

Tax Effects on Investments

Shijie Liu

PhD

Heriot-Watt University

"The copyright in this thesis is owned by the author. Any quotation from the thesis or use of any of the information contained in it must acknowledge this thesis as the source of the quotation or information."

Abstract

This doctoral thesis investigates empirically and theoretically the effect of tax on the composition of the optimal allocation of wealth to risky assets from various points of view. The first empirical chapter considers the effect of tax on a U.K. personal investor targeting domestic financial products. This research helps investors estimate the impact of a future tax change and maximize their portfolio return using a newly proposed optimization model and solution method. Following Bonami and Lejeune (2009), personal portfolios are constrained to meet or exceed a prescribed return threshold with a high confidence level and satisfy buy-in threshold and diversification constraints. Their model is improved by incorporating complex tax trading rules with withdrawal features that enhance those considered by Osorio et al. (2004, 2008). A solution based on Greedy methods is developed to deal with the proposed large-scale portfolio optimization problem. The empirical results report substantial non-linear tax effects on riskier assets and enhanced effects of withdrawal tax only when tax rates are high. The developed framework better enables investors to react to tax changes, and tax policy-makers to quantify the influence of tax changes on private investment preferences.

The second empirical chapter investigates the effect of an international transaction tax, the so-called ‘Tobin tax’, from the point of view of U.K., U.S., and E.U. personal investors targeting international financial products. This empirical research helps the policy maker to estimate the impact of Tobin tax on international capital flows and, therefore, assess the optimal way to introduce the new tax. An optimization model is proposed to maximize the expected net Sharpe ratio and find the optimal risky portfolio internationally. Complex trading and tax rules are considered. To examine the precise effects of different investment and transaction tax rules, a comparison of four tax settings is presented: source only, residence only, mixed with credit and mixed with

double taxation. The experimental results show that a source only tax union has more capital transits in international markets than a residence only tax union, and its optimal market portfolio is more sensitive to regional tax policy. In a mixed tax system, double taxation between residence- and source-taxed markets significantly reduces the attraction of the latter while its attraction is maintained with the credit method. Tobin tax can reduce the volatility of the market but the effect varies with tax rate, certain market specifications (e.g., expected returns and correlations with overseas markets) and investment tax rules. It does not depend on which side of the capital flow (inflow or outflow) is subject to Tobin tax. Finally, an agreement among countries to produce a consistent Tobin tax rate globally can significantly reduce the negative effect of Tobin tax on capital flows while retaining its positive effect on market stability in comparison to heterogeneous Tobin tax rates.

Finally, the third analytical chapter investigates theoretically the effect of tax from the point of view of an arbitrageur. This theoretical research addresses the condition of the existence of arbitrage opportunities on an after-tax basis, helping the policy maker improve the fairness and efficiency of markets by addressing effective tax policy. To track tax arbitrage, continuous time optimization models are developed with heterogeneous taxation between investors programmed with continuous rather than static income and capital gains (or losses). It is proved analytically that arbitrage opportunities exist for both perfectly correlated and non-perfectly correlated assets. For perfectly correlated assets, the analysis shows that tax arbitrage may exist, with the investor's top tax rate and some static asset parameters determining the existence of arbitrage opportunities. It is also proved that many of the equilibria obtained under income tax only are not optimal if investors are also subject to capital gains tax. For non-perfectly correlated assets, however, it is the market prices of cap and floor options on asset returns that decide the existence of tax arbitrage. In the government fixed-

income bond market, tax arbitrage between investors is difficult to eliminate unless investors are all subject to the same tax rates. But the return from this arbitrage can be limited if the government applies the same top tax rate to all investors.

Acknowledgement

First and foremost I want to thank my supervisors Prof. Andrew Adams and Dr Boulis Ibrahim. It has been an honor to be their Ph.D. student. I appreciate all their contributions of time and ideas to make my Ph.D. experience productive and stimulating. The joy and enthusiasm they have for their research was contagious and motivational for me, even during tough times in the Ph.D. pursuit.

I gratefully acknowledge the funding sources that made my Ph.D. work possible. I was funded by a James Watt Scholarship (provided by Heriot-Watt University) for all three years of study.

My time at Herriot-watt was made enjoyable in large part due to the many friends that became a part of my life.

Lastly, I would like to thank my family for all their love and encouragement. For my mother and grandmother who raised me with a love of science and supported me in all my pursuits.

Shijie Liu

Heriot-Watt University

August 2016

Declaration

In accordance with the appropriate regulations I hereby submit my thesis and I declare that:

- 1) The thesis embodies the results of my own work and has been composed by myself
- 2) Where appropriate, I have made acknowledgement of the work of others and have made reference to work carried out in collaboration with other persons
- 3) The thesis is the correct version of the thesis for submission and is the same version as any electronic versions submitted.
- 4) My thesis for the award referred to, deposited in the Heriot-Watt University Library, should be made available for loan or photocopying and be available via the Institutional Repository, subject to such conditions as the Librarian may require
- 5) I understand that as a student of the University I am required to abide by the Regulations of the University and to conform to its discipline.

Table of Contents

List of Tables.....	X
List of Figures	XI
Chapter 1 – Introduction.....	1
1.1 Objectives of the Thesis.....	1
1.2 Impact of Investment Taxes on an Individual’s Portfolio.....	2
1.2.1 U.K. tax rules	3
1.2.2 Improvements and contributions	6
1.2.3 Methodology	9
1.3 Impact of Tax on International Investments and Capital Flows.....	10
1.3.1 Market portfolio and mutual fund separation theorem	11
1.3.2 Tax rules	14
1.3.3 Improvements to the global portfolio optimization model	20
1.3.4 Methodology	22
1.4 Impact of Tax on Tax Arbitrage Opportunities.....	23
1.4.1 Tax arbitrage opportunities	25
1.4.2 Improvements to tax arbitrage model	26
1.4.3 Methodology	27
1.5 Structure of the Thesis	28
Chapter 2 - Literature Review	30
2.1 Impact of Tax on Portfolio Optimization	30
2.1.1 Tax constraints and mathematical programming.....	30
2.1.2 Algorithm for MINLP	35
2.2 Impact of Tax on Global Markets	36
2.2.1 General Tax	37
2.2.2 Tobin Tax.....	43
2.3 Tax Arbitrage	45
Chapter 3 - Tax Effects on Investment Portfolios: Large-Scale Optimization under Stochastic and Integer Constraints	49
3.1 Introduction.....	49
3.2 Post-tax Personal Investment Portfolio Optimization	54
3.3 Problem Constraints and Objective Functions	58
3.3.1 Basic trading constraints.....	58
3.3.2 Taxation	60

3.3.3 Estimation (stochastic) risk constraint	64
3.3.4 Single-period MINLP optimization	67
3.4 Solution Method	68
3.4.1 Improved Greedy for portfolio optimization	71
3.4.2 Improved Greedy for infeasibility (Improved Greedy)	74
3.4.3 Improved Greedy with precision (Improved Precision)	78
3.4.4 Computational results	80
3.5 Influence of Taxation on Personal Portfolio Management	87
3.5.1 Personal portfolio: data	87
3.5.2 Personal portfolio: results	88
3.5.3 Personal portfolio: implications	93
3.6 Conclusion	94
Chapter 4 - Quantitative Effects of Differential Tobin and Withholding Tax on Global Financial Markets	96
4.1 Introduction	96
4.2 Capital Asset Pricing Model with Heterogeneous Tax Rules	101
4.3. Tax Rules	102
4.4. Problem Constraints and Objective Functions	104
4.4.1. Basic trading constraints	104
4.4.2. Taxation	107
4.4.3. After-tax market portfolio	112
4.5. Influence of Taxation on Portfolio Management	113
4.5.1. Data, figures and experimental method	114
4.5.2. Residence-based and source-based tax systems	114
4.5.3. Tobin tax	123
4.6. Conclusion	135
Chapter 5 - Dynamic Tax Arbitrage for Perfectly Correlated and Non-perfectly Correlated Assets	138
5.1. Introduction	138
5.2. Tax Arbitrage Optimization	141
5.2.1. Model for Perfectly Correlated Assets	142
5.2.2. Model for Non-perfectly Correlated Assets	149
5.3. Tax Arbitrage for Perfectly Correlated Assets	151
5.3.1. Conditions for the Existence of Arbitrage Opportunities	152
5.3.2. Proof of Proposition 1	156

5.3.3. Further Special Cases	165
5.3.4. Implications.....	166
5.3.5. Equilibrium with Tax Arbitrage	167
5.4. Tax Arbitrage for Non-perfectly Correlated Assets.....	170
5.4.1. Single Investor.....	170
5.4.2. Multiple Investors	178
5.5. Conclusion	183
Chapter 6 – Conclusions.....	186
6.1 Personal Investment Tax in Portfolio Optimization	186
6.2 Tobin Tax in Global Financial Markets	187
6.3 Tax Arbitrage Optimization	189
6.4 Summary.....	191
References.....	193
Appendix A. Propositions and Proofs.....	205
Appendix B. More Detail on Restricted Global, General Global, and Local Tax Arbitrage	208

List of Tables

Table 3. 1 Notation.....	57
Table 3. 2 Main code of algorithm	73
Table 3. 3 Additional code for one-sided infeasibility	76
Table 3. 4 Additional code for two-sided infeasibility	78
Table 3. 5 Additional and replacement code to improve precision.....	80
Table 3. 6 Problem statistics	81
Table 3. 7 Computational results for instances showing an optimality gap.....	86
Table 3. 8 Investment settings for portfolio management.....	88
Table 4. 1 Notation.....	105
Table 5. 1 Notation.....	142
Table 5. 2 Arbitrage for Case One	164

List of Figures

Figure 3. 1 Process of integer allocation	74
Figure 3. 2 Process for one-sided infeasibility (inf: infeasible; fea: feasible).....	76
Figure 3. 3 Process for two-sided infeasibility.....	78
Figure 3. 4 Solution quality: percentage of instances in which a high-quality solution is returned within nine hours	83
Figure 3. 5 Speed and scalability: average computing time for a high-quality solution	85
Figure 3. 6 Influence of tax on equities.....	90
Figure 3. 7 Influence of tax on corporate bonds.....	90
Figure 3. 8 Influence of tax on commodities	90
Figure 3. 9 Difference between new and existing portfolio	91
Figure 3. 10 Influence of equity tax on investment account.....	92
Figure 3. 11 Influence of corporate bond tax on investment account	92
Figure 3. 12 Influence of commodity tax on investment account.....	92
Figure 4. 1 Residence-based vs. Source-based Investment Tax ($a=0.2$)	120
Figure 4. 2 Residence-based Investment Tax ($a=0.2$ or $a=0.4$)	123
Figure 4. 3 Tobin tax in residence-based only ($a=0.2$)	125
Figure 4. 4 Tobin tax in credit method ($a=0.2$)	129
Figure 4. 5 Tobin tax in source-based only ($a=0.2$).....	130
Figure 4. 6 Tobin tax in residence-based only ($a=0.2$ or $a=0.4$).....	133
Figure 4. 7 Tobin tax in residence-based only ($t=0.005$ vs $t=0;0.005;0.01$)	134
Figure 5. 1 Figure 5.1 No Arbitrage Opportunity	174
Figure 5. 2 Local Arbitrage Opportunity	175
Figure 5. 3 Restricted Global Arbitrage.....	176
Figure 5. 4 General Global Arbitrage	177
Figure 5. 5 No Arbitrage by Cooperation	179
Figure 5. 6 Local Arbitrage by Cooperation	180
Figure 5. 7 Restricted Global Arbitrage by Cooperation.....	181
Figure 5. 8 General Global Arbitrage by Cooperation.....	182
Figure 5. 9 No Global Arbitrage by Cooperation.....	183

Chapter 1 – Introduction

This chapter sets out the objectives of the thesis, provides an introduction to portfolio investment and corresponding tax rules, outlines the structure of the thesis and provides an overview of the remaining chapters.

1.1 Objectives of the Thesis

This thesis investigates the impact of taxes on the decision making process of financial investment and implications for governments and investors. Extant research on post-tax portfolio optimization is improved by introducing more tax rules. These rules are programmed as mathematical constraints into a portfolio optimization model. The optimal portfolio on an after-tax basis can then be obtained by solving the model. The quantitative impact of tax is assessed by observing a change of the optimal portfolio by a change of tax constraints. The three main objectives of the thesis are:

- To quantify the impact of investment taxes and corresponding tax benefits from investment bonds in an individual investor's portfolio by improving Osorio et al.'s (2008a) Model (Chapter 3)
- To quantify the impact of Tobin tax and withholding tax on capital flows between regional markets (Chapter 4)
- To quantify the impact of income tax and capital gains tax on single asset prices and arbitrage opportunities by improving Basak and Croitoru's (2001) model (Chapter 5)

The sections below expand on each of the above objectives. Section 1.2 introduces the tax rules on investment bonds, explains why the incorporation of tax constraints is important to portfolio optimization, and details the contribution of the work in Chapter 3. Section 1.3 introduces Tobin tax and withholding tax, explains why

an introduction of Tobin tax could lead to a change of capital flows between regional markets, and shows detailed contributions of the work in Chapter 4. Section 1.4 introduces the subject of tax arbitrage, states why it is important to improve Basak and Croitoru's model, and shows detailed contributions of the work in Chapter 5. Section 1.5 outlines the structure of the thesis.

1.2 Impact of Investment Taxes on an Individual's Portfolio

The first objective of this thesis is to quantify the impact of investment taxes and corresponding tax benefits from investment bonds in an individual's portfolio. Portfolio optimization is the process of choosing the proportions of various assets (e.g. commodities, bonds and equities) to be held in a portfolio, making the portfolio best match an investor's demands according to certain criteria. In modern portfolio theory, developed by Markowitz (1952), the optimal portfolios are obtained with an assumption that an investor wants to maximize a portfolio's expected return contingent on any given amount of risk, with risk measured by the standard deviation of the portfolio's rate of return.

However, in Markowitz's model, the calculation of total return from a portfolio is highly simplified. A lot of trading restrictions and costs are neglected. For example, there is no constraint to include a minimum purchasing amount of securities, annual management fees of investment accounts, transaction fees, and taxation. People try to improve Markowitz's work by including more real trading restrictions into the mean-variance model. According to Kolm et al. (2014), "there are some new trends and developments in the area of portfolio optimization, such as diversification methods, risk-parity portfolios, the mixing of different sources of alpha, and practical multi-period portfolio optimization". Details will be introduced in Chapter 2. The incorporation of trading rules (such as minimum purchasing amount) and trading costs

(such as management fees and transaction costs) has been solved. However, how to programme taxation into an optimization model is still a significant issue. Tax rules are so complex that it is difficult to simulate taxation precisely in a mathematical model. There are many heterogeneous tax rules internationally and across investors that it is challenging to consider all tax rules in one optimization model. Osorio et al. (2008a) discuss the benefits of tax rules for U.K. investment bonds, and their impact on the portfolio choices of U.K. investors. They consider the special tax treatments of investment bonds in a multi-stage portfolio optimization model. The complexity of tax rules along with other trading constraints makes the model hard to solve by extant algorithms, particularly when the number of asset classes considered increases to over 100. In fact, only three high-level asset classes, cash, bonds and equities, are considered by Osorio et.al (2008a). This simplification significantly reduces the model's level of complexity but ignores the correlation between individual bonds and equities. For example, the performance of a bond from an oil producer may be highly correlated with the performance of a share from the same company. Ignoring correlation may lead to sub-optimal allocation across asset classes. To solve this correlation issue, more low-level asset classes and their variance-covariance matrix is included in the proposed optimization model, and a new algorithm needs to be developed to solve the model. This is one of the main contributions of the thesis and will be discussed in detail in Chapter 3.

The following sections introduce relevant U.K. tax rules, explain how Osorio et al.'s (2008a) work can be improved, why the improvement is significant, how the improvement is made and the expected results from the improved model.

1.2.1 U.K. tax rules

According to Osorio et al. (2008a), the optimal portfolio is highly sensitive to tax. Consequently, tax constraints should be considered in a portfolio optimization exercise. However, effective tax rates vary with investors' conditions and products that investors purchase, and tax constraints largely increase the complexity of a portfolio optimization model. As a result, some level of abstraction is needed in progressing meaningfully with plausible analyses. Prior to setting out the abstraction, however, taxes on financial investments in the U.K. are introduced in detail. The information that follows relies heavily on the official U.K. government website (www.gov.uk)¹.

The total return from a financial investment is usually composed of income and capital gains. Investment income includes interest payments and dividends, and all investment income is subject to income tax. In general, however, most people earn a large portion of their total net income through employment. Capital gain is defined as the increase in the value of an asset above its purchase price. A capital loss is the decrease in the value of an asset below its purchase price. In the U.K., realised capital losses can be used to reduce the capital gains tax payment in the same tax year. If an investor's total taxable gains, following this reduction, are still above the tax-free allowance, this investor can deduct unused losses from previous tax years. If the losses reduce an investor's gain to the tax-free allowance, the investor can carry forward the remaining losses to a future tax year.

In addition, in financial markets, most investments are subject to an annual income tax payment while only capital gains tax can be deferred until the disposal of the assets. However, some insurance products, such as investment bonds, are subject to more complicated tax treatments whereby income tax can be deferred as well. Consequently, investments of these products may be more sensitive to tax constraints in the optimization process.

¹ This website introduces tax rules in the UK.

An investment bond is generally a single premium life insurance policy. However, it is an investment rather than insurance in the general sense. An insurance company will take the premium and invest it for income or capital gains which accrue until an investment bond holder withdraws money from the policy.² As the holder does not receive income from the policy, personal income tax is deferred. Thus, an investment bond is a potentially tax-efficient way of holding a range of investment funds in one portfolio. These funds offer a range of asset classes, helping investors diversify and to build their own portfolio indirectly.

Investment bonds are of two main types, on-shore and off-shore. Onshore bonds are policies whose funds are subject to U.K. tax. Corporation tax is payable at 20% on most of the income of onshore bonds. This part of tax is paid annually, and the payment of this tax is equivalent to investors having paid basic rate income and capital gains tax, so investors have no personal liability to basic rate income tax and capital gains tax on the proceeds from the bond.³ However, a liability to income tax and capital gains tax above the basic rate may arise if the bond is disposed of, or withdrawals are made.

Regarding withdrawals, there is a 5% annual allowance. If one withdraws more than 5% per policy year of the amount that one has paid into the investment bond, a tax payment is required on the excess amount. Otherwise, if one withdraws less than 5% per policy year of the amount that has been paid into the investment bond, the remaining withdrawal allowance is cumulative, and can be carried forward to future years, subject to the total cumulative 5% allowance amount not exceeding 100% of the amount paid into the investment bond. Tax rules on withdrawals from investment bonds are complicated. First, any tax liability on withdrawals is calculated on the amount withdrawn in excess of the accumulated 5% allowances. The gain in this excess is then

² For example, Prudential UK, one of U.K. largest insurance companies, founded in 1848.

³ In the U.K., tax is payable at the basic rate of 20 per cent on taxable income up to £31,785 (2015-16 tax year).

taxed if it falls into the higher rate (and the additional rate, where applicable) tax bracket when added to investors' taxable income for the tax year.⁴ In the proposed model of Chapter 3, withdrawal taxes are considered when calculating the total tax liability of the portfolio.

Offshore bonds are issued from tax havens outside the U.K. There is little or no tax charged on the insurance company's funds. Most offshore bonds offered to the U.K. investors are based in the Isle of Man, Dublin, Luxembourg or the Channel Islands. The income and capital gains of such a fund will normally be free of tax in the relevant jurisdiction. Hence, they are often referred to as 'gross roll-up'. Gross roll-up is actually something of an illusion for offshore bonds. This is because, the fund is likely to incur some withholding taxes on its underlying investments even if there may be no tax in the particular tax haven in which the insurance company is based. In other words, when investors dispose of all their off-shore bonds, cumulated income tax and capital gains tax have to be paid at the end. This special tax treatment on offshore bonds is also considered in the model of Chapter 3.

1.2.2 Improvements and contributions

As discussed in Section 1.2.1, tax constraints may have a significant impact on optimal portfolios and, hence, should be included in the optimization model. Besides tax, there are many other constraints that need to be considered. All trade orders should have a minimum purchasing amount, such as 1 share or 100 shares. As a result, in the optimization model, a binary variable should be used to make sure that if an asset is purchased, a minimum purchasing amount is applied. Similarly, investment risk,

⁴ If an investor has taxable income of more than £31,785, he/she will have to pay the higher rate of 40 per cent tax on the amount above £31,785 up to £150,000. If the investor have taxable income of more than £150,000, he/she will have to pay the additional rate of 45 per cent tax on the amount above this level

diversification requirements, management fees, and transaction costs should also be considered in the model. Details of these trading constraints are introduced in Chapter 3.

In Osorio et al.'s (2008) work, taxation of investment bonds in the U.K. is discussed. Income tax and capital gains tax are calculated to get the net portfolio return. The annual withdrawal allowance is also considered together with other trading rules, such as management fees and transaction costs. Mixed integer non-linear programming (MINLP) techniques are used to incorporate these market constraints. There are some extant algorithms able to solve it, but they lack efficiency and the precision of the optimal solution cannot be guaranteed when a large number of integer variables and non-linear constraints are introduced. As mentioned in Section 1.2.1, Osorio et al. (2008) deal with this issue by reducing the number of asset classes considered. Only 3 high-level asset classes (cash, bonds and equities) are used but the portfolio obtained from the model may be sub-optimal because correlations between low-level asset classes are ignored.

One of the main objectives of this thesis is to improve Osorio et al.'s model by developing a more advanced algorithm that allows for the inclusion of more low-level asset classes (up to 288). Also, their model is improved further by incorporating estimation risk and introducing more realistic trading rules.

Risk is a key feature of portfolio optimization. Many approaches have been developed to measure risk and uncertainty. Goldfarb and Iyengar (2003), for example, propose a robust factor model to manage risk. Other authors use historical asset return data to represent future risk (Bodnar and Schmid, 2007; Bonami and Lejeune, 2009; Lejeune, 2010) or assume that asset returns follow a normal distribution (Bodnar and Schmid, 2007). In Chapter 3, the approach used to measure risk falls within the Markowitz mean-variance framework. The classic Markowitz framework, however, assumes perfect knowledge of the expected returns of the assets and the variance-

covariance matrix. It assumes that there is no estimation error in both the expected returns and the variance of and covariance between assets. However, obtaining accurate estimates of these measures is difficult. There are many sources of estimation error in the process. For example, it may be impossible to obtain sufficient data samples, data may be unstable, and estimates of future returns will vary across investors (Mulvey and Erkan, 2003). This leads to the so-called ‘estimation risk’ (Bawa et al. 1979). It has been shown that estimation risk can lead to incorrect decisions on the composition of optimal portfolios (see, e.g., Ceria and Stubbs, 2006, and Cornuejols and Tütüncü, 2007). Very small differences in the value of measures can change the composition of portfolios significantly. Broadie (1993) and Chopra and Ziemba (2011) show that portfolio estimation risk is due mainly to errors in the estimation of expected returns than in the estimation of the variance-covariance matrix (Ceria and Stubbs 2006). Accordingly, one focus of Chapter 3 is on estimation risk of expected returns (Bonami and Lejeune, 2009) rather than the variance-covariance matrix (Lejeune and Samatlı-Pac, 2012). Since the algorithm proposed by Lejeune and Samatlı-Pac (2012) is tailored to the consideration of estimation risk of the variance-covariance matrix, the focus of the thesis is on the estimation risk of expected returns which makes their algorithm unsuitable for the proposed problem. A new algorithm is developed, instead, based on the work of Bonami and Lejeune (2009). This issue will be discussed further in Chapter 3.

The estimation risk of expected returns has attracted renewed interest in recent years, and several approaches to incorporate it into portfolio selection have been developed. We use Roy’s safety first risk criterion (Roy, 1959) that identifies as optimal the portfolio for which the probability of its return falling below a prescribed threshold is minimized to assess the estimation risk of expected returns. The constraint ensures that the total expected return exceeds the prescribed minimal level with a minimal

probability. In this thesis, stochastic constraints are used to introduce estimation risk. More detail regarding estimation risk will be introduced in Chapter 3.

1.2.3 Methodology

One of the main contributions of the thesis is the development of an efficient algorithm to solve large scale portfolio optimization with integer and non-linear constraints (MINLP). The covariance between low-level asset classes is also considered.

In the literature, algorithms for solving MINLP problems are often based on relaxation schemes. For the standard mean-variance portfolio problem, different approaches based on nonlinear Branch and Bound (B&B) algorithms (Bonami et al., 2008; Bonami and Lejeune, 2009) and outer approximations (Lejeune and Samatlı-Pa , 2012) have been discussed. However, the increased complexity of the proposed optimization problems due to the inclusion of taxes, probabilistic returns and the increased number of assets limits the use of these proposed algorithms. The tax withdrawal rules necessitate the re-evaluation of the entire objective function and constraints every instance in which the control variables are perturbed or integrality restrictions on the integer variables restored. In Bonami and Lejeune (2009), portfolio variance is the objective function, and their integer variable scoring process depends on a function of the specific contribution of each variable to the overall risk of the portfolio. This is estimated through the Lagrangian function, and for their simple variance-minimising objective function there is a direct link (mapping) between the control variables (asset weights) and the objective function (portfolio variance). This allows them to calculate the effect of a small change to the control variables with two simple equations. In the problems considered in this thesis, however, this is not possible since a change in the control variables does not map directly to the objective function, which is net-of-tax total returns. This is because changes in asset weights lead to complex

changes in tax and this in turn changes the total post-tax return used in the objective function.

In Chapter 3, a solution method for the proposed post-tax portfolio optimization problem is developed based on a Greedy algorithm, implemented with a new dynamic ranking procedure. The algorithm is used to find the optimal portfolios under various tax settings. The quantitative impact of tax on an individual's portfolio can be observed from the change of the optimal portfolio as various tax constraints are applied.

1.3 Impact of Tax on International Investments and Capital Flows

The second objective of this thesis is to quantify the impact of Tobin tax and withholding tax on capital flows between regional markets. Tobin tax is a type of transaction cost. It is still just an idea and not applied yet in any country in the world. In concept, if a Tobin tax is introduced, investors will be subject to such a tax if, and only if, capital is transferred from one country to another. This tax is different from income tax and capital gains tax. The payment is based on the total amount of capital that is transferred rather than the return on investments. In addition, withholding tax is specially designed for international investments. Returns from international investments may be subject to withholding tax. But this tax is calculated differently across countries. More detail regarding Tobin tax and withholding tax are introduced in Section 1.3.2.

The tax on an international investment is more complicated than the tax on a local investment. The tax payment on an international investment may include not only income tax and capital gains tax but also Tobin tax and withholding tax. These types of taxes may have a significant influence on investors' decision-making. The impact of Tobin tax on international investments has been investigated in the literature (more detail is given in Chapter 2). However, in most cases, the investigations ignore other types of tax, such as income tax, capital gains tax, or withholding tax. The optimal

portfolio, obtained while considering Tobin tax only and ignoring other major taxes, may be sub-optimal. This issue is investigated in detail in Chapter 4.

Section 1.3.1 introduces the concept of the market portfolio and the mutual fund separation theorem. Section 1.3.2 outlines the relevant tax rules in the U.K., U.S., and Eurozone. These tax rules include those relating to income tax, capital gains tax, Tobin tax and withholding taxes. Section 1.3.3 outlines the main contributions of the Chapter 4 and why they are significant. Section 1.3.4 introduces the methodology used to achieve the objective.

1.3.1 Market portfolio and mutual fund separation theorem

The capital asset pricing model (CAPM) was developed from the work of Markowitz (1952). It is used to determine a theoretically appropriate required rate of return for an asset. The theoretical price can then be compared with the market price to judge if the asset is relatively over-priced or under-priced.

In the CAPM, there is an important input that needs to be estimated - the expected return of the market portfolio. In practice, it is approximated by the rate of return of a market index (e.g. FTSE 100, S&P 500). In theory, the market portfolio is the portfolio that provides the highest excess return per unit of risk (usually referred to as the Sharpe Ratio). Given a required rate of return, a combination of the risk-free asset and the market portfolio will usually have a lower variance than investing in any risky portfolio on the minimum-variance frontier (apart from the market portfolio itself). The capital market line (CML) is derived by drawing a line from the risk-free rate of return tangent to the minimum-variance frontier of risky assets. The CML is superior to the minimum-variance frontier since it takes into account the risk-free asset in the portfolio. In addition, for a given risk-free rate, there is only one optimal portfolio (the tangent point on the minimum-variance frontier with CML) which can be combined with the

risk-free asset to achieve the lowest level of risk for a given required return. This is the market portfolio.

All rational investors can obtain a particular level of expected return efficiently by holding a linear combination of the risk-free asset and the market portfolio. Consequently, the market portfolio, in theory, is a portfolio containing the equilibrium weight of each risky asset in the market. The market value of an asset is simply equal to the asset's weight in the market portfolio multiplied by the sum of the aggregate market values of all assets.

If all resources are allocated efficiently and there is no foreign exchange risk, in equilibrium every asset should have only one price across regional markets. Otherwise an arbitrage opportunity will arise by buying the asset at a lower price in one country while selling it at a higher price in another country. As a result, the same required rate of return should be given by the CAPM for any single asset⁵. To get this required rate of return, an expected rate of market return is needed in the CAPM. However, this expected rate of market return is difficult to calculate after considering taxation. On an after-tax basis, the expected rate of market return varies across investors as investors are subject to different tax rates, leading to heterogeneous required rates of return even for the same asset. As a result, to get the right required rate of return for an asset on an after-tax basis, the right expected rate of market return must be obtained.

The standard mutual fund separation (Fisher's separation) theorem without tax states that, under certain conditions, the optimal portfolio for each investor can be constructed by holding each of certain mutual funds in appropriate ratios, where the number of mutual funds is smaller than the number of individual assets

⁵ A single asset refers to assets whose values are derived from the same resources and should be equal all the time even if they are traded in different regional markets. For example, the value of 1% share of Apple traded in the U.K. market should always be equal to the value of 1% share of Apple traded in the U.S. market (assuming there is no foreign exchange risk). So these two assets are the same and can be treated as a single asset.

in the portfolio. The selected mutual funds will be the same across investors, but the ratios held by each investor may vary. In standard portfolio theory the market portfolio and the risk-free asset are the two separating mutual funds, and investors hold proportions (either positive for going long or negative for going short). In this setup, the ratio of risky assets to the risk-free asset determines the overall return, and this relationship is clearly linear.

In the presence of taxes, Trauring (1979) develops a three-fund separation theorem leading to a three-term Capital Asset Pricing formula. He shows that an investor's optimal allocation is a linear combination of three identified risky portfolios, G, D, and E (not defined in detail in Trauring's paper), which are independent of investors, their utility function, and tax brackets. This independence does not, however, extend to the weights for each identified risky portfolio. These weights are functions of investor i 's income tax and capital gains tax brackets, and will be denoted by a_i , b_i , and c_i . As a result, investors with different tax brackets will hold different optimal risky portfolios, X_i . Trauring presents the following formula to calculate the market portfolio with tax:

$$\sum_{i=1}^n X_i = M = \left(\sum_{i=1}^n a_i\right) * G + \left(\sum_{i=1}^n b_i\right) * D + \left(\sum_{i=1}^n c_i\right) * E \quad (1.1)$$

where, n is number of investors in the market, X_i is the optimal portfolio of each investor i , and M is the so-called market portfolio. He further proves that the capitalization-weighted sum of all investors' optimal risky portfolios provides the market equilibrium condition. In Trauring (1979), however, the taxation process is still highly simplified. An asset's net return is calculated as $(1-t)*r$ where only fixed tax rates, t , are considered while more complex tax rules such as annual income tax, deferred capital gains tax, withholding tax on foreign investments, and transaction tax are not considered.

In Chapter 4, Trauring's work is extended by introducing more realistic tax conditions (e.g. withdrawal tax and Tobin tax) to the basic capital gains tax and income tax. Consequently, the proposed model can be used to investigate the impact of withholding tax and Tobin tax on international investments. Only the variation of income tax rate and capital gains tax rate across individual investors is considered in Trauring's work. Although the variation of withholding tax rules across countries can be reflected in the effective income tax rate and capital gains tax rate, his model is unable to show a direct link between withholding tax and basic investment tax. In other words, Trauring's model cannot show how withholding tax rules change the effective income tax and capital gains tax payments and, consequently, investment decisions. This issue is investigated in Chapter 4 where detailed withholding tax rules are discussed and incorporated in the proposed optimization model. The impact of withholding tax on investors' decision-making process and, consequently, on international capital flows is tested by observing the behaviour of the optimal portfolio under different withholding tax constraints. Tobin tax, a transaction tax for foreign investments, is also discussed in the model proposed in Chapter 4, while it does not feature in Trauring's work. Withholding tax and Tobin tax are now introduced in more detail in Section 1.3.2.

1.3.2 Tax rules

In addition to the U.K. tax rules which have been discussed in Section 1.2.1, tax rules in other countries and their impact on international financial markets are also investigated in this thesis. In this section, income tax and capital gains tax in the U.S. and the Eurozone, as well as Tobin tax and withholding tax in the U.K., the U.S. and the Eurozone are introduced.

a. Income tax and capital gains tax rules in the U.S. and the Eurozone

In the U.S., a tax is imposed on income by the Federal, most state, and many local governments. This tax payment is citizenship-based, and all U.S. citizens, regardless of place of residence, are obliged to report income and pay income tax every year. The rate of this income tax may increase as income increases. To present a manageable yet realistic model in Chapter 4, an effective tax rate is assumed and used to calculate effective tax payments. Taxable income is defined as total income less allowable deductions and although income is broadly defined, in this thesis the term ‘income’ is taken as that which is received on financial asset investments. Further, it is assumed that most investors would have already used up their allowable deductions before paying tax on investment income and, hence, all investors’ allowable deductions are considered as zero.

As in the U.K., capital gains in the U.S. are also taxable, and capital losses are allowed to reduce taxable capital gains. Investors must self-assess their income tax by filing tax returns. Due dates and other administrative procedures vary by jurisdiction. Typically, April 15 is the last day for individuals to file tax returns for Federal and many state and local returns. Tax as determined by the investor may be adjusted by the taxing jurisdiction. However, capital gains tax is paid only when realised, which usually occurs when the assets are sold. These general rules are broadly the same as those in the U.K.

In the Eurozone, tax is also charged on both income and capital gains. The Eurozone, officially called the Euro area, is a monetary union of 19 of the 28 E.U. members that use the euro as their common currency for local business activities. The European Central Bank (ECB) sets the monetary policy of the zone. It is governed by a president and all the heads of national central banks in the E.U.. One main task of the ECB is “to keep inflation under control. [Although] there is no common representation, governance or fiscal policy for the currency union, some co-operation does take place

through the Euro Group, which makes political decisions regarding the Eurozone and the Euro. The Euro Group is composed of the finance ministers of [the] Eurozone states, [but] in emergencies, national leaders also form the Euro Group”. As a result, although countries in the Eurozone have independent tax policies, co-operation on tax policy-making exists between Eurozone governments. In the discussion of Tobin tax, introduced later in this section, it is assumed that the Eurozone applies the same basic income and capital gains tax, and capital transactions within the Eurozone are not subject to a Tobin tax charge. In the Eurozone, Germany and France are two countries leading the group. Consequently, understanding their tax policy is important.

In Germany, all investment income of residents is subject to income tax but unlike the U.S., income tax is residence-based. Income is reported and taxed annually. “In January 2009, [however], Germany introduced a very strict capital gains tax (called ‘Abgeltungsteuer’) for shares, funds, certificates, [and bank interest]. Capital gains tax only applies to financial instruments (shares, bonds... etc.) that have been bought after 31 December 2008. Instruments bought before this date are exempt from capital gains tax (assuming that they have been held for at least 12 months), even if they are sold in 2009 or later, barring a change of law.” In addition, there is a tax-free allowance (‘Freistellungsauftrag’) on capital gains in Germany of €801 per person per year. Capital gains tax is due only when the asset is sold and the gains realized.

In France, capital gains on the sale of financial instruments (shares, bonds, and other financial products) are taxed at the marginal tax rate (up to 45%), plus 15.5% of social contributions (i.e. up to 60.5%). A deduction of 20% to 40% on the gross capital gain can be applied if the instrument has been held for at least 2 years. In addition, if a special account, the so-called PEA, is used to purchase shares, the capital gains are subject to social security taxes only. This is the case only if the PEA is held for more

than five years. There is a maximum amount, €152,000 that can be deposited in the PEA. As far as income tax is concerned, it is also residence-based and paid annually.

In Chapter 4, the tax rules of Germany and France are simplified and used as tax constraints to calculate tax liability for investments in the Eurozone. This is explained further in Chapter 4.

b. Withholding tax rules

Beside local investment taxes, such as basic income tax and capital gains tax, there are other types of tax to consider in respect of international investment. In Chapter 4, withholding tax and Tobin tax, as well as income tax and capital gains tax, are considered in examining the impact of tax on international capital flows.

Withholding tax is a type of levy deducted at source by some countries on income paid to foreign investors. In most regions, employment income is subject to withholding tax. Furthermore, payments of interest or dividends, sometimes even royalties, rental income and the sale of real estate are subject to withholding tax in many regions. International withholding tax rules vary with investors and countries. Many governments impose a levy on income from foreign investments, but there are no broad general rules for the taxing method. Existing variations range from double taxation (meaning that the same income is taxed by more than one country) to no tax at all. To avoid double taxation, countries that tax income generally use one of two systems: source-based or residence-based. In the source-based system, only local income, which is income from a source inside the country, is taxed. In the residence-based system, residents of the country are taxed on their worldwide (local and foreign) income, while non-residents are taxed only on their local income.

A residence-based tax system is often justified on the grounds that people and firms should contribute from all their income regardless of where it is earned, towards

the public services provided by the country where they live. Consequently, if all countries adopt the residence-based tax system (i.e. a residence-based tax union), the tax rate on an investment would not only be dependent on the asset and the country of investment, but also on the country of residence. All foreign investments are taxed firstly by the government in the country of investment and then by the government in the country of residence. Many countries, however, have tax treaties with each other to allow credits for the tax that residents have already paid to other countries on their foreign income. In addition, the residence-based tax system faces the daunting task of defining "residence" and characterizing the income of non-residents. Such definitions vary by country and investor, but usually involve the location of the investor's main residence/home and number of days the investor is physically present in the country.

In contrast to the residence-based tax system, a source-based tax system is usually justified on the grounds that the country which provides the opportunity to generate income or profits should have the right to tax it. Consequently, if all countries adopt the source-based tax system (i.e., a source-based tax union), all investors, regardless of their residence, would be subject to the same tax rules internationally. In other words, the tax rate on an investment would be dependent only on the asset and the country of investment (e.g., U.S. equities or Eurozone bonds), but not on the country of residence.

As discussed above, in a source-based tax union, tax payments are dependent on the country of investment only. In a residence-based tax union, however, tax payments are dependent on both the country of investment and the country of residence. It would be easy to model withholding tax mathematically if all countries are in one tax union, whether residence-based or source-based. In reality, however, this is not the case. International tax environment can be mixed where one country might adopt a residence-based tax system and another adopts a source-based tax system. Including different

withholding tax regimes in a single model increases the level of complexity. To deal with this issue, two different methods are used to calculate the tax payment, the double taxation method and the credit method.

In the double taxation method, tax is charged by two or more jurisdictions on the same investment. This method reduces the net return from foreign investments and, consequently, the investors' motivation to invest in foreign markets. In order to mitigate the burden of double taxation, many countries provide for tax relief on foreign investments. This is achieved using the credit method.

The credit method is used when two countries are in a tax union. For investors resident in the country with the residence-based tax system and investing in the country with the source-based tax system, the effective tax rate on the investment is the higher rate of the two countries. The double taxation method is used if the two countries are not in a tax union. For investors resident in the country with the residence-based tax system and investing in the country with the source-based tax system, the investment will be taxed by both governments. For investors resident in the country with a source-based tax system and investing in the country with a residence-based tax system, the investment will be taxed only by the government of the latter. Both methods will be incorporated into the optimization model to investigate the differential impact on international financial investments.

It follows that the optimal portfolio across countries using different withholding tax regimes may be sensitive to the local tax policy. In order to find an optimal portfolio for an international investor, it is necessary to take withholding tax into consideration in the optimization model. In Chapter 4, heterogeneous withholding tax is programmed as mathematical constraints in the proposed model to investigate its impact on international capital flows.

c. Tobin tax

In addition to withholding tax, there have been discussions on the introduction of a transaction tax on international capital movements and investments, the so-called Tobin Tax. The concept was introduced by James Tobin in the 1970s, and was initially defined as a tax on all spot conversions between two currencies, with the intention of imposing a penalty on short-term financial round-trip excursions into a foreign currency. As such, it was initially defined as a currency transaction tax levied only when transfers are made from one currency to another. The concept has since been extended to an international Tobin-style transaction tax on general wealth transfers across countries (or union, such as the Eurozone). This type of tax would be dependent on the size of the transaction. The concept has recently gained support among European governments. In 2013, the European Commission announced that a tax on financial transactions out of or into eleven EU countries would be introduced in 2014. The proposal was favoured by 11 EU members and passed in the European Parliament in 2012. It was then approved by the Council of the EU in 2013. The approval of the formal agreement on the details of the European Union financial transaction tax (EU FTT) by the European Parliament is still pending.

1.3.3 Improvements to the global portfolio optimization model

The main contribution of Chapter 4 is to extend Hanke et al.'s (2010) research by combining Tobin tax with investment taxes and withholding tax in a single optimization model.

The incorporation of Tobin tax may have a significant impact on investment decisions. It is expected to increase the cost of foreign investment, and will therefore decrease the incentive of a local investor to transfer money to a foreign market in search of higher asset returns. This impact may be large or small depending on the effective

rate of Tobin tax and how it is charged (e.g., only capital outflow to foreign markets is subject to tax, or only capital inflow to the local market is subject to tax). In Chapter 4, the main objective is to quantify this impact under different tax settings. As discussed in Section 1.3.2, Hanke et al. (2010) show that the impact of Tobin tax on trading volume depends on the size of the market. Accordingly, if Tobin tax affects trading volume and, hence, initiates capital flows, and since these are subject to local income tax and capital gains tax, then it is reasonable to expect interaction effects and sensitivity to income tax and capital gains tax. To investigate this sensitivity, it is necessary to include both investment tax (income tax and capital gains tax) constraints and withholding tax constraints along with Tobin tax constraints in a single mean-variance model. Obviously, this increases the level of complexity.

Withholding tax may also have an impact on international capital flows. Residence-based or source-based tax will change tax payments, leading to different net returns from invested assets, thereby changing investors' preferences for foreign investments. In Chapter 4, this impact is quantified by studying the behaviour of the optimal portfolio under different withholding tax settings. In addition, a further comparison between credit and double taxation methods can help us better understand the importance of an international tax union. Accordingly, a further objective of Chapter 4 is to investigate whether the introduction of the credit method improves capital flow between regional markets, and whether or not the double taxation method has an obvious lock-in effect on local capital.

An aggregated market portfolio can be obtained from the model by using the methodology introduced in Section 1.3.4. The impact of Tobin tax and withholding tax on international investments and global capital flows can then be tested by observing the aggregate market portfolio under different tax settings. With regard to withholding tax, initially it is assumed that all three regions, the U.K., the U.S., and the Eurozone,

are in a residence-based tax union. By changing the rate of income tax and capital gains tax in each country, the effect of tax on the aggregate market portfolio can be observed. The three regions' withholding tax rules are then changed from the residence-based to the source-based, one by one, and the impact is investigated. The impact on the aggregated market portfolio is also investigated for the same ranges of the rate of income tax and capital gains tax, and compared to previous observations.

With regard to Tobin tax, the tax rate is initially set to be 0. The aggregated market portfolio is observed under different rates of income tax and capital gains tax in each country. The rate of Tobin tax in each country is then increased incrementally and the differential impact on the aggregated market portfolio is noted under the same range of the rate of income tax and capital gains tax. The impact of withholding tax and Tobin tax can be quantified from these comparisons. From the experimental results, implications of the introduction of Tobin tax and the choice of withholding tax method can be obtained which should be of interest to governments.

1.3.4 Methodology

In Chapter 4, an improved mean-variance model that considers heterogeneous tax rules across investors is proposed. In addition, it is assumed that there are three types of investors: residents of the U.K., the Eurozone and the U.S. To pay more attention on heterogeneous tax across regional markets, all investors within the same country of residence are assumed to be subject to the same tax rules and therefore should hold the same optimal risky portfolio. This risky portfolio can be combined with the risk-free asset to maximize individual utility. The aggregated market portfolio under taxation is just the capitalization-weighted sum of the optimal risky portfolios for each type of investor.

$$Portfolio_{Market} = \frac{\sum Capitalization_i \times Portfolio_i}{\sum Capitalization_i} \quad (1.2)$$

As in Trauring (1979), the market portfolio with taxation is not a unique portfolio that all investors would hold. Nevertheless, it can provide the theoretical equilibrium of international markets if all investors are included. In other words, it can provide the percentage of a certain asset class that it should be included in the whole international market.

In Chapter 4, the aggregate market portfolio is obtained from the newly developed mean-variance model. The tax effect on the composition of the market portfolio and corresponding capital flows are investigated under different tax settings (Tobin tax and withholding tax on international investments).

1.4 Impact of Tax on Tax Arbitrage Opportunities

As discussed in Sections 1.2 and 1.3, tax constraints may have a significant impact on both individual investment decisions and international capital flows. However, in those two sections, the discussion is confined to the portfolio level. The impact of tax on a single asset (such as how heterogeneous tax rules are reflected in a unique fair price of an asset, and how an investor takes advantage of tax to secure an arbitrage opportunity on an after-tax basis) should also be discussed. This is the main objective and contribution of the analysis of Chapter 5.

One of the assumptions of the traditional CAPM, reviewed in Section 1.3, is that there is no tax. All prices obtained are on a before-tax basis. Attempts to extend the model by including simplified income tax were made (e.g., Elton and Gruber, 1978). This earlier work shows that the introduction of taxes will change an asset's fair price significantly. The fair price for a single asset, on a before-tax basis, varies across investors and countries because of heterogeneous tax rules. Having more than one fair

price for the same asset may lead to an arbitrage opportunity across regional markets. This is usually called ‘tax arbitrage’. In practice, however, tax rules are so complicated that it is hard to comprehensively include all taxes in a single pricing model. Accordingly, even if a tax arbitrage opportunity exists, it would be difficult to detect and take advantage of. Basak and Croitoru (2001) attempt to overcome this issue by introducing a highly generalized tax function to represent the complicated taxation process. The fair price of an asset on an after-tax basis is then discussed, in theory, through a theoretical model that they propose.

The main objective of the analysis in Chapter 5 is to extend the Basak and Croitoru (2001) model to a theoretical discussion on the existence of tax arbitrage. In Chapter 5, tax rules, asset prices, and the correlation between two assets are considered to formulate the conditions necessary for the existence of a tax arbitrage opportunity. The discussion centres around two cases. One considers tax arbitrage between two perfectly correlated assets, and the other considers the case between two non-perfectly correlated assets. The conditions for the existence of tax arbitrage are different for these two cases. The main contribution of this analysis stems from relaxing some of assumptions made by Basak and Croitoru (2001) that progresses the discussion towards a more complete treatment of this issue. First, the assumption of no capital gains tax is removed by introducing capital gains tax liability in the model. Second, the assumption that all tax arbitrage is dependent on the size of holding, and will disappear after enlarging the size to a certain level, is also removed. A tax arbitrage that is independent of the size of holding is defined and discussed.

The remainder of this section is organised as follows: Section 1.4.1 sets out the background of tax arbitrage, Section 1.4.2 introduces the main proposed improvements to Basak and Croitoru’s (2001) model, and Section 1.4.3 introduces the methodology.

1.4.1 Tax arbitrage opportunities

There are two types of arbitrage opportunities: riskless and risky arbitrage. One example of riskless arbitrage is the practice of taking advantage of a price difference for a single asset between two or more regional markets. This arbitrage is a transaction that involves no negative cash flow at any probabilistic or temporal state and a positive cash flow in at least one state. For instance, a riskless arbitrage is present when there is the opportunity to instantaneously and simultaneously buy an asset at a low price and sell it at a high price. An example of risky arbitrage is statistical arbitrage that refers to expected profit (though losses may occur).

The main arbitrage argument of arbitrage pricing theory is that if the price of an asset diverges from its fair value, arbitrage action should exert supply and demand forces that bring the price back into line. In other words, if the price reflects the fair value of an asset, there should be no arbitrage. In theory, the fair value of an asset can be obtained by finding the arbitrage-free equilibrium (Basak and Croitoru, 2001). This arbitrage-free equilibrium is also the precondition of many other pricing models, such as the Black-Scholes-Merton option pricing model.

However, the discussion of arbitrage-free equilibrium on a post-tax basis would be complicated as the taxing process varies across investors and countries. It is not easy to consider all tax rules in a single pricing model. That is why discussions of pricing models usually assume there is no tax. By using such a pricing model, arbitrage on a before-tax basis can be found and lead the asset's price to move around its before-tax fair value. This is because speculators will buy (or sell) assets when they are undervalued (or overvalued), driving their prices back to fair value. However, an arbitrage-free equilibrium on a before-tax basis may not be the same as an arbitrage-free equilibrium on an after-tax basis. Although it is difficult to consider all tax rules in a single pricing model, it is necessary to find the differences between the two equilibria.

Arbitrage on an after-tax basis, so-called Tax Arbitrage, is “the practice of profiting from differences between the ways that transactions are treated for tax purposes” (Basak and Croitoru 2001). Tax rules are so complex that many taxpayers can restructure their investments to minimize tax. In some cases, tax arbitrage is legal. For example, it is legal if investors profit from receiving revenues in a low tax country while incurring costs in a high tax region. In this way, tax payments can be reduced. In other cases, however, tax arbitrage is illegal. Tax arbitrage is likely to be very widespread. However, it is not easy to estimate the extent to which tax arbitrage is used.

The main consideration of the analysis in Chapter 5 is arbitrage on an after-tax basis. A single model including both capital gains tax and income tax is proposed to discuss the conditions for the existence of tax arbitrage.

1.4.2 Improvements to tax arbitrage model

Basak and Croitoru (2001) investigate the equilibrium implications of the existence of redundant securities with non-linear taxation, and the consequent opportunities for tax arbitrage. Heterogeneous tax rules across investors and countries lead to discrepancies in assets’ pre-tax market prices of risk. They show that this mispricing is set so that investors effectively cooperate to minimize aggregate tax payments, even though individually each investor may not minimize his own tax bill. In their work, the standard Brownian motion is used to simulate income changes and capital gains of an asset. Two main assumptions are used in their work. First, there is no capital gains tax. Second, there is no global tax arbitrage.⁶

Basak and Croitoru’s (2001) work is extended in Chapter 5 by relaxing these two assumptions. When Basak and Croitoru discuss tax arbitrage opportunities, they only consider income tax and assume that there is no tax on capital gains. In reality,

⁶ A tax arbitrage of which the existence is not dependent on the size of the investor’s position, is called ‘Global’ tax arbitrage.

though some of assets and investors may be exempt from capital gains tax, this assumption may lead the analysis to produce an incomplete conclusion. In Chapter 5, in order to re-check their conclusion in a more complete tax environment, their model is extended by incorporating capital gains tax. In addition, Basak and Croitoru (2001) assume that there always exists an arbitrage-free equilibrium in the market where a fair asset price can be found. This assumption is not always true as global arbitrage opportunities may exist in the market. Therefore, investors can enlarge their profit by increasing their holdings given enough market liquidity. There will be no arbitrage-free equilibrium for this case. To deal with this issue, in Chapter 5, the tax arbitrage is divided into three types and each is discussed separately. The solution of removing global tax arbitrage from the market is discussed from the point of view of a tax policy maker. Furthermore, Basak and Croitoru (2001) consider tax arbitrage between two assets that are perfectly correlated, but ignore the more realistic case of non-perfectly correlated assets. In Chapter 5, their work is extended to consider both cases.

1.4.3 Methodology

To achieve the third main objective of the thesis, in Chapter 5, tax is included in an analysis of statistical arbitrage. Sharpe ratios of assets on an after-tax basis are calculated and compared with each other. A simulation model with standard Brownian motion is proposed to determine the condition of the existence of tax arbitrage. In this process, both income tax and capital gains tax are considered as a generalisation.

Regarding perfectly correlated assets, the analysis in Chapter 5 proves that tax arbitrage opportunities may exist, but such arbitrage is dynamic and does not exist consistently. If there is no global tax arbitrage, an arbitrage-free equilibrium can be obtained by minimizing the sum of aggregate market income and capital gains tax payments. A proof is also presented that the arbitrage-free equilibrium obtained without

considering capital gains tax may be different from the equilibrium obtained with considering capital gains tax. It is therefore beneficial to incorporate capital gains tax when considering tax arbitrage and arbitrage-free equilibrium.

For non-perfectly correlated assets, it is not possible to completely offset an asset's risk by using other assets. To solve this issue, caps and floors on assets' income and capital gains are introduced to the model so that risk can be removed by short selling an asset whose cap is lower than the floor of the holding asset. The cap is a call option that sets a maximum future return for an underlying asset. The floor is a put option that sets a minimum future return for an underlying asset. Three new continuous-time optimization models are proposed to find conditions for the existence of local, global and restricted global arbitrage opportunities.⁷ These opportunities are further divided into two categories, type A and type B, depending on whether a strictly positive or only non-negative future net (after-tax) return will be realised for certain without an outflow of funds at any time. Further, given a set of tax rates and asset parameters, a new function, which requires asset holdings as inputs, is proposed to calculate an asset's marginal cap and floor for its total net return. The theoretical analytics show that the existence of tax arbitrage opportunities between non-perfectly correlated assets simply depends on the difference between assets' marginal caps and floors.

1.5 Structure of the Thesis

Chapter 2 reviews the theoretical literature of tax constraints in portfolio optimization and asset pricing. Chapter 3 extends Osorio et al.'s (2008) model by including more trading constraints and introducing a new algorithm to solve the model with a large number of asset classes. Chapter 4 investigates how the introduction of Tobin tax

⁷ Different from 'global' tax arbitrage, the existence of a local tax arbitrage is dependent on the size of investor's position. Too small or too large position may lead to the disappearance of the local tax arbitrage. In addition, if there is a tax arbitrage whose existence has a requirement on the minimum size of investor's position but no requirement on its maximum size, then this kind of tax arbitrage is called 'restricted global' tax arbitrage. More details will be introduced in Chapter 5.

changes the capital flow between regional markets taking into account local investment tax and international withholding tax. Chapter 5 improves Basak and Croitoru's (2001) work by including capital gains tax in the discussion of tax arbitrage, and extending their work by considering tax arbitrage between two non-perfectly correlated assets. Chapter 6 draws together the conclusions for theory, research and policy implications.

Chapter 2 - Literature Review

This chapter reviews the literature on the impact of tax on portfolio optimization and asset pricing. The existing literature on this subject can be divided into two groups, the impact at the micro level and the impact at the macro level. For the former, the discussion centres mainly on how to program tax constraints into personal portfolio optimization and how to solve this complex problem by an advanced algorithm to reduce total computing time. For the latter, the discussion mainly concerns how to include differential taxation rules to address fair value for an asset or asset class on an after-tax basis and how a change on investment tax rules leads to a new aggregate market portfolio and therefore a capital flow between regional markets.

2.1 Impact of Tax on Portfolio Optimization

This section presents a literature review on the impact of tax at a micro-level. The focus is mainly on how to program tax and other trading constraints into a personal portfolio optimization model and how to solve this complex problem using an advanced algorithm to reduce total computing time.

2.1.1 Tax constraints and mathematical programming

To quantify the impact of taxation on personal asset portfolio choice and composition, an optimization with tax and other real-market trading constraints needs to be developed.

Portfolio optimization has been studied using different models. For example, an important issue relating to basic mean-variance optimization is the uncertainty with problem parameters or the so-called estimation risk or uncertainty in the estimation of expected returns. Bonami and Lejeune (2009) minimize portfolio variance while simultaneously considering uncertainty in expected returns (estimation risk) and trading

restrictions modelled with integer constraints. They incorporate uncertainty in expected returns through a probabilistic constraint that follows Roy's (1952) safety first criterion. This identifies the optimal portfolio as that for which the probability of its return falling below a prescribed threshold is minimized. The portfolio's expected return is guaranteed to be above a prescribed minimum level with a high probability, typically $[0.7, 1)$. In addition, the three trading restrictions they consider are diversification, which ensures investments in a number of industrial sections, buy-in threshold, which prevents investors from holding small positions, and round-lot purchasing, which incorporates even-lot block trading behavior of institutional investors. Their research, however, ignores the substantial effects of taxation. Brandes et al. (2012) worked on the relationship between stock-specific transaction costs and portfolio optimization. They found that “the inclusion of stock-specific transaction costs at the portfolio construction stage permits higher turnover levels and allows portfolio managers to run larger portfolios without facing detrimental cost effects”. However, they did not include tax issue in the discussion either.

Taxation is important to investors and adds substantial complications to portfolio optimization problems. Feldstein (1976) analyzes the composition of 1799 households' portfolios and claims that personal income tax has a very powerful effect on individuals' demand for portfolio assets after adjusting for the effects of net wealth, age, sex and the ratio of human to nonhuman capitals. Constantinides and Scholes (1980) investigate the effect of capital gains tax on individual asset portfolios and conclude that, although riskless hedging by options and futures can eliminate capital gains tax, transaction costs blunt and may negate the strategy's effectiveness. As a result, capital gains tax does matter and renders the optimal investment strategy complex. Constantinides (1983 and 1984) extends the discussion on capital gains tax and claims that “tax law confers upon the investor a timing option - to realize capital losses and

defer capital gains. With the U.S. tax rate on long term gains and losses being about half the short term rate, the law provides a second timing option - to realize losses short term and gains long term, if at all". His simulations over the 1962–1977 period establish that investors subject to high tax should realize capital gains in the long term from high variance stocks and buy back stock so that potential future losses in short term can be realized. He also claimed that the small-firm anomaly cannot be explained by tax trading. But tax trading activities can predict a seasonal pattern in trading volume, and stock price volatilities, the so-called January effect, only if investors are irrational or ignorant of the price seasonality. Hubbard (1985) also reaches the conclusion that personal taxation has a significant impact on portfolio choice by analyzing U.S. cross-sectional data.

The above articles led to a discussion at the beginning of this century, of the impact of personal investment taxation. Dybvig and Koo (1996), Dammon *et al.* (2001 and 2004), De Miguel and Uppal (2005), and Birge and Yang (2007) investigate this issue using mathematical programming and all conclude that taxation constraints need to be incorporated in portfolio optimization to ensure a global optimal solution. Their work, however, does not consider real-market variations in tax rules within different accounts and across different countries and regions. In general, investment returns arise mainly in the form of income or capital gains and these are subject to different tax rates. Investors may withdraw funds as income either from returns or from initial invested capital. Tax rates also differ across investment accounts, investment assets and global regions. For example, to maximize tax advantage, some investment accounts have restrictions on the amount, timing and source of withdrawal from income or initial capital that investors can make. Some withdrawal limits increase over the investment horizon, while others are constant. Further, some taxes are payable immediately upon encashment of a certain type of income, while others can be deferred to the end of the

investment horizon. Tax rates and policies also vary across countries and global regions. These tax and withdrawal rules complicate portfolio optimization problems mainly in two ways: they introduce constraints that may involve binary or integer variables, which require integer programming, and cause an indirect mapping between the control variables (asset weights) and the portfolio optimization objective function. In other words, the objective function is neither linear nor perfectly convex with respect to asset weights in the portfolio.

Some recent papers on post-tax portfolio optimization improve long-term investment models by adding real-market features such as tax withdrawals (Osorio *et al.*, 2002, 2004a) and bank taper relief (Osorio *et al.*, 2008b).

Recently, Fischer and Gallmeyer (2016) discussed the performance of trading strategies with consideration of capital gains taxes and transaction costs. However, income tax is not included in his analysis. Huang (2008) included a retirement account which can defer the tax payment into the portfolio optimization and concluded that “investors place highly taxed assets in the tax-deferred account to maximize the tax benefit, and adjust their taxable portfolios to achieve the optimal risk exposure”. Similarly, tax is also considered in the work of Stein and Garland (2008) on investment management. The importance of tax to investment management is explained from high-level perspective. The thesis can be an addition to their work to explain the importance of tax by detailed analysis

a. Estimation Risk

The portfolio optimization literature discusses different approaches to measuring risk and uncertainty (Artzner *et al.*, 1999). Goldfarb and Iyengar (2003), for example, propose a robust factor model to manage risk. Other authors use historical data of asset

returns to represent future risk (Bodnar and Schmid, 2007; Bonami and Lejeune, 2009; Lejeune, 2010) and assume that asset returns follow a normal distribution (Bodnar and Schmid, 2007).

The classic Markowitz framework relies on perfect knowledge of the expected returns of the assets and the variance-covariance matrix (variance risk) and assumes that there is no estimation error. However, the expected returns and covariance between assets are not known and not observable. It is difficult to estimate them accurately. Indeed, a lot of possible sources of error affect their estimation leading to so-called estimation risk (Bawa et al. 1979) in portfolio selection. As stated by Mulvey and Erkan (2003), these sources include the impossibility of obtaining enough samples of data, data instability, and investors' differing estimates of future asset returns. Risk from estimation errors is a source of poor decisions because, as stated by Cornuejols and Tutuncu (2007) and Ceria and Stubbs (2006), the optimal portfolio composition is highly sensitive to the covariance between asset returns and their expected returns, and small changes in the moments of the returns can result in very different optimal portfolios.

Chopra and Ziemba (2011) and Broadie (1993) argue that portfolio estimation risk is mainly caused by errors in the estimation of asset returns but not to the same extent by the errors in estimating the variance of and covariance between asset returns (Ceria and Stubbs, 2006). Therefore, in Chapter 3, when investigating post-tax portfolio optimization, the main focus is on the estimation risk of expected returns (Bonami and Lejeune 2009) rather than the variance-covariance matrix (Lejeune and Samatli-Pac 2012). Since the algorithm proposed by Lejeune and Samatli-Pac (2012) is based on the reformulation of estimation risk of the variance-covariance matrix, the change to estimation risk of expected returns makes their algorithm unsuitable for the problem at hand. The issue of algorithm development will be discussed further in Section 2.1.2.

The estimation risk of expected returns has attracted renewed interest in recent years, and several approaches to incorporate it into portfolio selection have been developed. Here, Roy's safety first risk criterion (Roy, 1959; Bonami and Lejeune's, 2009) that identifies the optimal portfolio as being the one for which the probability of its return falling below a prescribed threshold is minimized, is used to assess the estimation risk of expected returns.

2.1.2 Algorithm for MINLP

As discussed above, to program tax and trading rules into a portfolio optimization model, probabilistic constraints are used with a large number of integer variables. The combination of integer variables and probabilistic constraints, results in mixed-integer nonlinear programming (MINLP) problems. These are challenging to solve, especially for large-scale problems such as the ones proposed (up to 288 assets and corresponding integer variables). Much work has been done on improving the efficiency of algorithms used to solve the mean-variance Markowitz model under MINLP. For example, Bienstock (1996) presents computational experience with a branch-and-cut algorithm to solve quadratic mixed integer programming problems (QMIP). Such problems arise in financial applications. His algorithm solves the largest real-life problems in a few minutes of run-time. Jobst *et al.* (2001) also examine the effects of applying buy-in thresholds, cardinality constraints and transaction round-lot restrictions to the portfolio selection problem under QMIP. To solve this challenging problem, they also propose alternative approaches. However, their methods cannot solve problems with non-linear constraints.

Konno and Yamamoto (2005) consider portfolio optimization problems with integer constraints. Such problems include, among others, mean-risk problems with non-convex transaction cost, minimal transaction unit constraints and cardinality

constraints on the number of assets in a portfolio. These problems, though practically very important, have been considered intractable because they require the solution of MINLP problems for which there are no efficient algorithms. In their paper, they show that these problems can be solved by the state-of-the-art integer programming methodologies if absolute deviation is used as the measure of risk. But their algorithm limits the format of risk constraint in the optimization.

Discussion of efficient algorithms for large-scale MINLP for portfolio optimization has become popular in recent years (Corazza and Favaretto, 2007; Gondzio and Grothey, 2007; Bonami and Lejeune, 2009; and Lejeune and Samatlı-Pac, 2012). However, as mentioned above, their work exhibits one or more of the following features that limit the applicability of their proposed algorithms to the general post-tax portfolio problems that this thesis considers. First, many ignore the uncertainty in problem parameters (estimation risk) (Bienstock, 1996). Second, the objective function considered is specific (Bonami and Lejeune, 2009). Third, the trading rules considered are simplified (Lejeune and Samatlı-Pac, 2012). A new method is, therefore, required for more complex portfolio optimization problems.

Previous work shows that branch and bound (henceforth B&B) methods, such as BONMIN, under most fractional branching rules exhibit higher precision for MINLP than approximating methods such as CPLEX (Bonami and Lejeune, 2009). However, since B&B methods search all possible solutions under a branching tree, it may require a large number of iterations to reach the optimal solution, and this reduces algorithm efficiency, particularly for large-scale MINLP.

2.2 Impact of Tax on Global Markets

This section reviews existing literature on the impact of tax at a macro level. The main focus is how to find market equilibrium (an optimal market portfolio) in the global

financial market under differential regional tax rules, and how a change in regional investment tax rules leads to market equilibrium and therefore a capital flow between regional markets.

2.2.1 General Tax

The existing literature on the macro-level impact of tax can be divided into two major categories - government tax revenues and international tax-driven capital flows.

a. Government tax revenues

Chang et al. (2002) use co-integration and vector auto-regression to assess the ‘Tax-and-Spend’, ‘Spend-and-Tax’, and ‘Fiscal Synchronization’ hypotheses for ten countries using annual time-series data over the period 1951 to 1996. The test results suggest unidirectional causality running from revenues to spending, supporting the ‘Tax-and-Spend’ hypothesis, for Japan, South Korea, Taiwan, UK, and the USA. The opposite relationship, supporting the ‘Spend-and-Tax’ hypothesis, holds only for Australia and South Africa.

Joulfaian and Mookerjee (1991) investigate the sources of growth in government revenues and expenditures in 22 OECD countries. A major conclusion is that reductions in spending are essential to reducing budget deficits and controlling government size.

Saunoris and Payneb (2010) estimate an asymmetric error correction model using a momentum threshold autoregressive approach and data from 1955 to 2009. Their analysis shows that government revenues are sensitive to short term changes in government expenditure and also to budgetary disequilibrium asymmetrically. Regarding the asymmetric adjustment, government revenues are more responsive to a worsening budget than they are to an improving budget. Creedy and Sanz-Sanzb (2011) investigate aggregate personal income tax revenue obtained from a multi-scheduler and

multi-regional personal income tax system, with revenue divided among central and regional governments. They conclude that aggregate income tax revenue can be expressed as a function of characteristics of the distribution of taxable income, making it possible to identify the sources of revenue differences among regions.

Creedy and Gemmell, (2006) demonstrate that it is important to find a reliable method to measure growth of tax revenues, for a tax system and also for its individual taxes, when designing tax policy. A change in tax parameter, such as income thresholds, tax rates, and allowances, is dependent on the expected automatic growth of tax revenue created from the tax system. It is so-called built-in flexibility, or revenue responsiveness, of the tax that generate these automatic revenue changes. In their book, this concept is approved by an invaluable review and synthesis of quantitative analysis. How this concept can be used to estimate revenue responsiveness across countries is demonstrated.

b. Global tax-driven capital flows

Papers in the second category often compare the two withholding tax systems: source- and residence-based taxes. Source-based taxation is justified on the basis that the country which provides the opportunity to generate income or profits should have the right to tax it. Thus, in a source-based tax system, all investments in a country will be taxed only by the government of that country no matter where the investor is from. Residence taxation, on the other hand, is based on the principle that people and firms should contribute to the public services provided for them by the country where they live, so they should be taxed on all their income, wherever it arises. Thus, in a residence-based tax system, all investments by an investor will be taxed only by the investor's country of residence no matter where the investment is located.

The two tax systems can be compared by observing changes in the external current⁸ and financial account balance⁹. A country's balance of payments consists of the financial account, financial account and the current account. In the current account, net factor income (income from overseas investments less payments to overseas investors), the balance of trade, and net cash transfers are the major components. A current account surplus increases a country's net foreign assets whereas a current account deficit reduces the country's net foreign assets.

Bovenberg (1992) explores how residence- and source-based taxes on capital income affect the external current account in small open economies. This effect is examined indirectly using the identity between the external current account balance and the difference between domestic saving and domestic investment¹⁰. The same method is also employed by Summers (1988), Sinn (1985), Slemrod (1988), Murphy (1986), Engel and Kletzer (1989), and Bovenberg (1989) in their discussion of the two tax systems. Summers (1988) examines the interactions between tax policy, international capital mobility, and international competitiveness, and concludes that tax policies which stimulate national investment without affecting national savings must inevitably lead to deterioration in a country's trade balance in the short and intermediate run. Sinn (1985) investigates the Accelerated Cost Recovery System (ACRS) of depreciation by the United States in the 1981 budgetary business tax cuts. This led to the US investment boom, the recent high world interest rates, the strength of the dollar, the US trade deficit and a massive redistribution of world capital towards the United States. Key stages in the process whereby the benefit for US corporations of a larger tax offset on

⁸ Current account: in the current account, net factor income (income from overseas investments less payments to overseas investors), the balance of trade, and net cash transfers are three main components.

⁹ Financial account: the financial account reflects net change in ownership of national assets.

¹⁰ The net capital outflow is the difference between domestic investment and domestic savings. It is calculated as the amount that local residents lend overseas minus the amount that foreigners lend to the home country. If this difference is positive, it means the economy is saving more than it's investing. The excess is lent to foreigners. If the net capital outflow is negative, this means the economy is financing this extra investment by borrowing from abroad.

depreciation could lead to such changes are explained in their paper. Murphy (1986) considers the interaction of saving and investment in determining the current account for a small open economy subject to productivity shocks. The analysis highlights the crucial role of the real exchange rate in macroeconomic adjustment and demonstrates the importance of inter-temporal substitution for determining the response of investment to anticipated productivity disturbances. Engel and Kletzer (1989) “examine a model of a small open economy in which there is free international mobility of financial capital, investment in capital goods and a [non-traded] good”. They explain why, even without restrictions on asset trades, there may be a correlation between investment and saving, and also why a country with high saving may nevertheless borrow from foreigners to finance its investment. Bovenberg (1989) uses “an intertemporal equilibrium model [to analyse] how lower source-based taxes on capital income impact trade performance and international competitiveness“. It shows that, depending on import shares and intertemporal and intratemporal substitution elasticities, capital accumulation may induce changes in the terms of trade and also real interest rates that increase domestic saving, even when financial capital can freely flow across borders.

Analysis of the current account and financial account balances can help us to understand the general impact of the two tax systems on international capital flows but it is difficult to deduce their impact on consumption as opposed to investment. Little research has been done on the impact of the two tax systems on global consumption. Articles concerning consumption tax relates to other tax issues (Bradford, 1995; Walker and Bloomfield, 1987; Brashares, 1999; Gordon et al, 2004; Rousslang, 2002). Bradford (1995) uses the same rate for income and consumption taxes to demonstrate why the issue of taxing ‘old capital’ or ‘old savings’ arises in the movement from an income to a consumption base. This indicates the trade-offs that must be considered as a consequence of this issue when assessing how changes in the price level, with or

without a transition, change the gains/losses distribution. Gordon et al. (2004) investigates the U.S tax system which has returned to the situation of the mid-1980's whereby its tax system raises little revenue from taxing investment income and capital gains. They conclude that, although the revenue from taxing investment income and capital gains is small, the benefits of a clean consumption tax have not been attained. There remain distortions to both saving and investment decisions, and distortions across capital assets, portfolios, corporate financing, and choice of organizational form under the patchwork of provisions that have been adopted. Rousslang (2002) disputes the conclusion of other authors "that a broad-based consumption tax would be more efficient if financial services to consumers, such as services for investment, loans and insurance, were exempted from the tax, even if taxing the financial services posed no special administrative burden. [He argues] that this conclusion rests on key assumptions and that alternative, equally plausible, assumptions support the conclusion that, [apart from] any special administrative burden, the tax rate on financial services to consumers should be at least as high as the tax rate on consumer goods".

Recently there has been work published on the comparison between source- and residence based taxes for investment, generally examining the differential impact of the two tax systems on the global capital allocation of real industry investments (Devereux et al., 2008; Devereux and Griffith, 2003; Fuest and Huber, 2004; Fuest et al., 2005). Devereux et al. (2008) test whether The Organisation for Economic Co-operation and Development "(OECD) countries compete with each other over corporate taxes in order to attract investment". They conclude that countries compete over the effective average rate of tax and the statutory rate of tax, which reflects governments' belief that international firms' choices of location are discrete. Devereux and Griffith (2003) consider the impact of taxation when investors face the location choice of multinationals, which depends on an effective average tax rate. They use data from 1999 to assess the

benefits of harmonising the treatment of dividends and statutory rate of tax in the E.U. They conclude that these benefits are conditional on the mobile investments' profitability. With low profitability, the effects of the co-ordination has no significant influence on the choice of effective tax rates across countries. Nevertheless, this effect is critical when the rate of profitability is high. As the choice of location is usually dependent on the decision of multinational companies with high profitability, the analysis indicates that this kind of co-ordination may have become more beneficial.

Fuest and Huber (2004) investigate the case that a lot of E.U. countries do not charge domestic tax on companies' foreign profits. They argue that with double taxation agreements, where foreign profits are not subject to domestic corporate tax, governments may use income tax on shareholder dividends to tax these profits. Unfortunately, this tax on shareholder dividends encourages the sale of domestic firms to foreigners. However, if double taxation relief can be used for domestic profits, domestic ownership may be preserved. Their results suggest that tax policy on dividends could contribute to the observed 'home bias' in equity portfolio investment.

Nevertheless, focusing on the impact on mergers and acquisitions together with Greenfield investments, Becher and Fuest (2011) reach a different conclusion. They claim that in previous work, the model considers Greenfield investment only and neglects the large part of international capital flows that take the form of mergers and acquisitions (M&A). Taking into account M&A investment leads to substantial changes in the efficiency properties of taxation. A similar conclusion is reached in other articles using different assumptions for the M&A market (Desai and Hines, 2004; Becker and Fuest, 2008, 2010).

Financial market investments account for a large part of global investments and should also be considered when comparing the two tax systems, but this topic has not been fully investigated. There have been studies on the impact of tax on financial

markets but these concern other tax issues. Some consider general transaction tax (Campbell and Kenneth, 1994; Edwards 1992; Hubbard, 1993), and others “Tobin” tax (Tornell, 1990; Reinhart, 1991).

2.2.2 Tobin Tax

As discussed in section 1.3.2c, the idea of a Tobin tax was introduced by James Tobin in the early 1970s (Tobin, 1978). It has been controversial among economists and politicians ever since (e.g. Haq et al., 1996; Habermeier and Kirilenko, 2003; Weaver et al., 2003). In 2013, the European Commission officially announced that a tax on financial transactions out of or into 11 EU countries would be introduced in 2014. The countries involved are Austria, Belgium, Estonia, France, Germany, Greece, Italy, Portugal, Slovakia, Slovenia and Spain. However, this proposal is still under discussion and has been postponed. The discussions are concerned with two matters. The first is about the impact of Tobin tax on market efficiency (the deviation of asset price from its fair value which is usually measured as price volatility in articles on Tobin tax), and the other concerns the impact on trading volume.

As regards market efficiency, some articles conclude that Tobin tax improves market efficiency by decreasing price volatility (Frankel, 1996; Pally, 1999; Ehrestein, 2002; Westerhoff, 2003; Ehrenstein et al., 2005; Cipriani and Guarino, 2008) while others conclude that Tobin tax reduces market efficiency by increasing price volatility (Kupiec, 1995; Aliber et al. 2003). The results are mainly derived from heterogeneous research methods and tax settings, for example the type of investors (long-term or short-term; speculators, fundamentalists or noise traders) and their motivation for trading. By experimental analysis on an artificial market with four types of traders, Mannaro et al. (2008) find that market efficiency decreases and price volatility increases. In contrast, Pally (1996) proposes a microeconomic model with two groups of risk-neutral traders

(fundamentalists and noise traders). He then uses the model to show that noise traders (speculators) create inefficiencies and higher costs for fundamentalists. He argues therefore that, although a Tobin tax would apply equally to noise traders and fundamentalists for a single trade, the overall impact is larger on noise traders as they trade more frequently. Consequently, noise trading would be reduced by a Tobin tax, and therefore market efficiency would be enhanced, which is contrary to the view expressed by Mannaro et al. (2008). So, there is no general agreement on the consequences of a Tobin tax on price volatility and market efficiency, although the work of Haberer (2006) may help to explain these apparent contradictions. In the work of Haberer (2006), a U-shaped relationship between market trading volume and price volatility is advocated. He concludes that market volume is reduced by a Tobin tax. But this new tax can have different impacts on price volatility, depending on the level of trading in the market.

As regards work on trading volume, all articles conclude that the introduction of Tobin tax would reduce the trading volume by decreasing transactions carried out by speculators (Haq et al., 1996; Weaver et al., 2003; Mannaro et al., 2008; Hanke et al., 2010). However, the analysis of Hanke et al. (2010) shows that, although a Tobin tax reduces market trading volume, the size of this reduction is heterogeneous and highly sensitive to the size of the market. In their work, they present an experiment with currency trading on two artificial markets, in which none, one, or both markets include a Tobin tax. Hanke et al. (2010) conclude that trading volume and trading activity are significantly affected if the Tobin tax is levied on the larger market, and the stronger influence of the tax on the larger market seems to be driven by drying up the hitherto very liquid large market. Some further questions arise from the work of Hanke et al. (2010). For example, is the impact of the Tobin tax on trading volume also sensitive to other external factors, such as market liquidity, market correlation, and investment tax

rules? In the literature, analysis of Tobin tax is isolated completely from other tax issues (e.g. income tax and capital gains tax). Is the impact of the Tobin tax on trading volume also sensitive to investment tax rules? When discussing Tobin tax, if investors are not free of income and capital gains tax, they should be included in the model to achieve the correct equilibria.

2.3 Tax Arbitrage

In this section, literature on tax arbitrage modelling and therefore post-tax asset pricing is presented. The expected pre-tax asset return and its risk are normally the main considerations in articles on portfolio management and asset pricing. In reality, however, heterogeneous taxation can significantly influence equilibrium prices. This heterogeneity may exist across different investors, securities and types of returns (capital gains or income). For example, some investors are subject to higher tax rates than others, derivative securities may follow a tax rule different from that of their underlying assets, and even the same asset may be subject to different taxation depending on the purpose for which it is held (e.g. retirement investing). All these features make the asymmetric treatment of taxes important in asset pricing but difficult to include in the mathematical programming. It is this complexity that makes most people decide to simplify asset pricing and portfolio management research by assuming constant tax rates. This problem is addressed by including tax heterogeneities in the model to establish the dynamic equilibrium of asset prices.

On a pre-tax basis, no mispricing of assets guarantees no arbitrage opportunities and thus investors cannot expect to make a profit without taking on risk when going long in one asset and short in another asset. However, with the inclusion of taxation, even if there is no pre-tax mispricing, differential tax treatment can lead to the existence of both attractive and unattractive securities on an after-tax basis and a tax arbitrage

opportunity whereby investors purchase a low-taxed asset financed by selling a high-taxed one. In this way, investors can reduce their tax bill or obtain an extra net positive profit without changing their current risk exposure (Basak and Croitoro, 2001).

Regarding differential taxes across high-income investors and low-income investors, Samuelson (1964) demonstrates that high rate tax-payers could, at the time of the study, reduce tax by purchasing bonds standing at a discount from low rate tax-payers at the expense of the government (lower tax income). Litzenger and Ramaswamy (1980), Dybvig and Ross (1986) and Ross (1987) show that the differential tax treatment on the rich and poor leads to a clientele effect on both quantity and price when purchasing assets, which increases complexity in determining equilibrium asset prices. Talmor (1989) also discusses the role of tax arbitrage in clientele effects on financial leverage. He argues that while total short sale constraints are often introduced to rule out tax arbitrage, such constraints are both unrealistic and conceptually problematic. Instead, milder constraints are advocated, which prevent tax arbitrage while still allowing short positions. It is demonstrated that a model with these constraints can support bond pricing as in the Miller equilibrium, although it leads to a richer set of tax clienteles.

Regarding taxation and asset pricing, Brennan (1970(1)) was the first to consider the pricing of assets under differential taxation of incomes and capital gains, and shows that there is a significant effect on asset pricing. Elton and Gruber (1978) also consider the effect of this differential taxation on portfolio composition and show that the inclusion of income tax changes the market equilibrium significantly. The implication of asymmetric tax on capital gains and income is discussed in later research on both asset pricing and portfolio management (Dammon and Spatt 1996; Basak and Gallmeyer 2003; Osorio et al. 2002, 2004(1), 2004(2), 2008(1) and 2008(2)). Dammon and Spatt (1996) explore the pricing and optimal trading of assets with transaction costs

and asymmetric capital gains tax. They claim that for assets that have been held for long periods, all capital gains below a certain threshold are realized by investors, and for assets that have been held for a short period, all gains and sometimes small losses are deferred by investors. Deferral of short-term losses may be contrary to common intuition. However, even without transaction costs, it may be optimal to defer short-term losses. Tax timing value is significantly lower under the strategies previously analysed than under the optimal trading strategy. Sialm (2009) also tried to find out if investors benefited from the tax burden of equities by observing data from 1913 to 2006, and its impact on asset pricing. They concluded that “an economically and statistically significant relation between before-tax abnormal asset returns and effective tax rates”.

Basak and Gallmeyer (2003) consider a dynamic asset pricing model with asymmetric dividend taxation and a unique risky asset. They study the dynamics of equilibrium security prices when agents face differential dividend taxation, and conclude that under logarithmic preferences, risk is transferred from the higher-taxed to the lower-taxed agent, and the interest rate decreases to counteract extra precautionary savings against this sub-optimally shared risk. Numerical analysis reveals further tax rate, time-to-horizon, and dividend risk effects. For most wealth allocations, the stock return volatility is increased above the no-tax benchmark. Osorio et al. (2002, 2004(1), 2004(2), 2008(1) and 2008(2)) try to include a differential personal investment tax constraint into portfolio optimization and conclude that the inclusion of tax leads to a significant change on the optimal portfolio composition. However, most of these papers ignore the tax effects of capital losses by assuming non-negative increases in market prices. This assumption ignores the long-term advantage of deferred tax and may lead to unrealistic conclusions.

To investigate the role of heterogeneous tax across time, Constantinides (1983) assumes that tax rates are higher in the short term than in the long term, and concludes

that investors will take advantage by realizing losses in the short term but gains in the long term. Dammon and Spatt (1996) and Osorio et al. (2004(1)) reach a similar conclusion in a multi-period model.

Finally, researchers have tried to determine equilibrium asset prices under heterogeneous tax brackets across investors. For instance, Dammon and Green (1987) implement a single-period model to reflect this heterogeneity in asset pricing. They consider the special case in which assets have static pay-offs, but are differentially taxed and conclude that the "no-tax arbitrage" condition simply requires that investors' tax rates must intersect. But this conclusion is reached under the assumption that an asset's capital gains are positive and its pre-tax pay-off is also positive. Jones and Milne (1992) conclude that equilibrium in the capital markets is not possible unless all countries adopt the same principle of international income taxation. Basak and Croitoru (2001) propose a time-continuous model to develop dynamic equilibria of asset prices between two heterogeneous agents when the presence of redundant, non-linearly taxed securities provides opportunities for tax arbitrage. Strobel (2001, 2005, 2012(a) and 2012(b)) programs heterogeneous income taxation for cross-country portfolio management to discover international tax arbitrage opportunities and proposes new tax-modified interest and put-call parity conditions. But when assets are assumed to be uncorrelated, authors usually adopt static but not continuous pay-offs to track possible arbitrage opportunities (Dammon and Green, 1987; Dammon and Spatt 1996; Strobel (2001, 2005, 2012(a) and 2012(b))). Basak and Croitorn (2001) consider arbitrage opportunities between correlated assets with returns assumed to be continuous, but only income tax (not capital gains tax) is included.

Chapter 3 - Tax Effects on Investment Portfolios: Large-Scale

Optimization under Stochastic and Integer Constraints

This chapter investigates the effects of income tax and capital gains tax on large investment portfolios. An optimization model is proposed to return portfolios meeting or exceeding a prescribed return threshold with a high confidence level and satisfying buy-in threshold and diversification constraints. Complex tax trading rules with withdrawal features of investment bonds are also incorporated. To implement this Mixed Integer Non-linear Programming (MINLP) model on large scale applications, a solution based on Greedy heuristics with newly introduced dynamic ranking and integer evaluation rules is proposed. Performance comparisons with extant MINLP branch and bound and approximation method solvers show that, while the latter fair well for small-scale MINLP problems of less than 72 asset classes, the proposed method retains good performance with up to 288 asset classes¹¹. A study on individual portfolio composition using the proposed model and solution method finds substantial non-linear tax effects on riskier assets and enhanced effects of withdrawal tax only when tax rates are high. In practice, the developed framework better enables investors to react to tax changes, and tax policy-makers to quantify the influence of tax changes on private investment preferences.

3.1 Introduction

Portfolio optimization has been studied using single and multistage stochastic programming with discrete asset choice constraints. An important issue relating to mean-variance optimization is the uncertainty in problem parameters or the so-called estimation risk or uncertainty in the estimation of expected returns. Bonami and Lejeune

¹¹ In this work, income and capital gains of an asset class (e.g. U.K. Telecom Equity) are calculated from historical data of a corresponding market index (e.g. U.K. Telecom Equity Index)

(2009) minimize portfolio variance while simultaneously considering uncertainty in expected returns (estimation risk) and trading restrictions modeled with integer constraints. They incorporate uncertainty in expected returns through a probabilistic constraint that follows Roy's (1952) safety first criterion. This identifies as optimal the portfolio for which the probability of its return falling below a prescribed threshold is minimized. The portfolio's expected return is above a prescribed minimum level with a high probability, typically $[0.7, 1)$. They also consider the following three trading restrictions: diversification, which ensures investments in a number of industrial sectors; buy-in threshold, which prevents investors from holding small positions; and round-lot purchasing, which incorporates even-lot block trading behavior of institutional investors. Bonami and Lejeune (2009), however, ignore the substantial effects of taxation, which are important to investors. Their work is extended in this important direction by solving portfolio optimization problems that incorporate their three trading restrictions as well as income and capital gains tax under a realistic set of tax rules.

Taxation complicates portfolio optimization problems, but can have significant and important effects on investor wealth (Feldstein 1976; Constantinides and Scholes 1980; Constantinides 1983, 1984; Hubbard 1985; Dybvig and Koo 1996; Dammon et al. 2001, 2004; De Miguel and Uppal 2005; Birge and Yang 2007). Most of this cited prior work, however, does not consider real-market variations in tax rules within different investment accounts and across different countries and regions. In reality, investment returns arise mainly in the form of income or capital gains and these are subject to different tax rates. Investors may withdraw funds as income either from returns or from initial invested capital. Tax rates also differ across investment accounts, investment assets and global regions. For example, to maximize tax advantage, some investment accounts have restrictions on the amount, timing and source of withdrawal from income or initial capital that investors can make. Further, some withdrawal limits increase over

the investment horizon, while others are constant, and some taxes are payable immediately upon encashment of a certain type of income, while others can be deferred to the end of the investment horizon. Moreover, tax rates and policies also vary across countries. These tax and withdrawal rules complicate portfolio optimization problems mainly by introducing constraints that may involve binary or integer variables, which require integer programming, and by causing an indirect mapping between the control variables (asset weights) and the portfolio optimization objective function.

Recent papers on post-tax portfolio optimization improve long-term investment models by adding real-market features such as tax withdrawals (Osorio et al. 2002, 2004a) and bank taper relief (Osorio et al. 2008b). These papers focus on the effect of taxes on portfolio allocation and deal with return uncertainty through scenario trees. Research on post-tax portfolio optimization can be extended further in five ways. First, a probabilistic constraint as in Bonami and Lejue (2009) is included to consider return uncertainty and estimation risk simultaneously with integer and other constraints that incorporate taxation and withdrawal rules. Second, the diversification constraint of Osorio et al. (2004a) is enhanced by requiring a portfolio to maintain a minimum number of assets, which is a regulatory requirement for some institutional investors. Third, the framework of Osorio et al. (2004a,b) is also improved by introducing a buy-in threshold constraint that avoids small investments in individual assets that are disallowed, cannot be purchased or are costly to maintain. Fourth, the withdrawal rules of Osorio et al. (2008b) is relaxed to allow for transactions between accounts, rather than within accounts only. Finally, as an extension to the study of Osorio et al. (2008b), the optimal portfolios under different tax rates (varying between 0 and 0.7) is obtained. The results present a non-linear relation between investor's optimal weight on a risky asset class and tax rate applied to the asset class in a complex and realistic tax and trading environment. The extended study provides an example of how the proposed

model and solution method enables investors and policy-makers to estimate the effects of tax rate changes.

The combination of integer and probabilistic constraints result in mixed-integer nonlinear programming (MINLP) problems. These are challenging to solve, especially for the considered large-scale problems. Much work has been done on improving the efficiency of algorithms used to solve the mean-variance Markowitz model under MINLP (Bienstock 1996; Konno and Yamamoto 2005; Jobst et al. 2001; Corazza and Favaretto 2007; Gondzio and Grothey 2007; Bonami and Lejeune 2009; Lejeune and Samatlı-Paç 2012). However, certain features can limit the applicability of these proposed algorithms to my proposed post-tax portfolio problem. First, many ignore the uncertainty in problem parameters (estimation risk) (Bienstock 1996). Second, the objective function considered is specific (Bonami and Lejeune 2009). Third, the trading rules considered are simplified (Lejeune and Samatlı-Paç 2012). A new method is therefore required for my proposed portfolio optimization problems.

Previous work shows that branch and bound (henceforth B&B) methods, such as BONMIN, under most fractional branching rules exhibits higher precision for MINLP than approximating methods such as CPLEX (Bonami and Lejeune 2009). However, the basic B&B method finds the optimal solution by ensuring that no other solutions can return a better result under a branching tree. To guarantee the optimality of returned solution, it may require a large number of iterations. This number will increase further exponentially as the number of integer variables rises. Therefore, for large-scale MINLP, the basic B&B cannot always return a valid solution within a given period of time. Improvements to the B&B method have been made to reduce its computing time but also make the new method specific to certain type of problems. For example, Bonami and Lejeune (2009) improve the B&B method by applying a new branching rule to largely reduce the required number of iterations when solving large-scale MINLP.

However, their new branching rule is only able to solve their proposed mean-variance model, where portfolio variance is minimized as the objective function while portfolio return is considered in the constraints. A new method based on Greedy heuristics by applying an improved ranking rule is proposed to find the optimal solution to each integer variable in sequence without a need to go back to previous steps. It makes the locally optimal choice for each integer variable approximate to a global optimum. The proposed method does not guarantee the optimality of solution to my proposed portfolio optimization problem, but it approximates a global optimal solution in a reasonable time with an acceptable optimality gap. In Section 3.4, experimental results show that the approximation returned by the proposed method is reliable in most cases, particularly for large-scale optimization problems. The classical Greedy heuristics (Chvatal, 1979), however, has its own disadvantages. In particular, its scope is limited to specific problems and its precision is highly dependent on the order of iteration. In response, a modification to Greedy is presented, and its performance is compared with that of BONMIN (the basic B&B method) and CPLEX (the approximating method).

One of main contributions of this chapter is that the proposed framework is used to investigate tax effects on personal portfolio investments. Specifically, the effects of changes in tax rates across asset classes in optimal portfolios of personal investors are investigated. Three accounts that are subject to different tax and withdrawal rules are considered, namely, offshore bonds, onshore bonds and unit trusts.¹² Total post-tax return is maximized subject to the constraints mentioned above. This will help investors to decide if they need change their portfolio after a new tax policy is released by the government. How big the impact of withdrawal tax on the optimal portfolio is also tested. This will help investors to assess whether it is worth considering withdrawal tax in the optimization process.

¹² Investors can transfer money into an account and then use the money in the account to purchase financial assets.

The remainder of this chapter is organized as follows. Section 3.2 presents the optimization problem and relevant settings for the micro level analysis. Section 3.3 presents the objective functions and constraints of the model. Section 3.4 presents the modified Greedy heuristic and tests its performance. Section 3.5 presents the empirical analyses of tax effects. Section 3.6 summarizes and concludes.

3.2 Post-tax Personal Investment Portfolio Optimization

Tax effects are tested by optimizing personal portfolios over a single period and, for a given level of risk, maximizing return net of taxes, management fees, and transaction costs. Investors diversify by both allocating their wealth across risky asset classes and locating their wealth across three investment 'accounts' that follow different tax and cash withdrawal rules. These accounts are offshore investment bonds, onshore investment bonds and unit trusts. Offshore bonds is a generic umbrella account for investments that benefit from certain tax concessions such as deferral, while unit trusts are assumed to contain only equity investments. Different tax rates and rules apply to income, capital gains and withdrawals. The general UK tax framework of Osorio et al. (2004a,b, 2008a,b) is adopted, which is still used by UK insurance companies, such as Prudential.¹³ Although some years have passed since the research of Osorio et al was published, the tax treatment on investment bonds is still the same nowadays. In this Chapter, Osorio et al.'s constraints are enhanced in a number of ways and flexibility is added to allow this setup to be applicable in other countries. This is discussed in Section 3.3.

Investors generally have two decisions to consider: initializing new portfolios or rebalancing existing portfolios. These are represented by two separate settings. First, on

¹³ Prudential plc is a British multinational life insurance and financial services company headquartered in London, United Kingdom. It was founded in London in May 1848. More details regarding tax treatment on investment bonds can be found at:
<http://www.pru.co.uk/investments/investment-articles/investments-and-tax/>

initializing new portfolios, investors are assumed to start with cash in hand, and the model's function is to optimize buying decisions in forming new risky asset portfolios within and across the three accounts. There is no demand for interim cash withdrawals and only new cumulative taxes need to be deducted from total return at the end of the period. Second, at rebalancing, investors are assumed to hold an existing portfolio that is bequeathed from the previous period with no cash in hand. Thus, the model's function is to optimize buy and sell decisions. New cumulative taxes, as well as old taxes accumulated from previous periods, need to be deducted from end-of-period total return.

In line with Osorio et al. (2004a), the tax structures for the three accounts are as follows:

- a. Offshore (investment) bonds
 - All taxes are cumulated and paid on total return at the end of the investment horizon.
 - Annual withdrawals up to 5% of the original investment are permitted, and associated taxes are deferred until the end of investment (encashment). Unused withdrawal allowances may be carried forward indefinitely.
 - Additional withdrawals beyond the annual 5% allowance limit may be made subject to an immediate tax payment at the encashment rate of t_{off} .
 - Withdrawals from the original capital are permitted only when all positive returns have been withdrawn. These withdrawals are not taxed.
- b. Onshore (investment) bonds
 - Part of the tax on total return is cumulated to the end of the investment and the rest is paid annually at the end of each period.
 - Tax on withdrawal is the same as offshore bonds, except that the encashment (only) tax rate is t_{on} .

c. Unit trusts

- Tax on capital gains is paid at the end of investment and this rate changes as its holding time increases. Thus, different assets in a portfolio may be subject to different tax rates depending on their time of purchase.
- Tax on income (e.g., dividends) is paid annually at the end of each period.
- Only return from the previous period is available for withdrawal at the beginning of the decision period.
- Withdrawals from the last period's income are not taxed at the encashment rate.
- Withdrawals from last year's capital gains are subject to an immediate tax at the encashment rate of CG_T , where the optimizing period is counted as $T + 1$.
- Capital withdrawals follow the same rules as for offshore bonds.

The notation adopted is defined in Table 3.1.

Table 3. 1 Notation

Input data	
$\mathbf{1}$	$= (1,1,\dots,1)'$
$\mathbf{u}'\mathbf{v}$	$= u_1v_1 + u_2v_2 + \dots + u_nv_n$ (Inner product)
$\mathbf{u} \circ \mathbf{v}$	$= (u_1v_1, u_2v_2, \dots, u_nv_n)'$ (Hadamard product)
$\mathbf{u./v}$	$= (u_1/v_1, u_2/v_2, \dots, u_n/v_n)'$
n_k	number of investment assets in account k
N_{min}	minimum number of assets with non-zero wealth in portfolio
L	total amount of wealth at the beginning of investment
L_k	amount of wealth at the beginning of investment in account k
f_k	percentage paid in management fee for account k
\mathbf{d}_j	sets of historical dividends or income returns of each asset in class j
\mathbf{g}_j	sets of possible capital gains of each asset in class j
$\bar{\mathbf{d}}_j$	expected dividends or income returns in class j
$\bar{\mathbf{g}}_j$	expected capital gains in class j
t_{off}	cumulative tax rate on gross returns of offshore bond
t_{on}	cumulative tax rate on gross returns of onshore bond
t_{an}	annual tax rate on gross returns from onshore bond
\mathbf{t}_{in}	income tax rate paid on dividends or income
G_T	capital gains tax rate in period t (changing by time)
c	transaction cost
$t_{off,e}$	tax rate on gross returns of offshore bond when underlying is equity
$t_{off,b}$	tax rate on gross returns of offshore bond when underlying is bond
$t_{off,c}$	tax rate on gross returns of offshore bond when underlying is commodity
\mathbf{O}_{ji}	net income from each asset of class j in previous year in unit trust
\mathbf{O}_{jg}	net gain from each asset of class j in previous year in unit trust
R_{kj0}	initial cumulative returns for account k of class j
X_{kj0}	initial accumulated tax in account k of class j
W_k	accumulated first withdrawal from account k
\mathbf{w}_{kj0}	initial amount of money held in each asset in account k of class j
T	number of past years of current investment
w_{min}	minimum position for each holding asset
\mathbf{U}_j	upper percentage bounds for asset j
R_{min}	minimum required return from investment
p_{min}	minimum required probability that real return exceeds required minimum level
Σ	covariance of assets
C_0	external funding at the beginning of period
Variables	
R_{kj1}	cumulative returns for account k after withdrawal of class j
X_{kj2}	final accumulated tax in account k of class j
V_{kj}	net redemption value obtained from account k of class j
\mathbf{w}_{kj1}	amount of money held in each asset after rebalance in account k of class j
\mathbf{w}_{kj2}	final amount of money held in each asset in account k of class j
\mathbf{i}_{kj}^b	amount of money spent to buy an asset in account k of class j
\mathbf{i}_{kj}^s	amount of money obtained when selling an asset in account k of class j

Table 3.1 Notation (cont.)

h_{kj}^1	first withdrawal from account k of class j
h_{kj}^2	excess withdrawal from account k of class j
h_{kj}^3	withdrawal taken from the original investment in account k of class j
I_{kj}	money spent to buy an asset of class j in account k when using withdrawal
y_k	$\in \{0,1\}$, binary variable for account k,
δ_k	$\in \{0,1\}$, binary variable for assets in account k

3.3 Problem Constraints and Objective Functions

3.3.1 Basic trading constraints

a. Internal trading budget

In the following sections, all asset classes in the equity markets are grouped together as a high-level asset class, called the equity asset class. Similarly, all asset classes in the bond markets are grouped together as a high-level asset class, called the bond asset class, and all asset classes in the commodity markets are grouped together as a high-level asset class, called the commodity asset class. An internal trading budget (balance) constraint ensures that for every account $k = 1, 2, 3$ the total selling proceeds from all three high-level asset classes $j=1, 2, 3$ ¹⁴, $\mathbf{1}' \mathbf{i}_{kj}^s$, are equal to the total buying costs, $\mathbf{1}' \mathbf{i}_{kj}^b$, so that

$$\sum_j (\mathbf{1}' \mathbf{i}_{kj}^b - \mathbf{1}' \mathbf{i}_{kj}^s) = 0, \quad \forall k = 1, 2, 3. \quad (3.1)$$

where $\mathbf{1}'$ is $(1, 1, 1, 1, 1, \dots)$, \mathbf{i}_{kj}^s is money received from the sale of assets in the high-level asset class j and account k, and \mathbf{i}_{kj}^b is the money spent to purchase assets in high-level asset class j and account k.

b. Diversification

¹⁴ $j=1$ is the equity asset class; $j=2$ is the bonds asset class; $j=3$ is the commodities asset class.

The following constraint sets an upper bound on the total value of each asset in a portfolio:

$$\sum_{k=1}^3 \mathbf{1}' \mathbf{w}_{kj1} \leq U_j \sum_{j=1}^3 \sum_{k=1}^3 \mathbf{1}' \mathbf{w}_{kj1} \quad \forall j = 1,2,3 \quad (3.2)$$

\mathbf{w}_{kj1} is the weight of asset class j in account k after rebalancing and U_j is the upper percentage of asset j in the portfolio. Therefore $U_j \sum_{j=1}^3 \sum_{k=1}^3 \mathbf{1}' \mathbf{w}_{kj1}$ is the upper bound and $\sum_{k=1}^3 \mathbf{1}' \mathbf{w}_{kj1}$ is the total wealth of each asset in all three accounts. By also setting a lower bound on the total number of assets in a portfolio, N_{min} , firm specific risk can be minimized in the portfolio,

$$\sum_{j=1}^3 \sum_{k=1}^3 \mathbf{1}' \delta_{kj} \geq N_{min}, \quad (3.3)$$

where the sum of the binary variables, $\delta_{kj} \in \{0,1\}$, counts this number. More details about this binary variable are introduced in the following paragraph. In their diversification constraint, Osorio et al. (2004a) do not stipulate a minimum number of assets, which is a beneficial consideration to individual investors in portfolio management. Evans and Archer (1968) conclude that an investor needs to construct a portfolio containing as little as 15 randomly selected stocks before the benefits of diversification, as measured by the standard deviation, are largely exhausted. Recent studies provide strong support for limiting the number of stocks an investor needs to hold to reduce portfolio risk to an acceptable level. Campbell et al. (2001) find a greater need for diversification. This need, they discover, is caused by the increased volatility of individual stocks, not increased volatility of the market. That has led to decreased correlations among individual stocks. Declining correlations among equities implies that the benefits of portfolio diversification have increased over time. The authors find that while a portfolio of about 20 stocks was sufficient to reduce the excess standard deviation of a portfolio to 10 percent in the 1960s, by the turn of the century that figure had risen to 50 stocks.

c. Buy-in threshold

This constraint requires a minimum purchasing volume for each asset, as small holdings are costly to maintain, disallowed or cannot be purchased. Thus,

$$\mathbf{w}_{kj1} \leq \delta_{kj} \sum_{j=1}^3 \sum_{k=1}^3 \mathbf{1}' \mathbf{w}_{kj1} \quad \forall k = 1,2,3; j = 1,2,3 \quad (3.4)$$

$$w_{min} \delta_{kj} \leq \mathbf{w}_{kj1} \quad \forall k = 1,2,3; j = 1,2,3. \quad (3.5)$$

If investors want to hold one asset in the new portfolio, the corresponding binary variable δ_{kj} must be valued at 1. In (3.5), $w_{min} \delta_{kj}$ defines the buy-in threshold requirement. w_{min} is the minimum weight that an asset can be hold in one account. This constraint does not feature in Osorio et al. (2004a,b, 2008a,b). The large number of integer variables introduced by constraints (3.3), (3.4) and (3.5) (equal to the number of assets, which could be hundreds) contributes to the complexity of the problem and, together with the non-linear stochastic risk constraint discussed below in Section 3.3.3, necessitates a new algorithm to solve the large-scale MINLP.

3.3.2 Taxation

The total tax liability is built up by calculating the impact of different tax rules on cumulative returns, withdrawals and wealth.

a. Cumulative returns

The remaining returns available for withdrawal in each account are:

$$\mathbf{R}_1 = \mathbf{R}_0 - \mathbf{h}^1 \mathbf{1} - (\mathbf{h}^2 \mathbf{1}) ./ (\mathbf{1} - \mathbf{t}) \quad (3.6)$$

$$\mathbf{R}_i = (R_{11i}, R_{12i}, R_{13i}, R_{21i}, R_{22i}, R_{23i}, R_{31i}, R_{32i}, R_{33i})' \quad \forall i = 0,1 \quad (3.7)$$

$$\mathbf{h}^i = (\mathbf{h}_{11}^i, \mathbf{h}_{12}^i, \mathbf{h}_{13}^i, \mathbf{h}_{21}^i, \mathbf{h}_{22}^i, \mathbf{h}_{23}^i, \mathbf{h}_{31}^i, \mathbf{h}_{32}^i, \mathbf{h}_{33}^i) \quad \forall i = 1,2,3 \quad (3.8)$$

$$\mathbf{t} = (t_{off,e}, t_{off,b}, t_{off,c}, t_{on,e}, t_{on,b}, t_{on,c}, G_{T,e}, G_{T,b}, G_{T,c})' \quad (3.9)$$

Here, both $\mathbf{h}^1\mathbf{1}$ and $\mathbf{h}^2\mathbf{1}$, in equation (3.6), represent total amounts of cash withdrawal from each account and high-level asset class. \mathbf{h}^1' and \mathbf{h}^2' is the transposed matrix of \mathbf{h}^1 and \mathbf{h}^2 . The difference is that the cash in \mathbf{h}^1 is within the tax free allowance, while the cash in \mathbf{h}^2 is beyond the tax free allowance and is therefore subject to an immediate tax payment. Here, \mathbf{h}^2 is the net amount. By using $\mathbf{h}^2\mathbf{1}$ divided by $\mathbf{1} - \mathbf{t}$, the amount of withdrawal beyond the tax free allowance before tax is returned. This amount plus the withdrawal within the tax free allowance is used to calculate the remaining cumulative return of the portfolio. R_{kj0} , in equation (3.7), is the total return of asset class j in account k available for withdrawal before rebalancing, while R_{kj1} is the total return of asset class j in account k available for further withdrawal after rebalancing. In addition, an upper bound on total withdrawals from asset returns is set by the constraint $R_{kj1} \geq 0$ ($\forall k = 1,2,3$).

b. Withdrawals

There are two types of withdrawal: one from returns and the other from initial capital. For return withdrawals from investment bonds there are immediate tax and withdrawal allowance limits, which increase with the time horizon:

$$\sum_j \mathbf{h}_{kj}^1 \mathbf{1} \leq 0.05 \times T \times L_k - W_k \quad \forall k = 1,2 \quad (3.10)$$

where T is the number of past years of current investment, L_k is the amount of wealth at the beginning of the investment in account k, and W_k is the accumulated first withdrawal from account k (i. e. \mathbf{h}_{kj}^1). Unlike investment bonds, withdrawals from the unit trust account are free from immediate tax and are only available from last year's income, \mathbf{O}_{ji} .

$$\mathbf{h}_{3j}^1 \leq \mathbf{O}_{ji} \quad \forall j = 1,2,3 \quad (3.11)$$

$$\mathbf{h}_{3j}^2 \leq \mathbf{O}_{jg} \quad \forall j = 1,2,3 \quad (3.12)$$

Initial capital, however, is available for encashment if, and only if, all available returns have been used up. Binary variables, $y_k \in \{0,1\}$, are subject to the following restrictions.

$$\sum_j \mathbf{h}_{kj}^1 \mathbf{1} - L \times y_k \leq 0 \quad \forall k = 1,2,3 \quad (3.13)$$

$$\sum_j R_{kj1} + L \times y_k \leq L \quad \forall k = 1,2,3 \quad (3.14)$$

\mathbf{h}_{kj}^3 is withdrawal from initial capitals of asset j in account k which is free of tax but is allowed only after the asset's returns are all withdrawn. L is the total amount of wealth at the beginning of the investment. According to (3.13) and (3.14), if investors wish to withdraw initial capital (\mathbf{h}_{kj}^3 is positive), the binary variable y_k will be 1, and R_{kj1} will then be equal to, or less than, 0. Thus, (3.13) and (3.14) introduce additional binary variables to the problem.

c. Wealth and external trading budget

The total wealth in each account after trading, and at the end of the period, are now calculated. In calculating the former, transactions both within and between accounts are counted.

$$\mathbf{w}_1 = \mathbf{w}_0 - [\mathbf{h}^1 + \mathbf{h}^2 / (1 - t) + \mathbf{h}^3] + (1 - c)(\mathbf{i}^b + \mathbf{I}) - \mathbf{i}^s \quad (3.15)$$

$$\mathbf{w}_i = (\mathbf{w}_{11i}, \mathbf{w}_{12i}, \mathbf{w}_{13i}, \mathbf{w}_{21i}, \mathbf{w}_{22i}, \mathbf{w}_{23i}, \mathbf{w}_{31i}, \mathbf{w}_{32i}, \mathbf{w}_{33i}) \quad \forall i = 0,1,2 \quad (3.16)$$

$$\mathbf{i}^b = (\mathbf{i}_{11}^b, \mathbf{i}_{12}^b, \mathbf{i}_{13}^b, \mathbf{i}_{21}^b, \mathbf{i}_{22}^b, \mathbf{i}_{23}^b, \mathbf{i}_{31}^b, \mathbf{i}_{32}^b, \mathbf{i}_{33}^b) \quad (3.17)$$

$$\mathbf{i}^s = (\mathbf{i}_{11}^s, \mathbf{i}_{12}^s, \mathbf{i}_{13}^s, \mathbf{i}_{21}^s, \mathbf{i}_{22}^s, \mathbf{i}_{23}^s, \mathbf{i}_{31}^s, \mathbf{i}_{32}^s, \mathbf{i}_{33}^s) \quad (3.18)$$

$$\mathbf{I} = (\mathbf{I}_{11}, \mathbf{I}_{12}, \mathbf{I}_{13}, \mathbf{I}_{21}, \mathbf{I}_{22}, \mathbf{I}_{23}, \mathbf{I}_{31}, \mathbf{I}_{32}, \mathbf{I}_{33}) \quad (3.19)$$

$$C_0 + \sum_{j=1}^3 \sum_{k=1}^3 (\mathbf{1}' \mathbf{h}_{kj}^1 + \mathbf{1}' \mathbf{h}_{kj}^2 + \mathbf{1}' \mathbf{h}_{kj}^3) = \sum_{j=1}^3 \sum_{k=1}^3 \mathbf{1}' \mathbf{I}_{kj} \quad (3.20)$$

Apart from the balance for internal, within account, trading, there is a balance for external trading across accounts, as shown in (3.20). With regard to the trading budget constraints (3.15), Osorio et al. (2004a) assume that all withdrawals during subsequent periods are held in cash and no transactions are allowed between accounts (Eq. (18) in

their paper). In contrast, cross-account transactions and cash withdrawal re-investments are included in this work, which is more realistic. In calculating wealth at the end of the period, both expected capital gains and income are considered, and corresponding annual tax payments and management fees are deducted.

$$\mathbf{w}_{112} = (1 - f_1)[(\mathbf{1} + \bar{\mathbf{d}}_e + \bar{\mathbf{g}}_e) \circ \mathbf{w}_{111}] \quad (3.21a)$$

$$\mathbf{w}_{122} = (1 - f_1)[(\mathbf{1} + \bar{\mathbf{d}}_b + \bar{\mathbf{g}}_b) \circ \mathbf{w}_{121}] \quad (3.21b)$$

$$\mathbf{w}_{132} = (1 - f_1)[(\mathbf{1} + \bar{\mathbf{d}}_c + \bar{\mathbf{g}}_c) \circ \mathbf{w}_{131}] \quad (3.21c)$$

$$\mathbf{w}_{212} = (1 - f_2) \left[\left(\mathbf{1} + (1 - t_{ane})(\bar{\mathbf{d}}_e + \bar{\mathbf{g}}_e) \right) \circ \mathbf{w}_{211} \right] \quad (3.22a)$$

$$\mathbf{w}_{222} = (1 - f_2) \left[\left(\mathbf{1} + (1 - t_{anb})(\bar{\mathbf{d}}_b + \bar{\mathbf{g}}_b) \right) \circ \mathbf{w}_{221} \right] \quad (3.22b)$$

$$\mathbf{w}_{232} = (1 - f_2) \left[\left(\mathbf{1} + (1 - t_{anc})(\bar{\mathbf{d}}_c + \bar{\mathbf{g}}_c) \right) \circ \mathbf{w}_{231} \right] \quad (3.22c)$$

$$\mathbf{w}_{312} = (1 - f_3)[(\mathbf{1} + (\mathbf{1} - t_{ine}) \circ \bar{\mathbf{d}}_e + \bar{\mathbf{g}}_e) \circ \mathbf{w}_{311}] \quad (3.23a)$$

$$\mathbf{w}_{322} = (1 - f_3)[(\mathbf{1} + (\mathbf{1} - t_{inb}) \circ \bar{\mathbf{d}}_b + \bar{\mathbf{g}}_b) \circ \mathbf{w}_{321}] \quad (3.23b)$$

$$\mathbf{w}_{332} = (1 - f_3)[(\mathbf{1} + (\mathbf{1} - t_{inc}) \circ \bar{\mathbf{d}}_c + \bar{\mathbf{g}}_c) \circ \mathbf{w}_{331}] \quad (3.23c)$$

Furthermore, since the tax rate on unit trusts decreases with the holding period, a set of rates, t_{ine} , t_{inb} , t_{inc} , are used to account for this feature.

d. Cumulative taxes

Finally, the total tax liability is calculated by adding deferred tax from previous periods to that of the current period.

$$X_{112} = X_{110} + t_{offe}(1 - f_1)[(\bar{\mathbf{d}}_e + \bar{\mathbf{g}}_e)' \mathbf{w}_{111}] - \{t_{offe}/(1 - t_{offe})\} \mathbf{1}' \mathbf{h}_{11}^2 \quad (3.24a)$$

$$X_{122} = X_{120} + t_{offb}(1 - f_1)[(\bar{\mathbf{d}}_b + \bar{\mathbf{g}}_b)' \mathbf{w}_{121}] - \{t_{offb}/(1 - t_{offb})\} \mathbf{1}' \mathbf{h}_{12}^2 \quad (3.24b)$$

$$X_{132} = X_{130} + t_{offc}(1 - f_1)[(\bar{\mathbf{d}}_c + \bar{\mathbf{g}}_c)' \mathbf{w}_{131}] - \{t_{offc}/(1 - t_{offc})\} \mathbf{1}' \mathbf{h}_{13}^2 \quad (3.24c)$$

$$X_{212} = X_{210} + t_{one}(1 - f_2)[(\bar{\mathbf{d}}_e + \bar{\mathbf{g}}_e)' \mathbf{w}_{211}] - \{t_{one}/(1 - t_{one})\} \mathbf{1}' \mathbf{h}_{21}^2 \quad (3.25a)$$

$$X_{222} = X_{220} + t_{onb}(1 - f_2)[(\bar{\mathbf{d}}_b + \bar{\mathbf{g}}_b)' \mathbf{w}_{221}] - \{t_{onb}/(1 - t_{onb})\} \mathbf{1}' \mathbf{h}_{22}^2 \quad (3.25b)$$

$$X_{232} = X_{230} + t_{onc}(1 - f_2)[(\bar{\mathbf{d}}_c + \bar{\mathbf{g}}_c)' \mathbf{w}_{231}] - \{t_{onc}/(1 - t_{onc})\} \mathbf{1}' \mathbf{h}_{23}^2 \quad (3.25c)$$

$$X_{312} = X_{310} + G_{T+1e}(1 - f_3)[\bar{\mathbf{g}}_e' \mathbf{w}_{311}] - \{G_{T+1e}/(1 - G_{Te})\} \mathbf{1}' \mathbf{h}_{31}^2 \quad (3.26a)$$

$$X_{322} = X_{320} + G_{T+1b}(1 - f_3)[\bar{\mathbf{g}}_b' \mathbf{w}_{321}] - \{G_{T+1b}/(1 - G_{Tb})\} \mathbf{1}' \mathbf{h}_{32}^2 \quad (3.26b)$$

$$X_{332} = X_{330} + G_{T+1c}(1 - f_3)[\bar{\mathbf{g}}_c' \mathbf{w}_{331}] - \{G_{T+1c}/(1 - G_{Tc})\} \mathbf{1}' \mathbf{h}_{33}^2 \quad (3.26c)$$

For unit trusts, the tax rate changes over time. Therefore separate rates are assumed for cumulative and immediate taxes, G_{T+1} and G_T , and calculate the final net return of each account by subtracting all deferred tax liabilities from account wealth.

$$V_{kj} = \mathbf{1}' \mathbf{w}_{kj2} - X_{kj2} \quad \forall k = 1, 2, 3; j = 1, 2, 3 \quad (3.27)$$

Here, V_{kj} is total net return from all asset classes in high-level asset class j and account k .

3.3.3 Estimation (stochastic) risk constraint

The portfolio optimization literature discusses different approaches to measuring risk and uncertainty (Artzner *et al.*, 1999). Goldfarb and Iyengar (2003), for example, propose a robust factor model to manage risk. Others use historical data of asset returns to represent future risk (Bonami and Lejeune 2009; Lejeune 2010) and assume that asset returns follow a normal distribution (Bodnar and Schmid 2007). In this analysis, the risk measurement falls within the Markowitz mean-variance framework. Bonami and Lejeune (2009) state that the classic Markowitz framework relies on perfect knowledge of expected returns and the variance-covariance matrix of the assets. This assumes that there is no estimation error. However, expected returns and variance-covariance matrices are unobservable and unknown. Obtaining accurate estimates of them is a challenge. Indeed, many possible sources of errors (e.g., impossibility of obtaining a sufficient number of data observations, instability of data and differing personal views

of decision-makers on future returns) affect estimation and lead to the so-called estimation risk in portfolio selection (Mulvey and Erkan 2003; Bawa et al. 1979). Estimation risk has been shown to be the source of erroneous decisions. As pointed out in Ceria and Stubbs (2006) and Cornuejols and Tütüncü (2007), the composition of the optimal portfolio is very sensitive to estimates of the moments of the return distribution, and minor perturbations in these estimates can result in the construction of different portfolios.

Broadie (1993), Chopra and Ziemba (2011) and Ceria and Stubbs (2006) show that estimation risk is due mainly to errors in estimating the mean of the return distribution. Hence, the focus of this analysis is the estimation risk of expected returns (as in Bonami and Lejeune, 2009) rather than the variance-covariance matrix of returns (as in Lejeune and Samatlı-Paç 2012). This makes the algorithm proposed by Lejeune and Samatlı-Paç (2012), which is based on the reformulation of estimation risk of the variance-covariance matrix, unsuitable for the problem under consideration. This issue will be discussed further in Section 3.4. The error in estimating expected returns has attracted renewed interest, and several approaches to incorporating it into portfolio selection have recently been developed. As in Bonami and Lejeune (2009), I adopt Roy's (1959) safety first risk criterion, which identifies as optimal the portfolio for which the probability of its return falling below a prescribed threshold is minimized, to incorporate the estimation risk of expected returns.

$$P\left(\sum_{j=1}^3 \sum_{k=1}^3 \xi_{kj}' \mathbf{w}_{kj1} \geq R_{min}\right) \geq p_{min}$$

The constraint ensures that total expected return, $\sum_{j=1}^3 \sum_{k=1}^3 \xi_{kj}' \mathbf{w}_{kj1}$, exceeds a prescribed minimal level R_{min} with a minimal probability p_{min} . The account returns (income and capital gains), $\xi_{k1} = \mathbf{d}_e + \mathbf{g}_e$, $\xi_{k2} = \mathbf{d}_b + \mathbf{g}_b$, $\xi_{k3} = \mathbf{d}_c + \mathbf{g}_c$, which are multiplied by the decision variables \mathbf{w} , are stochastic and not necessarily independent

across accounts. The constraint, however, requires transformation prior to incorporation in the model. First, for simplicity, the assumption of Bodnar and Schmid (2007) that the gross return follows a normal distribution, $N(\mu, \sigma^2)$, where μ is expected value and σ^2 is variance, is used. However, other, perhaps skewed, distributions are also possible (Bonami and Lejeune 2009). Second, since $\sum_{j=1}^3 \sum_{k=1}^3 \boldsymbol{\mu}_{kj}' \mathbf{w}_{kj1}$ is the average return, μ , and $\mathbf{w}_1' \boldsymbol{\Sigma} \mathbf{w}_1$ is equal to the variance σ^2 , where \mathbf{w}_1 is a vector $(\mathbf{w}_{111}, \mathbf{w}_{211}, \mathbf{w}_{311}, \mathbf{w}_{121}, \mathbf{w}_{221}, \mathbf{w}_{321}, \mathbf{w}_{131}, \mathbf{w}_{231}, \mathbf{w}_{331})$ of all asset weights at period 1, and $\boldsymbol{\Sigma}$ is variance-covariance matrix of all assets' rate of return, the normalized portfolio returns can be calculated as

$$\varphi \equiv \left(\sum_{j=1}^3 \sum_{k=1}^3 \boldsymbol{\xi}_{kj}' \mathbf{w}_{kj1} - \sum_{j=1}^3 \sum_{k=1}^3 \boldsymbol{\mu}_{kj}' \mathbf{w}_{kj1} \right) / \sqrt{\mathbf{w}_1' \boldsymbol{\Sigma} \mathbf{w}_1}$$

where $\boldsymbol{\mu}_{k1} = \bar{\mathbf{d}}_e + \bar{\mathbf{g}}_e$, $\boldsymbol{\mu}_{k2} = \bar{\mathbf{d}}_b + \bar{\mathbf{g}}_b$ and $\boldsymbol{\mu}_{k3} = \bar{\mathbf{d}}_c + \bar{\mathbf{g}}_c$. It follows that

$$\begin{aligned} P\left(\sum_{j=1}^3 \sum_{k=1}^3 \boldsymbol{\xi}_{kj}' \mathbf{w}_{kj1} \geq R_{min}\right) &= P(\varphi) \\ &\geq R_{min} - \sum_{j=1}^3 \sum_{k=1}^3 \boldsymbol{\mu}_{kj}' \mathbf{w}_{kj1} / \sqrt{\mathbf{w}_1' \boldsymbol{\Sigma} \mathbf{w}_1} \\ &= 1 - F_{(w)}\left(R_{min} - \sum_{j=1}^3 \sum_{k=1}^3 \boldsymbol{\mu}_{kj}' \mathbf{w}_{kj1} / \sqrt{\mathbf{w}_1' \boldsymbol{\Sigma} \mathbf{w}_1}\right) \end{aligned}$$

where $F_{(w)}$ is the cumulative probability distribution of the normalized portfolio return and $F_{(w)}^{-1}$ is its inverse. The probabilistic constraint is thus transformed into the following deterministic equivalent:

$$\begin{aligned} 1 - F_{(w)}\left(R_{min} - \sum_{j=1}^3 \sum_{k=1}^3 \boldsymbol{\mu}_{kj}' \mathbf{w}_{kj1} / \sqrt{\mathbf{w}_1' \boldsymbol{\Sigma} \mathbf{w}_1}\right) &\geq p_{min} \\ \Leftrightarrow F_{(w)}\left(R_{min} - \sum_{j=1}^3 \sum_{k=1}^3 \boldsymbol{\mu}_{kj}' \mathbf{w}_{kj1} / \sqrt{\mathbf{w}_1' \boldsymbol{\Sigma} \mathbf{w}_1}\right) &\leq 1 - p_{min} \\ \Leftrightarrow \sum_{j=1}^3 \sum_{k=1}^3 \boldsymbol{\mu}_{kj}' \mathbf{w}_{kj1} + F_{(w)}^{-1}(1 - p_{min}) \sqrt{\mathbf{w}_1' \boldsymbol{\Sigma} \mathbf{w}_1} &\geq R_{min} \end{aligned} \quad (3.28)$$

Note that constraint (3.28) also incorporates the variance and, hence, can be used to replace the basic variance risk constraint in the portfolio optimization model.

In the applications, $R_{min} = R_f$, the risk-free rate, is set. Confidence level, p_{min} , is then set to control the estimation risk to being below a certain level. In other words, the probability that actual return of the asset is below the estimated one is controlled under $1 - p_{min}$. This confidence level can then be modified according to the individual investor's risk preference. Further, as only extreme poor performance needs to be considered, a one-tail test is used. For instance, if 95% confidence is required, then p_{min} is set at 0.95 and $F_{(w)}^{-1}(1 - p_{min}) = -1.65$. Osorio et al. (2004a) incorporate return uncertainty through a scenario tree, but do not explicitly minimize risk, and this may lead to risky trading strategies. Osorio et al. (2004b, 2008a,b), however, explicitly minimize risk, but in stochastic quadratic programming problems. In this chapter, estimation risk rather than the basic variance risk is controlled by the probabilistic constraint to remain under a certain level. Note that (3.28) can also be interpreted as a constraint that imposes a pre-specified risk-return tradeoff, where the (negative) term $F_{(w)}^{-1}(1 - p_{min})$ measures this tradeoff in a manner similar to 'risk tolerance' of a utility function in mean-variance space. Specifically, it is the marginal rate of substitution between risk and return, and is allowed to vary according to the risk tolerance of the investor.

3.3.4 Single-period MINLP optimization

In contrast to the model of Bonami and Lejeune (2009), which minimizes variance, my objective is to maximize portfolio net (of tax) terminal wealth while holding estimation risk below a given level. This objective function resonates more clearly with investors who want to earn as much risk-adjusted net returns as possible while holding estimation risk below a certain level. It is important to mention that in Bonami and Lejeune (2009) and Lejeune and Samatlı-Paç (2012), portfolio variance risk is also set as the objective function. This is a convex function of the control variables (asset weights), and

perturbations in these variables map directly to the objective function (i.e., variances and covariances do not change with asset weights). With taxation, however, this direct mapping breaks down, and the objective function of terminal wealth has to be re-evaluated every time the control variables are perturbed, since tax liabilities differ for different asset weights. As a result, the method developed in Bonami and Lejeune (2009) and Lejeune and Samatlı-Paç (2012) is not applicable to my proposed optimization problem. This is expanded upon in Section 3.4.

The estimation risk constraint (with normally-distributed returns) provides an upper bound on portfolio risk. This bound will be lower (higher) when the required confidence level, p_{min} , is higher (lower). The proposed model will return the portfolio at the intersection of the boundary and the efficient frontier. In addition, the change of the objective function increases the sensitivity of portfolios to changes in tax rates. This enables us to better investigate the effects of tax on investment portfolios.

Combining all constraints, the optimization model for individual investors is formulated as the following MINLP problem:

$$\begin{aligned}
& \text{Maximize } w \quad \sum_{j=1}^3 \sum_{k=1}^3 V_{kj} \\
& \text{Subject to} \quad \text{constraints (3.1)–(3.28);} \\
& \quad V_{kj} \geq 0, k, j = 1, 2, 3; \\
& \quad X_{kj2}, R_{kj1} \geq 0, k, j = 1, 2, 3; \\
& \quad w_{kj1}, \mathbf{i}_{kj}^b, \mathbf{i}_{kj}^s, \mathbf{I}_{kj} \geq 0, k, j = 1, 2, 3; \\
& \quad \mathbf{h}_{kj}^1, \mathbf{h}_{kj}^2, \mathbf{h}_{kj}^3 \geq 0, k, j = 1, 2, 3; \\
& \quad y_{kj}, \boldsymbol{\delta}_{kj} \in \{0, 1\}, k, j = 1, 2, 3;
\end{aligned}$$

3.4 Solution Method

Initially, a non-linear B&B (BONMIN version 1.5) solver in OptiToolbox v.1.34, (MATLAB version) is used to solve the MINLP problem in Section 3.4. This, however,

is found to take a prohibitive amount of time, with no solution returned in over ten hours. Algorithms for solving MINLP problems are often based on relaxation schemes. For the standard mean-variance portfolio problem, different approaches based on nonlinear B&B algorithms (Bonami et al. 2008; Bonami and Lejeune 2009) and outer approximations (Lejeune and Samatlı-Paç 2012) have been proposed. However, the increased complexity of my optimization problems due to the inclusion of taxes, together with probabilistic returns and the large number of assets (up to 288), limits the use of these algorithms. The tax withdrawal rules necessitate the re-evaluation of the entire objective function and its constraints (3.1)–(3.28) at every iteration in which the control variables, \mathbf{w}_1 , are perturbed, or integrality restrictions on the integer variables restored. In Bonami and Lejeune (2009), portfolio variance is the objective function, and their integer variable scoring process depends on a function of the specific contribution of each variable to the overall risk of the portfolio. This is estimated through the Lagrangian function, and for their simple variance-minimizing objective function there is a direct link (mapping) between the control variables (asset weights) and the objective function (portfolio variance). This allows them to calculate the effect of a small change to the control variables with two simple equations (22 in their paper). In my model, however, this is not possible since a change in the control variables does not map directly to the objective function. Changes in asset weights lead to complex changes in tax and this in turn changes the total post-tax return used in the objective function. Consequently, an extension to the proposed solution of Bonami and Lejeune (2009) is required to make it applicable to the proposed large scale MINLP problem in this chapter. A further study on how sensitive an investor's optimal portfolio is to tax rates then can be completed.

A solution method based on Greedy heuristics with a new evaluation and ranking procedure is introduced. More details on this new procedure are given later in this

section. As optimality is not guaranteed with Greedy heuristics, a comparison is carried out and the optimality gaps against BONMIN (B&B) and CPLEX (approximation method) are reported.

In theory, the basic B&B algorithm (BONMIN) always returns the global optimal solution for all MINLP problems. However, in experiments, it is time consuming and has to be terminated before completion, returning an upper bound on the optimality gap, particularly for large-scale applications. One of the objectives of the proposed new solution method is to reduce the computing time for large-scale portfolio optimization to find an approximated solution that, in most cases, has an optimality gap no more than 2% against the solution returned by applying the basic B&B algorithm. Another objective is for this method to be flexible enough to solve large-scale post-tax portfolio optimization problems, not only the proposed model of Bonami and Lejeune (2009) but also the proposed model of this work.

First, the proposed MINLP problem in Section 3.3.4 is decomposed into two independent sub-problems, an integer programming problem and a non-linear programming problem, by non-linear relaxation. The non-linear component of the probabilistic constraint (3.28) is relaxed (removed).

$$\sum_{j=1}^3 \sum_{k=1}^3 \mu_{kj} w_{kj1} \geq R_{min} \quad (3.29)$$

The proposed new solution method based on the Greedy heuristics is used to solve the integer component and the interior point solver Ipopt (Gondzio and Grothey 2007) is used to solve the non-linear component.

The classical [basic] Greedy heuristic is an algorithm that makes the locally optimal choice at each stage approximate to the global optimum (Cormen et al., 1990). This method is applicable to the proposed portfolio optimization problem. The main difficulty in solving the problem is the large number of binary variables in constraints (3.3) – (3.5) and (3.13) – (3.14). By implementing Greedy heuristics, the sensitivity of

the objective function to each binary variable is evaluated and ranked, choosing the one with the highest sensitivity for valuation (giving the binary variable a value of 0 or 1) and then removing the binary variable from subsequent iterations. The algorithm repeats this process until all binary variables are removed and an optimal integer solution is returned.

As the optimal solution to each smaller instance of integer programming problem will provide an immediate output, the proposed algorithm doesn't consider the larger problem as a whole.¹⁵ Once a decision has been made, it is never reconsidered. Consequently, it reduces the number of iterations and therefore the computing time. However, it makes choices dependent only on choices made so far and not on future choices (non-anticipatory). It iteratively makes one greedy choice after another, reducing each given problem into a smaller one. This property makes the quality of solution highly sensitive to the ranking process. In fact, the order of allocation that depends on the ranking process determines whether the algorithm can return a global optimal solution. A proper ranking process is, therefore, required. This is the first major issue that needs to be solved. Further, to implement the Greedy heuristics for the proposed MINLP problem, the iterating process needs to be reprogrammed to make it suitable for the portfolio optimization problem, to deal with non-linear relaxation and to solve special cases when enumeration is pruned (stopped) by infeasibility. Solutions to these issues are described below.

3.4.1 Improved Greedy for portfolio optimization

The main Improved Greedy (IG) program is described in pseudo code in Table 3.2. The description focuses on the iteration and ranking processes (inserted comments are

¹⁵ In each iteration, only one binary variable is considered and solved, which is a smaller instance of the original integer programming problem.

preceded by the percentage sign, '%').¹⁶ The program starts by relaxing integer constraints only, thus reformulating the problem as a non-linear optimization (NonOpt). It then also relaxes the non-linear constraints and reformulates the problem as a linear optimization (LinOpt). LinOpt is solved first, returning an optimal objective function value (LinObj) of the linear problem (pseudo code line 03 in Table 3.2).¹⁷ All integer variables are then gathered into an 'Unsolved' set (line 04, Table 3.2). The program then calculates an 'impact score' (explained in the next paragraph) for each 'Unsolved' integer variable and selects the variable with the highest impact score (lines 05-12, Table 3.2). This is the variable that would be selected for integer restoration at this iteration. LinOpt is then solved twice, once when this selected variable is assigned an integer value of 0, and once when it is assigned an integer value of 1. The integer value that returns a higher solution is then assigned to the variable as its integer restoration at this iteration (lines 13-17, Table 3.2). The variable is then deemed 'Solved' and is removed from 'Unsolved'. Fig. 3.1 is a graphical depiction of the integer restoration process of lines 04-20 of Table 3.2. The program then moves forward to the next iteration and picks another 'Unsolved' variable for integer restoration (line 19, Table 3.2). This process is repeated until all 'Unsolved' integer variables are 'Solved' and their integer restrictions restored (line 20, Table 3.2). All integer solutions 'Solved' are then fed into the original MINLP problem and a final optimal solution is obtained by the interior point optimizer (Ipopt) in MATLAB v.7 (line 21, Table 3.2).

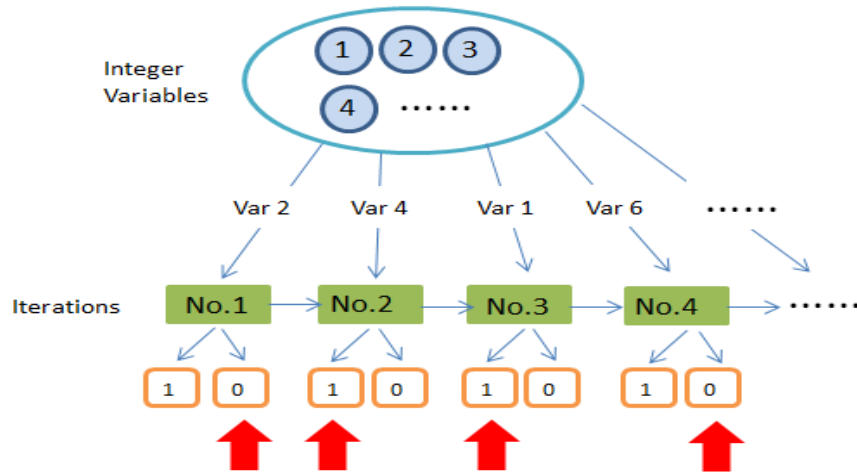
¹⁶ As the exact code in MATLAB v.7 is too long, only the pseudo code of the major idea is presented in Tables 3.2, 3.3, 3.4, and 3.5. Investors can apply it in MATLAB with minor changes.

¹⁷ For linear optimization we use `lp_solve` in OPTI Toolbox v.1.34 on MATLAB.

Table 3. 2 Main code of algorithm

```
01 Relax integer constraints only and record as "NonOpt"
02 Relax both non-linear and integer constraints and record as "LinOpt"
03 LinObj=Opti(LinOpt)
    % return optimal solution of LinOpt
04 Group all integer variables into set "Unsolved"
05 iteration=1
06 While iteration<=N
07 for i=1:N
08 if i∈Unsolved
09 Impact(i)=Gains(Var(i), Solved, LinOpt)
    % compute the impact of integer restoration of variable i
10 end
11 end
12 j=find(Max(Impact))
    % return the variable with the highest impact
13 if Opti(LinOpt, Solved, Var(j)=1)<=Opti(LinOpt, Solved, Var(j)=0)
14 Solved(j)=0
15 else
16 Solved(j)=1
17 end
18 Remove j from Unsolved
19 iteration=iteration+1
20 end
21 Obj=Opti(NonOpt, Solved)
    % return final solution with given integer values
```

Figure 3.1 Process of integer allocation



In ranking 'Unsolved' integer variables for selection (lines 08-10, Table 3.2) a formula similar to that used by Linderoth and Savelsbergh (1999) is adopted to evaluate how the restoration of the integrality condition impacts (decreases) the objective function of the portfolio optimization problem. This formula is $Impact(i) = (LinObj - Max(i)) + 2 * (LinObj - Min(i))$, where $Max(i)$ is the higher of two LP solutions of LinOpt obtained when the unsolved integer variable is assigned a value of 0 or 1. $Min(i)$ is the second (lower-value) solution.

3.4.2 Improved Greedy for infeasibility (Improved Greedy)

In theory, if no $Impact(i)$ could be returned for some variables, then the iterations would stop. This may occur when the integer restoration (at 0 or 1) for some integer variables renders the remaining LP problem infeasible. This issue affects the evaluating and ranking process and reduces the algorithm's precision and efficiency. To deal with it, a supplementary code is added for the two possible cases of infeasibility: one-sided and two-sided. One-sided infeasibility occurs when only one of the integers (0 or 1) assigned to an integer variable renders the remaining LP problem infeasible. Two-sided

infeasibility is rare and would occur if both integers (0 and 1) that could be assigned to a certain integer variable render the problem infeasible.

One-sided infeasibility is dealt with by switching the integer value for the problematic variable. If the LP problem is feasible when the variable is valued at 1 (or 0) but infeasible if it is valued at 0 (or 1), then the variable is integer valued at 1 (or 0) directly before the ranking process. This variable is then recorded as 'Solved' and removed from 'Unsolved'. The algorithm then continues the current iteration by evaluating and ranking all other 'Unsolved' variables. In this manner no more future iteration is required for this variable, and the resulting reduction in the total number of iterations improves the algorithm's efficiency. Table 3.3 describes the supplementary pseudo code and Fig. 3.2 provides an example of this process.¹⁸ Fig. 3.2 shows that while restoring the integrality restriction for a selected variable 2 in the first iteration, the algorithm finds variable 5 to be infeasible at the integer value of 1 but feasible at 0. It therefore chooses to evaluate variable 5 at 0 and moves it to 'Solved'. This is considered as a solution to an 'iteration' that takes precedence and is, thus, recorded as iteration 1. The algorithm then continues to solve the integrality restriction of variable 2 as iteration 2.

¹⁸ The digits on the left hand side of Table 3.3 are code line numbers relating to those in Table 3.2. For example, line 08.01 is line 01 of supplementary code to be inserted after line 08 in Table 3.2.

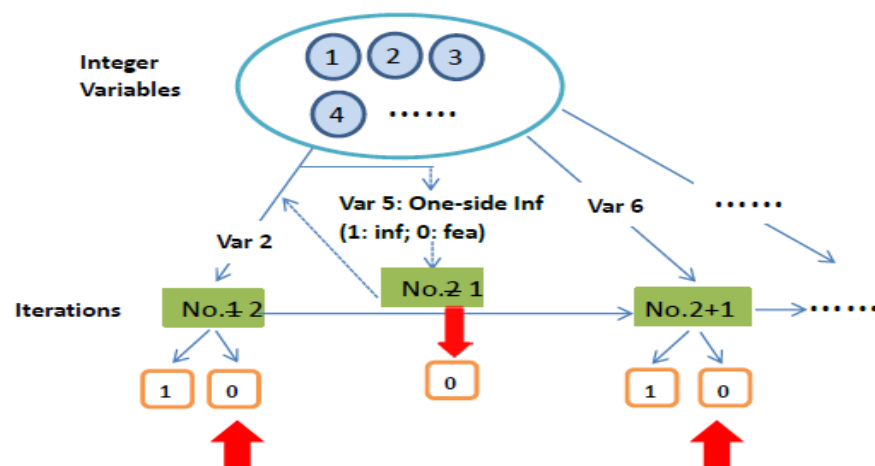
Table 3. 3 Additional code for one-sided infeasibility

```

08.01 inf=0
08.02 if Opti(LinOpt, Solved, Var(i)=1)==infeasible
    % Check whether the selected variable is feasible at 1 for linear problem
08.03 solved(i)=0
08.04 inf=inf+1
08.05 end
08.06 if Opti(LinOpt, Solved, Var(i)=0)==infeasible
    % Check whether the selected variable is feasible at 0 for linear problem
08.07 solved(i)=1
08.08 inf=inf+1
08.09 end
08.10 if inf==0
    % Two-sided feasibility
09.01 elseif inf==1
    % One-sided infeasibility only
09.02 iteration=iteration+1
09.03 Remove i from Unsolved
09.04 end

```

Figure 3. 2 Process for one-sided infeasibility (inf: infeasible; fea: feasible)



A more involved procedure is adopted to deal with two-sided infeasibility. If in the evaluation process the algorithm finds a variable infeasible at both 0 and 1, it reverts to the prior iteration (to the previous 'Solved' variable). It then switches the integer restoration of that iteration's variable to 0 (or 1) if it had been previously 'Solved' or restored to 1 (or 0) and treats the integer 1 (or 0) as 'cancelled' for this variable. If, however, the algorithm attempts to switch the integer to an already 'cancelled' value, it then reverts further back one iteration and repeats the check on the variable of that iteration. This process continues until an integer restoration is found to solve the infeasibility. All integer allocations and references to subsequent iterations will then be cleaned, and the algorithm is allowed to resume its iterations. In the extreme case where all previous iterations have 'cancelled' integers due to infeasibility, an error will be returned by the algorithm and the original MINLP problem is considered infeasible. Table 3.4 presents the relevant supplementary pseudo code, and Fig. 3.3 depicts an example of this idea. In Fig. 3.3, variable 6 at iteration 4 is found to be infeasible at both 0 and 1 (two-sided infeasibility). The algorithm clears this iteration and reverts back to the preceding iteration, 3. It switches the integer solution of variable 1 but finds the variable infeasible with this switch. The algorithm then clears this iteration and reverts back to iteration 2. It finds variable 4 not previously tested at integer 0 (as it had been previously 'Solved' at 1). It changes the integer restoration of this variable to 0 and refers to the previous solution of 1 as 'cancelled'. The algorithm then resumes the 'forward' iteration process. This improves the algorithm's precision by preventing it from returning an incorrect solution that the model is infeasible.

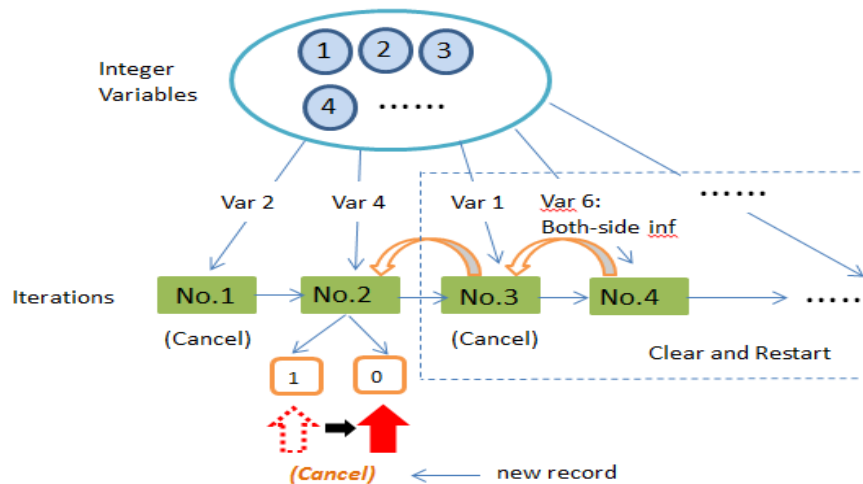
Table 3. 4 Additional code for two-sided infeasibility

```

09.03.01 elseif inf==2
    % Two-sided infeasibility
09.03.02 Revert back to the closest iteration which has no record of cancellation
09.03.03 Turn back to the status at the end of that iteration
09.03.04 Assign the opposite integer value to the variable of that iteration
09.03.05 Record this cancellation
09.03.06 Clean all records subsequent to this iteration
09.03.07 Terminate current iteration

```

Figure 3. 3 Process for two-sided infeasibility



3.4.3 Improved Greedy with precision (Improved Precision)

In addition, the effectiveness of the proposed solution may be affected by the initial non-linear relaxation imposed when searching for integer solutions. The relaxation used in the proposed algorithm is simple, and the two sub-problems, integer and non-linear relaxations, are separated completely. However, the gap between these two separate steps leads to some errors in integer valuation, especially when an extremely high confidence level ($p_{min} \geq 99.5\%$) is set to control risk. As these errors may downgrade the algorithm's performance, a more advanced relaxation method is applied to combine these two sub-problems and to improve precision.

This is carried out by adding a new test in the code whenever the variable with the highest impact is found. This test checks whether the integer solution obtained from a certain iteration is feasible and optimal with non-linear restoration. In the test, at the end of each ranking process, the top-ranked variable with non-linear, rather than linearized, constraints is re-optimized to ensure the variable is integer valued properly. In other words, the second evaluation stage within each iteration is carried out using an NLP rather than the LP of the improved Greedy described above. Thus, ranking is carried out by LP and integer evaluation is carried out by NLP. Table 3.5 shows the relevant code. The infeasibility test in the previous section is copied while changing the examined problem from "LinOpt" to "NonOpt". This code is used to eliminate the gap between integer and non-linear programming. Although the coding here is not complicated, this new relaxation method improves the algorithm's quality even though the repeated non-linearity test doubles the computing time.

Table 3. 5 Additional and replacement code to improve precision

```
12.01 inf=0
12.02 if Opti(NonOpt, Solved, Var(j)=1)==infeasible
    %Check whether selected variable is feasible for non-linear problem at 1
12.03 solved(j)=0
12.04 inf=inf+1
12.05 end
12.06 if Opti(NonOpt, Solved, Var(j)=0)==infeasible
    %Check whether selected variable is feasible for non-linear problem at 0
12.07 solved(j)=1
12.08 inf=inf+1
12.09 end
12.10 if inf==2
12.11 Follow the same action as in the case of two-sided infeasibility
12.12 elseif inf==0
13. if Opti(NonOpt, Solved, Var(j)=1)<=Opti(NonOpt, Solved, Var(j)=0)
17.01 end
```

3.4.4 Computational results

Similar to the work of Bonami and Lejeune (2009), the two new methods, the Improved Greedy and the Improved Precision, are compared with BONMIN B&B v.1.5, which is a NLP-based branch-and-bound algorithm from MATLAB/OPTI Toolbox v.1.34, and TOMLAB/CPLEX v.12.1, which is an approximation branch-and-bound approach that utilizes an interior point algorithm to solve second-order cone optimization problems. A similar structure of comparison among all four algorithms is used. For NLP problems, the two new algorithms use the interior point optimizer Ipopt v.3.10 in MATLAB/OPTI Toolbox v.1.34. All empirical work is performed on an IBM X201 with Intel Dual-Core i5 2.4GHz CPU, 3GB of RAM, and running Windows 7 and MATLAB 7.

The investigation on the effect of taxes in Section 3.5 uses actual financial market data sourced from DataStream. More detail on the data is provided in Section 3.5. The

performance comparison of algorithms conducted in this section, however, is carried out by constructing a test bed that uses simulated data to ensure that extreme cases that may not be present in the actual data are considered. The test bed itself consists of 108 portfolio optimization experiments/portfolios/instances divided into 6 size groups by number of assets: 18 instances/portfolios with 9 assets each, 18 instances with 18 assets each, 18 with 36, 18 with 72, 18 with 144, and 18 with 288. In each instance the prescribed minimum return level, R_{min} , is set equal to 5%, the tax rates for assets are all set equal to 40% or 60%, and the prescribed reliability level, p_{min} , is set between 60% and 99%. For simplicity, asset returns are assumed to follow a normal distribution.¹⁹ All problem instances are modeled using the AMPL modeling language. Table 3.6 presents problem statistics of the six portfolio groups. The largest problem considered contains 288 assets, 2604 variables, of which 291 are binary, and 4814 total constraints, of which 874 are non-linear inequalities and 589 are non-linear equalities (linear constraints with binary variables are classified as non-linear).

Table 3. 6 Problem statistics

Total Assets	Total Variables	Total Constraints	Binary Variables	Non-linear Inequality	Non-linear Equality
9	93	191	12	37	31
18	174	341	21	64	49
36	336	641	39	118	85
72	660	1241	75	226	157
144	1308	2441	147	442	301
288	2604	4814	291	874	589

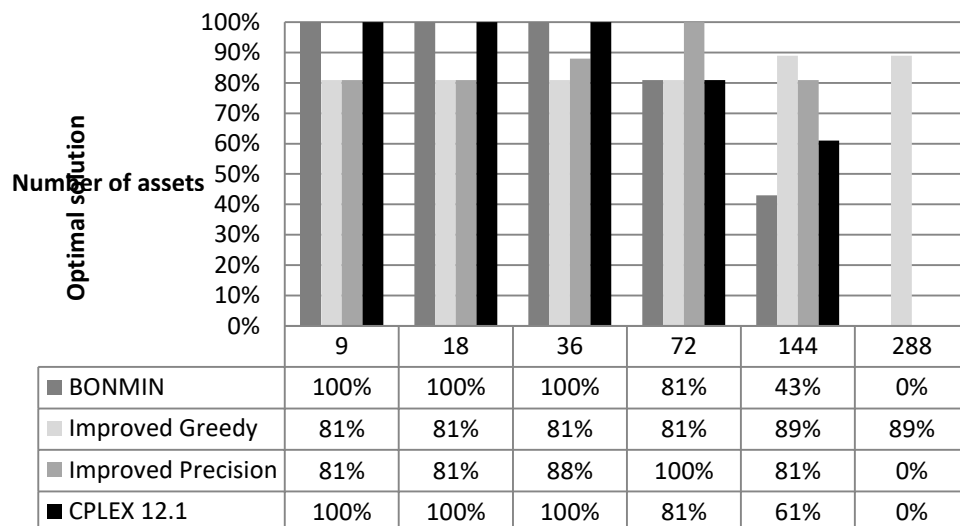
The comparison is conducted on both efficiency and precision, based on computing time and solution's quality, respectively. Bonami and Lejeune's procedure is adopted. The main results are given in Table 3.7. The 'Name' column lists the portfolio

¹⁹ This assumption can be relaxed to perhaps include positively-skewed distributions without altering the nature of the problem (see Bonami and Lejeune, 2009).

instances where one or more algorithm exhibits an optimality gap (digits to the left of the underscore denote the number of assets in a problem instance, and digits to the right denote the number, out of 18, of that portfolio instance). In each problem instance, four optimal solutions are returned by four algorithms. One solution to a particular problem instance is defined as 'high quality' if it is not worse than any of other three solutions. For these high quality solutions, there is a zero optimality gap (recorded as “*” in the columns of Table 3.7 entitled “Optimality Gap”). So, a solution to a particular problem instance is defined as 'low quality' if it exhibits a non-zero optimality gap relative to the high-quality solution of that problem instance. In Table 3.7 these are reported as a percentage, or denoted by 'NS' or 'INF'. A reported percentage figure measures how much lower the optimal objective function value of a low quality solution is relative to that obtained by a high quality solution. 'NS' denotes no solution returned within 9 hours, and 'INF' denotes an infeasible problem. Fig. 3.4 provides a summary of the algorithms solution quality reported in Table 3.8. The vertical axis is the proportion of experiments in which each respective algorithm returns a high-quality solution, and the horizontal axis is the number of assets in every set of experiments. BONMIN and CPLEX return high quality solutions in all experiments with 36 assets or less, while my two new algorithms return a high quality solution for around 81% of experiments of this size. The Improved Greedy exhibits a maximum optimality gap of 8% and an average of only 3.3% and when p_{min} of the stochastic opportunity constraint is set at a high level of 95% (Table 3.7). It returns an INF only in cases when p_{min} is set at the extremely high level of 99%. Such a high confidence level for stochastic risk enlarges the gap between integer and non-linear programming under linear relaxation and the algorithm will provide a wrong solution. In these two cases, the Improved Precision outperforms the Improved Greedy but still underperforms BONMIN and CPLEX. But when the number of assets increases beyond 36, a sudden decline is observed in the relative performance

of BONMIN and CPLEX. With 72 assets, the Improved Greedy retains its performance at 81%. BONMIN and CPLEX match this performance, but the Improved Precision outperforms all three algorithms, returning a high quality solution in 100% of cases. With 144 assets BONMIN's and CPLEX's performance decline to 43% and 61%, respectively, while the Improved Greedy performs at 89% and the Improved Precision performs at 81%. With 288 assets, only the Improved Greedy returns solutions, and it does so at its consistent level of performance of above 89%.

Figure 3. 4 Solution quality: percentage of instances in which a high-quality solution is returned within nine hours

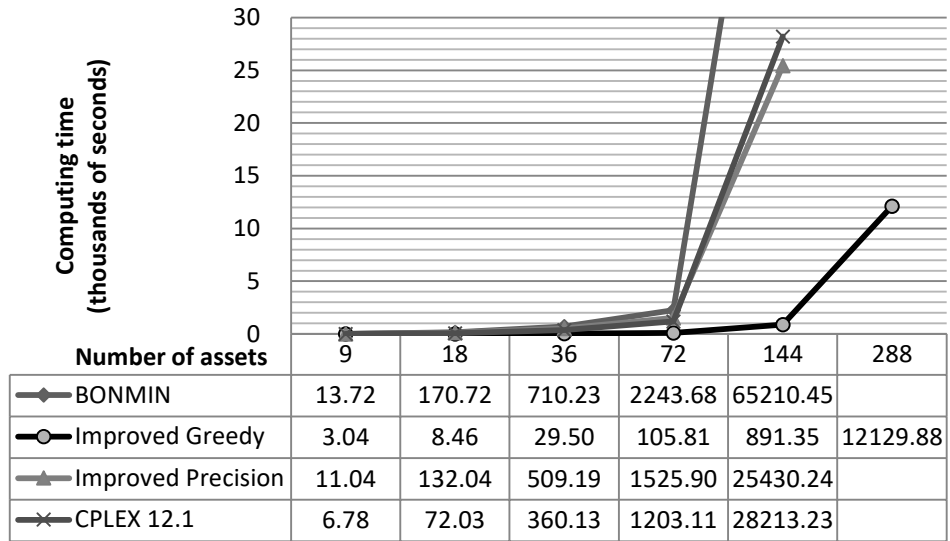


Thus, although the Improved Greedy does not attain the quality of BONMIN and CPLEX for small portfolio problems, it consistently provides high quality solutions, even for large portfolio problems, in over 81% of cases. Similar to Bonami and Lejeune (2009) and Lejeune and Samatlı-Paç (2012), my results show that neither BONMIN's B&B nor CPLEX are efficient for large-scale portfolio optimizations. The solution quality of the former decreases sharply from 100% to 0% as the number of assets increases from 36 to 288, mainly because time requirements terminate its iterating process before convergence, and no solution is returned. The solution quality of CPLEX

also decreases with increasing number of assets. This is because the general approximation method can only return a local optimal solution in a reasonable period of time as the number of integers increase. This holds an optimality gap from the solution found by the new algorithm.

Table 3.7 summarizes the quality of all solutions against number of assets for the six groups of experiments. Fig. 3.5 presents the average computing time expended in finding a high quality solution (only). An upper bound on computing time is set at 32400 seconds (9 hours). The figure shows that all four algorithms return a high-quality solution in a reasonably short time when no more than 72 assets are considered in the optimization. When this number doubles to 144, however, differences between the algorithms become apparent. Improved Greedy retains its efficiency while the time for the other three methods, particularly for BONMIN's B&B, increases dramatically. When the number of assets is increased to 288, only the Improved Greedy is able to return a solution and with an average computing time of 12,129.88 seconds (3.4 hours), which would normally be acceptable for a tax or a portfolio balancing exercise. None of the other algorithms could return a solution within an extended upper limit of 72,000 seconds (20 hours).

Figure 3. 5 Speed and scalability: average computing time for a high-quality solution



In summary, similar to the conclusion reached by Bonami and Lejeune (2009), a dynamic iterating process can solve large-scale portfolio optimizations under MINLP more efficiently than general B&B and approximation methods. The dynamic process is redesigned in a Greedy framework and tailored to more complex post-tax portfolio optimization problems. The computational results lead to the conclusion that the new solution method, Improved Greedy, is more reliable and efficient than BONMIN's B&B and CPLEX for the large-scale portfolio problems. Investors can obtain a high-quality solution within a reasonable period of time in most instances using PCs.

Table 3. 7 Computational results for instances showing an optimality gap

			BONMIN	CPLEX	Improved	Improved				BONMIN	CPLEX	Improved	Improved
			B&B	12.1	Greedy	Precision				B&B	12.1	Greedy	Precision
Name	Confidence	Tax	Optimality	Optimality	Optimality	Optimality	Name	Confidence	Tax	Optimality	Optimality	Optimality	Optimality
	Level	Rate (%)	Gap (%)	Gap (%)	Gap (%)	Gap (%)		Level	Rate (%)	Gap (%)	Gap (%)	Gap (%)	Gap (%)
	(p_{min} , %)							(p_{min} , %)					
9_15	95	40	*	*	7.1	4.2	144_16	95	60	NS	NS	*	NS
9_16	95	60	*	*	8.0	6.3	144_17	99	40	NS	NS	INF	NS
9_17	99	40	*	*	INF	9.8	144_18	99	60	NS	NS	INF	NS
9_18	99	60	*	*	INF	15.1	288_01	60	40	NS	NS	*	NS
18_15	95	40	*	*	2.5	0.5	288_02	60	60	NS	NS	*	NS
18_16	95	60	*	*	4.1	0.9	288_03	65	40	NS	NS	*	NS
18_17	99	40	*	*	INF	6.5	288_04	65	60	NS	NS	*	NS
18_18	99	60	*	*	INF	9.1	288_05	70	40	NS	NS	*	NS
36_15	95	40	*	*	1.4	*	288_06	70	60	NS	NS	*	NS
36_16	95	60	*	*	1.8	*	288_07	75	40	NS	NS	*	NS
36_17	99	40	*	*	INF	5.0	288_08	75	60	NS	NS	*	NS
36_18	99	60	*	*	INF	5.8	288_09	80	40	NS	NS	*	NS
72_15	95	40	*	11.3	0.7	*	288_10	80	60	NS	NS	*	NS
72_16	95	60	*	8.6	1.1	*	288_11	85	40	NS	NS	*	NS
72_17	99	40	NS	15.2	INF	*	288_12	85	60	NS	NS	*	NS
72_18	99	60	NS	6.8	INF	*	288_13	90	40	NS	NS	*	NS
144_11	85	40	NS	9.9	*	*	288_14	90	60	NS	NS	*	NS
144_12	85	60	NS	*	*	*	288_15	95	40	NS	NS	*	NS
144_13	90	40	NS	15.3	*	*	288_16	95	60	NS	NS	*	NS
144_14	90	60	NS	8.8	*	*	288_17	99	40	NS	NS	INF	NS
144_15	95	40	NS	NS	*	NS	288_18	99	60	NS	NS	INF	NS

3.5 Influence of Taxation on Personal Portfolio Management

In this section, influences of tax policy are examined under the set-up outlined in Sections 3.2 and 3.3. Market data are obtained from Thomson Reuters DataStream for these investigations.

3.5.1 Personal portfolio: data

Each of the high-level asset classes, equity, bond (corporate) and commodity segment of the U.K. market, is divided into subclasses. Equities are categorized by industry sector. All corporate bonds currently active in the market are divided into two groups first: investment grade and high yield. Each group is further divided into industrial subclasses (airline, technology, telecommunications ... etc.). Commodities are categorized by product type (e.g., oil, gold, copper, corn ... etc.). This generates 30 classes of U.K. shares, 7 of corporate bonds and 18 of commodities for each account (onshore investment bond, offshore investment bond and unit trust). All these subclasses are represented by a corresponding market index. Historical annual income (yield for bonds and dividend for equities) and capital gains of the market indexes are obtained from Datastream for the period 1990 through 2011. Bond market data (yield and capital gains) are from Barclay's bond index. Equity data (dividend and capital gains) are from FTSE for the U.K., US-DS Price Index for the U.S., and FTSEUR1ST 300 for the E.U. Data of commodities (capital gains) is from S&P commodity index. These are then used to calculate average returns and the covariance of related assets within their respective asset classes. All the relevant settings of the parameters in the model are presented in Table 3.8.

Table 3. 8 Investment settings for portfolio management

Taxation		Amount (%)
Offshore	t_{off}	40
Onshore	t_{an}	22
Onshore	t_{on}	18
unit trust	t_{in}	40
unit trust	G_T	40
unit trust	G_{T+1}	38
Fees		Amount (%)
Annual Fees	m_1 (Offshore)	1
	m_2 (Onshore)	2
	m_3 (Unit trust)	3
Transaction Costs	c	1
Initial Wealth		Amount (%)
New Investment	w_{10}, w_{20}, w_{30} (Portfolio)	0
	C_0 (Cash)	100
Rebalance	w_{10}, w_{20}, w_{30} (Portfolio)	100 (equally distributed)
	C_0 (Cash)	0

Note: t_{in} is a vector of tax rates applied to different underlying asset in the account

3.5.2 Personal portfolio: results

Initially, all tax rates are kept the same as specified in the initial settings. Each tax rate is then gradually varied in each round in order to test its effect on overall portfolio composition.

Figs. 3.6–3.8 show the resulting weights (change in wealth) at different tax rates for equities, corporate bonds and commodities, respectively. Figs. 3.10–3.12 show the resulting weights (change in wealth) at different tax rates for offshore bonds, onshore bonds and unit trusts, respectively. The horizontal dotted lines show the cases when optimization ignores taxes. The word 'new' denotes a set of applications with initial (beginning-of-period) wealth being held in cash and a new portfolio constructed by the

investor through optimization. In Fig. 3.9, the word 'existing' denotes a bequest portfolio that requires rebalancing at the beginning of the investment horizon.

Some important observations can be made from Figs. 3.6–3.8. First, the inclusion of taxes in portfolio optimization has a large effect, on average, on portfolio composition (asset weights). This general finding is consistent with previous research. What this analysis adds, however, is that when the effective tax rate is similar across assets (e.g., 40%), little or no difference is observed between the results of post-tax and pre-tax optimizations, but this difference changes dramatically when taxes for different products vary around 40%. Accordingly, the optimal investment strategy is sensitive to different tax rates across products. For example, in the setting when the corresponding tax decreases to 20%, investments in corporate bonds and commodities increase far more rapidly than those in equities. When taxes decrease further to 0%, it is the investments in equities instead, that increase more rapidly. Further, from the shape of the curve in Figs. 3.6-3.8, it is proved that the variation across different tax rates in the optimal weights for commodities is higher than that for corporate bonds and equities. This is a direct result of their relatively higher, on average, return and risk.

Figure 3. 6 Influence of tax on equities

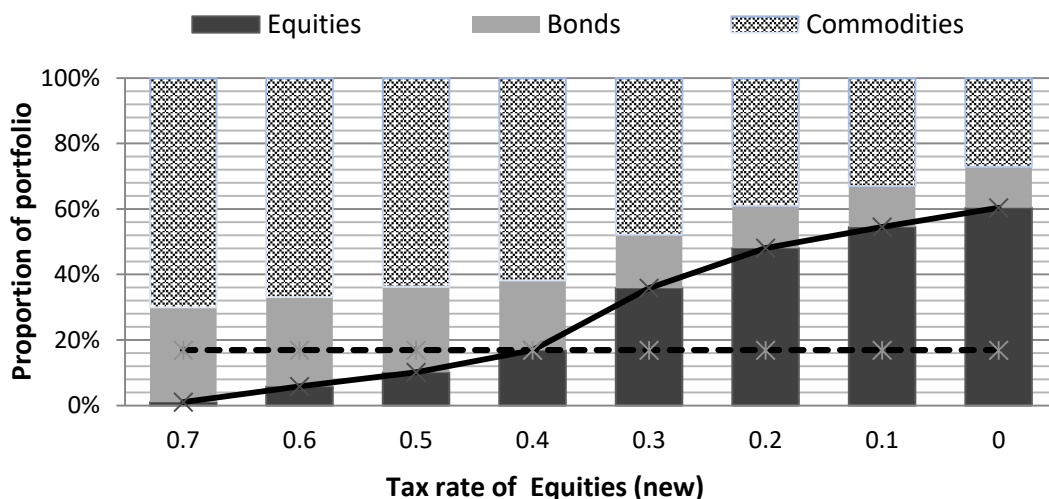


Figure 3. 7 Influence of tax on corporate bonds

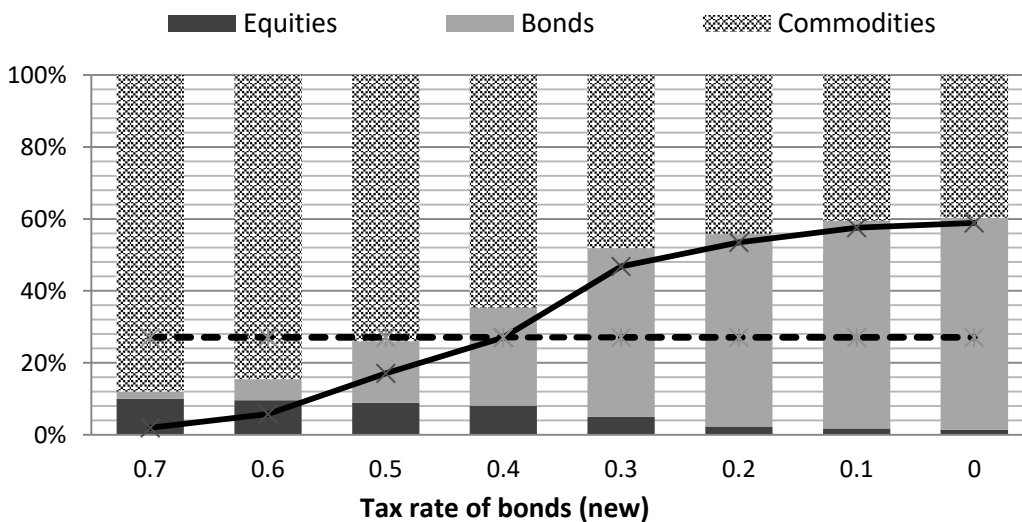
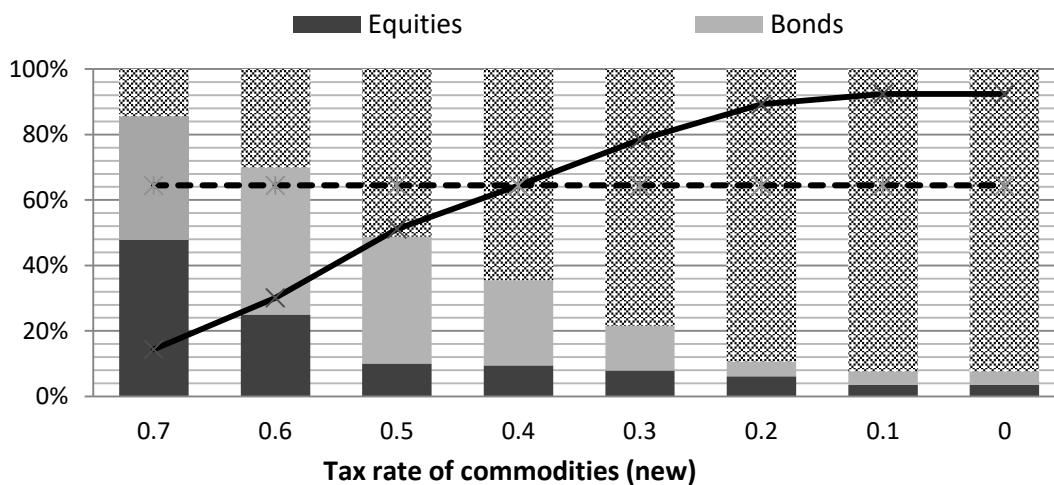
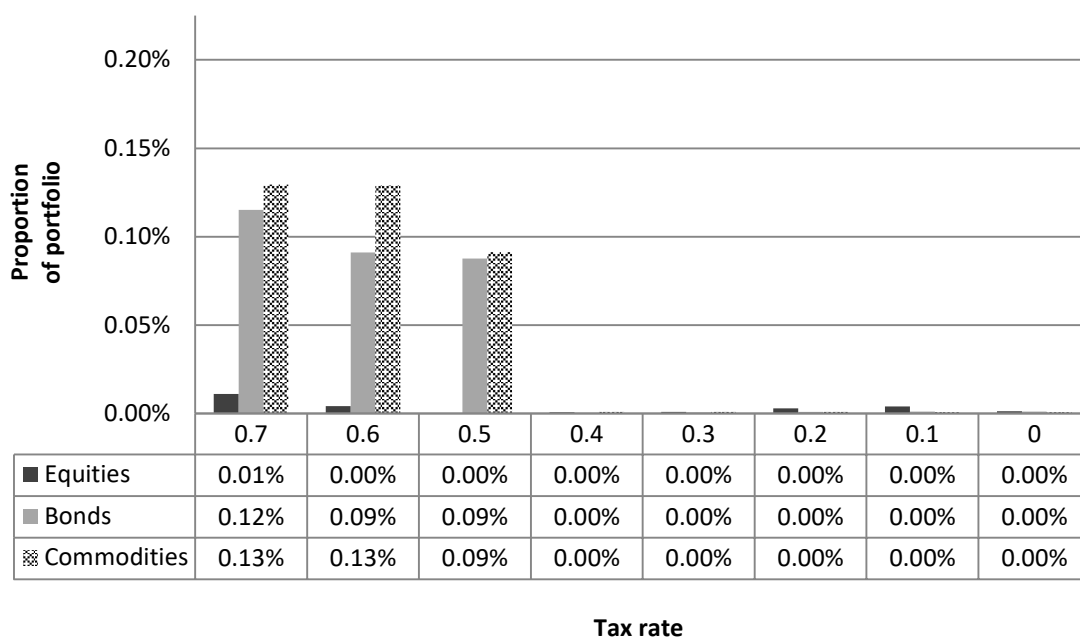


Figure 3. 8 Influence of tax on commodities



Second, Fig. 3.9 shows the difference in optimal portfolio composition between new and rebalanced investments. This represents the difference in weights of the three asset classes between the 'new' and 'existing' portfolios at different tax rates. One main observation from the figure is that the difference is small.

Figure 3. 9 Difference between new and existing portfolio



Finally, Figs. 3.10-3.12 show that the inclusion of different tax rules through investment accounts (onshore and offshore investment bonds and unit trusts) has a very limited effect on this single period portfolio optimization. As the tax rates on asset classes change (in analysis it is assumed that both income and capital gains tax rates on an asset class in each account are perturbed in equal increments), the total investment in each account remains almost the same. For example, in Fig. 3.11 the total investment in the onshore investment bond account remains around 35% no matter how the tax rate on equities changes.

Figure 3. 10 Influence of equity tax on investment account

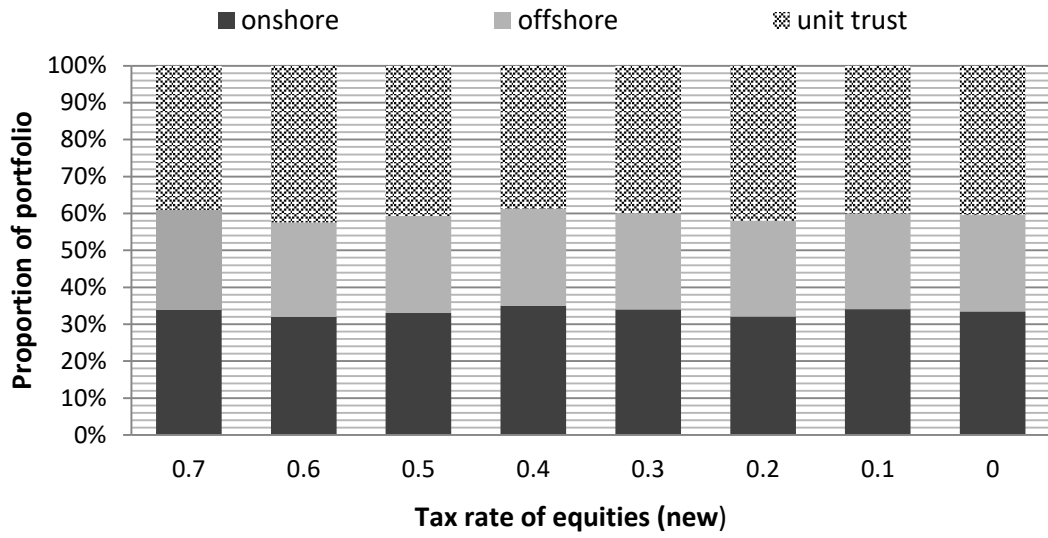


Figure 3. 11 Influence of corporate bond tax on investment account

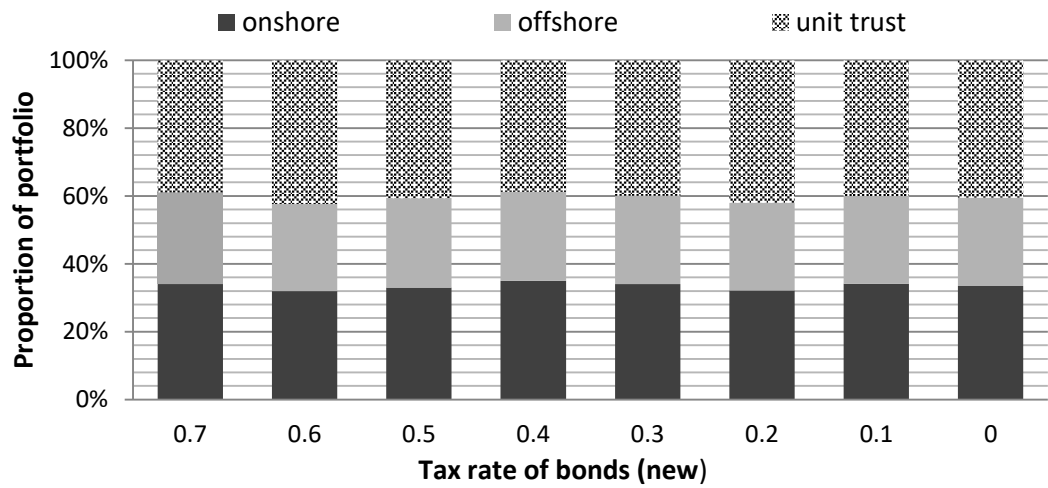
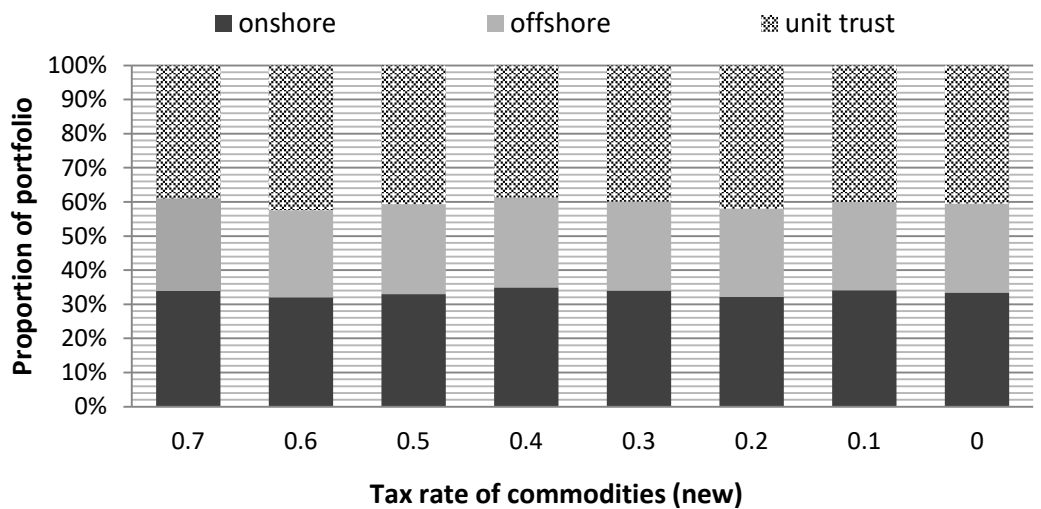


Figure 3. 12 Influence of commodity tax on investment account



3.5.3 Personal portfolio: implications

Some implications can be inferred from the empirical results. First, the weights at different tax rates (Figs. 3.6-3.8), show that taxation plays a more critical role in portfolio optimization for riskier investment assets (e.g. commodities). Second, a decrease in tax rates usually increases the optimal weights of assets, but this relationship is neither linear nor perfectly convex over the entire range of possible tax rates. The implication is that earlier studies that assume a perfectly convex relationship between tax rates and portfolio weights by simplifying tax rules impose simplistic tax consequences and err in their consideration of the true effect of taxes (e.g., Elton and Gruber, 1978).

Further, from Fig. 3.9, given that the two main factors that distinguish these two investments are transaction fees and withdrawal tax, it can be concluded that neither of these factors has a large effect on the final investment strategy. However, a difference exists at high tax rates and rises (especially for corporate bonds and commodities) as tax rates increase. Thus, even in a single-period model, there is a potential loss from instant withdrawal tax when the rate of this tax is high.

Finally, from Figs. 3.10-3.12, it is clear that in a single period optimization (new or rebalance), the long-term advantages of tax rules in onshore and offshore investment bond accounts are ignored. Thus, as the tax rates change for all three asset classes, only the portfolio composition in each account changes, while the total investment in each account remains more or less the same. In conclusion, long-term investors who use multi-stage optimization should consider investment accounts with different tax treatments, but short-term investors who use single-stage optimization can safely exclude them to simplify the model.

In summary, there is evidence that taxation plays a significant role in portfolio management, particularly when tax rates differ across products. Further, the relationship

between tax and optimal portfolio composition is neither linear nor convex. Thus, mathematical programming methods have an obvious advantage over theoretical methods in this area. In addition, withdrawal tax should not be ignored when related tax rates are high, even in a single-period investment horizon. Finally, investors' preference for certain assets is significantly influenced by its tax rate. Accordingly, tax policy setters can use the framework presented to estimate quantitative effects of the change of tax policy to avoid excessive capital outflow of, or excessive demand for, relevant financial products.

3.6 Conclusion

A post-tax portfolio optimization model is developed with integer-based trading constraints. In order to examine the real influence of income tax on portfolio management, a number of realistic trading constraints are incorporated. These include: the need for diversification, requirements on both the number of assets in a portfolio and the maximum holdings in single assets, round-lot buying, and taxation on cash withdrawals peculiar to the specific personal investments considered (the last two are modeled with integer variables). Bonami and Lejeune's (2009) model is used to account for the risk in estimating expected asset returns through a stochastic constraint that ensures the expected return of the portfolio exceeds a pre-specified threshold with a high confidence level. I am unaware of other research that considers so broad a range of trading rules for post-tax portfolio optimization in a single model. Hence, this is a more realistic simulation than prior work, and quantifies better the influence of tax on personal investments.

One key contribution of this chapter is that it innovates on the basic Greedy heuristic, making it available for post-tax portfolio optimization problems in which stochastic risk and realistic market restrictions modeled with integer constraints are

simultaneously considered. The combination of integer and nonlinear constraints explains the complexity involved in solving such problems under large-scale applications for which very few solvers are efficient. The efficacy of the approach on more than 50 problems containing up to 288 assets is evaluated, and the computational results provide evidence of its efficiency in two aspects: precision of solution and required computing time.

Regarding the role of tax in financial markets, it is found that income tax has an obvious impact on portfolio optimization for investors, which supports the conclusion of existing theoretical work. The mathematical programming in this chapter shows that realistic trading constraints, tax rates and portfolio composition have a complex relationship that is neither linear nor convex. Convexity assumptions often made in the literature to guarantee optimality, therefore, are not only unrealistic but also erroneous simplifications. This is the main advantage of this work over prior theoretical research on post-tax portfolio optimization. In addition, in the investigation on effects of withdrawal tax, it is found that for single-period optimization, this factor has a limited influence. Investors can simplify the optimization model by ignoring withdrawal tax without changing the optimal solution significantly. Finally, the analysis shows that investors' preference for a certain asset is significantly influenced by its tax rate. The model is, therefore, useful to tax policy setters.

Chapter 4 - Quantitative Effects of Differential Tobin and Withholding Tax on Global Financial Markets

This chapter investigates the quantitative effects of investment taxes and Tobin tax on capital flows between regional markets. An optimization model is proposed to maximize the expected net Sharpe ratio and find the optimal risky portfolio across tax jurisdictions. Complex trading and tax rules are considered. To examine the effects of different investment and transaction tax rules, a comparison of four tax settings is presented; namely, source only, residence only, mixed with credit, and mixed with double taxation. The experimental results show that a source only tax union has more capital transits between regional markets than a residence only tax union, and its optimal market portfolio is more sensitive to regional tax policy. In a mixed tax system, double taxation between residence- and source-taxed jurisdictions significantly reduces the attraction of the latter, while its attraction is maintained with the credit method. Tobin tax can reduce market volatility but the effect varies with its rate, certain market specifications (e.g., expected returns and correlations with overseas markets), and investment tax rules. This effect does not depend on which side of the capital flow (inflow or outflow) is subject to Tobin tax. Finally, an agreement between countries on a Tobin tax rate (policy) can significantly reduce the negative effect of Tobin tax on capital flows while retaining its positive effect on market stability in comparison to heterogeneous Tobin tax rates.

4.1 Introduction

As explained in 2.2.1, the extant literature on the macro-level impact of tax can be divided into two main categories. One is the analysis of government tax revenues (Saunoris and Payneb, 2010; Creedy and Sanz-Sanzb, 2011). The other, to which this

chapter relates, concerns international tax-driven capital flows, and often compares the two main withholding tax systems: source-based and residence-based. In a source-based tax system, all investments in a country are taxed only by the government of that country, regardless of the residency of the investor. In a residence-based tax system, all investments are taxed only by the investor's country of residence regardless of the investment location.

A comparison between these two tax systems can be made by observing the change in the two components of a country's balance of payments, namely the external current account and the financial account.²⁰ A current account surplus increases a country's net foreign assets whereas a current account deficit reduces them. Bovenberg (1992) explores how residence- and source-based taxes on income affect the external current account in small open economies.

Most research on the impact of tax employs statistical methods (Bovenberg, 1992, 1986; Becher and Fuest, 2008, 2010, 2011; Summers, 1988; Desai and Hines, 2004; Lipton and Sachs, 1983; Mutti and Grubert, 1985; Goulder and Eichengreen, 1989; Bovenberg and Goulder, 1989; and Keuschnigg, 1991), although some work adopts analytical methods, the so-called theoretical research (Becher and Fuest, 2011; Desai and Hines, 2004; Becker and Fuest, 2008; Becker and Fuest, 2010). The statistical method usually assumes a constant correlation coefficient between the investment decision and the corresponding tax factor. This assumption, however, does not always apply in the real world, particularly in a complex tax environment. For example, the results in Chapter 3 show that an increase in tax rate (e.g. 0.1) may lead to a quite different effect on investors' optimal portfolio with different initial rate (e.g. 0.4 to 0.5 and 0.5 to 0.6) even if all other factors are held constant. This proves that two identical

²⁰Current account: the current account consists of the balance of trade, net factor income (earnings on foreign investments minus payments made to foreign investors) and net cash transfers. Financial account: the financial account reflects net change in ownership of national assets.

changes in the tax rate (both increase by 0.1) of an asset class will not lead to the same change on its weight in the optimal portfolio if the initial tax rate is not the same. Therefore the correlation coefficient between the optimal weight on an asset class and the corresponding tax factor (tax rate) is not constant. Furthermore, although the theoretical research is better able to examine the variable sensitivity of investment decisions to the corresponding tax rules, it does not usually employ real market data. Chapter 4 uses a simulation method, mathematical programming. This research method is more appropriate than statistical and analytical methods for many investigations. In the mathematical programming approach, for example, the variable sensitivity of investment decisions to the corresponding tax settings can be investigated. And the mathematical work can use real market data.

In Chapter 4, an optimization model is first constructed to simulate the investment process of investors with different tax brackets, and the change in the aggregate demand of an asset class when its tax rate is cut or enhanced. The model is then used to compare the two withholding tax systems (residence- and source-based) by observing their differential tax-driven capital flows. In contrast to the literature on real industry investments, this thesis focuses on the capital flow for financial market investments. How a source-based tax system influences the volatility of global financial markets is investigated. The thesis also investigates whether a country applying source-based tax would benefit from joining an international tax union with other countries applying residence-based tax and if so, to what extent. Furthermore, whether a change from a pure residence-based tax union to a pure source-based tax union for developed countries would enhance the globalization of financial markets is also investigated.

Tobin tax has been a controversial topic among economists and politicians (e.g., Habermeier and Kirilenko, 2003). Researchers disagree on the consequences of a Tobin-style tax for price volatility and market efficiency. Some argue that such a tax

would improve market efficiency by decreasing price volatility (Ehrenstein et al., 2005; Cipriani and Guarino, 2008) while others argue that it would be detrimental to market efficiency by increasing price volatility (Aliber et al., 2003; Mannaro et al., 2008). Haberer (2006), however, presents a U-shaped relationship between price volatility and market trading volume. The likely reduction in market volume due to the introduction of a Tobin-style tax has different consequences for price volatility depending on relative market volume.

The majority of articles on trading volume conclude that the introduction of a Tobin tax would reduce trading volume by decreasing transactions carried out by speculators (Mannaro et al., 2008; Hanke et al., 2010). However, Hanke et al. (2010) show that the size of this reduction is highly sensitive to market capitalization.

A question which arises from the work of Hanke et al. (2010) is whether the impact of a Tobin tax on trading volume is sensitive to investment tax rules. In the extant literature, the impact of a Tobin tax is considered in isolation, completely separate from other taxes including income tax and capital gains tax. An improved model is proposed to investigate the economic impact of a Tobin tax. The work shows whether the implementation of a Tobin tax will impose an obvious capital locking effect on regional markets. It also shows whether this effect is the same for all countries. Does the effect vary with withholding tax rules (source-based or residence-based), market features (trending market or volatile market), the rate of Tobin tax (consistent or heterogeneous Tobin tax rate globally) and the way that Tobin tax is applied (tax on capital inflows only, capital outflows only or both)? In the simulation, realistic market settings are made by integrating the Tobin tax with other tax issues, such as income tax, capital gains tax and withholding tax.

Similar to the work by Mannaro et al. (2008), an artificial global financial market is set up in chapter 4 to integrate Tobin tax with other tax rules (both income tax and

capital gains tax), which is a significant improvement to the existing literature. The artificial market includes three regional sub-markets, E.U., U.K. and U.S., which are now in an international residence-based tax union, and all tax rates are set to be variable (the rate of the Tobin tax is set in a range of 0% to 1% while the rate of the investment tax²¹ is set in a range of 10% and 70%). The model is then used to test the differential effects of source- and residence-based tax systems on financial market investment. In this way, keeping all else the same, whether a change from source-based tax to residence-based tax and joining the international tax union would benefit a developing country which is currently applying source-based tax system can be investigated. Furthermore, whether a change from a residence-based tax union to a source-based tax union would improve the efficiency of global markets for developed countries is also investigated. The differential impact of the Tobin tax under heterogeneous investment tax rules (both tax rates and withholding tax applications) can also be tested.

In the future, as globalization progresses further, investors may change their residence more frequently than nowadays, and therefore source-based international tax systems will be more convenient (no need to assess an investor's country of residence for tax purposes). For example, under residence-based tax system, if an investor makes an investment in the U.S. market when living there and then becomes a U.K. resident, the investment would be taxed by U.S. government first and then taxed by the U.K. government after the investor moves to the U.K. The change of taxation increases the administrative workload and the complexity of global taxation, particularly when the investor changes his country of residence relatively frequently. Under source-based tax, however, no matter how often the investor changes his country of residence, the investment will always be taxed by the U.S. government, which simplifies global taxation and makes it easier to operate. It is therefore possible that in the future, even

²¹ Investment tax includes income tax and capital gains tax. In experiments, the effective rates of income tax and capital gains tax are always changed simultaneously and referred to as the investment tax rate.

developed countries (e.g. the U.S. or the U.K.) might use a source-based tax system rather than a residence-based tax system to simplify global taxation.

4.2 Capital Asset Pricing Model with Heterogeneous Tax Rules

Trauring (1979) develops a three-fund separation theorem leading to a three-term Capital Asset Pricing formula. The investor's optimal risky portfolio is a linear combination of three identified risky portfolios, G, D, and E, which are independent of investors, their utility function, and tax brackets.²² This independence does not, however, extend to the weights for each identified risky portfolio. These weights are functions of investor i 's income tax and capital gains tax brackets, and will be denoted by a_i , b_i , and c_i . As a result, investors with different tax brackets will have different optimal risky portfolios, Y_i . Trauring presents the following formula to calculate the market portfolio with tax:

$$\sum_{i=1}^n X_i = M = \left(\sum_{i=1}^n a_i \right) * (G\Gamma^{-1}) + \left(\sum_{i=1}^n b_i \right) * (D\Gamma^{-1}) + \left(\sum_{i=1}^n c_i \right) * (E\Gamma^{-1}) \quad (4.1)$$

where M is the so-called market portfolio, X_i is optimal portfolio for investor group i , Γ is the a matrix whose kj^{th} element is the ratio of covariance between asset k and j to variance of market, and n is the number of investor groups. He further proves that the capitalization-weighted sum of all investors' optimal risky portfolios provides the market equilibrium condition. In Trauring (1979), however, the taxation process is still highly simplified. An asset's net return is calculated as $(1-t)*r$ where only fixed tax rates (t) are considered while complex tax rules (annual income tax, deferred capital gains tax, withholding tax on foreign investment, and transaction tax) are not considered. In this chapter, Trauring's work is improved by including more complex and realistic tax settings under mathematical programming.

²² Detailed definition of these three independent risk portfolios can be found in Trauring (1979).

An optimization model is proposed to consider heterogeneous tax rules across investors. To keep the model tractable, all investors are divided into only three residence groups: U.K., Eurozone and U.S. Investors in the same group are assumed to be subject to the same tax rules and hold the same optimal risky portfolio. To determine the market portfolio under taxation, the capitalization-weighted sum of the three local optimal risky portfolios is calculated. Thus:

$$Portfolio_{Market} = \frac{\sum Capitalization_i \times Portfolio_i}{\sum Capitalization_i} \quad (4.2)$$

In equation (4.2), $Capitalization_i$ is the total market capitalization of regional market i . $Portfolio_i$ is the local optimal risky portfolio for regional market i . The total regional market capitalizations are then used as weights to calculate the weighted average optimal portfolio for the global market.

As in Trauring (1979), the market portfolio M is not a unique portfolio that all investors hold. In fact, it is a weighted sum of the optimal portfolios X_i of different investor groups. Nevertheless, it can still be used to provide the composition of the global risky financial market if all investors with different tax brackets hold their own optimal portfolio and are included in the calculation of the aggregated market portfolio.

4.3. Tax Rules

This section outlines the tax setup adopted. Basic investment tax (i.e., annual income tax and tax on realized capital gains) is introduced together with heterogeneous Tobin tax and foreign investment (withholding) tax rules.

First, the basic investment tax rules are set out. Each asset class (bonds, equities and commodities) in each country or region (U.K., U.S. and Eurozone, treating the Eurozone as a single country) is subject to an independent income tax rate and capital gains tax rate, which may differ across asset classes and countries. With regards to

income tax for bonds and equities, all income is assumed to be paid and taxed annually, and all net income is received in cash and can be used to purchase assets freely. Income tax is calculated and paid by asset, not by account.²³ With regard to capital gains tax for bonds, equities and commodities, a single-period model is used. All capital gains tax can be deferred if the holding assets are not sold or ‘disposed of,’ and withdrawals from a holding asset are subject to an instant capital gains tax payment (withdrawal tax) during the rebalancing process.²⁴ However, these deferred taxes will be calculated and deducted from total return to get net return at the end of the period in the model. The calculation is in respect of each individual asset, not for the whole account. All the tax rates are initially set at 40%, the middle of the range (also the average tax rate historically for individuals in the U.K. and the U.S.; see KPMG’s website), but are allowed to vary between 0% and 100%.²⁵ All returns from the risk-free asset (i.e., 3-month treasury bills), whether in the form of income or capital gains, are assumed to be free of tax.

As introduced in Section 1.3.2, source-based taxation is justified on the grounds that the country which provides the opportunity to generate income or profits should have the right to tax it. Residence-based taxation is justified on the grounds that people and firms should contribute towards the public services provided by the country where they live, on all their income regardless of the location of its source.

Consider two countries in a mixed international tax environment, one adopting a residence-based tax system and the other adopting a source-based tax system. Two different methods are required to calculate the effective tax rate: the credit method and the double taxation method.

²³Account: all assets subject to the same tax rules are put in one group (named as account) and are subject to the same tax constraints, e.g., U.K. bonds, U.S. equities, ..., etc.

²⁴ In this work, the withdrawal tax is the same as capital gains tax. All withdrawals from the sale of assets are subject to capital gains tax.

²⁵ Accessible at <http://www.kpmg.com/global/en/services/tax/tax-tools-and-resources/pages/individual-income-tax-rates-table.aspx>.

In the experiments, all three regions (U.K., U.S. and E.U.) are initially assumed to be a residence-based tax system, which conforms to the present situation. However, some of them are then replaced by a source-based tax system to investigate the impact of heterogeneous withholding tax on market performance.

Finally, when capital moves from one country to another, a Tobin tax may be payable. All transfers within one country are not subject to the Tobin tax. In the model, three different Tobin tax rules are investigated: ‘inflow tax only’ whereby investors are required to pay the Tobin tax to the country if and only if they transfer capital into it; ‘outflow tax only’ whereby investors are required to pay the Tobin tax to the country if and only if they transfer capital out of it; and ‘two-side tax’ whereby investors are required to pay the Tobin tax to the country if they transfer capital either into it or out of it. Some other assumptions are also applied.²⁶

4.4. Problem Constraints and Objective Functions

4.4.1. Basic trading constraints

All the notation used in the following exposition is explained in Table 4.1

²⁶The following are also assumed: all investors have the same available international asset classes; there is no exchange rate risk; total global wealth is constant; the regional supply of asset units is constant (i.e., there are no stock splits, rights issues or other new issues); an asset’s demand increases if its weight in the aggregated market portfolio is higher; and the asset supply is inelastic so a percentage change in asset weights corresponds to the percentage change in capitalization; international tax is simplified (ignoring annual exemption allowances); ‘Eurozone bond’ is index of French and German corporate and government bonds; ‘Eurozone’ uses one tax rate for one asset class; no capital flow restriction or friction; single period model; only consider risky assets; ‘assets’ are not securities but investible asset classes.

Table 4. 1 Notation

$\mathbf{1}$	$= (1,1,\dots,1)'$
$\mathbf{u}'\mathbf{v}$	$= u_1v_1 + u_2v_2 + \dots + u_nv_n$ (inner product)
$\mathbf{u} \circ \mathbf{v}$	$= (u_1v_1, u_2v_2, \dots, u_nv_n)'$ (Hadamard product)
n_k	number of investment assets in account k
f_{kj}	percentage paid in management fees for the account of class k and country j
$\bar{\mathbf{d}}_{kj}$	expected dividends or income returns for the account of class k and country j
$\bar{\mathbf{g}}_{kj}$	expected capital gains for the account of class k and country j
t_{cg}^{kj}	capital gains tax for the account of class k and country j
t_{cg}^{kjm}	capital gains tax for the account of class k and country j for group m
t_{in}^{kjm}	annual income tax for the account of class k and country j for group m
t_{in}^{kj}	annual income tax for the account of class k and country j
t_j^{in}	Tobin tax on capital inflows to the country j
t_j^{out}	Tobin tax on capital outflows from the country j
\mathbf{R}_{k0}^{jm}	initial cumulative capital gains for the account of class k and country j for group m
T_{k0}^{jm}	initial accumulated tax in the account of class k and country j for group m
\mathbf{w}_{k0}^{jm}	initial amount of wealth held in each asset for group m
Σ	covariance of assets
C_{0jm}	External funding for local investment at the beginning of the period
C_{0m}	external funding for overseas investment at the beginning of the period
\mathbf{R}_{k1}^{jm}	cumulative returns for the account of class k and country j after withdrawals
T_{k2}^{jm}	final accumulated tax for the account of class k and country j
TR_{kjm}	net redemption value obtained from the account of class k and country j
\mathbf{w}_{k1}^{jm}	amount of money held in each asset after rebalancing
\mathbf{w}_{k2}^{jm}	final amount of money held in each asset
\mathbf{l}_{kjm}^b	amount of money spent to buy an asset locally
\mathbf{i}_{kjm}^b	amount of money spent to buy an asset internationally
\mathbf{l}_{kjm}^s	amount of money obtained when selling an asset locally
\mathbf{i}_{kjm}^s	amount of money obtained when selling an asset internationally
\mathbf{x}_{kjm}^2	withdrawal from capital gains of assets for international transits
\mathbf{x}_{kjm}^3	withdrawal taken from the initial capital international transits
\mathbf{X}_{kjm}^2	withdrawal from capital gains of assets for local transits
\mathbf{X}_{kjm}^3	withdrawal taken from the initial capital local transits
δ_{kjm}	$\in [0,1]$, variable for assets
C_{2m}	income return at the end of the period

a. *Local trading budget*

A local trading budget (balance) constraint ensures that for all accounts, $k=1, 2, 3$ (note: 1 represents commodities, 2 represents bonds and 3 represents equities), in the same country $j, j=1, 2, 3$ (note: 1 represents the U.K., 2 represents the U.S. and 3 represents the Eurozone), the total selling proceeds from the local market, $\mathbf{1}'\mathbf{l}_{kjm}^s$, and the total external funding for the local investments, C_0 , are equal to the total buying costs within the local market, $\mathbf{1}'\mathbf{l}_{kjm}^b$, so that:²⁷

$$\sum_k \mathbf{1}'\mathbf{l}_{kjm}^b = \sum_k \mathbf{1}'\mathbf{l}_{kjm}^s + C_{0jm}; \quad j, m = 1, 2, 3 \quad (4.3)$$

All local transits within the same country or region are not subject to a Tobin tax. 'm' represents the group of investors who have the same country of residence (note: 1 represents U.K. residents, 2 represents U.S. residents and 3 represents Eurozone residents).

b. *International trading budget*

An international trading budget (balance) constraint ensures that for all accounts in all countries, the total international selling proceeds, $\mathbf{1}'\mathbf{i}_{kjm}^s$, and the total external funding for international transits, C_{0m} , are equal to the total international buying costs, $\mathbf{1}'\mathbf{i}_{kjm}^b$, so that

$$\sum_{k,j} \mathbf{1}'\mathbf{i}_{kjm}^b = \sum_{k,j} \mathbf{1}'\mathbf{i}_{kjm}^s + C_{0m}; \quad m = 1, 2, 3 \quad (4.4)$$

All international transits between different countries are subject to Tobin tax.²⁸

c. *Diversification and maximum holdings*

²⁷This budget includes all rebalancing activity within a single country, and we assume different asset classes are traded within different accounts.

²⁸This budget includes all rebalancing activities across countries, rebalancing investments from one country to another.

A diversification constraint is formulated by setting an upper bound on the value of each asset in a portfolio. Thus,

$$w_{k1}^{jm} \leq U * \delta_{kjm} \quad \forall k, j, m = 1, 2, 3, \quad (4.5)$$

$$0.0001 * \delta_{kjm} \leq w_{k1}^{jm} \quad \forall k, j, m = 1, 2, 3, \quad (4.6)$$

where U is the maximum holding weight for a single asset and is set equal to 0.05 in the optimization. If investors do not want to hold an asset in the new portfolio, the corresponding variable, $\delta_{kjm} \in [0, 1]$ in (4.6) must be equal to zero. If investors want to hold an asset in the new portfolio, the corresponding variable δ_{kjm} in (4.5) will be non-zero. By also setting a lower bound on the total number of assets in a portfolio, N_{min} , the firm-specific (or industry-specific) risk can be minimized in the market portfolio,

$$\sum_{k,j} \mathbf{1}' \delta_{kjm} \geq N_{min}, \quad m=1, 2, 3, \quad (4.7)$$

where the sum of the variables, $\delta_{kjm} \in [0, 1]$, counts this number. N_{min} is set at 20 in this work.

4.4.2. Taxation

The total tax liability is built up by calculating the impact of different tax rules on the cumulative returns, withdrawals and wealth.

a. Source- and residence-based tax systems

In a residence-based tax union, all investments are taxed by the country j in which the investment is made first, at rate t_{cg}^{kj} on capital gains and t_{in}^{kj} on incomes. If the investor is resident in a different country m with higher rates of tax t_{cg}^{km} and t_{in}^{km} , the gap will be taxed later. So the effective rate of tax on asset k in country j for an investor from country m is always the higher rate in the two countries on asset k .

$$t_{cg}^{kjm} = \max\{t_{cg}^{kj}, t_{cg}^{km}\}, \quad t_{in}^{kjm} = \max\{t_{in}^{kj}, t_{in}^{km}\} \quad (4.8)$$

In a source-based tax union, all investments are taxed by the invested country only.

So the effective rate of tax is always the rate in the invested country.

$$t_{cg}^{kjm} = t_{cg}^{kj}, t_{in}^{kjm} = t_{in}^{kj} \quad (4.9)$$

In a mixed tax system, the taxation between two countries, both under residence-based tax system, remains the same as previously. However, there are two different methods for calculating taxation between countries under different tax systems.

i. Credit method

The country with a source-based tax system is assumed to have a tax agreement with the other two countries. So if the country in which the investment is made (the ‘invested’ country) implements a source-based tax system while the investor is from a different country with residence-based tax system, the effective tax rate is:

$$t_{cg}^{kjm} = \max\{t_{cg}^{kj}, t_{cg}^{km}\}, \quad t_{in}^{kjm} = \max\{t_{in}^{kj}, t_{in}^{km}\} \quad (4.10)$$

Where t_{cg}^{km} is capital gains tax rate for account of class k and residence country of group m, while t_{in}^{km} is annual income tax rate for account of class k and residence country of group m. If the invested country implements a residence-based tax system while the investor is from a different country with a source-based tax system, the effective tax rate is:

$$t_{cg}^{kjm} = t_{cg}^{kj}, t_{in}^{kjm} = t_{in}^{kj} \quad (4.11)$$

ii. Double tax method

The country with a source-based tax system is assumed to have no tax agreement with the other two countries. So if the invested country implements a source-based tax system while the investor is from a different country with a residence-based tax system, the investment is taxed by both countries.

$$t_{cg}^{kjm} = 1 - (1 - t_{cg}^{kj})(1 - t_{cg}^{km}), t_{in}^{kjm} = 1 - (1 - t_{in}^{kj})(1 - t_{in}^{km}) \quad (4.12)$$

In equation (4.12), $(1 - t_{cg}^{kj})(1 - t_{cg}^{km})$ is the remaining rate of capital gains after tax payment by both countries. Then the effective tax rate for capital gains t_{cg}^{kjm} is calculated as 1 minus this remaining rate. A similar method is applied to get the effective tax rate for income.

If the invested country implements a residence-based tax system while the investor is from a different country with a source-based tax system, the effective tax rate is:

$$t_{cg}^{kjm} = t_{cg}^{kj}, t_{in}^{kjm} = t_{in}^{kj} \quad (4.13)$$

b. Cumulative capital gains

The remaining capital gains available for withdrawal in each account are calculated as:

$$R_{11}^{jm} = R_{10}^{jm} - \frac{x_{1jm}^2}{1 - t_{cg}^{1jm}} - X_{1jm}^2 / (1 - t_{cg}^{1jm}) \quad (4.14)$$

$$R_{21}^{jm} = R_{20}^{jm} - \frac{x_{2jm}^2}{1 - t_{cg}^{2jm}} - X_{2jm}^2 / (1 - t_{cg}^{2jm}) \quad (4.15)$$

$$R_{31}^{jm} = R_{30}^{jm} - \frac{x_{3jm}^2}{1 - t_{cg}^{3jm}} - X_{3jm}^2 / (1 - t_{cg}^{3jm}) \quad (4.16)$$

Here, x_{kjm}^2 represents the net amounts of cash withdrawal from the capital gains of each asset for international transits, while X_{kjm}^2 represents the net amounts of cash withdrawal from the capital gains of each asset for local transits. There are two sources of withdrawal: capital gains and initial capital. The difference represents the cash from the initial capital x_{kjm}^3 and X_{kjm}^3 , which is free of tax at encashment, while the cash in x_{kjm}^2 and X_{kjm}^2 are subject to an immediate tax payment. In addition, R_{k0}^{jm} is the

previously cumulated unrealized capital gains, and an upper bound on the total withdrawals is set at $\mathbf{R}_{k1}^{jm} \geq 0$ ($\forall k, j, m = 1, 2, 3$).

c. Withdrawals

When an asset is sold (or withdrawn), the gross amount from unrealized capital gains ($\mathbf{x}_{kjm}^2, \mathbf{X}_{kjm}^2$), which is proportional to the total gross amount ($\mathbf{i}_{kjm}^s / (1 - t_j^{out}), \mathbf{l}_{kjm}^s$), needs to be calculated first so that withdrawal tax on this amount can be calculated.

$\forall k = 1, 2, 3; \forall j = 1, 2, 3$

$$\mathbf{x}_{kjm}^2 = ((1 - t_{cg}^{kjm}) \mathbf{R}_{k0}^{jm} / (\mathbf{w}_{k0}^{jm} - t_{cg}^{kjm} * \mathbf{R}_{k0}^{jm})) \circ \mathbf{i}_{kjm}^s / (1 - t_j^{out}) \quad (4.17)$$

$$\mathbf{x}_{kjm}^3 = \mathbf{i}_{kjm}^s / (1 - t_j^{out}) - \mathbf{x}_{kjm}^2 \quad (4.18)$$

$$\mathbf{X}_{kjm}^2 = ((1 - t_{cg}^{kjm}) \mathbf{R}_{k0}^{jm} / (\mathbf{w}_{k0}^{jm} - t_{cg}^{kjm} * \mathbf{R}_{k0}^{jm})) \circ \mathbf{l}_{kjm}^s \quad (4.19)$$

$$\mathbf{x}_{kjm}^3 = \mathbf{l}_{kjm}^s - \mathbf{X}_{kjm}^2 \quad (4.20)$$

Since withdrawals for local transits are all exempt from Tobin tax, the term $(1 - t_j^{out})$ is not present in constraints (4.19) and (4.20).

d. Wealth

Next, the total wealth is calculated in each account after trading and at the end of the period. In calculating the former, the transactions between assets are considered.

$$\begin{aligned} \mathbf{w}_{11}^{jm} = \mathbf{w}_{10}^{jm} - \left(\frac{\mathbf{x}_{1jm}^2}{1 - t_{cg}^{1jm}} + \mathbf{x}_{1jm}^3 \right) - \left(\frac{\mathbf{X}_{1jm}^2}{1 - t_{cg}^{1jm}} + \mathbf{X}_{1jm}^3 \right) + (1 - t_j^{in}) \mathbf{l}_{1jm}^b \\ + \mathbf{l}_{1jm}^b \end{aligned} \quad (4.21)$$

$$\begin{aligned} \mathbf{w}_{21}^{jm} = \mathbf{w}_{20}^{jm} - \left(\frac{\mathbf{x}_{2jm}^2}{1 - t_{cg}^{2jm}} + \mathbf{x}_{2jm}^3 \right) - \left(\frac{\mathbf{X}_{2jm}^2}{1 - t_{cg}^{2jm}} + \mathbf{X}_{2jm}^3 \right) + (1 - t_j^{in}) \mathbf{l}_{2jm}^b \\ + \mathbf{l}_{2jm}^b \end{aligned} \quad (4.22)$$

$$w_{31}^{jm} = w_{30}^{jm} - \left(\frac{x_{3jm}^2}{1 - t_{cg}^{3jm}} + x_{3jm}^3 \right) - \left(\frac{X_{3jm}^2}{1 - t_{cg}^{3jm}} + X_{3jm}^3 \right) + (1 - t_j^{in}) t_{3jm}^b + l_{3jm}^b \quad (4.23)$$

A transit within the same country is distinguished from a transit between two countries. Tobin tax is applied to the latter transit only. In calculating wealth at the end of the period, both capital gains and incomes are considered, and the corresponding annual income tax payments and management fees are deducted (management fees are set as a parameter which can be either zero or positive).

$$w_{12}^{jm} = (1 - f_{1j})[(1 + \bar{g}_{1j}) \circ w_{11}^{jm}] \quad (4.24)$$

$$w_{22}^{jm} = (1 - f_{2j})[(1 + \bar{g}_{2j}) \circ w_{21}^{jm}] \quad (4.25)$$

$$w_{32}^{jm} = (1 - f_{3j})[(1 + \bar{g}_{3j}) \circ w_{31}^{jm}] \quad (4.26)$$

The income return should also be calculated and included in the total wealth.

$$C_{2m} = \sum_{k,j} (1 - f_{kj})[(1 - t_{in}^{kjm})(\bar{d}_{kj}' w_{k1}^{jm})] \quad (4.27)$$

Here, $k=2, 3$ only since there is no income from commodities.

e. Cumulative capital gains tax

Finally, the total tax liability is calculated by adding deferred tax from previous periods to that of the current period. Since this is a single-period model, all previous tax liabilities will have already been determined.

$$T_{12}^{jm} = T_{10}^{jm} + t_{cg}^{1jm}(1 - f_{1j})(\bar{g}_{1j}' w_{11}^{jm}) - \{t_{cg}^{1jm}/(1 - t_{cg}^{1jm})\} \mathbf{1}' x_{1jm}^2 - \{t_{cg}^{1jm}/(1 - t_{cg}^{1jm})\} \mathbf{1}' X_{1jm}^2 \quad (4.28)$$

$$T_{22}^{jm} = T_{20}^{jm} + t_{cg}^{2jm}(1 - f_{2j})(\bar{g}_{2j}' w_{21}^{jm}) - \{t_{cg}^{2jm}/(1 - t_{cg}^{2jm})\} \mathbf{1}' x_{2jm}^2 - \{t_{cg}^{2jm}/(1 - t_{cg}^{2jm})\} \mathbf{1}' X_{2jm}^2 \quad (4.29)$$

$$T_{32}^{jm} = T_{30}^{jm} + t_{cg}^{3jm} (1 - f_{3j}) (\bar{g}_{3j}' w_{31}^{jm}) - \{t_{cg}^{3jm} / (1 - t_{cg}^{3jm})\} \mathbf{1}' x_{3jm}^2 - \{t_{cg}^{3jm} / (1 - t_{cg}^{3jm})\} \mathbf{1}' X_{3jm}^2 \quad (4.30)$$

$$T_{k0}^{jm} = (\mathbf{1}' R_{k0}^{jm}) t_{cg}^{kjm} \quad (4.31)$$

In equation (4.28), $t_{cg}^{1jm} (1 - f_{1j}) (\bar{g}_{1j}' w_{11}^{jm})$ is the tax on the capital gains for the current period after management fees. This amount is then added to total cumulative tax, T_{10}^{jm} . As $\mathbf{1}' x_{1jm}^2$ is the net amount after tax, the expression $\{t_{cg}^{1jm} / (1 - t_{cg}^{1jm})\} \mathbf{1}' x_{1jm}^2$ is used to calculate the tax payment towards withdrawal for international transits. A similar method is used to calculate the tax payment towards withdrawals for local transits.

The final net return for each account by subtracting all the contingent capital gains tax from the account is calculated.

$$TR_{kjm} = \mathbf{1}' w_{k2}^{jm} - T_{k2}^{jm} \quad \forall k, j, m = 1, 2, 3 \quad (4.32)$$

4.4.3. After-tax market portfolio

In the CAPM, it is assumed that all investors are rational and are expected to hold the same market portfolio of risky assets (usually proxied by a comprehensive 'Index'), which maximizes the portfolio expected excess return over a risk-free rate per unit of portfolio risk (i.e., the Sharpe ratio). In this chapter, the objective function is the after-tax Sharpe ratio. For each group of investors, $m=1,2,3$, an independent risk-free rate of return on an after tax basis is introduced. This is because investors from different countries have access to different risk-free rates of return. The optimization models are used to find the optimal risky portfolio for each group of investors who have the same country of residence. For group $m=1, 2, 3$,

$$\text{Maximize } w (\sum_{k,j} TR_{kjm} + C_{2m} - R_f^m) / \sqrt{\mathbf{w}_{1m}' \Sigma \mathbf{w}_{1m}}$$

Subject to If residence-based (4.3)–(4.8), (4.14)–(4.32);

If source-based (4.3)–(4.7), (4.9), (4.14)–(4.32);

If mixed credit method (4.3)–(4.7), (4.10)–(4.11), (4.14)–(4.32);

If mixed double taxation (3)–(7), (12)–(32);

$$TR_{kjm} \geq 0, k = 1, 2, 3, j = 1, 2, 3; T_{k2}^{jm}, \mathbf{R}_{k2}^{jm} \geq 0, k =$$

$$1, 2, 3, j = 1, 2, 3;$$

$$w_{k1}^{jm}, i_{kjm}^b, i_{kjm}^s, l_{kjm}^b, l_{kjm}^s \geq 0, k = 1, 2, 3, j = 1, 2, 3;$$

$$x_{kjm}^2, x_{kjm}^3, X_{kjm}^2, X_{kjm}^3 \geq 0, k = 1, 2, 3, j = 1, 2, 3;$$

$$\delta_{kjm} \in [0, 1], k = 1, 2, 3, j = 1, 2, 3;$$

where R_f is the expected total wealth from investing in the risk-free asset. $\mathbf{w}_{1m}' \Sigma \mathbf{w}_{1m}$, where \mathbf{w}_1 is a vector ($w_{11}^{1m}, w_{11}^{2m}, w_{11}^{3m}, w_{21}^{1m}, w_{21}^{2m}, w_{21}^{3m}, w_{31}^{1m}, w_{31}^{2m}, w_{31}^{3m}$) of all the asset weights in period 1 (end of rebalancing), is equal to the variance σ^2 of the portfolio.

After obtaining the local optimal risky portfolio for each group of investors, the market portfolio is calculated using the following formula:

$$Portfolio_{Market} = \frac{\sum Capitalization_m \times Portfolio_m}{\sum Capitalization_m}$$

(Note: In the following work, the weight for Eurozone and U.S. investors is assumed to be 0.4 respectively, and the weight for U.K. investors is 0.2.)²⁹

4.5. Influence of Taxation on Portfolio Management

²⁹The weight for the E.U. investors here is the ratio of the total capital held by investors from the E.U. to the total capital held by all investors. The weight for the U.K. and the U.S. investors is derived in the same way.

4.5.1. Data, figures and experimental method

In the optimization, each equity, bond and commodity segment of the targeted market is divided into several subclasses. Commodities are based in all three markets (Eurozone, U.K. and U.S.) and are categorized by product type (e.g., oil, gold, copper, corn, ... etc.). All bonds currently available in the market are first divided into two groups: investment grade and high yield.³⁰ Each group is further divided into industrial subclasses (airline, technology, telecommunications ... etc.). Equities are also categorized by industry sector in all three markets. This generates 18 classes of U.K. commodities, 20 of U.S. commodities, 20 of Eurozone commodities, 7 of U.K. bonds, 24 of U.S. bonds, 24 of Eurozone bonds (mainly German and French bonds), 30 of U.K. shares, 40 of U.S. shares and 30 of Eurozone shares.³¹ All the historical annual dividends and capital gains of the asset classes are obtained from Datastream for the period 1990 to 2011. Data for bonds are from Barclay's bond index, including both government and corporate bonds. Data for equities are from the FTSE for the U.K., US-DS Price Index for the U.S., and the FTSEUR1ST 300 for the E.U. Data for commodities are from the S&P commodity index. Capital gains for each asset class are calculated as the change of index prices and adjusted by excluding corresponding income. After obtaining the optimal portfolio, investors can then invest in the corresponding index by purchasing index futures or holding the assets in each index directly.

4.5.2. Residence-based and source-based tax systems

In the experiments, investment tax rates for all asset classes are set initially at 40%. Then, in turn, each is changed incrementally from 10% to 70% so that the change of the

³⁰ Given that U.S. government bonds are often tax-free, they are not included in this work.

³¹ The same commodity asset should have the same price globally. Otherwise an arbitrage opportunity will exist. In this paper, commodities are grouped by country to show heterogeneous tax rules. For example, U.K. commodities represent commodities traded in the U.K. market and taxed by U.K. government. But the same returns of commodities are used in all three countries, U.K., U.S. and Eurozone. In reality, investors can use derivative contract or fund (e.g. ETFs) to invest in commodities indirectly. Investors are able to choose where to buy and therefore which country's tax to pay.

optimal portfolio, and therefore, the capital flow due to the change of tax rates, can be observed. In addition, it is assumed that the U.K., the U.S. and the Eurozone investors represent 100% of global markets. The proportion of total wealth of U.K. investors is taken to be 20%, while the proportion of total wealth of both Eurozone and U.S. investors is taken to be 40%.

In Fig.4.1, $a = \mathbf{R}_{k0}^{jm} / \mathbf{w}_{ko}^{jm}$. This is ratio of cumulative return to total asset weights.

A higher ratio here means more capital gains tax remaining for the payment at encashment, and therefore requires higher expected return to rebalance the portfolio. The three charts show how the optimal weight of the local market varies with the local investment tax rate under different withholding tax systems. In each chart, the curve ‘Residence Only’ is obtained when all three regions apply residence-based withholding tax and are in an international tax union. Both ‘Mixed (Credit Method)’ and ‘Mixed (Double Taxation)’ curves are obtained when the Eurozone and the U.K. apply residence-based withholding tax while the U.S. applies source-based withholding tax. However, the former assumes all three regions are still in an international tax union (Credit Method) while the latter assumes only the Eurozone and the U.K. are in a tax union (Double Taxation). The ‘Source Only’ curve is obtained when all three regions apply a source-based investment tax. In this tax system, all income is subject to a tax payment only in the country in which it is generated.

a. Residence only tax system

It can be seen from all three charts in Fig.1 that, for all three regional markets (Eurozone, U.K. and U.S.), the change of the market weight with the ‘Residence Only’ tax system is usually the smallest in comparison to the other three tax systems: ‘Source Only’, ‘Mixed (Credit Method)’ and ‘Mixed (Double Taxation)’. From the second chart, the weight of the U.K. market only changes from 60% to 62% as its investment tax rate is

cut from 40% to 30% with the ‘Residence-Only’ tax systems. This is much smaller than the changes with other tax systems, which are from 60% to 76% for ‘Source Only’, from 60% to 79% for ‘Mixed (Credit Method)’ and from 70% to 90% for ‘Mixed (Double Taxation)’. In addition, the difference in the change of market weight is usually obvious for ‘Residence Only’ tax, particularly in comparison to ‘Source Only’ tax. From the first chart in Fig.4.1, as the tax rate is cut from 40% to 10%, the weight of the Eurozone market increases from 30% to over 90% with ‘Source-Only’, while the weight increases only to 58% with ‘Residence Only’.

The experimental results also show that with ‘Residence Only’ tax, the change as the tax rate is cut is usually smaller than the change as the tax rate is increased. For example, from the third chart in Fig.4.1, the weight of the U.S. market rises from 21% to just 30% as the tax rate is cut to 10% while the weight decreases to almost 0% as the rate is increased to 70%. This is because with ‘Residence Only’ tax, the change in market weight is mainly due to rebalancing by local investors³² only when the rate falls below 40% while the change is due to rebalancing by overseas investors³³ as the rate increases above 40%. Under a ‘Residence Only’ tax system, when a regional market’s tax rate is cut below the other markets’ tax rate, only the local investor’s local investment³⁴ obtains the relatively low tax rate³⁵. As a result, such a tax cut will only benefit local investors and leads to a capital inflow to the local market due to rebalancing by local investors. In contrast, when the local market increases the tax rate above 0.4, local investors will always be taxed at the higher rate wherever they invest. So the local investors are not motivated to rebalance.

³² Local investor: the investor whose country of residence is the country we are discussing (or so-called focused market)

³³ Overseas investor: the investor whose country of residence is different from the country we are discussing

³⁴ Local investment: an investment allocated to the focused market

³⁵ For example, when the Eurozone market’s tax rate is cut to 0.1 while the rate for the other two markets is still 0.4, only Eurozone investors’ investment in Eurozone assets are taxed at 0.1; all other investors’ investments are still taxed at 0.4.

b. Source only tax system

With a ‘Source Only’ tax union, changes in the local market weight are much larger than changes in the other three tax systems (‘Residence Only’, ‘Mixed (Credit Method)’ and ‘Mixed (Double Taxation)’). This higher sensitivity leads to greater volatility of markets when a regional market changes its tax rules. This is because with a source-based tax system, the change of tax in the local market will affect all investors’ investments in the local market and will lead to a large amount of rebalancing globally. On the other hand, this feature enables the regional government to intervene in the local market in extreme cases.

c. Mixed (Credit Method) tax system

The U.S. market is assumed to apply source-based taxation while the other two markets still apply a residence-based tax. This setting is applied to test how a local tax change would affect global capital flow when heterogeneous withholding tax rules are used. In a ‘Mixed (Credit Method)’ tax system, it is assumed that the U.S. market is still in a tax union with the other two markets and double taxation is effectively eliminated. It can be seen from the first two charts in Fig.4.1 that when the tax rate is cut below 40%, the change of market weight is usually larger in a ‘Mixed (Credit Method)’ tax system than in a ‘Residence Only’ tax system for the Eurozone and the U.K. markets. For example, from the second chart in Fig.4.1, the ‘Mixed (Credit Method)’ curve always stands above the curve for ‘Residence Only’ as the tax rate is cut below 40%. In other words, the market will be more sensitive if a regional market with a source-based tax system is added to a ‘Residence Only’ tax union. This extra sensitivity is due to rebalancing by U.S. investors whose country of residence is assumed to apply a source-based tax system.

Consider a tax cut in the Eurozone and the U.K. markets, which are assumed to charge taxes based on investors' country of residence in both 'Residence Only' and 'Mixed (Credit Method)' tax systems. In a 'Residence Only' tax union, as the tax rate is cut, as mentioned previously, only local investors' local investments will be affected. In contrast, in a 'Mixed (Credit Method)' tax system, a tax rate cut in the Eurozone or the U.K. market will affect not only local investors but also U.S. investors. In a 'Mixed (Credit Method)' tax system, the U.S. market is assumed to use a source-based tax, and therefore the U.S. investors' investment in the Eurozone or the U.K. market is only taxed at the Eurozone or the U.K. tax rate. So such a tax cut in a 'Mixed (Credit Method)' tax system will lead to a larger rebalance for US investors when they hold assets in the other two markets. This implies a larger capital flow (from US to Eurozone and the U.K.) in global markets relative to the same cut in a 'Residence Only' tax system.

With regard to a tax increase in the Eurozone and the U.K. market, when the tax rate is set above 40% in the Eurozone or the U.K. market, the change of investment tax rate in a 'Mixed (Credit Method)' tax system will impose a similar effect on global markets as in a 'Residence Only' tax system. This is because the higher tax rate in a market using a residence-based tax system will increase the effective tax rate on the investment in that market to all investors regardless of whether it is in a 'Residence Only' tax system or a 'Mixed (Credit Method)' tax system. So markets will have the same sensitivity to the tax rate change in both tax systems.

In addition, a tax rate change in the U.S. market, which uses a residence-based tax in a 'Residence Only' system rather than a source-based tax in a 'Mixed (Credit Method)' tax system is observed. It is found that a tax cut (from 40% to a lower rate) in the U.S. market will lead to the same rebalancing for U.S. investors (holding more U.S. assets) regardless of whether it is in a 'Residence Only' tax system or a 'Mixed (Credit

Method)' tax system. On the other hand, a tax increase (from 40% to a higher rate) will lead to a greater rebalancing (holding more Eurozone and U.K. assets) and therefore larger capital flows (from the U.S. market to the other two markets) in a 'Mixed (Credit Method)' tax system rather than in a 'Residence Only' tax system.

In summary, if a market applies a source-based tax, it will be more sensitive to a tax increase than a tax cut. In contrast, if a market applies a residence-based tax, it will be more sensitive to a tax cut than a tax increase. These results will be of interest to both government and investors.

d. Mixed (Double Taxation) tax system

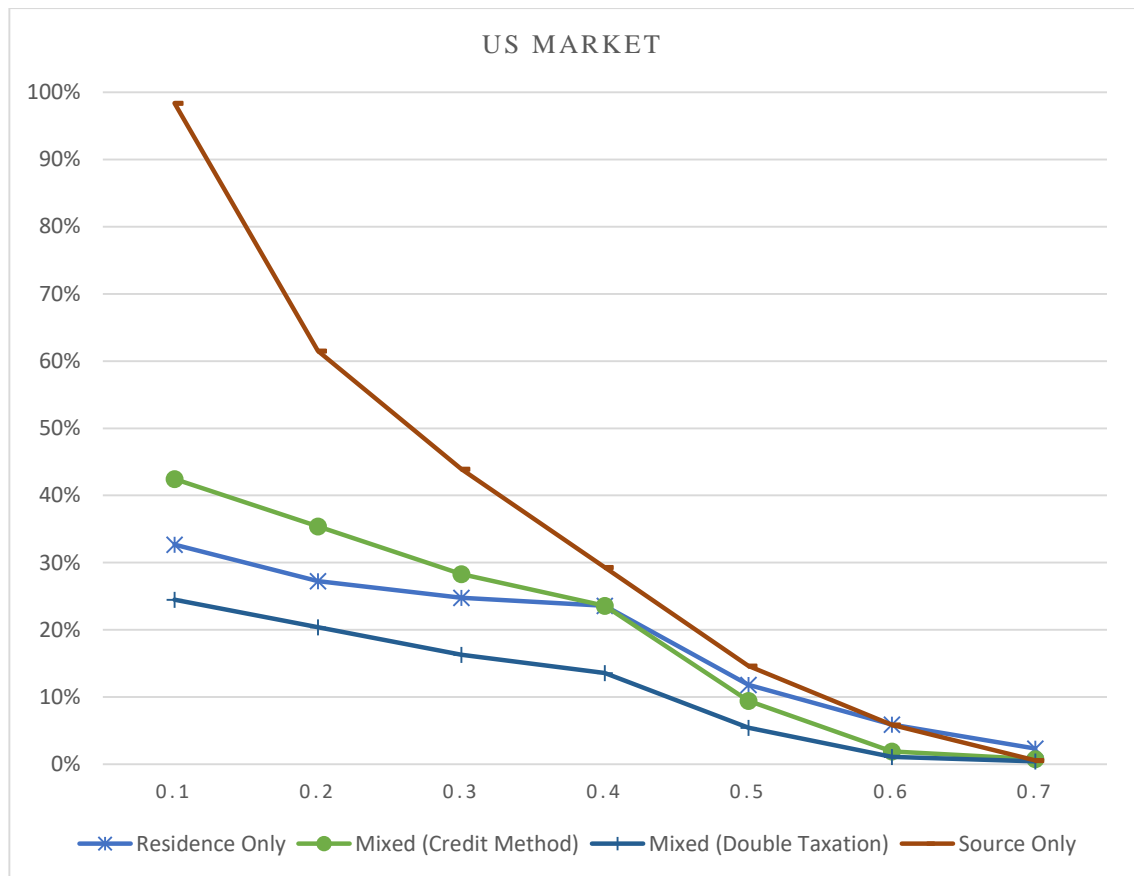
Again, the U.S. market applies a source-based taxation while the other two markets still apply a residence-based taxation. It can test whether a local tax change will affect global capital flows if heterogeneous withholding tax rules are used, and countries with different tax rules are not in a tax union. In Fig.4.1, the horizontal axis for all three charts is the investment tax rate (both income tax and capital gains tax) of the particular market and the vertical axis is the weight (or percentage) of that market to the global market in the obtained market portfolio. For example, if the weight for the Eurozone market in chart one is 20%, it means that the total summed weight of all assets the Eurozone assets in the obtained market portfolio is 20%. The three charts in Fig.4.1 show that in a 'Mixed (Double Taxation)' tax system, on average, the weight on the Eurozone and the U.K. markets is higher (the curves stand above the curve of the U.S. market) while the weight on the U.S. market is lower throughout the whole tax rate range in comparison to 'Residence Only' and 'Mixed (Credit Method)' tax systems. This is because in a 'Mixed (Double Taxation)' tax system there is double taxation on investments in the U.S. market from overseas investors (Eurozone and U.K. investors) and, therefore, its effective tax rate is always higher than that in 'Residence Only' and

‘Mixed (Credit Method)’ tax systems. Double taxation will roughly lead to a 40% decrease in the initial U.S. market weight, which largely reduces the attractiveness of the market with source-based tax. Apart from the lower initial market weight, the shape of the ‘Mixed (Double Taxation)’ curves is more or less the same as that for the ‘Mixed (Credit Method)’ in Fig. 4.1.

Figure 4. 1 Residence-based vs. Source-based Investment Tax ($\alpha=0.2$)³⁶



³⁶ In Fig.4.1, the three charts show the comparison of four withholding tax systems without Tobin tax: Residence Only, Mixed (Credit Method), Mixed (Double Taxation) and Source Only. The title of each chart is the focused regional market..

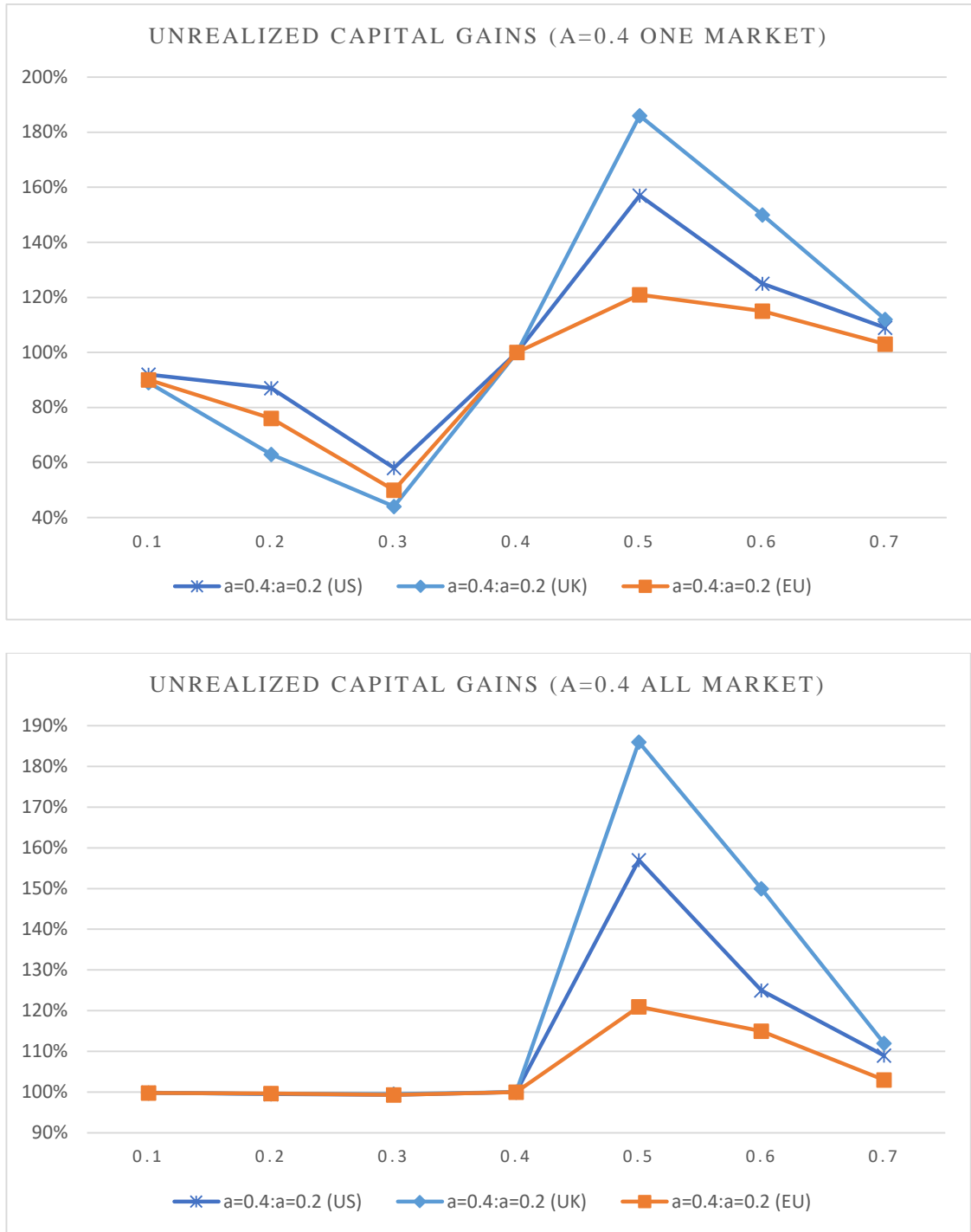


Another set of experiments are now carried out by doubling the initial unrealized capital gains amount in the global market portfolio (i.e. the parameter $a = R_{k0}^{jm} / W_{k0}^{jm}$ is changed from 0.2 to 0.4). First, the unrealized capital gains are doubled in only one market to obtain the first chart in Fig. 4.2. Next, the unrealized capital gains in all markets are doubled, and the second chart in Fig. 4.2 is obtained. The two charts of Fig. 4.2 show the differential impact of Residence Only withholding tax on global markets with two distinct unrealized capital gains amount, ‘a’. In detail, in the first chart, we assume only one regional market’s ‘a’ is changed from 0.2 to 0.4. In the second chart, we assume all three regional markets’ ‘a’ is changed from 0.2 to 0.4. The horizontal axis of these two charts is still the investment tax rate while the vertical axis is the ratio of the regional market’s weight with $a=0.4$ to its weight with $a=0.2$. In summary, when the tax rate is cut from 0.4 to 0.3, a vertical axis value below 100% means a reduced

amount of capital flowing into the particular market. Conversely, when the tax rate is raised from 0.4 to 0.5, a vertical axis value above 100% means a reduced amount of capital flowing out of that market. The two charts in Fig. 4.2 show that when the unrealized capital gains make up a higher proportion of the holding assets' value, keeping all else the same, the rebalancing amount is reduced. The second chart in Fig. 4.2 shows that as the tax rate is cut, the local market's weight decreases more when $a=0.4$ than when $a=0.2$.³⁷ This means that increased unrealized capital gains in foreign markets would lead to lower capital flows into the local market with the same tax cut and reduces the ability of governments to intervene in their local market using tax policy. On the other hand, it also shows that if an asset's expected net return increases, more unrealized capital gains in the market reduces the rebalancing activity and therefore the volatility (trading volume) of the market. In addition, when only one market's unrealized capital gains are doubled, the inflow of capital to this market will not be affected but the outflow of capital will be largely reduced (see the first chart in Fig. 4.2). In contrast, when all markets' unrealized capital gains are doubled, both the inflow and outflow of capital will be significantly reduced (roughly 50% on average, see the second chart in Fig. 4.2). A trending market (i.e., the real value of the asset in the market increases in the long term) usually creates more unrealized capital gains than a volatile but non-trending market (i.e., the real value of the asset in the market remains the same in the long term, but its price moves around its constant real value). These results lead to the conclusion that in the long-term, increasingly more capital will flow from the volatile market to the trending market, and the volatile market must offer a higher return to maintain investment capital. This conclusion is consistent with rational investors who would require higher returns in riskier (volatile) environments.

³⁷ The market of which tax rate is cut.

Figure 4. 2 Residence-based Investment Tax (a=0.2 or a=0.4)³⁸



4.5.3. Tobin tax

The rate of investment tax is now changed to simulate a change in the asset's expected net return, and an investigation is carried out into how the introduction of heterogeneous

³⁸ In Fig.4.2, the two charts show the differential impact of Residence Only withholding tax on global markets with two distinct unrealized capital gains amount, 'a'. The parameter 'a' is the assumed proportion of assets' unrealized capital gains in each regional market³⁸. In the first chart, we assume only one regional market's 'a' is changed from 0.2 to 0.4. In the second chart, we assume all three regional markets' 'a' is changed from 0.2 to 0.4. The horizontal axis of these two charts are still the investment tax rate while the vertical axis is the ratio of the regional market's weight with a=0.4 to its weight with a=0.2.

Tobin tax affects the rebalancing of investors and therefore capital flows between regional markets. From the nine charts in Figs. 4.3 to 4.5, we can see that, regardless of the withholding tax system, when the investment tax rate increases above 40%, the introduction of Tobin tax has a positive effect (i.e., it leads to an increase in the optimal market weight by reducing capital outflows). In contrast, when the tax rate is cut below 40%, Tobin tax has a negative effect (reduction in the optimal market weight by preventing capital inflows). For example, in the first chart of Fig. 4.3, for the ‘Tin&Tout=1%’ where both inflows and outflows of capital are Tobin tax charged, the curve increases above 100% as the investment tax rate is increased to 50% but decreases below 100% as the investment tax rate is cut to 30%.³⁹ In addition, whether using a consistent Tobin tax rate globally ‘Tin=0.5%’ improves market efficiency more than when using different Tobin tax rates in different regions ‘Tin=0;0.5%;1%’ is investigated.⁴⁰

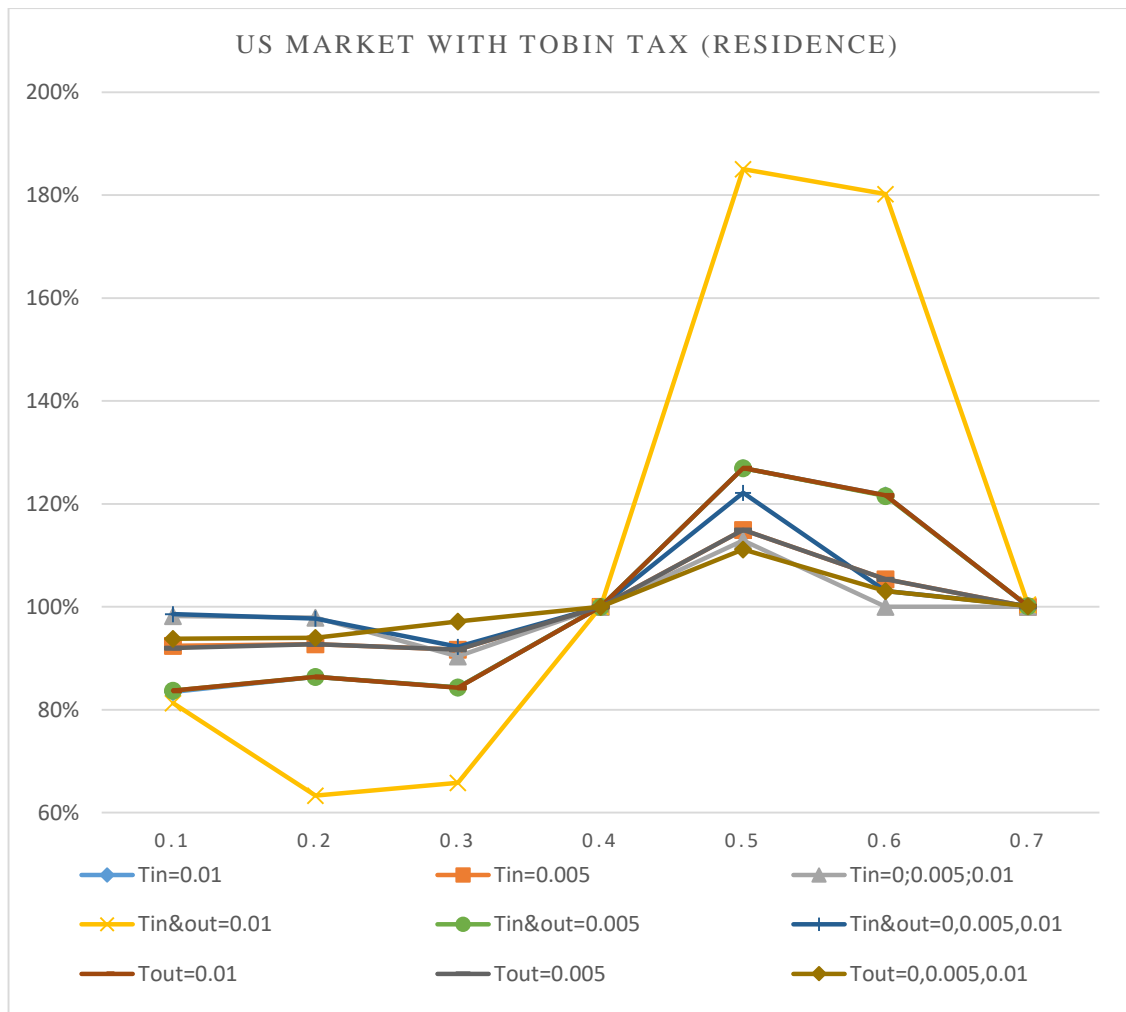
³⁹ The charge of the Tobin tax is divided into three groups: Tobin tax on capital inflows only ‘Tin’, Tobin tax on capital outflows only ‘Tout’, and Tobin tax on both capital inflows and outflows ‘Tin&Tout’.

⁴⁰ U.K.: Tin=0; U.S.: Tin=0.5%; E.U.: Tin=1%.

Figure 4.3 Tobin tax in residence-based only ($\alpha=0.2$)⁴¹



⁴¹ In Fig.4.3, the three charts show the comparison of different Tobin tax rules with a Residence Only withholding tax system. The title of each chart is the focused regional market. The horizontal axis is the investment tax rate while the vertical axis is the ratio of the focused market's weight with Tobin tax to its weight without Tobin tax.



From Fig. 4.3, it can be seen that with the same change in asset expected returns, investors' rebalancing strategy, and consequently capital flows between markets, are highly sensitive to Tobin tax. The third chart in Fig. 4.3 shows that for the 'Tin&Tout=0.5%' curve, Tobin tax can reduce the flow from the rebalancing process by 20%-40%. In extreme cases, the Tobin tax can even reduce the flow by up to 44% (see the 'Tin&Tout=1%' curve in the third chart of Fig. 4.3). This capital-lock effect is heterogeneous across different markets and different tax rate changes.⁴² For example, in the first chart of Fig. 4.3, as the investment tax rate is cut from 40% to 30%, Tobin tax 'Tin&Tout=1%' reduces the total capital inflows by 66% for the U.K. market (the optimal weight of the U.K. market is only 60% of the weight as Tin&Tout is changed from 0 to 1%). In contrast, in the second chart of Fig. 4.3, as the investment tax rate is

⁴² Reduce the amount of capital from both inflow and outflow.

cut from 40% to 30%, the same Tobin tax ‘ $T_{in} \& T_{out} = 1\%$ ’ reduces the total capital inflows by only 20% for the Eurozone market.

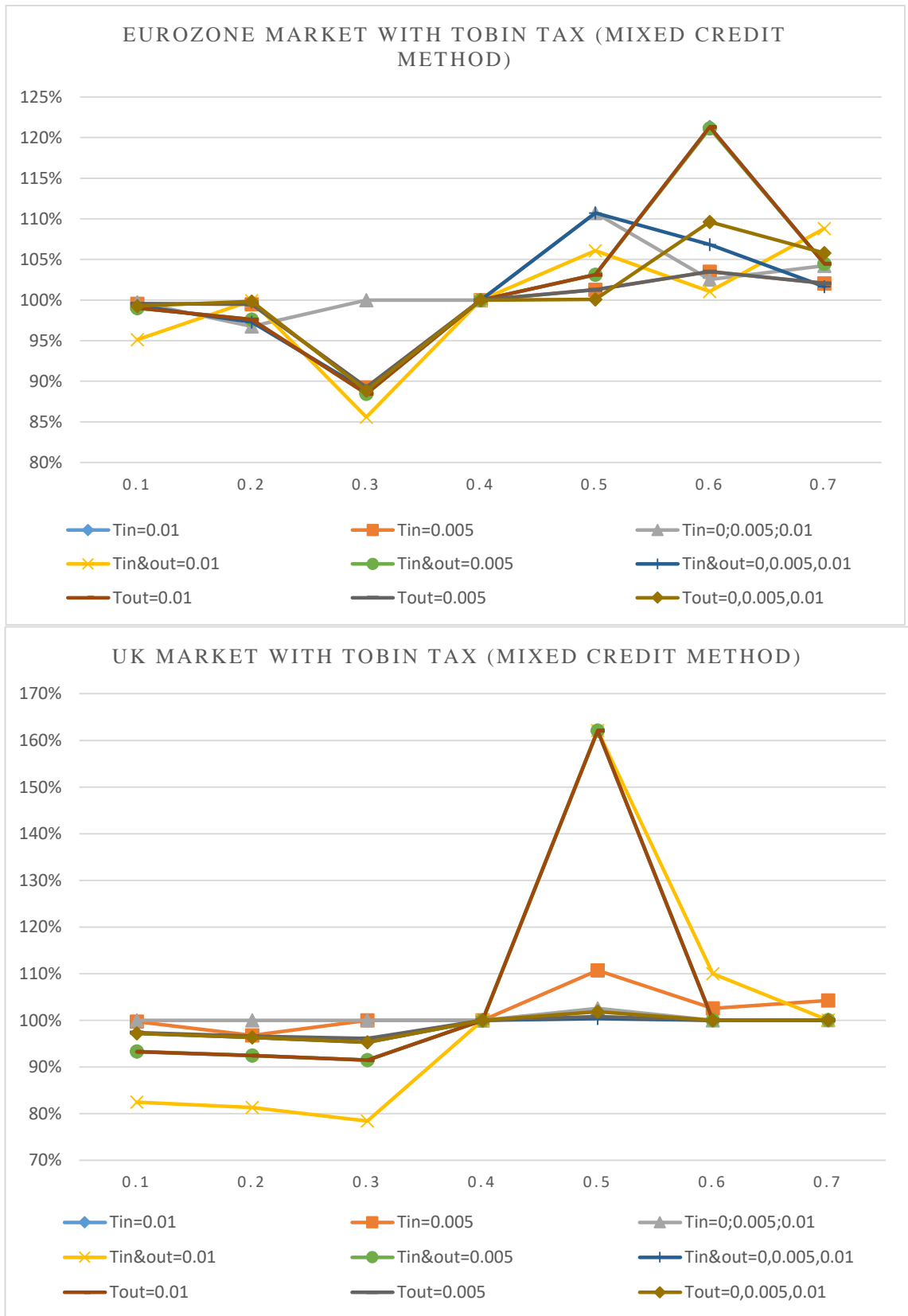
Fig. 4.3 shows that the market capital flow will usually be much more sensitive to Tobin tax with a small change in tax rate (e.g., a change from 40% to 50% or a change from 40% to 30% leads to greater rebalancing activity than larger changes). For example, in the first chart of Fig. 4.3, both the ‘ $T_{in} \& T_{out} = 1\%$ ’ and ‘ $T_{in} \& T_{out} = 0.5\%$ ’ curves reach their peak value as the investment tax rate is increased from 40% to 50%. This is because a small change in the investment tax rate, and consequently a small change in an asset’s expected return, will give investors little motivation to rebalance. When the motivation is small, the cost of Tobin tax is a major concern and may exceed the extra return (benefit) obtained from rebalancing, so a Tobin tax will be important to investors’ optimal portfolios. In contrast, when the motivation is large, the cost of Tobin tax is relatively small and rebalancing is beneficial. Charging a Tobin tax in this situation will, therefore, have little influence on investors’ optimal portfolios. In conclusion, the market will be more sensitive to the implementation of Tobin tax when asset returns are relatively stable and change only slowly and diminutively (a trending market) compared to a market in which asset returns change quickly and significantly (a volatile market). From the government’s point of view, the Tobin tax will reduce its ability to intervene in the market. This reduction (up to more than 50%) varies by the investment tax rules (Resident Only, Source Only or others) applied by the governments. Therefore, the introduction of Tobin tax hinders economic policy changes.

Fig. 4.3 shows that if the effective Tobin tax rate is similar, such as ‘ $T_{in} \& T_{out} = 0.5\%$ ’ and ‘ $T_{in} = 1\%$ ’, taxing both capital inflows and outflows and taxing only capital inflows or only capital outflows will have similar impact on markets. Thus, the impact of Tobin tax depends only on its effective rate but not on the flow of capital which is taxed. In detail, if major countries around the world decide to build a Tobin tax

union, an agreement on the effective rate of Tobin tax will suffice. Individual countries can then tailor their own Tobin tax rules (charging on inflows or outflows or both) to their individual circumstances.

Comparing the three withholding tax settings: ‘Residence Only’, ‘Mixed (Credit Method)’, and ‘Source Only’ (Figs. 4.3 to 4.5), it is clear that if the same Tobin tax rule is applied, the shape of the curves for each regional market remains roughly the same no matter which withholding tax system is used. However, the peak value (or the volatility) of the curves is not the same. So the impact of Tobin tax on markets is also dependent on other tax rules, such as withholding tax rules. As a result, when a government tries to predict the market response to the introduction of Tobin tax, taxes other than Tobin tax must also be considered. Ignoring investment taxes or withholding tax in an investigation of Tobin tax will lead to an inaccurate prediction of its impact.

Figure 4. 4 Tobin tax in credit method (a=0.2) ⁴³



⁴³ In Fig.4.4, the three charts show the comparison of different Tobin tax rules with Mixed (Credit Method) withholding tax system.

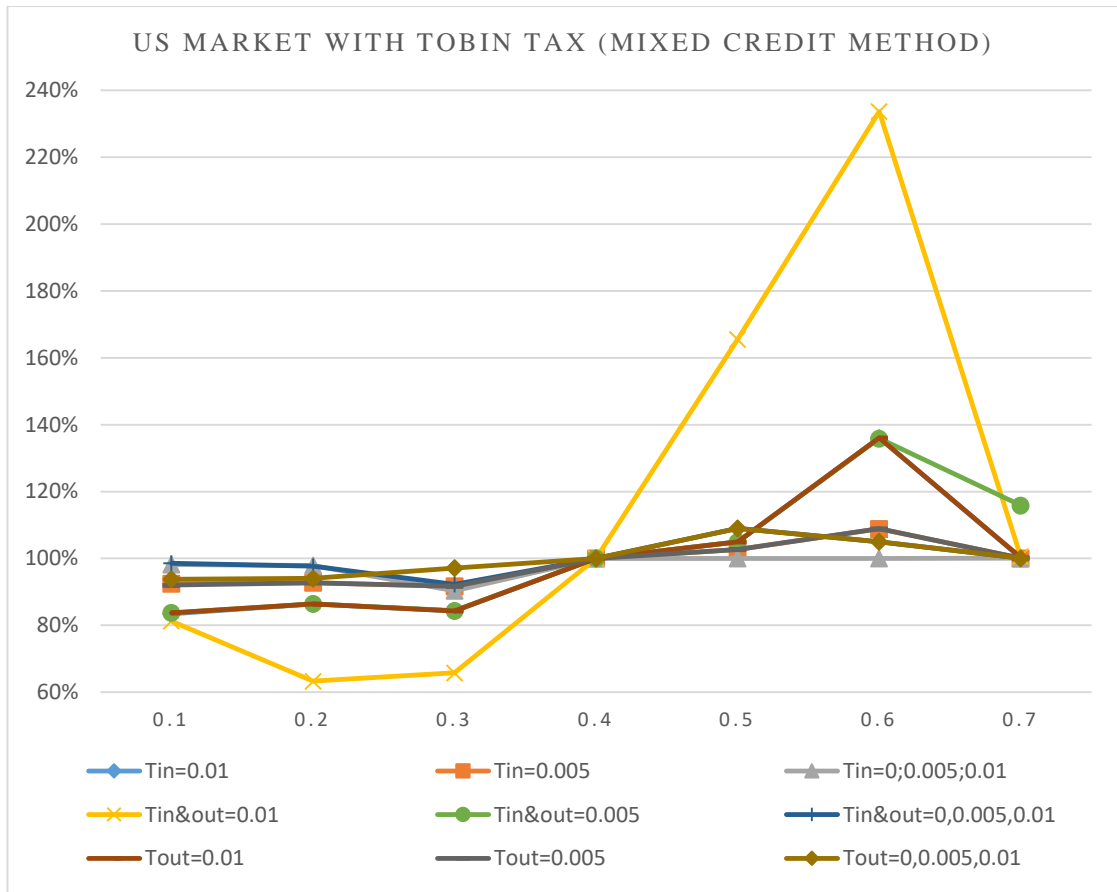
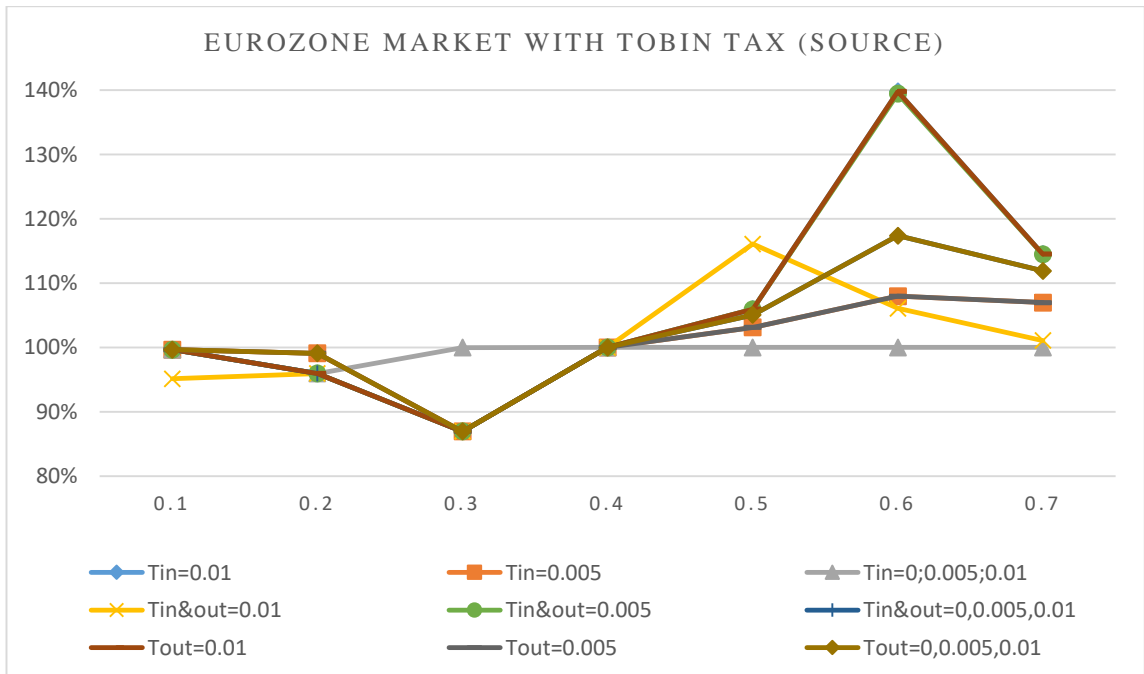
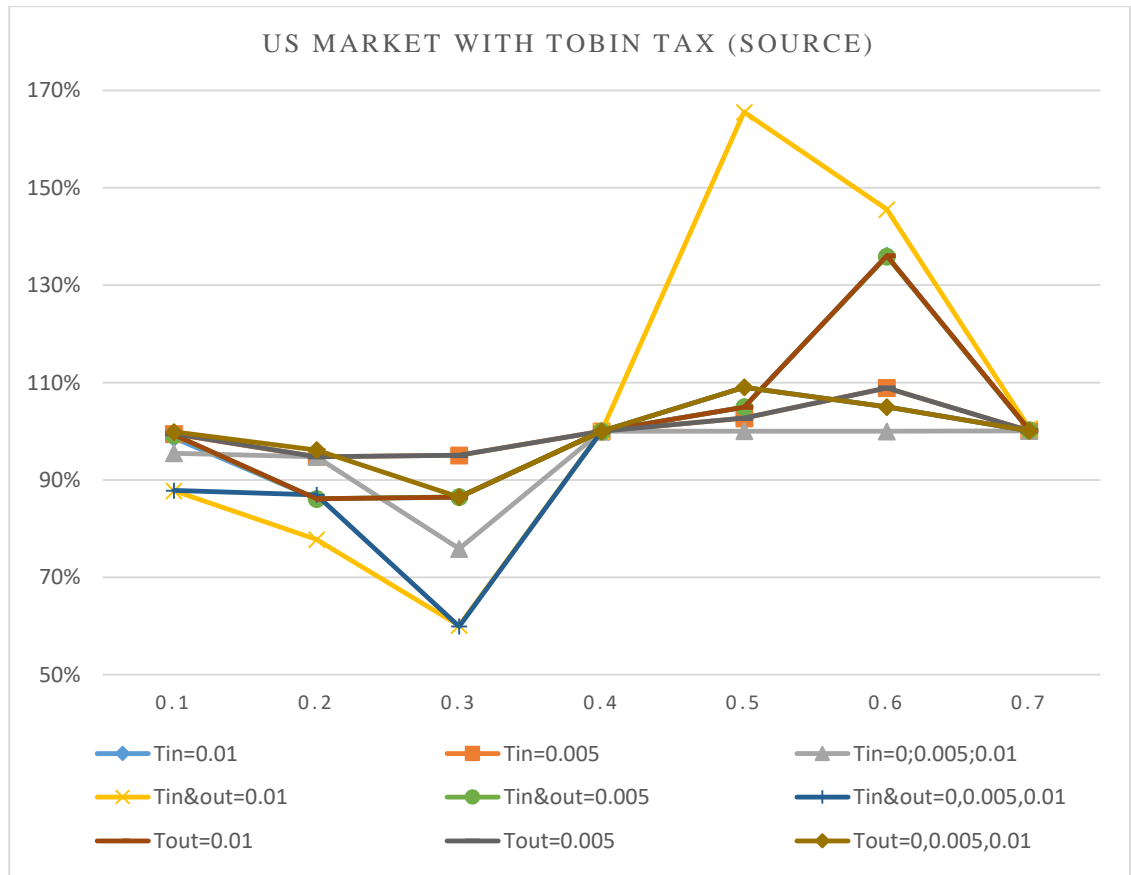
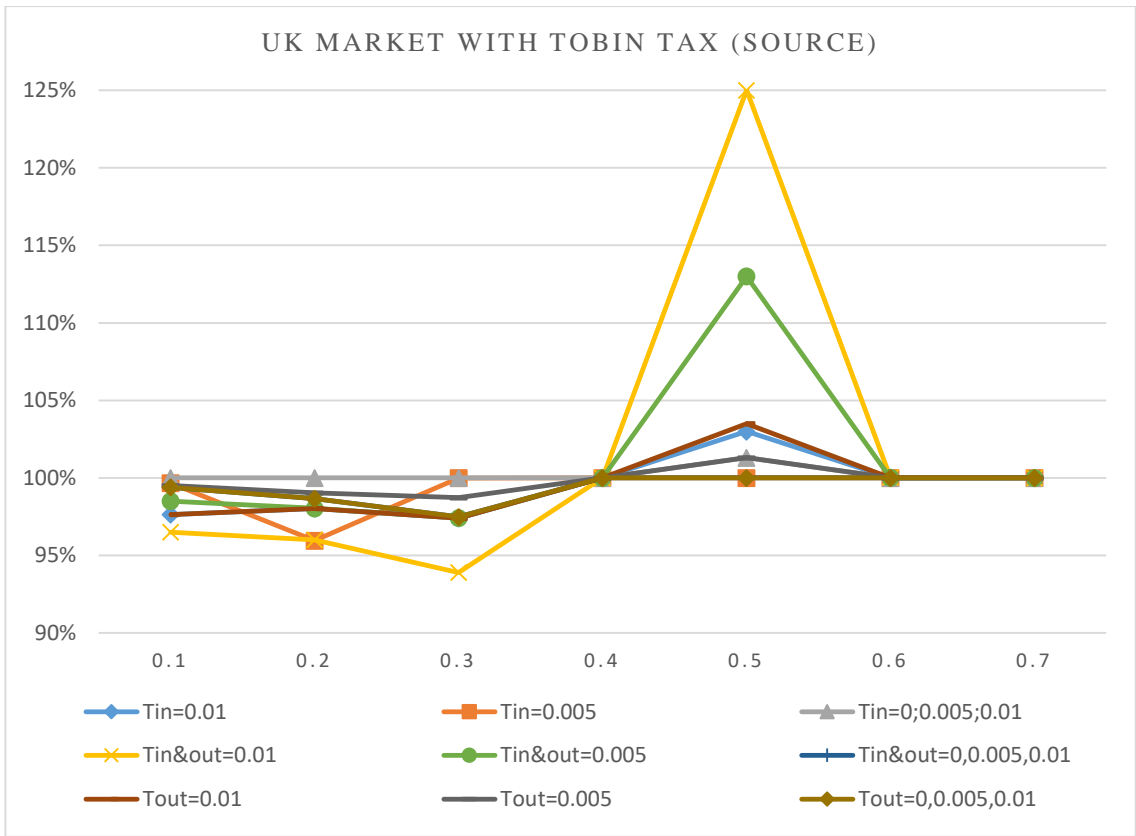


Figure 4.5 Tobin tax in source-based only ($\alpha=0.2$)⁴⁴

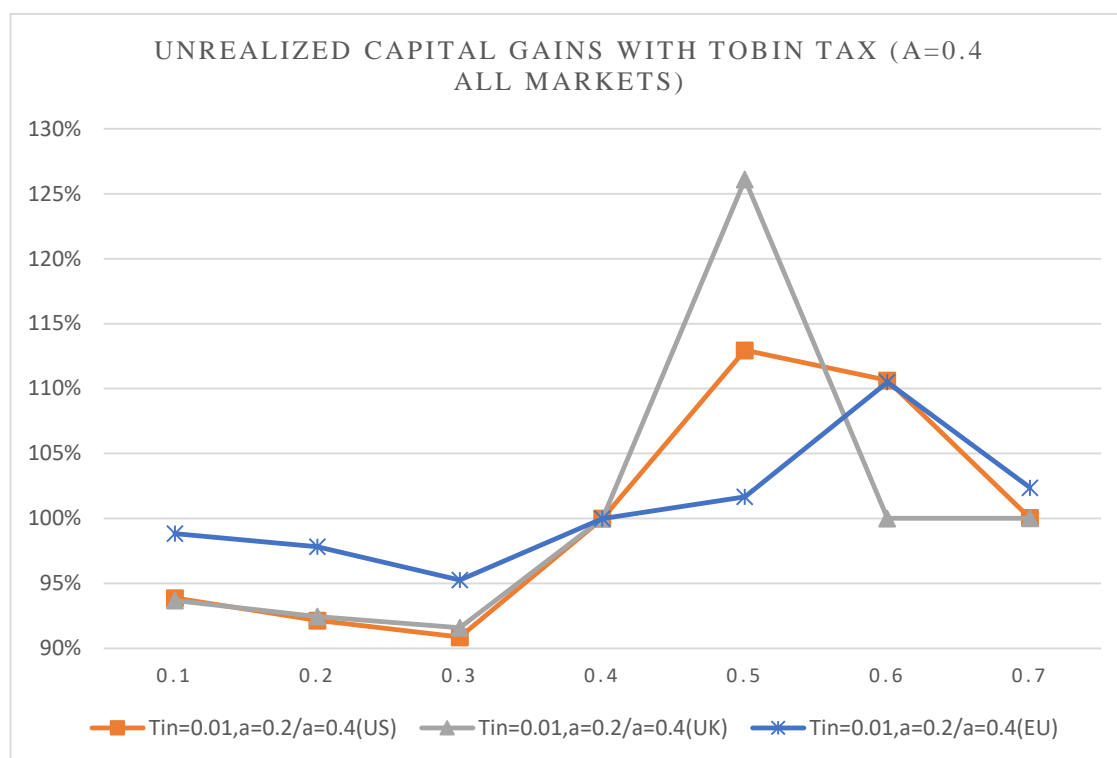


⁴⁴ In Fig.4.5, the three charts show the comparison of different Tobin tax rules with source-only withholding tax system. Other settings of these three figures are the same as Fig.4.3.



The impact of a Tobin tax is predicted with different amounts of unrealized capital gains in the market. Similar to the experiment in section 4.5.2, the parameter 'a' is changed from 0.2 to 0.4 and obtain Fig. 4.6. It is also assumed that the 'Residence Only' withholding tax system is being used. It is further assumed that Tobin tax is charged when capital flows into a country's market (capital inflows), and the rate is set at 1%. In Fig. 4.6, the vertical axis is the ratio of the local market weight when 'a' is 0.2 to the weight when 'a' is 0.4. With less unrealized capital gains ($a=0.2$), Tobin tax has a larger effect on rebalancing and therefore capital flows. But this difference is smaller when there is a large change in tax rate and therefore a large change in asset expected returns are made. So if asset expected returns change by a small amount, investors' rebalancing will be more sensitive to Tobin tax in a volatile market (with small unrealized capital gains), and the Tobin tax will greatly reduce the trading volume between markets. In contrast, in a trending market (with large unrealized capital gains), investors will be more concerned about the current portfolio's unrealized capital gains and their tax cost if redeemed, than the cost of Tobin tax. However, if asset expected returns change by a large amount, neither unrealized capital gains nor Tobin tax will be a major part of investors' rebalancing considerations.

Figure 4.6 Tobin tax in residence-based only (a=0.2 or a=0.4)⁴⁵



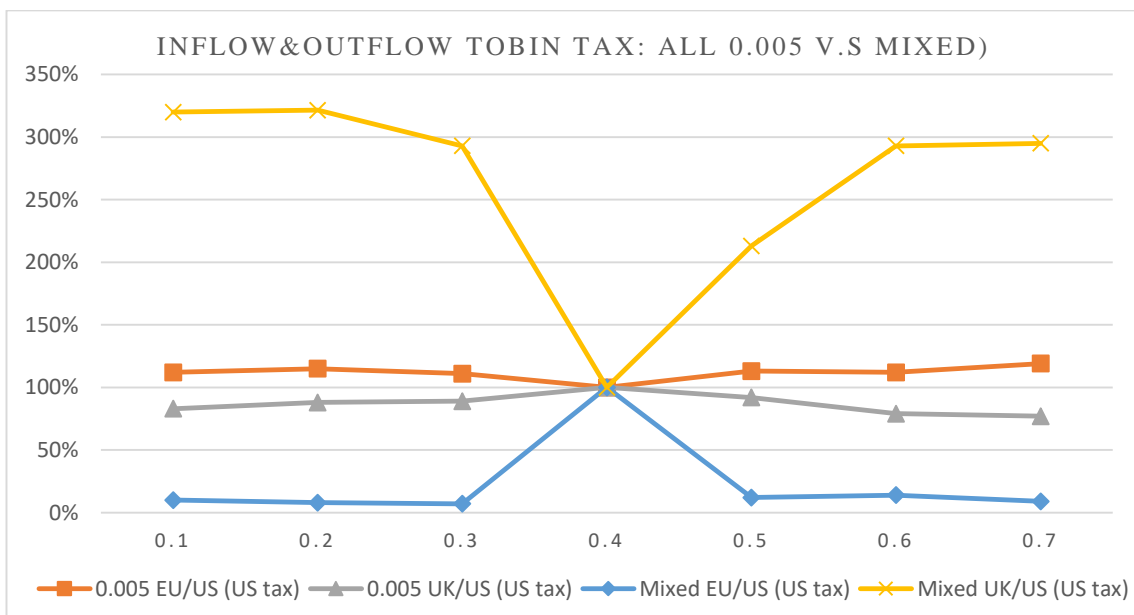
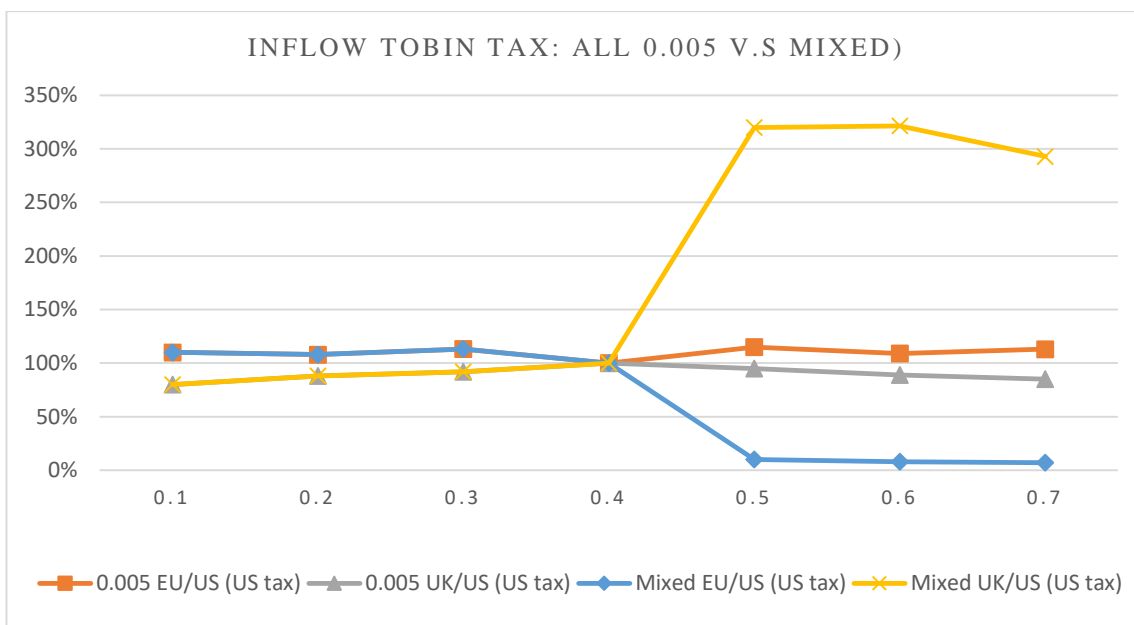
Finally, in Fig. 4.7, a comparison is made between using a consistent Tobin tax rate globally ('Tin=0.5%' and 'Tin&Tout=0.5%') and using heterogeneous Tobin tax rates globally ('Tin=0;0.5%;1%' and 'Tin&Tout=0;0.5%;1%').⁴⁶ It is assumed that a 'Residence Only' withholding tax system is being used. The vertical axis is the ratio of the percentage change in a target market (Eurozone or U.K. market) weight to the percentage change in the U.S. market weight when the investment tax rate is cut or increased. First, it is assumed that only capital inflows are subject to Tobin tax (the first chart of Fig. 4.7). It shows that when the U.S. investment tax rate is cut, there will be capital outflow from the Eurozone and the U.K. markets and into the U.S. market. However, since only the inflow will be taxed, the capital flow will be subject to the U.S. Tobin tax only. The effective tax rate should be the same regardless of whether

⁴⁵ Fig.4.6 shows the differential impact of Tobin tax on global markets with two distinct unrealized capital gains amount, 'a'. The parameter 'a' is changed from 0.2 to 0.4 in all three regional markets. The horizontal axis is the investment tax rate while the vertical axis is the ratio of the focused regional market's weight with a=0.2 to its weight with a=0.4.

⁴⁶ Here means the Tobin tax rate in the Eurozone is 0, the Tobin tax rate in the U.K. is 0.005 and the Tobin tax rate in the U.S. is 0.01

homogenous or heterogeneous Tobin tax rates are used globally. Thus, there will be no effect on the performance of markets.

Figure 4. 7 Tobin tax in residence-based only (t=0.005 vs t=0;0.005;0.01)⁴⁷



⁴⁷ Fig.4.7 shows the comparison of consistent and heterogeneous Tobin tax rules. In the consistent Tobin tax, all regional markets charge 0.5% on an international transit of wealth. In the heterogeneous Tobin tax, the U.K. market charges no Tobin tax, the U.S. market charges 0.5%, and the Eurozone market charges 1%. The first chart assumes only transits into a regional market will be charged a Tobin tax for that market while the second chart assumes both transits into or out of a regional market will be charged a Tobin tax for that market. The horizontal axis is the investment tax rate of the U.S. market while the vertical axis the ratio of the percentage change on the target market (the Eurozone or U.K. market) weight to the percentage change on U.S. market weight when investment tax rate is cut or increased.

In contrast, when the U.S. investment tax rate is increased, leading to a capital outflow from the U.S. market, heterogeneous Tobin tax rates will mean that all capital flowing into the U.K. market will be taxed at zero while capital flowing into the Eurozone market will be taxed at 1%. This difference in Tobin tax treatment will lead to increasingly more flow of capital into the U.K. market rather than into the Eurozone market, and consequently largely reduces the ability of the Eurozone market to attract overseas investment in the long-term. It can be seen from the first chart of Fig. 4.7 that when the tax rate is increased, on average, 300% more capital will flow into the U.K. market when heterogeneous Tobin tax rates are used. Furthermore, if both outflows and inflows of capital are subject to Tobin tax, the result will be different (see the second chart of Fig. 4.7). This difference occurs mainly when the U.S. investment tax rate is cut.

From the second chart in Fig. 4.7, when the U.S. tax rate is cut, the rebalancing amount is different from that using a consistent Tobin tax rate. More capital will flow out of the U.K. market and less capital will flow out of the Eurozone market if heterogeneous investment tax rates are used. As a result, heterogeneous Tobin tax rates will lead to higher volatility and trading volume in a low Tobin-taxed market (U.K) and lower volatility and trading volume in a highly Tobin-taxed market (Eurozone). This significantly reduces the ability of governments to intervene in the markets when necessary and reduces the ability to attract overseas investment.

4.6. Conclusion

This chapter investigates the quantitative effects of investment taxes and Tobin tax on capital flows between regional markets. A post-tax portfolio optimization model is developed with non-linear trading constraints and objective function. To undertake a quantitative examination of the influence of heterogeneous withholding and Tobin taxes on international financial markets, a broad range of the real-world trading constraints

are incorporated. So investor behaviour can be simulated more realistically than using a model with simplified trading constraints, and this influence is quantified by observing the rebalancing activities of rational investors under different tax settings.

On comparing residence- and source-based taxes on global investments, it is found that the global optimal portfolio is highly sensitive to a change in regional investment tax rates. This sensitivity depends on the size of the tax rate change, market specifications, and the international investment tax environment (Residence Only, Source Only, or Mixed). In a uniform tax policy across countries, a source only tax union will, on average, have more capital transits in international markets than would be the case with a Residence Only tax union, and its optimal market portfolio will be more sensitive to regional tax policy changes. In a mixed tax system, Mixed (Double Taxation) between residence- and source-taxed markets will significantly reduce the attractiveness of the latter to investors, while the Mixed (Credit Method) will perform much better (increasing the attractiveness of the market with a source-based tax by up to 20%- see the third chart in Fig. 4.1). The experimental results also suggest that volatile markets, which are usually accompanied by low unrealized capital gains, are more sensitive to a government's tax policy than trending markets.

Trading volume from rebalancing activities of rational investors (who seek to maximize the net Sharpe ratio) is highly sensitive to the implementation of a Tobin tax. This sensitivity varies with both market specifications and investment tax rules. A volatile market in a 'Mixed (Credit Method)' tax environment will be more sensitive to Tobin tax than a trending market in a 'Mixed (Double Taxation)' tax environment. Furthermore, the experiments show that the capital locking effects of Tobin tax is mainly dependent on its effective rate but not the taxation on the capital flow (taxing inflow only or outflow only), if a consistent Tobin tax rule is applied to all countries. When heterogeneous rules are used across regional markets, for a market with relatively

high Tobin tax rate, the inflow Tobin tax will have a much higher capital lock-out effect, and the outflow Tobin tax will have a much higher capital lock-in effect, in comparison to a consistent Tobin tax system. In other words, the capital locking effect of Tobin tax is enlarged significantly when heterogeneous Tobin tax rates are applied. As a result, it will be helpful if all countries reach an agreement on the implementation of Tobin tax. Otherwise, a relatively high Tobin tax will significantly reduce the appeal of local markets to foreign investors.

Chapter 5 - Dynamic Tax Arbitrage for Perfectly Correlated and Non-perfectly Correlated Assets

Continuous time optimization models are developed with heterogeneous taxation between investors programmed with continuous rather than static income and capital gains (or losses). It is proved analytically that tax arbitrage opportunities exist for both perfectly correlated and non-perfectly correlated assets. However, these opportunities are very sensitive to asset price changes and investment tax rules, and therefore difficult to track. For perfectly correlated assets, it is proved that tax arbitrage may exist, with the investor's top tax rate and some static asset parameters determining the existence of arbitrage opportunities. It is also proved that many of the equilibriums obtained under income tax only are not different from those incorporating capital gains tax if investors are subject to capital gains tax. For non-perfectly correlated assets, however, it is the market price of cap and floor options on assets' returns that determine the existence of tax arbitrage. In the government fixed-income bond market, tax arbitrage between investors is difficult to eliminate unless investors are all subject to the same tax rates. But the return from this arbitrage can be limited if the government applies the same top tax rate to all investors.

5.1. Introduction

The expected pre-tax asset return and its risk are normally the main considerations in research on portfolio management and asset pricing. In reality, however, heterogeneous taxation can significantly influence equilibrium prices (Basak and Croitoru, 2001). This heterogeneity may exist across different investors, securities and types of return (capital gains or income). Some investors, for example, are subject to higher tax rates than others; derivative securities may be subject to tax rules different from those applied to

their underlying assets; and even the same asset may be subject to different taxation depending on the purpose for which it is held (e.g., investing for retirement). These features make the asymmetric treatment of taxes important in asset pricing but challenging to include in mathematical programming. Mainly due to this complexity, researchers generally simplify work on asset pricing and portfolio management by assuming constant tax rates. In this chapter, I relax this assumption by including tax heterogeneities in determining the dynamic equilibrium of asset prices.

To investigate the role of heterogeneous tax over time, Constantinides (1983) assumes that tax rates are higher in the short term than in the long term. He shows that investors will take advantage by realizing losses in the short term but gains in the long term. Dammon and Spatt (1996) and Osorio et al. (2004) also show that the value of tax timing is significant to investors. Dammon and Spatt (1996), however, prove that even if there are no transaction costs, sometimes investors may also defer small losses when asymmetric capital gain tax rules are applied in the optimization.

Researchers have worked on finding equilibrium asset prices with heterogeneous tax brackets across investors. Basak and Croitoru (2001) propose a time-continuous model to develop dynamic equilibriums of asset prices between two heterogeneous agents when the presence of redundant, non-linearly taxed securities provides opportunities for tax arbitrage. They consider arbitrage opportunities between perfectly correlated assets with continuous returns but only income tax (not capital gains tax) is included (Basak and Croitoru 2001, Dammon and Green 1987, Dammon and Spatt 1996, Strobel 2005, Zuckerman 1989).

In this chapter, both capital gains and losses are considered, and continuous-time models with continuous returns are developed for both perfectly correlated and non-perfectly correlated assets. This enables many of the deficiencies in previous work to be corrected. For example, Basak and Croitorn discuss tax arbitrage opportunities when

assuming no capital gains or losses from the underlying assets. The work of Basak and Croitorn (2001) for perfectly correlated assets is extended by including capital gains tax and relaxing some of the assumptions. Capital gains tax is considered as part of total asset return tax while Basak and Croitorn assume there are no capital gains taxes. In their analysis of market equilibrium they assume that the existence of tax arbitrage is limited to an upper bound of the total amount of profit. In other words, any discovered tax arbitrage opportunity will disappear as more and more profit is secured as the arbitrage portfolio increases. The market equilibrium is reached when all arbitrage profit is realized and no more tax arbitrage profits can be found by investors. In this chapter, tax arbitrage which is limited by an upper bound is defined as a local tax arbitrage. In contrast, if there are mathematical constraints on tax rates and asset parameters (i.e. asset prices, expected returns and variance) which enable investors to generate profits not limited to any upper bound given enough liquidity in the market, this kind of arbitrage is defined as a global tax arbitrage opportunity. Since income must be positive while capital gains could be either positive or negative, it is more difficult to prove the existence of arbitrage opportunities when considering both capital gains tax and income tax rather than just income tax alone. This improvement increases the complexity but is necessary in considering tax arbitrage opportunities.

It is proved analytically in this chapter that arbitrage opportunities exist, but they are very sensitive to asset prices, tax rates and other parameters, and do not exist consistently. When no arbitrage opportunity exists, equilibrium can be achieved when the sum of aggregate market capital gains tax payments and income tax payments are minimized. It is also proved mathematically that when considering market equilibrium on an after-tax basis and investors are subject to capital gains tax, many equilibriums obtained under income tax only would not be attained after incorporating capital gains

tax. It is therefore necessary to incorporate capital gains tax when considering tax arbitrage opportunities and market equilibrium.

For non-perfectly correlated assets, caps and floors are included in the model. The cap is a call option that sets a maximum future return for an underlying asset. The floor is a put option that sets a minimum future return for an underlying asset. Three new continuous-time optimization models are proposed to find conditions for the existence of local, global and restricted global arbitrage opportunities⁴⁸. These opportunities are further divided into two categories, type A and type B, depending on whether a strictly positive or only non-negative future net (after-tax) return will be realised for certain without an outflow of funds at any time. Further, given a set of tax rates and asset parameters, a new function, which requires asset holdings as inputs, is proposed to calculate an asset's marginal cap and floor for its total net return. It is proved that the existence of tax arbitrage opportunities between non-perfectly correlated assets simply depends on the difference between assets' marginal caps and floors.

The remainder of this chapter is organized as follows. Section 5.2 describes the model for investors to optimize dynamic portfolio return under heterogeneous taxation. Section 5.3 analyses tax arbitrage and equilibrium for perfectly correlated assets. Section 5.4 discusses tax arbitrage and equilibrium for non-perfectly correlated assets with caps and floors and Section 5.5 concludes. Appendices A and B provide relevant definitions and proofs.

5.2. Tax Arbitrage Optimization

⁴⁸ Restricted tax arbitrage profit is not limited by an upper bound but limited to a minimum purchase of the selected assets. More details will be introduced in Section 4.

5.2.1. Model for Perfectly Correlated Assets

A new continuous-time optimization model for perfectly correlated assets based on the work of Basak and Croitoro (2001) is proposed. As in their work, heterogeneous taxation across investors and assets, and asymmetric tax treatment of long and short positions are considered in the model. However, to extend their work further, capital gains taxes are also included in the optimization.

Table 5. 1 Notation⁴⁹

$B(t)$	price of locally riskless "bond" at time t
$r(t)$	instantaneous interest rate of "bond" at time t
$S(t)$	price of the risky security with positive net supply of one share at time t
$P(t)$	price of the risky security P at time t
$S^*(t)$	dynamic capital gain of security S at time t
$P^*(t)$	dynamic capital gain of security P at time t
$\delta_i(t)$	income of risky security i at time t
$\mu_j(t)$	average total return of security j at time t
$\sigma_j(t)$	total volatility of security j at time t
$\mu_{\delta_j}(t)$	average income from security j at time t
$\sigma_{\delta_j}(t)$	income volatility of security j at time t
$\mu_j^c(t)$	average capital gains of security j at time t
$\sigma_{\delta_j}^c(t)$	capital gains volatility of security j at time t
$\alpha_j^i(t)$	agent i 's holding (in units) of security j at time t
$T^i(\cdot)$	the function of agent i 's tax bill on income
$CGT^i(\cdot)$	the function of agent i 's tax bill on capital gains
$T^{i*}(t)$	agent i 's marginal tax rate on income
$CGT^{i*}(t)$	agent i 's marginal tax rate on capital gains
$t_j^i(\cdot)$	the function of agent i 's taxable income from security j 's income
$cgt_j^i(\cdot)$	the function of agent i 's taxable income from security j 's capital gains
$t_j^{i*}(\cdot)$	agent i 's marginal tax rate on security j 's income
$cgt_j^{i*}(\cdot)$	agent i 's marginal tax rate on security j 's capital gains
τ_{j-}^i	agent i 's contribution rate to taxable income from security j 's income for
$cg\tau_{j-}^i$	agent i 's contribution rate to taxable capital change from security j 's capital losses

⁴⁹The table explains the meaning of every symbol used in the model

τ_{j+}^i	agent i 's contribution rate to taxable capital change from security j 's income
$cg\tau_{j+}^i$	agent i 's contribution rate to taxable income from security j 's capital gains
$\Delta_{S,P}(t)$	mispricing between two risky securities
$X^i(t)$	total wealth of agent i
$\Phi^i(t)$	risk exposure for agent i
$A_j(t)$	asset price of asset j at time t
$A_j^B(t)$	asset price with bound of asset j at time t
$\delta_{A_j}^B(t)$	income with bound of asset j at time t
$Flo_{\delta_{A_j}}^{tax}$	effective floor of income of asset j after tax
$Flo_{\delta_{A_j}}$	effective floor of income of asset j before tax
$Cap_{\delta_{A_j}}^{tax}$	effective cap of income of asset j after tax
$Cap_{\delta_{A_j}}$	effective cap of income of asset j before tax
\sum	to calculate a sum
$\tilde{W}_{i,j}(t)$	Brownian motion for income of asset j at time t
$W_{c,j}(t)$	Brownian motion for capital gains of asset j at time t
G_S	capital gains of asset S

a. *Continuous Time Model*

i. *Dynamic incomes (δ_S and δ_P) and prices (P and S)*

A continuous market is assumed, as in Basak and Croitoro (2001). Uncertainty is represented by a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \rho)$ within which a one-dimensional Brownian motion W is defined. It is assumed that all investors are homogeneous in their information (represented by \mathcal{F}_t) and beliefs (represented by ρ); all stochastic processes introduced are subject to \mathcal{F}_t ; all calculations with random variables hold ρ -a.s.; and all stochastic differential equations have a solution. Further, in the discussion of perfectly correlated assets, investors are assumed to trade three assets: a locally riskless 'bond' with price B returning an instantaneous interest rate r , and two risky assets with prices S and P paying dynamic incomes δ_S and δ_P . The underlying dynamics are governed by the following processes:

$$dB(t) = B(t)r(t)dt \quad (5.1)$$

$$d\delta_S(t) = \delta_S(t)[\mu_{\delta_S}(t)dt + \sigma_{\delta_S}(t)dW(t)] \quad (5.2)$$

$$d\delta_P(t) = \delta_P(t)[\mu_{\delta_P}(t)dt + \sigma_{\delta_P}(t)dW(t)] \quad (5.3)$$

$$dS(t) + \delta_S(t)dt = S(t)[\mu_S(t)dt + \sigma_S(t)dW(t)] \quad (5.4)$$

$$dP(t) + \delta_P(t)dt = P(t)[\mu_P(t)dt + \sigma_P(t)dW(t)] \quad (5.5)$$

where the processes r , μ_S , μ_P , σ_S , σ_P are determined endogenously (given the parameters) in equilibrium. Investor i 's holding (in a unit) of asset j at time t is denoted by $\alpha_j^i(t)$, $j \in \{S, P\}$.

ii. Correlation

Brownian motion W is employed to generate both assets' dynamic incomes and capital gains. The processes guarantee that the two risky assets' correlation coefficient equals one and income is correlated with capital gains (the same process is used in Basak and Croitoro (2001)).

iii. Tax on income

Investor i is taxed on income received from the two risky assets. At time t , he or she pays the instantaneous amount

$$T^i(t_S^i(\alpha_S^i(t)\delta_S(t)) + t_P^i(\alpha_P^i(t)\delta_P(t))) \quad (5.6)$$

where $T^i(\bullet)$, $t_S^i(\bullet)$ and $t_P^i(\bullet)$ are functions of income tax. The argument of the tax bill $T^i(\bullet)$ is referred to as investor i 's total taxable income, and the argument of the taxable income $t_j^i(\bullet)$ is referred to as total income from asset j . In addition, the formula for the marginal tax rate will be $T^{i*}(S, P) = T^{i*}(t_S^{i*}) + T^{i*}(t_P^{i*})$. For convenience, investor i 's income tax bill at time t is denoted by $T^i(t)$, taxable income from asset j at time t by

$t_j^i(t)$, and their derivatives (marginal tax rate) by $T^{i*}(t)$ (or T^{i*}) and $t_j^{i*}(t)$ (or t_j^{i*}). As a result, the effective marginal tax rates can be calculated as $T^{i*}(t_S^{i*})$, $T^{i*}(t_P^{i*})$. For example, if the marginal tax rate is 40% on asset S and 28% on asset P , we can set $T^{i*}(t)$ to be 40%, $t_S^{i*}(t)$ to be 100% and $t_P^{i*}(t)$ to be 70%, so that the effective tax rate $T^{i*}(t) \times t_j^{i*}(t)$ is 40% for asset S and 28% for asset P . This effective marginal tax rate can then be used to calculate tax payment by multiplying incomes from the asset.

iv. *Heterogeneous income tax on long and short positions*

Following the work of Basak and Croitorn, the heterogeneous treatment of long and short positions is considered in the model using the following equation:

$$t_j^i(\alpha_j^i(t)\delta_j(t)) = (\alpha_j^i(t)\delta_j(t))^+ \tau_{j+}^i + (\alpha_j^i(t)\delta_j(t))^- \tau_{j-}^i \quad (5.7)$$

In equation (5.7), the total income tax from holding $\alpha_j^i(t)$ units of asset j for investor i is equal to the sum of the tax on both long and short positions. The long positions $(\alpha_j^i(t)\delta_j(t))^+$ are taxed at long-position tax rate τ_{j+}^i , and short positions $(\alpha_j^i(t)\delta_j(t))^-$ are taxed at short-position tax rate τ_{j-}^i . This setting may cause the taxation function to be non-differentiable at zero income. Apart from this special case, all functions $T^i(\bullet)$, $t_S^i(\bullet)$ and $t_P^i(\bullet)$ are assumed to be continuously differentiable.

v. *Capital gains*

The Basak and Croitorn (2001) model is improved by including taxes on capital gains. All capital gains functions are assumed to be continuously differentiable and are defined as:

$$G_S = S(t_2) - S(t_1) = \int_{t_1}^{t_2} 1dS(t) \quad (5.8)$$

$$G_p = P(t_2) - P(t_1) = \int_{t_1}^{t_2} dP(t) \quad (5.9)$$

Thus, dynamic capital gain is expressed as $dS(t)$ and $dP(t)$ (or $S^*(t)dt$ and $P^*(t)dt$).

vi. The accrual capital gains taxation

The continuous tax system developed by Zuckerman (1989) in which contingent capital gains tax liabilities are assessed continuously is applied. For a pure risk-free arbitrage opportunity, investors need to complete buy and sell orders simultaneously to avoid an outflow of capital, and sell the portfolio as a whole to realize a risk-free return. The whole process should happen in a short period of time (after a small change of asset price or new income is received) to guarantee the riskless profit. Thus, all capital gains liabilities are assumed to be realized and paid immediately when the portfolio is sold. As no capital gains tax is deferred over a long time in the process of generating tax arbitrage profits, the time benefit of capital gains tax (the tax payment is deferred until the sale of asset) is not important in an arbitrage opportunity and therefore ignored in this chapter. Capital gains tax is simplified by applying Zuckerman's model, but the incorporation of capital gains tax in a discussion of arbitrage opportunities is still a challenging task. This is because capital gains could turn out to be either positive or negative at the end while income from an asset with a long position must be non-negative. So the calculation of capital gains tax is different from that of income tax. A discussion of tax arbitrage with capital gains tax is also different from that with income tax only. This difference cannot be shown from calculation formulas in this section but can be shown from the discussion of tax arbitrage opportunities in Section 5.3 and 5.4. One of main contributions of this chapter is to find tax arbitrage with capital gains tax and income tax together. Further details are given in Sections 5.3 and 5.4.

vii. *Tax on capital gains*

Investor i has a contingent tax liability on capital gains received from the two risky assets. At time t the investor will have a contingent liability equal to:

$$CGT^i(cgt_S^i(\alpha_S^i(t)S^*(t)) + cgt_P^i(\alpha_P^i(t)P^*(t))) \quad (5.10)$$

where $CGT^i(\bullet)$, $cgt_S^i(\bullet)$ and $cgt_P^i(\bullet)$ are functions of taxation. The argument of the tax liability $CGT^i(\bullet)$ is referred to as investor i 's taxable capital gains, and the argument of the taxable capital gains $cgt_j^i(\bullet)$ is referred to as the total capital gains from asset j . For convenience, investor i 's tax liability on capital gains at time t is sometimes denoted by $CGT^i(t)$, taxable capital gains from asset j at time t by $cgt_j^i(t)$, and their derivatives (marginal tax rate) by $CGT^{i*}(t)$ (or CGT^{i*}) and $cgt_j^{i*}(t)$.

viii. *Heterogeneous tax treatment of gains and losses*

The heterogeneous treatment of capital gains and losses is included in the model, as follows:

$$cgt_S^i(\alpha_S^i S^*(t)) = (\alpha_S^i(t)S^*(t))^+ cgt_{S_+}^i + (\alpha_S^i(t)S^*(t))^- cgt_{S_-}^i \quad (5.11)$$

$$cgt_P^i(\alpha_P^i P^*(t)) = (\alpha_P^i(t)P^*(t))^+ cgt_{P_+}^i + (\alpha_P^i(t)P^*(t))^- cgt_{P_-}^i \quad (5.12)$$

In (5.11) and (5.12), capital gains $(\alpha_S^i(t)S^*(t))^+$ and $(\alpha_P^i(t)P^*(t))^+$ are taxed at tax rates $cgt_{S_+}^i$ and $cgt_{P_+}^i$, and capital losses $(\alpha_S^i(t)S^*(t))^-$ and $(\alpha_P^i(t)P^*(t))^-$ are taxed at tax rates $cgt_{S_-}^i$ and $cgt_{P_-}^i$. This setting may cause the taxation function to be non-differentiable at zero capital gains. The solution to this non-differentiability has previously been developed (Basak and Croitoru 2001). For non-zero capital gains, the functions $CGT^i(\bullet)$, $cgt_S^i(\bullet)$ and $cgt_P^i(\bullet)$ are assumed to be continuously differentiable.

b. *Dynamic Net Return*

In the optimization, the total dynamic net return from the portfolio is based on the value of asset holdings, dynamic income and capital gains. Given current wealth $X^i(t)$, the net incremental amount for the current period is obtained by summing the return from the riskless bond, $[X^i(t) - \alpha_s^i(t)S(t) - \alpha_p^i(t)P(t)]r(t)dt$, with the total return from the two risky assets, $\alpha_s^i(t)[dS(t) + \delta_s(t)dt]$ and $\alpha_p^i(t)[dP(t) + \delta_p(t)dt]$, and deducting total tax liabilities on incomes and capital gains $T^i(t)$ and $CGT^i(t)$. Thus,

$$\begin{aligned}
dX^i(t) = & [X^i(t) - \alpha_s^i(t)S(t) - \alpha_p^i(t)P(t)]r(t)dt + \alpha_s^i(t)[dS(t) + \delta_s(t)dt] \\
& + \alpha_p^i(t)[dP(t) + \delta_p(t)dt] - T^i(t_s^i(\alpha_s^i(t)\delta_s(t)) + t_p^i(\alpha_p^i(t)\delta_p(t)))dt \\
& - CGT^i(cgt_s^i(\alpha_s^i(t)S^*(t)) + cgt_p^i(\alpha_p^i(t)P^*(t)))dt
\end{aligned} \tag{5.13}$$

In addition, after replacing variables for income and capital gains, $dS(t)$, $\delta_s(t)dt$, $dP(t)$ and $\delta_p(t)dt$, with corresponding Brownian motion variables in (5.13), a new equation with stochastic variable W is obtained as follows:

$$\begin{aligned}
dX^i(t) = & X^i(t)r(t)dt + \alpha_s^i(t)S(t)\{[\mu_s(t) - r(t)]dt + \sigma_s(t)dW(t)\} \\
& + \alpha_p^i(t)P(t)\{[\mu_p(t) - r(t)]dt + \sigma_p(t)dW(t)\} - T^i(t)dt - CGT^i(t)dt
\end{aligned} \tag{5.14}$$

Using the Markowitz mean-variance model, the objective of investors will be to maximize the expected value of total net dynamic return in a finite time period $[0, T]$,

$$Max \ E\left[\int_0^T dX^i(t)\right] \tag{5.15}$$

while risk is considered in a risk constraint.

c. Risk Constraint

If the total risk (volatility of return) is equal to zero, this portfolio is called as an arbitrage portfolio when its expected return is larger than the free-risk rate. If total risk is not equal to zero, investors need to constrain it to be under a certain level. To control

for this factor, a risk constraint is added. Since the correlation coefficient between two assets is equal to 1, so the total volatility of the portfolio should be equal to the linear combination of the two assets' own volatility. Consequently, a function is defined to reflect the total portfolio risk as follows:

$$\Phi^i(t) \equiv \alpha_s^i(t)S(t) + (\sigma_p(t)/\sigma_s(t))\alpha_p^i(t)P(t) \leq \Phi_{max}^i \quad (5.16)$$

Here, $\Phi^i(t) \times \sigma_s(t)$ is the portfolio return volatility.

5.2.2. Model for Non-perfectly Correlated Assets

For non-perfectly correlated assets, another optimization model is proposed based on the work of Dammon and Green (1987) in which heterogeneous taxation across investors and assets is considered. Their work is extended by programming the problem in continuous-time (pay-offs from assets are not static but continuous), and adding to the model heterogeneous taxation on income and capital gains, taxation of capital losses and asymmetric tax treatment of long and short positions.

It is assumed that investors are able to invest in a set of underlying assets whose prices A_j follow Brownian motion. Two independent Brownian motions, $W_{i,j}$ and $W_{c,j}$, for each asset's income and capital gains, are introduced so that the asset returns are non-perfectly correlated.

a. Dynamic Returns with Bounds

(i) Dynamic income and capital gains

As in subsection 5.2.1a, it is assumed that there is no fixed correlation between capital gains and income. Stochastic income and prices are generated using different W for different assets j , $j \in \{1, 2, \dots\}$.

$$d\delta_{A_j}(t) = \delta_{A_j}(t)[\mu_{\delta_{A_j}}(t)dt + \sigma_{\delta_{A_j}}(t)dW_{i,j}(t)] \quad (5.17)$$

$$dA_j(t) + \delta_{A_j}(t)dt = A_j(t)[\mu_{A_j}^c(t)dt + \sigma_{A_j}^c(t)dW_{c,j}(t)] \quad (5.18)$$

(ii) *Caps and floors*

In Dammon and Green (1987), static pay-offs from assets are used when tracing arbitrage opportunities. In this work, pay-offs are assumed to be continuous. Basak and Croitoro (2001) discuss tax arbitrage under continuous pay-offs. However, they assume perfect correlation (1 or -1) between selected assets. In reality, most assets have non-perfect correlation between each other. As a result, the use of Basak and Croitoro's model is restricted. To relax this restriction, it is assumed that there is a cap and a floor on each underlying asset's income and capital gains instead of perfect correlation between two selected assets.⁵⁰ In practice, this assumption can be achieved by purchasing derivative instruments (e.g. collars). New dynamic income and capital gains with bounds are then defined as:

$$\delta_{A_j}^B(t)dt = \text{Max}\{\text{Flo}_{\delta_{A_j}}(t), \text{Min}\{\text{Cap}_{\delta_{A_j}}(t), \delta_{A_j}(t)dt\}\} \quad (5.19)$$

$$dA_j^B(t) = \text{Max}\{\text{Flo}_{A_j}(t), \text{Min}\{\text{Cap}_{A_j}(t), dA_j(t)\}\} \quad (5.20)$$

b. *Optimization Model and Arbitrage Opportunity*

Investors want to maximize net portfolio returns $dX^i(t)$ at time t , calculated as follows:

$$dX^i(t) = \sum_j \alpha_{A_j}^i(t)[dA_j^B(t) + \delta_{A_j}^B(t)dt] - T^i[\sum_j t_{A_j}^i(\alpha_{A_j}^i(t)\delta_{A_j}^B(t))]dt \\ - CGT^i[\sum_j cgt_{A_j}^i(\alpha_{A_j}^i(t)A_j^{B*}(t))]dt \quad (5.21)$$

⁵⁰ Here, a cap is defined as a derivative whereby the seller makes payments at the end of each period in which the asset price exceeds the strike price. Similarly, a floor is defined as a derivative whereby the buyer receives payments at the end of each period in which the asset price is below the strike price.

If there is a portfolio guaranteeing that:

$$dX^i(t) \geq 0 \quad \forall dA_j^B(t), \delta_{A_j}^B(t)dt \quad (5.22)$$

$$X^i(t) = \sum_j \alpha_{A_j}^i(t) A_j(t) = 0 \quad (5.23)$$

where there are values of $dA_j^B(t)$ and $\delta_{A_j}^B(t)dt$ such that $dX^i(t) > 0$, then a tax arbitrage opportunity exists for investor i . Note that if such a set of asset holdings exists but the portfolio is limited to a small size to retain the arbitrage opportunity, only a local tax arbitrage opportunity exists. If the portfolio is not limited to a small size, there is a global tax arbitrage opportunity in which investors are able to secure a large amount of risk-free return on an after-tax basis. Section 5.4 contains more analysis of this type of opportunity.

5.3. Tax Arbitrage for Perfectly Correlated Assets

In this section, tax arbitrage is discussed when two risky assets are positively perfectly correlated. This can happen when both financial assets (including derivatives) are derived from the same underlying asset. In Basak and Croitoro (2001), the equilibrium between two perfectly correlated assets in the market is achieved on the assumption that there is equilibrium between two investors when neither of them can realize more risk-free profit by enlarging current holdings further. This assumption, however, does not always hold in the market. In this section, constraints on the existence of tax arbitrage are determined so that the arbitrage opportunity can be found quickly according to given asset parameters and tax functions (rates). In theory, a global tax arbitrage is an arbitrage opportunity for which an investor can always increase his/her risk-free profit by enlarging current holdings given enough liquidity. When a global tax arbitrage opportunity exists, rational investors will want to realize as much risk-free profit as possible, and there will be no equilibrium in the market at that moment. Subsequently, a

large number of buying orders will quickly change the asset price and therefore remove the global tax arbitrage opportunity. In Section 5.3.1, conditions for the existence of tax arbitrage are discussed. The result can help investors to judge if there is a tax arbitrage opportunity between two given assets in the current market and how much risk-free profit can be made at most. In Section 5.3.2, equilibrium between the two assets is discussed and compared to the equilibrium found by Basak and Croitorn (2001). The comparison will help us to assess the importance of capital gains tax when discussing tax arbitrage.

5.3.1. Conditions for the Existence of Arbitrage Opportunities

As discussed in Section 5.2.1, investors want to maximize dynamic net return within a given risk budget (see equation (5.15)). The best way to achieve this objective is to find riskless arbitrage opportunities. In other words, investors want to determine a portfolio between two perfectly correlated assets for which risk $\Phi^i(t)$ (see equation (5.16)) is 0 while its dynamic return $dX(t)$ is larger than that from a riskless ‘bond’ with price B . As in Basak and Croitorn (2001), on a pre-tax basis, if two assets are perfectly positively correlated, an arbitrage opportunity exists when there is a pre-tax mispricing between these two assets. This mispricing is defined as:

$$\Delta_{S,P}(t) = (\mu_S(t) - r(t))/\sigma_S(t) - (\mu_P(t) - r(t))/\sigma_P(t) \neq 0 \quad (5.24)$$

In other words, there is a pre-tax mispricing if two assets’ risk premiums (expected return minus risk-free rate) per unit of volatility risk are different.

Similarly, on an after-tax basis, a so-called tax arbitrage opportunity exists when there is mispricing between two assets.

$$\theta_S(\alpha_S^i(t), \alpha_P^i(t), t) > \theta_P(\alpha_S^i(t), \alpha_P^i(t), t) \quad \forall \delta_S(t), \delta_P(t), S^*(t), P^*(t) \quad (5.25)$$

or,

$$\theta_S(\alpha_S^i(t), \alpha_P^i(t), t) < \theta_P(\alpha_S^i(t), \alpha_P^i(t), t) \quad \forall \delta_S(t), \delta_P(t), S^*(t), P^*(t) \quad (5.26)$$

where

$$\begin{aligned} \theta_S &= [\mu_S(t) - r(t)] / \sigma_S(t) \\ &- [\delta_S(t) t_S^{i*}(\alpha_S^i(t) \delta_S(t)) T^{i*}(t_S^i(\alpha_S^i(t) \delta_S(t)) + t_P^i(\alpha_P^i(t) \delta_P(t)))] / (S(t) \sigma_S(t)) \\ &- [S^*(t) cgt_S^{i*}(\alpha_S^i(t) S^*(t)) CGT^{i*}(cgt_S^i(\alpha_S^i(t) S^*(t)) + cgt_P^i(\alpha_P^i(t) P^*(t)))] / (S(t) \sigma_S(t)) \quad (5.27) \end{aligned}$$

$$\begin{aligned} \theta_P &= [\mu_P(t) - r(t)] / \sigma_P(t) \\ &- [\delta_P(t) t_P^{i*}(\alpha_P^i(t) \delta_P(t)) T^{i*}(t_S^i(\alpha_S^i(t) \delta_S(t)) + t_P^i(\alpha_P^i(t) \delta_P(t)))] / (P(t) \sigma_P(t)) \\ &- [P^*(t) cgt_P^{i*}(\alpha_P^i(t) P^*(t)) CGT^{i*}(cgt_S^i(\alpha_S^i(t) S^*(t)) + cgt_P^i(\alpha_P^i(t) P^*(t)))] / (P(t) \sigma_P(t)) \quad (5.28) \end{aligned}$$

θ_S and θ_P are risk premiums per unit of volatility risk on an after-tax basis for assets S and P. Using θ_S as an example, $[\mu_S(t) - r(t)] / \sigma_S(t)$ is pre-tax risk premium per unit of risk. The following expression

$$\delta_S(t) t_S^{i*}(\alpha_S^i(t) \delta_S(t)) T^{i*}(t_S^i(\alpha_S^i(t) \delta_S(t)) + t_P^i(\alpha_P^i(t) \delta_P(t)))$$

calculates the income tax on investment of asset S. The income tax cost per unit of risk is calculated by dividing this income tax by asset price and volatility risk, $S(t) \sigma_S(t)$. In addition, the following expression

$$S^*(t) cgt_S^{i*}(\alpha_S^i(t) S^*(t)) CGT^{i*}(cgt_S^i(\alpha_S^i(t) S^*(t)) + cgt_P^i(\alpha_P^i(t) P^*(t)))$$

calculates the capital gains tax on investment of asset S. The capital gains tax cost per unit of risk is calculated by dividing this capital gains tax by asset price and volatility risk, $S(t) \sigma_S(t)$. Finally, the risk premium per unit of risk on an after-tax basis is calculated by deducting income tax cost per unit of risk and capital gains tax cost per unit of risk from pre-tax risk premium per unit of risk, as shown in equation (5.27).

In contrast to the pre-tax analysis, mispricing on an after-tax basis depends on marginal tax rates and therefore on asset returns and asset holdings. If the after-tax mispricing of the arbitrage opportunity is non-zero and retains the same sign under any

possible returns from certain asset holdings (see inequalities (5.25) and (5.26)), then it is defined as a tax arbitrage opportunity.

Henceforth, assets are assumed to be neither free of tax on income nor free of tax on capital gains. The two assets are perfectly positively correlated, so their income and capital gains should have a linear relationship. In statistics, the correlation coefficient is a measure of the linear correlation between two variables X and Y, giving a value between +1 and -1 inclusive. There will be a linear relationship between two variables if their correlation coefficient is 1. This measurement was developed by Karl Pearson (1895) from a related idea introduced by Francis Galton in the 1880s. So, with an assumption that two assets' returns are perfectly positively by correlated (correlation coefficient is 1), we obtain:

$$\delta_S(t) = a_\delta \cdot \delta_P(t) + b_\delta \quad (5.29)$$

$$S^*(t) = a_{cg} \cdot P^*(t) + b_{cg} \quad (5.30)$$

$$\delta_S(t) + S^*(t) = a_{total} \cdot (\delta_P(t) + P^*(t)) + b_{total} \quad (5.31)$$

where, a_δ , a_{cg} , a_{total} , b_δ , b_{cg} , b_{total} are all constant, and a_δ , a_{cg} , a_{total} are equal to $\rho_\delta \cdot \sigma_{S_\delta} / \sigma_{P_\delta}$, $\rho_{cg} \cdot \sigma_{S^*} / \sigma_{P^*}$, $\rho_{total} \cdot \sigma_S / \sigma_P$ respectively. Since ρ_δ , ρ_{cg} , ρ_{total} are all 1, a_δ , a_{cg} , a_{total} should all be equal to the ratio of corresponding volatility, $\sigma_{S_\delta} / \sigma_{P_\delta}$, $\sigma_{S^*} / \sigma_{P^*}$, σ_S / σ_P . We then obtain the following expression for after-tax mispricing from (5.27) and (5.28),

$$\begin{aligned} \theta_S - \theta_P = & \frac{\mu_S(t) - r(t)}{\sigma_S(t)} - \frac{\delta_S(t) t_S^{i*}(\alpha_S^i(t) \delta_S(t)) T^{i*}(t_S^i(\alpha_S^i(t) \delta_S(t)) + t_P^i(\alpha_P^i(t) \delta_P(t)))}{S(t) \sigma_S(t)} \\ & - \frac{S^*(t) cgt_S^{i*}(\alpha_S^i(t) S^*(t)) CGT^{i*}(cgt_S^i(\alpha_S^i(t) S^*(t)) + cgt_P^i(\alpha_P^i(t) P^*(t)))}{S(t) \sigma_S(t)} \\ & - \left[\frac{\mu_P(t) - r(t)}{\sigma_P(t)} - \frac{\delta_P(t) t_P^{i*}(\alpha_P^i(t) \delta_P(t)) T^{i*}(t_S^i(\alpha_S^i(t) \delta_S(t)) + t_P^i(\alpha_P^i(t) \delta_P(t)))}{P(t) \sigma_P(t)} \right] \end{aligned}$$

$$\begin{aligned}
& - \frac{P^*(t) cgt_p^{i*}(\alpha_p^i(t) P^*(t)) CGT^{i*}(cgt_s^i(\alpha_s^i(t) S^*(t)) + cgt_p^i(\alpha_p^i(t) P^*(t)))}{P(t) \sigma_p(t)} \Big] \\
& = \Delta_{S,P}(t) - [(\delta_S(t) t_S^{i*}(\bullet)) / (S(t) \sigma_S(t)) - (\delta_P(t) t_P^{i*}(\bullet)) / (P(t) \sigma_P(t))] T^{i*} \\
& \quad - [(S^*(t) cgt_S^{i*}(\bullet)) / (S(t) \sigma_S(t)) - (P^*(t) cgt_P^{i*}(\bullet)) / (P(t) \sigma_P(t))] CGT^{i*} \\
& = \Delta_{S,P}(t) - [b_\delta t_S^{i*}(\bullet) / (S(t) \sigma_S(t)) \\
& \quad + \delta_P(t) (a_\delta t_S^{i*}(\bullet) / (S(t) \sigma_S(t)) - t_P^{i*}(\bullet) / (P(t) \sigma_P(t)))] \cdot T^{i*} \\
& \quad - [b_{cg} cgt_S^{i*}(\bullet) / (S(t) \sigma_S(t)) \\
& \quad + P^*(t) (a_{cg} cgt_S^{i*}(\bullet) / (S(t) \sigma_S(t)) - cgt_P^{i*}(\bullet) / (P(t) \sigma_P(t)))] \cdot CGT^{i*} \\
& = \Delta_{S,P}(t) - MisTax_{in} - MisTax_{cg} \tag{5.32}
\end{aligned}$$

where

$$\begin{aligned}
MisTax_{in} &= \left[\frac{b_\delta t_S^{i*}(\bullet)}{S(t) \sigma_S(t)} + \delta_P(t) \left(\frac{a_\delta t_S^{i*}(\bullet)}{S(t) \sigma_S(t)} - \frac{t_P^{i*}(\bullet)}{P(t) \sigma_P(t)} \right) \right] \cdot T^{i*} \\
MisTax_{cg} &= \left[\frac{b_{cg} cgt_S^{i*}(\bullet)}{S(t) \sigma_S(t)} + P^*(t) \left(\frac{a_{cg} cgt_S^{i*}(\bullet)}{S(t) \sigma_S(t)} - \frac{cgt_P^{i*}(\bullet)}{P(t) \sigma_P(t)} \right) \right] \cdot CGT^{i*}
\end{aligned}$$

Here, \bullet represents the corresponding taxable income and capital gains from asset S, and

$$T^{i*} = T^{i*}(t_S^i(\alpha_S^i(t) \delta_S(t)) + t_P^i(\alpha_P^i(t) \delta_P(t)))$$

$$CGT^{i*} = CGT^{i*}(cgt_S^i(\alpha_S^i(t) S^*(t)) + cgt_P^i(\alpha_P^i(t) P^*(t)))$$

This leads to the following properties giving the conditions for the existence of tax arbitrage (case one).⁵¹

Proposition 1 Given two assets' parameter values, $a_\delta, b_\delta, a_{cg}, b_{cg}, S(t), P(t), \sigma_S(t), \sigma_P(t)$, and tax functions of investor i , $t_S^i(\cdot), t_P^i(\cdot), T^i(\cdot), cgt_S^i(\cdot), cgt_P^i(\cdot), CGT^i(\cdot)$, if for all scenarios in Table A.2 ($\forall i, j$), the following inequality always holds

$$\Delta_{S,P}(t) - MisTax_{in} - MisTax_{cg} \geq 0$$

⁵¹ $MisTax_{in}$ refers to net income tax of the portfolio and $MisTax_{cg}$ refers to net capital gains tax of the portfolio.

then there is a tax arbitrage opportunity between the two assets (case one – referred to (5.35)).

5.3.2. Proof of Proposition 1

From equation (5.32), it follows that there will be a tax arbitrage opportunity if and only if $\theta_S - \theta_P$ is positive (or negative) for any possible value of capital gain $P^*(t)$. A precondition that

$$a_{cg} cgt_S^{i^*}(\bullet) / (S(t)\sigma_S(t)) - cgt_P^{i^*}(\bullet) / (P(t)\sigma_P(t)) = 0$$

must hold. This is because capital gains $P^*(t)$ could be either a large positive or negative figure, which makes it difficult to predict the value of

$$P^*(t)(a_{cg} cgt_S^{i^*}(\bullet) / (S(t)\sigma_S(t)) - cgt_P^{i^*}(\bullet) / (P(t)\sigma_P(t)))$$

in $MisTax_{cg}$ and consequently makes it difficult to keep the value of $\theta_S - \theta_P$ positive (or negative) for any value of capital gains. If $a_{cg} cgt_S^{i^*}(\bullet) / (S(t)\sigma_S(t)) - cgt_P^{i^*}(\bullet) / (P(t)\sigma_P(t))$

is zero for any value of capital gains, then

$$P^*(t)(a_{cg} cgt_S^{i^*}(\bullet) / (S(t)\sigma_S(t)) - cgt_P^{i^*}(\bullet) / (P(t)\sigma_P(t)))$$

will also be zero for any value of capital gains, and therefore the value of $MisTax_{cg}$ will be independent of $P^*(t)$.

Consequently, $\theta_S - \theta_P$ will always be positive (or negative) under any value of capital

gains. In other words, its sign will be constant for any value of capital gains. However,

when asset holdings and short sale amounts are not large enough, the volatility of

capital gains will change the argument (\bullet) and make their marginal tax rate $cgt_j^{i^*}(\bullet)$

move between its lower and upper bounds, making it impossible to retain the equality

$$P^*(t)(a_{cg} cgt_S^{i^*}(\bullet) / (S(t)\sigma_S(t)) - cgt_P^{i^*}(\bullet) / (P(t)\sigma_P(t))) = 0. \text{ Thus, local arbitrage opportunities in}$$

which investors expect to obtain a limited riskless return do not exist for perfectly

correlated assets. So it is only necessary to discuss global arbitrage opportunities.

In theory, for a global arbitrage opportunity, investors can expect to secure a large riskless positive return by enlarging their positions. In practice, this enlargement is subject to real market restrictions, such as the number of assets in the market. Thus all conditions to secure a large profit mentioned in the following discussion are theoretical. Investors can increase the return by enlarging the size of the global arbitrage portfolio until some real market practical limit is reached. In the following analysis, therefore, this chapter seeks to prove mathematically that there may exist a portfolio (stated in Proposition 1) with a large size (M) which has a positive (or negative) net marginal mispricing, and further enlargement will not change the mispricing.⁵² So investors can keep enlarging the portfolio to lock in more and more positive risk-free return (global arbitrage opportunity). It can be seen from equation (5.32) that the value of the mispricing depends on marginal tax rates, $t_s^{i*}(\bullet), t_p^{i*}(\bullet), cgt_s^{i*}(\bullet), cgt_s^{i*}(\bullet), CGT^{i*}, T^{i*}$, and the marginal tax rates depend on the argument of the corresponding tax function. To fix these marginal tax rates at a certain level, a minimum size of holding, M , is set on each asset such that even with the smallest non-zero income or capital gain (e.g., £0.01), the investor will be subject to the top tax rate. In other words, M is assumed large enough so that $t_s^{i*}(M \cdot \delta_s) = t_s^{i*}(M)$ for any non-zero value of δ_s . For example, if the investor pays the top income tax rate when income is over £20,000, this minimum size will be $20,000/0.01=2,000,000$. The minimum size for capital gains (both long and short positions) is also calculated, and the larger value of these minimum sizes is set as the value of M . If both income and capital gains tax rates are constant (independent of the total return), M could be any value. In the following proof, the argument of M means the investor is subject to the top tax rate (i.e., $t_s^{i*}(M)$). In other words,

$$\alpha_s^i(t) = M; \tag{5.33}$$

⁵² In the mathematical proof, we assume that the market can provide enough liquidity so that all buying or selling orders can be executed at the current price.

In addition, setting the portfolio risk equal to 0, the following is obtained from equation (5.16):

$$\alpha_S^i(t) \times S(t) \times \sigma_S(t) + \alpha_P^i(t) \times P(t) \times \sigma_P(t) = 0 \quad (5.34)$$

As a result, there will be a global arbitrage opportunity if and only if:⁵³

Case One:

$$\begin{aligned} \theta_S(+M, -M \cdot S(t)\sigma_S(t) / (P(t)\sigma_P(t)), t) &\geq \theta_P(+M, -M \cdot S(t)\sigma_S(t) / (P(t)\sigma_P(t)), t) \\ \forall \delta_S(t), \delta_P(t), S'(t), P'(t) \end{aligned} \quad (5.35)$$

Case Two:

$$\begin{aligned} \theta_S(-M, +M \cdot S(t)\sigma_S(t) / (P(t)\sigma_P(t)), t) &\leq \theta_P(-M, +M \cdot S(t)\sigma_S(t) / (P(t)\sigma_P(t)), t) \\ \forall \delta_S(t), \delta_P(t), S'(t), P'(t) \end{aligned} \quad (5.36)$$

In both cases, the inequality exists for at least one possible outcome.

In these two cases, all marginal tax rates, $t_S^{i*}(\bullet)$, $t_P^{i*}(\bullet)$, $cgt_P^{i*}(\bullet)$, $cgt_S^{i*}(\bullet)$, $T^{i*}(\bullet)$ and $CGT^{i*}(\bullet)$, have three possible outcomes depending on the argument $(+M)$, $(-M)$ and (0) . In addition, from equation (5.32), it follows that the mispricing on an after-tax basis, $\theta_S - \theta_P$, can be divided into three independent parts: pre-tax mispricing between assets, different income taxes and different capital gains taxes. Since the first part is only dependent on constant parameters, only the second and third parts need to be discussed in detail in the proof. (Note: since the proof of Case Two is simply the proof of Case One when S is set as P and P is set as S, only the proof of Case One is given).

a. Differential Tax on Capital Gains

In this section, different capital gains taxes which are dependent on $S^*(t)$ and $P^*(t)$ are discussed. As capital could increase or decrease, the arguments $\alpha_S(t)S^*(t)$ and

⁵³In arbitrage, investors will always hold the under-priced asset while short selling the overpriced asset.

$\alpha_p(t)P^*(t)$ could be positive, negative or zero. Different scenarios are discussed dependent on the value of capital gains.

i. When $S^(t) \neq 0$ and $P^*(t) \neq 0$*

As discussed above, to guarantee a positive mispricing in (5.32), the first condition which must hold is that

$$a_{cg} cgt_S^{i*}(M)/(S(t)\sigma_S(t)) - cgt_P^{i*}(M)/(P(t)\sigma_P(t)) = 0.$$

In addition, if $b_{cg} = 0$, according to equation (5.30), $S^*(t)$ and $P^*(t)$ will always have the same sign. So the condition

$$a_{cg} cgt_S^{i*}(M)/(S(t)\sigma_S(t)) - cgt_P^{i*}(M)/(P(t)\sigma_P(t)) = 0$$

can be transformed to

$$a_{cg} cgt_S^{i*}(-M)/(S(t)\sigma_S(t)) - cgt_P^{i*}(-M)/(P(t)\sigma_P(t)) = 0$$

and

$$a_{cg} cgt_S^{i*}(+M)/(S(t)\sigma_S(t)) - cgt_P^{i*}(+M)/(P(t)\sigma_P(t)) = 0 \quad (5.37)$$

However, if $b_{cg} = 0$ does not hold, the arguments (M) for the two assets could be different (one is $(+M)$ and the other is $(-M)$) depending on the value of their capital gains $S^*(t)$ and $P^*(t)$. Thus, the marginal tax rate for positive and negative capital gains needs to be the same,

$$cgt_S^{i*}(+M) = cgt_S^{i*}(-M) \quad (5.38)$$

$$cgt_P^{i*}(+M) = cgt_P^{i*}(-M) \quad (5.39)$$

such that the following condition can be retained:

$$a_{cg} cgt_S^{i*}(M)/(S(t)\sigma_S(t)) - cgt_P^{i*}(M)/(P(t)\sigma_P(t)) = 0$$

Capital gains taxes are then given by:

$$MisTax_{cg} = [(b_{cg}/(S(t)\sigma_S(t)))cgt_S^{i*}(+M \text{ or } -M)] \cdot CGT^{i*}(+M, -M \text{ or } 0) \quad (5.40)$$

Now, the value of $MisTax_{cg}$ only depends on certain asset parameters and is static at several specific amounts.

ii. When $S^*(t) = 0$

Since the two assets are perfectly positively correlated, we obtain

$$P^*(t) = -b_{cg} / a_{cg} \quad (5.41)$$

Substituting this into equation 5.32 and setting the argument of $cg t_p^{i*}$ and CGT_p^{i*} as $\frac{b_{cg}}{a_{cg}}$,

we have total marginal capital gains taxes:

$$MisTax_{cg} = \left\{ -\frac{b_{cg}}{a_{cg}} cg t_p^{i*} \left[M \cdot \frac{b_{cg}}{a_{cg}} \right] / (P(t) \sigma_p(t)) \right\} \times CGT_p^{i*} \left(M \cdot \frac{b_{cg}}{a_{cg}} \right) \quad (5.42)$$

iii. When $P^*(t) = 0$

Since the two assets are perfectly positively correlated, we obtain

$$S^*(t) = b_{cg} \quad (5.43)$$

Total marginal capital gains taxes are then given by:

$$MisTax_{cg} = \{ b_{cg} cg t_s^{i*} [M \cdot (b_{cg})] / (S(t) \sigma_s(t)) \} \times CGT_s^{i*} [M \cdot (b_{cg})] \quad (5.44)$$

In conclusion, with the constraint on the marginal capital gains tax rate (equations 5.38 and 5.39), the total capital gains tax ($MisTax_{cg}$) is changed to be static (when $M = +M, -M$ or 0) rather than continuous. This amount is dependent on the value of the asset parameters, such as b_{cg} , $S(t)$ and $\sigma_s(t)$. With static capital gains tax payments, it is possible to find an arbitrage opportunity.

b. *Differential Tax on Income*

In this section, different income taxes which are dependent on $\delta_s(t)$ and $\delta_p(t)$ are discussed. Since both $\delta_s(t)$ and $\delta_p(t)$ must be non-negative, the discussion of income tax is different from that of capital gains tax.

$$\delta_s(t) = a_\delta \cdot \delta_p(t) + b_\delta \quad (5.45)$$

$$\delta_s(t), \delta_p(t) \geq 0, a_\delta = \rho_\delta \cdot \sigma_{S_\delta} / \sigma_{P_\delta} = 1 \cdot \sigma_{S_\delta} / \sigma_{P_\delta} \geq 0^{54} \quad (5.46)$$

From equation (5.45) and (5.46), we obtain:

$$\delta_s(t) \geq \max\{0, b_\delta\} \quad (5.47)$$

$$\delta_p(t) \geq \max\{0, -b_\delta / a_\delta\} \quad (5.48)$$

Dependent on the value of b_δ , we can divide the discussion of income tax into three scenarios, (i) to (iii), as below.

$$i. \quad b_\delta > 0 \Rightarrow \delta_s(t) \geq b_\delta ; \delta_p(t) \geq 0$$

Similar to the analysis of capital gains tax, the bound of $MisTax_{in}$ is now discussed depending on the value of $\delta_p(t)$.

1) When $\delta_p(t) \neq 0$

An arbitrage opportunity will exist, if and only if

$$a_\delta t_S^{i*}(+M) / (S(t)\sigma_S(t)) - t_P^{i*}(-M) / (P(t)\sigma_P(t)) \leq 0 \quad (5.49)$$

From equation (5.32), (5.33) and (5.35), the following can be obtained

$$\begin{aligned} MisTax_{in} &= \left[\frac{b_\delta t_S^{i*}(\bullet)}{S(t)\sigma_S(t)} + \delta_p(t) \left(\frac{a_\delta t_S^{i*}(\bullet)}{S(t)\sigma_S(t)} - \frac{t_P^{i*}(\bullet)}{P(t)\sigma_P(t)} \right) \right] \cdot T^{i*} \\ &= \left[\frac{b_\delta t_S^{i*}(+M)}{S(t)\sigma_S(t)} + \delta_p(t) \left(\frac{a_\delta t_S^{i*}(+M)}{S(t)\sigma_S(t)} - \frac{t_P^{i*}(-M)}{P(t)\sigma_P(t)} \right) \right] \cdot T^{i*} \end{aligned}$$

⁵⁴ In other words, since income from both assets, S and P, are perfectly positively correlated, a_δ must be positive.

As $t_S^{i^*}(+M)$, $\frac{b_\delta t_S^{i^*}(+M)}{S(t)\sigma_S(t)}$ and $a_\delta t_S^{i^*}(+M)/(S(t)\sigma_S(t)) - t_P^{i^*}(-M)/(P(t)\sigma_P(t))$ are constant,

the value of the $MisTax_{in}$ is only changed by $\delta_p(t)$ and T^{i^*} . Since the marginal tax rate T^{i^*} is non-negative, there will be a constant upper bound for the value of $MisTax_{in}$, if and only if equation (5.49) holds, as below

$$MisTax_{in} \leq \lim_{\delta_p \rightarrow 0^+} \{ [b_\delta t_S^{i^*}(+M)/(S(t)\sigma_S(t)) + \delta_p(t)(a_\delta t_S^{i^*}(+M)/(S(t)\sigma_S(t)) - t_P^{i^*}(-M)/(P(t)\sigma_P(t))] \cdot T^{i^*} \} \quad (5.50)$$

Considering the argument of T^{i^*} , the value of $MisTax_{in}$ is

$$\leq [t_S^{i^*}(+M)b_\delta/(S(t)\sigma_S(t))]T^{i^*}(t_S^i(+M \cdot \delta_S(t)) + t_P^i(-M \cdot \delta_P(t))) \quad (5.51)$$

Therefore⁵⁵

$$t_S^i(+M \cdot \delta_S(t)) + t_P^i(-M \cdot \delta_P(t)) = +M, -M \text{ or } 0$$

and

$$MisTax_{in} \leq \max_{m \in \{+M, 0, -M\}} \{ [t_S^{i^*}(+M) \cdot b_\delta / (S(t)\sigma_S(t))] \cdot T^{i^*}(m) \} \quad (5.52)$$

The constant upper bound for $MisTax_{in}$ in equation (5.52) leads to a constant lower bound for $\theta_S - \theta_P$. Now, to guarantee the arbitrage opportunity, we only need to consider this lower bound to see whether it will make the expression $\theta_S - \theta_P$ non-negative for any value of $\delta_S(t)$ and $\delta_P(t)$.

2) When $\delta_P(t) = 0$

Income tax payments are constant at

$$MisTax_{in} = [t_S^{i^*}(+M) \cdot b_\delta / (S(t)\sigma_S(t))] \cdot T^{i^*}(+M) \quad (5.53)$$

⁵⁵Here, when $\delta_S(t) \neq 0$, $t_S^{i^*} \equiv t_S^{i^*}(+M)$, and $t_S^{i^*}(+M) \cdot b_\delta / (S(t)\sigma_S(t))$ is constant. Thus the value of the $MisTax_{in}$ above is only changed by $\delta_p(t)(a_\delta t_S^{i^*}(+M)/(S(t)\sigma_S(t)) - t_P^{i^*}(-M)/(P(t)\sigma_P(t)))$

$$\text{ii. } b_\delta < 0 \Rightarrow \delta_S(t) \geq 0; \delta_P(t) \geq -b_\delta / a_\delta$$

1) When $\delta_S(t) \neq 0$

Similar to the analysis in (i), to guarantee the existence of an arbitrage opportunity, we require:

$$t_S^{i*}(+M)/(S(t)\sigma_P(t)) - t_P^{i*}(-M)/(P(t)\sigma_P(t)) \leq 0 \quad (5.54)$$

so that there is a constant upper bound for income taxes.

$$\begin{aligned} \text{MisTax}_{in} &\leq \max_{m \in \{+M, 0, -M\}} \left\{ \lim_{\delta_S \rightarrow 0^+} \{t_S^{i*}(+M) \bullet b_\delta / (S(t)\sigma_S(t)) + \delta_P(t)\} \right. \\ &\quad \left. \times [a_\delta t_S^{i*}(+M)/(S(t)\sigma_S(t)) - t_P^{i*}(-M)/(P(t)\sigma_P(t))] \times T^{i*}(m) \right\} \\ &= \max_{m \in \{+M, 0, -M\}} \left\{ [t_S^{i*}(+M) \bullet b_\delta / (S(t)\sigma_S(t)) + (-b_\delta / a_\delta)] \right. \\ &\quad \left. \times [a_\delta t_S^{i*}(+M)/(S(t)\sigma_S(t)) - t_P^{i*}(-M)/(P(t)\sigma_P(t))] \times T^{i*}(m) \right\} \end{aligned} \quad (5.55)$$

The argument of $t_S^{i*}(\cdot)$ depends on the value of income from asset S. In equation (5.55), the income from asset S is set to be infinitely close to but not equal to 0. Thus, multiplied by a large holding amount, the argument of $t_S^{i*}(\cdot)$ should still be +M but not 0 (see definition of +M in section 5.3.2).

2) When $\delta_S(t) = 0$

Since the income is 0, $\delta_P(t) = -b_\delta / a_\delta$. From (5.55), income tax payments are constant at

$$\text{MisTax}_{in} = [-b_\delta / a_\delta] (-t_P^{i*}(-M)/(P(t)\sigma_P(t))) \cdot T^{i*}(-M) \quad (5.56)$$

$$\text{iii. } b_\delta = 0 \Rightarrow \delta_S(t) = a \cdot \delta_P(t) \geq 0$$

1) When $\delta_S(t) \neq 0$ and $\delta_P(t) \neq 0$

To guarantee the existence of an arbitrage opportunity, we require

$$a_{\delta} t_S^{i*}(+M)/(S(t)\sigma_S(t)) - t_P^{i*}(-M)/(P(t)\sigma_P(t)) \leq 0 \quad (5.57)$$

so there is a constant upper bound for income taxes.

$$\begin{aligned} MisTax_{in} &\leq \max_{m \in [+M, 0, -M]} \left\{ \lim_{\delta_p \rightarrow 0^+; \delta_S \rightarrow 0^+} [t_S^{i*}(+M) \cdot b_{\delta} / (S(t)\sigma_S(t)) \right. \\ &\quad \left. + \delta_p(t) \times (a_{\delta} t_S^{i*}(+M)/(S(t)\sigma_S(t)) - t_P^{i*}(-M)/(P(t)\sigma_P(t))) \right] \times T^{i*}(m) \} \\ &= 0 \end{aligned} \quad (5.58)$$

which leads to a constant lower bound for $\theta_S - \theta_P$. To guarantee the arbitrage opportunity, we only need to consider this lower bound to see whether it will make the expression $\theta_S - \theta_P$ non-negative for any value of $\delta_S(t)$ and $\delta_P(t)$.

2) When $\delta_S^i(t) = 0$ and $\delta_P^i(t) = 0$

Income tax payments are constant at 0. All possible scenarios of the values of $MisTax_{cg}$ and $MisTax_{in}$ have been discussed, and the upper bound for each of them in each scenario has been found. To check if the value of mispricing between two assets on an after-tax basis is non-negative consistently for any value of income and capital gains (case one tax arbitrage), we just need to make sure that in each scenario, even when the values of $MisTax_{cg}$ and $MisTax_{in}$ are at their maximum, $\theta_S - \theta_P$ is non-negative, as shown in Proposition 1. Discussions in section 5.3.2 are summarized in Table 5.2 and support Proposition 1.

Table 5. 2 Arbitrage for Case One

Constraints	
1:	$\frac{a_{cg} cgt_S^{i*}(M)}{S(t)\sigma_S(t)} - \frac{cgt_P^{i*}(M)}{P(t)\sigma_P(t)} = 0$ ⁵⁶ see eq. (5.37)
2:	$\frac{a_{\delta} t_S^{i*}(+M)}{S(t)\sigma_S(t)} - \frac{t_P^{i*}(-M)}{P(t)\sigma_P(t)} \leq 0$ see eq. (5.49)
3:	$\theta_S(+M, -M, t) - \theta_P(+M, -M, t) \equiv \Delta_{S,P}(t) - MisTax_{in}(j) - MisTax_{cg}(i) \geq 0, \forall i, j$

⁵⁶ If $b_{cg} \equiv 0$, we only require $\frac{a_{cg} cgt_S^{i*}(+M)}{S(t)\sigma_S(t)} = \frac{cgt_P^{i*}(+M)}{P(t)\sigma_P(t)}$ and $\frac{a_{cg} cgt_S^{i*}(-M)}{S(t)\sigma_S(t)} = \frac{cgt_P^{i*}(-M)}{P(t)\sigma_P(t)}$

i	$S^*(t); P^*(t)$	$MisTax_{cg}(i)$
i=1	$S^*(t) \neq 0$	$= [\frac{b_{cg}}{S(t)\sigma_S(t)} cgt_S^{i*} (+M \text{ or } -M)] \cdot CGT^{i*} (+M, +M \text{ or } 0)$ ⁵⁷ (5.40)
i=2	$S^*(t) = 0$	$= \{ -\frac{b_{cg}/a_{cg}}{P(t)\sigma_P(t)} cgt_P^{i*} [-M \cdot (b_{cg} / a_{cg})] \} \times CGT^{i*} (-M \cdot (b_{cg} / a_{cg}))$ (5.42)
i=3	$P^*(t) = 0$	$= \{ \frac{b_{cg}}{S(t)\sigma_S(t)} cgt_S^{i*} [+M \cdot (b_{cg})] \} \times CGT^{i*} [+M \cdot (b_{cg})]$ (5.44)
If $b_\delta > 0$		
j	$\delta_S(t); \delta_P(t)$	$MisTax_{in}(j)$
j=1	$\delta_P(t) \neq 0$	$\leq \max_{m \in \{+M, 0, -M\}} \{ [\frac{b_\delta}{S(t)\sigma_S(t)} t_S^{i*} (+M)] \times T^{i*}(m) \}$ (5.52)
j=2	$\delta_P(t) = 0$	$= [\frac{b_\delta}{S(t)\sigma_S(t)} t_S^{i*} (+M)] \cdot T^{i*} (+M)$ (5.53)
If $b_\delta < 0$		
j	$\delta_S(t); \delta_P(t)$	$MisTax_{in}(j)$
j=1	$\delta_S(t) \neq 0$	$\leq \max_{m \in \{+M, 0, -M\}} \{ [\frac{b_\delta}{S(t)\sigma_S(t)} t_S^{i*} (+M) + (b_\delta / a_\delta) \times (\frac{a_\delta t_S^{i*} (+M)}{S(t)\sigma_S(t)} - \frac{t_P^{i*} (-M)}{P(t)\sigma_P(t)})] \times T^{i*}(m) \}$ (5.55)
j=2	$\delta_S(t) = 0$	$= [b_\delta / a_\delta] (-\frac{t_P^{i*} (-M)}{P(t)\sigma_P(t)}) \cdot T^{i*} (-M)$ (5.56)
If $b_\delta = 0$		
j	$\delta_S(t); \delta_P(t)$	$MisTax_{in}(j)$
j=1	$\delta_S(t) \neq 0$	≤ 0 (5.57)
j=2	$\delta_S(t) = 0$	$= 0$ (5.58)

5.3.3. Further Special Cases

The discussion above is based on the assumption that both assets are subject to income tax and capital gains tax. However, there are cases in which assets may be tax-free with respect to either income tax or capital gains tax.

If one of the assets is free of capital gain tax ($cgt_j^i(M) = 0$)

⁵⁷For this case, the argument of $cgt_S^{i*}(\cdot)$ has two possible values, $+M$ and $-M$. The argument of $CGT^{i*}(\cdot)$ has three possible values, $+M$, 0 and $-M$. Thus in total, the whole expression has six possible values. If arbitrage exists, all six values must hold the sign of after-tax mispricing unchanged.

According to Constraint 1 in Table 5.2, if a tax arbitrage opportunity exists, the other asset must also be free of capital gains tax. Thus, the existence of tax arbitrage depends only on income taxes.

If one of the assets is free of income taxes ($t_j^i(M) = 0$)

(a) If $t_s^i(M) = 0$ and $t_p^i(M) \neq 0$

According to Constraint 2 in Table 5.2, only an arbitrage opportunity in Case One may exist.

(b) If $t_s^i(M) \neq 0$ and $t_p^i(M) = 0$

According to Constraint 2 in Table 5.2, only an arbitrage opportunity in Case Two may exist.

(c) If $t_s^i(M) = 0$ and $t_p^i(M) = 0$

The existence of a tax arbitrage opportunity depends only on capital gains taxes.

5.3.4. Implications

According to Table 5.2, if the arbitrage opportunity in Case One exists, the values of the parameters must satisfy the three constraints, 1, 2 and 3. In the constraints, capital gains tax and income tax are calculated separately by different functions according to their different tax rules. There are two findings from the above analysis. First, a local arbitrage opportunity in which investors expect to obtain a finite riskless return does not exist between perfectly correlated assets without caps and floors. This is because the uncertainty of the sign and unbounded positive size of capital gains make it impossible to retain the condition

$$a_{cg} cgt_s^{i*}(\cdot) / (S(t)\sigma_s(t)) - cgt_p^{i*}(\cdot) / (P(t)\sigma_p(t)) = 0$$

with a small holding. So a riskless return cannot be guaranteed without restrictions imposed on capital gains. Second, a global arbitrage opportunity exists between perfectly correlated assets without caps and floors but does not exist consistently depending on asset prices. Only when the prices of the two assets satisfy the condition

$$a_{cg} cgt_s^{i*}(M)/(S(t)\sigma_s(t)) - cgt_p^{i*}(M)/(P(t)\sigma_p(t)) = 0$$

with a large holding can a global arbitrage opportunity exist, and the top tax rate needs to be considered in the constraints as well. With these constraints, investors can quickly determine whether there is a global tax arbitrage opportunity between two perfectly correlated assets with given asset parameters and tax functions.

5.3.5. *Equilibrium with Tax Arbitrage*

In real markets, investors usually optimize portfolios based on expectations derived from assets' historical performance. In this continuous-time model, however, it is assumed that all optimal portfolios are achieved when current returns are used as expectations of the future in the optimization (note: this return is dynamic and should follow a random walk in future). Under this assumption, there will be a dynamic equilibrium between risky assets for investors when there is no global arbitrage opportunity. In equilibrium, the after-tax *mispricing* should be zero so that there is no motivation for investors to change the current portfolio to get a better return under the risk budget.

a. *Equilibrium for a Single Investor*

In equilibrium on an after-tax basis, two perfectly correlated assets' after-tax risk premium per unit of volatility risk must be the same. Consequently, the following equation is obtained (the definition of this premium can be obtained from equation (5.27) and (5.28)):

$$\theta_S(\alpha_S^i(t), \alpha_P^i(t), t) = \theta_P(\alpha_S^i(t), \alpha_P^i(t), t) \quad (5.59)$$

Including the pre-tax mispricing $\Delta_{S,P}(t)$ in this equation, we obtain the following expression:

$$\begin{aligned} \Delta_{S,P}(t) &= \Delta_{S,P}(t) - (\theta_S(\alpha_S^i(t), \alpha_P^i(t), t) - \theta_P(\alpha_S^i(t), \alpha_P^i(t), t)) \\ &= [(\delta_S(t)t_S^{i*}(\alpha_S^i(t)\delta_S(t)))/(S(t)\sigma_S(t)) - (\delta_P(t)t_P^{i*}(\alpha_P^i(t)\delta_P(t)))/(P(t)\sigma_P(t))] \\ &\quad \times T^{i*}(t_S^i(\alpha_S^i(t)\delta_S(t)) + t_P^i(\alpha_P^i(t)\delta_P(t))) \\ &\quad + [(S^*(t)cgt_S^{i*}(\alpha_S^i(t)S^*(t)))/(S(t)\sigma_S(t)) - (P^*(t)cgt_P^{i*}(\alpha_P^i(t)P^*(t)))/(P(t)\sigma_P(t))] \\ &\quad \times CGT^{i*}(cgt_S^i(\alpha_S^i(t)S^*(t)) + cgt_P^i(\alpha_P^i(t)P^*(t))) \end{aligned} \quad (5.60)$$

Proposition 3 in Appendix A.1 presents the resulting condition on rational holdings $(\hat{\alpha}_S^i, \hat{\alpha}_P^i)$ in equilibrium which is derived from equation (5.60).

b. Market Equilibrium

This section assumes that there is an equilibrium between two investors, and pre-tax mispricing $\Delta_{S,P}(t)$ exists. This allows us to explain the role of mispricing in financial markets and the properties of equilibrium. The following analysis assumes that the taxation functions are continuously differentiable. According to the conclusion of Basak and Croitoro (2001), with given risk exposure budget for investor 1 and 2, $\Phi^1(t)$ and $\Phi^2(t)$, the mispricing has to be such that two investors' holdings, α_S^i and α_P^i ($i=1, 2$), determined from non-satiation, clear the financial markets. This yields the expression for mispricing reported in Proposition 4 in Appendix A.2. The proposition on market equilibrium in Basak and Croitoro (2001) is improved by including tax on capital gains.

As in the conclusion of Basak and Croitoro (2001) at the individual level, the pre-tax mispricing compensates each investor for differences in taxation across securities: they will adjust their portfolio holdings to the equilibrium level until the compensation

is exact. By doing this a single investor will in general not simply choose the portfolio strategy that minimizes the amount of tax paid. However, at the market level, investors' demand/behavior is aggregated to that of a representative investor in the market. Adopting this macro view, it can be proved that the pre-tax mispricing in the market equilibrium is set so that the representative investor effectively minimizes the amount of aggregate tax paid. In this work, after including capital gains tax, it is found that in equilibrium, the representative investor minimizes the aggregate amount of income tax payments as well as capital gains tax payments, $T^1(t) + T^2(t) + CGT^1(t) + CGT^2(t)$, as proved in Proposition 5 in Appendix A.3.

From a mathematical viewpoint, the inclusion of capital gains tax leads to a simple conclusion that equilibrium is obtained when aggregate income tax is minimized if only income tax is considered, and equilibrium is obtained when aggregate tax on both income and capital gains is minimized if both income and capital gains tax are considered. However, this conclusion has a significant implication in finance. It is assumed that there is only one asset in the market with a constant supply N , and that there are two investors, A and B, subject to different tax rates (A has a high tax rate on income but a low tax rate on capital gains while B has a low tax rate on income but a high tax rate on capital gains). We further assume that both investors are passive and the market will decide on the distribution of the asset weight between them. Market equilibrium is obtained when the total net return of both investors are maximized (Pareto Optimality). In fact, no matter how the market distributes the asset weight, the total gross return from both investors is constant and independent of the distribution. The heterogeneous tax rates between the two investors, however, means that the total tax payment from both investors and therefore their total net return depends on the distribution of the asset weight. For example, more asset weight to investor A will decrease the total capital gains tax but increase the total income tax, and more asset

weight to investor B will decrease the total income tax but increase the total capital gains tax. If the conclusion of Basak and Croitor (2001) is adopted, then the market always distributes all asset weight to investor B to minimize the aggregate income tax payment, and this distribution is regarded as equilibrium in the market. Nevertheless, this equilibrium should never exist since an efficient market will find the best trade-off between income tax and capital gains tax to minimize the aggregate tax payment. As a result, investors as a whole can get maximum net reward from the market. This conclusion is also applicable when there are more than two assets and more than two investors in the market. So the conclusion of Basak and Croitor (2001) does not hold, and my work proves that in the discussion of market equilibrium on an after-tax basis, if investors are subject to capital gains tax, market equilibrium obtained with an income tax only assumption would not be applicable after incorporating capital gains tax.

5.4. Tax Arbitrage for Non-perfectly Correlated Assets

This section extends the model introduced in Section 5.2.2 to determine the constraints that will allow the existence of tax arbitrage opportunities between non-perfectly correlated assets with caps and floors.

5.4.1. Single Investor

It is first assumed that there is only one investor and two risky assets.

a. Local Arbitrage with Static Payoff

In local arbitrage, an investor obtains a small riskless return by holding one asset and short-selling the other, and this return cannot be increased by enlarging the size of the arbitrage portfolio. If the payoff of assets is static, then proposition 6 below can be

obtained. $A_j^B(t)$ and $\delta_{A_j}^B(t)$ are ‘asset price with bound’ and ‘income with bound’, respectively.

Proposition 6. There will be a local tax arbitrage opportunity if and only if there is a set of asset holdings, $\alpha_{A_j}^i(t)$, for investor i , that solves the following system of inequalities:

$$\begin{aligned} dX^i(t) &= \sum_j \alpha_{A_j}^i(t) [dA_j^B(t) + \delta_{A_j}^B(t)dt] - T^i \left[\sum_j t_{A_j}^i (\alpha_{A_j}^i(t) \delta_{A_j}^B(t)) \right] dt \\ &\quad - CGT^i \left[\sum_j cgt_{A_j}^i (\alpha_{A_j}^i(t) A_j^{B^*}(t)) \right] dt \\ &\geq 0 \quad \forall dA_j^B(t), \delta_{A_j}^B(t)dt, j=1,2 \end{aligned} \quad (5.61)$$

$$X^i(t) = \sum_j \alpha_{A_j}^i(t) A_j(t) = 0 \quad (5.62)$$

In (5.61),

$$\sum_j \alpha_{A_j}^i(t) [dA_j^B(t) + \delta_{A_j}^B(t)dt]$$

is the pre-tax total return from the portfolio in the period dt while

$$T^i \left[\sum_j t_{A_j}^i (\alpha_{A_j}^i(t) \delta_{A_j}^B(t)) \right] dt \quad \text{and} \quad CGT^i \left[\sum_j cgt_{A_j}^i (\alpha_{A_j}^i(t) A_j^{B^*}(t)) \right] dt$$

are income and capital gains tax payments respectively. Thus, the expression calculates the net return from the portfolio in the period dt . If this net return is non-negative for any value of income and capital gains in the period, there is an arbitrage opportunity on an after-tax basis.

Proposition 7. The local tax arbitrage opportunities can be divided into two types.

1. *Arbitrage type A*

Suppose that holdings exist such that the dynamic return of the portfolio is

$$dX^i(t) > 0 \quad \forall dA_j^B(t), \delta_{A_j}^B(t)dt, j=1,2 \quad (5.63)$$

In arbitrage type A, the investor expects to obtain a positive riskless return.

2. Arbitrage type B

Suppose that holdings exist such that there is at least one possible return

$dA_j^B(t), \delta_{A_j}^B(t)dt, j=1,2$ for which

$$dX^i(t) = 0 \quad (5.64)$$

In arbitrage type B, the investor only has a chance of obtaining a positive riskless return. In other words, the portfolio will return either 0 or a positive net profit to arbitrageurs. A non-negative rather than a positive net profit is guaranteed.

b. Local Arbitrage with Continuous Payoff

In the model given by equations (5.63) and (5.64), if the optimal solution returns a non-negative objective function value, the portfolio obtained is a tax arbitrage portfolio. This optimization method is applicable for assets with static payoffs. There is, however, a technical issue for assets with continuous payoffs. How can one use finite constraints to present infinite possible payoffs (continuous payoffs)? To deal with this issue, cap and floor options are introduced. As in Section 5.2, cap and floor options can set a boundary on both income and capital gains for each asset. Without taxation, it is easy to confirm the existence of arbitrage opportunities if one asset's floor of total return is above the other asset's cap of total return. However, with taxation, the assets' caps and floors will depend on marginal tax rates and consequently the total amount of asset holdings. Thus, the marginal after-tax floor and cap on incomes, $Flo_{\delta_{A_j}}^{tax}(t, \alpha_{A_j}^i(t))$ and $Cap_{\delta_{A_j}}^{tax}(t, \alpha_{A_j}^i(t))$, and on capital gains, $Flo_{\delta_{A_j}}^{tax}(t, \alpha_{A_j}^i(t))$ and $Cap_{\delta_{A_j}}^{tax}(t, \alpha_{A_j}^i(t))$ at time t are introduced. Their values are calculated as follows:

$$Flo_{\delta_{A_j}}^{tax}(t, \alpha_{A_j}^i(t)) =$$

$$\frac{Flo_{\delta_{A_j}}(t)\{1-t_{A_j}^i[\alpha_{A_j}^i(t)Flo_{\delta_{A_j}}(t)]T^i[\sum_j t_{A_j}^i(\alpha_{A_j}^i(t)Flo_{\delta_{A_j}}(t))]\}}{A_j(t)} \quad (5.65)$$

$$Flo_{A_j}^{tax}(t, \alpha_{A_j}^i(t)) = \frac{Flo_{A_j}(t)\{1-cgt_{A_j}^i[\alpha_{A_j}^i(t)Flo_{A_j}(t)]CGT^i[\sum_j t_{A_j}^i(\alpha_{A_j}^i(t)Flo_{A_j}(t))]\}}{A_j(t)} \quad (5.66)$$

$$Cap_{\delta_{A_j}}^{tax}(t, \alpha_{A_j}^i(t)) = \frac{Cap_{\delta_{A_j}}(t)\{1-t_{A_j}^i[\alpha_{A_j}^i(t)Cap_{\delta_{A_j}}(t)]T^i[\sum_j t_{A_j}^i(\alpha_{A_j}^i(t)Cap_{\delta_{A_j}}(t))]\}}{A_j(t)} \quad (5.67)$$

$$Cap_{A_j}^{tax}(t, \alpha_{A_j}^i(t)) = \frac{Cap_{A_j}(t)\{1-cgt_{A_j}^i[\alpha_{A_j}^i(t)Cap_{A_j}(t)]CGT^i[\sum_j t_{A_j}^i(\alpha_{A_j}^i(t)Cap_{A_j}(t))]\}}{A_j(t)} \quad (5.68)$$

Note that in practice, since the higher tax rate from a higher return is only applied to the incremental amount, the inclusion of taxes will not change the ranking order of returns: the cap and floor on a pre-tax basis are still the cap and floor after deducting tax payments.

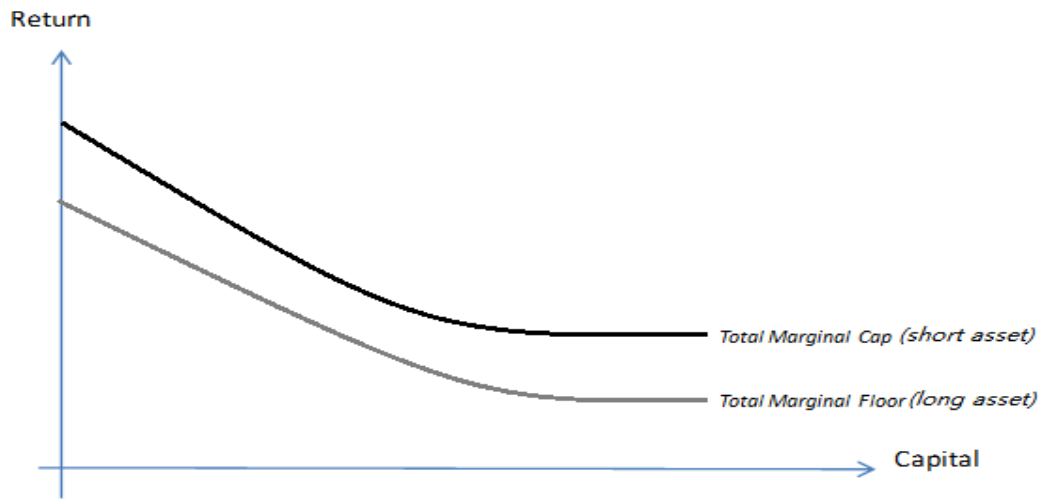
Expressions for the marginal after-tax floor and cap on total return (total marginal floor and cap) are obtained as follows:

$$Flo_{total}^{tax}(t, \alpha_{A_j}^i(t)) = Flo_{\delta_{A_j}}^{tax}(t, \alpha_{A_j}^i(t)) + Flo_{A_j}^{tax}(t, \alpha_{A_j}^i(t)) \quad (5.69)$$

$$Cap_{total}^{tax}(t, \alpha_{A_j}^i(t)) = Cap_{\delta_{A_j}}^{tax}(t, \alpha_{A_j}^i(t)) + Cap_{A_j}^{tax}(t, \alpha_{A_j}^i(t)) \quad (5.70)$$

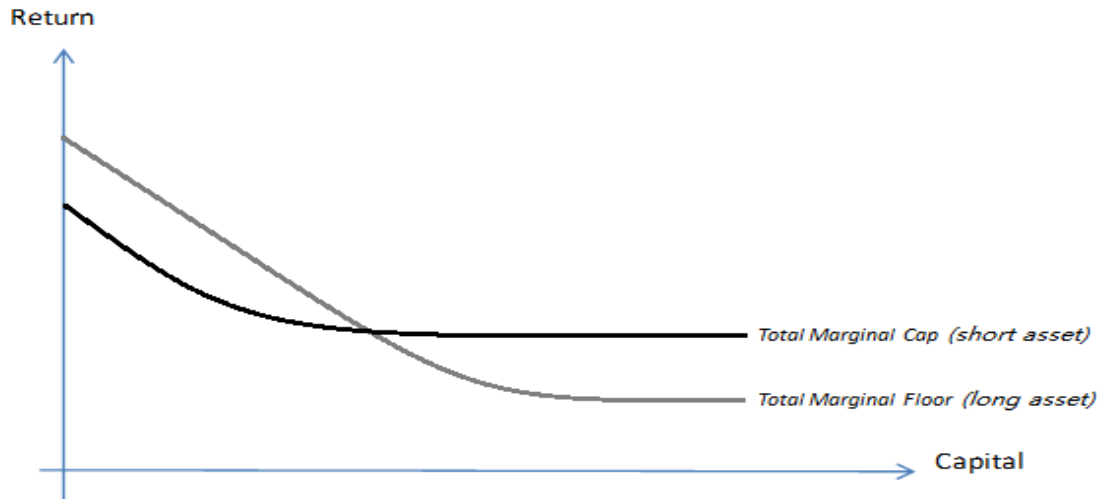
If long and short position tax treatments are the same, a tax arbitrage opportunity exists simply when a long-position asset's total marginal floor is above a short-position asset's total marginal cap. As shown in Fig.5.1, no tax arbitrage opportunity exists if the long-positioned asset's total marginal floor stands below the short-positioned asset's total marginal cap for any amount of capital.

Figure 5.1 Figure 5.1 No Arbitrage Opportunity



Furthermore, as shown in Fig.5.2, a local arbitrage opportunity exists if the total marginal floor stands above at the beginning and intersects with the total marginal cap later as capital increases. In this case, investors are able to obtain a finite riskless return without an outflow of capital at the beginning and are willing to enlarge arbitrage portfolio proportionately (keep the same composition) until the intersection point.

Figure 5. 2 Local Arbitrage Opportunity



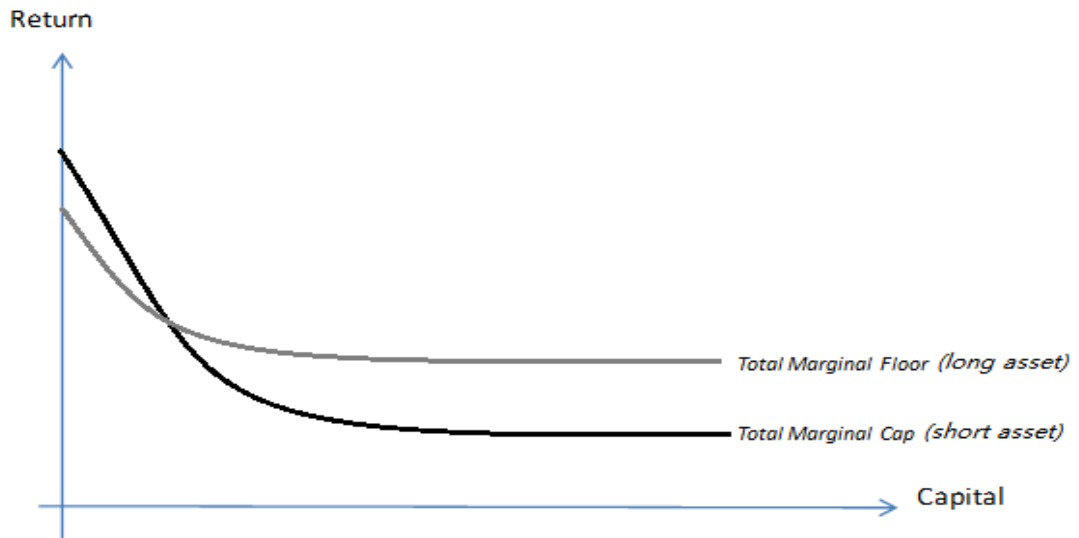
c. *Global Arbitrage*

i. *Restricted global arbitrage*

In restricted global arbitrage, an investor can guarantee a non-negative riskless return by holding a large amount of one asset and short-selling a large amount of the other. This portfolio can be enlarged but cannot be downsized to a small amount. This arbitrage opportunity exists if and only if there is a set of asset holdings, $\alpha_{A_j}^i(t)$, for investor i , that satisfy the system of inequalities given in Appendix B.1.

Similar to the analysis in the Section 5.4.1, as shown in Fig.5.3, a restricted global arbitrage opportunity exists if the total marginal floor intersects the total marginal cap and remains above it as capital increases. In this case, investors can achieve a risk-free return after intersection and are willing to enlarge the arbitrage portfolio to secure a riskless profit.

Figure 5.3 Restricted Global Arbitrage

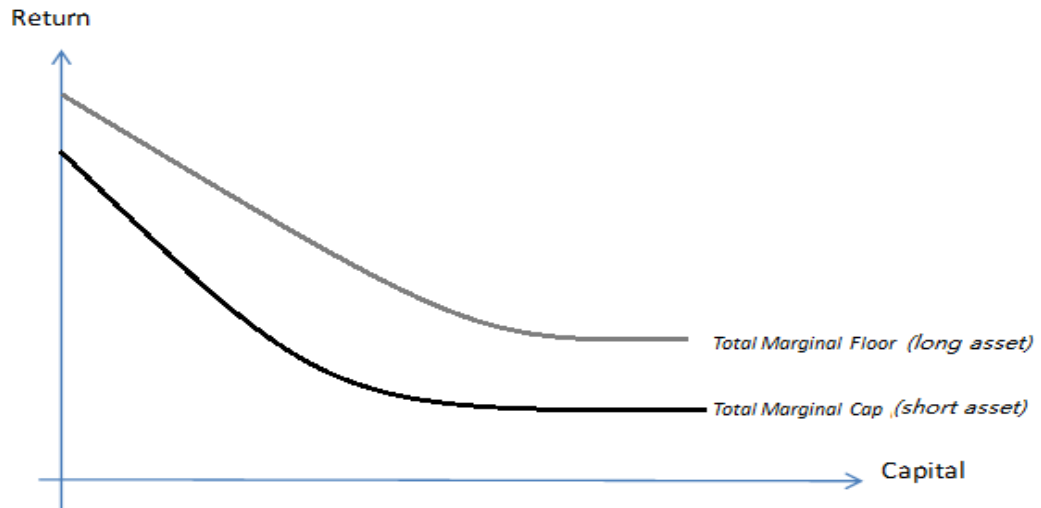


ii. *General global arbitrage*

In general global arbitrage, an investor expects to obtain a non-negative riskless return by holding one asset and short selling the other, and the portfolio can be either enlarged or downsized. In other words, this portfolio can be multiplied by a large (e.g. 1000) or small (e.g. 0.001) figure without eliminating the arbitrage opportunity. This arbitrage opportunity exists if and only if there is a set of asset holdings, $\alpha_{A_j}^i(t)$, for the investor i , that satisfy the system of inequalities given in Appendix B.2.

For general global arbitrage, as shown in Fig.5.4, the opportunity exists if the total marginal floor always stands above the total marginal cap for any amount of capital. In this case, investors can obtain a risk-free return from the beginning and are willing to enlarge the arbitrage portfolio to secure a large riskless profit.

Figure 5. 4 General Global Arbitrage



d. Implications

In summary, a tax arbitrage opportunity may exist between two non-perfectly correlated assets under a continuous time model if there are caps and floors on their incomes and capital gains. In real markets, investors usually get continuous time returns but not static returns from assets, and caps and floors can be set easily by purchasing a collar (short-selling a call option and using the proceeds to purchase a put option). According to Section 5.4.1, individual investors can quickly identify tax arbitrage opportunities by comparing the assets' total marginal caps and floors. If they can find two assets such that the long-positioned asset's total marginal floor stands above the short-positioned asset's total marginal cap, a tax arbitrage opportunity will exist between the two assets. If the floor remains above the cap only for a small amount of capital, the arbitrage is local. If, on the other hand, the floor stands above the cap only for a large amount of capital, it is restricted global arbitrage. If the floor is always above the cap for any amount of capital, there is a general global arbitrage opportunity.

5.4.2. Multiple Investors

Arbitrage opportunities between two investors with heterogeneous taxation but a single asset should be discussed too. It is found that an opportunity may exist between two investors if the low-taxed investor holds a long position while the high-taxed investor holds a short position. The constraints on tax policy are also determined from the government's viewpoint to avoid different types of arbitrage opportunities and identify the market equilibrium for which there is no tax arbitrage opportunity.

a. Local Arbitrage

In local arbitrage between multiple investors, a finite riskless return from a single asset can be obtained between two investors when one investor holds a long position while the other investor holds a short position. This arbitrage opportunity exists in the market if and only if for a certain asset $j \in \{1, 2, \dots\}$ there is a set of asset holdings between investors $i = 1, 2$, $\alpha_{A_j}^i(t)$, that satisfy the system of inequalities given in Appendix B.3.

Similarly, defining the marginal after-tax cap and floor for each investor as

$Flo_{\delta_{A_j}}^{tax}(t, \alpha_{A_j}^i(t))$, $Flo_{A_j}^{tax}(t, \alpha_{A_j}^i(t))$, $Cap_{\delta_{A_j}}^{tax}(t, \alpha_{A_j}^i(t))$ and $Cap_{A_j}^{tax}(t, \alpha_{A_j}^i(t))$, we can obtain

their precise values as follows: for investor i , $i = 1, 2$

$$Flo_{\delta_{A_j}}^{tax}(t, \alpha_{A_j}^i(t)) = Flo_{\delta_{A_j}}(t) \{1 - t_{A_j}^{i'} [\alpha_{A_j}^i(t) Flo_{\delta_{A_j}}(t)] T^{i'} [t_{A_j}^i (\alpha_{A_j}^i(t) Flo_{\delta_{A_j}}(t))]\} / A_j(t) \quad (5.71)$$

$$Flo_{A_j}^{tax}(t, \alpha_{A_j}^i(t)) = Flo_{A_j}(t) \{1 - cgt_{A_j}^{i'} [\alpha_{A_j}^i(t) Flo_{A_j}(t)] CGT^{i'} [t_{A_j}^i (\alpha_{A_j}^i(t) Flo_{A_j}(t))]\} / A_j(t) \quad (5.72)$$

$$Cap_{\delta_{A_j}}^{tax}(t, \alpha_{A_j}^i(t)) = Cap_{\delta_{A_j}}(t) \{1 - t_{A_j}^{i'} [\alpha_{A_j}^i(t) Cap_{\delta_{A_j}}(t)] T^{i'} [t_{A_j}^i (\alpha_{A_j}^i(t) Cap_{\delta_{A_j}}(t))]\} / A_j(t) \quad (5.73)$$

$$Cap_{A_j}^{tax}(t, \alpha_{A_j}^i(t)) =$$

$$Cap_{A_j}(t) \{ 1 - cgt_{A_j}^i [\alpha_{A_j}^i(t) Cap_{A_j}(t)] CGT^i [t_{A_j}^i(\alpha_{A_j}^i(t) Cap_{A_j}(t))] \} / A_j(t) \quad (5.74)$$

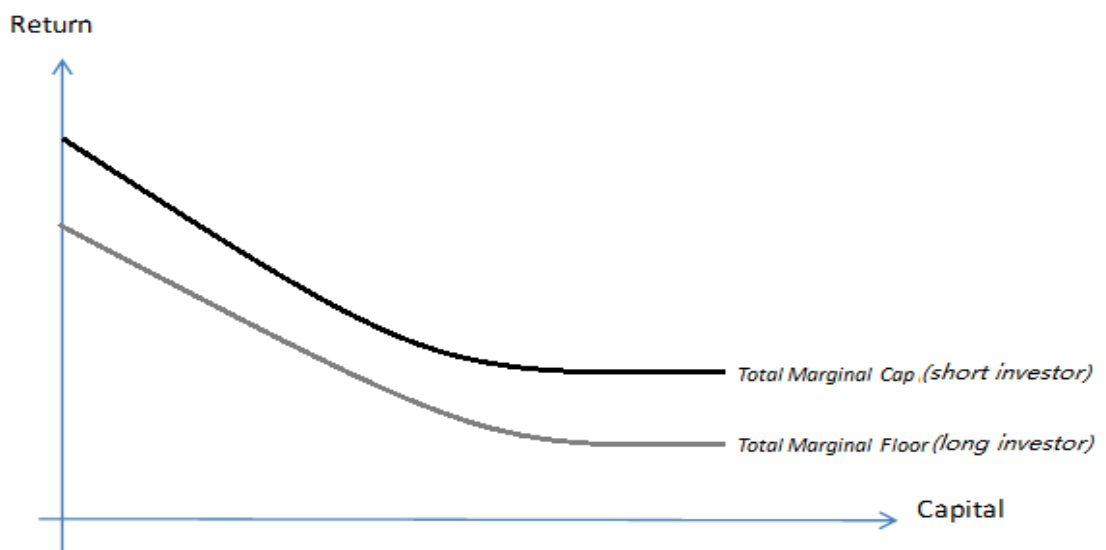
The total marginal floor and cap is then obtained as follows:

$$Flo_{total}^{tax}(t, \alpha_{A_j}^i(t)) = Flo_{\delta_{A_j}}^{tax}(t, \alpha_{A_j}^i(t)) + Flo_{A_j}^{tax}(t, \alpha_{A_j}^i(t)) \quad (5.75)$$

$$Cap_{total}^{tax}(t, \alpha_{A_j}^i(t)) = Cap_{\delta_{A_j}}^{tax}(t, \alpha_{A_j}^i(t)) + Cap_{A_j}^{tax}(t, \alpha_{A_j}^i(t)) \quad (5.76)$$

A tax arbitrage opportunity exists when the long-position investor's total marginal floor is above the short-position investor's total marginal cap. As shown in Fig.5.5, no tax arbitrage opportunity exists if the long-position investor's total marginal floor stands below the short-position investor's total marginal cap for any amount of capital.

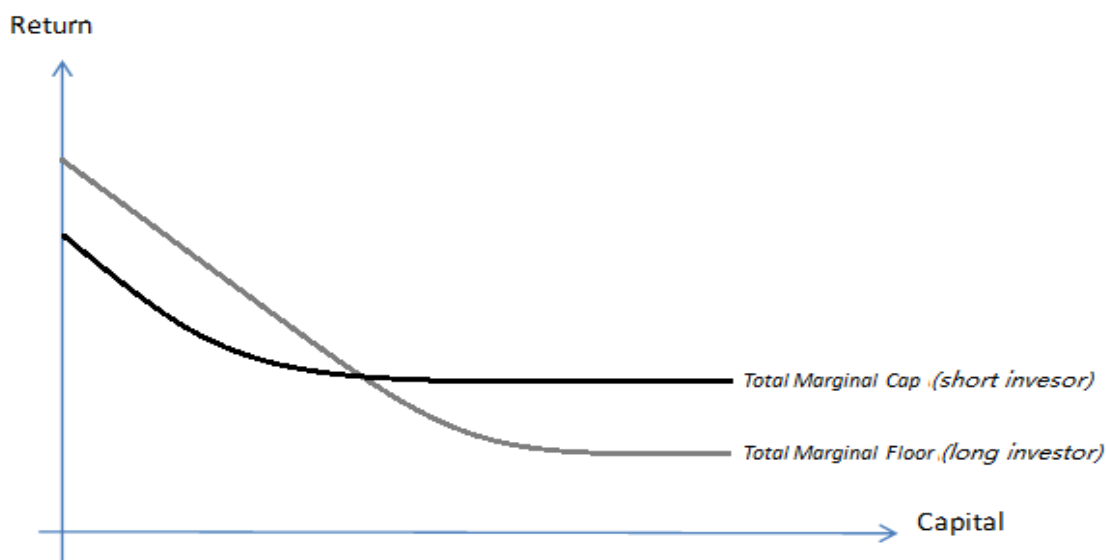
Figure 5. 5 No Arbitrage by Cooperation



Furthermore, as shown in Fig.5.6, a local arbitrage opportunity exists if the total marginal floor stands above the total marginal cap at the beginning and intersects it later as capital increases. In this case, a positive risk-free return can be obtained and the

arbitrage portfolio between two investors can be enlarged proportionately until the intersection point⁵⁸.

Figure 5. 6 Local Arbitrage by Cooperation



b. *Global Arbitrage*

i. *Restricted global arbitrage*

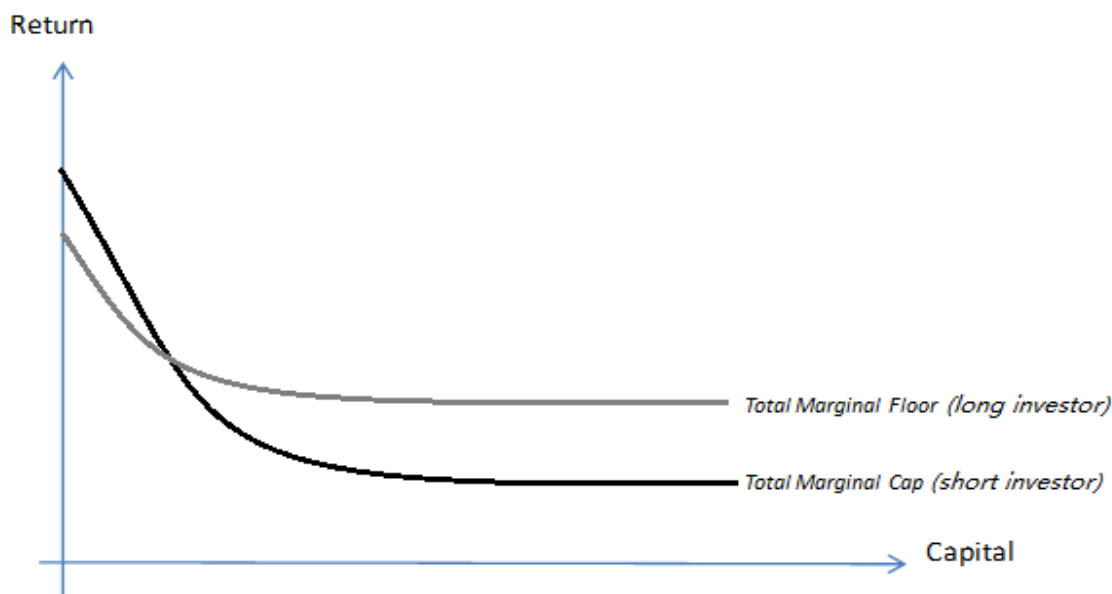
In restricted global arbitrage, a riskless return can be obtained on a single asset between two investors, but the portfolio between two investors can only be enlarged but not downsized. This arbitrage opportunity exists in the market if and only if for a certain asset $j \in \{1, 2, \dots\}$, there is a set of asset holdings between investors $i = 1, 2$, $\alpha_{A_j}^i(t)$, that satisfy the system of inequalities given in Appendix B.4.

As in Section 5.4.2, and as shown in Fig. 5.7, a restricted global arbitrage opportunity exists if the long-position investor's total marginal floor intersects the short-

⁵⁸ On a before tax basis, it is realistic to find two assets such that one's floor of return is larger than the other's cap of return. However, on an after tax basis, when an asset is tax free (e.g. Treasury bills or some other tax-exempt product) while the other is subject to the top tax rate (the investor must pay top tax rate on taxable profit), then it is possible that the tax-exempt asset's floor is larger than the other asset's cap.

position investor's total marginal cap and remains above it as capital increases. In this case, a risk-free return can be obtained after the intersection point and the arbitrage portfolio between two investors can be enlarged to secure a large risk-free return.

Figure 5. 7 Restricted Global Arbitrage by Cooperation



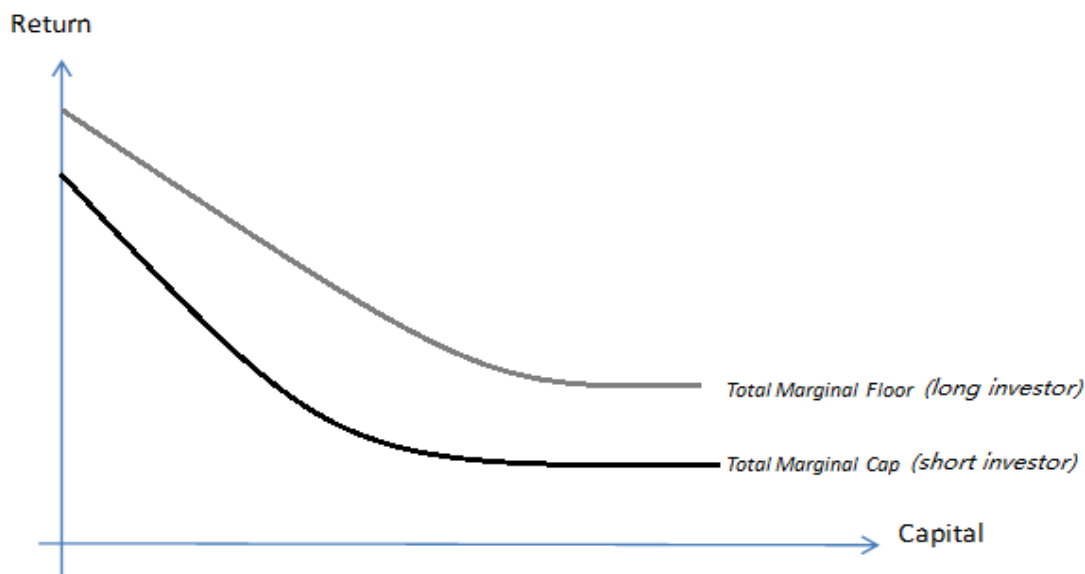
ii. General global arbitrage

In general global arbitrage, a riskless return on a single asset can be obtained between two investors, and this portfolio between two investors can be either enlarged or downsized. If so, this portfolio can be multiplied by a large or small figure without eliminating the arbitrage opportunity. This arbitrage opportunity in the market exists if and only if for a certain asset $j \in \{1, 2, \dots\}$ there is a set of asset holdings between investors $i = 1, 2$, $\alpha_{A_j}^i(t)$, satisfying the system of inequalities given in Appendix B.5.

For general global arbitrage, as shown in Fig. 5.8, the opportunity exists if the long-position investor's total marginal floor always stands above the short-position investor's total marginal cap for any capital. In this case, a risk-free return can be

obtained from the beginning and the arbitrage portfolio between two investors can be enlarged to secure a large risk-free return.

Figure 5. 8 General Global Arbitrage by Cooperation



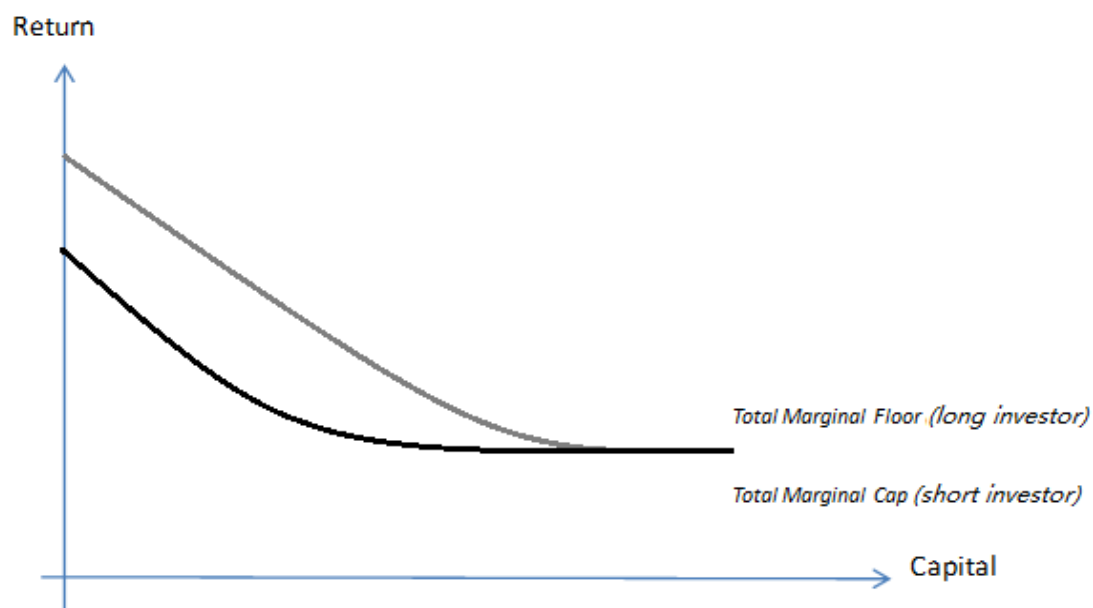
c. Implications

In summary, from the investor's viewpoint, in addition to the arbitrage opportunity between two assets, tax arbitrage may also exist on a single asset between two investors if they are subject to different tax codes, and the long-positioned investor's marginal floor stands above the short-positioned investor's marginal cap. As in Section 5.4.1, all arbitrage opportunities can be divided into three classes: local, restricted global and general global arbitrage, and two types, A and B.

From the government's viewpoint, for fixed-income assets for which the cap is always equal to the floor on a pre-tax basis, the inclusion of heterogeneous taxation will make the long-positioned investor's total marginal floor stand above the short-positioned investor's total marginal cap, at least during the initial period as shown in Fig.5.9. The local arbitrage cannot be eliminated completely unless all investors are

subject to the same tax code. The government, however, can remove the possibility of a large risk-free return between any two of investors by applying the same top tax rate to all investors. This will make the marginal total floor and cap equal after a certain amount of capital, as shown in Fig. 5.9.

Figure 5.9 No Global Arbitrage by Cooperation



5.5. Conclusion

Two improved continuous-time models are proposed for after-tax portfolio optimization. One applies to assets with perfect correlation and the other applies to non-perfectly correlated assets with caps and floors. Asset returns are assumed to be continuous and the Brownian motion process is used to simulate market prices. Both capital gains tax and income tax are included, and differential taxation for investors with long and short positions is considered.

For perfectly correlated assets, the work of Basak and Croitoro (2001) is extended by including tax on capital gains. In contrast to their analysis, which mainly concerns

market equilibrium when no global tax arbitrage opportunity exists, mathematical constraints on tax rates and asset parameters (i.e. asset price, variance and expected return) are determined to prove the existence of global tax arbitrage opportunities with no restriction on capital gains and income. An arbitrage portfolio between perfectly correlated assets exists when all constraints are satisfied. Since income must be non-negative while capital gains could be either positive or negative, it is more difficult to prove the existence of arbitrage opportunities with both capital gains tax and income tax than with income tax only. This improvement increases the level of complexity but is necessary for an analysis of tax arbitrage opportunities. It is concluded that only a global but no local tax arbitrage opportunity may exist. It is also concluded that since asset prices (referred to as ask prices when purchasing and bid prices when selling) which are dynamic over time are in the tax arbitrage opportunity constraints, so the arbitrage opportunities between perfectly correlated assets are dynamic as well and do not exist consistently. When no arbitrage opportunity exists, as in Basak and Croitoro (2001), investors will act to reduce aggregate market income tax payments, and market equilibrium will be achieved when the sum of income tax payments are minimized. However, according to my work, investors will act to reduce the aggregate of both market income and market capital gains tax payments, and market equilibrium will be achieved when the sum of both income tax and capital gains tax payments rather than income tax payment only are minimized. This proves that, with regard to market equilibrium on an after tax basis, if investors are subject to capital gains tax, equilibrium asset prices obtained under income tax only would not apply after incorporating capital gains tax.

For non-perfectly correlated assets with caps and floors, three new continuous-time optimization models are proposed to find conditions for the existence of local, global and restricted global arbitrage opportunities. These opportunities are further

divided into two categories, type A and type B, depending on whether a strict positive or only nonnegative future net after-tax return will be obtained with certainty without an outflow of funds at any time. On the other hand, given tax rates and asset parameters, a new function, which requires asset holdings as inputs, is proposed to calculate an asset's marginal cap and floor on its total net return. It is concluded that the existence of tax arbitrage opportunities between non-perfectly correlated assets simply relies on the difference between assets' marginal caps and floors. A single investor can expect to receive a risk-free return when the long-position asset's marginal floor stands above the short-position asset's marginal cap. In addition, an arbitrage opportunity exists between two investors if for the same asset, the long-positioned investor's marginal floor stands above the short-positioned investor's marginal cap. In the fixed-income market, It shows that a local (finite) tax arbitrage opportunity between investors is difficult to eliminate unless they are all subject to the same tax policy, but a global (large risk-free return) tax arbitrage opportunity can be avoided if the government applies the same top tax rate to all investors.

Chapter 6 – Conclusions

Mathematical programming is used to quantify the effect of tax on investments, capital flows and arbitrage. By using proposed models, the impact of taxes on private portfolio optimization, global market performance and asset pricing are investigated. The experimental results demonstrate the importance of tax in all three fields and provide some useful conclusions for investors and governments.

6.1 Personal Investment Tax in Portfolio Optimization

In Chapter 3, to investigate the impact of personal investment taxes on private portfolio optimization, a post-tax portfolio optimization model with integer-based trading constraints is developed. In order to examine the real influence of income tax on portfolio management, many real-world trading constraints are considered. These include the need for diversification, requirements on both the number of assets in a portfolio and the maximum holdings in single assets, round-lot buying, and taxation on cash withdrawals (the last two are modeled with integer variables). The proposed model also accounts for the risk in estimating expected asset returns through the introduction of a stochastic constraint, by which the expected return of the portfolio exceeds the threshold with a high confidence level.

One key contribution of this thesis is that it innovates on the basic Greedy algorithm, making it available for post-tax portfolio optimization problems in which stochastic risk and real-life market restrictions modelled with integer constraints are simultaneously considered. The combination of integer and nonlinear constraints reflects the complexity involved in solving such problems under large-scale applications for which very few solvers are efficient. The efficacy of the approach is evaluated on more than 50 problems containing up to 288 assets, and the computational results

provide evidence of its efficiency in two aspects: precision of solution and required computing time.

It is found that income tax has a clear impact on portfolio optimization for investors, which supports the conclusion from existing theoretical work. It also shows that with real trading constraints, tax rates and portfolio composition have a complex relationship that is neither linear nor convex. Convexity assumptions often made in the literature to guarantee global optimality, therefore, are not only unrealistic but also erroneous simplifications. This is the main advantage of my work over prior theoretical research on post-tax portfolio optimization. In addition, in the investigation on effects of withdrawal tax, it is found that for single-period optimization, this factor has a very limited influence. Investors can simplify the optimization model by ignoring withdrawal tax without changing the optimal solution significantly. Finally, the analysis proves that investors' preference for certain assets is significantly influenced by taxation. Governments should estimate its quantitative effects before carrying out a new tax policy to avoid an unexpected capital outflow or excessive demand in relevant financial products.

6.2 Tobin Tax in Global Financial Markets

In Chapter 4, to investigate the impact of Tobin tax capital flows between regional markets, a post-tax portfolio optimization model with non-linear trading constraints and objective function is developed. To undertake a better examination of the influence of heterogeneous withholding and Tobin tax on international financial markets, almost all the real-world trading constraints are considered in the model. These include the need for diversification, requirements for both the number of assets in a portfolio and the maximum holdings in single assets, annual tax, deferred capital gains tax and taxation on cash withdrawals. So, investor behaviour can be simulated better. This influence is

quantified by observing the rebalancing activities of rational investors under different tax settings.

Regarding the comparison between residence- and source-based taxes on international investments, it is found that the global optimal portfolio is highly sensitive to a change in regional investment tax rates. This sensitivity depends on the size of the change, market specifications and the international investment tax environment (residence only, source only or mixed tax). In a pure tax environment, the source only tax union will on average have more capital transits in international markets than the residence only tax union, and its optimal market portfolio will be more sensitive to regional tax policy changes. In a mixed tax system, double taxation between residence- and source-taxed markets will significantly reduce the attractiveness of the latter while the credit method will perform much better (increasing the attractiveness of the market with a source-based tax by up to 20%). In addition, experimental results suggest that volatile markets are usually accompanied by less unrealized capital gains and therefore are more sensitive to a government's tax policy than trending markets.

As regards the Tobin tax, market trading volume from rebalancing activities of rational investors (who seek to maximize the net Sharpe ratio) is highly sensitive to the implementation of Tobin tax. This sensitivity not only varies by market specifications but also varies by investment tax rules. A volatile market in a “Mixed with credit method” tax environment will be more sensitive to Tobin tax than a trending market in a “Mixed with double taxation” tax environment. Furthermore, experiments show that the capital locking effects of Tobin tax is mainly dependent on its effective rate but not the side of taxation (tax is levied on capital inflow only, outflow only, or both) if a consistent Tobin tax rule is applied globally. When the rule is heterogeneous across countries, for the market with relatively high Tobin tax rate, inflow Tobin tax will have a much higher capital lock-out effect, and outflow Tobin tax will have a much higher

capital lock-in effect, in comparison to a consistent Tobin tax system across countries. In other words, the capital locking effect of Tobin tax is enlarged significantly when heterogeneous Tobin tax rates are applied globally. As a result, it will be very helpful if all countries can reach a deal on the implementation of Tobin tax. Otherwise the relatively high Tobin tax will significantly reduce the attractiveness of the local market to overseas investors.

6.3 Tax Arbitrage Optimization

In Chapter 5, to investigate tax arbitrage opportunities and therefore post-tax asset pricing, two improved continuous-time models for after-tax portfolio optimization are proposed. One applies to assets with perfect correlation and the other applies to uncorrelated assets with caps and floors. It is assumed that asset returns are continuous over time and use the Brownian motion process to simulate market prices. Both capital gains tax and income tax are included, and differential taxation for investors with long and short positions is also considered.

For correlated assets, the work of Basak and Croitoro (2001) is extended by including tax on capital gains. In contrast to their analysis which mainly concerns market equilibrium when no global tax arbitrage opportunity exists, mathematical constraints on tax rates and asset parameters (i.e. asset price, variance and expected return) are determined in this thesis to prove the existence of global tax arbitrage opportunities with no restriction on capital gains and income. An arbitrage portfolio between correlated assets can be achieved when all constraints are satisfied. Since income must be positive while capital gains could be either positive or negative, it is more difficult to prove the existence of arbitrage opportunities with both capital gains tax and income tax than with income tax only. This improvement increases the level of complexity but is necessary for the analysis of tax arbitrage opportunities. It is

concluded that only a global but no local tax arbitrage opportunity may exist. It is also concluded that since asset prices (referred to as purchase prices) which are dynamic over time are restricted in the constraint, tax arbitrage opportunities between correlated assets are also dynamic and do not exist consistently. When no arbitrage opportunity exists, in contrast to conclusions of Basak and Croitoro, investors will act to reduce aggregate market tax payments and market equilibrium will be achieved when the sum of capital gains tax payments and income tax payments rather than income tax payments only are minimized. This finding has a significant implication to financial markets and relevant research. It also proves that in the discussion of market equilibrium on an after tax basis, if investors are subject to capital gains tax, many equilibriums obtained under income tax only would be sub-optimal.

For uncorrelated assets with caps and floors, three new continuous-time optimization models are developed to achieve local, global and restricted global arbitrage opportunities. These opportunities are further divided into two categories, type A and type B, depending on whether a strict positive or only nonnegative future net return on a tax basis will be returned for sure without an outflow of funds at any time. On the other hand, given tax rates and asset parameters, a new function, which takes asset holdings as an input, is proposed to calculate an asset's marginal cap and floor on its total net return. According to the analysis, it is found that the existence of tax arbitrage opportunities between uncorrelated assets simply relies on the difference between assets' marginal caps and floors. A single investor can expect to receive a risk-free return when the long-positioned asset's marginal floor stands above the short-positioned asset's marginal cap. In addition, an arbitrage opportunity exists between two investors if for the same asset, the long-positioned investor's marginal floor stands above the short-positioned investor's marginal cap. In the fixed-income market, a local tax arbitrage opportunity between investors is difficult to eliminate unless they are all

subject to the same tax policy but a global tax arbitrage opportunity can be avoided efficiently if the government applies the same top tax rate to all investors.

6.4 Summary

In this thesis, the impact of tax in financial markets is quantified from three aspects using mathematical programming.

First, as regards the impact of tax on private portfolio optimization, it is proved that under a single-period optimization model, differential tax rules across investment accounts (i.e. the way the investment return is taxed in offshore bond account, onshore bond account and unit trust account) have limited impact on the optimal portfolio composition. Differential tax rates across asset classes (i.e. equities, bonds and commodities), however, have a significant influence on the optimal portfolio composition. As a result, taxation constraints should be included in a portfolio optimization model. However, these constraints can be simplified only to reflect differential tax rates over asset classes if a single-period model is used.

Second, as regards the impact of tax on capital flows between regional markets, it is proved that not only local investment tax but also Tobin-style international transaction tax has a remarkable impact on global market balance. The experimental results show that different ways of charging Tobin tax lead to different capital locking effects. Tobin tax on capital inflows imposes a capital lock-out effect (i.e. stop capital flowing into local markets), while Tobin tax on capital outflows imposes a capital lock-in effect (i.e. stop capital flowing out of local markets). In addition, experiment results show that Tobin tax will lead to a volatile global market if its application is heterogeneous across countries. An agreement across countries on Tobin tax can reduce this effect.

Finally, as regards the impact of tax on asset pricing, it is proved mathematically that differential taxes across asset classes and investors may lead to a tax arbitrage opportunity when basic CAPM is used to price assets. The arbitrage opportunity can exist between both correlated assets and uncorrelated assets. In the analysis, for these two cases, different constraints are found to help investors to check the existence of arbitrage opportunities. In addition, the work also proves that local tax arbitrage opportunities are difficult to eliminate completely. However, global tax arbitrage opportunities can be avoided if the same top tax rate is applied across all asset classes and investors.

References

- Aliber, R.Z., Chowdhry, B. and Yan, S., 2003. Some evidence that a Tobin tax on foreign exchange transactions may increase volatility. *Review of Finance*, 7(3), pp.481-510.
- Basak, S. and Croitoru, B., 2001. Non-linear taxation, tax-arbitrage and equilibrium asset prices. *Journal of Mathematical Economics*, 35(2), pp.347-382.
- Basak, S. and Gallmeyer, M., 2003. Capital market equilibrium with differential taxation. *European Finance Review*, 7(2), pp.121-159.
- Bawa, V.S., Brown, S.J. and Klein, R.W., 1979. Estimation risk and optimal portfolio choice. North-Holland Publ.NORTH-HOLLAND PUBL. CO., N. Y., 190 pp.
- Becker, J. and Fuest, C., 2008. Tax competition—Greenfield investment versus mergers and acquisitions. *Regional Science and Urban Economics*, 41(5), pp.476-486.
- Becker, J. and Fuest, C., 2010. Taxing foreign profits with international mergers and acquisitions. *International Economic Review*, 51(1), pp.171-186.
- Bienstock, D., 1996. Computational study of a family of mixed-integer quadratic programming problems. *Mathematical programming*, 74(2), pp.121-140.
- Birge, J.R. and Yang, S., 2007. A model for tax advantages of portfolios with many assets. *Journal of Banking & Finance*, 31(11), pp.3269-3290.
- Bodnar, T. and Schmid, W., 2007. The distribution of the sample variance of the global minimum variance portfolio in elliptical models. *Statistics*, 41(1), pp.65-75.
- Brennan, M.J., 1970. Taxes, market valuation and corporate financial policy. *National Tax Journal*, 23(4), pp.417-427.

- Bonami, P. and Lejeune, M.A., 2009. An exact solution approach for portfolio optimization problems under stochastic and integer constraints. *Operations research*, 57(3), pp.650-670.
- Bonami, P., Biegler, L.T., Conn, A.R., Cornuéjols, G., Grossmann, I.E., Laird, C.D., Lee, J., Lodi, A., Margot, F., Sawaya, N. and Wächter, A., 2008. An algorithmic framework for convex mixed integer nonlinear programs. *Discrete Optimization*, 5(2), pp.186-204.
- Bovenberg, A. and Goulder, L.H., 1993. Promoting investment under international capital mobility: an intertemporal general equilibrium analysis. NBER working paper, (w3139).
- Bovenberg, A.L., 1989. The effects of capital income taxation on international competitiveness and trade flows. *The American Economic Review*, pp.1045-1064.
- Bovenberg, A.L., 1992. Residence-and source-based taxation of capital income in an overlapping generations model. *Journal of Economics*, 56(3), pp.267-295.
- Bradford, D.F., 1995. Consumption taxes: Some fundamental transition issues (No. w5290). National Bureau of Economic Research.
- Brandes, Y., Domowitz, I., & Serbin, V., 2012. Transaction Costs and Equity Portfolio Capacity Analysis. *Oxford Handbooks Online*. Retrieved 29 Apr. 2017
- Brashares, E., 1999. Replacing the federal income tax with a consumption tax. *Public Administration and Public Policy*, 72, pp.549-578.
- Brennan, M.J., 1970. Taxes, market valuation and corporate financial policy. *National tax journal*, 23(4), pp.417-427. Brennan, M.J., 1970(2). Taxes, market valuation and corporate financial policy. *National Tax Journal*, 23(4), pp.417-427.

- Brennan, M.J., 1970. Taxes, market valuation and corporate financial policy. *National Tax Journal*, 23(4), pp.417-427.
- Broadie, M., 1993. Computing efficient frontiers using estimated parameters. *Annals of Operations Research*, 45(1), pp.21-58.
- Campbell, J.Y. and Froot, K.A., 1994. International experiences with securities transaction taxes. In *The internationalization of equity markets* (pp. 277-308). University of Chicago Press.
- Campbell, J.Y., Lettau, M., Malkiel, B.G. and Xu, Y., 2001. Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. *The Journal of Finance*, 56(1), pp.1-43.
- Ceria, S. and Stubbs, R.A., 2006. Incorporating estimation errors into portfolio selection: Robust portfolio construction. *Journal of Asset Management*, 7(2), pp.109-127.
- Chang, T., Liu, W.R. and Caudill, S.B., 2002. Tax-and-spend, spend-and-tax, or fiscal synchronization: new evidence for ten countries. *Applied Economics*, 34(12), pp.1553-1561.
- Chopra, V.K. and Ziemba, W.T., 2011. The effect of errors in means, variances, and covariances on optimal portfolio choice. *The Kelly Capital Growth Investment Criterion: Theory and Practice*, 3, p.249.
- Chvatal, V., 1979. A greedy heuristic for the set-covering problem. *Mathematics of operations research*, 4(3), pp.233-235.
- Cipriani, M. and Guarino, A., 2008. Transaction costs and informational cascades in financial markets. *Journal of Economic Behavior & Organization*, 68(3), pp.581-592.

- Constantinides, G.M. and Scholes, M.S., 1980. Optimal liquidation of assets in the presence of personal taxes: Implications for asset pricing. *The Journal of Finance*, 35(2), pp.439-449.
- Constantinides, G.M., 1983. Capital market equilibrium with personal tax. *Econometrica: Journal of the Econometric Society*, pp.611-636.
- Constantinides, G.M., 1984. Optimal stock trading with personal taxes: Implications for prices and the abnormal January returns. *Journal of Financial Economics*, 13(1), pp.65-89.
- Corazza, M. and Favaretto, D., 2007. On the existence of solutions to the quadratic mixed-integer mean–variance portfolio selection problem. *European Journal of Operational Research*, 176(3), pp.1947-1960.
- Creedy, J. and Gemmell, N., 2006. *Modelling tax revenue growth*. Edward Elgar Publishing.
- Creedy, J. and Sanz-Sanz, J.F., 2011. Modelling aggregate personal income tax revenue in multi-schedular and multi-regional structures. *Economic Modelling*, 28(6), pp.2589-2595.
- Dammon, R.M. and Green, R.C., 1987. Tax arbitrage and the existence of equilibrium prices for financial assets. *The Journal of Finance*, 42(5), pp.1143-1166.
- Dammon, R.M. and Spatt, C.S., 1996. The optimal trading and pricing of securities with asymmetric capital gains taxes and transaction costs. *Review of Financial Studies*, 9(3), pp.921-952.
- Dammon, R.M., Spatt, C.S. and Zhang, H.H., 2001. Optimal consumption and investment with capital gains taxes. *Review of Financial Studies*, 14(3), pp.583-616.

- Dammon, R.M., Spatt, C.S. and Zhang, H.H., 2004. Optimal asset location and allocation with taxable and tax - deferred investing. *The Journal of Finance*, 59(3), pp.999-1037.
- DeMiguel, V. and Uppal, R., 2005. Portfolio investment with the exact tax basis via nonlinear programming. *Management Science*, 51(2), pp.277-290.
- Desai, M.A. and Hines Jr, J.R., 2004. Old rules and new realities: Corporate tax policy in a global setting. *National Tax Journal*, pp.937-960.
- Devereux, M.P. and Griffith, R., 2003. Evaluating tax policy for location decisions. *International Tax and Public Finance*, 10(2), pp.107-126.
- Devereux, M.P., Lockwood, B. and Redoano, M., 2008. Do countries compete over corporate tax rates?. *Journal of Public Economics*, 92(5), pp.1210-1235.
- Dybvig, P. and Koo, H.K., 1996. Investment with taxes. Washington University.
- Dybvig, P.H. and Ross, S.A., 1986. Tax clienteles and asset pricing. *The Journal of Finance*, 41(3), pp.751-762.
- Edwards, F.R., 1992. Taxing transactions in futures markets: objectives and effects. *Journal of Financial Services Research*, 7(1), pp.75-91.
- Ehrenstein, G., 2002. Cont-Bouchaud percolation model including Tobin tax. *International Journal of Modern Physics C*, 13(10), pp.1323-1331.
- Ehrenstein, G., Westerhoff, F. and Stauffer, D., 2005. Tobin tax and market depth. *Quantitative Finance*, 5(2), pp.213-218.
- Elton, E.J. and Gruber, M.J., 1978. Taxes and portfolio composition. *Journal of Financial Economics*, 6(4), pp.399-410.

- Engel, C.M. and Kletzer, K.M., 'Saving and Investment in an Open Economy with Non-Traded Goods', *International Economic Review*, 4 (1989), 735–52.
- Evans, J.L. and Archer, S.H., 1968. Diversification and the reduction of dispersion: an empirical analysis. *The Journal of Finance*, 23(5), pp.761-767.
- Feldstein, M., 1976. Personal taxation and portfolio composition: An econometric analysis. *Econometrica: Journal of the Econometric Society*, pp.631-650.
- Fischer, M & Gallmeyer, M 2016, 'Heuristic Portfolio Trading Rules with Capital Gain Taxes' *Journal of Financial Economics*, vol 119, no. 3, pp. 611–625. DOI: 10.1016/j.jfineco.2016.01.024
- Frankel, J.A., 1996. How well do foreign exchange markets function: might a Tobin Tax help? (No. w5422). National Bureau of Economic Research.
- Fuest, C. and Huber, B., 2004. Why do countries combine the exemption system for the taxation of foreign profits with domestic double taxation relief?. *Journal of International Economics*, 62(1), pp.219-231.
- Fuest, C., Huber, B. and Mintz, J., 2005. Capital mobility and tax competition. Now Publishers Inc.
- Goldfarb, D. and Iyengar, G., 2003. Robust portfolio selection problems. *Mathematics of Operations Research*, 28(1), pp.1-38.
- Gondzio, J. and Grothey, A., 2007. Solving non-linear portfolio optimization problems with the primal-dual interior point method. *European Journal of Operational Research*, 181(3), pp.1019-1029.
- Gordon, R., Kalambokidis, L., Rohaly, J. and Slemrod, J., 2004. Toward a consumption tax, and beyond. *The American Economic Review*, 94(2), pp.161-165.

- Goulder, L.H. and Eichengreen, B., 1989. Savings promotion, investment promotion, and international competitiveness. In Trade policies for international competitiveness (pp. 5-52). University of Chicago Press.
- Haberer, M., 2006. Regulierung internationaler Finanzmärkte durch Transaktionssteuern: Die Wirkung einer Tobin-Steuer auf Handelsvolumen und Wechselkursvolatilität (Doctoral dissertation).
- Habermeier, K. and Kirilenko, A.A., 2003. Securities transaction taxes and financial markets. *IMF Economic Review*, 50(1), pp.165-180.
- Hanke, M., Huber, J., Kirchler, M. and Sutter, M., 2010. The economic consequences of a Tobin tax—an experimental analysis. *Journal of Economic Behavior & Organization*, 74(1), pp.58-71.
- Haq, M., Kaul, I. and Grunberg, I. eds., 1996. The Tobin tax: coping with financial volatility. Oxford University Press.
- Huang, J., 2008. Taxable and tax-deferred investing: a tax-arbitrage approach. *Review of Financial Studies*, 21(5), pp.2173-2207.
- Hubbard, R.G., 1985. Personal taxation, pension wealth, and portfolio composition. *The Review of Economics and Statistics*, pp.53-60.
- Hubbard, R.G., 1993. Securities transactions taxes: Tax design, revenue, and policy considerations. *Tax Notes*, pp.985-1000.
- Internal Revenue Service. <http://www.irs.gov/>.
- Jobst, N.J., Horniman, M.D., Lucas, C.A. and Mitra, G., 2001. Computational aspects of alternative portfolio selection models in the presence of discrete asset choice constraints. *Quantitative Finance*, 1(5), pp.489-501.

- Jones, C. and Milne, F., 1992. Tax arbitrage, existence of equilibrium, and bounded tax rebates. *Mathematical Finance*, 2(3), pp.189-196.
- Joulfaian, D. and Mookerjee, R., 1991. Dynamics of government revenues and expenditures in industrial economies. *Applied Economics*, 23(12), pp.1839-1844.
- Keuschnigg, C., 1991. The transition to a cash flow income tax. *Swiss Journal of Economics and Statistics*, 127(2), pp.113-140.
- Kolm, P.N., Tütüncü, R. and Fabozzi, F.J., 2014. 60 Years of portfolio optimization: Practical challenges and current trends. *European Journal of Operational Research*, 234(2), pp.356-371.
- Konno, H. and Yamamoto, R., 2005. Integer programming approaches in mean-risk models. *Computational Management Science*, 2(4), pp.339-351.
- KPMG. <http://www.kpmg.com/>.
- Kupiec, P.H., 1995. A securities transactions tax and capital market efficiency. *Contemporary Economic Policy*, 13(1), p.101.
- Lejeune, M.A. and Samatli-Pağ, G., 2012. Construction of risk-averse enhanced index funds. *INFORMS Journal on Computing*, 25(4), pp.701-719.
- Lejeune, M.A., 2010. A VaR Black–Litterman model for the construction of absolute return fund-of-funds. *Quantitative Finance*, 11(10), pp.1489-1501.
- Linderoth, J.T. and Savelsbergh, M.W., 1999. A computational study of search strategies for mixed integer programming. *INFORMS Journal on Computing*, 11(2), pp.173-187.

- Litzenberger, R.H. and Ramaswamy, K., 1980. Dividends, short selling restrictions, tax - induced investor clienteles and market equilibrium. *The Journal of Finance*, 35(2), pp.469-482.
- Mannaro, K., Marchesi, M. and Setzu, A., 2008. Using an artificial financial market for assessing the impact of Tobin-like transaction taxes. *Journal of Economic Behavior & Organization*, 67(2), pp.445-462.
- Markowitz, H., 1952. Portfolio selection. *The journal of finance*, 7(1), pp.77-91.
- Mulvey, J. and Erkan, G., 2003. Decentralized risk management for global P/C insurance companies. *Applications of Stochastic Programming MPS-SIAM Series on Optimization*, Philadelphia, pp.503-530.
- Murphy, R.G., 1986. Productivity shocks, non-traded goods and optimal capital accumulation. *European Economic Review*, 30(5), pp.1081-1095.
- Mutti, J. and Grubert, H., 1985. The taxation of capital income in an open economy: the importance of resident-nonresident tax treatment. *Journal of Public Economics*, 27(3), pp.291-309.
- Osorio, M.A., Gulpınar, N. and Rustem, B., 2008a. A mixed integer programming model for multistage mean–variance post-tax optimization. *European Journal of Operational Research*, 185(2), pp.451-480.
- Osorio, M.A., Gulpınar, N. and Rustem, B., 2008b. A general framework for multistage mean-variance post-tax optimization. *Annals of Operations Research*, 157(1), pp.3-23.
- Osorio, M.A., Gulpınar, N., Rustem, B. and Settergren, R., 2004a. Post-tax optimization with stochastic programming. *European Journal of Operational Research*, 157(1), pp.152-168.

- Osorio, M.A., Gülpınar, N., Rustem, B. and Settergren, R., 2004b. Tax impact on multi - stage mean - variance portfolio allocation. *International Transactions in Operational Research*, 11(5), pp.535-554.
- Osorio, M.A., Settergren, R., Rustem, B. and Gulpınar, N., 2002. Post tax optimal investments. In *Financial Engineering, E-commerce and Supply Chain* (pp. 153-174). Springer US.
- Pearson, K., 1895. Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, 58, pp.240-242.
- Reinhart, V., 1991. The 'Tobin tax,' asset accumulation, and the real exchange rate. *Journal of International Money and Finance*, 10(3), pp.420-431.
- Ross, S.A., 1987. Arbitrage and martingales with taxation. *Journal of Political Economy*, 95(2), pp.371-393.
- Rousslang, D.J., 2002. Should financial services be taxed under a consumption tax? Probably. *National Tax Journal*, pp.281-291.
- Roy, A.D., 1952. Safety first and the holding of assets, *Econometrica* 20: 431–449.
- Samuelson, P., 1964. Tax deductibility of economic depreciation to insure invariant valuation. *Journal of Political Economy* 72, 604-606.
- Saunoris, J.W. and Payne, J.E., 2010. Tax more or spend less? Asymmetries in the UK revenue–expenditure nexus. *Journal of Policy Modeling*, 32(4), pp.478-487.
- Sialm, C., 2009. Tax changes and asset pricing. *The American Economic Review*, 99(4), pp.1356-1383.
- Sinn, H.W., 1985. Why taxes matter: Reagan's accelerated cost recovery system and the US trade deficit. *Economic Policy*, 1(1), pp.239-247.

- Stein, D. M. and Garland, J. P. 2008. Investment Management for Taxable Investors. Handbook of Finance. II:1:13.
- Strobel, F., 2001. International tax arbitrage, tax evasion and interest parity conditions. Research in Economics, 55(4), pp.413-427.
- Strobel, F., 2005. International tax arbitrage, financial parity conditions and preferential capital gains taxation. Quantitative Finance, 5(2), pp.219-226.
- Strobel, F., 2012(a). International tax arbitrage, currency options and put-call parity conditions. Journal of International Financial Markets, Institutions and Money, 22(3), pp.473-486.
- Strobel, F., 2012(b). International tax arbitrage and residence vs. source-based capital income taxation. Research in Economics, 66(4), pp.391-397.
- Summers, L.H., 1988. Tax policy and international competitiveness. In International aspects of fiscal policies (pp. 349-386). University of Chicago Press.
- Talmor, E., 1989. Tax arbitrage restrictions and financial leverage clienteles. Journal of Banking & Finance, 13(6), pp.831-838.
- Tobin, J., 1978. A proposal for international monetary reform. Eastern economic journal, 4(3/4), pp.153-159.
- Tornell, A., 1990. Real vs. financial investment can Tobin taxes eliminate the irreversibility distortion?. Journal of Development Economics, 32(2), pp.419-444.
- Trauring, M., 1979. A capital asset pricing model with investors taxes and three categories of investment income. Journal of Financial and Quantitative Analysis, 14(03), pp.537-545.

- Walker, C.E. and Bloomfield, M.A. eds., 1987. *The Consumption tax: a better alternative?*. Ballinger Pub. Co.
- Weaver, J., Dodd, R. and Baker, J., 2003. *Debating the Tobin tax: New rules for global Finance coalition*. Washington DC.
- Westerhoff, F., 2003. Heterogeneous traders and the Tobin tax. *Journal of Evolutionary Economics*, 13(1), pp.53-70.
- Zuckerman, D., 1989. Capital gains taxation and the lock-in effect: A probabilistic framework. *Operations Research Letters*, 8(2), pp.101-106.

Appendix A. Propositions and Proofs

As in the work of Basak and Croitorn (2001), investor i will divide his/her composite risk exposure between S and P in such a way that he/she is either indifferent to marginal shifts from one security to the other, or any shift yields negative gain. Proposition A.1 presents the resulting condition on rational holdings.

A.1. Proposition 3⁵⁹

Let $\Phi^i(t)$ and $\Delta_{S,P}(t)$ be given. If taxation is continuously differentiable, investor i is indifferent between all pairs,

$$(\hat{\alpha}_S^i, \hat{\alpha}_P^i) \equiv (\hat{\alpha}_S^i(\Phi^i(t), \Delta_{S,P}(t), t), \hat{\alpha}_P^i(\Phi^i(t), \Delta_{S,P}(t), t)) \quad (\text{A.1})$$

leading to the same value for his risk budget $\Phi^i(t)$ such that

$$\theta_S^i(\hat{\alpha}_S^i(t), \hat{\alpha}_P^i(t), t) = \theta_P^i(\hat{\alpha}_S^i(t), \hat{\alpha}_P^i(t), t) \quad (\text{A.2})$$

A.2. Proposition 4⁶⁰

Assume that market equilibrium exists; that there is a net supply N of asset S ; and that investors' composite risk budget sharing are given by $\Phi^1(t)$ and $\Phi^2(t)$. The pre-tax mispricing is then given by

$$\Delta_{S,P}(t) = \left\{ \frac{\delta_S(t) t_S^{1*} (\hat{\alpha}_S^1(t) \delta_S(t))}{S(t) \sigma_S(t)} - \frac{\delta_P(t) t_P^1 [(\Phi^1(t) - \hat{\alpha}_S^1(t) S(t)) \sigma_S(t) \delta_P(t) / P(t) \sigma_P(t)]}{P(t) \sigma_P(t)} \right\} \\ \times T^{1*} [t_S^1 (\hat{\alpha}_S^1(t) \delta_S(t)) + t_P^1 ((\Phi^1(t) - \hat{\alpha}_S^1(t) S(t)) \sigma_S(t) \delta_P(t) / P(t) \sigma_P(t))]$$

⁵⁹The proof of Proposition 3 follows from Proposition 3.1 in Basak and Croitorn (2001), with only obvious changes in notation. Q.E.D.

⁶⁰A proof of Proposition 4 follows from Proposition 4.1 in Basak and Croitorn (2001), with only obvious changes in notation. Q.E.D.

$$\begin{aligned}
& + \left\{ \frac{S^*(t) cgt_s^{1*} (\hat{\alpha}_s^1(t) S^*(t))}{S(t) \sigma_s(t)} - \frac{P^*(t) cgt_p^{1*} [(\Phi^1(t) - \hat{\alpha}_s^1(t) S(t)) \sigma_s(t) P^*(t) / P(t) \sigma_p(t)]}{P(t) \sigma_p(t)} \right\} \\
& \times CGT^{1*} [cgt_s^1 (\hat{\alpha}_s^1(t) S^*(t)) + cgt_p^1 ((\Phi^1(t) - \hat{\alpha}_s^1(t) S(t)) \sigma_s(t) P^*(t) / P(t) \sigma_p(t))] \quad (A.3)
\end{aligned}$$

where $\hat{\alpha}_s^1(t) = \hat{\alpha}_s^1(\Phi^1(t), \Phi^2(t), t)$ satisfies:

$$\begin{aligned}
& \left\{ \frac{\delta_s(t) t_s^{1*} (\hat{\alpha}_s^1(t) \delta_s(t))}{S(t) \sigma_s(t)} - \frac{\delta_p(t) t_p^{1*} [(\Phi^1(t) - \hat{\alpha}_s^1(t) S(t)) \sigma_s(t) \delta_p(t) / P(t) \sigma_p(t)]}{P(t) \sigma_p(t)} \right\} \\
& \times T^{1*} [t_s^1 (\hat{\alpha}_s^1(t) \delta_s(t)) + t_p^1 ((\Phi^1(t) - \hat{\alpha}_s^1(t) S(t)) \sigma_s(t) \delta_p(t) / P(t) \sigma_p(t))] \\
& + \left\{ \frac{S^*(t) cgt_s^{1*} (\hat{\alpha}_s^1(t) S^*(t))}{S(t) \sigma_s(t)} - \frac{P^*(t) cgt_p^{1*} [(\Phi^1(t) - \hat{\alpha}_s^1(t) S(t)) \sigma_s(t) P^*(t) / P(t) \sigma_p(t)]}{P(t) \sigma_p(t)} \right\} \\
& \times CGT^{1*} [cgt_s^1 (\hat{\alpha}_s^1(t) S^*(t)) + cgt_p^1 ((\Phi^1(t) - \hat{\alpha}_s^1(t) S(t)) \sigma_s(t) P^*(t) / P(t) \sigma_p(t))] \\
& = \left\{ \frac{\delta_s(t) t_s^{2*} ((N - \hat{\alpha}_s^1(t)) \delta_s(t))}{S(t) \sigma_s(t)} - \frac{\delta_p(t) t_p^{2*} [(\Phi^2(t) - (N - \hat{\alpha}_s^1(t))) \sigma_s(t) \delta_p(t) / P(t) \sigma_p(t)]}{P(t) \sigma_p(t)} \right\} \\
& \times T^{2*} [t_s^2 ((N - \hat{\alpha}_s^1(t)) \delta_s(t)) + t_p^2 ((\Phi^2(t) - (N - \hat{\alpha}_s^1(t))) \sigma_s(t) \delta_p(t) / P(t) \sigma_p(t))] \\
& + \left\{ \frac{S^*(t) cgt_s^{2*} ((N - \hat{\alpha}_s^1(t)) S^*(t))}{S(t) \sigma_s(t)} \right. \\
& \left. - \frac{P^*(t) cgt_p^{2*} [(\Phi^2(t) - (N - \hat{\alpha}_s^1(t))) \sigma_s(t) P^*(t) / P(t) \sigma_p(t)]}{P(t) \sigma_p(t)} \right\} \times CGT^{2*} [\\
& cgt_s^2 ((N - \hat{\alpha}_s^1(t)) S^*(t)) + cgt_p^2 ((\Phi^2(t) - (N - \hat{\alpha}_s^1(t))) \sigma_s(t) P^*(t) / P(t) \sigma_p(t))] \quad (A.4)
\end{aligned}$$

A.3. Proposition 5 and Proof

In market equilibrium, the mispricing adjusts so that, among all pairs of portfolio holdings (α_s^i, α_p^i) , $i=1,2$, that satisfy investors' risk budget $\Phi^i(t)$ and clear financial markets, investors choose the one that minimizes total tax on both income and capital gains $T^1(t) + T^2(t) + CGT^1(t) + CGT^2(t)$.

Proof: In equilibrium, total tax is given by

$$\begin{aligned}
& T^1[t_s^1(\hat{\alpha}_s^1(t)\delta_s(t)) + t_p^1((\Phi^1(t) - \hat{\alpha}_s^1(t)S(t))\sigma_s(t)\delta_p(t) / P(t)\sigma_p(t))] \\
& + CGT^1[cgt_s^1(\hat{\alpha}_s^1(t)S^*(t)) + cgt_p^1((\Phi^1(t) - \hat{\alpha}_s^1(t)S(t))\sigma_s(t)P^*(t) / P(t)\sigma_p(t))] \\
& + T^2[t_s^2((N - \hat{\alpha}_s^1(t))\delta_s(t)) + t_p^2((\Phi^2(t) - (N - \hat{\alpha}_s^1(t))S(t))\sigma_s(t)\delta_p(t) / P(t)\sigma_p(t))] \\
& + CGT^2[cgt_s^2((N - \hat{\alpha}_s^1(t))S^*(t)) + cgt_p^2((\Phi^2(t) \\
& - (N - \hat{\alpha}_s^1(t))S(t))\sigma_s(t)P^*(t) / P(t)\sigma_p(t))] \tag{A.5}
\end{aligned}$$

The first-order condition of the problem consisting of minimizing this expression with respect to $\alpha_s^1(t)$ is (A.4). Q.E.D.

Appendix B. More Detail on Restricted Global, General Global, and Local Tax Arbitrage

B.1. Restricted Global Tax Arbitrage Opportunity for Single Investors

This arbitrage opportunity exists if and only if there is a set of asset holdings, $\alpha_{A_j}^i(t)$, for investor i , satisfying the following system of inequalities:

$$\begin{aligned}
 dX^i(t) &= \sum_j +M \cdot \alpha_{A_j}^i(t) [dA_j^B(t) + \delta_{A_j}^B(t)dt] - T^i \left[\sum_j t_{A_j}^i (+M \cdot \alpha_{A_j}^i(t) \delta_{A_j}^B(t)) \right] dt \\
 &\quad - CGT^i \left[\sum_j cgt_{A_j}^i (+M \cdot \alpha_{A_j}^i(t) A_j^{B*}(t)) \right] dt \\
 &\geq 0 \quad \forall dA_j^B(t), \delta_{A_j}^B(t) dt, j=1,2
 \end{aligned} \tag{B.1}$$

$$X^i(t) = \sum_j \alpha_{A_j}^i(t) A_j(t) = 0 \tag{B.2}$$

(Note: different from the inequality (5.64), $+M$ represent a very large number. It is multiplied with $\alpha_{A_j}^i(t)$ to make sure that the total number of holding shares in arbitrage portfolio exceeds the minimum requirement.)

B.2. General Global Tax Arbitrage Opportunity for Single Investor

This arbitrage opportunity exists if and only if there is a set of asset holdings, $\alpha_{A_j}^i(t)$, for investor i , satisfying the following system of inequalities:

$$\begin{aligned}
 dX^i(t) &= \sum_j \lambda \cdot \alpha_{A_j}^i(t) [dA_j^B(t) + \delta_{A_j}^B(t)dt] - T^i \left[\sum_j t_{A_j}^i (\lambda \cdot \alpha_{A_j}^i(t) \delta_{A_j}^B(t)) \right] dt \\
 &\quad - CGT^i \left[\sum_j cgt_{A_j}^i (\lambda \cdot \alpha_{A_j}^i(t) A_j^{B*}(t)) \right] dt \\
 &\geq 0 \quad \forall \lambda \in (0, Liquidity); dA_j^B(t), \delta_{A_j}^B(t) dt, j=1,2
 \end{aligned} \tag{B.3}$$

$$X^i(t) = \sum_j \alpha_{A_j}^i(t) A_j(t) = 0 \quad (\text{B.4})$$

(Note: different from the inequality (5.61), λ is multiplied with $\alpha_{A_j}^i(t)$ to make sure that the number of holding shares in arbitrage portfolio can be cut or increased.)

B.3. Local Tax Arbitrage Opportunity for Multiple Investors

This arbitrage opportunity exists in the market if and only if for a certain asset $j \in \{1, 2, \dots\}$, there is a set of asset holdings between investors $i = 1, 2, \dots$, $\alpha_{A_j}^i(t)$, satisfying the following system of inequalities:

$$\begin{aligned} dX_j(t) = \sum_i \{ & \alpha_{A_j}^i(t) [dA_j^B(t) + \delta_{A_j}^B(t) dt] - T^i [t_{A_j}^i (\alpha_{A_j}^i(t) \delta_{A_j}^B(t))] dt \\ & - CGT^i [cgt_{A_j}^i (\alpha_{A_j}^i(t) A_j^{B*}(t))] dt \} \\ & \geq 0 \quad \forall dA_j^B(t), \delta_{A_j}^B(t) dt \end{aligned} \quad (\text{B.5})$$

$$X_j(t) = \sum_i \alpha_{A_j}^i(t) A_j(t) = 0 \quad (\text{B.6})$$

B.4. Restricted Global Tax Arbitrage Opportunity for Multiple Investors

This arbitrage opportunity exists in the market if and only if for a certain asset $j \in \{1, 2, \dots\}$, there is a set of asset holdings between investors $i = 1, 2, \dots$, $\alpha_{A_j}^i(t)$, satisfying the following system of inequalities:

$$\begin{aligned} dX_j(t) = \sum_i \{ & +M \cdot \alpha_{A_j}^i(t) [dA_j^B(t) + \delta_{A_j}^B(t) dt] - T^i [t_{A_j}^i (+M \cdot \alpha_{A_j}^i(t) \delta_{A_j}^B(t))] dt \\ & - CGT^i [cgt_{A_j}^i (+M \cdot \alpha_{A_j}^i(t) A_j^{B*}(t))] dt \} \\ & \geq 0 \quad \forall dA_j^B(t), \delta_{A_j}^B(t) dt \end{aligned} \quad (\text{B.7})$$

$$X_j(t) = \sum_i \alpha_{A_j}^i(t) A_j(t) = 0 \quad (\text{B.8})$$

(Note: different from the inequality (B.5), $+M$ is multiplied with $\alpha_{A_j}^i(t)$ to make sure that the number of holding shares in arbitrage portfolio exceeds the minimum requirement.)

B.5. General Global Tax Arbitrage Opportunity for Multiple Investors

This arbitrage opportunity in the market exists if and only if for a certain asset $j \in \{1, 2, \dots\}$, there is a set of asset holdings between investors $i = 1, 2, \dots$, $\alpha_{A_j}^i(t)$, satisfying the following system of inequalities:

$$\begin{aligned} dX_j(t) = \sum_i \{ & \lambda \cdot \alpha_{A_j}^i(t) [dA_j^B(t) + \delta_{A_j}^B(t) dt] - T^i [t_{A_j}^i (\lambda \cdot \alpha_{A_j}^i(t) \delta_{A_j}^B(t))] dt \\ & - CGT^i [cgt_{A_j}^i (\lambda \cdot \alpha_{A_j}^i(t) A_j^{B'}(t))] dt \} \\ \geq 0 \quad \forall \lambda \in (0, \text{Liquidity}); & dA_j^B(t), \delta_{A_j}^B(t) dt \end{aligned} \quad (\text{B.9})$$

$$X_j(t) = \sum_i \alpha_{A_j}^i(t) A_j(t) = 0 \quad (\text{B.10})$$

(Note: different from the inequality (B.5), λ is multiplied with $\alpha_{A_j}^i(t)$ to make sure that the number of holding shares in arbitrage portfolio can be cut or increased.)