorithms.

where $F(x) := \mu' x - \lambda \mathcal{R}(x)$ denotes the objective function of the *GRB* problem, γ is an annealing parameter that is used to concentrate the $p^*(x)$ in the proximity of the global optimum, and $\mathbb{I}_{\mathcal{F}}(\cdot)$ denotes the indicator function of the set $\mathcal{F} = \{x \in \mathcal{X} \mid h_k(x) := \sum_{i \in \mathcal{M}_k} RC_i(x) - \beta_k \mathcal{R}(x) = 0, k = 1, \ldots, s\}$. Since the feasible region $\mathcal{F}$ of the *GRB* problem is, typically, very "small"; $p^*(x_{t+1})$ is likely to be zero, for most candidate points x_{t+1} and these points will be rejected in the acceptance-rejection step (6.12). Therefore, using MCMC to sample *only* from the feasible region is very difficult, and particularly so for high-dimensional problems.

One possible approach to overcoming these difficulties is to allow the MCMC iterates x_t to be *infeasible*, but to "direct" them towards the feasible region by adding a term which penalizes infeasibility to our definition of p^*. In particular, we could define

$$P_c(x) := F(x) + \frac{1}{2} c \|h(x)\|_2^2$$

where c is a negative constant and now use $p^*(x) = e^{\gamma P_c(x)}$ as the unnormalized density. The main difficulty with the penalty approach is that it is very sensitive to the value of the penalty parameter c. This is a well-known phenomenon, and the augmented Lagrangian algorithm was introduced in order to circumvent this numerical instability.

The AL-MCMC Algorithm

In the augmented Lagrangian approach, we define the time[4] t target distribution to be $p_t^*(x) := e^{\gamma_t \mathcal{L}_{c_t}(u_t, x)}$ where the augmented Lagrangian function of the *GRB* problem is

[4]We note that since p_t^* now changes with each iteration, there is no longer a fixed target stationary distribution for our algorithm.

defined as:

$$\begin{aligned}\mathcal{L}_{c_t}(u_t, x) &:= F(x) + u_t' h(x) + \frac{1}{2} c_t \|h(x)\|_2^2 \\ &= \mu' x - \lambda \mathcal{R}(x) + \sum_{k=1}^{s} u_{t,k} \Big(\beta_k \mathcal{R}(x) - \sum_{i \in \mathcal{M}_k} RC_i(x) \Big) \\ &+ \frac{1}{2} c_t \left(\sum_{k=1}^{s} \Big(\beta_k \mathcal{R}(x) - \sum_{i \in \mathcal{M}_k} RC_i(x) \Big)^2 \right)\end{aligned}$$

where $u_t = (u_{t,1}, \ldots, u_{t,s}) \in \mathbb{R}^s$ is a vector of time t Lagrange[5] multipliers. Let $d_{c_t}(u) := \max_{x \in \mathcal{X}} \mathcal{L}_{c_t}(u, x)$ denote the dual objective.

The initial vector of Lagrange dual multipliers u_0 and the penalty parameter c_0 are specified exogenously. The values for dual multipliers u_t and the non-increasing penalty parameter c_t for $t \geq 1$ are chosen adaptively during the course of the simulation. In particular, we decrease c_t by a predetermined value ϵ_c when there is no improvement in constraint violations over a particular iteration. When there is an improvement in constraint violation, we do not update c_t but instead update the Lagrange multipliers using the first order conditions, i.e. we set

$$u_{t+1} = u_t - \epsilon_u \nabla d_{c_t}(u_t) \tag{6.13}$$

where $d_{c_t}(u)$ denotes the dual function, and ϵ_u is a given step size. We chose not to update *both* c_t and u_t in every iteration in order to ensure that we leave the current location only after adequately exploring its neighborhood. The update of duals u_t or the penalty parameter c_t occurs every iteration whether or not the candidate x_{t+1} is accepted. We note that (6.13) represents the steepest descent iteration for minimizing d_{c_t} but one may choose other methods such as Newton's method for updating the Lagrange multipliers (see Appendix C.1). Furthermore, one can use other criteria for updating c_t. See Bertsekas (1996) for a detailed discussion of the augmented Lagrangian method.

We use a multivariate normal distribution as our proposal distribution for generating a

[5]See Appendix C.1 for further details on the augmented Lagrangian functions.

candidate value of x_{t+1}. In particular, we generate $z_t^* \sim N(0, \sigma_t^q I)$ and take

$$x_{t+1} = x_t + z_t^*$$

as our candidate point which is then accepted with probability $\alpha(x_t, x_{t+1})$. Note that the jump size variance term, σ_t^q, is allowed to vary with t. In fact, we decrease σ_t^q by a factor of κ where $0 < \kappa < 1$ as we get closer to the feasible region. More specifically, we decrease σ_t^q only if the percentage drop in the size of the constraint violations[6] is larger than a predetermined value δ.

In each iteration, irrespective of whether the proposed sample x_{t+1} is accepted or rejected, the annealing parameter γ_t is increased according to

$$\gamma_t = \sigma_\gamma \gamma_{t-1}$$

where σ_γ is a predetermined value. Thus, the AL-MCMC algorithm is a simulated annealing algorithm[7] where by forcing $\lim_{t \to \infty} \gamma_t = \infty$ we hope to drive samples towards the global optimum of the *GRB* problem.

The AL-MCMC algorithm attempts to combine the best aspects of the augmented Lagrangian method and the MCMC method. The augmented Lagrangian term guides the Markov chain towards a feasible region, while the acceptance-rejection step in the MCMC method attempts to ensure that the iterates do not get trapped in poor local maxima of the *GRB* problem.

A complete specification of our AL-MCMC algorithm is given in Algorithm 2. A feasible sample point with the highest value of $F(\cdot)$ is probably most suitable to be used as a starting point for a non-linear optimization routine. However, as the direct sampling of a feasible point is overly difficult for the *GRB* problem, the last point obtained by the algorithm is

[6] See Algorithm 2 for precise details. Depending on the specific problem under consideration, one may choose to modify this step or simply to keep σ_t^q constant across all t.

[7] See Van Laarhoven and Arts (1987) for further details

then fed to a non-linear optimization routine with the goal of quickly finding a good nearby local maximum.

We note that our algorithm is a heuristic algorithm that we hope is capable of producing good starting points for a non-linear optimization solver. We expect this algorithm to be useful for general non-convex optimization problems beyond the GRB problem of this chapter. There is also further scope for improvement. For example, we could use a more sophisticated MCMC algorithm as compared to the Metropolis-Hastings. For example, if we suspect that $F(\cdot)$ or $L_{c_t}(\cdot)$ is multi-modal then hybrid MCMC methods such as Hamiltonian MCMC should be superior. It is also possible to tailor the proposal distributions, $q(x_{t+1}|x_t)$, for the problem at hand. Note also that while it is not explicitly stated, it of course makes sense to keep track of the best feasible sample that has been obtained during the execution of the algorithm.

We therefore propose the following procedure to solve the *GRB* problem:

Step 1. Generate an initial vector x_0 to be used as the starting point of the Markov chain and choose values of $\gamma_0, \sigma_\gamma, \epsilon_c, \epsilon_u, \delta, \sigma_0^q, c_0, u_0$ and κ to be used as parameters for the AL-MCMC algorithm (Algorithm 2).

Step 2. Perform the AL-MCMC algorithm to obtain an initial point x_s to be fed to a non-linear optimization routine.

Step 3. Solve the *GRB* problem using a non-linear optimization solver with x_s obtained from Step 1 as the initial guess.

The AL-MCMC algorithm described in this section can be further enhanced by using a set of different random starting points x_0 for generating Markov chains. For instance, in our numerical experiments we used antithetic starting points to generate several values of x_s[8].

[8]Readers interested in antithetic variates in Monte Carlo techniques can refer to Robert and Casella (2004).

Algorithm 2 AL-MCMC

Choose $x_0, \gamma_0, \sigma_\gamma, \epsilon_c, \epsilon_u, \delta, \sigma_0^q, c_0, u_0, \kappa$.
for $t = 0 : T$ **do**
  Draw a candidate sample x_{t+1} from the proposal $q(x_{t+1}|x_t)$.
  Let $\alpha(x_t, x_{t+1}) = \min\left\{1, \frac{q(x_t|x_{t+1})p^*(x_{t+1})}{q(x_{t+1}|x_t)p^*(x_t)}\right\}$ where $\frac{p^*(x_{t+1})}{p^*(x_t)} = e^{\gamma_t(\mathcal{L}_{c_t}(u_t,x_{t+1})-\mathcal{L}_{c_t}(u_t,x_t))}$.
  if $\alpha \geq 1$ **then**
    $x_{t+1} \leftarrow x_{t+1}$ # Accept the candidate
  else
    Draw $p \sim \mathbb{U}[0, 1]$
    if $p \leq \alpha$ **then**
      $x_{t+1} \leftarrow x_{t+1}$ # Accept the candidate
    else
      $x_{t+1} \leftarrow x_t$ # Reject the candidate
    end if
  end if
  $\gamma_{t+1} \leftarrow \sigma_\gamma \gamma_t$ # Update the annealing parameter
  if $\|h(x_{t+1})\|_2^2 < \|h(x_t)\|_2^2$ **then**
    $u_{t+1} \leftarrow u_t - \epsilon_u \nabla d_{c_t}(u_t)$ where $d_{c_t}(u) = \max_{x \in \mathcal{X}} \mathcal{L}_{c_t}(u, x)$ # Update the Lagrange multipliers
    if $\frac{\|h(x_t)\|_2^2}{\|h(x_t+1)\|_2^2} - 1 > \delta$ **then**
      $\sigma_{t+1}^q \leftarrow \kappa \sigma_t^q$ # Update the jump size
    end if
  else
    $c_{t+1} \leftarrow c_t + \epsilon_c$ # Update the penalty parameter
  end if
end for

6.3 Numerical Results

We now present numerical results for the two proposed approaches: the SDP relaxation and the AL-MCMC algorithm. We first describe a simple example with the goal of illustrating the potential effectiveness of the AL-MCMC algorithm. We then turn to discuss the performance of the two approaches when they are tested on *GRB* problems with the number of assets ranging from 7 to 100. All the results presented in this section are based on percentage returns, i.e. returns are multiplied by 100, unless otherwise stated. Note also that the term "optimal solution" generally denotes a local optimum.

6.3.1 Numerical Results for a Toy Example

Our first problem[9] is a 5-asset problem with a variance-covariance matrix of percentage returns:

$$\Sigma = \begin{bmatrix} 94.868 & 33.750 & 12.325 & -1.178 & 8.778 \\ 33.750 & 445.642 & 98.955 & -7.901 & 84.954 \\ 12.325 & 98.955 & 117.265 & 0.503 & 45.184 \\ -1.178 & -7.901 & 0.503 & 5.460 & 1.057 \\ 8.778 & 84.954 & 45.184 & 1.057 & 34.126 \end{bmatrix}.$$

We also assumed that the expected returns of these assets are identical so that $\mu = \mu_0\mathbf{1}$. Suppose now we want to compute a long-only risk parity portfolio with minimum variance and that we apply the AL-MCMC algorithm to solve this problem. We used a single Markov chain of 5,000 points, i.e. $T = 5,000$ in Algorithm 2, and x_0 was generated uniformly from the 5-dimensional unit cube. We also used the following parameters:

- initial annealing parameter $\gamma_0 = 1$ with $\sigma_\gamma = 1.0007$;
- initial penalty parameter $c_0 = -10,000$ with $\epsilon_c = 0$;
- jump size $\sigma_0^q = 0.5$ with $\kappa = 0.75$;
- threshold parameter for updating σ_t^q, $\delta = 0.01$;
- initial Lagrange multipliers $u_0 = \mathbf{0}$ with $\epsilon_u = 0.01$.

Since this problem is relatively simple with just five constraints, we did not need to update c during the course of the algorithm. $x_{5000} = [0.1245; 0.0467; 0.0833; 0.6133; 0.1323]$ is the last point obtained from the AL-MCMC algorithm. If we specify the feasibility tolerance to 10^{-4}, this point is, in fact, the optimal risk parity solution.[10] Without a

[9] This is the same example presented in Bai *et al.* (2013).

[10] One can readily check that x^* is indeed the optimal risk parity solution by solving Problem 6.3 directly.

use of a non-linear optimization routine, the AL-MCMC algorithm was able to discover a reasonably good risk parity solution. The running time[11] for the algorithm was 1.31 seconds.

When the algorithm was applied without the penalty parameter, i.e. $c_t = 0$ for all t, or without the Lagrange multipliers, i.e. $u_t = \mathbf{0}$ for all t, it failed to converge to a risk parity solution. All of its sampled points violated the risk parity constraints by more than 10^{-4}, and hence, the help of a non-linear optimization routine was necessary for finding an optimal risk parity solution. When its last point was supplied to a non-linear optimization routine, the optimal risk parity solution was found successfully. These results demonstrate the potential advantage of incorporating the augmented Lagrangian method into the MCMC algorithm.

6.3.2 Numerical Results for the *GRB* Problem

For more general *GRB* problems we focused on the portfolio volatility risk measure $\mathcal{R}(x) := \sqrt{x'\Sigma x}$ and assumed a risk aversion parameter λ of 1. Expected asset returns μ, covariance matrices Σ, and risk budgets β are all generated randomly. In particular, we sampled μ from N(0, I), and for Σ, we first generated a matrix $V \in \mathbb{R}^{d\times d}$ using a standard normal distribution, and then converted it into a symmetric positive semidefinite matrix by multiplying it by its transpose; i.e. $\Sigma = V'V$. We generated risk-budgets $\beta = (\beta_1, \ldots, \beta_s)$ from $\mathbb{U}^s(0,1)$ and normalized them such that $\sum_{k=1}^{s} \beta_k = 1$.

We considered four test cases listed in Table 6.1 under two different scenarios. In the first scenario, we assumed that $\mu = \mu_0\mathbf{1}$, i.e. all assets have identical returns. In the second scenario, we let assets have different returns.

In order to evaluate the AL-MCMC algorithm, we generated 5 antithetic pairs of random points. For each pair (x, x_{ANTI}) of random points, we first sampled $x = (x_1, \ldots, x_d)$

[11] All our experiments were performed using `Matlab` on an Intel Core i5-680 (3.60GHz), 64-bit operating system.

Table 6.1: Test Case Descriptions

Test Case	Number of assets (d)	Number of subsets (s)	Max. function evals. for `fmincon`
1	7	3	3,000
2	30	5	5,000
3	50	5	7,000
4	100	10	20,000

from $\mathbb{U}^d(0,1)$ and set $x_{\text{ANTI}} = (1 - x_1, \ldots, 1 - x_d)$. We used each of these 10 points as the starting point x_0 for generating a single Markov chain with a length of $1,000$. Therefore, we simulated a total of 10 Markov chains. The final[12] point, i.e. the $1,000$-th point, from each chain is then used as the initial starting point of a non-linear optimization solver. In our experiments we used the `fmincon` solver with the interior point method in `Matlab` as our non-linear optimization solver. The maximum number of function evaluations allowed for `fmincon` for each test case is specified in the final column of Table 6.1. We also used `Matlab` and `CVX` (Boyd and Grant, 2008, 2014) for solving the SDP relaxation of the *GRB* problem (6.8).

6.3.2.1 Identical Returns: $\mu = \mu_0 \mathbf{1}$

In this scenario, we assume $\mu = \mu_0 \mathbf{1}$, in which case, the *GRB* problem becomes a minimum variance problem subject to the risk budgeting constraints. We solved this problem using five different approaches. First, we computed an SDP lower bound, and second, we used the solution obtained from the SDP relaxation as the initial point to be fed to `fmincon`. We refer to this as the SDP-`fmincon` approach. Next, we solved the problem using the AL-MCMC-`fmincon` approach, i.e. we simulated 10 Markov chains starting from 10 random initial points generated using the antithetic random variate method described earlier, and used the $1,000$-th iterate for each chain as the initial point for a call to `fmincon`. In order to benchmark the contribution of the MCMC algorithm, we solved the *GRB* problem using

[12]The rationale behind choosing the final point is that due to the risk budgeting constraints, sampled points are most likely to be infeasible, and an infeasible sample point attaining the highest value of $F(\cdot)$ is not necessarily the best point in terms of its proximity to the optimum solution.

`fmincon` starting from $10,000$ random starting points distributed according to $\mathbb{U}^d(0,1)$. In addition, we also considered the alternating linearization backtracking (ALM-BTKR) approach to solving risk parity problems.

The ALM-BTKR approach was introduced by Bai *et al.* (2013) where the risk parity problem was formulated as the quadratic least-squares problem:

$$\begin{aligned} \min_{x\in\mathcal{X},\theta} \quad & \textstyle\sum_{k=1}^{s}(\sum_{i\in\mathcal{M}_k} x_i(\Sigma x)_i - \theta)^2 \\ \text{subject to} \quad & a_i \le x_i \le b_i,\ i = 1,\dots,d. \end{aligned} \tag{6.14}$$

Note that risk parity is achieved when (6.14) has an optimal value of zero. This approach can be easily extended to the case where the risk of the various asset classes $\mathcal{M}_k$ are not equal by scaling θ as follows.

$$\begin{aligned} \min_{x\in\mathcal{X},\theta} \quad & \textstyle\sum_{k=1}^{s}(\sum_{i\in\mathcal{M}_k} x_i(\Sigma x)_i - \beta_k\theta)^2 \\ \text{subject to} \quad & a_i \le x_i \le b_i,\ i = 1,\dots,d. \end{aligned} \tag{6.15}$$

We stress that (6.14) and (6.15) are, in fact, feasibility problems. An optimal objective function value of zero in each problem simply indicates the discovery of a feasible risk parity or risk budgeting portfolio, but not necessarily one that has minimum risk.

In Table 6.2 we report the following metrics for each of the four methods: SDP-`fmincon`, AL-MCMC-`fmincon`, `fmincon` and ALM-BTKR.

- $\min \tilde{F}(x^*)$ denotes the best (feasible) solution of the 10 solutions obtained by the AL-MCMC-`fmincon` algorithm, and the $10,000$ solutions obtained by the `fmincon` algorithm. For ALM-BTKR, we report $\tilde{F}(x^*_{ALM-BTKR})$ where $x^*_{ALM-BTKR}$ is the solution obtained by solving (6.15). For SDP-`fmincon`, the single solution obtained from the method is reported.

- The range of the obtained (feasible) solutions. We do not report a solution range for ALM-BTKR and SDP-`fmincon` since each method yields only a single solution.

- The SDP lower bound.
- The number of failures. A failure occurs when `fmincon` does not return a feasible solution. The feasibility tolerance of 10^{-6} is used.
- t (sec). We report the amount of time taken to obtain $\min \tilde{F}(x^*)$ for the first time over the 10 trials for the AL-MCMC-`fmincon` algorithm and over the 10,000 trials for the `fmincon` algorithm. For ALM-BTKR and SDP-`fmincon`, t represents the total execution time as each only yields a single solution.

The main issue with reporting the execution time t in the manner described above is that, in order to determine the $\min \tilde{F}(\cdot)$, we first need to compute all 10 solutions for the AL-MCMC-`fmincon` algorithm and 10,000 solutions for the `fmincon` algorithm. However, given that AL-MCMC-`fmincon` uses only 10 random points and `fmincon` uses 10,000 random points, comparing the total execution time of each method is also problematic. The advantage of reporting t in the above way is that it allows us for a more fair comparison of AL-MCMC-`fmincon` with `fmincon`. Note that all solutions are reported with a precision of four decimal places.

From the results reported in Table 6.2, the portfolio $x^*_{ALM-BTKR}$ is not the minimum variance portfolio. In most cases, $\tilde{F}(x^*_{ALM-BTKR})$ is at least 50% larger than the solutions obtained by the other two methods. Also, due to the backtracking component of ALM-BTKR, the method may become prohibitively slow for higher dimensional problems. For Test Case 1 with 7 assets, the ALM-BTKR method takes less than 25 seconds, but for Test Case 4 with 100 assets the execution time is more than 330 seconds.

When all assets are assumed to have identical expected returns, the SDP relaxation appears to provide a fairly effective lower bound, against which we can compare solutions obtained from other methods. For example, in Set 3 of the Test Case 2, the difference between the SDP relaxation and the AL-MCMC-`fmincon` solution is just 0.25, and therefore, we know that the AL-MCMC solution is close to the global optimum. Since the *GRB* problem is non-convex, having an effective lower bound on its objective function is very

Table 6.2: Numerical Results for the Case of $\mu = \mu_0 \mathbf{1}$

In this table, we report results for the four methods: SDP-`fmincon`, AL-MCMC-`fmincon`, `fmincon`, and ALM-BTKR, when $\mu = \mu_0 \mathbf{1}$. The first and second columns contain the test set number and the name of the algorithm respectively. The third column reports the best solution obtained. The fourth, fifth and sixth columns report the range of the obtained solutions, the SDP lower bound and the number of failures respectively. The last column reports the execution time in seconds required to obtain the best solution.

Test Case 1: 7 Assets and 3 Subsets

Set	Method	min $\tilde{F}(x^*)$	Solution range	SDP lower bound	No. of failures	t (sec)
1	SDP-fmincon	4.83	–	4.19	0	12.78
1	AL-MCMC-fmincon	4.69	[4.69, 5.15]	4.19	0	2.81
1	fmincon	4.69	[4.69, 8.21]	4.19	9	63.64
1	ALM-BTKR	N/A	–	4.19	1	0.05
2	SDP-fmincon	33.17	–	31.72	0	3.64
2	AL-MCMC-fmincon	33.09	[33.09, 34.16]	31.72	0	18.50
2	fmincon	33.09	[33.09, 229.82]	31.72	194	100.26
2	ALM-BTKR	111.00	–	31.72	0	16.84
3	SDP-fmincon	25.76	–	25.57	0	2.06
3	AL-MCMC-fmincon	25.78	[25.78, 31.91]	25.57	0	16.90
3	fmincon	25.77	[25.77, 98.65]	25.57	840	3,253.72
3	ALM-BTKR	N/A	–	25.57	1	21.37

Test Case 2: 30 Assets and 5 Subsets

Set	Method	min $\tilde{F}(x^*)$	Solution range	SDP lower bound	No. of failures	t (sec)
1	SDP-fmincon	38.59	–	38.06	0	3.44
1	AL-MCMC-fmincon	38.59	[38.59, 38.59]	38.06	0	2.54
1	fmincon	38.59	[38.59, 328.61]	38.06	968	5.48
1	ALM-BTKR	100.61	–	38.06	0	49.39
2	SDP-fmincon	68.54	–	67.03	0	2.27
2	AL-MCMC-fmincon	68.54	[68.54, 68.54]	67.03	0	1.37
2	fmincon	68.54	[68.54, 320.29]	67.03	613	3.28
2	ALM-BTKR	105.58	–	67.03	0	47.82
3	SDP-fmincon	39.80	–	39.55	0	2.55
3	AL-MCMC-fmincon	39.80	[39.80, 39.80]	39.55	0	2.32
3	fmincon	39.80	[39.80, 348.25]	39.55	642	8.29
3	ALM-BTKR	112.50	–	39.55	0	51.58

Test Case 3: 50 Assets and 5 Subsets

Set	Method	min $\tilde{F}(x^*)$	Solution range	SDP lower bound	No. of failures	t (sec)
1	SDP-fmincon	38.24	–	36.97	0	3.65
1	AL-MCMC-fmincon	38.24	[38.24, 38.24]	36.97	0	2.68
1	fmincon	38.25	[38.25, 328.33]	36.97	2,339	4.17
1	ALM-BTKR	79.89	–	36.97	0	106.72
2	SDP-fmincon	52.58	–	51.46	0	2.99
2	AL-MCMC-fmincon	52.58	[52.58, 52.58]	51.46	0	2.16
2	fmincon	52.58	[52.58, 392.55]	51.46	1,862	2.51
2	ALM-BTKR	101.35	–	51.46	0	92.30
3	SDP-fmincon	56.24	–	51.77	0	3.57
3	AL-MCMC-fmincon	56.24	[56.24, 56.24]	51.77	0	2.86
3	fmincon	56.24	[56.24, 366.45]	51.77	2,584	2.08
3	ALM-BTKR	109.11	–	51.77	0	106.66

Test Case 4: 100 Assets and 10 Subsets

Set	Method	min $\tilde{F}(x^*)$	Solution range	SDP lower bound	No. of failures	t (sec)
1	SDP-fmincon	49.21	–	46.82	0	15.43
1	AL-MCMC-fmincon	49.21	[49.21, 49.21]	46.82	0	8.69
1	fmincon	49.21	[49.21, 310.90]	46.82	3,569	8.27
1	ALM-BTKR	103.64	–	46.82	0	333.02
2	SDP-fmincon	55.33	–	53.76	0	14.84
2	AL-MCMC-fmincon	55.33	[55.33, 103.99]	53.76	0	7.98
2	fmincon	55.33	[55.33, 314.89]	53.76	3,213	26.88
2	ALM-BTKR	110.39	–	53.76	0	354.29
3	SDP-fmincon	50.55	–	48.28	0	14.15
3	AL-MCMC-fmincon	50.55	[50.55, 50.55]	48.28	0	8.19
3	fmincon	50.55	[50.55, 314.43]	48.28	2,940	16.55
3	ALM-BTKR	94.75	–	48.28	0	361.89

informative. The SDP-`fmincon` method moreover exhibits comparable performance to the AL-MCMC-`fmincon` method in most cases except for Set 1 and Set 2 of the Test Case 1. In these cases, the optimal solutions of the SDP-`fmincon` method are 0.14 and 0.08 higher than those of the AL-MCMC-`fmincon` method.

The AL-MCMC-`fmincon` method generally has better performance than `fmincon`. The execution time of `fmincon` is inconsistent. For example, the execution time of `fmincon` in Test Case1 with 7 assets ranges anywhere from 63.64 seconds to $3,253.72$ seconds to obtain an optimal solution. For all three sets of the Test Case 3, `fmincon` failed to find a feasible solution near or over $3,000$ times. This suggests it was somewhat lucky that `fmincon` quickly converged to good solutions for Test Case 3. In contrast, the AL-MCMC-`fmincon` method was able to find an optimal solution within 20 seconds for all test cases, and in the majority of cases, it took less than 10 seconds. Furthermore, the number of failures for AL-MCMC-`fmincon` is zero; suggesting that the AL-MCMC approach is able to generate a good initial point for `fmincon`.

The range of solutions computed by `fmincon` is also much wider than the range for AL-MCMC-`fmincon`. The larger test cases clearly highlight the fact that `fmincon` significantly underperforms AL-MCMC-`fmincon`. For instance, in Set 2 Test Case 4, the `fmincon` solutions range from 55.33 to 314.89 whereas the AL-MCMC-`fmincon` solutions range from 55.33 to 103.99. This is actually the widest solution range we see for AL-MCMC-`fmincon`. In other test cases, solution ranges are very narrow for the AL-MCMC-`fmincon` algorithm.

We also compared the performance of the AL-MCMC algorithm against AL-MCMC but with the Lagrange multipliers $u_t \equiv 0$ for all t, i.e. a pure penalty method, and AL-MCMC but with the penalty parameter $c_t \equiv 0$ for all t, i.e. a pure dual method. We found that the AL-MCMC-`fmincon` algorithm performed better than both of these alternatives. These results further demonstrate the merit of integrating the augmented Lagrangian method with the MCMC algorithm. We report the results for the AL-MCMC algorithm with $u_t \equiv 0$ and $c_t \equiv 0$ in Appendix C.2.

6.3.2.2 General Expected Returns

The next set of results are for the case where expected returns are not identical across assets. Table 6.3 presents upper bounds for the *GRB* problem that were obtained using the SDP relaxation. We also report the maximum constraint violation

$$\max_{k=1,\ldots,s} \left| \sum_{i \in \mathcal{M}_k} \frac{RC_i(x^*_{SDP})}{\mathcal{R}(x^*_{SDP})} - \beta_k \right| \tag{6.16}$$

of the optimal SDP solution, x^*_{SDP}.

Table 6.3: Numerical Results for the SDP Relaxation

Table 6.3 presents the upper bound on the objective function, $F(x) = \mu' x - \mathcal{R}(x)$, of the *GRB* problem obtained via the SDP relaxation. The first and second columns contain the test set and case number, respectively. The third and fourth columns contain the number of assets (d) and the number of the subsets (s) considered in each test case, respectively. The sixth column contains the maximum constraint violation of the optimal SDP solution, i.e. (6.16), and the final column reports the execution time (in seconds) of `CVX` for solving the SDP relaxation.

Set	Test Case	d	s	Upper bound	Max. constraint violation	t (sec.)
1	1	7	3	0.91	0.09	0.55
2	1	7	3	103.82	0.76	0.38
3	1	7	3	33.68	0.39	0.38
1	2	30	5	−22.73	0.24	0.58
2	2	30	5	−14.28	0.27	0.56
3	2	30	5	28.38	0.63	0.54
1	3	50	5	5.41	0.52	0.92
2	3	50	5	24.71	0.15	0.96
3	3	50	5	−3.51	0.39	1.00
1	4	100	10	−5.58	0.19	6.51
2	4	100	10	−15.62	0.10	6.29
3	4	100	10	3.25	0.21	6.19

We see that the maximum constraint violation of x^*_{SDP} is quite large in certain cases. For example, it is[13] 0.76 for Set 2 Test Case 1, 0.63 for Set 3 Test Case 2 and 0.52 for Set 1 Test Case 3. In these cases, the SDP upper bounds are likely to be slack since the SDP solutions are far from being feasible.

We can further test this hypothesis by using x^*_{SDP} as the initial point for `fmincon` to solve the *GRB* problem as we did in the previous section, i.e. SDP-`fmincon` approach. Table 6.4 presents these results together with the results of the AL-MCMC-`fmincon` procedure and the `fmincon` starting from $10,000$ random points. In Table 6.4 we report the following

[13] Recall that the β_k's are all positive and sum to 1 so that a violation of 0.76 is indeed quite large.

metrics.

- max $F(x^*)$ denotes the best (feasible) solution of the 10 solutions from the AL-MCMC-`fmincon` algorithm, and the $10,000$ solutions from the `fmincon` algorithm. For SDP-`fmincon`, the single solution obtained from the method is reported.

- The range of the obtained (feasible) solutions. Note that we do not report a solution range for the SDP-`fmincon` method since this method produces just a single solution.

- The SDP upper bound.

- A failure occurs when `fmincon` does not return a feasible solution. The feasibility tolerance of 10^{-6} is used.

- t (sec). Consistent with the previous section, note that we report the amount of time taken to obtain max $F(x^*)$ for the first time in the 10 trials for the AL-MCMC-`fmincon` algorithm and the $10,000$ trials for the `fmincon` algorithm.

All solutions are reported with a precision of four decimal places. It is clear that the overall performance of the AL-MCMC-`fmincon` method is superior to the other two methods. In comparison to the SDP-`fmincon` approach, we note that the AL-MCMC-`fmincon` method was able to find an optimal solution to all test cases. The SDP-`fmincon` method failed to find even a feasible solution for Set 2 Test Case 1 and Set 1 Test Case 3. It is interesting to note that these failures occurred when the maximum constraint violations of the SDP relaxation solutions were noticeably large (see Table 6.3). This suggests that an upper bound obtained from the SDP relaxation may turn out to be slack when the optimal SDP solution violates the risk budgeting constraints by a large value. In the previous section, we saw that the SDP relaxation provides a relatively tight bound when all assets have identical expected returns.

In comparison to `fmincon`, AL-MCMC-`fmincon` exhibits a more stable and consistent performance. It is apparent from Table 6.4 that the solution ranges given by `fmincon` are very wide in general. For example, the solutions obtained from `fmincon` for Set 2 Test Case

Table 6.4: Numerical Results for the Case of $\mu \neq \mu_0 \mathbf{1}$

Table 6.4 presents the numerical results for the three methods: SDP-`fmincon` method, AL-MCMC-`fmincon` method and `fmincon`, when $\mu \neq \mu_0 \mathbf{1}$. The first and second columns contain the test set number and the name of the algorithm, respectively. The third column reports the best solution, i.e. $\max F(x^*)$, and the fourth column reports the range of the obtained solutions. The fifth column reports the upper bound on the objective function $F(x)$ obtained by the SDP relaxation. The sixth column reports the number of failures. The final column reports the execution time (in seconds).

Test Case 1: 7 Assets and 3 Subsets

Set	Method	$\max F(x^*)$	Solution range	SDP upper bound	No. of failures	t (sec)
1	SDP-`fmincon`	0.83	–	0.91	0	1.28
1	AL-MCMC-`fmincon`	0.83	[0.56, 0.83]	0.91	0	8.54
1	`fmincon`	0.83	[−1.24, 0.83]	0.91	13	128.25
2	SDP-`fmincon`	N/A	–	103.82	1	2.03
2	AL-MCMC-`fmincon`	63.89	[49.42, 63.89]	103.82	0	4.48
2	`fmincon`	63.88	[−235.38, 63.88]	103.82	990	6,398.85
3	SDP-`fmincon`	30.31	–	33.68	0	0.88
3	AL-MCMC-`fmincon`	30.31	[30.31, 30.31]	33.68	0	0.58
3	`fmincon`	30.31	[−83.04, 30.31]	33.68	568	2.17

Test Case 2: 30 Assets and 5 Subsets

Set	Method	$\max F(x^*)$	Solution range	SDP upper bound	No. of failures	t (sec)
1	SDP-`fmincon`	−25.33	–	−22.73	0	1.91
1	AL-MCMC-`fmincon`	−25.33	[−25.33, −25.33]	−22.73	0	2.37
1	`fmincon`	−25.33	[−309.20, −25.33]	−22.73	640	15.13
2	SDP-`fmincon`	−17.45	–	−14.28	0	3.25
2	AL-MCMC-`fmincon`	−17.45	[−17.45, −17.45]	−14.28	0	2.38
2	`fmincon`	−17.45	[−403.05, −17.45]	−14.28	601	11.29
3	SDP-`fmincon`	N/A	–	28.38	1	N/A
3	AL-MCMC-`fmincon`	14.84	[14.83, 14.84]	28.38	0	1.92
3	`fmincon`	14.84	[−332.36, 14.84]	28.38	1,136	1.75

Test Case 3: 50 Assets and 5 Subsets

Set	Method	$\max F(x^*)$	Solution range	SDP upper bound	No. of failures	t (sec)
1	SDP-`fmincon`	N/A	–	5.41	1	N/A
1	AL-MCMC-`fmincon`	−0.30	[−0.30, −0.30]	5.41	0	2.52
1	`fmincon`	−0.30	[−345.38, −0.30]	5.41	2,253	2.30
2	SDP-`fmincon`	17.86	–	24.71	0	3.63
2	AL-MCMC-`fmincon`	17.86	[17.86, 17.86]	24.71	0	1.97
2	`fmincon`	17.86	[−402.00, 17.86]	24.71	2,261	2.15
3	SDP-`fmincon`	−7.94	–	−3.51	0	3.85
3	AL-MCMC-`fmincon`	−7.94	[−7.94, −7.94]	−3.51	0	2.17
3	`fmincon`	−7.94	[−302.78, −7.94]	−3.51	2,610	1.94

Test Case 4: 100 Assets and 10 Subsets

Set	Method	$\max F(x^*)$	Solution range	SDP upper bound	No. of failures	t (sec)
1	SDP-`fmincon`	−9.35	–	−5.58	0	14.48
1	AL-MCMC-`fmincon`	−9.35	[−9.35, −9.35]	−5.58	0	7.00
1	`fmincon`	−9.35	[−353.13, −9.35]	−5.58	2,652	8.00
2	SDP-`fmincon`	−17.28	–	−15.62	0	13.97
2	AL-MCMC-`fmincon`	−17.27	[−17.28, −17.27]	−15.62	0	7.43
2	`fmincon`	−17.27	[−302.76, −17.27]	−15.62	1,878	8.43
3	SDP-`fmincon`	0.67	–	3.25	0	15.06
3	AL-MCMC-`fmincon`	0.67	[0.67, 0.67]	3.25	0	8.24
3	`fmincon`	0.67	[−273.36, 0.67]	3.25	3,025	43.22

1 range from -235.38 to 63.89, and `fmincon` took $6,398.85$ seconds discover an optimum despite Test Case 1 being a low dimensional problem with only 7 assets. This suggests that when `fmincon` uses an unfavorable starting point, x_s, the solution it obtains can be very far from a good local optimum. The results of Set 3 Test Case 4 further demonstrates this. Of $10,000$ random starting points, `fmincon` failed to find a feasible solution $3,025$ times. Considering the wide solution range of $[-273.36, 0.67]$, many of the failed as well as the feasible solutions are likely to be far from an optimum. In many cases, the stand-alone `fmincon` has a considerable number of failures and so in general one would need to try many random points as the starting point, x_s.

In contrast, the AL-MCMC-`fmincon` method yields solutions whose ranges are much narrower. Except for Set 1 and Set 2 of Test Case 1, all the obtained solutions are very close to the best solution. Moreover, the AL-MCMC-`fmincon` method had zero failures for all test cases. This means that the starting points generated by the AL-MCMC algorithm are much more favorable than random starting points. Also note that in all cases, AL-MCMC-`fmincon` finds an optimal solution in less than 10 seconds. Based on these observations we can conclude that AL-MCMC-`fmincon` appears to be a much more reliable tool for solving the general GRB problem with non-identical expected asset returns.

6.4 Conclusions

In this chapter we proposed a generalized risk budgeting (GRB) approach to portfolio construction. In comparison with the existing risk-based asset allocation techniques, our approach provides investors with more flexibility in that it allows investors to optimize a risk-return profile and to define risk budgets for possibly overlapping subsets of assets. Minimum variance, risk parity and risk budgeting strategies are therefore special cases of GRB strategies.

Although we showed that the GRB problem can be formulated as a convex optimization problem in an important special case, the general GRB problem is a non-convex optimization problem. We introduced an SDP relaxation for bounding the optimal value of the GRB problem. When all assets have identical expected returns, our numerical results suggested that this SDP bound was quite tight, and could therefore be used to assess the quality of solutions produced by other approaches. Our main contribution in this chapter is a simulation-based algorithm that combines augmented Lagrangian optimization ideas with MCMC methods. The goal of this algorithm is to compute a candidate solution in the neighborhood of the optimum, or a very good local optimal solution of the GRB problem. This candidate solution could then be used as the starting point for a standard non-linear optimization solver. In several numerical experiments our AL-MCMC algorithm was indeed successful in finding very good starting points.

We also note that our AL-MCMC approach is a general solution approach for solving non-convex optimization problems. The augmented Lagrangian algorithms is a very popular algorithm for computing local optimum solutions for non-convex problem. Combining this algorithm with the MCMC method opens up the possibility of converging to the *global* optimal solution, or at least providing a good starting-point for a non-linear optimization routine. In addition, this approach can be implemented very easily and is computationally fast. We expect it to be of particular use for non-convex problems with small feasible regions where computing a good starting point is challenging. We intend to apply this approach to such problems in future research.

Chapter 7

Conclusions

This dissertation has set out new quantitative approaches to security selection and portfolio optimization. In the first part of the dissertation, we proposed two security selection strategies based on the notions of efficiency: (i) a quantitative stock selection strategy based on operational efficiency and (ii) a quantitative currency selection strategy based on macroeconomic efficiency. The key ideas underlying these strategies were the following.

- There is a potential positive link between firm's operational efficiency and its stock performance;
- There is a plausible relationship between country's macroeconomic efficiency and its currency performance.

This study therefore has not only provided an empirical analysis of efficiency-based approaches to security selection, but it has also provided an empirical investigation of the aforementioned relationships.

We must stress that the proposed investment models were simple and intuitive, which we believe is essential in quantitative modeling. Our strategy consisted of two basic steps: (i) quantifying operational (macroeconomic) efficiency of firms (countries) being considered for investment, and (ii) taking market positions based on their estimated efficiency. In quantifying efficiency, we systematically integrated various financial ratios (macroeconomic variables) of firms (countries) into a single score by means of data envelopment analysis (DEA). Our methodical approach resulted in a comprehensive and interpretable measure of efficiency soundly built on firm (country) fundamentals. The merit of our security selection strategies is in utilizing such a measure as a guidepost for making investment decisions.

Through a thorough testing of the proposed stock and currency selection strategies in empirical settings provided respectively by the U.S. Information Technology (IT) sector, and the countries of G10 and emerging market currencies, we deduced that:

- stocks of efficient firms yield superior and yet, less volatile returns relative to those of inefficient firms; and
- currencies of efficient countries have stronger performance than those of inefficient countries.

We also provided empirical evidence that operational efficiency has significant explanatory power in describing the cross-sectional attributes of the U.S. IT stock returns; and explained that the risk of efficiency loss over time is the most probable source of excess returns of investing in efficient firms. With regard to the currency selection strategy, its successful performance depicted in the empirical results suggests that macroeconomic variables are indeed the crucial factors influencing exchange rate dynamics.

In the first part of the dissertation, we also detailed the development of the data-driven, joint variable selection method for DEA. Our simulation study showed that its performance was more robust and favorable compared to the widely-used variable selection methods in the DEA literature. The real-world example we presented, furthermore, emphasized the importance as well as the virtue of using variable selection algorithms that are tailored for DEA when building DEA models.

In the second part of the dissertation, we introduced a generalized risk budgeting (GRB) approach to portfolio construction. Considering that minimum variance, risk parity and risk budgeting portfolios are all special instances of a GRB portfolio, our study essentially provided a general framework for constructing risk-based portfolios. This framework offers a much greater degree of flexibility in the way it allows for: (i) short sales of assets, (ii) mean-risk portfolio optimization, and (iii) risk budgets to be defined for possibly overlapping subsets of assets. Such flexibility certainly makes our approach more attractive compared to the prevailing risk-based approaches to portfolio optimization.

Our solution approaches to the GRB portfolio optimization problem demonstrated the potential benefits of using a SDP relaxation and a numerical sampling algorithm for solving difficult non-linear optimization problems. For instance, this study's numerical examples illustrated that relatively tight bounds on the optimal objective function value of the GRB problem could be obtained using the SPD relaxation. Theses examples also confirmed the effectiveness of the AL-MCMC algorithm, which deliberately incorporates the augmented Lagrangian (AL) method into the Markov chain Monte Carlo (MCMC) method, in finding a favorable point that could be used as an initial point to be supplied to a standard non-linear optimization routine for solving the GRB problem. Most methods of optimization are sensitive to initial points; for example, interior-point methods are known to perform poorly if the starting point is unfavorable (Gertza *et al.*). On that account, a noble advantage of the proposed AL-MCMC algorithm is that it is a generic method for determining an initial point and can serve as a heuristic tool for solving any optimization problem.

"Sound financial decision making is a quantitative trade-off between risk and return." This is a broadly accepted and understood idea by the investing public nowadays. There is no doubt that quantitative algorithms and models have become invaluable tools for building investment portfolios. Along with the tremendous increase in the use of such tools, quantitative portfolio construction has evolved into its own discipline and further a growing area of research. It is to this very area, this dissertation has made contributions.

Part III

Bibliography

Bibliography

J. S. Abarbanell and B. J. Bushee. Fundamental Analysis, Future Earnings, and Stock Prices. *Journal of Accounting Research*, 35(1):1 – 24, 1997.

N. Adler and B. Golany. Including Principal Component Weights to Improve Discrimination in Data Envelopment Analysis. *Journal of the Operational Research Society*, 53(9):985 – 991, 2002.

S. N. Afriat. The Construction of a Utility Function from Expenditure Data. *International Economic Review*, 8:67 – 77, 1967.

S. N. Afriat. Efficiency Estimation of Production Functions. *International Economic Review*, 13(3):568 – 598, 1972.

I. Alam and R. Sickles. The Relationship between Stock Market Returns and Technical Efficiency Innovations: Evidence from the U.S. Airline Industry. *Journal of Productivity Analysis*, 9:35 – 51, 1998.

Y. Amihud. Illiquidity and Stock Returns: Cross-Section and Time-Series Effects. *Journal of Financial Markets*, 5:31 – 56, 2002.

A. Artzner, F. Delbaen, J. M. Eber, and D. Heath. Coherent Measures of Risk. *Mathematical Finance*, 9(3):203 – 228, 1999.

N. K. Avkiran. Decomposing Technical Efficiency and Window Analysis. *Studies in Economics and Finance*, 22(2):61 – 91, 2004.

X. Bai, K. Scheinberg, and R. Tutuncu. Least-Squares Approach to Risk Parity in Portfolio Selection. *Working Paper*, 2013.

R. D. Banker, A. Charnes, and W. W. Cooper. Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science*, 30:1078 – 1092, 1984.

P. W. Bauer, A. N. Berger, G. D. Ferrier, and D. B. Humphrey. Consistency Conditions for Regulatory Analysis of Financial Institutions: A Comparison of Frontier Efficiency Techniques. *Journal of Economics and Business*, 50:85 – 114, 1998.

D. P. Bertsekas. *Constrained Optimization and Lagrange Method*. Athena Scientific, Belmont, MA, 1996.

D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific Belmont, MA, 1999.

J. Bilson. The Monetary Approach to the Exchange Rate – Some Empirical Evidence. *IMF Staff Paper*, 25(March):48 – 75, 1978.

F. Black and R. Litterman. Global Portfolio Optimization. *Financial Analysts Journal*, 48:28 – 43, 1992.

J. M. Boughton. Test of Performance of Reduced-Form Exchange Rate Models. *Journal of International Economics*, 23:41 – 56, 1987.

W. F. Bowlin, A. Charnes, W. W. Cooper, and H. D. Sherman. Data Envelopment Analysis and Regression Approaches to Efficiency Estimation and Evaluation. *Annals of Operations Research*, 2:59 – 94, 1985.

M. Boyd and M. Grant. Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited, 2008. `http://stanford.edu/~boyd/graph_dcp.html`.

M. Boyd and M. Grant. CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`, 2014.

S. Boyd and L. Vandenberghe. Semidefinite Programming Relaxations of Non-Convex Problems in Control and Combinatorial Optimization. In V. Roychowdhuri A. Paulraj, and and C. Schaper, editors, *Communications, Computation, Control and Signal Processing: A Tribute to Thomas Kailath*, pages 279 – 288. Kluwer, 1997.

R. A. Brealey and S. C. Myers. *Principles of Corporate Finance*. McGraw-Hill, 1991.

B. Bruder and T. Roncalli. Managing Risk Exposures Using the Risk Budgeting Approach. *SSRN*, 2012.

M. M. Carhart. On Persistence Mutual Fund Performance. *Journal of Finance*, 52:57 – 82, 1997.

L. K. C. Chan, N. Jegadeesh, and J. Lakonishok. Momentum Strategies. *Journal of Finance*, 51(5):1681 – 1713, 1996.

V. Charles and M. Kumar, editors. *Data Envelopment Analysis and Its Applications to Management*. Cambridge Scholars Publishing, Newcastle, UK, 2012.

A. Charnes, W. W. Cooper, and E. Rhodes. Measuring the Efficiency of Decision-Making Units. *European Journal of Operations Research*, 2:429 – 444, 1978.

A. Charnes, C. T. Clark, W. W. Cooper, and B. Golany. A Developmental Study of Data Envelopment Analysis in Measuring the Efficiency of Maintenance Units in the U. S. Air Forces. *Annals of Operations Research*, 2:95 – 112, 1985.

A. Charnes, W. W. Cooper, D. Divine, G. A. Klopp, and J. Stutz. An Application of Data Envelopment Analysis to U. S. Army Recruitment Districts. *Research Report CCS*, 436, 1985.

A. Charnes, W. W. Cooper, B. Golany, L. Seiford, and J. Stutz. Foundations of Data Envelopment Analysis for Pareto-Koopmans Efficient Empirical Production Functions. *Journal of Econometrics*, 30(1–2):91 – 107, 1985.

A. Charnes, W. W. Cooper, A. Y. Lewin, and L. M. Seiford. *Data Envelopment Anlaysis: Theory, Methodology, and Applications*. Kluwer Academic Publishers, Norwell, MA, 1995.

W. C. Chen and A. L. Johnson. The Dynamics of Performance Space of Major League Baseball Pitchers 1871 – 2006. *Annals of Operational Research*, 181(1):287 – 302, 2010.

W. W. Cooper, K. S. Park, and J. T. Pastor. RAM: Range Adjusted Measure of Inefficiency for Use with Additive Models, as Relations to Other Models and Measures in DEA. *Journal of Productivity Analysis*, 11:5 – 42, 1999.

W. W. Cooper, L. M. Seiford, and K. Tone. *Data Envelopment Analysis: A Comprehensive Text with Models, References and DEA-Solver Software*. Kluwer Academic Publishers, Norwell, MA, 2000.

W. W. Cooper, L. M. Seiford, and J. Zhu. *Handbook on Data Envelopment Analysis*. Kluwer Academic Publishers, Norwell, MA, 2004.

V. DeMiguel, L. Garlappi, and R. Uppal. Optimal versus Naive Diversification: How Inefficient Is the 1/N Portfolio Strategy? *The Review of Financial Studies*, 22:1915 – 1953, 2009.

E. Dimson and M. Mussavian. Market Efficiency. *The Current State of Business Disciplines*, 3:959 – 970, 2000.

R. Dornbusch. Expectations and Exchange Rate Dynamics. *Journal of Political Economy*, 84(6):1161 – 1176, 1976.

P. D. Easton and M. E. Zmijewski. SEC Form 10K/10Q Reports and Annual Reports to Shareholders: Reporting Lags and Squared Market Model Prediction Errors. *Journal of Accounting Research*, 31(1):113 – 129, 1993.

J. Eckstein and D.P. Bertsekas. On the Douglas-Rachford Splitting Method and the Proximal Point Algorithm for Maximal Monotone Operators. *Mathematical Programming*, 55(1):293 – 318, 1992.

N. C. P. Edirisinghe and X. Zhang. Portfolio Selection under DEA-based Relative Financial Strength Indicators: Case of U.S. Industries. *Journal of the Operational Research Society*, 59:842 – 856, 2008.

D. D. Evanoff and P. R. Israilevich. Scale Elasticity and Efficiency for U.S. Banks. *Working Paper Series, Issues in Financial Regulation: Federal Reserve Bank of Chicago*, pages 91 – 115, 1991.

E. F. Fama and K. R. French. The Cross-Section of Expected Stock Returns. *Journal of Finance*, 47(2):427 – 465, 1992.

E. F. Fama and K. R. French. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*, 33(1):3 – 56, 1993.

E. F. Fama and K. R. French. Profitability, Investment, and Average Returns. *Journal of Financial Economics*, 82:491 – 518, 2006.

E. F. Fama and J. MacBeth. Risk Return and Equilibrium: Empirical Tests. *Journal of Political Economy*, 81:607 – 636, 1973.

M. J. Farrell. The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society Series A (General)*, 120(3):253 – 290, 1957.

S. Fischer. Exchange Rate Regimes and Crisis Prevention in Emerging Market Economics. In *4th Annual Conference of the Brookings Wharton Papers on Financial Services*. International Monetary Fund, 2001.

J. A. Frankel. On the Mark: A Theory of Floating Exchange Rates Based on Real Interest Rate Differentials. *American Economic Review*, 69:610 – 622, 1979.

T. J. Frecka and C. F. Lee. Generalized Financial Ratio Adjustment Processes and Their Implications. *Journal of Accounting Research*, 21(1):308 – 316, 1983.

J. A. Frenkel. A Monetary Approach to the Exchange Rate: Doctrinal Aspects and Empirical Evidence. *Scandinavian Journal of Economics*, 78(2):200 – 224, 1976.

H. O. Fried, C. A. K. Lovell, and S. S. Schmidt. *The Measurement of Productive Efficiency and Productivity Growth*. Oxford University Press, New York, NY, 2008.

L. Friedman and Z. Sinuany-Stern. Combining Ranking Scales and Selecting Variables in the DEA Context: The Case of Industrial Branches. *Computers and Operations Research*, 25(9):781 – 791, 1998.

B. Frijns, D. Margaritis, and M. Psillaki. Firm Efficiency and Stock Returns. *Journal of Productivity Analysis*, 37(3):295 – 306, 2012.

D. Gabay and B. Mercier. A Dual Algorithm for the Solution of Nonlinear Variational Problems via Finite Element Approximation. *Computers & Mathematics with Applications*, 2(1):17 – 40, 1976.

D. Galagedera and P. Silvapulle. Experimental Evidence on Robustness of Data Envelopment Analysis. *Journal of the Operational Research Society*, 54:654 – 660, 2003.

M. Gertza, J. Nocedal, and A. Sartenar. A Starting Point Strategy for Nonlinear Interior Methods. *Applied Mathematics Letters*, Aug.

R. Glowinski and A. Marroco. Sur l'approximation, par elements finis d'ordre un, et la resolution, par penalisation-dualite d'une classe de problemes de Dirichlet non lineares. *Rev. Francaise d'Automat. Inf. Recherche Operationelle*, (9):41 – 76, 1975.

B. Golany and Y. Roll. An Application Procedure for DEA. *OMEGA*, 17(3):237 – 250, 1989.

E. Grifell-Tatje and C. A. K. Lovell. The Sources of Productivity Change in Spanish Banking. *European Journal of Operational Research*, 98(2):364 – 380, 1997.

M. R. Hestenes. Multiplier and Gradient Methods. *Journal of Optimization Theory and Applications*, 4(5):303–320, 1969.

P. Hooper and J. Morton. Fluctuations in the Dollar: A Model of Nominal and Real Exchange Rate Determination. *Journal of International Money and Finance*, 1:39 – 56, 1982.

J. O. Horrigan. The Determination of Long-Term Credit Standing with Financial Ratios: Empirical Research in Accounting: Selected Studies. *Journal of Accounting Research*, 4:44 – 62, 1966.

C. Hsiao. *Analysis of Panel Data*. Cambridge Universirty Press, Cambridge, UK, 2003.

P. Isard. *Exchange Rate Economics*. Cambrige University Press, New York, NY, 1995.

P. Isard. Uncovered Interest Parity. *Interantional Monetary Fund*, 1996.

N. Jegadeesh and S. Titman. Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *Journal of Finance*, 48:65 – 91, 1993.

N. Jegadeesh. Evidence of Predictable Behavior of Security Returns. *Journal of Finance*, 45:881 – 898, 1990.

L. Jenkinson and M. Anderson. A Multivariate Statistical Approach to Reducing the Number of Variables in Data Envelopment Analysis. *European Journal of Operational Research*, 147:51 – 61, 2003.

A. L. Johnson and T. Kuosmanen. Data Envelopment Analysis as Nonparametric Least-Squares Regression. *Operations Research*, 58(1):149 – 160, 2010.

B. D. Jordan, S. A. Ross, and R. W. Westerfield. *Fundamentals of Corporate Finance*. McGraw-Hill/Irwin, New York, NY, 9th edition, 2010.

T. C. Koopmans. Analysis Production as an Efficient Combination of Activities. In T. C. Koopmans, editor, *Activity Analysis of Production and Allocation*. John Wiley & Sons, New York, NY, U.S., 1951.

G. Kozmetsky, S. Thore, and F. Phillips. DEA of Financial Statements Data: The U.S. Computer Industry. *Journal of Productivity Analysis*, 5:229 – 248, 1994.

L. Kuntz and A. Vera. Modular Organization and Hospital Performance. *Health Services Management Research*, 20(1):48 – 58, 2007.

J. Lakonishok, A. Shleifer, and R. W. Vishny. Contrarian Investment, Extrapolation, and Risk. *Journal of Finance*, 49(5):1541 – 1578, 1994.

B. Lehmann. Fads, Martingales, and Market Efficiency. *Quarterly Journal of Economics*, 105:1 – 28, 1990.

B. Lev. Industry Average as Targets for Financial Ratios. *Journal of Accounting Research*, pages 290 – 299, Autumn 1969.

J. Lewellen. Predicting Returns with Financial Ratios. *Journal of Financial Economics*, 74:209 – 235, 2004.

J. Liu, S. Ji, and J. Ye. SLEP: Sparse Learning with Efficient Projections. *Arizona State University*, 2009.

J. S. Liu, L. Y. Y. Louis, and B. J. Y. Lin W. Lu. Data Envelopment Analysis 1978 – 2010: A Citation-based Literature Survey. *OMEGA*, 41(1):3 – 15, 2013.

C. A. K. Lovell and J. T. Pastor. Units Invariant and Translation Invariant DEA Models. *Operations Research Letters*, 18(3):147 – 151, 1995.

S. Maillard, T. Roncalli, and J. Teiletche. The Properties of Equally Weighted Risk Contribution Portfolios. *Journal of Portfolio Management*, 36(4):60 – 70, 2010.

A. J. McNeil, R. Frey, and P. Embrechts. Princeton University Press, Princeton, NJ, 2005.

R. Meese and K. Rogoff. Empirical Exchange Rate Models of the Seventies: Do they Fit Out of Sample? *Journal of International Economics*, 14:3 – 24, 1983.

L. Meier, S. Van De Geer, and P. Buhlmann. The Group Lasso for Logistic Regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):53 – 71, 2008.

R. Michaely, R. H. Thaler, and K. L. Womack. Price Reactions to Dividend Initiations and Omissions: Overreaction or Drift? *Journal of Finance*, 50:573 – 608, 1995.

N. R. Nataraja and A. L. Johnson. Guidelines for Using Variable Selection Techniques in Data Envelopment Analysis. *European Journal of Operational Research*, 215:662 – 669, 2011.

G. X. Nguyen and P. E. Swanson. Firm Characteristics, Relative Efficiency, and Equity Returns. *Journal of Financial and Quantitative Analysis*, 44(1):213 – 236, 2009.

J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer Verlag, 1999.

R. Novy-Marx. The Other Side of Value: The Gross Profitability of Premium. *NBER Working Paper*, (15940), 2012.

J. Ou and S. H. Penman. Financial Statement Analysis and the Prediction of Stock Returns. *Journal of Accounting and Economics*, 11:295 – 329, 1989.

Y. A. Ozcan and M. J. McCue. Development of a Financial Performance Index for Hospitals: DEA Approach. *Journal of the Operational Research Society*, 47:18 – 26, 1996.

J. C. Paradi and P. Phille. Financial Performance of Ontario (Canada) Credit Unions: An Application of DEA in the Regulatory Environment. *European Journal of Operational Research*, 139:339 – 350, 2002.

J. T. Pastor and J. Ruiz. Variables with Negative Values in DEA. In J. Zhu and W. W. Cook, editors, *Modeling Data Irregularities and Structural Complexities in Data Envelopment Analysis*, pages 63 – 84. Springer, Berlin, Germany, 2007.

J. T. Pastor, J. L. Ruiz, and I. Sirvent. A Statistical Test for Nested Radial DEA Models. *Operations Research*, 50(4):728 – 735, 2002.

J. T. Pastor. New Additive DEA Models for Handling Zero and Negative Data. *Working Paper*, 1994.

A. C. Paynor and I. M. D. Little. *Higgledy Piggledy Growth Again: An Investigation of the Predictability of Company Earnings and Dividends in the U. K., 1951-1961*. A.M. Kelley, New York, NY, 1966.

D. W. Pearce. *The MIT Dictionary of Modern Economics*. The MIT Press, Cambridge, MA, U.S., 3rd edition, 1986.

S. H. Penman. An Evaluation of Accounting Rate of Return. *Journal of Accounting, Auditing, and Finance*, 6(2):233 – 255, 1991.

M. J. D. Powell. A Method for Nonlinear Constraints in Minimization Problems. In R. Fletcher, editor, *Optimization*. Academic Press, New York, NY, 1972.

E. Qian. On the Financial Interpretation of Risk Contribution: Risk Budgets Do Add Up. *Journal of Investment Management*, 4(4):41 – 51, 2006.

T. Qin and I. Song. Joint Variable Selection for Data Envelopment Analysis via Group Sparsity. *SSRN*, 2013.

Z. Qin, K. Scheinberg, and D. Goldfarb. Efficient Block-coordinate Descent Algorithms for the Group Lasso. *Mathematical Programming Computation*, 5:143 – 169, 2013.

B. Rees. *Financial Analysis*. Prentice Hall, London, UK, 2nd edition, 1995.

A. Reisman. Data Envelopment Analysis: Preface to a Special Issue of Scio-Econonic Planning Sciences. *Scio-Economic Planning Sciences*, 38:151 – 121, 2003.

C. P. Robert and G. Casella. *Controlling Monte Carlo Variance*. Springer, New York, NY, 2004.

R.T. Rockafellar. The Multiplier Method of Hestenes and Powell Applied to Convex Programming. *Journal of Optimization Theory and Applications*, 12(6):555–562, 1973.

J. Ruggiero. Impact Assessment of Input Omission on DEA. *International Journal of Information Technology & Decision Making*, 4(3):359 – 368, 2005.

G. J. Schinasi and P. A. V. B. Swamy. The Out-of-sample Forecasting Performance of Exchange Rate Models When Coefficients Are Allowed to Change. *Journal of International Money and Finance*, 8(3):375 – 390, 1989.

T. R. Sexton, R. H. Silkman, and Hogan A. J. *Data Envelopment Analysis: Critique and Extensions*. Jossey-Bass, San Francisco, CA, 1986.

A. Shleifer. *Inefficient Markets: An Introduction to Behavioral Finance*. Oxford University Press, Oxford, UK, 2000.

L. Simar and P. W. Wilson. Testing Restrictions in Nonparametric Efficiency Models. *Communications in Statistics*, 30(1):159 – 184, 2001.

P. Smith. Data Envelopment Analysis Applied to Financial Statements. *Omega Interational Journal of Management Science*, 18:131 – 138, 1990.

F. Spinu. An Algorithm for Computing Risk Parity Weights. *SSRN*, 2013.

E. Thanassoulis. A Comparison of Regression Analysis and Data Envelopment Analysis as Alternative Methods for Performance Assessments. *Journal of the Operational Research Society*, 44(11):1129 – 1144, 1993.

R. Tibshirani. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996.

H. Tulkens and O. Vanden Eeckaut. Non-parametric Efficiency, Progress and Regress Measures for Panel Data: Methodological Aspects. *European Journal of Operational Research*, 80:474 – 499, 1995.

T. Ueda and Y. Hoshiai. Application of Component Analysis for Parsimonious Summarization of DEA Inputs and/or Outputs. *Journal of the Operations Research Society of Japan*, 40(4):466 – 478, 1997.

E. van den Berg, M. Schmidt, M.P. Friedlander, and K. Murphy. Group Sparsity via Linear-time Projection. Technical report, Technical Report TR-2008-09, Department of Computer Science, University of British Columbia, 2008.

P. J. Van Laarhoven and E. H. L. Arts. Reidel Publishers, Amsterdam, 1987.

H. Wang and P. Schmidt. One Step and Two Step Estimation of the Effects of Exogenous Variables on Technical Efficiency Levels. *Journal of Productivity Analysis*, 18:129 – 1044, 2002.

J. W. Wilcox. The P/B-ROE Valuation Model. *Financial Analysts Journal*, 35(1):58 – 66, 1984.

W. T. Woo. The Monetary Approach to Exchange Rate Determination Under Rational Expectations: The Dollar-Deutschmark Rate. *Journal of International Economics*, 18(1/2):1 – 16, 1985.

S.J. Wright, R.D. Nowak, and M.A.T. Figueiredo. Sparse Reconstruction by Separable Approximation. *IEEE Transactions on Signal Processing*, 57(7):2479–2493, 2009.

Q. Yeh. The Application of Data Envelopment Analysis in Conjuction with Financial Ratios for Bank Performance Evaluation. *Journal of the Operational Research Society*, 47(8):980 – 988, August 1996.

M. Yuan and Y. Lin. Model Selection and Estimation in Regression with Grouped Variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67, 2006.

P. Yue. Competition Between U.S. and Japanese Electronic Companies. *The University of Texas at Austin: IC2 Institute*, 1991.

P. Yue. Data Envelopment Analysis and Commerical Banking Performance: A Primer with Applications to Missouri Banks. *Federal Reserve Bank of St. Luis Review*, 74-1, 1992.

H. Zou and T. Hastie. Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.

Part IV

Appendices

Appendix A

Quantitative Stock Selection Based on Operational Efficiency

A.1 The Complete List of Industries

Table A.1 provides the complete list of industries considered in this study. For the classifying firms into different groups of industries, we use the global industry classification standard (GICS). GICS is widely accepted among investment researchers, portfolio and asset managers as one of the most commonly-used industry classifications in the world. It is developed by S&P and Morgan Stanley Capital International (MSCI) Barra. DEA and multiple linear regression models are applied to each industry in the level 3 classification. Due to a very few number of firms classified under Electronic Equipment, Instruments & Components and Office Electronics, these two industries are grouped together as one industry.

Table A.1: GICS Information Technology (IT) Sector Breakdown

Table A.1 presents the list of industry groups (level 2), industries (level 3) and industry sub-groups (level 4) of S&P IT sector as per GICS.

Industry Groups (Level 2)	Industries (Level 3)	Industry Sub-Groups (Level 4)
Semiconductors &	Semiconductors	Semiconductors
Semiconductor Equipment	Semiconductor Equipment	Semiconductor Equipment
Software &	IT Services	Data Processing & Outsourced Services
Services		IT Consulting & Other Services
	Internet Software & Services	Internet Software & Services
	Software	Application Software
		Home Entertainment Software
		Systems Software
Technology Hardware &	Communication Equipment	Communication Equipment
Equipment	Computers & Peripherals	Computer Hardware
		Computer Storage & Peripherals
	Electronic Equipment, Instruments	Electronic Components
	& Components	Electronic Equipment & Instruments
		Electronic Manufacturing Services
		Technology Distribution
	Office Electronics	Office Electronics

A.2 Formulae for Financial Ratios

This section provides formulae used for calculating each of the financial ratios used in this study.

- Accounts payable turnover: calculated as purchases divided by average accounts payable. Purchases are the sum of ending inventory and cost of goods sold minus beginning Inventory. Average accounts payable is the sum of beginning and ending accounts payable divided by 2;
- Accounts receivable turnover: calculated as trailing 12 month net sales divided by year over year average accounts receivable. Trailing 12 month values are calculated by adding the most recent four quarters;
- Asset turnover: calculated as net sales divided by average total assets. Average is the average of the beginning balance and ending balance;
- Asset to equity: calculated as total assets divided by shareholder's equity;

- Common equity to total asset: calculated as total common equity divided by total assets;
- Long term debt to total equity: calculated as long-term borrowings divided by total shareholder's equity;
- Total debt to common equity: calculated as the sum of short-term and long-term borrowings divided by total shareholder's equity;
- Total debt to total asset: calculated as the sum of short-term and long-term borrowings divided by total assets;
- Cash ratio: calculated as the sum of cash, near cash items and marketable securities divided by current liabilities;
- Current ratio: calculated as current assets divided by current liabilities;
- Quick ratio: calculated as liquid assets divided by current liabilities, where liquid assets are the sum of cash, near cash items, marketable securities, short-term investments, accounts receivables and notes receivables;
- Price to book ratio: calculated as market price divided by book value per share;
- Earning before interests and taxes (EBIT) margin: calculated as trailing 12 month operating income (loss) divided by trailing 12 month net sales. Trailing values are calculated using the most recent four quarters;
- Gross margin: calculated as net sales minus cost of goods sold, divided by net sales;
- Profit margin: calculated as net income divided by net sales;
- Return on asset: calculated by dividing trailing 12 month net income (losses) minus trailing 12 month total cash preferred dividends by average assets. Trailing values are calculated using the most recent four quarters;

- Return on equity: calculated as trailing 12 month net income (losses) minus trailing 12 month cash preferred dividends, divided by average of total common equity. Total common equity is the sum of share capital, additional paid in capital and retained earnings. Average is the average of the beginning balance and ending balance.

A.3 The LASSO-based Variable Selection Algorithm

LASSO (Tibshirani, 1996) is a regularization method in statistics that is often used for model selection, where it works by penalizing models based on the number of their parameters. LASSO minimizes the residual sum of squares subject to the sum of the absolute value of the coefficients being less than a constant.

Suppose data is generated by a linear regression model

$$Y_n = X_n \beta^n + \epsilon_n,$$

where $\epsilon_n = (\epsilon_1, \ldots, \epsilon_n)^T$ is a vector of i.i.d. random variables with mean 0 and variance σ^2, $Y_n = (y_1, ..., y_n)^T$ is a response vector, X_n is an $n \times l$ design matrix and $\beta^n = (\beta_1^n, \ldots, \beta_l^n)^T$ is a vector of model coefficients. LASSO estimates $\hat{\beta}^n = (\hat{\beta}_1^n, \ldots, \hat{\beta}_l^n)^T$ are then defined by

$$\hat{\beta}^n(\lambda) = \arg\min_{\beta} ||Y_n - X_n\beta||_2^2 + \lambda||\beta||_1,$$

where $|| \cdot ||_1$ stands for the l_1 norm of a vector. The parameter $\lambda \geq 0$ controls the amount of regularization applied to the estimates. Setting $\lambda = 0$ reverses the LASSO problem to the ordinary least-squares (OLS) problem while assigning a very large to λ will completely shrink β^n to 0, thus leading to the empty or null model. In general, moderate values of λ will cause shrinkage of the solutions towards 0, and some coefficients may end up being exactly 0.

By applying the same idea, we modify the LASSO method for DEA models. With n DMUs, l inputs $X = x_{k,j} \in \mathbb{R}^{l \times n}$ and 1 output $Y = y_{1,j} \in \mathbb{R}^{1 \times n}$, the DEA weighted additive model (WAM) version of the LASSO estimates for a particular $\text{DMU}_p, p \in \{1, \ldots, n\}$ is obtained by solving the following LP for some $\lambda > 0$:

$$
\begin{aligned}
\min \quad & \left(-y_{1,p} v_1' + \sum_{k=1}^{l} x_{k,p} u_k' + w \right) && + \quad \lambda \sum_{k=1}^{l} u_k' \\
\text{subject to} \quad & -y_{1,j} v_1' + \sum_{k=1}^{l} x_{k,j} u_k' + w && \geq \quad \frac{y_{1,j}}{(l+1) R_{1,p}^{+}} - \sum_{k=1}^{l} \frac{x_{k,j}}{(l+1) R_{k,p}^{-}}, \ j = 1, \ldots, n, \\
& v_1' && \geq \quad 0, \\
& u_k' && \geq \quad 0, \ k = 1, \ldots, l \\
& && w, \quad \text{free}
\end{aligned}
$$

where $R_{r,p}^{+} = \max_{j=1,\ldots,n}\{y_{r,j}\} - y_{1,p}$, and $R_{k,p}^{-} = x_{k,p} - \min_{j=1,\ldots,n}\{x_{k,j}\}$, which can be seen as the ranges of possible improvements in each variable for DMU_p, and u' and v' are the shifted values of u and v in the dual of the original DEA WAM to obtain zero lower bounds. To avoid problems with zeros, the slacks corresponding to variables with $R_{1,p}^{+} = 0$ and/or $R_{k,p}^{-} = 0$ for some $k = 1, \ldots, l$ are ignored. Similar to the LASSO, the parameter $\lambda \geq 0$ controls the amount of regularization applied to the estimates. Setting $\lambda = 0$ and shifting the values of u' and v' back to u and v respectively reverses the LASSO-based DEA WAM problem to the dual of the original DEA WAM problem.

We solve the LASSO-based DEA WAM problem for each DMU_j, $j = 1, \ldots, n$ for each quarter in the estimation period in our empirical study. Then, by using a heuristic approach, we select the input variables that are most relevant to all the DMUs across all time periods. Our heuristic approach is simply based on cross-sectional time-series averages of the LASSO-based DEA WAM input weights (or coefficients). One can further refine this by applying the group LASSO (see Chapter 5 (Qin and Song, 2013)).

A.4 Results of the Factor Model Analyses

This section provides the results of the three factor models: (i) the CAPM, (ii) the Fama-French three-factor model, and (iii) the six-factor model, with and without the transaction costs. For each model, we report the factor loading estimates for:

- the top and bottom decile efficiency-based portfolios;
- the top and bottom decile residual-based portfolios; and
- the top-minus-bottom spread of the efficiency-based portfolios.

A.4.1 CAPM

Table A.2: Excess Returns on CAPM

Table A.2 presents α and factor loading estimate from the CAPM

$$(r_{p,t} - r_{f,t}) = \alpha_p + \beta_{p,r_m - r_f}(r_{m,t} - r_{f,t}) + \epsilon_{p,t}$$

where $r_{p,t}$ is the return of the portfolio p, $r_{m,t}$ is the market return and $r_{f,t}$ is the risk-free return (or 1-month T-bill rate) at time t for efficiency- and residual-based equally-weighted portfolios. Panel A reports estimates for the top- and bottom-decile efficiency-based equally-weighted portfolios, Panel B reports estimates for the top- and bottom-decile residual-based equally-weighted portfolios, and Panel C reports estimates for the top-minus-bottom spread of the efficiency-based equally-weighted portfolios. The t-stats are shown in parenthesis. *, **, *** represent significance at 10%, 5% and 1% respectively.

Panel A. Efficiency-based Equally-Weighted Portfolio

Top-Decile Portfolio	α	$\beta_{r_m - r_f}$
1-Month	10.97**	1.22***
	(2.42)	(16.70)
3-Month	10.22***	1.21***
	(2.21)	(15.70)
6-Month	11.68***	1.26***
	(2.52)	(16.36)
Bottom-Decile Portfolio	α	$\beta_{r_m - r_f}$
1-Month	−0.65	1.61***
	(−0.11)	(16.89)
3-Month	1.62	1.55***
	(0.29)	(16.20)
6-Month	−0.73	1.52***
	(−0.13)	(15.64)

Panel B. Residual-based Equally-Weighted Portfolio

Top-Decile Portfolio	α	$\beta_{r_m - r_f}$
1-Month	2.89	1.7***
	(0.46)	(16.24)
3-Month	15.89**	1.74***
	(2.44)	(16.48)
6-Month	11.22*	1.73***
	(1.79)	(16.52)
Bottom-Decile Portfolio	α	$\beta_{r_m - r_f}$
1-Month	10.87**	1.54***
	(1.93)	(17.05)
3-Month	3.17	1.54***
	(0.62)	(17.64)
6-Month	5.05	1.54***
	(0.90)	(15.99)

Panel C. Top-minus-bottom Spread

	α	$\beta_{r_m - r_f}$
1-Month	13.17**	−0.31***
	(2.16)	(−3.13)
3-Month	9.95*	−0.26***
	(1.75)	(−2.72)
6-Month	11.71**	−0.17**
	(2.43)	(−2.12)

A.4.2 Fama and French Three-Factor Model

Table A.3: Excess Returns on Fama and French Factors

Table A.3 presents α and factor loading estimates from the Fama-French three-factor model

$$(r_{p,t} - r_{f,t}) = \alpha_p + \beta_{p,m}(r_{m,t} - r_{f,t}) + \beta_{p,SMB}SMB_t + \beta_{p,HML}HML_t + \epsilon_{p,t}$$

where $r_{p,t}$ is the return of the portfolio p, $r_{m,t}$ is the market return, SMB_t is the size factor, HML_t is the value factor, and $r_{f,t}$ is the risk-free return (or 1-month T-bill rate) at time t for efficiency- and residual-based equally-weighted portfolios. Panel A reports estimates for the top- and bottom-decile efficiency-based equally-weighted portfolios, Panel B reports estimates for the top- and bottom-decile residual-based equally-weighted portfolios, and Panel C reports estimates for the top-minus-bottom spread of the efficiency-based equally-weighted portfolios. The t-stats are shown in parenthesis. *, **, *** represent significance at 10%, 5% and 1% respectively.

Panel A. Efficiency-based Equally-Weighted Portfolio

Top-Decile Portfolio	α	$\beta_{r_m - r_f}$	β_{SMB}	β_{HML}
1-Month	8.99**	1.10***	0.66***	−0.35***
	(2.23)	(16.29)	(5.34)	(−3.32)
3-Month	7.89*	1.13***	0.62***	−0.46***
	(1.92)	(15.64)	(4.94)	(−3.78)
6-Month	10.38**	1.20***	0.52***	−0.44***
	(2.42)	(15.75)	(3.82)	(−3.41)
Bottom-Decile Portfolio				
1-Month	0.14	1.48***	0.83***	−1.00***
	(0.03)	(20.42)	(6.30)	(−8.96)
3-Month	0.83	1.51***	0.68***	−1.01***
	(0.19)	(18.88)	(4.85)	(−7.5)
6-Month	−1.41	1.46***	0.78***	−1.02***
	(−0.33)	(18.16)	(5.44)	(−7.56)

Panel B. Residual-based Equally-Weighted Portfolio

Top-Decile Portfolio	α	$\beta_{r_m - r_f}$	β_{SMB}	β_{HML}
1-Month	0.05	1.51***	1.07***	−0.58***
	(0.01)	(16.38)	(6.32)	(−4.08)
3-Month	11.74**	1.60***	1.03***	−0.70***
	(2.19)	(17.27)	(6.30)	(−4.48)
6-Month	6.94	1.55***	1.13***	−0.49***
	(1.33)	(16.43)	(6.74)	(−3.07)
Bottom-Decile Portfolio				
1-Month	9.79**	1.41***	0.76***	−0.60***
	(2.00)	(17.27)	(5.07)	(−4.74)
3-Month	0.76	1.45***	0.76***	−0.63***
	(0.18)	(18.49)	(5.47)	(−4.75)
6-Month	2.76	1.43***	0.89***	−0.68***
	(0.58)	(16.41)	(5.72)	(−4.65)

Panel C. Top-minus-bottom Spread

	α	$\beta_{r_m - r_f}$	β_{SMB}	β_{HML}
1-Month	11.19*	−0.27***	−0.26	0.60***
	(1.92)	(−2.74)	(−1.49)	(4.01)
3-Month	9.67*	−0.26***	−0.19	0.43***
	(1.70)	(−2.61)	(−1.07)	(2.57)
6-Month	11.62***	−0.16**	−0.32**	0.54***
	(2.55)	(−1.99)	(−2.22)	(4.02)

A.4.3 Six-Factor Models

Table A.4: Excess Returns of Efficiency- and Residual-based Portfolios on Six-Factors

Table A.4 presents α and factor loading estimates from the six-factor model

$$(r_{p,t} - r_{f,t}) = \alpha_p + \beta_{p,r_m - r_f}(r_{m,t} - r_{f,t}) + \beta_{p,SMB}SMB_t + \beta_{p,HML}HML_t + \beta_{p,MOM}MOM_t + \beta_{p,LTR}LTR_t + \beta_{p,STR}STR_t + \epsilon_{p,t}$$

where $r_{p,t}$ is the return of the portfolio p, $r_{m,t}$ is the market return, SMB_t is the size factor, HML_t is the value factor, MOM_t is the Carhart momentum factor, LTR_t is the long term reversal factor, STR_t is the short term reversal factor and $r_{f,t}$ is the risk-free return (or 1-month T-bill rate) at time t for efficiency- and residual-based equally-weighted portfolios. Panel A reports estimates for the top- and bottom-decile efficiency-based equally-weighted portfolios, Panel B reports estimates for the top- and bottom-decile residual-based equally-weighted portfolios, and Panel C reports estimates for the top-minus-bottom spread of the efficiency-based equally-weighted portfolios. The t-stats are shown in parenthesis. *, **, *** represent significance at 10%, 5% and 1% respectively.

Panel A. Efficiency-based Equally-Weighted Portfolio

Top-Decile Portfolio	α	$\beta_{r_m-r_f}$	β_{SMB}	β_{HML}	β_{MOM}	β_{LTR}	β_{STR}
1-Month	9.20**	1.09***	0.55***	−0.56***	−0.05	0.38***	0.00
	(2.31)	(13.40)	(4.24)	(−4.13)	(−0.87)	(2.48)	(−0.04)
3-Month	8.35**	1.13***	0.52***	−0.67***	−0.05	0.40***	−0.05
	(2.06)	(13.22)	(3.89)	(−4.6)	(−0.72)	(2.55)	(−0.48)
6-Month	11.38***	1.13***	0.43***	−0.65***	−0.16**	0.40***	−0.05
	(2.74)	(13.03)	(3.08)	(−4.33)	(−2.36)	(2.51)	(−0.50)
Bottom-Decile Portfolio							
1-Month	0.22	1.38***	0.82***	−1.02***	−0.14**	0.04	0.08
	(0.05)	(15.67)	(5.82)	(−7.01)	(−2.00)	(0.26)	(0.84)
3-Month	1.15	1.39***	0.64***	−1.07***	−0.17**	0.15	0.08
	(0.26)	(14.73)	(4.31)	(−6.63)	(−2.33)	(0.86)	(0.70)
6-Month	−0.36	1.34***	0.67***	−1.24***	−0.23***	0.42***	0.02
	(−0.09)	(14.92)	(4.72)	(−8.05)	(−3.33)	(2.56)	(0.20)

Panel B. Residual-based Equally-Weighted Portfolio

Top-Decile Portfolio	α	$\beta_{r_m-r_f}$	β_{SMB}	β_{HML}	β_{MOM}	β_{LTR}	β_{STR}
1-Month	0.56	1.43***	0.98***	−0.74***	−0.17*	0.32	−0.01
	(0.11)	(12.79)	(5.52)	(−4.03)	(−1.92)	(1.52)	(−0.12)
3-Month	12.88***	1.55***	0.87***	−1.04***	−0.17**	0.63	−0.11***
	(2.50)	(14.60)	(5.25)	(−5.69)	(−2.07)	(3.22)	(−0.90)
6-Month	8.38*	1.48***	1.06***	−0.74***	−0.22***	0.46**	−0.22*
	(1.67)	(13.91)	(6.25)	(−4.03)	(−2.69)	(2.34)	(−1.73)
Bottom-Decile Portfolio							
1-Month	9.95**	1.43***	0.67***	−0.78***	−0.01	0.31*	−0.03
	(2.02)	(14.29)	(4.19)	(−4.69)	(−0.10)	(1.66)	(−0.25)
3-Month	1.07	1.45***	0.65***	−0.83***	−0.04**	0.39	−0.01
	(0.25)	(15.48)	(4.42)	(−5.18)	(−0.54)	(2.24)	(−0.11)
6-Month	3.61	1.35***	0.81***	−0.85***	−0.16**	0.33*	−0.02
	(0.77)	(13.35)	(5.03)	(−4.90)	(−2.09)	(1.76)	(−0.21)

Panel C. Top-minus-bottom Spread

	α	$\beta_{r_m-r_f}$	β_{SMB}	β_{HML}	β_{MOM}	β_{LTR}	β_{STR}
1-Month	10.74*	−0.20*	−0.38**	0.40**	0.08	0.34	0.07
	(1.84)	(−1.71)	(−2.03)	(2.03)	(0.92)	(1.53)	(0.54)
3-Month	9.41*	−0.17	−0.27	0.29	0.13	0.25	0.00
	(1.65)	(−1.40)	(−1.45)	(1.44)	(1.42)	(1.15)	(−0.01)
6-Month	11.25**	−0.12	−0.31**	0.58***	0.08	−0.07	0.01
	(2.45)	(−1.21)	(−2.01)	(3.51)	(1.11)	(−0.39)	(0.13)

A.4.4 Results of the Factor Model Analyses After Transaction Costs

Table A.5: Excess Returns After Transaction Costs on CAPM, Fama-French, and Six Factors

Table A.5 presents α and factor loading estimates from the CAPM (Panel A), the Fama-French three-factor model (Panel B), and the six-factor model (Panel C) for top-decile efficiency- and residual-based equally-weighted portfolios after incorporating transaction costs of 50 basis points. The t-stats are shown in parenthesis. *, **, *** represent significance at 10%, 5% and 1% respectively.

Panel A. CAPM

Efficiency-based Portfolio	α	$\beta_{r_m-r_f}$
1-Month	9.95**	1.22***
	(2.21)	(16.75)
3-Month	9.22**	1.21***
	(2.00)	(15.74)
6-Month	10.64**	1.26***
	(2.32)	(16.41)
Residual-based Portfolio	α	$\beta_{r_m-r_f}$
1-Month	0.63	1.70***
	(0.10)	(16.16)
3-Month	13.20**	1.73***
	(2.07)	(16.53)
6-Month	8.73	1.73***
	(1.41)	(16.54)

Panel B. Fama-French Three-Factor Model

Efficiency-based Portfolio	α	$\beta_{r_m-r_f}$	β_{SMB}	β_{HML}
1-Month	7.99**	1.10***	0.66***	−0.35***
	(2.00)	(16.34)	(5.36)	(−3.32)
3-Month	6.94*	1.13***	0.62***	−0.46***
	(1.70)	(15.71)	(4.94)	(−3.82)
6-Month	9.37**	1.20***	0.51***	−0.43***
	(2.20)	(15.80)	(3.80)	(−3.39)
Residual-based Portfolio	α	$\beta_{r_m-r_f}$	β_{SMB}	β_{HML}
1-Month	−2.19	1.51***	1.07***	−0.58***
	(−0.42)	(16.26)	(6.29)	(−4.02)
3-Month	9.12*	1.60***	1.03***	−0.69***
	(1.73)	(17.33)	(6.34)	(−4.47)
6-Month	4.52	1.55***	1.13***	−0.48***
	(0.88)	(16.43)	(6.75)	(−3.02)

Panel C. Six-Factor Model

Efficiency-based Portfolio	α	$\beta_{r_m-r_f}$	β_{SMB}	β_{HML}	β_{MOM}	β_{LTR}	β_{STR}
1-Month	8.18**	1.09***	0.55***	−0.55***	−0.05	0.37**	0.00
	(2.07)	(13.43)	(4.25)	(−4.12)	(−0.86)	(2.45)	(0.00)
3-Month	7.39*	1.13***	0.51***	−0.67***	−0.04	0.40***	−0.05
	(1.84)	(13.32)	(3.89)	(−4.64)	(−0.65)	(2.57)	(−0.52)
6-Month	10.35***	1.13***	0.42***	−0.64***	−0.15**	0.40***	−0.06
	(2.51)	(13.09)	(3.07)	(−4.33)	(−2.32)	(2.51)	(−0.56)
Residual-based Portfolio	α	$\beta_{r_m-r_f}$	β_{SMB}	β_{HML}	β_{MOM}	β_{LTR}	β_{STR}
1-Month	−1.73	1.42***	0.99***	−0.74***	−0.16*	0.32	−0.01
	(−0.33)	(12.68)	(5.50)	(−3.98)	(−1.89)	(1.50)	(−0.04)
3-Month	10.22**	1.55***	0.87***	−1.03***	−0.16**	0.62	−0.12
	(2.02)	(14.70)	(5.28)	(−5.67)	(−1.94)	(3.21)	(−0.97)
6-Month	5.93	1.48***	1.06***	−0.73***	−0.21***	0.45**	−0.23*
	(1.20)	(13.95)	(6.28)	(−3.98)	(−2.66)	(2.31)	(−1.85)

A.4.5 Variable Definitions and Additional Results for the Fama-MacBeth Regressions

This section provides the definitions of the control variables used in the Fama-MacBeth regressions. The definitions of the proxies for accruals (ACCR), leverage (LEV), illiquidity (ILLIQ), momentum (MOM) and reversal (REV) are as follows.

- Accruals (ACCR): We use annual changes in the working capital as the proxy for accruals.

- Leverage (LEV): We use the long term debt to equity ratio as the proxy for leverage.

- Illiquidity (ILLIQ): The illiquidity variable for each stock in month t is defined as the ratio of the absolute monthly stock return to its dollar trading volume:

$$ILLIQ_{i,t} = \frac{|r_{i,t}|}{DVOL_{i,t}} \tag{A.1}$$

  where $r_{i,t}$ is the return on stock i in month t, and $DVOL_{i,t}$ is the respective monthly trading volume in dollars (Amihud, 2002).

- Momentum (MOM): The (intermediate) momentum variable for each stock in month t is defined as the cumulative return on the stock over the previous 11 months starting 2 months ago; i.e. the cumulative return from month $t-12$ to month $t-2$ (Jegadeesh and Titman, 1993).

- Reversal (REV): The (short term) reversal variable for each stock in month t is defined as the return on the stock over the previous month; i.e. the return in month $t-1$ (Jegadeesh, 1990; Lehmann, 1990).

A.5 Sensitivity Analysis for the Exponentially Weighted Moving Average (EWMA) Smoothing Parameter

This section compares the performances of the efficiency-based portfolios formed using different smoothing parameters for computing the exponentially weighted moving average (EWMA) of efficiency scores. We consider 4-, 8-, 10-, 12-, and 16-quarters as smoothing parameters. As can be seen from Table A.6, the performance of the efficiency-based portfolio is not so much sensitive to the changes in the EWMA smoothing parameter.

Table A.6: Results of the EWMA Smoothing Factor Sensitivity Analysis

Table A.6 presents the results of EWMA smoothing factor sensitivity analysis. Panel A, B and C report descriptive statistics, risk-return tradeoff measures and risk measures of the top-decile efficiency-based equally-weighted portfolio returns based on various EWMA smoothing factors (4-, 8-, 10-, 12- and 16-quarter) for 1-, 3- and 6-month lagged investments respectively.

Panel A. 1-Month Lagged Investment

	Descriptive Statistics				Risk-Return Tradeoff Measures		Risk Measures		
Quarters	Total Return (%)	Vol. (%)	Kurt.	Skew.	Sharpe Ratio	Sortino Ratio	VaR (%)	CVaR (%)	MDD (%)
4Q	12.93	24.88	3.07	−0.02	0.43	0.69	36.51	46.63	48.70
8Q	14.01	24.86	3.03	0.02	0.48	0.77	36.21	45.82	48.17
10Q	13.16	24.49	3.10	−0.02	0.45	0.72	35.84	47.39	46.74
12Q	12.97	24.58	3.10	−0.03	0.44	0.70	36.04	49.58	45.86
16Q	13.86	24.74	3.22	−0.04	0.47	0.75	36.06	48.02	50.34

Panel B. 3-Month Lagged Investment

	Descriptive Statistics				Risk-Return Tradeoff Measures		Risk Measures		
Quarters	Total Return (%)	Vol. (%)	Kurt.	Skew.	Sharpe Ratio	Sortino Ratio	VaR (%)	CVaR (%)	MDD (%)
4Q	15.15	24.4	3.11	−0.01	0.54	0.88	35.19	44.71	48.50
8Q	15.45	24.58	2.91	0.02	0.55	0.90	35.40	44.19	49.17
10Q	13.73	23.98	3.02	−0.03	0.49	0.78	34.90	45.63	48.58
12Q	14.11	23.72	2.88	0.01	0.51	0.83	34.39	45.68	46.22
16Q	15.23	23.84	3.03	−0.07	0.55	0.89	34.28	41.75	49.20

Panel C. 6-Month Lagged Investment

	Descriptive Statistics				Risk-Return Tradeoff Measures		Risk Measures		
Quarters	Total Return (%)	Vol. (%)	Kurt.	Skew.	Sharpe Ratio	Sortino Ratio	VaR (%)	CVaR (%)	MDD (%)
4Q	15.67	24.99	3.25	0.23	0.55	0.94	36.00	42.97	47.29
8Q	14.33	25.19	3.02	0.06	0.49	0.81	36.64	44.98	49.64
10Q	14.53	25.27	3.05	−0.06	0.50	0.80	36.71	47.61	50.50
12Q	14.35	25.06	2.95	−0.06	0.50	0.80	36.43	48.88	48.67
16Q	15.07	25.04	3.21	0.15	0.52	0.87	36.22	44.83	47.51

A.6 Performance Results of the Long-Short Equally-Weighted Portfolios

This section presents the performance of the long-short equally-weighted portfolios. Two types of long-short portfolios are considered: 1X0-X0 portfolios and leveraged neutral portfolios. With regard to 1X0-X0 portfolios, we consider 100-0, 110-10, 120-20, 140-40 and 150-50 portfolios, and their respective performance metrics are shown in Table A.7. With regard to leveraged neutral portfolios, we consider leverage values ranging from 0.5 to 2, and the performance metrics of the corresponding portfolios are shown in Table A.8.

Table A.7: Performance Results of the 1X0-X0 Portfolios

Table A.7 presents descriptive statistics and risk-return trade-off measures of the 1X0-X0 portfolio returns. An 1X0-X0 portfolio has 1X0% exposure to long positions consist of the constituents of the top efficiency decile and X0% exposure to short positions consist of the constituents of the bottom efficiency decile. Panel A, B and C report descriptive statistics and risk-return tradeoff measure of various 1X0-X0 portfolio returns for 1-, 3- and 6-month lagged investments respectively.

Panel A. 1-Month Lagged Investment

	Descriptive Statistics				Risk-Return Tradeoff Measures	
Strategy	Total Return (%)	Volatility (%)	Kurtosis	Skewness	Sharpe Ratio	Sortino Ratio
100-0	12.93	24.88	3.07	−0.02	0.43	0.69
110-10	14.27	24.84	3.30	0.11	0.49	0.80
120-20	15.59	24.98	3.61	0.25	0.54	0.92
130-30	16.87	25.28	3.96	0.38	0.58	1.03
140-40	18.12	25.75	4.34	0.51	0.62	1.12
150-50	19.33	26.37	4.70	0.62	0.65	1.20

Panel B. 3-Month Lagged Investment

	Descriptive Statistics				Risk-Return Tradeoff Measures	
Strategy	Total Return (%)	Volatility (%)	Kurtosis	Skewness	Sharpe Ratio	Sortino Ratio
100-0	15.15	24.4	3.11	−0.01	0.54	0.88
110-10	16.04	24.42	3.31	0.10	0.57	0.96
120-20	16.91	24.59	3.54	0.22	0.6	1.04
130-30	17.74	24.9	3.80	0.34	0.63	1.12
140-40	18.54	25.35	4.05	0.46	0.65	1.18
150-50	19.3	25.93	4.30	0.57	0.67	1.23

Panel C. 6-Month Lagged Investment

	Descriptive Statistics				Risk-Return Tradeoff Measures	
Strategy	Total Return (%)	Volatility (%)	Kurtosis	Skewness	Sharpe Ratio	Sortino Ratio
100-0	15.67	24.99	3.25	0.23	0.55	0.94
110-10	17.11	24.85	3.28	0.32	0.61	1.08
120-20	18.64	24.88	3.41	0.43	0.67	1.23
130-30	20.17	25.02	3.60	0.54	0.73	1.39
140-40	21.70	25.24	3.84	0.66	0.78	1.55
150-50	23.22	25.56	4.10	0.77	0.83	1.71

Table A.8: Performance Results of the Leveraged Neutral Portfolios

Table A.8 presents descriptive statistics and risk-return tradeoff measures of the leveraged neutral portfolios. A leveraged neutral portfolio earns risk-free return on its collateral and has 100% exposure to long positions consist of the constituents of the top efficiency decile and has 100% exposure to short positions consist of the constituents of the bottom efficiency decile. Panel A, B and C report descriptive statistics and risk-return tradeoff measure of various leveraged neutral portfolio returns for 1-, 3- and 6-month lagged investments respectively.

Panel A. 1-Month Lagged Investment

	Descriptive Statistics				Risk-Return Tradeoff Measures	
Leverage	Total Return (%)	Volatility (%)	Kurtosis	Skewness	Sharpe Ratio	Sortino Ratio
0.50	7.77	9.50	4.47	0.57	0.59	1.07
0.75	10.37	14.21	4.48	0.56	0.58	1.05
1.00	12.79	18.93	4.49	0.56	0.56	1.01
1.25	15.03	23.65	4.49	0.55	0.55	0.98
1.50	17.06	28.37	4.49	0.55	0.53	0.95
1.75	18.88	33.09	4.49	0.55	0.51	0.91
2.00	20.48	37.81	4.49	0.55	0.49	0.87

Panel B. 3-Month Lagged Investment

	Descriptive Statistics				Risk-Return Tradeoff Measures	
Leverage	Total Return (%)	Volatility (%)	Kurtosis	Skewness	Sharpe Ratio	Sortino Ratio
0.50	5.96	8.88	3.84	0.07	0.44	0.71
0.75	7.67	13.29	3.84	0.07	0.42	0.68
1.00	9.21	17.70	3.85	0.06	0.40	0.65
1.25	10.55	22.12	3.85	0.06	0.38	0.62
1.50	11.69	26.53	3.85	0.06	0.36	0.58
1.75	12.63	30.94	3.85	0.06	0.34	0.55
2.00	13.34	35.35	3.85	0.06	0.32	0.51

Panel C. 6-Month Lagged Investment

	Descriptive Statistics				Risk-Return Tradeoff Measures	
Leverage	Total Return (%)	Volatility (%)	Kurtosis	Skewness	Sharpe Ratio	Sortino Ratio
0.50	7.16	7.36	3.58	0.49	0.71	1.30
0.75	9.64	11.02	3.58	0.48	0.70	1.28
1.00	12.02	14.68	3.57	0.48	0.69	1.26
1.25	14.31	18.34	3.57	0.48	0.67	1.23
1.50	16.49	22.00	3.57	0.47	0.66	1.21
1.75	18.57	25.66	3.57	0.47	0.65	1.18
2.00	20.54	29.33	3.57	0.47	0.63	1.16

A.7 Performance of the Efficiency Decile Portfolios

The performance measures of all ten efficiency-decile portfolios are presented in Table A.9. The first column gives the decile number. The decile 1 represents the most efficient firms while the decile 10 represents the least efficient firms.

Table A.9: Performance Results of the Efficiency Decile Portfolios

Table A.9 presents descriptive statistics and risk-return tradeoff measures of the efficiency decile portfolios. Panel A, B and C report descriptive statistics and risk-return tradeoff measure of the efficiency decile portfolio returns for 1-, 3- and 6-month lagged investments respectively.

Panel A. 1-Month Lagged Investment

	Descriptive Statistics				Risk-Return Tradeoff Measures	
Decile	Total Return (%)	Volatility (%)	Kurtosis	Skewness	Sharpe Ratio	Sortino Ratio
1	12.93	24.88	3.07	−0.02	0.43	0.69
2	13.64	28.77	2.87	−0.15	0.40	0.62
3	10.83	28.45	3.62	−0.01	0.31	0.47
4	8.82	30.81	3.12	0.10	0.22	0.34
5	6.58	29.05	3.12	−0.19	0.15	0.23
6	2.77	28.95	3.13	−0.40	0.02	0.03
7	6.39	29.65	2.86	−0.11	0.14	0.21
8	6.35	29.85	2.93	−0.19	0.14	0.21
9	1.16	32.06	3.48	0.19	−0.03	−0.04
10	−0.45	32.84	3.95	−0.23	−0.08	−0.11

Panel B. 3-Month Lagged Investment

	Descriptive Statistics				Risk-Return Tradeoff Measures	
Decile	Total Return (%)	Volatility (%)	Kurtosis	Skewness	Sharpe Ratio	Sortino Ratio
1	15.15	24.40	3.11	−0.01	0.54	0.88
2	12.28	28.85	3.11	−0.02	0.35	0.56
3	17.14	28.09	4.18	0.22	0.54	0.89
4	15.78	30.04	3.30	0.03	0.46	0.74
5	7.73	29.22	3.30	−0.23	0.19	0.29
6	8.83	28.02	2.85	−0.29	0.24	0.36
7	9.14	29.21	2.93	−0.14	0.24	0.37
8	5.69	29.09	3.06	−0.14	0.13	0.18
9	11.24	30.64	3.69	0.12	0.30	0.48
10	5.86	31.04	3.52	0.14	0.12	0.19

Panel C. 6-Month Lagged Investment

	Descriptive Statistics				Risk-Return Tradeoff Measures	
Decile	Total Return (%)	Volatility (%)	Kurtosis	Skewness	Sharpe Ratio	Sortino Ratio
1	15.67	24.40	3.25	0.23	0.55	0.94
2	13.79	28.85	2.93	0.06	0.42	0.69
3	9.54	28.09	3.38	−0.05	0.26	0.40
4	12.67	30.04	3.63	0.15	0.38	0.62
5	11.63	29.22	3.16	0.04	0.34	0.54
6	6.83	28.02	3.09	−0.32	0.17	0.25
7	4.67	29.21	3.15	−0.29	0.09	0.13
8	7.22	29.09	3.31	−0.17	0.19	0.28
9	5.82	30.64	3.23	−0.13	0.13	0.19
10	2.16	31.04	3.84	−0.11	0.01	0.01

A.8 Formulae for Performance Measures

- Annualized Sharpe Ratio:

$$\text{Annualized Sharpe Ratio} = \frac{E[R - Rf]}{\sigma} = \frac{E[R - Rf]}{\sqrt{var[R - Rf]}}$$

where R is the annualized asset return and Rf is the annualized risk free rate of return. σ is annualized the standard deviation of the excess of the asset return. The annualized asset return is computed as follows.

$$R = \left(\frac{S_n}{S_0}\right)^{12/n} - 1$$

where $S_n = S_0 \Pi_{i=1}^{n}(1 + R_i)$, n is the number of months in the investment period, S_0 is the initial investment, e.g. a constant, and R_i is the rate of return for month i. The annualized standard deviation is computed as:

$$\sigma = \sigma_m \sqrt{12}$$

where σ_m is the monthly standard deviation. The annualized risk free rate of return Rf is computed in the same way the annualized asset return R is computed.

- Annualized Sortino Ratio:

$$\text{Annualized Sortino Ratio} = \frac{R - T}{DR}$$

where R is the annualized asset return, T is the target return, i.e. annualized risk free rate of return, Rf, and DR is the annualized downside risk. DR is computed as follows:

$$DR = DR_m \sqrt{12}$$

where

$$DR_m = \sqrt{\frac{\sum_{i=1}^{n} \min\{R_i - Rf_i, 0\}^2}{n}},$$

n is the number of months in the investment period, R_i is the rate of return for month i and Rf_i is the monthly risk free rate for month i.

- Annualized VaR_α: Given a confidence level $\alpha \in (0,1)$, the VaR of a portfolio at the confidence level α is given by the smallest number l such that the probability that the loss L exceeds l is at most $(1-\alpha)$. Mathematically, if L is the loss of a portfolio, then $\mathrm{VaR}_\alpha(L)$ is the level α-quantile; i.e.

$$\mathrm{VaR}_\alpha(L) = \inf\{l \in \mathbb{R} : P(L > l) \leq 1 - \alpha\} = \inf\{l \in \mathbb{R} : F_L(l) \geq \alpha\}.$$

  $\mathrm{VaR}_{\alpha,m}$ is derived using monthly returns of a portfolio: in particular, we used `portvrisk` function in `Matlab`. The respective annualized VaR_α is computed as $\mathrm{VaR}_\alpha = \mathrm{VaR}_{\alpha,m}\sqrt{12}$. Note that `portvrisk` calculates VaR_α using a normal distribution.

- Annualized CVaR_α:

$$\mathrm{CVaR}_\alpha = E[R_i | R_i \leq \mathrm{VaR}_{\alpha,m}]\sqrt{12}.$$

Appendix B

Quantitative Currency Selection Based on Macroeconomic Efficiency

B.1 Structural Exchange Rate Models

In monetary approach, exchange rate is defined as the relative price of two currencies, and therefore is determined by the relative supply and demand for those currencies. The two key underlying assumptions of this approach are the purchasing power parity (PPP) and uncovered interest rate parity (UIP) conditions.

B.1.1 Purchasing Power Parity (PPP)

The purchasing power parity (PPP) states that exchange rates between two currencies are in equilibrium when their purchasing power is the same in each of the two countries. This means that the exchange rate between the currencies of two countries should equal the ratio of the national prices levels; i.e.

$$S_{i,j} = P_i / P_j \tag{B.1}$$

where $S_{i,j}$ is the nominal exchange rate measured in currency i per unit of currency j, P_i and P_j are the respective price levels of countries i and j. By taking logarithms of both sides of (B.1), we obtain an alternative representation of the PPP condition,

$$s_{i,j} = p_i - p_j.$$

B.1.2 Uncovered Interest Rate Parity (UIP)

The interest rate parity states that the domestic interest rate should equal the foreign interest rate plus the expected change of the exchange rate. If the foreign exchange market participants are risk-neutral and have rational expectations, then the interest rate differential between two countries, say a country i and a country j, should be offset by the expected change in the logarithm of the nominal exchange rate. Accordingly, the interest rate of a country i, r_i, is represented as

$$r_i = r_j + f_{i,j} - s_{i,j} \tag{B.2}$$

where r_j is the interest rate of country j, $f_{i,j}$ is the forward exchange rate and $s_{i,j}$ is the logarithm of the current spot nominal exchange rate. This is known as the covered interest rate parity (CIP). The uncovered interest rate parity (UIP) is the interest rate parity without a forward contract to hedge exchange rate risk. As “expectations of future exchange rates are not directly observable” in the market (Isard, 1996), UIP operates under the assumption that current forward rate $f_{(i,j),t}$ at time t will equal the expected exchange rate:

$$f_{(i,j),t} = E[s_{(i,j),t+1}]. \tag{B.3}$$

Using (B.3), we can rewrite (B.2) as follows

$$r_{i,t} = r_{j,t} + E[s_{(i,j),t+1}] - s_{(i,j),t}.$$

By rearranging, we obtain the following specification of UIP:

$$\Delta E[s_{(i,j),t+1}] = E[s_{(i,j),t+1}] - s_{(i,j),t} = r_{i,t} - r_{j,t} \tag{B.4}$$

where $\Delta E[s_{(i,j),t+1}]$ can be viewed as the expected rate of depreciation of the domestic currency of a country i.

B.1.3 The Flexible Price Monetary Model

The demand for money is assumed to be dependent on real output, the price level and the level of the nominal interest rate. Accordingly, monetary equilibrium in a country at time t is given by:

$$m_t = p_t + \phi o_t - \lambda r_t,$$

where:

- m_t is the logarithm of the money demand;
- o_t is the logarithm of the real output;
- p_t is the logarithm of the price level;
- r_t is the nominal interest rate;
- ϕ and λ are parameters.

Suppose the parameters in the monetary equilibrium equations of a country i and a country j are equal. Then, by taking the difference of the monetary equilibrium equation of a country i and that of a country j, we obtain a relative money demand function,

$$(m_{i,t} - m_{j,t}) = (p_{i,t} - p_{j,t}) + \phi(o_{i,t} - o_{j,t}) - \lambda(r_{i,t} - r_{j,t}) \tag{B.5}$$

where the variables of a country i and those of a country j are denoted by subscripts i and j respectively. Under the assumption of PPP, we can rewrite (B.5) as

$$s_{(i,j),t} = (m_{i,t} - m_{j,t}) - \phi(o_{i,t} - o_{j,t}) + \lambda(r_{i,t} - r_{j,t}). \tag{B.6}$$

where $s_{(i,j),t}$ is the price of a currency j in a currency i. This equation (B.6) is known as the fundamental flexible price monetary equation and can be expressed with an error term as follows:

$$s_{(i,j),t} = \alpha + \beta_m(m_{i,t} - m_{j,t}) + \beta_o(o_{i,t} - o_{j,t}) + \beta_r(r_{i,t} - r_{j,t}) + \epsilon_t$$

where α is constrained to be 0, β_m is constrained to be 1, β_o is constrained to be negative and β_i is constrained to be positive.

B.1.4 The Sticky Price Monetary Model

The sticky price monetary model extends the flexible price monetary model to allow for sustained inflation differentials across countries. In doing so, it allows for slow domestic price adjustment and consequent deviations from PPP. In other words, it assumes that prices are fixed in the short run, and they adjust slowly towards the long run equilibrium.

The sticky price monetary model assumes that both PPP and UIP conditions hold. By combining the UIP (B.4) and the relative money demand function (B.5) and solving for the relative price levels, we obtain:

$$(p_{i,t} - p_{j,t}) = (m_{i,t} - m_{j,t}) - \phi(o_{i,t} - o_{j,t}) + \lambda(\Delta E[s_{(i,j),t+1}]). \tag{B.7}$$

Then, the PPP assumption gives us that the expected rate of depreciation of the currency i is equal to the expected inflation differential:

$$\Delta E[s_{(i,j),t+1}] = \Delta E[p_{i,t+1}] - \Delta E[p_{j,t+1}].$$

Under the assumption of rational expectations, stable system and exogenous income growth (for simplicity equal to zero, so $o_{i,t} - o_{j,t} = \bar{o}_i - \bar{o}_j$ where $(\bar{o}_i - \bar{o}_j)$ is the long run equilibrium relative money supply), the expected inflation rate is equal to the rationally expected monetary growth rate. A benchmark specification of the money supply process is that monetary growth follows a random walk. Then the rationally expected future relative monetary growth rate is simply the current relative monetary growth rate represented by $(\pi_{i,t} - \pi_{j,t})$. By combining this with (B.7), we obtain the following monetary equation of exchange rate determination:

$$s_{(i,j),t} = (m_{i,t} - m_{j,t}) - \phi(\bar{o}_i - \bar{o}_j) + \lambda(\pi_{i,t} - \pi_{j,t}). \quad \text{(B.8)}$$

The sticky price monetary model further assumes that in the short run, when the exchange rate deviates from its equilibrium value, it is expected to close that gap with a speed adjustment θ. In the long run, when the exchange rate lies on its equilibrium path, it is expected to increase at $(\pi_{i,t} - \pi_{j,t})$. So, the expected change $\Delta E[s_{(i,j),t+1}]$ in exchange rate can be expressed as

$$\Delta E[s_{(i,j),t+1}] = -\theta(s_{(i,j),t} - \bar{s}_{i,j}) + (\pi_{i,t} - \pi_{j,t}) \quad \text{(B.9)}$$

where $\bar{s}_{i,j}$ is the long run equilibrium exchange rate. By combining (B.9) with the UIP condition, we obtain

$$s_{(i,j),t} - \bar{s}_{i,j} = -(1/\theta)[(r_{i,t} - \pi_{i,t}) - (r_{j,t} - \pi_{j,t})]. \quad \text{(B.10)}$$

This equation shows that the gap between the exchange rate and its equilibrium value is proportional to the real interest rate differential. By putting (B.8), which represents the long run monetary equilibrium path, and (B.10), which represents the short run overshooting effect, together and assuming that the current equilibrium money supply levels are given by their current actual levels, we obtain a general monetary equation of exchange rate determination:

$$s_{(i,j),t} = \beta(m_{i,t} - m_{j,t}) - \phi(o_{i,t} - o_{j,t}) + \lambda(\pi_{i,t} - \pi_{j,t}) - (1/\theta)[(r_{i,t} - \pi_{i,t}) - (r_{j,t} - \pi_{j,t})].$$

By rearranging, we obtain:

$$s_{(i,j),t} = \beta(m_{i,t} - m_{j,t}) - \phi(o_{i,t} - o_{j,t}) - (1/\theta)(r_{i,t} - r_{j,t}) + (\lambda + 1/\theta)(\pi_{i,t} - \pi_{j,t}). \quad \text{(B.11)}$$

(B.11) can be expressed with an error term as follows:

$$s_{(i,j),t} = \alpha + \beta_m(m_{i,t} - m_{j,t}) + \beta_o(o_{i,t} - o_{j,t}) + \beta_r(r_{i,t} - r_{j,t}) + \beta_\pi(\pi_{i,t} - \pi_{j,t}) + \epsilon_t \quad \text{(B.12)}$$

where α is constrained to be 0, β_m is constrained to be 1, β_o is constrained to be negative and β_r is constrained to be negative and β_π is constrained to be positive. The major difference between (B.12) and the flexible price monetary model is in the inclusion of inflation rate.

B.1.5 The Sticky Price Asset Model

The sticky price monetary model is extended by the sticky price asset model, which allows for large and sustained changes in the long-run real exchange rate. As the long-run real exchange rate changes are assumed to be correlated with the unanticipated shocks in the current account, the sticky price asset model includes current account balance z as an additional fundamental variable. The formal representation of the model is given by:

$$s_{(i,j),t} = \alpha + \beta_m(m_{i,t} - m_{j,t}) + \beta_o(o_{i,t} - o_{j,t}) + \beta_r(r_{i,t} - r_{j,t}) + \beta_\pi(\pi_{i,t} - \pi_{j,t}) + \beta_z(z_{i,t} - z_{j,t}) + \epsilon_t,$$

where α is constrained to be 0, β_m is constrained to be 1, β_o and β_r are constrained to be negative and β_π and β_z are constrained to be positive.

B.2 Data

The data set used in this study is primarily obtained from the O.E.C.D. database. It consists of:

- quarterly observations of money supply, real output, interest rate, inflation, current account, real effective exchange rate (REER) of the countries in the currency universe;
- monthly observations of nominal exchange rates between the U.S. and the countries in the currency universe; and
- biennial observations of trade weights between the countries in the currency universe with each other,

covering 1994 – 2012. Money supply, real output, REER and current account are seasonally adjusted with a base year of 2000. We should therefore note that our in-sample test uses some data that were not available at the time of the analysis. The below list provides variable descriptions and data sources.

- Money supply (m)
  - Data source: O.E.C.D. main economic indicators
  - Series: Monetary aggregates – M3 broad money
- Real output (o)
  - Data source: O.E.C.D. quarterly national accounts
  - Series: Gross domestic product – expenditure approach
- Interest rates (3 months) (r)
  - Data source: O.E.C.D. monthly monetary and financial Statistics
  - Series: Short-term interest rates, percent per annum
- Inflation (π)

  - Data source: O.E.C.D. key short-term economic indicators
  - Series: Consumer prices – growth on the same period of the previous year (year-on-year changes in consumer prices index)

- Current account (z)
  - Data source: O.E.C.D. main economic indicators
  - Series: Balance of payments – current account balance

- Real effective exchange rate (REER) (E)
  - Data source: J. P. Morgan
  - Series: Trade weighted real exchange rate index

- Nominal exchange rate (S)
  - Data source: Bloomberg market data
  - Series: Spot nominal exchange rate

- Trade weights (w)
  - Data source: Bank for International Settlement (BIS)
  - Series: Trade weights for broad indices consisting of 61 economies

B.3 Supplementary Empirical Results

B.3.1 In-Sample Results

This section provides the results for the in-sample tests of the two DEA model specifications, namely the multiple output DEA specification and the extended multiple output DEA specification, with various window sizes. Table B.1 and Table B.2 present these results.

Table B.1: In-Sample Performance of the Multiple Output DEA Specification with Different Window Sizes

Table B.1 presents performance metrics of the efficiency-based portfolios constructed from the multiple output DEA specifications of the three structural models: the flexible price monetary model, the sticky price monetary model, and the sticky price asset model, with different window sizes.

	Flexible Price Monetary			Sticky Price Monetary			Sticky Price Asset		
Total Return (%)	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
1Q	10.47	11.30	10.23	10.93	11.63	10.03	7.67	8.13	7.61
2Q	12.40	10.96	9.47	11.11	10.69	10.78	10.42	11.05	10.22
4Q	12.49	10.63	9.53	11.70	10.94	10.09	12.35	12.24	10.51
8Q	10.63	10.65	9.48	10.25	10.48	8.99	10.08	10.26	10.55
16Q	11.19	11.70	10.12	10.40	10.37	9.76	10.72	10.36	9.22
Volatility (%)	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
1Q	8.11	8.33	8.13	7.85	8.05	8.05	4.65	4.75	4.80
2Q	8.08	8.19	7.87	7.62	7.89	8.47	5.23	5.47	5.87
4Q	7.00	7.13	6.74	7.21	7.59	7.42	5.68	5.92	5.96
8Q	7.43	7.62	7.72	7.16	7.16	7.40	7.16	7.30	7.26
16Q	7.58	7.80	7.76	7.40	7.58	7.94	7.21	7.32	7.04
Kurtosis	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
1Q	2.24	2.20	2.13	2.14	2.09	2.21	2.14	2.40	2.39
2Q	2.39	2.03	2.19	2.49	2.35	2.39	2.31	2.10	2.38
4Q	2.43	2.42	2.42	2.53	2.49	2.47	2.18	2.14	2.34
8Q	2.86	2.79	2.48	2.54	2.43	2.29	2.68	2.65	2.57
16Q	2.46	2.29	2.00	2.27	2.34	2.13	3.04	2.87	2.22
Skewness	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
1Q	0.11	0.05	0.11	0.11	0.14	0.30	-0.03	0.11	0.14
2Q	−0.01	0.02	−0.01	−0.14	0.07	0.07	−0.21	0.02	0.25
4Q	−0.33	−0.19	0.09	0.12	0.20	0.07	−0.05	0.06	−0.06
8Q	0.06	0.18	0.10	0.12	0.06	0.12	0.17	0.22	0.29
16Q	0.21	0.11	0.22	0.03	0.00	0.06	−0.14	0.00	0.24
Sharpe Ratio	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
1Q	1.04	1.11	1.00	1.13	1.19	0.98	1.20	1.28	1.14
2Q	1.28	1.09	0.93	1.19	1.09	1.02	1.59	1.64	1.38
4Q	1.55	1.41	1.20	1.58	1.72	1.32	1.30	1.37	1.31
8Q	1.15	1.13	0.95	1.14	1.18	0.93	1.12	1.12	1.16
16Q	1.20	1.24	1.03	1.13	1.10	0.96	1.20	1.13	1.01

Table B.2: In-Sample Performance of the Extended Multiple Output DEA Specification with Different Window Sizes

Table B.2 presents performance metrics of the efficiency-based portfolios constructed from the extended multiple output DEA specifications of the three structural models: the flexible price monetary model, the sticky price monetary model, and the sticky price asset model, with different window sizes.

	Flexible Price Monetary			Sticky Price Monetary			Sticky Price Asset		
Total Return (%)	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
1Q	9.03	9.54	7.68	7.61	8.99	8.44	6.67	7.94	7.20
2Q	12.31	11.47	11.47	12.91	11.68	12.14	10.23	9.55	9.88
4Q	12.87	12.14	10.74	12.47	13.18	11.63	10.17	10.96	10.16
8Q	12.24	12.28	11.76	13.44	13.60	13.25	9.97	10.64	9.63
16Q	12.49	12.52	11.21	11.85	12.50	10.78	9.59	9.60	10.21

Volatility (%)	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
1Q	6.79	7.24	7.03	7.03	7.38	7.25	6.24	6.48	6.36
2Q	7.52	8.00	7.94	6.63	7.03	7.46	6.07	6.14	6.07
4Q	6.98	7.17	7.19	6.58	6.47	7.20	6.25	6.52	6.13
8Q	6.92	7.05	7.65	7.24	7.25	7.50	5.44	5.71	5.76
16Q	7.68	7.77	7.97	7.67	7.77	8.03	5.98	6.34	6.31

Kurtosis	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
1Q	2.32	2.24	2.34	2.25	2.20	2.31	2.65	2.50	2.48
2Q	2.28	2.14	2.44	2.26	2.42	2.45	2.56	2.71	2.48
4Q	2.36	2.40	2.47	2.41	2.48	2.69	2.87	2.85	2.48
8Q	2.29	2.48	2.25	2.53	2.49	2.10	2.39	2.43	2.44
16Q	2.23	2.14	2.07	2.23	2.14	2.03	2.33	2.57	2.33

Skewness	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
1Q	0.00	0.05	0.03	0.02	0.06	0.10	0.19	0.20	0.19
2Q	−0.04	0.09	0.33	0.07	0.19	0.3	0.37	0.45	0.35
4Q	−0.03	−0.01	0.13	0.12	0.21	0.22	−0.02	0.02	0.24
8Q	−0.18	0.05	0.12	−0.12	0.05	0.03	0.09	0.21	0.22
16Q	−0.01	0.02	0.06	0.04	0.02	0.08	0.04	0.01	0.12

Sharpe Ratio	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
1Q	1.02	1.03	0.79	0.79	0.94	0.87	0.74	0.91	0.80
2Q	1.36	1.18	1.18	1.63	1.37	1.34	1.34	1.22	1.28
4Q	1.55	1.41	1.20	1.58	1.72	1.32	1.30	1.37	1.31
8Q	1.47	1.45	1.26	1.57	1.59	1.48	1.45	1.50	1.30
16Q	1.35	1.35	1.14	1.27	1.34	1.08	1.26	1.19	1.28

B.3.2 Out-of-Sample Results

This section presents the out-of-sample performance of the benchmarks, against which the efficiency-based portfolios are evaluated. The performance metrics of the two U.S. dollar indices and the equally weighted portfolio that consists of all currencies in the sample are shown in Table B.3, and those of the top and bottom residual-based portfolios are shown in Table B.4. The top and bottom residual-based portfolios are comprised of the currencies ranked in the top 25% and the bottom 25% of the sample respectively.

Table B.3: Out-of-Sample Performance of the Market Benchmarks

Table B.3 presents performance metrics of the two U.S. dollar indices: DXY Index and USTW$ Index and the equally weighted portfolio (EW) that consists of all currencies in the sample for 1-, 3-, and 6-month lagged investments.

Return (%)	1-Month	3-Month	6-Month
EW	6.47	4.92	5.61
DXY	−3.47	−3.00	−2.06
USTW$	−3.71	−3.30	−2.81

Volatility (%)	1-Month	3-Month	6-Month
EW	10.60	10.68	10.76
DXY	9.46	9.93	9.99
USTW$	6.41	6.56	6.55

Kurtosis	1-Month	3-Month	6-Month
EW	3.87	3.66	3.62
DXY	3.26	3.07	2.98
USTW$	3.22	3.05	3.02

Skewness	1-Month	3-Month	6-Month
EW	-0.79	−0.68	−0.70
DXY	0.18	0.27	0.22
USTW$	0.59	0.58	0.52

Sharpe Ratio	1-Month	3-Month	6-Month
EW	0.46	0.33	0.41
DXY	−0.54	−0.45	−0.32
USTW$	−0.83	−0.72	−0.61

Table B.4: Out-of-Sample Performance of the Top and Bottom Residual-based Portfolios

Table B.4 presents performance metrics of the residual-based portfolios for the flexible price monetary model, the sticky price monetary model and the sticky price asset model for 1-, 3-, and 6-month lagged investments.

Panel A: Total Return

	Flexible Price Monetary			Sticky Price Monetary			Sticky Price Asset		
Return (%)	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Top	6.19	5.31	6.09	6.65	5.28	5.93	7.24	4.82	6.26
Bottom	5.46	3.89	5.86	5.70	3.51	5.82	5.49	3.78	4.99
Volatility (%)	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Top	9.91	9.38	9.01	10.03	9.52	9.04	10.27	10.22	9.29
Bottom	11.12	11.22	10.75	11.26	11.32	11.18	11.48	11.40	11.57
Kurtosis	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Top	3.42	3.35	3.18	3.47	3.26	3.17	3.23	3.66	3.02
Bottom	3.79	3.37	3.24	3.49	3.39	3.01	3.42	3.35	3.14
Skewnes	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Top	−0.42	−0.51	−0.37	−0.43	−0.51	−0.34	−0.45	−0.68	−0.27
Bottom	−0.97	−0.60	−0.52	−0.75	−0.65	−0.36	−0.74	−0.68	−0.58
Sharpe Ratio	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Top	0.46	0.41	0.54	0.50	0.40	0.53	0.55	0.33	0.55
Bottom	0.35	0.22	0.44	0.36	0.18	0.41	0.34	0.21	0.33

Panel B: Price Return

	Flexible Price Monetary			Sticky Price Monetary			Sticky Price Asset		
Return (%)	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Top	2.65	1.89	2.69	3.11	1.87	2.55	3.63	1.37	2.80
Bottom	1.13	−0.34	1.66	1.59	−0.40	1.80	1.49	−0.14	1.13
Volatility (%)	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Top	9.93	9.38	9.05	10.06	9.52	9.08	10.29	10.21	9.33
Bottom	11.14	11.24	10.79	11.27	11.34	11.23	11.48	11.37	11.59
Kurtosis	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Top	3.37	3.30	3.17	3.41	3.20	3.16	3.19	3.59	3.00
Bottom	3.75	3.37	3.21	3.45	3.36	3.02	3.42	3.37	3.18
Skewnes	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Top	−0.40	−0.47	−0.35	−0.41	−0.47	−0.31	−0.42	−0.64	−0.24
Bottom	−0.97	−0.62	−0.53	−0.75	−0.63	−0.37	−0.74	−0.69	−0.60
Sharpe Ratio	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month	1-Month	3-Month	6-Month
Top	0.11	0.05	0.17	0.15	0.05	0.15	0.20	−0.01	0.17
Bottom	−0.04	−0.16	0.04	0.00	−0.16	0.05	−0.01	−0.14	−0.01

Appendix C

A Generalized Risk Budgeting Approach to Portfolio Construction

C.1 The Augmented Lagrangian Method

The augmented Lagrangian function was originally introduced by Hestenes (1969) and Powell (1972) and used in the quadratic penalty method for solving an equality constrained optimization problem of the form:

$$\begin{aligned} \min_{x} \quad & F(x) \\ \text{subject to} \quad & h_k(x) = 0, \quad k = 1, \ldots, s. \end{aligned} \tag{C.1}$$

Instead of directly solving (C.1), the quadratic penalty method solves a sequence of problems of the form:

$$\min_{x} \quad \mathcal{L}_{c_t}(x, u_t) = F(x) + u_t' h(x) + \frac{1}{2} c_t \|h(x)\|_2^2 \tag{C.2}$$

where $\{u_t\}$ is a bounded sequence in $\mathbb{R}^s$ and $\{c_t\}$ is a penalty parameter sequence satisfying

$$0 < c_t < c_{t+1} \quad \forall \quad t \text{ and } c_t \to \infty.$$

In the original version of the penalty method the multipliers u_t are set to zero for all t, and the success of the method depends on sequentially increasing the penalty parameter to infinity. It is possible to improve the performance of the method, however, by employing nonzero multipliers u_t and by updating them in an intelligent manner after the time t minimization of $\mathcal{L}_{c_t}(x, u_t)$.

Extensive empirical research has shown that the penalty function method is quite reliable and usually converges to at least a local minimum of the original problem. The failure of this method is usually due to the fact that unconstrained minimization of $\mathcal{L}_{c_t}(x, u_t)$ becomes increasingly ill-conditioned as $c_t \to \infty$. By suitably updating the multipliers, u_t, the difficulties due to ill-conditioning can be significantly mitigated (Bertsekas, 1996).

Suppose then we wish to update u_t after solving (C.2) at time t. We do this by setting

$$u_{t+1} := u_t + c_t h(x_t). \tag{C.3}$$

The initial vector u_0 is chosen arbitrarily and the sequence $\{c_t\}$ may be either preselected or determined adaptively during the algorithm. Note that if $u_t + c_t h(x_t)$ does not belong to a pre-specified bounded open set known to contain u^*, then u_t may be left unchanged in the next time step. From a duality perspective, (C.3) represents a steepest ascent iteration. More specifically, we can define the dual function as

$$\begin{aligned} d_c(u) &= \min_x \mathcal{L}_c(u, x) \\ &= \min_x F(x) + u'h(x) + \frac{1}{2}c\|h(x)\|_2^2. \end{aligned}$$

Noting that $\nabla d_c(u) = h(x)$ we see that (C.3) can be written as

$$u_{t+1} = u_t + c_t \nabla d_c(u_t)$$

which is sometimes referred to as the first order iteration method of multipliers. One could also update u_t using the second order iteration method of multipliers. For instance, one could implement the second order multiplier iteration using Newton's method for maximizing the dual function d_c. This would lead to updates of u_t of the form:

$$u_{t+1} = u_t - [\nabla^2 d_{c_t}(u_t)]^{-1} \nabla d_{c_t}(u_t). \tag{C.4}$$

Note that since the second order methods can be expensive to compute and store the Hessian matrix, (C.4) is not implemented in our numerical experiments.

C.2 Numerical Results for When (i) $u_t = 0$ and (ii) $c_t = 0$ for All t

In this Appendix we present numerical results for (i) the AL-MCMC method but with $u_t = 0$ for all t and (ii) the AL-MCMC method but with $c_t = 0$ for all t. we refer to these restrictions as the Penalty-MCMC and Lagrangian-MCMC (L-MCMC) methods, respectively. We compare their performances against the full AL-MCMC method. For comparison purposes, we use the same starting points for generating Markov chains across all three methods. Table C.2 and Table C.1 report the results for the two scenarios: (i) $\mu = \mu_0 \mathbf{1}$ and (ii) $\mu \neq \mu_0 \mathbf{1}$, respectively.

As can be seen from both tables, AL-MCMC-`fmincon` has a superior and more consistent performance than Penalty-MCMC-`fmincon` and L-MCMC-`fmincon`. For example, the solution ranges of Penalty-MCMC-`fmincon` and L-MCMC-`fmincon` are often very wide. For example, when $\mu \neq \mu_0 \mathbf{1}$, in Set 2 Test Case 4, their solution ranges were [−102.97,

-17.27] and [-102.47, -17.27] respectively while all of the AL-MCMC-`fmincon` solutions have the range of [$-17.28, -17.27$]. Also, in Set 2 Test Case 1 for the same scenario, the best solutions of Penalty-MCMC-`fmincon` and L-MCMC-`fmincon` are 0.62 and 0.17 lower than the best solution of AL-MCMC-`fmincon` respectively. Moreover, while AL-MCMC-`fmincon` had zero failures in all test cases, the other two methods had occasional failures. We can make similar observations when $\mu = \mu_0 \mathbf{1}$.

The results in this Appendix therefore demonstrate the advantage of incorporating the augmented Lagrangian method, as opposed to the penalty method or the Lagrangian multipliers method, into our MCMC algorithm.

Table C.1: Numerical Results for the Case of $\mu = \mu_0 \mathbf{1}$

Table C.1 presents numerical results for the three methods: AL-MCMC-`fmincon`, Penalty-MCMC-`fmincon` and L-MCMC-`fmincon`, when $\mu = \mu_0 \mathbf{1}$. The first and second columns contain the test set number and the name of the algorithm respectively. The third column reports the best solution obtained. The fourth column reports the range of the obtained solutions. The fifth column reports the number of solutions that failed to attain the same value as the best solution $\min \tilde{F}(x^*)$. The final column reports the amount of time taken to obtain the best solution.

Test Case 1: 7 Assets and 3 Subsets

Set	Method	$\min \tilde{F}(x^*)$	Solution range	No. of failures	t (sec)
1	AL-MCMC-`fmincon`	4.69	[4.69, 5.15]	0	2.81
1	Penalty-MCMC-`fmincon`	4.71	[4.71, 5.48]	0	1.26
1	L-MCMC-`fmincon`	4.71	[4.71, 6.18]	0	7.55
2	AL-MCMC-`fmincon`	33.09	[33.09, 34.16]	0	18.50
2	Penalty-MCMC-`fmincon`	33.16	[33.16, 64.81]	0	14.35
2	L-MCMC-`fmincon`	33.10	[33.10, 40.15]	0	14.88
3	AL-MCMC-`fmincon`	25.78	[25.78, 31.91]	0	16.90
3	Penalty-MCMC-`fmincon`	26.64	[26.64, 35.17]	2	9.28
3	L-MCMC-`fmincon`	29.79	[29.79, 31.70]	6	4.61

Test Case 2: 30 Assets and 5 Subsets

Set	Method	$\min \tilde{F}(x^*)$	Solution range	No. of failures	t (sec)
1	AL-MCMC-`fmincon`	38.59	[38.59, 38.59]	0	2.54
1	Penalty-MCMC-`fmincon`	38.59	[38.59, 38.60]	0	2.91
1	L-MCMC-`fmincon`	38.59	[38.59, 38.60]	0	2.41
2	AL-MCMC-`fmincon`	68.54	[68.54, 68.54]	0	1.37
2	Penalty-MCMC-`fmincon`	68.54	[68.54, 68.54]	0	2.30
2	L-MCMC-`fmincon`	68.54	[68.54, 68.54]	0	2.34
3	AL-MCMC-`fmincon`	39.80	[39.80, 39.80]	0	2.32
3	Penalty-MCMC-`fmincon`	39.80	[39.80, 39.80]	0	1.70
3	L-MCMC-`fmincon`	39.80	[39.80, 39.80]	0	9.70

Test Case 3: 50 Assets and 5 Subsets

Set	Method	$\min \tilde{F}(x^*)$	Solution range	No. of failures	t (sec)
1	AL-MCMC-`fmincon`	38.24	[38.24, 38.24]	0	2.68
1	Penalty-MCMC-`fmincon`	38.24	[38.24, 38.24]	1	2.40
1	L-MCMC-`fmincon`	38.24	[38.24, 38.24]	1	3.32
2	AL-MCMC-`fmincon`	52.58	[52.58, 52.58]	0	2.16
2	Penalty-MCMC-`fmincon`	52.58	[52.58, 52.58]	0	2.26
2	L-MCMC-`fmincon`	52.58	[52.58, 52.58]	0	2.28
3	AL-MCMC-`fmincon`	56.24	[56.24, 56.24]	0	2.86
3	Penalty-MCMC-`fmincon`	56.24	[56.24, 56.24]	0	2.60
3	L-MCMC-`fmincon`	56.24	[56.24, 56.24]	0	2.35

Test Case 4: 100 Assets and 10 Subsets

Set	Method	$\min \tilde{F}(x^*)$	Solution range	No. of failures	t (sec)
1	AL-MCMC-`fmincon`	49.21	[49.21, 49.21]	0	8.69
1	Penalty-MCMC-`fmincon`	49.21	[49.21, 49.21]	0	10.13
1	L-MCMC-`fmincon`	49.21	[49.21, 49.23]	2	30.50
2	AL-MCMC-`fmincon`	55.33	[55.33, 103.99]	0	7.98
2	Penalty-MCMC-`fmincon`	55.33	[55.33, 103.07]	0	10.02
2	L-MCMC-`fmincon`	55.33	[55.33, 55.86]	2	13.18
3	AL-MCMC-`fmincon`	50.55	[50.55, 50.55]	0	8.19
3	Penalty-MCMC-`fmincon`	50.55	[50.55, 50.55]	0	9.81
3	L-MCMC-`fmincon`	50.55	[50.55, 50.57]	2	11.59

Table C.2: Numerical Results for the Case of $\mu \neq \mu_0 \mathbf{1}$

Table C.2 presents numerical results for the three methods: AL-MCMC-`fmincon`, Penalty-MCMC-`fmincon` and L-MCMC-`fmincon`, when $\mu \neq \mu_0 \mathbf{1}$. The first and second columns contain the test set number and the name of the algorithm respectively. The third column reports the best solution obtained. The fourth column reports the range of the obtained solutions. The fifth column reports the number of failures. The final column reports the amount of time taken to obtain the best solution.

Test Case 1: 7 Assets and 3 Subsets

Set	Method	$\max F(x^*)$	Solution range	No. of failures	t (sec)
1	AL-MCMC-`fmincon`	0.83	[0.56, 0.83]	0	8.54
1	Penalty-MCMC-`fmincon`	0.81	[0.70, 0.81]	0	1.86
1	L-MCMC-`fmincon`	0.83	[−1.10, 0.83]	0	9.44
2	AL-MCMC-`fmincon`	63.89	[49.42, 63.89]	0	4.48
2	Penalty-MCMC-`fmincon`	63.27	[45.01, 63.27]	0	10.22
2	L-MCMC-`fmincon`	63.72	[54.06, 63.72]	0	21.47
3	AL-MCMC-`fmincon`	30.31	[30.31, 30.31]	0	0.58
3	Penalty-MCMC-`fmincon`	30.31	[30.31, 30.31]	1	0.94
3	L-MCMC-`fmincon`	30.31	[30.31, 30.31]	0	0.76

Test Case 2: 30 Assets and 5 Subsets

Set	Method	$\max F(x^*)$	Solution range	No. of failures	t (sec)
1	AL-MCMC-`fmincon`	−25.33	[−25.33, −25.33]	0	2.37
1	Penalty-MCMC-`fmincon`	−25.33	[−25.33, −25.33]	0	3.44
1	L-MCMC-`fmincon`	−25.33	[−25.33, −25.33]	0	1.90
2	AL-MCMC-`fmincon`	−17.45	[−17.45, −17.45]	0	2.38
2	Penalty-MCMC-`fmincon`	−17.45	[−17.45, −17.45]	0	2.03
2	L-MCMC-`fmincon`	−17.45	[−17.45, −17.45]	0	2.11
3	AL-MCMC-`fmincon`	14.84	[14.83, 14.84]	0	1.92
3	Penalty-MCMC-`fmincon`	14.84	[14.78, 14.84]	0	4.60
3	L-MCMC-`fmincon`	14.84	[14.78, 14.84]	0	2.07

Test Case 3: 50 Assets and 5 Subsets

Set	Method	$\max F(x^*)$	Solution range	No. of failures	t (sec)
1	AL-MCMC-`fmincon`	−0.30	[−0.30, −0.30]	0	2.52
1	Penalty-MCMC-`fmincon`	−0.30	[−0.30, −0.30]	1	2.34
1	L-MCMC-`fmincon`	−0.30	[−0.32, −0.30]	1	2.69
2	AL-MCMC-`fmincon`	17.86	[17.86, 17.86]	0	1.97
2	Penalty-MCMC-`fmincon`	17.86	[17.84, 17.86]	0	2.60
2	L-MCMC-`fmincon`	17.86	[17.86, 17.86]	0	2.44
3	AL-MCMC-`fmincon`	−7.94	[−7.94, −7.94]	0	2.17
3	Penalty-MCMC-`fmincon`	−7.94	[−7.94, −7.94]	1	2.61
3	L-MCMC-`fmincon`	−7.94	[−7.94, −7.94]	0	2.07

Test Case 4: 100 Assets and 10 Subsets

Set	Method	$\max F(x^*)$	Solution range	No. of failures	t (sec)
1	AL-MCMC-`fmincon`	−9.35	[−9.35, −9.35]	0	7.00
1	Penalty-MCMC-`fmincon`	−9.35	[−9.35, −9.35]	0	9.96
1	L-MCMC-`fmincon`	−9.35	[−203.60, −9.35]	2	8.69
2	AL-MCMC-`fmincon`	−17.27	[−17.28, −17.27]	0	7.43
2	Penalty-MCMC-`fmincon`	−17.27	[−102.97, −17.27]	0	7.38
2	L-MCMC-`fmincon`	−17.27	[−102.47, −17.27]	0	9.83
3	AL-MCMC-`fmincon`	0.67	[0.67, 0.67]	0	8.24
3	Penalty-MCMC-`fmincon`	0.67	[0.67, 0.67]	0	9.33
3	L-MCMC-`fmincon`	0.67	[0.56, 0.67]	2	10.55