Return Predictability and Its Implications for Portfolio Selection

Min Zhu

Business School

The University of Sydney

A thesis submitted in fulfilment of the requirements

for the degree of Doctor of Philosophy

September 2012

Statement of Originality

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes. I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

Min Zhu

Acknowledgements

Firstly, I am deeply grateful to my advisors, Maxwell Stevenson and Ross Sparks. They have been supportive of the direction that I have taken and let me be in control. I could always rely on them for wise advice. This thesis would not have been possible without their support and encouragement.

In addition to my advisors I have been fortunate to have many other great researchers here at USYD. Graham Partington has given me lots of good advice on various issues from my research to career development. David Johnstone has introduced me some interesting research questions. Juan Yao has always been stimulating. With her, we have discussed and exchanged ideas on more than one occasion.

Special thanks go to John Donnelly at CSIRO for his generous help on adminstration issues at CSIRO. Without him, the close link with CSIRO would not have been possible. This close link has proved to be a valuable resource in my PhD studies.

I thank the Quantitative Equity Products Team at Schroders for offering me the opportunity to working on their research problems. This joint journey has proved to be fruitful. I have gained practical knowledge of investing from working with the team members, David Philpotts, Richard Lawson, Michael O'Brien and Rupeng Chen. I appreciate Russell Investment for providing data on Russell indices. Ziqian Huang and Amanda Lin at Russell Investment have put up with my endless data inquiries. Many thanks go to Manfred Gilli and Enrico Schumann at the University of Geneva for their help on the heuristic optimization algorithm.

Multiple generations of PhD students at Business School USYD have made

my time here both more enjoyable and more productive. Many thanks go to Kaiying, Nurhidayah, Rui, Susana, Raza, Paristoo and Dianelle, amongst many others.

It is a pleasure to thank research student program managers Maryann and Mandy for looking after us and making the research centre a second home to us. Financial support from the Capital Markets Cooperative Research Centre (CMCRC) is also gratefully acknowledged.

And finally, I am eternally obliged to my partner Yougan for his understanding and support through the duration of my studies. He has always been here for me when I got frustrated from the PhD research. His patience and humor have rescued me from peril more times than I can recall.

Publications and Working Papers Relevant to the Thesis

1. **Zhu, M.** (2012). Jackknife for Bias Reduction in Predictive Regressions. Forthcoming *Journal of Financial Econometrics.*

2. **Zhu, M.**, D. Philpotts and M.J. Stevenson (2012). Classification and Regression Trees and Their Use in Financial Modeling. Book Chapter. Forthcoming in F.J. Fabozzi (Eds.) *Encyclopedia of Financial Models.* Wiley.

3. **Zhu, M.**, D. Philpotts, R. Sparks and M.J. Stevenson (2011). A hybrid Approach to Combining CART and Logistic Regression for Stock Ranking. *Journal of Portfolio Management*, Fall **38**(1), 100 - 109.

4. **Zhu, M.**, D. Philpotts, and M.J. Stevenson. The Benefits of Tree-based Models for Stock Selection. Forthcoming *Journal of Asset Management.*

5. **Zhu, M.**. Return Prediction and Portfolio Selection: A Distributional Approach. *Working paper.*

Abstract

This thesis inquires into a range of issues in return predictability and its implications. First, the thesis investigates estimation bias in predictive regressions. This research stresses the importance of accounting for the bias when studying predictability. To tackle the problem of biased estimation, a general and convenient method based on the jack-knife technique is proposed. The proposed method reduces the bias for both single- and multiple-regressor models and for both short- and long-horizon regressions. Compared with the existing bias-reduction methods in the literature, the proposed method is more stable, robust and flexible. More importantly, it can successfully reduce the estimation bias in long-horizon regressions, whereas the existing bias-reduction methods in the literature cease to work. The effectiveness of the proposed method is demonstrated by simulations and empirical estimates of common predictive models in finance. Empirical results show that the significant predictive variables under ordinary least squares become insignificant after adjusting for the finite-sample bias. These results cast doubt on conclusions drawn in earlier studies on the return predictability by these variables.

Next, this thesis examines the predictability of return distributions. It provides detailed insights into predictability of the entire stock and

bond return distributions in a quantile regression framework. The difficulty experienced in establishing predictability of the conditional mean through lagged predictor variables does not imply that other parts of the return distribution cannot be predicted. Indeed, many variables are found to have significant but heterogenous effects on the return distributions of stocks and bonds. The thesis establishes a quantile-copula framework for modelling conditional joint return distributions. This framework hinges on quantile regression for marginal return distributions and a copula for the return dependence structure. The framework is shown to be flexible and general enough to model a joint distribution while, at the same time, capturing any non-Gaussian characteristics in both marginal and joint returns.

The thesis then explores the implications of return distribution predictability for portfolio selection. A distribution-based framework for portfolio selection is developed which consists of the joint return distribution modelled by the quantile-copula approach and an objective function accommodating higher-order moments. Threshold-accepting optimisation technique is used for obtaining optimal allocation weights. This proposed framework extends traditional moment-based portfolio selection in order to utilise the whole predicted return distribution.

The last part of the thesis studies nonlinear dynamics of cross-sectional stock returns using classification and regression trees (CART). The CART models are demonstrated to be a valuable alternative to linear

regression analysis in identifying primary drivers of the stock returns. Moreover, a novel hybrid approach combining CART and logistic regression is proposed. This hybrid approach takes advantage of the strengths in both CART and linear parametric models. An empirical application to cross-sectional stock return prediction shows that the hybrid approach captures return dynamics better than either a standalone CART or a logistic model.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

Return predictability is of profound importance in many fields of finance including asset pricing, portfolio management and risk management and, hence, has been one of the most researched areas in finance for decades. Academia and practitioners have done a huge amount of theoretic and empirical work over the past few decades. Despite the vast literature on return predictability, John Cochrane, the president of the American Finance Association, devoted his recent presidential address to return prediction (Cochrane, 2011). Cochrane claims the journey to understand returns is only beginning at both the time-series and cross-sectional levels. Return prediction, or more generally understanding returns, is an evergreen research area in finance.

The research in this thesis aims to contribute to understanding returns. It explores a range of research issues, including estimation bias in predictive regressions, return distribution predictability and its implications for portfolio selection

and the use of nonlinear approaches for return prediction. The selected topics reflect my research interests developed over the past three years. This chapter introduces the background and motivation of these problems. It also reviews some important literature in these areas.

1.1.1 Predictive regressions and estimation bias

Based on lagged predictor variables, stock returns are assumed to be predictable in the current conditional asset pricing literature. The economic method employed in a typical study is a predictive regression which is an ordinary least squares (OLS) regression of the stock return, r_t, on lagged predictor variables. Early studies reviewed by Fama (1970) used such models to examine market efficiency. In the past few decades, numerous other studies have used predictive regressions to investigate predictability of returns.

Many variables have been found to predict returns. The most prominent ones are financial ratios such as the dividend yield (Campbell, 1987; Fama and French, 1988; Hodrick, 1992; Ang and Bekaert, 2007; Cochrane, 2008; Lettau and Van Nieuwerburgh, 2008; Binsbergen and Koijen, 2010), the earnings-price ratio (Rozeff, 1984; Campbell and Shiller, 1988; Lamont, 1998), the book-to-market ratio (Kothari and Shanken, 1997; Pontiff and Schall, 1998) and accruals (Sloan, 1996; Fama and French, 2008). But other variables have also been found to be powerful predictors, such as short-term interest rates (Fama and Schwert, 1977; Campbell, 1987; Breen *et al.*, 1989; Ang and Bekaert, 2007), inflation (Fama, 1981; Fama and Schwert, 1977; Campbell and Vuolteenaho, 2004), consumption-to-wealth ratio (Lettau and Ludvigson, 2001, 2005) and net stock issues (Ikenberry

et al., 1995; Loughran and Ritter, 1995; Daniel and Titman, 2006; Pontiff and Woodgate, 2008).

Moreover, academia claim that the predictive component is stronger over long horizons than over short horizons. Fama and French (1988) find that return forecast t-statistics rise with horizon, suggesting that long-horizon return regressions offer greater statistical evidence for return predictability. Cochrane (1999) calls the long-horizon predictability one of the three most important facts in finance. This fact is emphasised repeatedly in other studies, including Fama (1998), Campbell (2001) and Barberis and Thaler (2003), among others. The conventional wisdom in the literature is that long-run regressions produce more accurate results by strengthening the signal coming from the data while eliminating the noise.

These conclusions, at both short and long horizons, however, have been subject to great statistical scrutiny on the grounds that the persistence of the predictor variables, and the correlation of the innovations of the regressors with those of returns, might bias the regression coefficients and affect t-statistics (Nelson and Kim, 1993; Stambaugh, 1999; Torous *et al.*, 2004; Campbell and Yogo, 2006). Many of the predictor variables used in predictive regressions are highly persistent. For example, at a monthly frequency, all four common regressors, the dividend yield, the earnings-price ratio, the book-to-market ratio and the short-term rate, have a first-order auto-regression structure with an auto-regression coefficient near one (See, for example, Campbell and Yogo, 2006). This leaves the unit root problem wide open. In such situations, the OLS estimators of the coefficients exhibit finite-sample bias which can greatly contaminate statistical inference. A further problem is the possibility of data mining. Ferson *et al.* (2003) and Ferson *et al.* (2008) study the combined effects of data mining and

spurious regression in the context of predictive regressions. They find that the effects of data mining and spurious regression interact and reinforce each other, leading to large regression bias.

Nelson and Kim (1993) stress that "the estimated biases are large enough to affect inference in practice, and should be accounted for when studying predictability". To be concrete, Elliott and Stock (1994) provide Monte Carlo evidence which suggests a 20% size distortion in t-tests caused by estimation bias for plausible parameter values and sample sizes in a one-period regression of returns on the dividend yield. Stambaugh (1999) derives the exact finite-sample bias expression for one-period single-regressor regressions. He reports the bias equals one-third of the OLS estimate when NYSE returns are regressed on the dividend yield over the period 1927 to 1996. Other studies, including Goyal and Welch (2003), Amihud and Hurvich (2004), Lewellen (2004), Campbell and Yogo (2006) and Ang and Bekaert (2007), conclude that the statistical evidence of predictability is weaker or even disappears once tests are adjusted for estimation bias.

Spurious regression bias is also an issue for long-horizon regressions. As pointed out by Valkanov (2003), regressing a long-run variable on a short-run variable yields inconsistent estimates, and spurious regression relations may be found between two independent variables. Lanne (2002) declares that evidence of predictability over both short and long horizons is spurious and follows from a neglected near unit root problem. Torous *et al.* (2004) find that evidence of predictability is reliable at shorter horizons but non-existent at long horizons after accounting for finite-sample biases. Boudoukh *et al.* (2008) show that both OLS coefficient estimates and R^2 are proportional to the horizon even under the null

hypothesis of no predictability.

The predictability of stock returns is therefore still an open question. A difficulty with understanding the rather large body of literature on predictability is the muddle caused by estimation bias in predictive regressions. When facing the problem of biased estimation in predictive regressions, most of the attention in the finance literature has been directed at constructing valid tests of significance (see, for example, Nelson and Kim, 1993; Valkanov, 2003; Lewellen, 2004; Campbell and Yogo, 2006). Much less attention has been given to the problem of obtaining better estimators. Whereas scaling the critical value in t-test upwards to get conservative confidence intervals can help defend somewhat against spurious regression, obtaining bias-reduced estimates is a more direct way to address the problem. In addition, accurate estimates have important economic and practical value in out-of-sample forecasts, which is often the ultimate purpose of predictive regressions in practice. Therefore, there is an urgent need for alternative econometric methods for correcting the bias and conducing valid inference.

1.1.2 Return distribution predictability and its implication for portfolio selection

Because predictive regressions can offer only a conditional mean relationship between returns and predictor variables, analysis of return predictability with predictive regressions limits inquiries to the conditional mean only. Historically, the return predictability literature had an almost exclusive focus on the conditional mean of returns. Over the past two decades, much literature has emerged which explores the predictability of stock return volatility. For example, Poon

and Granger (2003) review the practice of forecasting volatility in financial markets. This extends the predictability inquiries to the second moment. Return predictability, however, should investigate more than the first two moments.

First, there is considerable evidence, both theoretical and empirical, that return distributions are in general not normal and, hence, cannot be adequately characterised by the first two moments alone. Merton (1982) shows that if instantaneous returns are normal, then the price process is lognormal and, unless the measurement interval is very small, the simple returns are not normal. Numerous studies, including Fama (1965), Harvey and Zhou (1993), Chen *et al.* (2001), Cont (2001), Hueng and McDonald (2005), Chiang and Li (2007) and Post *et al.* (2008) have found significant asymmetries in empirical asset returns. Apparently, the first two moments do not reveal a comprehensive picture of returns.

Second, in many areas of financial economics, knowledge is required of either the entire return distribution or other parts of the distribution than the conditional mean. In asset pricing, higher-order moments such as skewness and kurtosis have proven useful to explain variation in stock returns. For example, studies have investigated the skewness preference in investor investment decisions and its impacts on asset pricing through the work of Harvey and Siddique (2000), Brunnermeier *et al.* (2007), Mitton and Vorkink (2007), Barberis and Huang (2008), Boyer *et al.* (2010) and others. Studies on kurtosis preference include Fang and Lai (1997), Dittmar (2002), Guidolin and Timmermann (2008), among others. In risk management, focus is usually on the lower tails of the return distribution; however, in portfolio management, investors generally require an estimate of the entire distribution of future returns. Hence, understanding

return predictability in more detail has great economic importance in many areas of financial economics.

A natural tool to investigate whether a return distribution is predictable using lagged predictor variables is quantile regression. Quantile regression has been introduced by Koenker and Bassett Jr (1978, 1982). The basic ideas, though, go back to the earliest work on regression by Boscovich in the mid-18th century and to Edgeworth at the end of the 19th century (Edgeworth, 1888, see also the introduction on the history of quantile regression by Koenker, 2005). Generalising the common linear regression framework by shifting the focus from the conditional mean to conditional quantiles allows quantile regression to provide a conditional distribution view instead of a mere conditional mean. Quantile regression has gradually become a complementary approach to the traditional mean regression methods.

Recently, both practitioners and academia have started to use quantile regression to investigate return distribution predictability. Cenesizoglu and Timmermann (2008) study whether the distribution of the S&P 500 monthly returns is predictable using lagged economic variables. By employing a quantile regression framework, they find the significant predictability, both in sample and out-of-sample, of the entire stock return distribution. Ma and Pohlman (2008) demonstrate that under some strict assumptions and a symmetric loss function, the use of quantile regression leads to better return forecasts. Investment practitioners Gowlland *et al.* (2009) show that factor effects are not constant across return distributions using cross-sectional stock data. They advocate to use quantile regression as a tool in quantitative investing for better understanding and controlling factor risks. Pedersen (2010) uses quantile regression to examine the

7

predictability of the S&P 500 monthly returns and monthly returns of the US 5-year Treasury bonds. He reports strong empirical evidence of distribution predictability of both stock and bond returns.

Return predictability is especially pertinent to portfolio selection. For an individual investor with a given utility, portfolio selection is essentially a comparison of the future investment return distributions. However, the difficulty in obtaining joint return distributions often leads to approximate distributions with a few individual moments. This has resulted in a large body of literature focusing on moment-based analysis of portfolio selection. For example, the classic mean-variance framework by Markowitz (1952) uses the first two moments of the distribution of returns. Portfolio selection with a few higher moments has also been considered in the literature, such as three-moments, mean-variance-skewness portfolio selection (see, for example, de Athayde and Flôres, 2004; Briec et $al.$, 2007; Mencía and Sentana, 2009), and four-moments, mean-variance-skewness-kurtosis portfolio selection (see, for example, Jurczenko et $al.$, 2006; Guidolin and Timmermann, 2008).

Despite the tractability and economic appeal of such moment-based models, Brockett and Kahane (1992) point out that investors do not, in general, have preferences that can be translated into a function of the first N moments of the return distribution. Further, the use of individual moments for portfolio selection ignores the fact that portfolio characteristics are jointly defined by all higher-order moments instead of a few individual moments. Statistically, it is also extremely difficult to establish that an effect is caused by, say, the third moment as opposed to all moments of order three or higher. This strongly suggests that any portfolio selection approach based on a few individual moments is myopic.

The research on return distribution predictability is in its infancy. Financial theory and empirical studies need to be expanded. The resulting conditional probability distributions would be useful in many applications, especially in portfolio management. Exploration of return distribution predictability on portfolio management is, therefore, of great interest.

1.1.3 Nonlinear return prediction and CART

The vast majority of the literature examines stock return predictability in a linear regression framework. The standard asset pricing models such as the Capital Asset Pricing Model (CAPM) (Sharpe, 1964; Lintner, 1965; Mossin, 1966), the Arbitrage Pricing Theory (APT) (Ross, 1976), the Fama-French three-factor-model (Fama and French, 1993) and the Carhart four-factor model (Carhart, 1997) all assume a linear relationship between the mean returns and the factor loadings. However, there is no priori reason to believe that asset returns respond in a linear fashion to risk factors. Indeed, there is increasing evidence that asset returns may be better characterised by a model which allows for nonlinear behaviour. Hsieh (1991) finds strong evidence of nonlinearity in stock returns. Bansal and Viswanathan (1993) and Bansal *et al.* (1993) propose a nonlinear arbitrage-pricing model which relaxes the linearity restriction of the APT. They find that this nonlinear APT model is more adequate in explaining the returns. Hiemstra and Jones (1994) find evidence of nonlinear causality from volume to returns. Based on their findings, they advocate future research should consider nonlinear theoretical mechanisms and empirical regularities when devising and evaluating models of the joint dynamics of stock prices and trading volume.

Dittmar (2002) reports substantial benefits in modelling the pricing kernel as a nonlinear function of the market return.

The frequency of large moves in stock markets during the Global Financial Crisis between 2007 and 2009 is much greater than would be expected under a normal distribution. This again reveals the inadequacy of a linear regression framework for return analysis. After the Global Financial Crisis, interest in nonlinear dynamics has increased in both the financial press and the academic literature. The motivation behind financial practitioners' interest in nonlinear dynamics, though, is slightly different from that of academia; it is more from the model risk diversification point of view as explained below.

Historically, linear factor models have been widely accepted and used by financial practitioners. According to a series of surveys of modelling techniques among large asset managers in the United States and Europe by the Chartered Financial Analyst (CFA) Institute [1], linear regression remains the primary workhorse for financial modelling, being used by the vast majority of the firms surveyed. Given the widespread use of linear factor models and similar data sources by many firms, there was a high degree of commonality in the trading strategies used by quantitative investors prior to the "quant-shock" of July-August 2007 (Ang, 2008; Khandani and Lo, 2011). The poor relative returns experienced by the vast majority of quantitatively orientated asset managers between 2007 and 2009 has exposed the risk introduced by common forecasting tools.

As nonlinear modelling techniques are not at present widely used within the investment community, they are appealing in the context of offering a high degree

---

[1]The survey results can be found in the monographs by Fabozzi *et al.* (2006) and Fabozzi *et al.* (2008).

of model diversification. Because linear factors are relatively easily determined, the same factors will tend to be used by a large number of managers, eliminating the profit from using them. Thus, unexploited profit opportunities are most likely to be found in the nonlinearities of the market.

Many types of nonlinear modelling techniques have been applied to or have potential to be applied to the problem of predicting returns. Among them, a decision tree technique called classification and regression trees (CART) contains certain properties which make it suitable for return prediction. Proposed by Breiman *et al.* (1984), CART is a nonlinear and non-parametric modelling technique that does not impose the stringent assumptions required by classical regression analysis. It is robust, flexible and distribution-free. Instead of taking a "black box" approach, CART models are intuitive and straightforward and allow for economic interpretation. More importantly, CART models are well suited to identifying any complex interactions in the data.

Although the approach is not widely utilised within the investment community, the applications of CART to financial markets nevertheless include the classification of financially distressed firms by Frydman *et al.* (1985), asset allocation by Sorensen *et al.* (1998), equity style timing by Kao and Shumaker (1999) and stock selection by Sorensen *et al.* (2000).

One of the challenges facing both academia and practitioners in cross-sectional stock return analysis is, as pointed out by Cochrane (2011), how to identify the primary drivers of stock returns within a plethora of new variables. Cochrane (2011) claims that methods other than the Fama-French style of regression should be used. The statistical properties of CART make it suitable for cross-sectional return prediction. A systematic evaluation of tree-based models for cross-sectional

stock return forecasting is, therefore, of interest to both academics and practitioners.

1.2   Objectives of Thesis

Building on the background and motivation in the previous section, this thesis has four main objectives.

First, motivated by the importance of correcting the finite-sample bias in predictive regressions, the thesis aims to contribute to an active recent literature on alternative econometric methods for reducing the bias. Specifically, this thesis proposes a general and convenient method based on the jackknife technique for bias reduction. The method works for both single- and multiple-regressor models and for both short- and long-horizon regressions. As the existing methods do not work in long-horizon regressions, the research fills the gap in the literature by suggesting better estimators for long-horizon regressions. This part of the research also aims to provide a comprehensive evaluation of all the available bias-reduction methods in the literature, as well as to re-examine some popular empirical evidence used to support the claim of return predictability in the literature after accounting for the finite-sample bias.

Second, motivated by the limitation in the current practice of return predictability, this thesis goes beyond predictability of the conditional mean and variance and examines whether stock and bond return distributions are more generally predictable. For this purpose, a quantile regression framework is employed and a wide range of lagged economic state variables are considered in order to predict different quantiles of stock and bond return distributions. The thesis

contributes to the literature by providing new empirical evidence of distribution predictability using monthly returns of two broad-based indices, the Russell 1000 Index and the US Aggregate Bond Index. Compared to the S&P 500 Index and the 5-year Treasury bonds studied by Cenesizoglu and Timmermann (2008) and Pedersen (2010), these two broad-based indices are more comprehensive and unbiased barometers for the US stock and bond markets. Given their wide recognition in investment communities, predictability of these indices has academic value as well as significant economic value to investors.

Additionally, the thesis proposes a quantile-copula framework to model a joint distribution of asset returns, with the quantile approach to extract systematic information in marginal distributions and copulas to capture the dependence structure.

Third, motivated by the limitation of moment-based portfolio selection and the empirical evidence of return distribution predictability, the thesis develops a distribution-based framework for portfolio selection by incorporating predicted return distributions. More specifically, this distribution-based portfolio selection framework includes a joint return distribution modelled by the quantile-copula approach and an objective function which is a generalisation of the Omega measure introduced by Shadwick and Keating (2002). The solutions can be obtained by a heuristic optimisation technique. This distribution-based portfolio selection overcomes some limitation of moment-based portfolio selection approaches.

Last, motivated by limitation of linear regression based approaches for asset pricing, this thesis investigates the use of a nonlinear model, CART, to analyse cross-sectional stock returns. As a joint work with the Quantitative Equity Prod-

uct Team at Schroder Investment Management [1], this part of the research is more practitioner-oriented. The research aims to provide a comprehensive analysis of the strengths and weaknesses of CART for stock return prediction over the linear regression based approaches. Further, to take advantage of the strengths in both the CART and linear parametric models, a novel hybrid approach combining CART and logistic regression is also proposed and tested for cross-sectional return prediction.

1.3 Structure of Thesis

The organisation of this thesis is as follows. Chapter 2 analyses the finite-sample bias in predictive regressions. The existing bias-reducing methods in the literature are reviewed and their pros and cons are discussed. A new bias-reduction method based on the jackknife technique is proposed for obtaining bias-reduced estimates. A systematic comparison of the performance of the proposed estimator with the existing ones is carried out by simulations. An empirical application to equity premium prediction using the dividend yield and the short rate highlights the differences between the results derived from the standard approach and those from the bias-reduced estimator. The significant predictive variables under ordinary least squares become insignificant after adjusting for the finite-sample bias. These results cast doubt on conclusions drawn in earlier studies about the significance of these two variables.

Chapter 3 introduces quantile regression and uses it as a tool to investigate predictability of the return distributions of both stocks and bonds. The difficulty

---

[1]http://www.schroders.com/qep/home/

experienced in establishing predictability of the conditional mean through economic state variables does not imply that other parts of the return distribution cannot be predicted. Indeed, many variables in the range of considered economic state variables are found to have significant but heterogenous effects on the return distributions of stocks and bonds. For a sufficiently fine grid of quantiles, an entire marginal distribution of asset returns can be traced out.

Extending the work in Chapter 3, Chapter 4 develops a quantile-copula framework to model a joint return distribution. Further, Chapter 4 generalises the Omega measure introduced by Shadwick and Keating (2002) based on the prospect theory. This generalised Omega measure is then used as an objective function for asset allocation. A heuristic optimisation technique called threshold accepting algorithm is also introduced in Chapter 4. The joint distribution modelled by the quantile-copula framework, along with the proposed objective function and threshold accepting algorithm, makes possible distribution-based portfolio selection that utilises all of the underlying return distribution information. The proposed portfolio selection framework is illustrated by an empirical application to asset allocation between stocks and bonds.

Chapter 5 introduces the CART modelling technique. Both theoretical and empirical comparisons of CART with linear regression based approaches for stock return analysis are given. CART is demonstrated to be a valuable alternative to linear regression analysis in identifying the primary drivers of the stock returns.

Chapter 6 discusses limitations of the CART model for return prediction and proposes a novel hybrid approach to combining CART and logistic regression. This hybrid approach takes advantage of the strengths in both CART and linear parametric models, and results in enhanced predictions of cross-sectional stock

returns. An empirical application to US stock data demonstrates that, in comparison with tree-based models and logistic regression, the proposed hybrid approach enhances portfolio returns over time without introducing significant risks.

Finally, Chapter 7 summarises the key findings and suggests directions for future research.

Chapter 2

Predictive Regression for Return Prediction and Bias Reduction

2.1   Predictive Regressions

Predictive regressions for stock returns have long been a staple of financial economics. The simplest predictive regression is single-regressor and one-period regression which is to regress the stock return, r_t, on a lagged predictor variable, x_{t-1},

$$r_t = \alpha + \beta x_{t-1} + u_t,$$

where u_t is an error term. As reviewed in Section 1.1.1, examples of lagged predictor variables include the dividend yield, the earnings-price ratio, the book-to-market ratio, and various measures of the interest rate. Many of these variables behave as highly persistent time series and their disturbance terms are contemporaneously correlated with those of returns. These characteristics can be captured

mathematically by a class of predictive regressions specified as follows

$$r_t = \alpha + \beta x_{t-1} + u_t, \tag{2.1}$$

where x_{t-1} is a first-order autoregressive process,

$$x_t = \phi + \rho x_{t-1} + v_t. \tag{2.2}$$

The bivariate error terms (u_t, v_t) follows a joint normal distribution with mean 0 and a covariance matrix

$$\begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix}. \tag{2.3}$$

A number of scholars have studied the statistical properties of this class of predictive regressions, including Stambaugh (1999), Amihud and Hurvich (2004), Campbell and Yogo (2006), among others.

Another popular predictive model is a long-horizon prediction regression which regresses future p-period returns onto a one-period predictor variable, captured by models of the form

$$r_{t+p} = \alpha_p + \beta_p x_{t-1} + u_{t+p}, \tag{2.4}$$

where $r_{t+p} = \sum_{i=0}^{p-1} r_{t+i}$, a moving summation. When $p = 1$, it is the one-period predictive regression (2.1). As reviewed in Section 1.1.1, though it is not without controversy, the strongest evidence of the return predictability cited so far comes from long-horizon predictive regressions.

2.2 Biases in Predictive Regressions

In the one-period predictive regression specified by 2.1 to 2.3, the OLS coefficient estimates are subject to finite-sample biases as illustrated below.

The relationship between u_t and v_t can also be written as $u_t = \xi v_t + \epsilon_t$, where $\xi = \sigma_{uv}/\sigma_v^2$ and ϵ_t are independently and identically distributed (i.i.d.) errors, which are independent of v_t, $i.e.$, $E(\epsilon_t|v_1, v_2, ...v_T) = 0$. In this setup, the marginal mean is $E(x_t) = \phi/(1-\rho)$, and the marginal variance is $var(x_t) = \sigma_v^2/(1-\rho^2)$.

Let X be the design matrix whose t-th row is $(1, x_{t-1})$, and $R = (r_1, r_2, ..., r_T)'$. The OLS estimator of the regression coefficients in (2.1) is

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} = (X'X)^{-1}X'R$$

with the variance given by

$$\sigma_{ols}^2 = (X'X)^{-1}X'\sigma_u^2. \tag{2.5}$$

Denoting $\bar{x} = \sum_{t=1}^{T} x_{t-1}/T$ and making use of the fact that $r_t = \alpha + \beta x_{t-1} + u_t$, the bias in $\hat{\beta}$ is

$$\hat{\beta} - \beta = \frac{\sum_{t=1}^{T}(x_{t-1} - \bar{x})u_t}{\sum_{t=1}^{T}(x_{t-1} - \bar{x})^2}.$$

Using $E(u_t|v_t) = \xi v_t$ and $v_t = x_t - \phi - \rho x_{t-1}$, the finite-sample bias in $\hat{\beta}$ is

$$E(\hat{\beta}) - \beta = \xi E\left\{ \frac{\sum_{t=1}^{T}(x_{t-1} - \bar{x})E(x_t|v_t)}{\sum_{t=1}^{T}(x_{t-1}^2 - \bar{x}^2)} - \rho \right\}. \tag{2.6}$$

Let $\hat{\rho}$ be the OLS estimator of ρ,

$$\hat{\rho} = \frac{\sum_{t=1}^{T}(x_{t-1} - \bar{x})x_t}{\sum_{t=1}^{T}(x_{t-1}^2 - \bar{x}^2)},$$

and the bias in $\hat{\beta}$ can be written as

$$\mathrm{E}(\hat{\beta}) - \beta = \xi\{\mathrm{E}(\hat{\rho}) - \rho\}.$$

According to Marriott and Pope (1954) and Kendall (1954), under the assumptions of normality and AR(1) for x_t, the bias in $\hat{\rho}$ can be expressed as

$$\mathrm{E}(\hat{\rho}) - \rho = -(1 + 3\rho)/T + O(1/T^2), \tag{2.7}$$

and it follows that

$$\mathrm{E}(\hat{\beta}) - \beta = -\frac{(1 + 3\rho)}{T}\xi + O\left(\frac{1}{T^2}\right). \tag{2.8}$$

The result in (2.8) appears in Stambaugh's paper (Stambaugh, 1999). As we can see, the bias is proportional to ξ and the autoregressive coefficient ρ but inversely proportional to sample size T.

The problem of estimation bias is more severe in long-horizon predictive regressions than in one-period predictive regressions. As demonstrated by Boudoukh *et al.* (2008), the bias is increasing with the horizon even under the null hypothesis of no predictability. Not only the magnitude of the bias is larger, the bias mechanism is much more complex in long-horizon regressions due to the fact that the returns r_{t+p} become more persistent as p increases. In this application, apart

from the finite-sample bias that arises due to lagged stochastic regressors, there is also the spurious regression bias related to the classic studies of Yule (1926) and Granger and Newbold (1974). This additional bias is caused by autocorrelated errors induced by a highly persistent dependent variable series. These two types of biases reinforce each other in the regression and makes long-horizon regressions even more troublesome when it comes to parameter estimation and statistical inference.

2.3   Existing Methods for Bias Reduction

This section reviews the existing methods in the finance literature for bias reduction in predictive regressions before introducing a new approach in the next section.

The plug-in method by Stambaugh (1999) and the augmented regression method by Amihud and Hurvich (2004) are two existing and fundamental approaches in the literature for obtaining bias-reduced coefficient estimates. Many other methods and tests essentially use one or the other to reduce bias. For example, the Q-statistic of Campbell and Yogo (2006) uses the plug-in method in the numerator to adjust the OLS estimator for the bias. The hypothesis testing in a multiple-regressor regression setup by Amihud *et al.* (2009) uses the augmented regression method to reduce bias. Both methods are developed under the conditions of one-period predictive regressions, and their suitability for long-horizon regressions is not clear. The details of these two methods are given below.

2.3.1 Plug-in method

Motivated by the finite-sample theory of Stambaugh (1999), researchers usually solve the bias problem by a plug-in method. This involves directly estimating the bias using Stambaugh's bias expression (2.8) and then adjusting the OLS estimator for the bias. Denote $\hat{\beta}$ and $\hat{\rho}$ the OLS estimators of β and ρ, respectively. The bias-reduced estimator by the plug-in method is

$$\hat{\beta}_{Plug-in} = \hat{\beta} + \frac{(1+3\hat{\rho})}{T}\hat{\xi}, \tag{2.9}$$

where $\hat{\xi} = \sum \hat{u}_t \hat{v}_t / \sum \hat{v}_t^2$, and $\hat{u}_t$, $\hat{v}_t$ are the residuals from OLS regressions in (2.1) and (2.2), respectively.

The plug-in method, however, suffers from a severe drawback because it relies on availability of explicit bias expressions. As Stambaugh's bias expression is only for single-regressor regressions, there is no plug-in version available for multiple-regressor models.

2.3.2 Augmented regression

Another proposed method for bias reduction in the finance literature is the augmented regression by Amihud and Hurvich (2004). This method consists of two steps: in the first step, the first-order autoregressive coefficients of the predictor variables are estimated and the corresponding residual errors are then calculated; in the second step, the dependent variable is regressed on the predictor variables and their corresponding residual errors from the previous step.

More specifically, the procedure of the augmented method is:

i) Estimate model (2.2) and obtain the OLS estimators of the coefficients,

$\hat{\rho}$ and $\hat{\phi}$. Based on the bias expression in Kendall (1954), construct the bias-corrected estimators of the coefficients ρ and ϕ as $\hat{\rho}^c = \hat{\rho} + (1+3\hat{\rho})/T + 3(1+3\hat{\rho})/T^2$ and $\hat{\phi}_t^c = (1 - \hat{\rho}^c) \sum_{t=1}^{T} x_t/T$. Obtain the residuals $\hat{v}_t^c$ as $\hat{v}_t^c = x_t - (\hat{\phi}_t^c + \hat{\rho}^c x_{t-1})$.

ii) Obtain $\hat{\beta}_{Augmented}$ as the coefficient of x_{t-1} in an OLS regression of y_t on x_{t-1} and $\hat{v}_t^c$, with intercept, $y_t = \alpha + \beta x_{t-1} + \kappa v_t^c + e_t$.

Amihud and Hurvich (2004) show that the OLS estimator $\hat{\beta}_{Augmented}$ is bias-reduced, and the method works for both single- and multiple-regressor models[1]. However, this regression-based bias-reduction method in practice is heavily dependent on correct model specification or, otherwise, spurious results may be obtained.

2.4 Jackknife for Bias Reduction

2.4.1 Ordinary jackknife estimator

The jackknife technique was originally proposed by Quenouille (1949, 1956) for bias reduction. One of the crucial assumptions required by the jackknife is that samples are i.i.d.. We now briefly explain why the jackknife technique works under an i.i.d. assumption, but does not work in predictive regressions.

Suppose we have a sample $\mathcal{S} = (S_1, \cdots, S_T)$ and an estimator $\hat{\theta} = f(\mathcal{S})$. Schucany *et al.* (1971) show that for many common statistics, including most

---

[1]For multi-regressor cases, the estimation procedure is more complicated when there are linear interdependencies among multiple regressors. It involve an iterative estimation procedure in the first step to obtain $\hat{\rho}^c$ and $\hat{v}_t^c$. For more details, refer to Amihud and Hurvich (2004) and Amihud *et al.* (2009).

maximum likelihood estimates, the bias of $\hat{\theta}$ is of the form

$$\mathrm{E}(\hat{\theta}) - \theta = \frac{a}{T} + \frac{b}{T^2} + \cdots, \tag{2.10}$$

where θ is the true underlying value. Under *the assumption that S_1 to S_T are i.i.d. random variables*, a and b do not depend upon T, *i.e.*, constants.

A jackknife estimator has the property that it removes the order $1/T$ term from the bias form (2.10). This is achieved by focusing on the sub-samples that leave out one observation at a time, *i.e.*,

$$\mathcal{S}_{(-t)} = (S_1, S_2, \cdots, S_{t-1}, S_{t+1}, \cdots, S_T),$$

for $t = 1, \cdots, T$. Let $\hat{\theta}_{(-t)} = f(\mathcal{S}_{(-t)})$, the estimator of the same functional form as $\hat{\theta}$ but computed from the sub-sample $\mathcal{S}_{(-t)}$, and define

$$\hat{\theta}_t = T\hat{\theta} - (T-1)\hat{\theta}_{(-t)}.$$

It is easy to see that $\hat{\theta}_t$ is an estimate of θ with the bias term $O(1/T)$ being removed, because

$$\mathrm{E}\hat{\theta}_t = T\left(\theta + \frac{a}{T} + \frac{b}{T^2} + \cdots\right) - (T-1)\left(\theta + \frac{a}{T-1} + \frac{b}{(T-1)^2} + \cdots\right)$$
$$= \theta + O\left(\frac{1}{T^2}\right).$$

The jackknife estimator is the mean of the $\hat{\theta}_t$, $t = 1, \cdots, T$,

$$\hat{\theta}_{JK} = \sum_{t=1}^{T} \hat{\theta}_t / T = T\hat{\theta} - (T-1) \sum_{t=1}^{T} \hat{\theta}_{(-t)} / T. \qquad (2.11)$$

Clearly, $\hat{\theta}_{JK}$ has the similar bias expression as $\hat{\theta}_t$, that is, reducing the OLS bias by a factor of $O(1/T)$, but with a smaller variance.

Under the predictive regression setup specified by (2.1) to (2.3), however, the observations are correlated. Deleting observations from the middle of the time series certainly violates the correlation structure of the data. As a consequence, instead of being constants, the values of a and b in the bias form (2.10) depend on which observation is removed. Hence, the ordinary jackknife estimator (2.11) can no longer reduce bias. This claim is given by the following theorem.

Theorem 2.4.1. *Suppose $\hat{\theta}_{JK}$ is the estimator of β obtained from jackknifing the predictive regression specified by (2.1) to (2.3), we have $E(\hat{\beta}_{JK} - \beta) = O(T^{-1})$.*

Proof. See the Appendix A. □

2.4.2 Moving-block jackknife (MBJK) estimator

To preserve the correlation structure of the data, an alternative is to use a moving block of length l. Let the block shift by one observation each time, resulting in a set of $k = T - l + 1$ sub-samples of the form,

$$\mathcal{S}_i = \{(r_i, x_{i-1}), \cdots, (r_{i+l}, x_{i+l-1})\},$$

for $i = 1, \cdots, k$. All these sub-samples preserve the autocorrelation structure in the regressors and the cross-correlation structure between the regressors and the

returns.

Let $\hat{\beta} = f(\mathcal{S})$ and $\hat{\beta}_{(i)} = f(\mathcal{S}_i)$, the OLS estimates of the slope coefficient by the full sample and the i-th block sample $\mathcal{S}_i$, respectively. Define

$$\hat{\beta}_i = \frac{1}{T-l}(T\hat{\beta} - l\hat{\beta}_{(i)}),$$

for $i = 1, \cdots, k$. Under the condition $l = O(T)$ (*i.e.*, the block size l cannot be too small), each $\hat{\beta}_i$ is a bias-reduced estimate of β as

$$
\begin{aligned}
\mathrm{E}\hat{\beta}_i &= \frac{1}{T-l}\left\{T\left(\beta + \frac{a}{T} + \frac{b}{T^2} + \cdots\right) - l\left(\beta + \frac{a}{l} + \frac{b}{l^2} + \cdots\right)\right\} \\
&= \beta - \frac{b}{Tl} + \cdots \\
&= \beta + O\left(\frac{1}{T^2}\right).
\end{aligned}
$$

The MBJK estimate is the mean of the $\hat{\beta}_i$,

$$\hat{\beta}_{MBJK} = \sum_{i=1}^{k}\hat{\beta}_i/k = \frac{T}{k-1}\hat{\beta} - \frac{l}{k-1}\sum_{i=1}^{k}\hat{\beta}_{(i)}/k, \qquad (2.12)$$

which removes the order $1/T$ bias term from the OLS bias of the form (2.8).

Akin to any non-parametric bias-reduction technique, there is a bias-variance trade-off in the MBJK estimator controlled by the block size l. A heuristic argument is that $\hat{\beta}_i$ based on a large l is more accurate than that based on a small l, and hence $\hat{\beta}_{MBJK}$, the average of $\hat{\beta}_i$s, is more accurate based on a large l. The large block size l, however, results in a small k. Note that the variance of $\hat{\beta}_{MBJK}$ is inverse to k, $\mathrm{var}(\hat{\beta}_{MBJK}) = \mathrm{var}(\hat{\beta}_i)/k$. Therefore, the variability of $\hat{\beta}_{MBJK}$ increases with l. These heuristics are supported by the simulation results below.
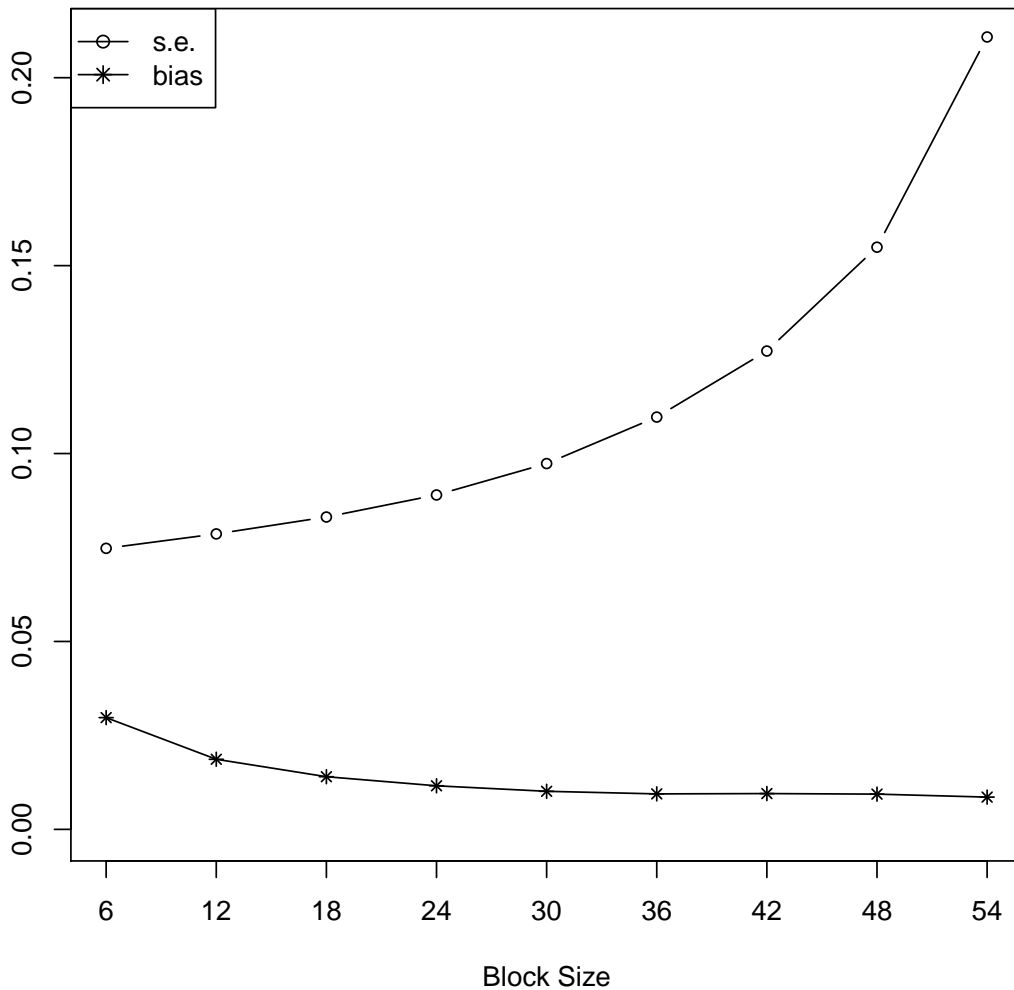
Figure 2.1: Bias-variance trade-off controlled by block size l in the MBJK estimator. Data generated from the models (2.1) to (2.3). Parameter values are $T = 60$, $\rho = 0.99$, and $\delta = -0.95$. Nine block sizes $l = 6, 12, 18, \cdots, 54$ are considered. For each block size, two statistics – the bias and standard error (s.e.) of the MBJK estimator – are reported based on 10,000 samples.

We generate data from the model specified by (2.1) to (2.3). The correlation between u_t and v_t, δ, takes -0.9. The autoregressive coefficient ρ is set to 0.99,

and all other coefficients (α, β and ϕ) are 0. The experiment is run for a sample size $T = 60$, and nine block sizes $l = 6, 12, 18, \cdots, 54$, varying from one-tenth to nine-tenths of T. For each l, 10,000 samples are generated to calculate the bias and standard error of the MBJK estimator.

Figure 2.4.2 depicts the relation between the block size and the bias and standard error of the MBJK estimate of the slope coefficient. As l increases, the bias of the MBJK estimator decreases, but its variability increases. Whereas the bias diminishes at a decreasing rate, the variability increases at an accelerating rate. The bias-variance trade-off in the MBJK estimator controlled by l is obvious. Instead of disadvantaging the method, this trade-off brings it great flexibility in tackling various problems. For example, in short-horizon (*i.e.*, one period ahead) return predicting, the use of $l = 0.3T$, as revealed in the simulation results, has a satisfactory performance in bias reduction without a substantial increase in estimation variance [1]. In the cases where the magnitude of the finite-sample bias is severe, such as in long-horizon regressions, a large l can be used to achieve a good bias reduction at the cost of increasing variability. In summary, depending on the severity of the bias, the length l can be chosen in a discretionary sense in practice.

2.5 Comparison of Estimators

This section carries out simulations to systematically study the finite-sample performance of the OLS estimator and the three bias-reduction estimators for predictive regressions, that is, the BMJK estimator proposed in this chapter and the

---

[1]Similar simulation results are produced for $T = 120$ and are not reported here.

two existing ones in the literature, the plug-in estimator as well as the augmented regression estimator.

The predictive regression with one predictor variable is by far the most studied and commonly used in the literature and, hence, is the focus of the simulation studies. Results on bivariate regressions are also presented. For the MBJK estimator, the block size l is fixed at $0.3T$ in most of the simulation studies below, unless otherwise specified.

2.5.1 Single-factor predictive models

The model specified by (2.1) to (2.3) is used to generate data for the single-regressor case. The correlation between u_t and v_t, $\delta = \sigma_{uv}/(\sigma_u \sigma_v)$, takes three different values: -0.8, -0.9, and -0.95. This negative value assumption is without loss of generality because the sign of the β is unrestricted. The autoregressive coefficient ρ is set to either 0.95, 0.99, or 0.999. These values for δ and ρ are realistic according to Stambaugh (1999) and Campbell and Yogo (2006). The sample size, T, is equal to 60, 120, or 360. The innovation terms u_t and v_t are of unit variances. The true parameter values for α, β, and ϕ are all set to 0 in all simulations.

For each combination of the parameter values listed above, 10,000 samples are generated. From each set of generated returns and regressor values, four slope estimates by the four approaches — the OLS, the augmented regression, the plug-in method, and the MBJK method — are calculated. The average bias and the root mean squared error (RMSE) are then calculated across the 10,000 samples for each estimator.

$\delta = -0.8$ (bias with RMSE in parentheses)

Method	$T=60,\ \rho=0.95$	$T=60,\ \rho=0.99$	$T=60,\ \rho=0.999$	$T=120,\ \rho=0.95$	$T=120,\ \rho=0.99$	$T=120,\ \rho=0.999$	$T=360,\ \rho=0.95$	$T=360,\ \rho=0.99$	$T=360,\ \rho=0.999$
OLS	0.064 (0.098)	0.069 (0.095)	0.068 (0.094)	0.031 (0.052)	0.035 (0.051)	0.035 (0.049)	0.009 (0.022)	0.011 (0.017)	0.012 (0.017)
Augmented	0.013 (0.080)	0.017 (0.072)	0.016 (0.071)	0.005 (0.044)	0.009 (0.038)	0.009 (0.036)	0.001 (0.020)	0.002 (0.014)	0.003 (0.012)
Plug-in	0.015 (0.079)	0.019 (0.071)	0.018 (0.070)	0.005 (0.044)	0.010 (0.038)	0.009 (0.035)	0.001 (0.020)	0.002 (0.014)	0.003 (0.012)
MBJK	0.008 (0.085)	0.011 (0.078)	0.010 (0.076)	0.001 (0.047)	0.004 (0.041)	0.003 (0.039)	0.000 (0.021)	0.000 (0.015)	0.001 (0.013)

$\delta = -0.9$ (bias with RMSE in parentheses)

Method	$T=60,\ \rho=0.95$	$T=60,\ \rho=0.99$	$T=60,\ \rho=0.999$	$T=120,\ \rho=0.95$	$T=120,\ \rho=0.99$	$T=120,\ \rho=0.999$	$T=360,\ \rho=0.95$	$T=360,\ \rho=0.99$	$T=360,\ \rho=0.999$
OLS	0.071 (0.103)	0.077 (0.104)	0.078 (0.103)	0.034 (0.055)	0.039 (0.054)	0.040 (0.053)	0.010 (0.022)	0.013 (0.019)	0.013 (0.018)
Augmented	0.015 (0.081)	0.020 (0.075)	0.019 (0.074)	0.005 (0.045)	0.010 (0.040)	0.010 (0.038)	0.001 (0.020)	0.003 (0.014)	0.003 (0.013)
Plug-in	0.016 (0.081)	0.020 (0.074)	0.020 (0.073)	0.005 (0.044)	0.010 (0.040)	0.011 (0.038)	0.001 (0.020)	0.003 (0.014)	0.003 (0.013)
MBJK	0.009 (0.086)	0.011 (0.080)	0.011 (0.079)	0.000 (0.048)	0.004 (0.043)	0.004 (0.040)	-0.001 (0.020)	0.000 (0.015)	0.000 (0.014)

$\delta = -0.95$ (bias with RMSE in parentheses)

Method	$T=60,\ \rho=0.95$	$T=60,\ \rho=0.99$	$T=60,\ \rho=0.999$	$T=120,\ \rho=0.95$	$T=120,\ \rho=0.99$	$T=120,\ \rho=0.999$	$T=360,\ \rho=0.95$	$T=360,\ \rho=0.99$	$T=360,\ \rho=0.999$
OLS	0.076 (0.108)	0.082 (0.109)	0.082 (0.107)	0.036 (0.056)	0.042 (0.056)	0.042 (0.055)	0.011 (0.023)	0.013 (0.019)	0.014 (0.019)
Augmented	0.016 (0.084)	0.021 (0.078)	0.020 (0.076)	0.006 (0.045)	0.011 (0.040)	0.010 (0.038)	0.001 (0.020)	0.003 (0.014)	0.004 (0.013)
Plug-in	0.017 (0.083)	0.022 (0.077)	0.022 (0.075)	0.006 (0.045)	0.011 (0.039)	0.011 (0.038)	0.001 (0.020)	0.003 (0.014)	0.004 (0.013)
MBJK	0.010 (0.090)	0.012 (0.082)	0.013 (0.079)	0.001 (0.047)	0.005 (0.043)	0.004 (0.041)	0.000 (0.021)	0.000 (0.015)	0.001 (0.014)

Table 2.1 provides an overview of the finite-sample properties of the four estimating methods. For each combination of the parameter values, both bias and RMSE are reported, with the latter given in parentheses. Although finite-sample biases of the OLS estimator tend to be substantial, they become moderate or even negligible for all three bias-reduced estimators. The two parametric bias-reduced methods — the augmented method and the plug-in method — perform similarly in terms of both bias and RMSE. The non-parametric method, the jackknife procedure, is consistently and substantially better than its parametric counterparts in reducing bias, especially in the cases where $T = 60$ and $T = 120$. The MBJK method also reduces RMSE of the OLS estimator.

However, the MBJK method has slightly larger RMSEs than its parametric counterparts. The evaluation of the accuracy of an estimator of some parameter using RMSE, or equivalently mean squared error (MSE), is common in the literature. The MSE decomposes into a sum of squared bias and variance of the estimator (MSE = Bias^2 + Var), both quantities are important when evaluating an estimator. However, the MSE imposes an arbitrary judgment as to the relative importance of bias and variance (*e.g.*, Rosenberg and Guy, 1995). As pointed out by Simonoff (1993), a more useful way is to evaluate bias and variance based on their use which may vary in different applications. For example, bias can be much more critical than variance in pricing continuous time contingent claims such as bond options. This is because a small bias in the mean reversion parameter using least squares or maximum likelihood can translate into pricing biases which are economically too significant to ignore (Phillips and Yu, 2009; Yu, 2012). Rosenberg and Guy (1995) discuss prediction criteria for asset beta and conclude that bias is more a relevant criterion for stock selection purpose while valuing

convertible assets cares more about variance.

2.5.2 Robustness assessment

Given the noisy nature of financial data sets, robustness is an important advantage for any financial econometric model. We are particularly interested in assessing the impact of outliers, heteroscedasticity, and model misspecification on the bias and RMSE of the estimators.

First, we investigate the performance of the estimators in the presence of either outliers or heteroscedasticity, the two most common issues for financial data. For tractability, the parameter values of ρ and δ are fixed at 0.99 and -0.9, respectively.

The outlier scenario we consider is as follows. The return innovation u_t follows a standard normal $N(0, 1)$, which is contaminated by random shocks from $N(0, 4)$ distribution. In the simulations, different contamination rates are considered, namely, $\varphi = 1\%, 5\%, 10\%, 20\%$, and 30%. For each contamination rate, 10,000 samples are generated and the corresponding bias and RMSE are computed.

Two cases of heteroscedasticity are considered in the simulations. In the first case (Scenario 1), the conditional volatility of r_{t+1} given x_t, σ_{u_t}, changes over time. We let σ_{u_t} take three different values: 0.5 for the first one-third, 1 for the second one-third, and 1.5 for the last one-third. In the second case (Scenario 2), the value of σ_{u_t} changes through x_t. In our simulations, we use

$$\sigma_{u_t} = \max(0.5, (\min(0.4|x_t|, 1.5))).$$

That is, the volatility changes with x_t but is constrained within the interval

[0.5, 1.5]. Again, 10,000 samples are generated for each case to compute the corresponding bias and RMSE.

Second, we look into the issue of model misspecification. Financial time series show complex properties, and obtaining a good model to describe a series is generally a very challenging task. Up to now, we had assumed that the regressor is an AR(1) process. However, there is no compelling theoretical reason to believe it should always be the case. Indeed, De Santis (2007) used AR(2) to model the dynamics of the dividend yield, consumption growth, and dividend growth. Therefore, a model misspecification can occur when an AR(1) process is used to model a true underlying AR(2) process. We examine to what extent this type of model misspecification affects the bias and RMSE of the estimators. In the simulations, the correlation between two innovation processes δ takes the value -0.9. We generate the regressor samples using two AR(2) processes, $x_t = 0.5x_{t-1} + 0.4x_{t-2} + v_t$ (Scenario 1) and $x_t = 0.2x_{t-1} + 0.6x_{t-2} + v_t$ (Scenario 2), but use an AR(1) model to fit the data. We generate 10,000 samples for each scenario and compute the corresponding bias and RMSE.

Table 2.2 summarizes the results. For all four estimators, the finite-sample biases do not seem to increase with the contamination rate. However, the RMSE increases as the contamination rate increases. In both outlier and heteroscedasticity cases, the performance of the two parametric estimators is similar except that the plug-in estimator seems to have slightly larger biases in the small sample size ($T = 60$). But the difference is no longer distinct as the sample size increases. In these cases, all three bias-reduction methods remove the bias well, with the jackknife procedure delivering the least biased estimates but higher RMSE than the other two.

Table 2.2: Assessment of model robustness. The table reports the mean bias and root mean squared error (RMSE) of slope estimates from the four estimating approaches; the OLS, augmented regression, plug-in method, and MBJK. The RMSE is given in parentheses. Columns 2-6 list results in presence of outliers, with return innovations being contaminated by random shocks from $N(0,4)$ at various contamination rates. Columns 7-8 are for the heteroscedasticity scenarios as described in Section 2.5.2. Columns 9-10 are for the model misspecification scenarios as described in Section 2.5.2. The sample size, T, is equal to 60, 120, or 360. The correlation of the two innovation processes δ takes -0.9. All results are based on 10,000 simulations.

	Contamination rate					Heteroscedasticity		Misspecification	
	1%	5%	10%	20%	30%	Scenario1	Scenario2	Scenario1	Scenario2
					$T=60$				
OLS	0.078	0.077	0.077	0.077	0.076	0.081	0.070	0.079	0.056
	(0.108)	(0.116)	(0.127)	(0.147)	(0.162)	(0.104)	(0.100)	(0.122)	(0.111)
Augmented	0.021	0.019	0.019	0.018	0.018	0.016	0.010	0.038	0.035
	(0.082)	(0.093)	(0.108)	(0.131)	(0.150)	(0.081)	(0.076)	(0.108)	(0.110)
Plug-in	0.023	0.021	0.021	0.020	0.020	0.018	0.013	0.039	0.035
	(0.082)	(0.092)	(0.106)	(0.129)	(0.148)	(0.080)	(0.076)	(0.107)	(0.109)
MBJK	0.011	0.012	0.012	0.011	0.011	0.006	0.004	0.019	0.016
	(0.089)	(0.102)	(0.117)	(0.139)	(0.160)	(0.087)	(0.085)	(0.108)	(0.109)
					$T=120$				
OLS	0.040	0.040	0.041	0.038	0.041	0.043	0.036	0.040	0.029
	(0.056)	(0.061)	(0.068)	(0.076)	(0.087)	(0.059)	(0.053)	(0.072)	(0.073)
Augmented	0.011	0.010	0.012	0.009	0.011	0.010	0.007	0.019	0.018
	(0.042)	(0.049)	(0.056)	(0.067)	(0.079)	(0.045)	(0.040)	(0.065)	(0.072)
Plug-in	0.012	0.011	0.012	0.010	0.012	0.011	0.008	0.019	0.018
	(0.042)	(0.049)	(0.056)	(0.066)	(0.079)	(0.045)	(0.040)	(0.065)	(0.072)
MBJK	0.004	0.004	0.005	0.002	0.005	0.001	0.001	0.005	0.005
	(0.046)	(0.053)	(0.062)	(0.071)	(0.085)	(0.050)	(0.045)	(0.065)	(0.072)
					$T=360$				
OLS	0.012	0.013	0.013	0.013	0.012	0.015	0.012	0.013	0.010
	(0.019)	(0.022)	(0.025)	(0.028)	(0.032)	(0.023)	(0.019)	(0.032)	(0.038)
Augmented	0.002	0.003	0.003	0.003	0.003	0.003	0.002	0.006	0.006
	(0.015)	(0.018)	(0.022)	(0.026)	(0.030)	(0.018)	(0.015)	(0.032)	(0.038)
Plug-in	0.003	0.003	0.003	0.003	0.003	0.004	0.002	0.006	0.006
	(0.015)	(0.018)	(0.022)	(0.026)	(0.030)	(0.018)	(0.015)	(0.032)	(0.038)
MBJK	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000
	(0.016)	(0.020)	(0.023)	(0.027)	(0.031)	(0.019)	(0.016)	(0.031)	(0.038)

In the scenarios of model misspecification, the augmented and the plug-in methods perform much worse than in the other robust scenarios, reducing less biases. Especially in the case where the the second lag has a bigger impact on the current value of the regressor than the first lag (*i.e.*, Scenario 2), the two parametric estimators do not seem to significantly correct the OLS biases. Furthermore, the two methods also lose their advantage in RMSE and no longer produce the lowest RMSEs. The jackknife estimator, on the contrary, behaves quite well. It shows great robustness to the model misspecification, not only consistently reducing bias significantly but also producing the lowest estimation uncertainty.

2.5.3 Long-horizon predictive models

We study the effectiveness of the three bias-reduction methods in long-horizon regression as specified by (2.4) through simulations. For the jackknife procedure, apart from the MBJK estimator with $l = 0.3T$, the MBJK estimator with $l = 0.5T$ is also considered because of the large magnitude of the bias in this application.

To keep things tractable, the parameters ρ and δ are fixed at 0.99 and -0.9, respectively. The horizon p takes three different values of 3, 6, and 12, which captures the common applications of long-horizon forecasting. Although some of the scenarios, such as forecasting returns for the next 12 periods using 60 samples, are unlikely to be encountered in reality, they provide extreme conditions to test the model performance. We generate samples from the model specified by (2.1) to (2.3). In order to get T pairs of (r_{t+p}, x_t), $T+p-1$ samples are generated, and

the moving sums of length p are calculated and re-matched with the predictor variable.

Table 2.3 presents the results based on 10,000 repetitions for each scenario. One of the most striking features is that the finite-sample biases are much more severe compared with short-horizon regressions, *i.e.*, $p = 1$. Even when $T = 360$, the bias of the OLS when $p = 12$ is more than ten times as large as that in the short-horizon forecasting. The bias decreases with the sample size but increases with the horizon once the sample size is fixed. As pointed out by Valkanov (2003), long-horizon regressions produce inconsistent estimates and tend to give "significant" results, regardless of whether there is a structural relation between the underlying variables. Again, the simulation results highlight the danger of interpreting the OLS results naively in long-horizon regressions.

Again, the augmented method and the plug-in method perform similarly. However, to our disappointment, they do not reduce much of the bias. On the contrary, the MBJK method does a much better job in reducing bias, especially with $l = 0.5T$. Although the RMSE of the MBJK estimator with $l = 0.5T$ is larger than that with $l = 0.3T$, it is still smaller than that in the other three methods. To make the comparison more transparent, Table 2.4 lists the percentage reduction over the OLS bias for each of the three bias-reduction methods. The augmented method and the plug-in method can reduce only 4% to 26% of the OLS bias across all combinations. The MBJK estimator slashes the bias from 25% up to 92% with $l = 0.3T$ and from 39% to 97% with $l = 0.5T$. If excluding the extreme scenarios that are unrealistic in practice, such as the combinations $(T = 60, p = 6)$, $(T = 60, p = 12)$, and $(T = 120, p = 12)$, the MBJK estimator with $l = 0.5T$ does a decent job reducing 76% to 97% of the bias.

Table 2.3: Finite-sample performance in long-horizon regressions. The table reports the mean bias and root mean squared error (RMSE) of slope estimates from the four estimating approaches; the OLS, augmented regression, plug-in method, and MBJK. The RMSE is in parentheses. The MBJK estimators with two different block sizes, $0.3T$ and $0.5T$, are considered. The horizon p takes three different values; 3, 6, and 12. The sample size, T, is equal to 60, 120, or 360. The parameter values of ρ and δ are fixed at 0.99 and -0.9, respectively. All results are based on 10,000 simulations.

| | | OLS | Augmented | Plug-in | MBJK | |
					$l = 0.3T$	$l = 0.5T$
$T = 60$	$p = 3$	0.219	0.163	0.164	0.079	0.053
		(0.276)	(0.248)	(0.246)	(0.219)	(0.241)
	$p = 6$	0.370	0.330	0.328	0.206	0.149
		(0.467)	(0.446)	(0.442)	(0.400)	(0.420)
	$p = 12$	0.599	0.574	0.570	0.451	0.365
		(0.726)	(0.716)	(0.711)	(0.692)	(0.700)
$T = 120$	$p = 3$	0.115	0.088	0.087	0.027	0.018
		(0.153)	(0.137)	(0.136)	(0.122)	(0.135)
	$p = 6$	0.214	0.189	0.188	0.075	0.050
		(0.278)	(0.263)	(0.262)	(0.221)	(0.250)
	$p = 12$	0.368	0.348	0.347	0.203	0.146
		(0.470)	(0.460)	(0.459)	(0.401)	(0.456)
$T = 360$	$p = 3$	0.038	0.028	0.028	0.003	0.001
		(0.055)	(0.050)	(0.050)	(0.046)	(0.051)
	$p = 6$	0.076	0.065	0.065	0.008	0.005
		(0.107)	(0.102)	(0.101)	(0.089)	(0.101)
	$p = 12$	0.138	0.130	0.129	0.031	0.018
		(0.199)	(0.194)	(0.193)	(0.168)	(0.191)

Table 2.4: Bias reduction in long-horizon regressions. The table lists the percentage reduction over the OLS bias for each of the three bias-reduction methods considered. Results are computed based on Table 2.3.

		Augmented	Plug-in	MBJK $l = 0.3T$	$l = 0.5T$
	$p = 3$	26%	25%	64%	76%
$T = 60$	$p = 6$	11%	11%	44%	60%
	$p = 12$	4%	5%	25%	39%
	$p = 3$	23%	24%	77%	84%
$T = 120$	$p = 6$	12%	12%	65%	77%
	$p = 12$	5%	6%	45%	60%
	$p = 3$	26%	26%	92%	97%
$T = 360$	$p = 6$	14%	14%	89%	93%
	$p = 12$	6%	7%	78%	87%

2.5.4 Multi-factor predictive models

We now consider bias reduction in multi-factor predictive models. Although less studied than single-factor regressions, multi-factor predictive regressions are vastly popular among practitioners. There is no plug-in version available in multi-factor regressions due to the unavailability of explicit bias forms in this setup.

We consider a two-factor predictive model specified as

$$r_t = \alpha + \beta_1 x_{t-1} + \beta_2 z_{t-1} + u_t,$$

$$x_t = \phi_1 + \rho_1 x_{t-1} + v_{1t}, \tag{2.13}$$

$$z_t = \phi_2 + \rho_2 z_{t-1} + v_{2t}.$$

The trivariate error terms (u_t, v_{1t}, v_{2t}) follow a joint normal distribution with

mean 0 and a covariance matrix

$$\Sigma = \begin{pmatrix} 1 & \delta_1 & \delta_2 \\ \delta_1 & 1 & \psi \\ \delta_2 & \psi & 1 \end{pmatrix}. \tag{2.14}$$

In the simulations, the parameters α, β_1, β_2, ϕ_1 and ϕ_2 all take the value 0. We consider three representative cases. In the first case, we set $(\rho_1, \rho_2) = (0.99, 0.99)$ and $(\delta_1, \delta_2, \psi) = (-0.9, -0.9, 0.8)$. That is, two regressors are highly persistent, highly endogenous, and highly correlated. This setup corresponds to the case of using two financial ratios, such as the dividend yield and the earnings ratio, in a bivariate regression. Apparently, collinearity is an issue here. Although not favored by academics, this type of regression is very common among practitioners.

In the second case, we let $(\rho_1, \rho_2) = (0.99, 0.99)$ and $(\delta_1, \delta_2, \psi) = (-0.9, 0, 0.4)$. That is, the two regressors are highly persistent, but the first one is highly endogenous, whereas the second one is exogenous, and the two are of moderate correlation. This setup corresponds to a regression favoured by Ang and Bekaert (2007), using the dividend yield and the short rate as predictor variables. They argue that the predictability of the dividend yield is considerably enhanced when jointly used with the short rate.

In the third case, we assume $(\rho_1, \rho_2) = (0.99, 0.1)$ and $(\delta_1, \delta_2, \psi) = (-0.9, 0.4, 0)$. That is, the first regressor is highly persistent and is of high negative correlation with the returns, whereas the second regressor is not very persistent and is of positive correlation with the returns. In addition, the two regressors are independent. This setup corresponds to the case of predicting individual stock returns using the dividend yield of the stock and a lagged market return.

Table 2.5: Finite-sample performance on multivariate regressions. The table reports the mean bias and root mean squared error (RMSE) of slope coefficients from the three estimating approaches; the OLS, augmented regression, and MBJK. The RMSE is in parentheses. Samples are generated from the models (2.13) and (2.14). The top row specifies the correlations between the innovations to the two regressors and those to the returns. The second row gives the correlation of the two innovation processes of the two regressors. The third row is the values of the autoregressive roots ρ_1 and ρ_2. Columns 2-3, 4-5, and 6-7 are for the three cases discussed in Section 2.5.4, respectively. The sample size, T, is equal to 60, 120, or 360. All results are based on 10,000 samples.

	$\delta_1 = -0.9, \delta_2 = -0.9$ $\psi = 0.8$ $\rho_1 = 0.99, \rho_2 = 0.99$		$\delta_1 = -0.9, \delta_2 = 0$ $\psi = 0.4$ $\rho_1 = 0.99, \rho_2 = 0.99$		$\delta_1 = -0.9, \delta_2 = 0.4$ $\psi = 0$ $\rho_1 = 0.99, \rho_2 = 0.1$	
	β_1	β_2	β_1	β_2	β_1	β_2
			$T = 60$			
OLS	0.106	0.005	-0.213	-0.236	-0.038	-0.024
	(0.218)	(0.192)	(0.286)	(0.302)	(0.075)	(0.139)
Augmented	0.022	0.003	-0.044	-0.047	-0.046	-0.005
	(0.107)	(0.091)	(0.166)	(0.180)	(0.169)	(0.262)
MBJK	0.015	0.000	-0.024	-0.028	0.000	-0.001
	(0.207)	(0.198)	(0.220)	(0.221)	(0.075)	(0.140)
			$T = 120$			
OLS	0.054	0.003	-0.108	-0.120	-0.019	-0.014
	(0.114)	(0.101)	(0.140)	(0.149)	(0.039)	(0.095)
Augmented	0.011	0.001	-0.022	-0.025	-0.026	-0.002
	(0.053)	(0.036)	(0.087)	(0.095)	(0.088)	(0.178)
MBJK	0.004	0.000	-0.002	-0.004	-0.001	0.002
	(0.108)	(0.107)	(0.113)	(0.114)	(0.038)	(0.096)
			$T = 360$			
OLS	0.017	0.001	-0.034	-0.038	-0.006	-0.005
	(0.040)	(0.038)	(0.050)	(0.053)	(0.014)	(0.054)
Augmented	0.003	0.000	-0.006	-0.006	-0.007	0.000
	(0.019)	(0.013)	(0.033)	(0.033)	(0.032)	(0.101)
MBJK	0.000	0.000	0.000	0.001	0.000	0.000
	(0.039)	(0.037)	(0.042)	(0.043)	(0.014)	(0.054)

For each case, 10,000 samples are generated to compute the bias and RMSE. The results are reported in Table 2.5, with columns 2-3 for the first case, columns 4-5 for the second case, and columns 6-7 for the third case. Interestingly, when there is high collinearity (*i.e.*, the first case), the OLS estimator has significant finite-sample biases for the first regressor but negligible biases for the second regressor. The performance of the augmented approach is mixed in the three cases considered. It performs well in the first case, eliminating biases significantly and delivering very low RMSEs. However, it performs badly in the third case. In this case, compared with the OLS estimator, the augmented approach not only increases the biases of the first regressor but also doubles the estimating uncertainty for both regressors. In sharp contrast, the jackknife procedure is stable across all three scenarios, consistently delivering the least biased estimates with RMSEs of the same scale as, or much less than those of the OLS estimator.

2.5.5 Comparison summary

We conclude this section with a summary of the simulation findings. The simulations reveal the substantial finite-sample biases of the OLS estimator. This bias issue is especially severe in long-horizon regressions as the magnitude of the bias increases rapidly with the horizon. This supports the claim by Nelson and Kim (1993) that "the estimated biases are large enough to affect inference in practice, and should be accounted for when studying predictability". The simulations also show that the proposed jackknife estimator offers substantial improvements over the OLS estimator and enables reductions in both bias and RMSE.

Of the three bias-reduction approaches considered, the MBJK method almost

always produces the least biased estimates. Furthermore, the MBJK method possesses several statistical properties which distinguish it to its parametric counterparts.

First and foremost, the MBJK method is the only method which can reduce bias in long-horizon regressions. This is a major advantage over its alternatives. Because of the complex bias mechanism in long-horizon regressions, it is difficult to tackle the problem of biased estimation and so far there are no exiting remedies in the literature for this application. Indeed, the augmented and the plug-in approaches cease to work in this application. However, the proposed jackknife procedure can provide an effective solution to address the estimation problem in these cases (Table 2.3).

Second, as a non-parametric method, the MBJK method possesses great robustness. Both the augmented regression and the plug-in method require certain parametric assumptions to be able to provide satisfactory performance. As demonstrated by the simulations, the MBJK method works well in the model misspecification cases and outperforms the augmented regression and the plug-in method. In multi-regressor regressions, the plug-in method does not apply, while the augmented regression can produce results even worse than the OLS estimator (in terms of both bias and RMSE) in certain cases. The MBJK method, however, is stable and consistent. This robustness is an attractive feature and offers advantages over the standard methods given the noise nature of financial data sets.

Third, the MBJK method is simple and computationally easy to implement. It involves only a linear combination of a series of the OLS estimators using various subsets of the data. This computational advantage is more evident in

multi-regressor models as the augmented regression requires a complex iterative estimation procedure to obtain the solution.

Although the MBJK method is more superior in reducing bias, we also notice that it often has slightly larger RMSEs than its two counterparts except for the long-horizon regressions. However, this drawback does not undermine the value of adopting the MBJK approach given its various advantages listed above. Furthermore, as mentioned in Section 2.5.1, the relevance of RMSE can be critically dependent on the application context. Maybe a more appropriate way to evaluate an estimator is to incorporate a relative importance of bias and variance into the MSE criterion through the weighting as suggested by Lin and Tu (1995).

2.6 Empirical Illustration

We illustrate the proposed jackknifing procedure using some common predictive models in finance, namely, predicting the equity premiums by either the lagged dividend yield, or the short rate, or both. The dividend yield and the short rate receive great attention in the stock return prediction literature. For example, Lewellen (2004) reported strong evidence for predictive power of the dividend yield, whereas Campbell and Yogo (2006) found evidence that the short rate predicts returns. Ang and Bekaert (2007) also found the short rate robust in predicting returns. Furthermore, they argued that the dividend yield works better together with the short rate. We illustrate the proposed method on these predictive models and highlight the differences between the OLS estimates and the bias-adjusted estimates.

2.6.1 Data

As pointed out by Ang and Bekaert (2007), interest rate data are hard to interpret before the 1951 Treasury Accord, as the Federal Reserve pegged interest rates during the 1930s and the 1940s. Second, a number of studies have identified parameter instability during the 1990s. For example, Paye and Timmermann (2006) identified a significant structural break in the coefficient of the dividend yield around the 1990s. Goyal and Welch (2003) found that predictability by the dividend yield is not robust with the inclusion of the 1990s. Ang and Bekaert (2007) documented the coefficient for the dividend yield is twice as large if estimated from a sample that excludes the 1990s than if it was estimated from a sample inclusive of the 1990s. Hence, we focus on the post-Accord period, starting from January, 1952 up to December, 1989 for the analysis.

The data used in this section are the monthly return series of the S&P 500 Index. We briefly introduce the data and their sources below [1].

Stock returns (R_t): Monthly S&P 500 index returns from 1952 to 1989 are from the Center for Research in Security Prices (CRSP). They are continuously compounded returns on the index, including dividends.

Risk-free rate r_f: The risk-free rate from 1952 to 1989 is the T-bill rate.

Dividend yield dy: Dividends are 12-month moving sums of dividends paid on the S&P 500 Index. The original data are from Robert Shiller's website [2]. The dividend yield is the dividends divided by the index level.

Short-term rate tbl: The short rate is the secondary market rates of 3-month T-bills from the economic research database at the Federal Reserve Bank at St.

---

[1] For a detailed data description, refer to Welch and Goyal (2008)

[2] http://aida.econ.yale.edu/ shiller/data.htm.

Table 2.6: Summary statistics, 01/1952 – 12/1989 (456 months). The table reports summary statistics of the equity premiums, dividend yields (dy), and short rates (tbl). The column Kurt reports excess kurtosis. The last column CI is the 95% confidence interval for the estimated autoregressive coefficient $\hat{\rho}$.

	Mean	Median	Std Dev	Skew	Kurt	$\hat{\rho}$	CI
r	5.66%	7.09%	14.54%	-0.58	2.96	0.06	(-0.04, 0.15)
dy	3.92%	3.65%	0.92%	0.74	-0.57	0.98	(0.97, 1.00)
tbl	5.48%	5.08%	3.11%	0.93	0.82	0.99	(0.97, 1.00)

Louis [1].

Following the usual convention, the equity premium is computed as

$$r_t = \log(1 + R_t) - \log(1 + r_f).$$

Summary statistics of the series are presented in Table 2.6. The return numbers reported in the table are annualised. The returns are the most variable, whereas the volatility of the dividend yield is the lowest. The return series is also characterised by fat tails as evidenced by the large excess kurtosis. There is no strong evidence of autocorrelation in the equity premiums as indicated by the insignificant $\hat{\rho}$. On the other hand, the instruments, both the dividends and short rates exhibit high persistence, with the 95% confidence interval of the autoregressive coefficient being [0.97, 1].

2.6.2 Empirical results

We fit the two univariate regressions and the bivariate regression by the OLS and the jackknife procedure. Both short-horizon (at the one-month horizon) and

---

[1]http://research.stlouisfed.org/fred2/categories/22

long-horizon predictability (horizons of 3, 6, and 12 months) are examined. To reduce the bias, the MBJK estimator with the block length $l = 0.3T$ is used for the short-horizon regressions, whereas the MBJK estimator with $l = 0.5T$ is used for the long-horizon regressions. Table 2.7 overviews the regression results. For each horizon, it lists the OLS estimates and the MBJK estimates of the predictor variable coefficients. The t-statistics are computed using Newey-West (Newey and West, 1987) standard errors with $p + 1$ lags.

The OLS results reveal that (i) the OLS coefficient estimates are proportional to the horizon and (ii) the predictive ability of the dividend yield is considerably enhanced when coupled with the short rate in the regression. These results are consistent with the findings by Boudoukh et $al.$ (2008) and Ang and Bekaert (2007). The OLS results also convey a well-celebrated message – strong return predictability. Except for an insignificant short rate coefficient at $p = 12$ in a univariate regression, all other coefficient estimates are significant at the 5% level, with many of them being significant even at the 1% level. The predictability is especially pronounced for the bivariate regression in long horizons. This strong statistical evidence of predictability, however, vanishes completely after removing finite-sample biases, as shown in the last three columns of Table 2.7. It indicates that the finite-sample bias explains the bulk of apparent predictability. These empirical results cast doubt on the conclusions drawn in earlier studies regarding the predictive power of the dividend yield and the short rate.

Table 2.7: Regression results for the period 01/1952 – 12/1989 (456 months). The univariate regressions regress the equity premiums on the dividend yields (dy) or short rates (tbl). The bivariate regression uses two regressors, dy and tbl. For the results, 'OLS' reports the standard OLS estimates, and 'MBJK' reports the bias-adjusted estimates. The MBJK estimator with $l = 0.3T$ is used for the short-horizon regressions ($p = 1$), whereas the MBJK estimator with $l = 0.5T$ is used for the long-horizon regressions ($p > 1$). The t-statistics are computed using Newey-West standard errors.

		OLS			MBJK		
		Coefficient	t-stat	p-value	Coefficient	t-stat	p-value
		Univariate Regression					
	dy	0.489	2.251	0.025	0.244	1.125	0.261
	tbl	-0.164	-2.842	0.005	-0.025	-0.436	0.663
$p = 1$		Bivariate Regression					
	dy	0.739	3.162	0.002	0.314	1.345	0.179
	tbl	-0.233	-3.771	0.000	-0.028	-0.447	0.657
		Univariate Regression					
	dy	1.524	2.602	0.010	0.785	1.341	0.181
	tbl	-0.419	-2.628	0.009	-0.040	-0.251	0.802
$p = 3$		Bivariate Regression					
	dy	2.193	3.381	0.001	0.235	0.362	0.718
	tbl	-0.626	-3.707	0.000	0.208	1.230	0.219
		Univariate Regression					
	dy	3.295	3.000	0.003	1.390	1.265	0.206
	tbl	-0.706	-1.988	0.047	-0.119	-0.334	0.738
$p = 6$		Bivariate Regression					
	dy	4.504	3.504	0.001	0.480	0.373	0.709
	tbl	-1.130	-3.113	0.002	0.300	0.827	0.409
		Univariate Regression					
	dy	6.628	3.302	0.001	3.165	1.577	0.115
	tbl	-1.066	-1.523	0.128	-0.360	-0.515	0.607
$p = 12$		Bivariate Regression					
	dy	8.638	3.456	0.001	2.155	0.862	0.389
	tbl	-1.879	-2.675	0.008	0.051	0.072	0.943

2.7 Summary

In the class of predictive regressions studied in this Chapter, a rate of return is regressed on a lagged stochastic regressor, which is autoregressive with errors that are correlated with the errors of the regression model. The OLS estimator exhibits the finite-sample bias, which potentially leads to an incorrect conclusion that the lagged variable has predictive power while in fact it does not. This chapter provides a non-parametric method based on the jackknife technique to reduce estimation bias. Simulations show that the method is highly effective in a broad range of model specifications. It reduces the bias for both single- and multiple-regressor models and for both short- and long-horizon regressions. It offers substantial improvements over the OLS estimator and enables reductions in both bias and mean square error, so the gains from bias reduction are not lost in variance increases. Compared with the other available counterparts in the literature, the proposed method is more general and stable. It is particularly useful in long-horizon regressions for which the alternative bias-reduction methods do not work. It also performs well in the situations with outliers, heteroscedasticity, and model misspecification.

The usefulness of the method is also illustrated in the empirical estimates of the common predictive models in finance which examine the predictive power of the dividend yield and the short rate. The significant predictive variables under the OLS become insignificant after adjusting for the bias in both univariate and bivariate regressions, for both short and long horizons. These discrepancies are large and suggest that bias reduction in predictive regressions is important in practical applications.

Chapter 3

Quantile Regression for Return Prediction

3.1 Modelling the Return Distribution

3.1.1 Quantile models

The predictive regression (2.1) can also be written as

$$r_t = \alpha + \beta x_{t-1} + \sigma_u \epsilon_t, \tag{3.1}$$

where σ_u is a conditional volatility and ϵ_t is a return innovation which follows a standard normal distribution. To emphasise the time-series aspect of the data, we allow σ_u to be time varying and denote it by σ_t hereafter.

Let $\mathcal{F}_{t-1}$ be the information available at time $t-1$ and Φ be the standard normal distribution. Then the implied τ-th conditional quantile of r_t by model

(3.1) is

$$Q_\tau(r_t|\mathcal{F}_{t-1}) = \alpha + \beta x_{t-1} + \sigma_t \Phi^{-1}(\tau) \equiv \alpha_\tau + \beta x_{t-1}, \ \tau \in (0,1),$$

where $\Phi^{-1}(\tau)$ is the τ-th quantile of the standard normal distribution. Across the distribution of r_t, the only parameter varying with τ is the location α_τ which is determined solely by a mean effect α and the conditional volatility σ_t. Therefore, despite the linear regression model (3.1) being designed only to capture the conditional mean effect, in an ideal Gaussian world it provides a complete view of the future return.

In real life, however, a single mean curve and the associated conditional volatility are rarely adequate summaries of the relationship between returns and covariates. As reviewed in Section 1.1.2, the asset returns are commonly observed to exhibit non-Gaussian features, and the investors' interest goes well beyond mean and variance. By replacing the Gaussian distribution assumption with a general distribution F_ϵ for the return innovation ϵ_t, even in the simplest case in which the covariate effect is constant across all quantiles, the τ-th conditional quantile of r_t,

$$Q_\tau(r_t|\mathcal{F}_{t-1}) = \alpha + \sigma_t F_\epsilon^{-1}(\tau) + \beta x_{t-1},$$

is no longer determined solely by the mean effect and σ_t. Instead it involves estimation of the distribution F_ϵ.

Furthermore, there is no compelling theoretical reason to believe that β should be constant across quantiles. If some of the slope coefficients change with the quantile τ, then this is indicative of some form of heteroscedasticity. This can occur when economic state variables not only affect the conditional mean, but are

also linked to the conditional variance (as is often the case in finance applications). An example is given below.

$$r_t = \alpha + \beta x_{t-1} + \sigma_t \epsilon_t,$$

$$\sigma_t = \gamma_0 + \gamma_1 x_{t-1},$$

where γ_1 measures the effect of x_{t-1} on the volatility σ_t. This specification implies that the τ-th quantile takes the following form,

$$Q_\tau(r_t | \mathcal{F}_{t-1}) = \alpha + \gamma_0 F_\epsilon^{-1}(\tau) + (\beta + \gamma_1 F_\epsilon^{-1}(\tau)) x_{t-1} \equiv \alpha_\tau + \beta_\tau x_{t-1}.$$

Therefore, if a variable correlates positively with the volatility ($\gamma_1 > 0$), the slope coefficient increases as τ increases from 0 to 1. A reverse pattern should arise for variables correlating negatively with the volatility ($\gamma_1 < 0$).

The model

$$Q_\tau(r_t) = \alpha_\tau + \beta_\tau x_{t-1} + u_t, \ \tau \in (0,1) \tag{3.2}$$

is the so-called quantile regression introduced by Koenker and Bassett Jr (1978, 1982). A comprehensive introduction to quantile regression is presented in Koenker (2005). Proposed as an alternative to the ordinary least squares approach, quantile regression enjoys certain advantages over traditional approaches. It does not need to specify an error distribution, which can be difficult in some cases. Compared with the predictive regression model (3.1), the quantile regression (3.2) is less stringent and general, with the only assumption on u_t being $F_u(0) = \tau$. It provides a "distributional" perspective rather than merely a conditional mean view. By varying τ from 0 to 1, a complete picture of covariate effect on return

distribution is obtained.

Quantile regression has not been employed in finance until quite recently. Most applications of quantile regression in finance are on value at risk (VaR) modelling (see, for example, Taylor, 1999; Chernozhukov and Umantsev, 2001; Engle and Manganelli, 2004; Giacomini and White, 2006; Adrian and Brunnermeier, 2011). Using quantile regression to explore return predictability is a relatively new approach in the return prediction literature. Section 1.1.2 provides a review of the literature in this area. Strong empirical evidence of distribution predictability of the monthly returns of both the S&P 500 index and the US 5-year Treasury bonds is documented in the literature.

3.1.2 Estimation

For convenience, we denote $\boldsymbol{\beta}_\tau = (\alpha_\tau, \beta_{1,\tau})^T$, the regression parameter. For the return observation r_t, $X_{t-1} = (1, x_{t-1})^T$ is the associated covariate vector. The quantile regression model (3.2) can be written as

$$Q_\tau(r_t) = X_{t-1}^T \boldsymbol{\beta}_\tau + \epsilon_t.$$

Following the seminal work of Koenker and Bassett Jr (1978, 1982), the estimator of the parameter $\boldsymbol{\beta}_\tau$, $\hat{\boldsymbol{\beta}}_\tau$ is obtained by minimizing the following loss function,

$$L(\boldsymbol{\beta}_\tau) = T^{-1} \sum_{t=1}^{T} \rho(r_t - Q_\tau(r_t | \boldsymbol{\beta}_\tau)), \tag{3.3}$$

where

$$\rho(u) = \begin{cases} (1-\tau)|u| & \text{if } u \le 0 \\ \tau|u| & \text{if } u > 0 \end{cases}, \tag{3.4}$$

is the so-called tick loss function. The parameters are estimated using linear programming as proposed by Koenker and d'Orey (1987).

Parameter estimation uncertainty is required in order to perform statistical inference, such as in-sample predictability tests. One way to estimate the standard errors of $\hat{\boldsymbol{\beta}}_\tau$ is to use the score function for $\boldsymbol{\beta}_\tau$ and the delta method (see, for example, White, 1982; Koenker, 2005).

The estimating function or the score function for $\boldsymbol{\beta}_\tau$ is $U(\boldsymbol{\beta}_\tau) = L'(\boldsymbol{\beta}_\tau)$, the derivative of the loss function (3.3). As the derivative of the tick loss function $\rho(u)$ is $\psi(u) = \tau - \text{I}(u < 0)$ for any $u \ne 0$, the score function for $\boldsymbol{\beta}_\tau$ is

$$U(\boldsymbol{\beta}_\tau) = T^{-1} \sum_{t=1}^{T} X_{t-1}\{\tau - \text{I}(r_t - Q_\tau(r_t|\boldsymbol{\beta}_\tau))\}.$$

Suppose that $f_t(\cdot)$ is the density function of ϵ_t and $f_t(0) > 0$. In this case, we have

$$B = \text{cov}\{U(\boldsymbol{\beta}_\tau)\} = \tau(1-\tau)T^{-2} \sum_{t=1}^{T} X_{t-1}X_{t-1}^T, \tag{3.5}$$

and

$$A = \partial\text{E}\{U(\boldsymbol{\beta}_\tau)\}/\partial\boldsymbol{\beta} = T^{-1} \sum_{t=1}^{T} X_{t-1}X_{t-1}^T f_t(0). \tag{3.6}$$

The asymptotic covariance of $\hat{\boldsymbol{\beta}}_\tau$ by the delta method is given by

$$\Lambda = A^{-1}B(A^{-1})^T.$$

In order to obtain Λ, we need to estimate both A and B. However, it is not easy to obtain an estimator for A as it involves the density functions of ϵ_t. If U is smooth, we can estimate A by $\hat{A} = \partial U(\boldsymbol{\beta}_\tau)/\partial \boldsymbol{\beta}_\tau$ evaluated at $\hat{\boldsymbol{\beta}}_\tau$. If U is not smooth (as in quantile regression) and, as such, a derivative does not exist at certain points, the evaluation of A will depend on the unknown underlying density function. This makes it difficult to obtain an estimate for the covariance matrix of $\hat{\boldsymbol{\beta}}_\tau$.

The "unsmoothness" function is, in general, a "curse" in statistical inference. This may partially explain why the quantile approach is not so widely used. For the purpose of statistical inference, we use a method introduced by Wang *et al.* (2009) for obtaining the standard errors of the regression coefficients. This method eliminates unsmoothness in quantile estimation functions, and is more robust and less computationally intensive than the widely used bootstrap methods. The details of the method are given in Appendix B.

3.2   Data Description

We examine predictability of the entire stock and bond return distribution through the use of quantile regression (3.2). This section presents the data sets for this purpose. Two broad-based indices are chosen to represent stock and bond returns. They are the Russell 1000 Index, which is constructed and maintained by Russell Investments; and the US Aggregate Bond Index, which has been constructed by the now-defunct Lehman Brothers and is currently maintained by Barclays Capital. Compared to the commonly used stock index, the S&P 500 Index, the Russell 1000 Index offers a more comprehensive representation of the US stock

market, while the US Aggregate Bond Index is a well-recognized barometer for investment-grade bonds being traded in the US. The details of the data sets are listed below.

3.2.1 Stock data

Monthly stock returns are the simple returns on the Russell 1000 Index, including dividends, from 1979:01 to 2011:02, where the starting date is dictated by data availability. Eleven economic state variables are considered as potential candidates to predict the return distribution. These variables fall into two broad categories.

- Index characteristic variables

  - Cross-sectional volatility (cv), 1996:07 to 2011:02. It measures the cross-sectional return dispersion of the components in the Russell 1000 Index.

  - Dividend yield (dy), available at quarterly frequency from 1979:03 to 1986:12 and monthly frequency from 1987:01 to 2011:02. It is calculated as the 12-month moving sum of dividends paid on the Russell 1000 Index divided by the index level.

  - Price-to-book ratio (p/b), 1986:12 to 2011:02. It is calculated as the cap-weighted sum of the index components' price-to-book ratios.

  - Price-to-earnings ratio (p/e), 1986:12 to 2011:02. It is calculated as the cap-weighted sum of the index components' price-to-earnings ratios. Negative earnings are excluded from the calculation.

- Earnings-per-share growth forecast (*epsgf*), 1986:12 to 2011:02. It is calculated as the cap-weighted sum of the index components' I/B/E/S consensus earnings-per-share growth rates in the long term (typically five years).

- Broad market variables

  - Market volatility (*vix*), 1990:01 to 2011:02. It measures the market's expectation of stock market volatility over the next 30-day period. The index is calculated and disseminated by the Chicago Board Options Exchange.

  - Three-month T-Bill rate (*tbl3m*), 1979:01 to 2011:02. It is from the economic research database at the Federal Reserve Bank at St. Louis (FRED). The 3-month T-Bill secondary market rate serves as a proxy for expectations of future economic activity.

  - Inflation (*infl*), 1979:01 to 2011:02. It is the last 12-month rate change on the Consumer Price Index. The Consumer Price Index (All Urban Consumers) is from the Bureau of Labor Statistics. Because inflation is released in the following month, there is a one-month lag before using it in the monthly regressions.

  - Default yield spread (*dfy*), 1979:01 to 2011:02. It is the difference between BAA- and AAA-rated corporate bond yields. The corporate bond yields are from FRED. The default yield spread captures the effect of default premium, which tracks the long-term business cycle conditions, higher during recessions and lower during expansions.

– Term spread (*tms*), 1979:01 to 2011:02. It is approximated by the difference between the yields on 10-year Treasuries and 3-month Treasuries. The yields on 10-year Treasuries again are from FRED.

– Consumer sentiment index (*cs*), 1979:01 to 2011:02. It is from Datastream. The index is constructed by the University of Michigan.

3.2.2   Bond data

The monthly bond returns are the simple returns on the US Aggregate Bond Index from 1976:01 to 2011:02, where the starting date is dictated by data availability. As for the predictor variables, five of the broad market variables described above, *tbl3m, infl, dfy, tms* and *cs*, are also used to predict bond returns. To match with the bond data, the time series of the these state variables are from 1976:01 to 2011:02 for *tbl3m, infl, dfy* and *tms*, and from 1978:01 to 2011:02 for *cs* due to data availability. Researchers have also identified exchange rate change as a risk factor in bond returns (see, for example, Chow *et al.*, 1997). Following their insights, a trade-weighted exchange rate is also considered for predicting the bond returns.

• Trade-weighted exchange index (*twex*), 1976:01 to 2011:02. It is from FRED. This index is a weighted average of the price of various currencies relative to the dollar, which accurately reflects the strength of the dollar relative to other world currencies.

Table 3.1 reports descriptive statistics for the stock and bond returns as well as the predictor variables. The stock and bond returns possess typical features

Table 3.1: This table reports summary statistics for the stock and bond data sets. In the table, kurtosis represents excess kurtosis, JB stands for the Jarque-Bera statistic, p-val is the p-value of the Jarque-Bera statistic and N is the number of observations. The top panel reports the summary statistics for the stock data, while the bottom panel for the bond data. R_s and R_b are the monthly returns of the stock and bond indices, respectively.

	median	mean	std. err	skewness	kurtosis	min	max	JB	p-val	N
				Panel A: Stock 1979:01 – 2011:02						
R_s	0.014	0.010	0.045	-0.714	2.138	-0.217	0.129	108.398	0	386
cv	6.999	7.844	2.667	1.223	1.072	4.582	18.060	53.808	0	176
dy	0.020	0.022	0.008	0.488	-0.900	0.010	0.041	21.194	0	291
p/b	2.706	2.834	0.882	1.071	0.480	1.576	5.430	59.261	0	291
p/e	17.360	17.860	4.017	0.657	0.069	10.030	28.500	21.247	0	291
epsgf	0.120	0.126	0.019	1.558	2.378	0.095	0.197	189.667	0	291
vix	19.490	20.390	7.844	1.564	3.928	10.420	59.890	272.638	0	254
tbl3m	0.050	0.054	0.034	0.686	0.443	0	0.163	33.907	0	386
infl	0.031	0.039	0.029	1.868	3.593	-0.021	0.148	437.809	0	386
dfy	0.010	0.011	0.005	1.704	3.378	0.006	0.034	375.426	0	386
tms	0.019	0.018	0.013	-0.557	-0.203	-0.026	0.044	20.672	0	386
cs	90.250	86.280	13.206	-0.448	-0.597	51.700	112.000	18.506	0	386
				Panel B: Bond 1976:01 – 2011:02						
R_b	0.007	0.007	0.016	0.664	6.303	-0.061	0.113	739.571	0	422
tbl3m	0.051	0.054	0.033	0.681	0.653	0	0.163	40.799	0	422
infl	0.033	0.041	0.029	1.594	2.740	-0.021	0.148	314.641	0	422
dfy	0.010	0.011	0.005	1.753	3.677	0.006	0.034	459.547	0	422
tms	0.019	0.018	0.013	-0.602	-0.072	-0.026	0.044	25.731	0	422
cs	89.650	86.070	13.084	-0.412	-0.592	51.700	112.000	16.941	0	398
twex	94.310	96.050	14.334	0.799	0.641	70.340	143.900	52.849	0	422

of asset returns. Mean (annualised) returns are 12.42% for stocks and 8.15% for bonds, both series exhibit leptokurtosis; moreover stocks have negative skewness while bonds display positive skewness. Annualised volatilities are 15.69% for stocks and 5.67% for bonds. The Jarque-Bera statistic strongly rejects the hypothesis of normal distribution for both asset returns, as well as all the economic state variables considered.

3.3   Univariate Quantile Regression Results

Following the convention, the mean effects of the state variables are first examined using the predictive regression (3.1) on the full sample. Table 3.2 reports the estimated slope coefficients by the OLS for both the stocks and bonds. Then the distribution predictability is investigated using the quantile regression (3.2), also on the full sample. Table 3.3 presents the coefficient estimates of the state variables at the eleven chosen quantiles ranging from 0.05 to 0.95.

For stock returns, most of the state variables considered appear to have little ability to predict the mean according to the OLS estimation results. However, according to the quantile regression, six out of the eleven variables, namely, *cv, dy, p/e, epsg, dfy* and *vix*, show non-negligible effects on various parts of the return distribution. To gain some intuition, Figure 3.1 shows the effects of these six variables at finer quantile grids. Each plot in the figure depicts one variable coefficient in the quantile regression model. The solid line with filled dots represents the point estimates, $\hat{\beta}_{1j,\tau} : \tau = 0.05, 0.1, 0.15, \cdots, 0.95$ for the j-th variable, $j = 1, \cdots, 6$. The shaded gray area depicts 90% pointwise confidence bands. Superimposed on the plot is a dashed line representing the OLS estimate of the

Table 3.2: This table reports the mean effects of the economic state variables considered using predictive regression. The coefficient estimates of *cs, cv, p/b, p/e, vix* and *twex* have been multiplied by 100.

| Stock | | Bond | |
Variable	Estimate	Variable	Estimate
tbl3m	0.0364	tbl3m	0.0677***
infl	-0.0192	infl	0.0051
dfy	0.1654	dfy	0.3637**
tms	0.0473	tms	0.1039*
cs	-0.0081	cs	-0.0027
cv	-0.2363^*	twex	0.0209***
dy	0.6001^*		
p/b	-0.3807		
p/e	-0.1204^*		
epsgf	-0.2375^*		
vix	-0.0043		

∗ indicates significance at the 10% level
∗∗ indicates significance at the 5% level
∗ ∗ ∗ indicates significance at the 1% level

mean effect of the variable, with two dotted lines again representing a 90% confidence interval for this coefficient. The solid horizontal line is the zero line. The horizontal axis lists quantiles running from 0.05 through 0.95.

If assumptions for the standard linear regression model hold, the quantile slope estimates should fluctuate randomly around a constant level, with only the intercept parameter systematically increasing with τ. However, none of the slope estimates of the six variables can be described as random fluctuations here. In fact, the quantile slope estimates of the variables such as *cv, dfy* and *vix* follow a systematic pattern with negative values in the left tail and positive values in the right tail. These three variables are significant in the tail parts of the distribution, but have little impact in the middle. It seems that large positive and negative

60

impacts of the variables in the tails cancel each other out, which leads to barely significant results from the conditional mean estimates as indicated by the OLS estimates. The variable *dy* appears to have a significant positive effect in the lower tail, which implies the worst-case scenarios of the stock returns can be somewhat mitigated as dividend yield increases. Two earnings-related variables, *p/e* and *epsgf*, show overall negative effects and affect the middle parts of the distribution significantly.

As for the bonds, the conditional mean seems to be more predictable than that of the stocks as more variables come out significant in the linear regression analysis. Moreover, all six variables considered contribute to predicting the distribution. Figure 3.2 presents an intuitive summary of the quantile regression results for the bond data.

The OLS estimates, again, are far from an adequate summary of the variable effects on the bond returns. The slope estimates of the variables, *tbl3m, infl, dfy* and *twex*, all systematically increase with τ. While both the OLS and quantile results indicate that an increase in the value of *tms* is likely to increase the bond returns, the quantile analysis tells a more detailed story of how this variable affects the bond returns. The variable *tms* affects the lower to middle quantiles significantly, but not the upper part of the distribution. The effect of consumer sentiment (*cs*) is mainly in the left tail.

In summary, for both the stocks and bonds, the heterogeneous effects of the state variables on the returns are self-evident. It does not matter whether the heterogeneity arises from the volatility channel, as discussed in Section 3.1.1, or from more complicated channels, the quantile regression analysis provides a much richer picture than the conditional mean approach.

Table 3.3: This table reports the estimated slope coefficients of economic state variables obtained from quantile regression models for both the stock and bond returns. The coefficients are estimated at the quantiles $\tau = 0.05, 0.1, \cdots, 0.9$ and 0.95. The significance of the coefficient estimates is based on the induced smoothing method proposed by Wang et al. (2009). The coefficient estimates of *cv*, *p/b*, *p/e*, *vix*, *cs* and *twex* have been multiplied by 100.

	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	0.95
					Quantile						
					Stock						
cv	−1.679***	−0.885***	−0.982***	−0.640*	−0.616*	−0.429	−0.062	0.078	0.313	0.403*	0.282
dy	1.609***	0.846	0.522	0.861*	1.013**	0.506	0.445	0.561	0.305	0.303	0.383
p/b	−0.258	−0.571	−0.400	−0.749	−0.929	−0.726	−0.670	−0.018	−0.032	−0.202	−0.276
p/e	−0.027	−0.155	−0.107	−0.198**	−0.252**	−0.197*	−0.226**	−0.119	−0.025	−0.093	−0.069
epsgf	−0.437	−0.689**	−0.224	−0.453**	−0.554**	−0.384*	−0.336	−0.009	0.121	0.229	0.053
vix	−0.479***	−0.311***	−0.147*	−0.154*	0.011	0.013	0.128*	0.209***	0.227***	0.237***	0.277***
tbl3m	0.139	0.086	−0.006	0.011	−0.072	−0.044	0.002	0.057	0.014	0.005	−0.043
infl	0.034	−0.208	−0.044	0.001	−0.044	−0.020	0.002	0.054	0.070	0.003	−0.099
dfy	−1.711*	−1.795**	−0.338	−0.911	−0.359	0.021	0.571	0.850	1.207	1.860**	2.170***
tms	0.326	0.028	0.141	−0.003	0.116	0.017	−0.045	−0.099	−0.056	0.120	0.250
cs	0.073**	0.036	0.002	−0.016	−0.018	−0.007	−0.031	−0.021	−0.016	−0.012	−0.033
					Bond						
tbl3m	−0.183***	−0.091**	−0.068*	−0.030	−0.032	0.036	0.075*	0.102***	0.134***	0.169***	0.191***
infl	−0.246***	−0.203***	−0.116***	−0.100***	−0.066	−0.066	0.015	0.028	0.069	0.137***	0.345***
dfy	−1.481***	−0.837***	−0.457	−0.164	0.002	0.076	0.283	0.488*	0.747***	1.168***	1.360***
tms	0.124	0.241**	0.199**	0.142*	0.164**	0.151**	0.112	0.069	0.101	0.091	0.106
cs	0.048***	0.025**	0.009	0.006	0.007	0.003	0.001	0.002	−0.008	−0.014	−0.019
twex	−0.014	−0.007	0.009	0.006	0.011	0.017**	0.019***	0.021***	0.022***	0.037***	0.046***

∗ indicates significance at the 10% level
∗∗ indicates significance at the 5% level
∗ ∗ ∗ indicates significance at the 1% level

Figure 3.1: This figure plots the slope coefficient estimates for the stock data. The solid line with the filled dots gives the coefficients of state variables estimated from the quantile regression, with the shaded grey area depicting a 90% confidence interval. The dashed line gives the OLS estimate of mean effect, with two dotted lines again representing a 90% confidence interval for this coefficient. The solid horizontal line is the zero effect line. The coefficient estimates and corresponding confidence bands of *cv, p/e* and *vix* have been multiplied by 100.
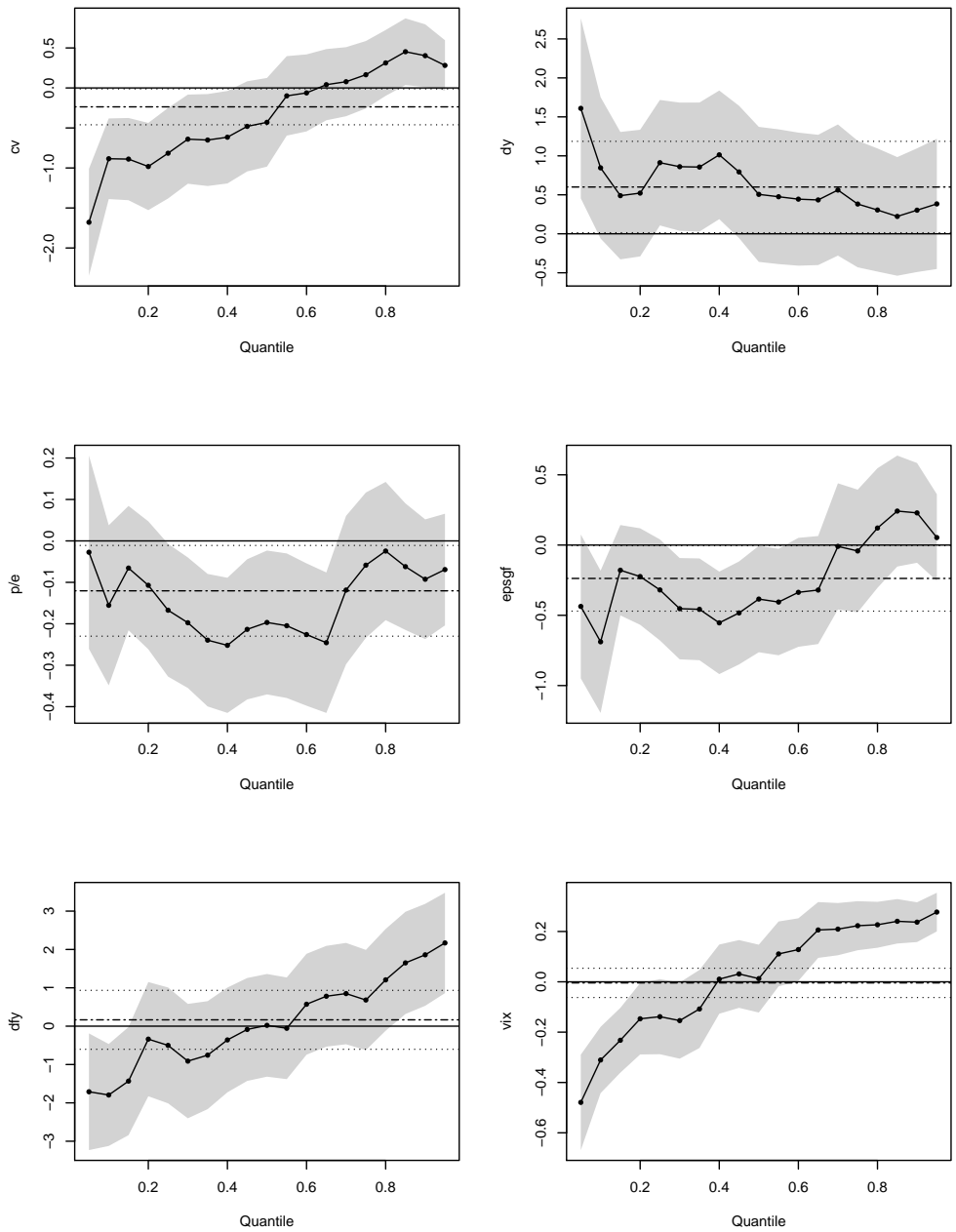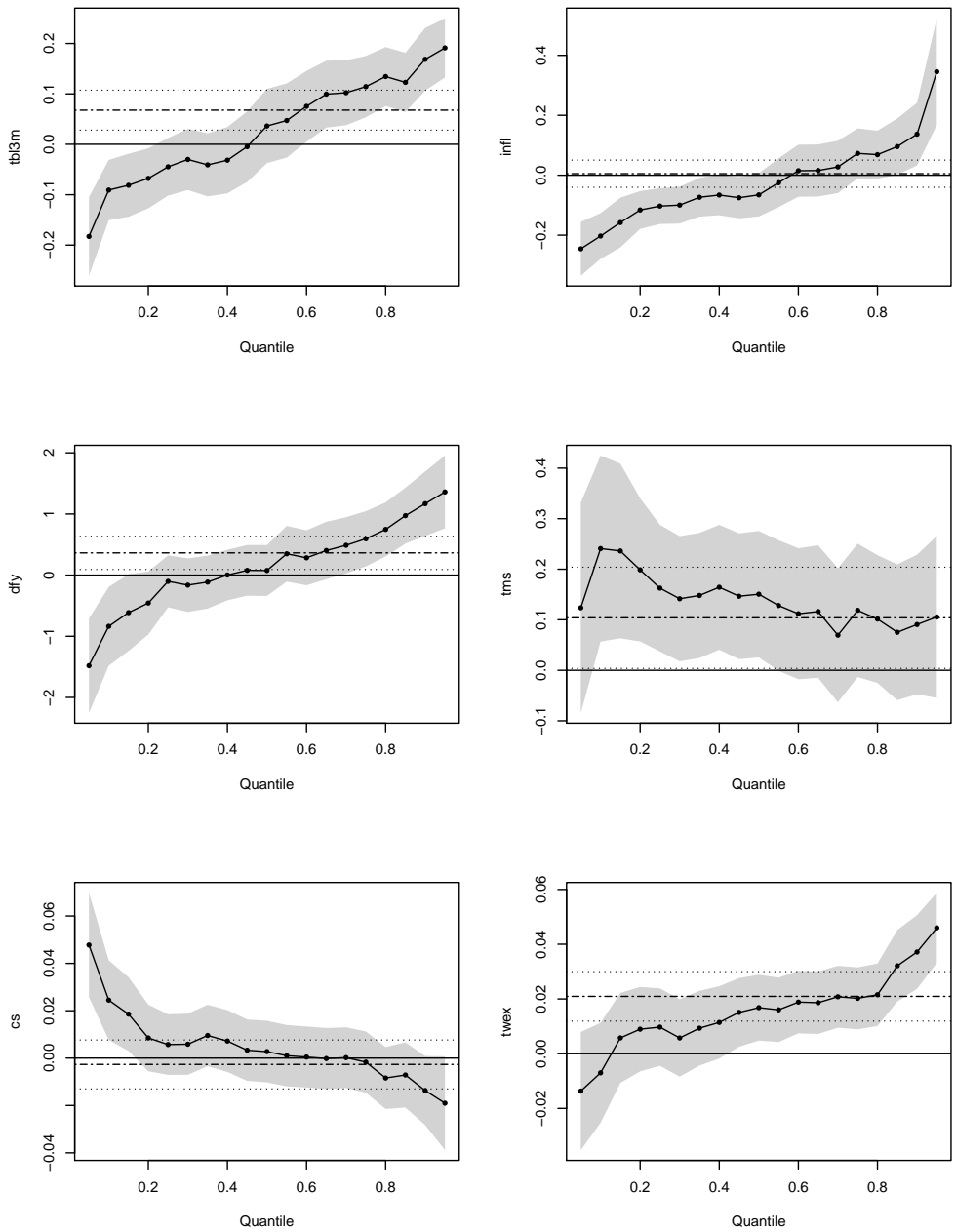
Figure 3.2: This figure plots the slope coefficient estimates for the bond data. The solid line with the filled dots gives the coefficients of state variables estimated from the quantile regression, with the shaded grey area depicting a 90% confidence interval. The dashed line gives the OLS estimate of mean effect, with two dotted lines again representing a 90% confidence interval for this coefficient. The solid horizontal line is the zero effect line. The coefficient estimates of *cs* and *twex* have been multiplied by 100.

64

3.4 Marginal Distributions by Model Combination

Two sets of variables are identified in the univariate analysis above to be useful in predicting return distribution: *cv, dy, p/e, epsg, dfy* and *vix* for the stock return and *tbl3m, infl, dfy, tms, cs* and *twex* for the bond return. To get a good return distribution forecast using the information available at the end of February 2011, an equal-weighted combination of the forecasts from each of the univariate quantile models is used,

$$Q_\tau^* = \frac{1}{n} \sum_{i=1}^n Q_\tau^i, i = 1, 2, \cdots, 6, \tag{3.7}$$

where Q_τ^i is the conditional τ-th quantile associated with the univariate model i. This equal-weighted combination is applied to both the stocks and the bonds. The use of the forecast combination is based on the following two reasons. Firstly, a multiple regression model that incorporates many predictor variables does not seem to work well in practice. Gains from using more variables in regression are likely to be outweighed by increasing parameter uncertainty. Indeed, Rapach *et al.* (2010) find that multi-factor regression model performs worse than single-factor predictive regression models in the case of forecasting equity premium. Secondly, the forecasting literature often shows that a simple average is difficult to outperform in a variety of settings in economics and finance (see, for example, Timmermann, 2006).

In order to get a distribution forecast for the returns in March 2011, a sufficiently fine grid of quantiles needs to be estimated. For this purpose, we use

Stock Return Distribution

Figure 3.3: Conditional distribution of stock returns forecasted at the end of February 2011. The mean and standard error of the distribution are 1.073% and 0.0403. Superimposed on the plot is a dashed line representing the normal distribution with the same mean and standard error.

a software package developed by Koenker [1] to automatically find all τ values at

---
[1]A quanitle regression package in R is available at http://www.econ.uiuc.edu/ roger/research/rq/rq.html.
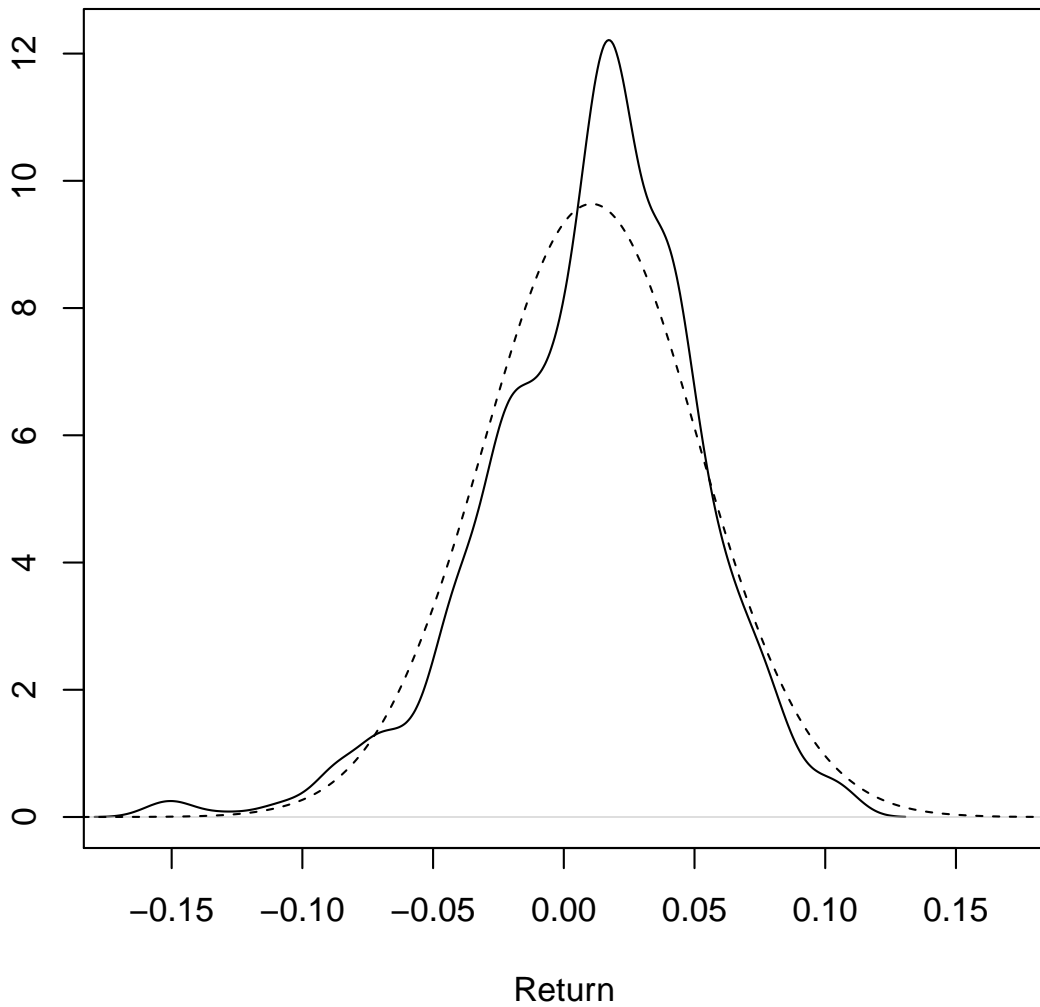
Bond Return Distribution
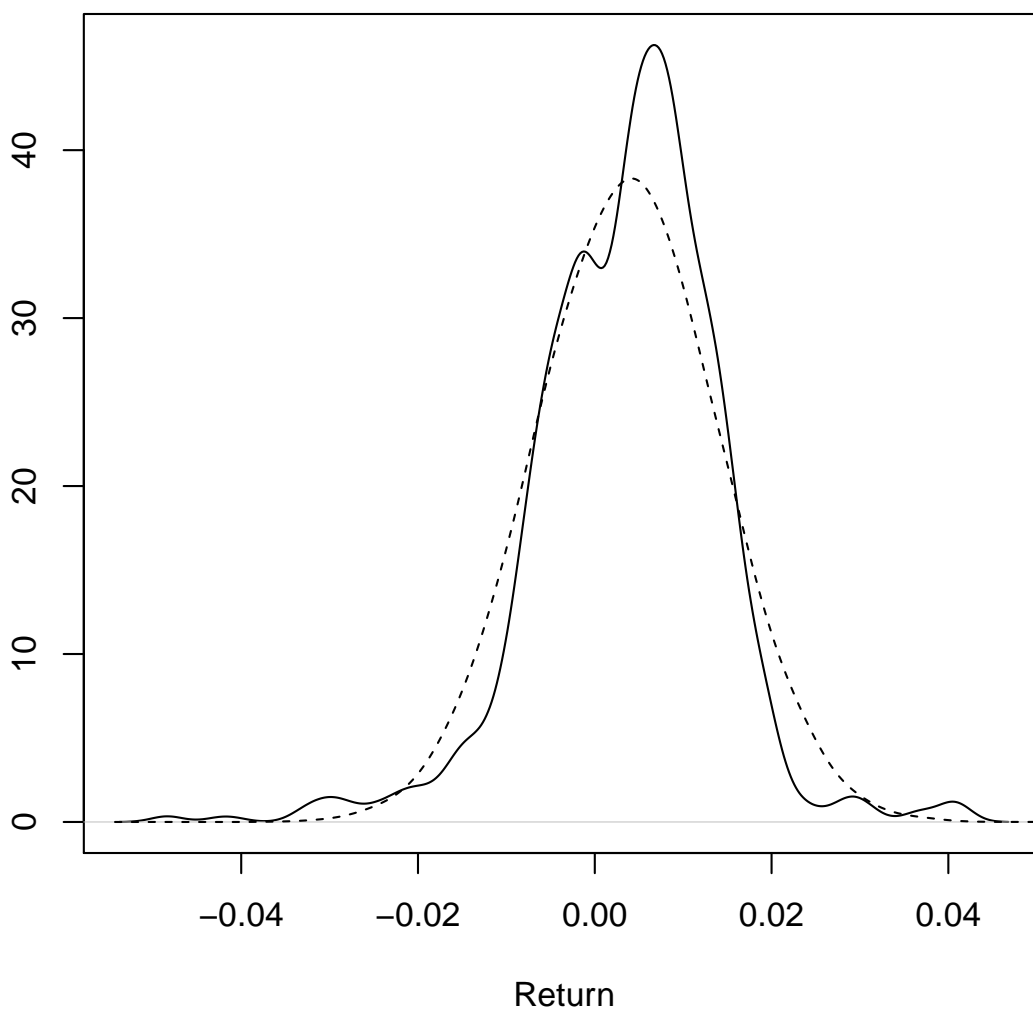
Figure 3.4: Conditional distribution of bond returns forecasted at the end of February 2011. The mean and standard error of the distribution are 0.374% and 0.0103. Superimposed on the plot is a dashed line representing the normal distribution with the same mean and standard error.

which the parameter estimates change. With these estimated quantile functions,

the inverse cumulative distribution functions for both the stock and bond returns

can be constructed. In order to get empirical return distributions, we generate 100,000 random returns for both the stocks and bonds. This is achieved by inverse transform sampling. First, we generate a random number u from the standard uniform distribution in the interval $[0, 1]$. Then, we find the closest quantile τ to u at which we have a solution, and compute the corresponding return. This return can be regarded as the random number drawn from the distribution described by the quantile functions. For both the stocks and bonds, we repeat the process 100,000 times to get 100,000 random returns.

The conditional stock return distribution at the end of February 2011 based on the 100,000 samples has mean 0.011, standard error 0.040, skewness -0.644 and excess kurtosis 1.178. Compared with the unconditional distribution summarised in Table 3.1, the conditional return is of higher mean, lower volatility, slightly lower downside risk and smaller excess kurtosis. Figure 3.3 depicts the conditional distribution of the stock returns (the solid line). The superimposed dash line is a normal distribution with the same mean and standard error as the stock returns. Using the normal distribution to approximate the conditional stock return distribution will result in underestimating downside risk and overestimating upside potential.

The conditional bond return distribution at the end of February 2011 based on the 100,000 samples has mean 0.004, standard error 0.010, skewness -0.477 and excess kurtosis 2.729. Compared with the unconditional bond distribution summarised in Table 3.1, the conditional return is of lower mean, lower volatility but larger downside risk and smaller excess kurtosis. Figure 3.4 depicts the conditional distribution of the bond returns (the solid line). Again the dash line is a normal distribution with the same mean and standard error as the bond

returns. Compared with the normal distribution, the conditional distribution of the bond returns has fatter tails and more concentrated mass in the middle of the distribution.

3.5 Out-of-sample Predictability

The above empirical studies report strong in-sample evidence of predictability of the full return distributions. It is not clear how much weight should be placed on out-of-sample statistics in judging the predictability of returns. Several authors have argued that poor out-of-sample performance is not evidence against predictability per se but is only evidence of the difficulty in exploiting predictability with trading strategies (see, for example, Inoue and Kilian, 2005; Cochrane, 2008). Nevertheless, we address the out-of-sample predictability of the full return distribution in this section. However, given the short return time series for both the stocks and bonds (a little more than 30 years) and even shorter time series for a number of predictor variables, the main purpose here is more of demonstrating how to assess out-of-sample predictability of a return distribution forecast than drawing concrete conclusions on out-of-sample predictability.

We use prevailing quantile to represent "no predictability". A prevailing quantile model is the quantile model (3.2) with no predictor variable,

$$Q_\tau(r_t|\mathcal{F}_{t-1}) = \beta_{0,\tau} + u_t. \tag{3.8}$$

A prevailing quantile model is calculated using historical return series only. It is the quantile equivalent to the prevailing mean used by Goyal and Welch (2003)

and Welch and Goyal (2008) for assessing out-of-sample mean return predictability.

To examine the out-of-sample predictability, we use data up to 1999:12 as the initial estimation sample and retain the period from 2000:01 to 2011:02 as the out-of-sample evaluation period. This out-of-sample period includes the burst of the dot-com bubble in 2000-2001 and the Global Financial Crisis in 2007-2009, and therefore can be considered a challenging period. One-step-ahead forecasts are generated for returns in 2000:01. In the following month (January 2000), the data window expands to include 2000:01, the parameters of the quantile regression models are re-estimated and then used to predict returns for 2000:02, and so forth, up to the end of the sample. This process generates a set of 133 out-of-sample forecasts for each quantile τ.

We assess three quantile forecasting models for both the stocks and the bonds, namely (i) six univariate quantile regression models (3.2) using the factors identified in Section 3.3; (ii) an equal-weighted combination of the forecasts from univariate quantile regression models as described in Section 3.4; (iii) a prevailing quantile model with no predictor variable (3.8).

We consider quantiles $\tau = 0.05, 0.1, 0.3, 0.5, 0.7, 0.9$ and 0.95. Similar to Cenesizoglu and Timmermann (2008), we use two statistics to measure model fit. The first one is out-of-sample coverage ratio, *i.e.*, the percentage of times that actual returns fall below the predicted τ-quantile. If a model is correctly specified, this coverage ratio should be equal to τ. The second measure is out-of-sample tick loss. The tick loss is based on the tick objective function (3.4) and is computed

Table 3.4: This table reports out-of-sample coverage probability which is the proportion of actual stock/bond returns in the out-of-sample period (2000:01 – 2011:02) that fall below the predicted quantile. For both the stocks and bonds, the first six rows report results for the univariate quantile regressions using the predictor variables listed in each row. The row "Combined" reports results for the equal-weighted combination of the six univariate quantile forecasts. The row "Prevailing" reports results for the prevailing quantile forecasts. The parameters of the forecasting models are estimated recursively using an expanding window of data.

	Quantile						
	0.05	0.1	0.3	0.5	0.7	0.9	0.95
	Stock						
cv	0.083	0.143	0.316	0.549	0.774	0.910	0.932
dy	0.113	0.158	0.301	0.496	0.737	0.902	0.940
p/e	0.120	0.165	0.353	0.526	0.744	0.887	0.932
epsgf	0.103	0.138	0.353	0.571	0.744	0.872	0.940
dfy	0.098	0.165	0.368	0.556	0.724	0.910	0.970
vix	0.083	0.135	0.353	0.571	0.744	0.932	0.960
Combined	0.088	0.133	0.331	0.541	0.739	0.902	0.959
Prevailing	0.128	0.165	0.361	0.556	0.752	0.910	0.920
	Bond						
tbl3m	0.068	0.120	0.308	0.489	0.624	0.850	0.887
infl	0.053	0.105	0.338	0.541	0.782	0.940	0.970
dfy	0.015	0.045	0.286	0.526	0.812	0.977	1.000
tms	0.030	0.068	0.293	0.549	0.805	0.977	0.985
cs	0.015	0.030	0.278	0.504	0.812	0.970	0.992
twex	0.038	0.068	0.248	0.466	0.722	0.910	0.970
Combined	0.043	0.070	0.286	0.496	0.744	0.947	0.960
Prevailing	0.023	0.060	0.293	0.519	0.820	0.977	0.985

Table 3.5: This table reports out-of-sample loss which is computed under the tick loss function over the out-of-sample period (2000:01 – 2011:02). For both the stocks and bonds, the first six rows report results for the univariate quantile regressions using the predictor variables listed in each row. The row "Combined" reports results for the equal-weighted combination of the six univariate quantile forecasts. The row "Prevailing" reports results for the prevailing quantile forecasts. The parameters of the forecasting models are estimated recursively using an expanding window of data.

	Quantile						
	0.05	0.1	0.3	0.5	0.7	0.9	0.95
				Stock			
cv	0.705	1.167	2.263	2.525	2.203	0.997	0.563
dy	0.893	1.405	2.282	2.417	2.090	1.037	0.583
p/e	0.892	1.422	2.246	2.365	2.068	1.034	0.588
epsgf	0.877	1.396	2.243	2.398	2.121	1.051	0.600
dfy	0.851	1.345	2.316	2.422	2.062	0.994	0.574
vix	0.718	1.225	2.314	2.506	2.018	0.871	0.491
Combined	0.765	1.263	2.254	2.420	2.084	0.982	0.534
Prevailing	0.883	1.409	2.305	2.425	2.085	1.031	0.579
				Bond			
tbl3m	0.195	0.289	0.506	0.551	0.468	0.278	0.194
infl	0.181	0.278	0.508	0.544	0.469	0.241	0.161
dfy	0.188	0.287	0.506	0.559	0.488	0.269	0.176
tms	0.193	0.299	0.509	0.554	0.483	0.279	0.187
cs	0.197	0.291	0.504	0.551	0.492	0.286	0.195
twex	0.185	0.285	0.509	0.534	0.449	0.250	0.158
Combined	0.182	0.281	0.502	0.541	0.459	0.241	0.153
Prevailing	0.188	0.285	0.504	0.545	0.476	0.276	0.186

as follows,

$$L_\tau = \sum_t (\tau - I\{r_t - Q_\tau(r_t) < 0\})(r_t - Q_\tau(r_t)).$$

Leitch and Tanner (1991) find that compared with conventional measures such as mean squared errors, this tick loss is more closely related to the possibility of making economic profits from return forecasts.

Table 3.4 reports the out-of-sample coverage ratio for all the models considered. Firstly, we notice that the coverage ratios are not so close to their corresponding τ values, especially in the tails. This inaccuracy can be due to the fact that we have short time series and only 133 out-of-sample forecasts. For univariate models, there is a certain degree of variation in the results across models and quantiles. By using the absolute deviation between the coverage ratio and the corresponding τ value as the measure of closeness, we find that no single univariate model is consistently better than the simple prevailing quantile model which assumes no predictability. On the other hand, although the equal-weighted quantile combination is not the best for all the quantiles considered, it is very robust and delivers a satisfactory performance. It consistently outperforms the prevailing quantile model for almost all the quantiles for the stock and bond data. The only exception is at $\tau = 0.3$ for the bond data, where the equal-weighted quantile combination slightly underperforms the prevailing quantile model.

Table 3.5 lists the out-of-sample loss under the tick loss function. Again, none of the single univariate quantile models can consistently outperform the prevailing quantile model. But the equal-weighted quantile combination consistently outperforms the prevailing quantile for all the quantiles and for both the stocks and bonds. Especially in the tails, it outperforms by large margins, which indicates

the relatively good predictive ability in the tails of the distributions.

The out-of-sample results show that the simple equal-weighted model combination adds value and improves upon the individual univariate quantile models. The out-of-sample analysis also suggests that the return distributions in general are predictable by economic state variables through the simple equal-weighted model combination with the tails of the distributions being especially predictable.

3.6   Summary

The traditional focus on return predictability is the conditional mean, which is insufficient and does not reveal a complete picture of the returns. This chapter investigates the predictability of return distributions in a quantile regression framework. The use of quantile regression allows us to examine the predictability of any specific part of the return distributions. For a sufficiently fine grid of quantiles, the entire distribution can be traced out.

We carry out empirical studies to investigate the return distribution predictability of the Russell 1000 Index and the US Aggregate Bond Index. These studies report strong empirical evidence of the predictability of different parts of the distribution other than mean both in-sample and out-of-sample. In the in-sample study, a number of economic state variables show significant but heterogenous effects on various parts of return distributions, which are especially pronounced for the bond returns. In the out-of-sample study, the distributions are predictable by economic state variables through the simple equal-weighted model combination. For both the in-sample and the out-of-sample studies, the evidence of predictability is strongest in the tails of the distributions.

This chapter complements the literature on return distribution predictability by providing further empirical evidence using the two broad-based indices. Given the wide recognition of these two indices in investment communities, predictability of their returns has academic value as well as a significant economic value to investors. The empirical analysis also demonstrates that quantile regression for predicting distributions is flexible and general enough to capture any non-Gaussian characteristics in asset returns.

Chapter 4

Joint Return Distribution Modelling and Distribution-based Portfolio Selection

4.1 Modelling Joint Return Distribution

4.1.1 Copulas for return dependence

Chapter 3 provides strong empirical evidence of predictability of return distributions. It shows the distribution in general is more predictable than the conditional mean of stock and bond returns. The marginal return distribution of an asset can be obtained using the proposed quantile regression model (3.2) and choosing a sufficiently fine grid of quantiles. For many applications in economics and finance, however, marginal distributions are not enough. This is the case especially for portfolio management as almost all investment decisions involve more than

76

one asset. It is therefore of great importance to investigate joint return distributions. To model a joint return distribution, we first consider modelling return dependence using copulas.

A copula is a multivariate function with one-dimensional margins being uniform on $[0, 1]$. The term *copula* was introduced by Sklar (1959). However, the idea of copula dates back to Hoeffding (1940, 1941), who established the best possible bounds for these functions and studied measures of dependence that are invariant under strictly increasing transformations. Nelsen (2006) provides a comprehensive introduction to the copula theory.

The fundamental theorem in copula theory is established by Sklar (1959).

Theorem 4.1.1. *(Sklar's Theorem, 1959): Let F be a distribution function on R^k with one-dimensional distribution $F_1, \cdots, F_k$. Then there is a copula C such that*

$$F(x_1, \cdots, x_k) = C(F_1(x_1), \cdots, F_k(x_k)). \tag{4.1}$$

If F is continuous, then C in (4.1) is unique and is given by

$$C(u_1, \cdots, u_k) = H(F_1^{-1}(u_1), \cdots, F_k^{-1}(u_k))$$

for $u = (u_1, \cdots, u_k) \in R^k$, where $F_i^{-1}(u_i) = inf\{x : F_i(x) \geq u_i\}, i = 1, \cdots, k$. Conversely, if C is a copula on $[0, 1]^k$ and $F_1, \cdots, F_k$ are distribution functions on R, then the distribution function defined in (4.1) is a distribution function on R^k with one-dimensional margin $F_1, \cdots, F_k$.

Although the theory of copulas was established in 1959, it was not until the 1970s that copulas were used in the modelling of data. Since the pioneering

work of Embrechts *et al.* in 1999, copula models have enjoyed steadily increasing popularity in finance. For example, Li (2000) studies the problem of default correlation in credit risk models using copulas, while Bouyé *et al.* (2000) and Cherubini *et al.* (2004) discuss various applications of the copula theory to financial problems. Patton (2006) further proposes extensions of the copula theory to allow for conditioning variables and employs it to construct flexible models of the conditional dependence structure of exchange rates. Ammann and Süss (2009) apply the skewed Student's t copula to generate meta-skewed Student's t distributions. They find that the asymmetry property of the copula helped to improve description of the dependence structure between equities returns. Fischer *et al.* (2009) consider constructing high-dimensional copulas to sufficiently capture the characteristics of financial returns.

Compared with measures such as correlation, which can only capture linear dependence, a copula is a more sophisticated and complete description of the dependence structure of asset returns. The most frequently used copula families are elliptical copulas and Archimedean copulas. The copulas in the elliptical family include the Gaussian copula and Student's t copula. The copulas in the Archimedean family include the Clayton, Frank, Gumbel and Ali-Mikhail-Haq (AMH) copulas. Their detailed functional forms are given below. Details and properties of these copulas are given in Nelsen (2006).

Gaussian copula The Gaussian copula takes the form

$$
\begin{aligned}
C(u, v; \rho) &= \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v)) \\
&= \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho}} \exp\left\{ \frac{-(x^2 - 2\rho xy + y^2)}{2(1-\rho^2)} \right\} dx dy,
\end{aligned}
$$

where Φ_ρ denotes the standard bivariate normal cumulative distribution function with correlation ρ, and Φ is the standard normal cumulative distribution function. The normal copula is flexible in that it allows for equal degrees of positive and negative dependence.

Student's t copula The bivariate t-copula has two dependence parameters, degrees of freedom ν and correlation ρ,

$$
\begin{aligned}
C(u, v; \rho, \nu) &= t_{\rho,\nu}(t_\nu^{-1}(u), t_\nu^{-1}(v)) \\
&= \int_{-\infty}^{t_\nu^{-1}(u)} \int_{-\infty}^{t_\nu^{-1}(v)} \frac{\Gamma(\frac{\nu+2}{2})}{\nu\pi\Gamma(\frac{\nu}{2})\sqrt{1-\rho^2}} \left(1 + \frac{x^2 + y^2 - 2\rho xy}{\nu(1-\rho^2)}\right) dx dy, \\
&\qquad\qquad\qquad\qquad \rho \in (-1, 1), \nu > 2,
\end{aligned}
$$

where t_ν is the probability density function of a student's t distribution with degree of freedom ν, and t_ν^{-1} is the inverse cumulative distribution function of a student's t distribution. The parameter ν controls the heaviness of the tails. For $\nu < 3$, the variance does not exist, and for $\nu < 5$, the fourth moment does not exist. As $\nu \to \infty$, $C(u, v; \rho, \nu) \to C(u, v; \rho)$.

Clayton copula The Clayton copula takes the form

$$
C(u, v; \theta) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, \qquad \theta \in [-1, \infty) \setminus \{0\}.
$$

The Clayton copula cannot account for negative dependence. It exhibits strong left-tail dependence and relatively weak right-tail dependence.

Gumbel copula The Gumbel copula takes the form

$$
C(u, v; \theta) = \exp\left\{-\left[(-\ln u)^\theta + (-\ln v)^\theta\right]^{1/\theta}\right\}, \qquad \theta \in [1, \infty).
$$

Similar to the Clayton copula, Gumbel does not allow negative dependence, but, in contrast to the Clayton, the Gumbel exhibits strong right-tail dependence and relatively weak left-tail dependence.

Frank copula The Frank copula takes the form

$$C(u, v; \theta) = -\frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right).$$
$$\theta \in (-\infty, \infty) \setminus \{0\}.$$

The Frank copula permits negative dependence between marginals. Similar to the Gaussian and Student's t copulas, dependence is symmetric in both tails for this copula.

Ali-Mikhail-Haq (AMH) copula The AMH copula takes the form

$$C(u, v; \theta) = \frac{uv}{1 - \theta(1 - u)(1 - v)}, \ \theta \in [-1, 1].$$

The AMH copula allows both negative and positive dependence.

4.1.2 Quantile-copula approach for joint returns

Sklar's theorem states that a joint distribution can be expressed in terms of its respective marginal distributions and a dependence function C that binds them together. In other words, copulas can be used to piece together joint distributions when only marginal distributions can be specified. Therefore, combining non-parametric marginal distributions modelled by quantile regressions and a dependence structure modelled by a copula, a joint return distribution can be easily obtained.

This quantile-copula approach for joint return modelling is general and flexible. It allows rich non-Gaussian joint distribution families to be generated. The conventional way of modelling joint asset returns in the literature is through the use of multivariate elliptical distributions. For example, Ang and Bekaert (2002), Guidolin and Timmermann (2007) and Mencía and Sentana (2009) use mixtures of multivariate normal distributions; while Harvey *et al.* (2010) consider a multivariate skewed normal distribution. Jondeau and Rockinger (2006) and Adcock (2010) employ a skewed Student's t distribution to allow for more dispersion. Despite limited choice of multivariate elliptical distributions, these joint distributions also suffer from various parametric constraints. For example, the skewness of both the multivariate Student's t and normal distribution is a function of the first two moments, not a separate parameter. Our quantile-copula approach for joint distributions does not have such constraints and is, therefore, more flexible.

The quantile-copula approach for joint return distribution modelling is computationally cheap to implement. Pedersen (2010) proposes modelling the joint asset returns through the use of multivariate quantile regression, *i.e.,* a single regression model with more than one response variable. Although the multivariate quantile regression can yield some insights into the joint asset returns, it has disadvantages. The parameter estimation for multivariate quantile regression is challenging, especially computationally as the estimates are often unstable and not unique. On top of that, constructing a joint distribution using multivariate quantile regression is not a straightforward task. It involves estimation of sufficiently fine grid of multidimensional quantiles, where the computational intensity increases exponentially with the number of dimensions. In contrast, the well-developed copula theory is a more convenient vehicle than the less-developed

multivariate quantile regression approach.

The implications of return distribution foreseeability reach throughout finance but are yet to be explored. Return distribution predictability is especially pertinent to portfolio selection. As reviewed in Section 1.1.2, instead of using full distributions, a majority of the finance literature on portfolio selection reduces to approximate distributions with a few individual moments. Now, given that joint return distributions can be modelled in the quantile-copula framework, we seek to extend moment-based portfolio selection to a distribution-based exercise. For this purpose, we need to have an objective function which utilises the full underlying distribution information, as well as a proper optimisation algorithm for obtaining a solution.

4.2   Portfolio Selection: Objective Function

4.2.1   Omega measure

Let r be return of a portfolio over a period. The return r can be decomposed as

$$r = L + \max(r - L, 0) - \max(L - r, 0), \tag{4.2}$$

where L is a benchmark return or a reference point so that a return above L is considered as a gain and a return below L as a loss. Therefore, $\max(r - L, 0)$ and $\max(L - r, 0)$ can be viewed as upside potential and downside risk of the portfolio, respectively. Adjusting the upside potential and downside risk by their

82

corresponding probabilities, we get

$$E[\max(r - L, 0)] = \int_L^\infty (r - L)f(r)dr,$$

and

$$E[\max(L - r, 0)] = \int_{-\infty}^L (L - r)f(r)dr,$$

which are the upper partial moment (UPM) of order 1 and the lower partial moment (LPM) of order 1, respectively. The ratio of the UPM and LPM of order 1 is the so-called "Omega" by Shadwick and Keating (2002),

$$\Omega(L) = \frac{E[\max(r - L, 0)]}{E[\max(L - r, 0)]} = \frac{UPM_1(L)}{LPM_1(L)}. \qquad (4.3)$$

The Omega measure is the ratio of the expected upside of the asset over the benchmark (over-performance) and its expected downside. The Omega can also be viewed as the ratio of the payout of a "virtual" call option $E[\max(r - L, 0)]$ over the payout of a "virtual" put option $E[\max(L - r, 0)]$. Shadwick and Keating (2002) describe the Omega measure as a probability adjusted ratio of gains to losses and state that, for a benchmark L, the simple rule of preferring more to less implies that an asset with a high value of Omega is a better investment than one with a lower value.

The Omega measure provides a succinct summary of financial performance/risk of a portfolio. Instead of using only a few individual moments to represent a return distribution, as the quadratic utility does, it incorporates all the higher-order moments of a return distribution. In addition, the Omega also takes into account the benchmark return against which a given outcome will be viewed as a gain or

a loss. Even in the case in which returns are normally distributed, this provides additional information which mean and variance alone do not encode. This can lead to significantly different portfolio optimisations than are produced by the classical mean-variance portfolio analysis.

In the following section, we show the connection between the Omega measure and prospect theory and propose a general version of the Omega measure according to the prospect theory to incorporate asymmetric preference for gains and losses.

4.2.2   Prospect theory and generalised Omega measure

Darsinos and Satchell (2004) establish the connection between prospect theory and the Omega measure which we summarise in this section.

Prospect theory is developed by Kahneman and Tversky (1979) as a psychologically more accurate description of preferences compared to expected utility theory. It describes how people choose between probabilistic alternatives and evaluate potential losses and gains. In short, people evaluate outcomes relative to a reference point L which is variable and often dependent on initial wealth v_0. Then people make decision by maximizing the gains based on the potential outcomes and their respective probabilities.

In prospect theory, a full description of decision making involves specifying the utility function relative to a reference point $L(v_0)$, which is usually initial wealth (v_0) dependent. Letting $U(\cdot)$ be utility, with v being final wealth, it is

usual to write the prospect utility function as:

$$
\begin{aligned}
U(v) &= U_1(v - L(v_0)), \text{ if } v > L(v_0) \\
&= -U_2(L(v_0) - v), \text{ if } v \leq L(v_0),
\end{aligned}
\tag{4.4}
$$

where $U_1(\cdot)$ and $U_2(\cdot)$ are increasing functions. Denoting $r = v/v_0$ and $L = L(v_0)/v_0$ and letting $v_0 = 1$ without loss of generality, the prospect utility function (4.4) can be written as

$$
\begin{aligned}
U(r) &= U_1(r - L), \text{ if } r > L \\
&= -U_2(L - r), \text{ if } r \leq L.
\end{aligned}
\tag{4.5}
$$

As suggested by prospect theory, investors more often have an asymmetric preference for gains and losses. Thus, we include an extra parameter into equation (4.5):

$$
\begin{aligned}
U(r) &= U_1(r - L), \text{ if } r > L \\
&= -\lambda U_2(L - r), \text{ if } r \leq L,
\end{aligned}
\tag{4.6}
$$

where $\lambda > 0$. Such a specification is called a loss aversion utility function.

If we set

$$
U_1(r - L) = E(r - L | r > L)\text{pr}(r > L) = E[\max(r - L, 0)],
$$
$$
U_2(L - r) = E(L - r | r \leq L)\text{pr}(r \leq L) = E[\max(L - r, 0)],
$$

the Omega measure is the ratio of U_1 over U_2 with a symmetric preference for gains

85

and losses, *i.e.*, $\lambda = 1$. Of course, other performance measures can be generated by specifying different increasing functions for U_1 and U_2 and combining them differently, such as taking minus instead of taking ratios.

One way to generalise the Omega measure to take account of investors' loss attitude is

$$\Omega_1(L, \lambda) = E[\max(r - L, 0)] - \lambda E[\max(L - r, 0)].$$

It is actually the regret-reward measure studied by Dembo and his colleagues (see, for example, Dembo and Rosen, 1999; Dembo and Mausser, 2000). However, this measure suffers a critical drawback. Taking the expectations of the equation (4.2), we have

$$E(r) = L + E[\max(r - L, 0)] - E[\max(L - r, 0)].$$

By setting $L = \mu$, the mean of the return, we get

$$\mu = \mu + E[\max(r - \mu, 0)] - E[\max(\mu - r, 0)],$$

or

$$E[\max(r - \mu, 0)] - E[\max(\mu - r, 0)] = 0.$$

Therefore, the regret measure fails to rank return performances when $\lambda = 1$ and $L = \mu$ because it is 0 for any return distribution under these conditions.

An appropriate generalisation of the Omega function (4.3) to incorporate loss aversion can be

$$G_\Omega(L, \lambda) = \log(E[\max(r - L, 0)]) - \lambda\log(E[\max(L - r, 0)]), \lambda > 0. \qquad (4.7)$$

The loss aversion parameter $\lambda < 1$ corresponds to risk seekers, who tend to be lured by a large potential gain and place a relatively small weight on potential losses. And $\lambda > 1$ applies to a conservative investor, who views losses more seriously and are willing to trade some of their average returns for a decreased chance that they will experience a large loss. When $\lambda = 1$, G_Ω is the log-version of the Omega, and maximizing it is equivalent to maximizing the Omega. This proposed measure G_Ω does not break down when $\lambda = 1$ and $L = \mu$ as the regret-reward measure does.

4.3 Portfolio Selection: Optimisation Technique

We consider a one-period portfolio optimisation problem. There are J assets, and an investor holds them from time $t = 0$ until time $t = T$. Let $\mathcal{W}$ be a vector of length J which stores the weights of assets in the investor's portfolio, and let R be a vector of length J which stores the returns of assets from $t = 0$ to $t = T$. The investor's return will then be given by $r_T = \mathcal{W}'R$. Our objective is to maximize G_Ω, or alternatively, minimize $-G_\Omega$,

$$\min_{\mathcal{W}}(-G_\Omega(r_T|L,\lambda))$$

$$\text{s.t. } \sum_{j=1}^{J} w_j = 1, \text{ and}$$

$$w_j^{inf} \leq w_j \leq w_j^{sup}, j \in J,$$

where w_j is the j-th element of the weight vector $\mathcal{W}$, and w_j^{inf} and w_j^{sup} are the minimum and maximum holding sizes for the j-th asset. We do not include

a return constraint here because the objective function, $-G_\Omega(r_T|L, \lambda)$, already includes a measure of reward.

At first sight, minimizing $-G_\Omega(r_T|L, \lambda)$ is a difficult task, as the resulting optimisation problem is not convex. This means optimisation performed on the generalised Omega may lead to a rough and even discontinuous objective surface which can no longer be handled by linear or quadratic programming. In the optimisation literature, heuristic optimisation techniques are proposed as a way out of this problem. Several authors have investigated the application of heuristic optimisation techniques to portfolio selection, including but not limited to Dueck and Winker (1992), Chang *et al.* (2000), Beasley *et al.* (2003), Maringer (2005), Gilli *et al.* (2006) and Gilli and Schumann (2010). Amongst all heuristic optimisation techniques, the threshold-accepting algorithm is one of the most popular procedures. We briefly introduce the threshold-accepting algorithm here. Gilli *et al.* (2006) and Gilli and Schumann (2010) provide a more general and detailed exposition.

Belonging to the class of local search algorithms, the threshold-accepting algorithm starts with a random feasible solution and then explores its neighborhood in the solution space by moving from its current position, accepting a new solution if and only if it improves the objective function according to a certain threshold. The implementation of the algorithm requires the definition of the search space, the objective function, the neighborhood and the threshold sequence.

We illustrate the algorithm in an allocation problem of two assets with no short-selling constraint. The search space is all the bivariate combination $\mathcal{W} = (w_1, w_2)$ subject to $w_1 \geq 0$ and $w_1 \geq 0$ and $w_1 + w_2 = 1$. The objective function is $f(w; L, \lambda) = -G_\Omega(w; L, \lambda)$. The neighborhood of a solution w^c, $\mathcal{N}(w^c)$ is defined

using ϵ-spheres:

$$\mathcal{N}(w^c) = \{x^n | x^n \in \mathcal{W}, \| x^n - x^c \| < \epsilon\}.$$

The threshold-accepting algorithm comprises two main procedures.

1. Generate the threshold sequence ψ which is of length n_r, in a descending order and decreases toward 0. The procedure is as follows:

   Randomly choose weights $w^c \in \mathcal{W}$

   for $i = 1 : n_d$ do

     compute $w^n \in \mathcal{N}(w^c)$ and $\delta_i = |f(w^n) - f(w^c)|$

     $w^c = w^n$

   end for

   compute empirical distribution F of δ based on $\delta_i, i = 1, 2, \cdots, n_d$

   compute threshold sequence $\psi_k = F^{-1}(\frac{n_r - k}{n_r}), k = 1, 2, \cdots, n_r$

2. Search for the best solution by iterating through all values of ψ.

   Randomly generate current solution $w^c \in \mathcal{W}$

   for $i = 1 : n_r$ do

     for $j = 1 : n_s$ do

      Generate w^n and compute $\delta = f(w^n) - f(w^c)$

      if $\delta < \psi_i$, then $w^c = w^n$

     end for

   end for

   $w^{opt} = w^c$

In order to explore the search space more efficiently, the algorithm may be restarted m times by repeating the above two procedures using different starting values, w^c. The final solution is then taken to be the best solution amongst all restarts.

A classic local search stops at the first local minimum that it finds, which may not be the global optimal. The threshold-accepting search overcomes this problem by allowing uphill moves through the greater than 0 thresholds in ψ, hence it also accepts new solutions which lead to a deterioration in the objective function. This allows the algorithm to walk away from local minima.

4.4   Empirical Illustration

We use the stock and bond data as described in Chapter 3 to illustrate the effectiveness of the proposed quantile-copula approach for joint distribution modelling and the portfolio selection under the generalised Omega measure.

4.4.1   Joint return distribution of stocks and bonds

Chapter 3 obtains the marginal return distributions of the Russell 1000 Index and the US Aggregated Bond Index using the quantile regression. In order to get the joint return distribution, we use copulas to model the dependence structure of the stock and bond returns.

The Spearman correlation of the stock and bond returns for the period from 1979:01 to 2011:02 is 0.230. Any further anticipation of independence of the two series can be erased by a simple linear regression. By regressing the monthly stock return at time t on the corresponding bond returns at time t, a highly

significant slope coefficient 0.608 with standard error 0.133 is obtained. This strongly suggests that two returns are not independent.

Table 4.1: This table reports copula fitting results using maximum likelihood.

| Copula | Likelihood | Parameter | | |
		Estimate	Std.err	p-value
Gaussian	12.0665	0.2469	0.0443	0
Student's t	13.5789	0.2457	0.0525	0
Gumbel	13.9588	1.1809	0.0449	0
Clayton	6.6609	0.2312	0.0657	0.0004
Frank	10.8405	1.4694	0.3053	0
AMH	9.6284	0.5414	0.0959	0

All the copulas listed in Section 4.1.1 are considered for modelling the dependence of the stock and bond returns. In the case of uniform marginal, the copula is equivalent to the joint cumulative distribution function. The model parameters can thus be estimated using the maximum likelihood method. Table 4.1 reports the copula fitting results, including log-likelihood values and parameter estimates. The degree of freedom of the Student's t copula is estimated to be 10.6. It can be seen that the Gumbel copula attained the greatest log-likelihood value amongst all copulas considered. Hence the Gumbel copula is used to model the dependence of the two return series. The implication of the Gumbel copula is that the stock and bond returns exhibit strong right-tail dependence and relatively weak left-tail dependence.

Using the Gumbel copula dependence structure coupled with the marginal distribution modelled through the equal-weighted quantile model combination (3.7), the joint return distribution of the stocks and bonds can be obtained empirically.

4.4.2 Optimal asset allocation

The asset allocation between the stocks and bonds are carried out using the joint distribution modelled by the quantile-copula approach, the G_Ω measure and the threshold-accepting optimisation technique. The parameters in the threshold-accepting algorithm take the values $n_r = 10, n_d = 5000$ and $n_s = 2000$. For the optimisation problem considered in this section, the solutions are of little difference amongst restarts, hence m is taken to be 1. However, a large value of m, eg. $m = 60$, is preferred for optimisation problems with rough and non-continuous surfaces, for example, portfolio selection with Value-at-Risk (VaR) (see, for example, Gilli *et al.*, 2006; Gilli and Schumann, 2010).

Four values of the loss aversion parameter are considered, namely $\lambda = 1, 1.2, 1.4$ and 1.6, corresponding to an increasing loss aversion toward portfolio selection. Firstly, $n = 100,000$ independent paired random numbers are generated from the estimated Gumbel copula and denoted as $(u_i, v_i), i = 1, 2, \cdots, n$. Denote the cumulative distribution functions of the stocks and bonds as F_s and F_b, respectively, which are obtained using the quantile model combination (3.7). Then n pairs of returns for the stocks and bonds can be generated by $r_{s,i} = F_s^{-1}(u_i)$ and $r_{b,i} = F_b^{-1}(v_i)$, for $i = 1, 2, \cdots, n$. These generated paired returns are believed to be a good representation of the joint distribution. It follows that portfolio selection by maximizing G_Ω with the no short-selling constraint can be performed using the 100,000 samples. [1]

Figure 4.1 shows how the stock weights in the optimal portfolios change with an investor's benchmark return. The benchmark returns span from 0 to 1.2%,

---

[1]Several more sets of 100,000 joint return samples are also generated, on which the optimal weights differ only after the fourth digital. Therefore the sample size of 100,000 is big enough to obtain reliable optimisation results.
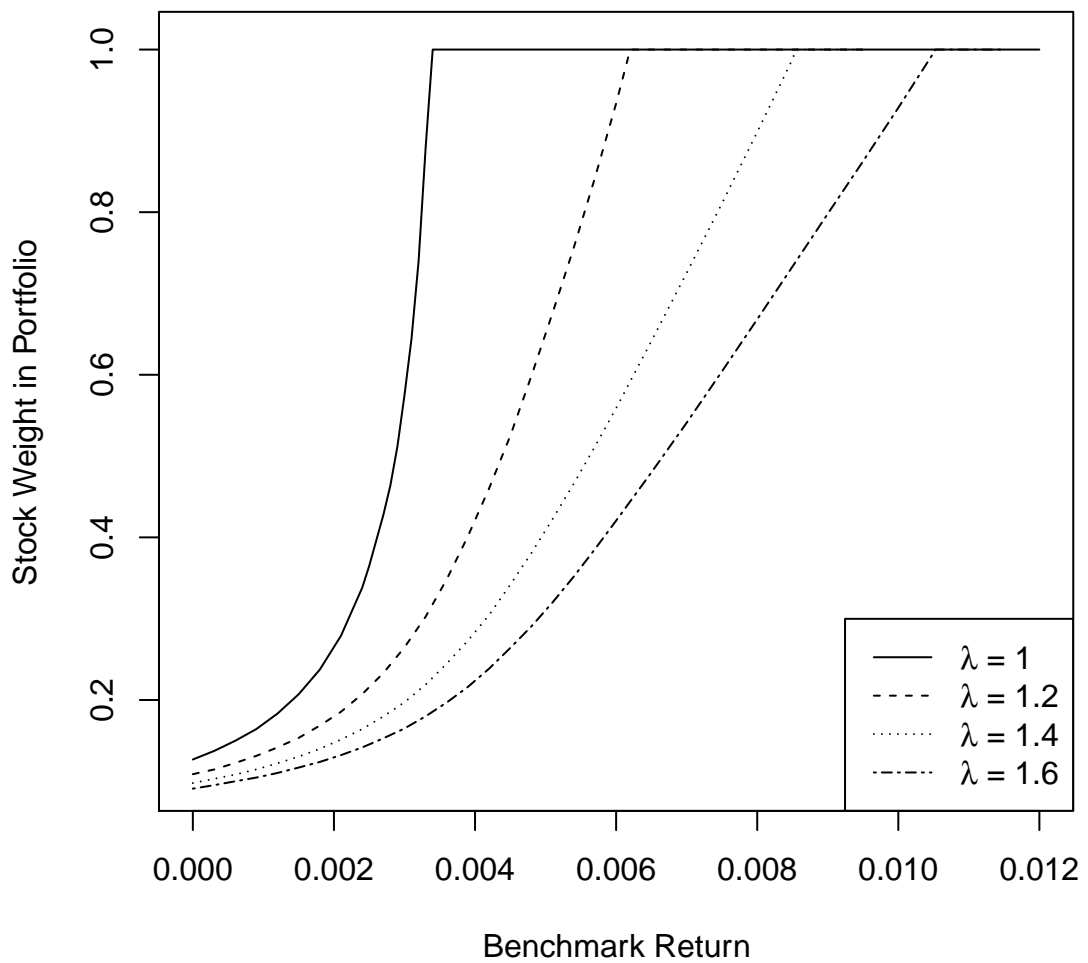
Figure 4.1: This figure plots the stock weight from maximizing G_Ω. The x axis is benchmark return L which spans from 0 to 1.2%. At the end of February 2011, the risk-free rate is 0.011%, the mean forecasted stock return is 1.073% and the mean forecasted bond return is 0.374%. Different types of lines correspond to different levels of risk aversion which increases from left to right, namely $\lambda = 1, 1.2, 1.4,$ and 1.6.

covering a few important numbers, including the risk-free rate at the end of February 2011 (0.011%), the expected future stock return (1.073%) and the expected future bond return (0.374%). The first pronounced feature is that the optimal weight in stocks increases with the investor's benchmark return. It is hardly surprising. As the required return level increases, it leaves no option but to hold more and more assets which can offer more upside potential, in this case, stocks. Besides the required return, the investor's attitude toward loss also plays a big role in allocating the assets. For a given benchmark return, the more loss-averse the investor is, the less inclined he is to hold risky asset stocks. For example, for the benchmark return $L = 0$, the stock weights decline from 13.09% to 9.29% as λ increases from 1 to 1.6.

4.5 Summary

This chapter develops a general and flexible framework for modelling joint return distributions. The building blocks of the framework are quantile regressions and copulas. Under the copula theory, a joint distribution can be decomposed into two separate parts, marginal distributions and a dependence structure. Quantile regressions are used to model marginal distributions of asset returns, while copulas are employed to capture dependence structure across asset returns. The proposed framework is very flexible in reproducing statistical features of returns. It also remains tractable even when several assets are considered.

This chapter also develops a distribution-based portfolio selection framework to make use of the predicted joint return distributions. The proposed framework for portfolio selection can be viewed as an improved version of the classical

mean-variance portfolio analysis, with marginal return distributions modelled by quantile regressions replacing mean of returns, dependence structure modelled by copulas replacing covariance matrix of returns, the generalised Omega replacing quadratic preference and the threshold-accepting optimisation algorithm replacing quadratic programming. The portfolio selection is intuitively appealing and empirically implementable.

Chapter 5

Decision Trees for Return Prediction

5.1   An Introduction to CART

Classification and regression trees (CART) are non-parametric modelling techniques that essentially use recursive partitioning techniques to separate observations in a binary and sequential fashion. There are two varieties: (1) classification trees when the dependent variable is categorical and (2) regression trees when the dependent variable is continuous. We begin by introducing the standard tree terminology. The root is the top node which includes all observations in the learning sample. The splitting condition at each node is expressed as an "if-then-else" rule that is determined by a specific splitting criterion. The splitting node is also called parent and the two descendant sub-nodes are called children. A node keeps splitting until a terminal node or leaf is reached.

The fundamental idea behind CART is to recursively partition the space until

all the sub-spaces are sufficiently homogenous to apply simple models to them. This is in contrast to linear regressions which are global models where a single predictive formula is imposed over the entire data space. When the data set has multiple features which interact in complicated and nonlinear ways, as is often the case with financial data, a single global model may not adequately capture the underlying relations.

There are two major steps in the CART analysis: (1) build a tree using a recursive splitting of nodes and (2) prune the tree in order to obtain the optimal tree size so as to prevent over-fitting. Each of these two steps will be discussed in more details below. Breiman *et al.* (1984) provide a detailed overview of the theory and methodology of CART, including a number of examples from many disciplinary areas. There are also many software packages that implement the CART algorithm. Popular ones include R packages such as *rpart* and *tree* and the Matlab function *classregtree*.

5.1.1   Binary recursive partitioning

Let $\mathcal{L}$ be a learning sample, $\mathcal{L} = (x_1, y_1), \cdots, (x_n, y_n)$, where x_i is a vector of attributes, y_i is the response which can be categorical or continuous, and n is the number of observations. The attribute vector x_i belongs to $\mathcal{X}$, the attribute space. The tree building algorithm involves repeatedly splitting subsets of $\mathcal{L}$ into two descendant subsets, beginning with $\mathcal{L}$ itself. For a continuous variable x_i, the allowed splits are of the form $x_i < c$ versus $x_i \geq c$. For categorical variables the levels are divided into two classes. Therefore, for a categorical variable with K levels, there are $2^{K-1} - 1$ possible splits disallowing the empty split and ignoring
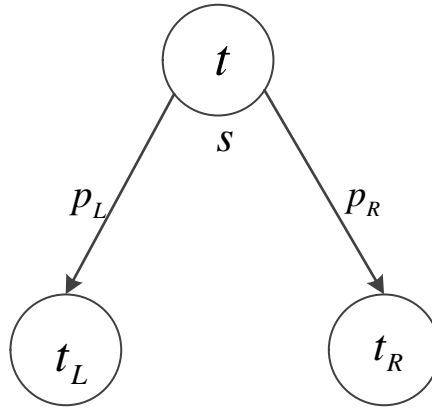
Figure 5.1: A split generates two children of the node t, denoted by t_L and t_R. A proportion p_L of the initial data go into the left child and a proportion of p_R go into the right child.

the order.

In choosing the best splitting rule, CART seeks to maximize the average purity of the two child nodes. Hence some criterion measuring data homogeneity or, alternatively, impurity should be introduced. These impurity measures are loosely classed splitting criteria. Let us introduce, for any node t, a measure $i(t)$ that signifies the impurity of the node. Suppose that a candidate split s divides the node into t_L and t_R such that a proportion p_L of the cases in t go into t_L and a proportion p_R go into t_R (see Figure 5.1). Then the goodness of the split is defined to be the decrease of impurity

$$\Delta i(s, t) = i(t) - p_L i(t_L) - p_R i(t_R).$$

For an arbitrary node t and a set of splitting candidates S, the optimal split is

chosen to be the one

$$s^* = \max_{s \subset S} \Delta i(s, t).$$

In other words, the optimal split is the one that reduces impurity by the greatest amount.

The idea for classification and regression trees is quite similar in terms of partitioning methods. Both are based on impurity reducing. However, they use different measures of impurity to decode the split.

In a classification problem, suppose that we want to classify data into K classes. At each node t of a classification tree we have a probability distribution $p_{tk}, k = 1, \cdots, K$, over all K categories. The probabilities are conventionally estimated from the node proportions, such that $p_{tk} = n_{tk}/n_t$, where n_{tk} is the number of observations in the k-th class, and n_t is the sample size at node t.

The two most common measures of impurity for classification trees are the Gini index

$$i(t) = \sum_{j \neq k} p_{tj} p_{tk} = 1 - \sum_k p_{tk}^2,$$

and entropy or information

$$i(t) = -\sum_k p_{tk} \log(p_{tk}),$$

where $0 \log(0) = 0$.

For regression trees, the most popular impurity measure is

$$i(t) = \sum_{j=1}^{n_t} (y_{tj} - \mu_t)^2,$$

99

where the constant μ_t for node t is estimated by average value of the training data falling into node t.

5.1.2 Tree pruning

However, the use of partitioning rules alone cannot guarantee a useful tree model. If reducing impurity is the only goal in tree induction, we will eventually end up with a maximal tree which has one observation or one class in each leaf, whichever reaches first. This kind of tree adapts too well to the features of the learning sample and has a very high risk of being over-fitted. Tree pruning is a way to improve the robustness of the model by trading off in-sample fitting against out-of-sample accuracy. This is particularly important if the model is being used to make predictions.

The best-known procedure for tree pruning is the cost-complexity pruning proposed by Breiman *et al.* (1984). Let T be a tree and its size be the number of terminal nodes. The optimal tree is the one which minimizes the following cost-complexity measure

$$R_\alpha(T) = R(T) + \alpha \, \text{size}(T),$$

where α is a complexity parameter to penalize tree size, and R is the cost which is commonly taken as misclassification errors in classification cases and deviance in regression cases. For a given value of the complexity parameter α, an optimal tree can be determined. In general, finding the optimal value for α would require an independent set of data, *i.e.*, a testing sample. This requirement is often avoided in practice by using a cross validation procedure.

5.2   CART versus Linear Weighting Approaches

Influenced by classical asset pricing theories such as the Capital Asset Pricing Model (CAPM)) (Sharpe, 1964; Lintner, 1965) and the Arbitrage Pricing Theory (APT) (Ross, 1976), the relationship between stock returns and various risk factors are conventionally taken to be linear. The weaknesses of such a linear weighting approach of mapping returns to risk factors include:

- The classical regression model operates under the assumption that the data follows a Gaussian distribution. However, it is now widely recognised that stock returns are not always normally distributed, particularly over shorter-term time horizons, and can display both fat tails and skewness (see Section 1.1.2);

- It is assumed that stock returns respond in a linear fashion to a change in a predictor variable. There is actually no compelling theoretical reason to believe this is the case, and empirical observation suggests that this assumption is often violated in reality (see Section 1.1.3);

- Such models can be distorted by multicollinearity, outliers and missing values in the data, typical issues for financial data sets. Furthermore, the linear framework is not particularly efficient at identifying the interaction between important predictors, particularly when the data set is noisy.

Compared to the various linear weighting type of approaches, CART offers a number of benefits in data exploration. In particular, it has a very high degree of interpretability. CART efficiently compresses a large volume of data into an easy to understand graphical form which identifies the essential characteristics.

Being non-parametric CART does not require any assumptions to be made about the underlying distribution of the variables being modelled. It is well suited to identifying any nonlinearities and complex interactions in the data. In contrast to parametric models, CART is computationally fast at selecting predictors and particularly efficient at capturing their interactions. It is less affected by missing values and multicollinearity. Furthermore, CART is quite robust in the presence of outliers and well suited to noisy data sets, both of which tend to be features of financial data.

The CART approach also departs from traditional modelling methods by determining a hierarchy of input variables which may be closer to the human decision-making processes. Indeed, a key strength of CART over the classical modelling methods is that it allows one to represent various types of interactions between variables, particularly conditional relevance (see, for example, Van Der Smagt and Lucardie, 1991). Conditional relevance occurs if a factor is relevant only when it is conditioned upon some other factor. For example, only if a certain condition is met by the first high-level attribute is a second attribute taken into consideration. The same holds for the next attribute in the tree hierarchy, and so on.

The weaknesses of the CART modelling technique will be discussed in Chapter 6 together with a remedy. Using the North American stock data, the rest of the chapter aims to provide an evaluation of CART for cross-sectional stock return forecasting and compares it with the linear regression approaches.

5.3 Data Description

We use monthly stock data from 1986:12 to 2010:08 for the North American stock markets including US and Canada but excluding financial stocks as defined by the Global Industry Classification Standard (GICS) [1]. In order to avoid the over representation of potentially less liquid companies, to be included in the universe each stock must have also had a market capitalisation in excess of \$1bn in 2010 and its equivalent historically. The data are constructed from a variety of vendors including Exshare, Worldscope and the Institutional Brokers Estimate System (I/B/E/S). Using cross-sectional data stacked monthly, the number of total observations in the panel data set amounts to 279,188 (or 980 stocks per month on average).

At the end of each month, forward total stock returns including dividends are calculated. Using the median return of all sample companies in the same period as a benchmark return, the excess returns are then computed as the total returns minus the benchmark returns.

In deciding which stock characteristics to include as possible determinants of future returns, attention is given to those variables that accord with investment intuition and have been found to be important in prior studies. We focus on stock characteristics from the following broad categories:

- **Value factors**: These factors are measures of firm value, such as dividends, earnings, book value and cash flows. A long list of academic literature documents the value phenomenon. Chapter 2 has listed some references

---

[1]Financial stocks are excluded due to their different accounting structure which makes comparisons with non-financials troublesome, although similarly structured stock selection models can also be applied within the sector.

to prior studies which claim ratios such as dividends-to-price, earnings-to-price and book-to-price may be associated with systematic variation in stock returns. Lakonishok *et al.* (1994) find that firms' cash flow streams may also be associated with systematic variation in stock returns.

- **Profitability factors**: This group of factors include ratios such as return-on-assets, return-on-equity, cash return-on-equity, pre-tax margins and asset turnover ratio. Literature documenting predicting power of profitability factors in subsequent stock returns includes Dechow *et al.* (2001), Campbell and Thompson (2008) and Chen *et al.* (2011).

- **Financial strength factors**: These factors measure debt attainability of a firm. They include measures on a firm's level of debt, such as debt-to-equity ratio and debt-to-market capitalization ratio, and measures on a firm's ability to service its debt, such as interest cover and free cash flow-to-debt ratio. There are many empirical studies focusing on firm financial strength and stock performance, such as Bhandari (1988), Barbee Jr *et al.* (1996) and Campbell *et al.* (2008).

- **Momentum factors**: Momentum factors refer to stocks's past return performance. The momentum effects are well documented in the finance literature. As early as in 1980s, De Bondt and Thaler (1985, 1987) demonstrate that stock returns over three to five years have explanatory power over future returns. Later, Jegadeesh and Titman (1993, 2001) find that returns over three to twelve months also have predictive power over future returns. In the industry, momentum strategy enjoys vast popularity.

- **Analyst forecast factors**: These factors are constructed from the I/B/E/S database. This group of factors include consensus forecasts of analysts and revisions in the analyst forecasts. Analyst forecasts have been studies in the finance and accounting literature and have found to relate to stock returns (see, for example, Trueman, 1994). Analyst forecasts revisions, especially earnings revisions have drawn considerable attention in the literature. Studies carried out by scholars such as Givoly and Lakonishok (1980) and Lys and Sohn (1990) show that an investor who acts upon analysts' earnings revisions can consistently outperform a buy-and-hold policy after transaction costs.

A broad spectrum of stock characteristics from the categories above are selected, as reported in Table 5.1. Instead of using raw values, we use rank orders in order to improve the robustness of the analysis. At each month, the rank order for each variable are computed by firstly ranking all n stocks available that month according to the corresponding variable value. Then the resulting rank is divided by n to scale it between 0 and 1. It is well known that a lot of the financial ratios are highly correlated which makes data analysis and statistical inference a challenge. In order to overcome the high correlation in financial variables, nine composite factors are promoted as potential explanatory variables which are constructed as an equally weighted average of the underlying variables. Details of the composite factor construction are also described in Table 5.1.

Table 5.2 illustrates the spearman rank correlation matrix for the composite factors over the whole sample period. It is clear that the correlations among the composites are quite low with most of the correlations between -0.2 to 0.2 and the largest value being 0.41 (HIST.GROWTH and PROF). Table 5.2 also highlights

Table 5.1: This table lists the composite factors for return prediction. The left column displays the composite factors with their abbreviations given in parentheses. The right column reports the method and the variables used to form composite factors.

Composite factor	Description
Value (VAL)	An equally weighted average of dividends-to-price ratio, cash flows-to-price ratio, sales-to-price and book-to-price ratios.
Profitability (PROF)	An equally weighted average of return-on-assets ratio, return-on-equity ratio, cash return-on-equity ratio, pre-tax margins and asset turnover.
Leverage (LEVERAGE)	An equally weighted average of debt-to-equity ratio and debt-to-market capitalization ratio.
Debt Service (DEBT.SERVICE)	An equally weighted average of interest cover and free cash flows-to-debt ratio.
Momentum (MOM)	An equally weighted average of 6-month return momentum and 12-month return momentum.
Stability[a] (STAB)	An equally weighted average of earnings, sales and cash flows stability over the previous 5 years.
Historic Growth[b] (HIST.GROWTH)	An equally weighted average of 3 year historic growth in earnings, sales and cash flows.
Forward Growth[c] (FWD.GROWTH)	An equally weighted average of I/B/E/S forecasted earnings growth expectation for FY1 and FY2.
Earnings Revisions[d] (EREV)	An equally weighted average of the 3 month change in I/B/E/S forecasted earnings expectations for FY1 and FY2.

[a] Sales, earnings and cash flows stabilities are measured by the t-statistic of the slope coefficient of a regression of previous five historic sales, earnings and cash flows on the same month on the number of years. Therefore, stability measures the strength of the linear trend in the underlying time-series.

[b] Historic growth in sales, earnings or cash flows is measured by the percentage change of current sales, earnings or cash flows over that of three years before.

[c] Forward growth is the percentage change of the IBES consensus earnings expectation (either for Financial Year1 or Financial Year 2) over the current earnings.

[d] Earnings Revisions (EREV) is defined as the number of analysts in IBES adjusting earnings forecast up minus the number of analysts adjusting earnings forecast down scaled by the total number of analysts following a firm in IBES.

Table 5.2: This table reports the spearman rank correlation matrix for the composite factors over the whole sample period (December 1986 – August 2010). It is clear that the correlation values among the composites are quite low with most of the correlations between -0.2 to 0.2 and the largest value being 0.41 (HIST.GROWTH and PROF).

	VAL	PROF	STAB	LEVE-RAGE	DEBT. SERVICE	MOM	FWD. GROWTH	HIST. GROWTH	EREV
VAL	1	0.03	0.02	-0.37	-0.01	-0.15	-0.04	-0.18	-0.12
PROF	0.03	1	0.38	-0.37	0.02	0.14	0.02	0.41	0.03
STAB	0.02	0.38	1	-0.2	0.03	0.15	0.02	0.21	-0.01
LEVERAGE	0.37	-0.37	-0.2	1	-0.01	-0.19	-0.01	-0.16	-0.07
DEBT.SERVICE	-0.01	0.02	0.03	-0.01	1	-0.01	0	-0.01	-0.19
MOM	-0.15	0.14	0.15	-0.19	-0.01	1	-0.11	0.05	0.18
FWD.GROWTH	-0.04	0.02	0.02	-0.01	0	-0.11	1	0.02	-0.14
HIST.GROWTH	-0.18	0.41	0.21	-0.16	-0.01	0.05	0.02	1	0.04
EREV	-0.12	0.03	-0.01	-0.07	-0.19	0.18	-0.14	0.04	1

Table 5.3: Univariate analysis over the period December 1986 to August 2010. "t-stat" is the t-statistic of the slope coefficient of a univariate regression of monthly returns on a lagged composite factor. Annualised portfolio returns are reported over the whole period. Portfolios are re-balanced monthly and transaction costs are not taken into account.

Composite factor	t-stat	Annualised Return (%)			
		High	Medium	Low	Hi-Lo
VAL	12.73	16.09	13.64	9.77	5.93
PROF	8.09	14.62	13.47	11.23	3.47
EREV	7.36	15.92	11.98	11.71	4.52
MOM	5.53	12.95	14.63	11.53	1.75
LEVERAGE	4.85	14.82	12.86	11.82	2.56
STAB	1.95	13.52	12.38	13.75	0.76
FWD.GROWTH	-0.12	13.41	12.17	13.97	-0.59
HIST.GROWTH	-0.49	13.79	12.40	12.60	-0.25
DEBT.SERVICE	-1.38	12.49	13.59	13.55	-0.62

some key cross-sectional relationships. Cheap companies, as determined by the Value composite, tend to be characterised by some evidence of distress. For example, they are positively correlated with leverage and also tend to suffer from poor sentiment (negative price momentum and earnings revisions) as well as low historic growth. This is in line with the rationalist approach of Fama and French (1996) who suggest the value premium is a reward for taking on additional risk (*i.e.*, they are more leveraged). However, it is not inconsistent with the more behaviourally focused suggestion posited by Lakonishok *et al.* (1994) that the premium is instead an artefact of systematic misvaluation caused by investors extrapolating recent news flow.

To gain an insight into the performance of the individual composite factor, we employ univariate analysis. More specifically, we regress one month forward

excess returns on each individual composite factor respectively and report the t-statistic of the slope coefficient. We also partition the composite factors into three equally sized portfolios of stocks, thereby forming a "high", "medium" and a "low" portfolio each month. The ordering is such that the "high" portfolio is typically associated a priori with a positive outcome (*e.g.*, cheap, strong positive revisions, high stability, strong growth, high debt service etc) although for the leverage composite factor, a "high" value actually implies poor financial strength (*e.g.*, high leverage).

Table 5.3 shows the univariate t-statistic and the annualised returns to each of the three portfolios over the entire period, as well as the High-Low portfolio return. The univariate data analysis suggests that Value, Profitability and Earnings Revisions are the most significant determinants of forward returns followed by the historic momentum and leverage. Whilst they are not statistically significant, one interpretation of the modest negative returns to high historic and forward growth is that investors tend to overpay for growth and are subsequently disappointed on average when high growth rates prove unsustainable Haugen (2009). We would also highlight the positive return to leverage which suggests that, historically, the more highly leveraged companies have actually outperformed. This is not a disguised banking effect as financial stocks are excluded from this data set.

5.4 Results: Linear Weighting Approaches

Our purpose is to promote the benefits of the decision tree methodology compared to more traditional modelling approaches. As such, we initially build two versions of a linear combination of composite factors in order to provide benchmarks

against which to compare the CART model with. These models are built over a subset of the full historical period available (December 1986 to April 2007) in order to provide a reasonable period for out-of-sample testing. The end-date is also chosen deliberately as we are particularly interested in monitoring the performance of the competing models during the quant shock of mid-2007 and the subsequent Global Financial Crisis.

The first set of linear weights is derived from a multi-factor regression of monthly excess returns on the lagged composite factors. The coefficients from this regression are then converted into a weight vector which is used to rank all stocks considered. As shown in the first column of Table 5.4 and consistent with the univariate analysis results (Table 5.3), the regression based weighting scheme favours Value, Profitability, Earnings Revisions, Momentum and Leverage, whilst it penalises companies with high stability or high historic growth.

The second version of the linear combination of factors takes into account the historic return and volatility of each composite factor by using standard mean-variance analysis to create an optimal weighting scheme. Table 5.4 reports the results. As with the linear regression model above, this approach also favours the high return factors (Value, Earnings Revisions and Momentum) but places a greater emphasis on Stability at the expense of Profitability as the latter has historically been more volatile. As Momentum has historically been lowly correlated with Value, its weight in the mean-variance framework is also boosted.

Whilst the precise choice of factors and optimal weighting schemes vary considerably in practice, the dominant emphasis upon both the Valuation and Momentum signals in the two linear weighting stock selection models is consistent with the traditional focus of many quantitative managers. In essence, it is rela-

Table 5.4: This table reports factor weights from two linear weighting schemes using the data from December 1986 to April 2007.

Composite factor	Regression Based Weights	Mean-Variance Optimised Weights
VAL	26.6%	47.0%
PROF	24.1%	10.7%
EREV	21.7%	28.1%
MOM	16.3%	37.5%
LEVERAGE	14.2%	-8.0%
DEBT.SERVICE	1.4%	-13.6%
STAB	-2.8%	12.8%
FWD.GROWTH	4.7%	-6.3%
HIST.GROWTH	-6.2%	-8.2%

tively easy to generate a Value and Momentum themed stock selection model in practice.

5.5 Results: CART Model

Using exactly the same inputs as the linear models above, we then build a classification tree with the purpose of predicting subsequent stock performance. Stocks are sorted into two groups, "outperformers" for those with positive excess returns and "underperformers" for the remainder. The induced categorical variable is then used as the dependant variable in the subsequent modelling process. One of the benefits of working with categorical responses instead of raw returns lies in the fact that it alleviates the impact of extreme returns that may have. As with the linear alternatives, the tree model is built with data up to and including April 2007 whilst the data between May 2007 and August 2010 are reserved for out-of-sample testing. By way of example, Figure 5.2 graphically illustrates the
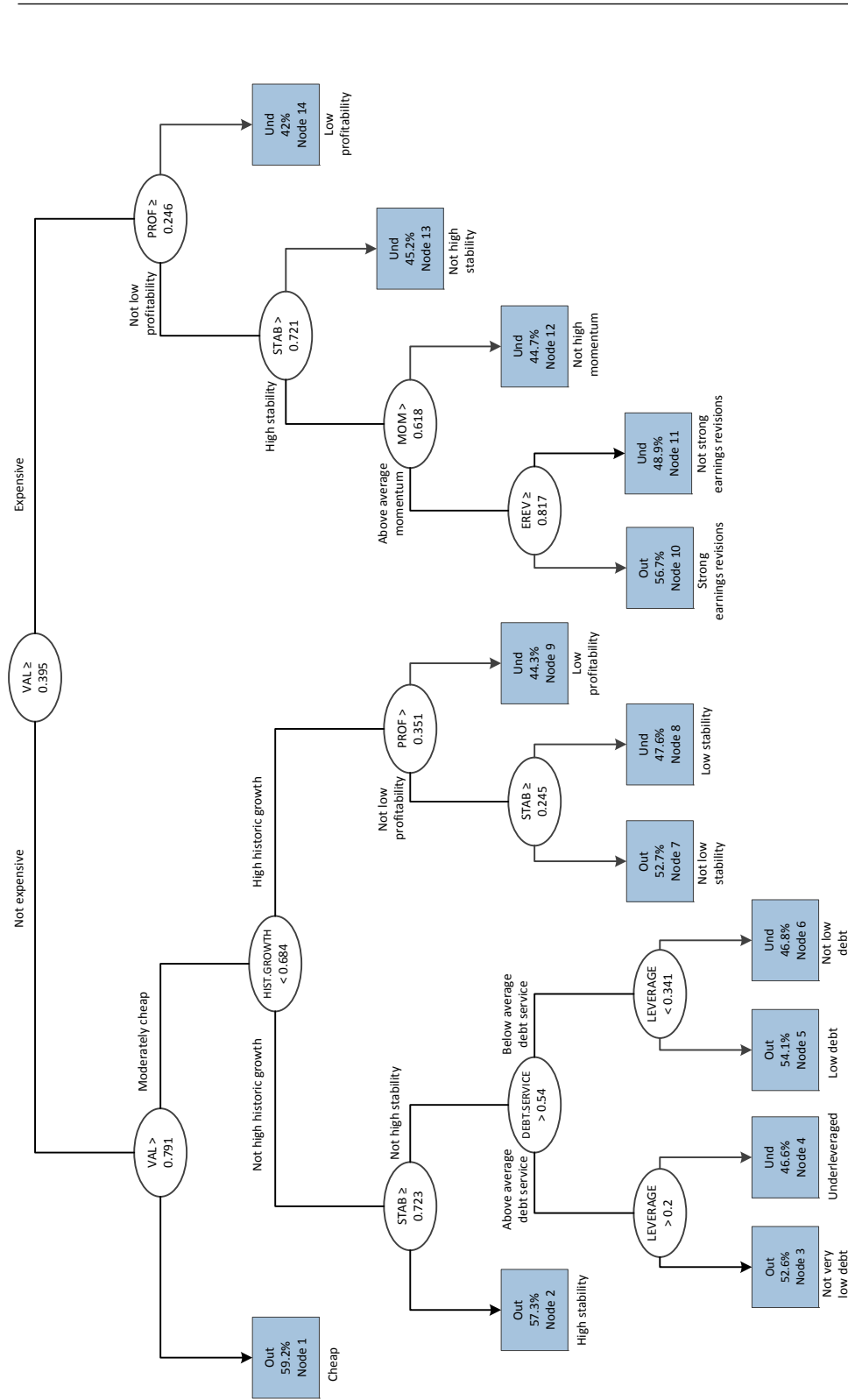
Figure 5.2: Decision Tree for North American stocks built using data from December 1986 to April 2007 to model the chance of a stock outperforming the benchmark. The dependent variable is set as an "outperformer" (Out) if a stock subsequently achieves a higher return than the market, and "underperformer" (Und) otherwise. The outperforming probabilities are reported in percentage at each terminal node along with the splitting criteria.

hierarchical structure of a stock selection tree built at April 2007.

The first observation to note is that the primary split is on Value and, more specifically, the distinction between those stocks that are relatively expensive (the right hand branch) and the not-expensive stocks. One of the most attractive nodes splits again on high Value and therefore identifies cheap stocks as having a 59.2% probability of outperforming the benchmark (Node 1). In contrast, the worst performing stocks are characterised by being expensive and exhibiting low profitability (Node 14). Companies with these attributes only have a 42% chance of outperforming. Similarly, Node 13 suggests that poor stability is also a reason to penalise expensive stocks even if their profitability is not particularly weak. Technology stocks make up almost a third of this node.

As a benefit of identifying conditional relevance, the tree is able to distinguish the exception to the rule. For example, whilst both of the linear weighting approaches indicate that Value is the most important driver of stock returns, the tree model suggests that stocks which are not cheap still have a good chance of outperforming the benchmark providing that they are blessed with profitability, stability in earnings, strong momentum and are also associated with strong earnings revisions (Node 10).

The decision tree framework also highlights the nonlinear behavior of the stock returns to the underlying predictor variables. For example, stocks in Nodes 3 and 5 have similar outperforming probabilities but are of opposite preference with regard to leverage. Conditional on above-average debt cover, Node 3 actually prefers some degree of leverage and more significantly penalizes overly conservative firms (with too low leverage). Recovery stocks within the most cyclical area of the market such as consumer discretionary, industries and material comprise

Table 5.5: This table reports the Spearman rank correlation of model outputs out-of-sample (May 2007 - August 2010).

	Regression	Mean-Var	CART
Regression	1	0.83	0.56
Mean-Var	0.83	1	0.57
CART	0.56	0.57	1

almost half of Node 3. In contrast, leverage is a characteristic to be avoided amongst firms that cannot service their debts (Node 5).

5.6 Results: Model Comparison

The effectiveness of the three alternate approaches in terms of their stock selection abilities is then assessed out-of-sample. Specifically, we rank all stocks each month based upon the fitted values implied by the three models from May 2007 until August 2010, thereby deliberately covering a historical period that has been associated with relatively poor performance for most quantitative managers. Table 5.5 reports the rank correlation of the model predictions during this period. The rank correlation between the two linear weighting approaches is relatively high at 0.83. In contrast, the predictions generated by the tree model offer clear diversification benefits, exhibiting a much lower rank correlation of 0.56 and 0.57 with the two linear weighting approaches.

Next, we use portfolio strategy to compare the models out-of-sample. This is assessed by forming two portfolios by splitting the stocks equally on the predicted outperformance probabilities each month: a "long" portfolio (those expected to outperform) and a "short" portfolio (those expected to underperform). We report

the following performance measures.

- **Time series hit rate**: It is computed as $\frac{T^+}{T}$, where T^+ is the number of months with above benchmark returns in the out-of-sample period, and T is the number of months in the out-of-sample period. Time series hit rate is the proportion of months that the portfolio outperformed the benchmark.

- **Excess return**: It refers to annualised portfolio excess return over the benchmark,

$$r_e = \left[(\Pi_{t=1}^T(1 + r_{pt}) - \Pi_{t=1}^T(1 + r_{bt}))^{1/T} \right]^{12/freq},$$

where r_{pt} and r_{bt} are the returns of a portfolio and the benchmark at time t, and $\Pi_{t=1}^T(1 + r_p t)$ and $\Pi_{t=1)}^T(1 + r_{bt})$ are the cumulative returns over T periods for the portfolio and the benchmark, respectively. The power term $12/freq$ is to annualise return, where $freq$ is the frequency in the month that returns are recorded. In our case, return is in monthly frequency, therefore, $freq = 1$. In the case of quarterly return, we have $freq = 3$, and so on.

- **Tracking error**: It is annualised and computed as

$$\sigma = \sqrt{\frac{12}{freq}} sd(r_{pt} - r_{bt}),$$

where, sd is the standard deviation and, again, $freq$ is the frequency in the month that returns are recorded.

- **information ratio**: It is the annualised mean of the return difference

between a portfolio and the benchmark divided by the annualised tracking error as defined above,

$$IR = \frac{12/freq \sum_{t=1}^{T}(r_{pt} - r_{bt})}{\sqrt{\frac{12}{freq}} sd(r_{pt} - r_{bt})} = \sqrt{\frac{12}{freq}} \frac{\sum_{t=1}^{T}(r_{pt} - r_{bt})}{sd(r_{pt} - r_{bt})}.$$

• **Stock holding period**: Stock holding period is a measure of portfolio turnover. The longer the holding period is, the lower are the portfolio turnover and the transaction costs. A stock's holding period in a portfolio is defined as the average holding periods in the backtest. For example, stock A stays in a portfolio for three consecutive periods, and then drops out. Later it comes back into the portfolio for another two consecutive periods. The holding period for stock A in the portfolio is taken to be $h = 2.5$, the average of its holding periods in the portfolio. For any portfolio, let n be the number of stocks ever entered into that portfolio at backtest, and $h_1, h_2, \cdots, h_n$ be the corresponding holding period series. We report the mean and median of the holding period based on the portfolio holding period series.

Table 5.6 reports the annualised excess return (the benchmark chosen as the median return of all stocks considered), the tracking error, the information ratio, time series hit rate (TS hit rate) and portfolio holding period information of the various strategies. Both of the "long" linear weighting schemes outperform the benchmark slightly in the out-of-sample period before transaction costs, albeit modestly (0.9% and 1.1%) whilst the long portfolios derived from the tree model actually perform relatively well, outperforming by 2.6% with similar relative risk.

Table 5.6: This table reports performance of three stock ranking models in the out-of-sample period (May 2007 – August 2010). At each month, two portfolios are formed by splitting the stocks equally based on the model outputs: a "long" portfolio (those expected to outperform) and a "short" portfolio (those expected to underperform). Portfolio performance measures reported are annualised excess returns, tracking errors, the information ratios (IR), time series hit rate (TS hit rate) and stock holding period. Portfolios are re-balanced monthly and transaction costs are not taken into account.

Portfolio	Model	Excess Return(%)	Tracking Error (%)	IR	TS hit rate	Holding Period	
						Mean	Median
Long	Regression	0.9	2.3	0.40	0.46	9.1	6.7
	Mean-Var	1.1	3.1	0.37	0.46	9.6	5.0
	CART	2.6	2.9	0.89	0.57	10.1	6.0
Short	Regression	-0.9	2.5	-0.38	0.53	5.1	5.7
	Mean-Var	-1.2	3.1	-0.37	0.54	6.6	5.0
	CART	-2.8	3.4	-0.82	0.43	7.5	4.5

The same outcome is also identified when comparing the performance of the short portfolios.

The tree model is slightly more consistent in its outperformance with a time series hit rate of 57% whilst the success rates of the linear weighting approaches are both less than 50% suggesting positive but erratic returns. Furthermore, the holding periods of the portfolios derived from the tree model are no shorter than those of the portfolios by the multi-factor regression and the mean-variance analysis. This suggests the good performance of the tree model is not achieved though higher turnover.

A closer inspection of the difference between the tree-based portfolios and the stock selection implied by the two linear weighting approaches suggests that the improved performance of the tree model arises from the capture of the nonlinear

relationship between debt sustainability and leverage during the Global Financial Crisis. The multi-factor predictive regression extrapolates the historic relationship prior to 2007 between higher leverage and stock outperformance whilst the mean-variance optimised weights simply penalised companies that are more capable of servicing their debt. In contrast, the tree model is more specific in identifying the nonlinear interaction between leverage and debt service. Specifically, leveraged companies are preferred only if the firm has a sufficient level of profitability that it is capable of servicing its debt, but if this is not the case then leveraged companies are specifically avoided.

Although we have only focused upon a particular and a rather unusual period of history, the out-of-sample performance for the tree model since 2007 is encouraging compared to the linear weighting approaches and suggests that there is a role for non-traditional stock selection models, if only to help diversify model risk. Our research using sector specific models suggests that the tree models are not unambiguously superior in every period to a linear alternative. However, in all cases they offer a high degree of diversification from standard modelling approaches.

5.7   Summary

The identification of the common factors in stock performance has historically been the domain of linear modelling approaches despite a growing awareness of the nonlinearities in market returns. The non-parametric and nonlinear modelling approach CART provides a convenient way to explore nonlinear dynamics of the stock returns. We have observed that when applied to the North American stock

data, CART approach can generate a very different model from the traditional approaches. It identifies nonlinear structural relationships which are intuitive.

For asset management in the real world, the widespread use of linear modelling methodologies among quantitative asset managers, taken together with the similarity in data sources and risk models may in turn have contributed towards model risk in financial markets. This leads to a high degree of commonality in investment decisions. As a less used technique, CART is appealing in the context of potentially offering a degree of model diversification.

Chapter 6

A Hybrid Approach for Return Prediction

6.1 Motivations

This chapter continues the investigation of the use of CART for cross-sectional stock return prediction from the previous chapter. In this chapter, we discuss the major weaknesses of CART and propose a hybrid model to combining CART with a linear model, logistic regression, for improving model performance.

Chapter 5 demonstrates the effectiveness of CART for stock return prediction. This nonlinear and non-parametric model enjoys various benefits over the traditional linear weighting approaches. However, the use of CART is not without critics.

The major criticism against CART lies in the recursive nature of the tree building process. Local optimisation at each step in the sequential node-splitting process does not guarantee global optimisation of the overall tree structure. In
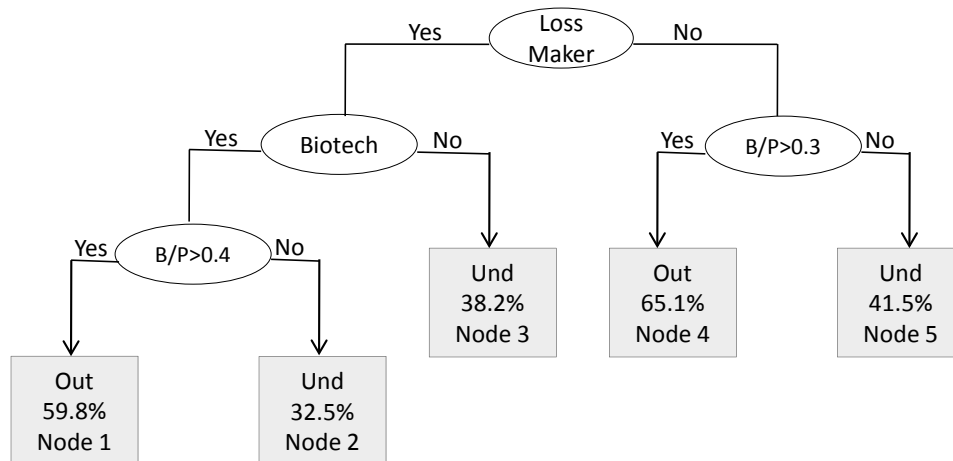
Figure 6.1: A hypothetical example of a classification tree for stocks in health care sector. OutPerf stands for outperforming and UndPerf underperforming. B/P is the book-to-price ratio.

other words, when determining the hierarchy of rules that form the tree, the algorithm is not aware of the nature of branches further down the tree, including the terminal nodes. The resulting tree structure therefore does not guarantee global optimisation. Several alternatives to CART have developed to address these problems, such as random forests (Breiman, 2001).

Discretization of continuous variables may be another possible problem in CART solutions. We illustrate this point using a hypothetic stock prediction example. As shown in Figure 6.1, the chance of a stock generating a return over and above a relevant benchmark in a specific period (*e.g.*, three months from now) is determined by the joint effects of whether the company is currently making negative earnings (Loss Maker), whether it is a biotechnology company (Biotech), and the value of its book-to-price ratio (B/P). The majority vote is

endorsed as the final predicted class, as listed in the terminal nodes. Displayed with the predicted classes are the proportions of outperformers in each terminal node. This proportions are used as a proxy for probability of outperforming.

In this example, B/P is the only continuous input variable and all others are categorical. CART discretizes the continuous variable B/P according to two different thresholds (0.4 and 0.3) at two splitting nodes. Two consequences follow. First, the responses are not sensitive to B/P changes within a node, as all members in a node share the same outperforming probability regardless of variation in B/P. This leads to a noticeable characteristic of the tree model which is the discontinuous resultant probabilities. The five-node tree can only produce five different probabilities. Second, it causes over-sensitiveness of responses to a continuous variable close to a boundary. In our example, it means a small change in the value of B/P can lead to a unproportionately large change in response. For example, conditional on the company not being a loss maker, when the value of B/P changes from 0.301 to 0.3, the probability of a stock outperforming the benchmark falls from 0.65 to 0.41 and leads to a downgrade from out-performer to under-performer. While dicretization of continuous variables and assigning the same output to all members in a node is parsimonious; on the other hand, it could oversimplify the complexity of the real data. The resulting abrupt shift from one node to the other may not always be realistic.

The parametric linear counterpart of CART is logistic regression. Belonging to the generalised linear model family, logistic regression has been used extensively in economics and finance, most commonly when forecasting a specific event such as bankruptcy or the probability of default (see, for example, Ohlson, 1980). In contrast to CART, the logistic regression is highly effective at capturing any

122

global features of a data set. It also allows a continuous variable to influence outcomes continuously and, therefore, produces a smooth response surface. A small change in a predictor variable yields a small change in the predicted probability. However, being a linear technique, logistic regression shares the usual weakness of the classical modelling approach. Specifically, it requires a valid mean function assumption, is affected by multicollinearity and is sensitive to outliers and missing data.

While CART and logistic regression are traditionally regarded as competitors for modelling data, they are essentially complementary to each other after noting the strengths and weaknesses associated with these two models. It is natural to combine these two approaches in order to better serve the purpose of predicting future stock returns. A desirable modelling outcome would, therefore, be to uncover the nonlinearities in the data set whilst still producing a relatively smooth probability surface in a globally optimal model. Accordingly, a hybrid approach to combining CART and logistic regression introduced in the following section can be a way to achieve the goal.

6.2 The Hybrid Approach: Combining CART and Logistic Regression

The first step is to apply CART to uncover the high-order interactions in data. Suppose that there are n stocks that require classification. Let X be the matrix of continuous attributes of the stocks which is of dimension $n \times m_1$. The j-th column $X_{\cdot j}$, a vector of length n, is the j-th continuous attribute of all the stocks, such as

book to price ratio (B/P), and the i-th row $X_{i.}$, a vector of length m_1, records all the continuous attributes for the i-th stock. Let D_j be the corresponding discrete representation of $X_{.j}$, where $j = 1, 2, \cdots, m_1$. The collection of all variables, denoted as D, consists of m_1 members which are discretized continuous variables, and m_2 members which are discrete by nature, such as industry membership. Assume that the optimal tree built on D is of K terminal nodes with n_k stocks in the k-th terminal node, $k = 1, 2, \cdots, K$. Let s_k be the number of outperformers in the node k. The relative frequency of outperformers, $p_k = s_k/n_k$, is assigned to all n_k members in the k-th node as their probability to outperform a benchmark. The probability p_k can be regarded as a consensus belief which is voted by all n_k members in the node. For the i-th stock in the k-th node, a more realistic probability model is to incorporate a perturbation into the consensus probability,

$$p_{ki} = p_k + \epsilon_{ki}, i = 1, 2, \cdots, n_k,$$

where ϵ_{ki} is a stock-specific effect acknowledging the difference within members.

The above discussion naturally leads to the second step, fine-tunning p_k for each individual stock to get p_{ki}. We model ϵ_{ki} through logistic regression. In this step, only the continuous attributes are used. In the scale of the logit transformation, we have

$$\text{logit}(p_{ki}) = \text{logit}(p_k) + X^k_{(i.)}\boldsymbol{\beta}, \tag{6.1}$$

where $X^k_{i.}$ denotes the i-th row in X^k, the matrix of continuous attributes of the n_k stocks in the k-th terminal node, and $\boldsymbol{\beta}$ is a vector of coefficients with length m_1. In the case that $X^k_{i.}\boldsymbol{\beta}$ is negligible, meaning that finer scales do not provide more information than that is captured by p_k, the adjusted probability remains

as p_k – the initial probability by CART. There are two possible ways to proceed from here.

- Fit a terminal-node specific logistic model. In this case, $\boldsymbol{\beta}$ in the logistic regression (6.1) is actually $\boldsymbol{\beta}_k$ as it is specific to the terminal node in question. The approach requires a large volume of data in order to cope with the segmentation but is more appropriate when working with data that incorporates structural breaks.

- Fit one universal logistic model using all the data whilst imposing the tree model from the first step. Compared with the first approach, this one may compromise terminal node-specific features with globally dominant ones but is likely to be more robust as a result.

In the remainder of this chapter, we illustrate the hybrid approach using the second method above to ensure that any global features in the data set not captured by the initial CART model have an opportunity to be incorporated. Computationally, the parameter $\boldsymbol{\beta}$ in the logistic regression (6.1) is estimated by including the logit transformation of p_k, $\mathrm{logit}(p_k)$, as an offset in a logistic regression, meaning that the coefficient of $\mathrm{logit}(p_k)$ in the regression is constrained to unity.

6.3 Data Description

We test the proposed method using the stock data for the North American markets. The details of the data construction and the input variables have already been described in Section 5.3. In order to enable an extensive backtest, we con-

sider the defensive stocks in the North American markets. The defensive stocks comprise four GICS sectors, namely consumer staples, health care, telecommunication services and utilities. To further reduce the sample size to a manageable level, we use quarterly instead of monthly stock data. The quarterly data are from 1986:12 to 2010:06, and three-month-forward stock returns including dividends are calculated for each stock in the data set at the end of March, June, September and December of each year. The sample consists of an average of 295 companies per quarter and 27,994 observations in total.

For each period, the quarterly stock returns are compared with the median return of all sample companies in that period to classify the stocks into two groups, "outperformers" for those with returns above the median return and "underperformers" for the rest. This induced categorical variable is then used as the dependant variable in the subsequent modelling process. Essentially, the model is forecasting the probability of a stock outperforming its peer group in the subsequent three month period. As mentioned in Chapter 5, working with a categorical dependent variable rather than raw returns alleviaters impacts of outliers and, therefore, improves model robustness.

6.4   Backtest Setup

We use data from 1986:12 to 2000:09 as the initial estimation sample, and retain the period from 2000:12 to 2010:06 for the out-of-sample evaluation. We use recursive estimation starting with the data from 1986:12 up to the time of the forecast in order to generate a series of one-quarter-ahead forecasts. We then build several portfolios based on these forecasts and track their performance relative to the

benchmark. Specifically, three portfolios are formed such that the first portfolio (P1) contains the third of stocks with the lowest outperforming probabilities (*i.e.*, the underperforming portfolio), the second portfolio (P2) contains the next third (*i.e.*, the market portfolio) and the final portfolio (P3) contains the remaining third of stocks with the highest probability of outperforming as derived from the model. We use an equal weighting scheme for portfolio construction. Portfolios are re-balanced quarterly and profits are reinvested. For simplicity, transaction costs and taxes are not considered for portfolio performance evaluation.

The first and the third portfolios (P1 and P3) are natural choices for short and long portfolios, respectively, and are therefore of special interest to us. We examine the risk-adjusted excess returns of these portfolios. We use two sets of risk factors: the Fama-French (FF) three factors (Fama and French, 1993) which are factors related to market (MKT), value (HML) and size (SMB), and Carhart's four factors, which are the FF factors plus an additional momentum factor (MOM, Carhart, 1997). The data for the risk factors are sourced from Kenneth French's website [1]. More specifically, we carry out linear regression analysis of portfolio returns on the risk factors

$$r_p - r_B = \alpha + \sum_{m=1}^{L} \lambda_m \beta_m,$$

where r_p is the portfolio return, r_B is the benchmark return, β_m is the loading of portfolio p on factor m, λ_m is the risk premium associated with factor m, and L takes value either three or four depending on the risk model used. The intercept, α, from the risk analysis is the risk-adjusted return. A significant α indicates

---

[1] $http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.$

skills in portfolio construction.

As it is of particular interest as to whether the CART-Logistic hybrid approach is superior to either a standalone CART or a logistic regression when estimated separately, we also compare the out-of-sample performance of the proposed hybrid approach to these simpler alternatives that nevertheless form the building blocks to the hybrid model. Furthermore, we also incorporate a second tree-based technique known as random forest which is an ensemble approach proposed by Breiman (2001) that overcomes the weaknesses of CART listed in Section 6.1. Instead of using one decision tree, the random forest approach grows a collection of trees (the forest) and the final output is derived from averaging across the output of the individual trees. The details of the random forest approach is given in Appendix C.

To facilitate the model comparison, using the same principle as the formation of portfolios on the hybrid model outputs, three portfolios are also formed for each competing model at each out-of-sample period. We then compare portfolio performance by different models. All the portfolio performance measures listed in Section 5.6 are used, which are the annualised excess returns of the portfolios, cross-sectional and time series hit rates, information ratio as well as the holding period. However, the value for $freq$ takes 3 instead of 1 as we use quarterly data here. In addition, we also report the panel hit rate:

- **Panel hit rate**: It is a time series average of cross-sectional hit rate, defined as

$$\frac{1}{T}\sum_{t=1}^{T}\frac{N_t^+}{N_t},$$

  where N_t is the number of stock in a portfolio at quarter t, N_t^+ is the number

of outperforming stocks in that portfolio, and T is the number of quarters in the out-of-sample period.

6.5 Results: Hybrid Model

At the end of each quarter, the nine composite factors are bucketed into three equally sized bins. A CART model is built using the data from 1986:12 to 2000:09 and subsequent models are built at the end of each quarter up to 2010 using an expanding window of data. Missing values in the CART models are handled using surrogate splitting [1]. We then use a logistic regression to fine-tune the CART based probabilities. Unlike CART, the logistic regression is not as adept at dropping insignificant factors, so an AIC-variable selection procedure is used to select the important variables. The stocks with missing information are assigned with CART predicted probabilities.

As an example of the trees in the backtest, Figure 6.5 illustrates the last tree grown at the end of June, 2010. As it is shown, the chance of a stock generating a return above the benchmark, *i.e.*, an equally weighted defensive stock returns, is primarily determined by the joint effects of Profitability, Value and Stability although other factors are relevant further down the decision hierarchy. Indeed, the first observation to note is that the primary split is on Profitability and, more specifically, the distinction between those stocks that are profitable (the left hand branch) and the not so profitable stocks. Overall, the tree model prefers profitable and cheap stocks (Node 1) and attaches them with a 56.3% probability

---

[1]Surrogate splitters are back-up rules that closely mimic the action of primary splitting rules in the cases of missing primary splitters. The use of surrogate splitting effectively minimizes the ad-hoc handling of missing values.
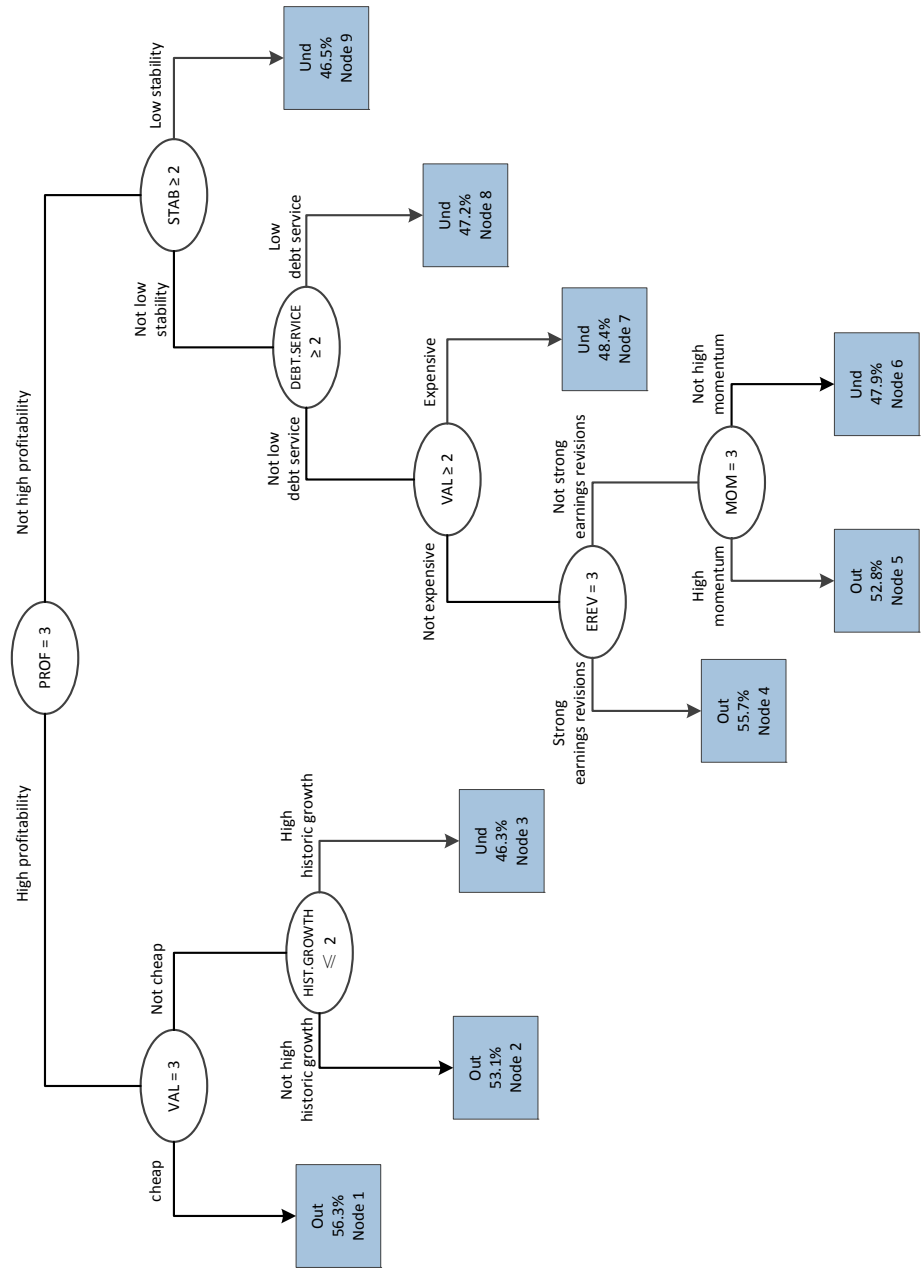
Figure 6.2: The decision tree at June 2010. The outperforming probabilities are reported at the terminal nodes. Out stands for outperforming and Und underperforming. Factors are split into three equally sized categories each period with bucket 3 representing a high value of the underlying factor and bucket 1 representing a low value (*e.g.*, VAL = 3 represents the cheapest category of stocks).

of outperforming the overall universe. On top of exhibiting low profitability, the underperforming stocks also have characteristics such as low stability (Node 9) and poor debt service (Node 8).

Table 6.1: This table reports the estimated coefficients from the logistic regression at June 2010. A stepwise AIC-variable selection procedure is used to select the important variables and the final model has three composite factors. P-values are in parentheses.

Variable	Coefficient	P-value
EREV	0.086	0.004***
VAL	0.057	0.014**
PROF	-0.091	0.001***

$*$ indicates significance at the 10% level
$**$ indicates significance at the 5% level
$***$ indicates significance at the 1% level

There are also exceptions to the general rule which is one of the benefits of tree-based methods. For example, attractively priced stocks with high stability and adequate debt service also outperform, particularly if they have strong earnings revisions (Node 4) or high momentum (Node 5) even if their profitability is weak. Furthermore, the composite factor Value has occurred at two different splitting nodes with different splitting values. The interactions between Value and the other composites are neatly captured by the hierarchical structure of the tree.

The second step of the hybrid approach is to use a logistic regression to adjust the probabilities produced by the CART model. This step has the ability to incorporate linear factor effects as well as any global influences that the tree model may have missed, as well as produce somewhat smoother response surface.

Following the AIC-variable selection procedure, three composite factors, Earnings Revisions, Value and Profitability, are identified in this step as being signif-

icant in adjusting the tree-predicted probabilities. Table 6.1 lists the estimated coefficients from the logistic regression at June 2010. On top of the tree-predicted probabilities, the logistic regression adjusts downwards the weights on Profitability (*i.e.*, it has a negative coefficient) while it increases the emphasis upon Earnings Revisions and Value. This suggests that the CART model is overly influenced by Profitability which is the primary splitting rule but does not adequately capture the linear effects arising from analyst sentiment in particular.

Table 6.2: This table reports risk analysis results using the FF three factors and Carhart four factors for the portfolios P1 and P3. The risk-adjusted return α and the coefficients for the corresponding risk factors are listed with p-values in parentheses.

	Risk analysis using the FF three factors		Risk analysis using the Carhart four factors	
	P1	P3	P1	P3
α	-0.011 (0.009)***	0.009 (0.003)***	-0.008 (0.042)**	0.007 (0.016)**
MKT	0.133 (0.004)***	-0.072 (0.034)**	0.082 (0.064)*	-0.042 (0.229)
SMB	0.087 (0.267)	-0.060 (0.312)	0.013 (0.857)	-0.017 (0.775)
HML	-0.227 (0.000)***	0.103 (0.012)**	-0.318 (0.000)***	0.156 (0.001)***
MOM	—	—	-0.137 (0.006)***	0.080 (0.036)**

$*$ indicates significance at the 10% level
$**$ indicates significance at the 5% level
$***$ indicates significance at the 1% level

The risk analysis results for the portfolios P1 and P3 are summarized in Table 6.2. At the 5% significance level, the risk-adjusted returns (α) of the long and short portfolios are statistically significant for both sets of the risk factors. Figure
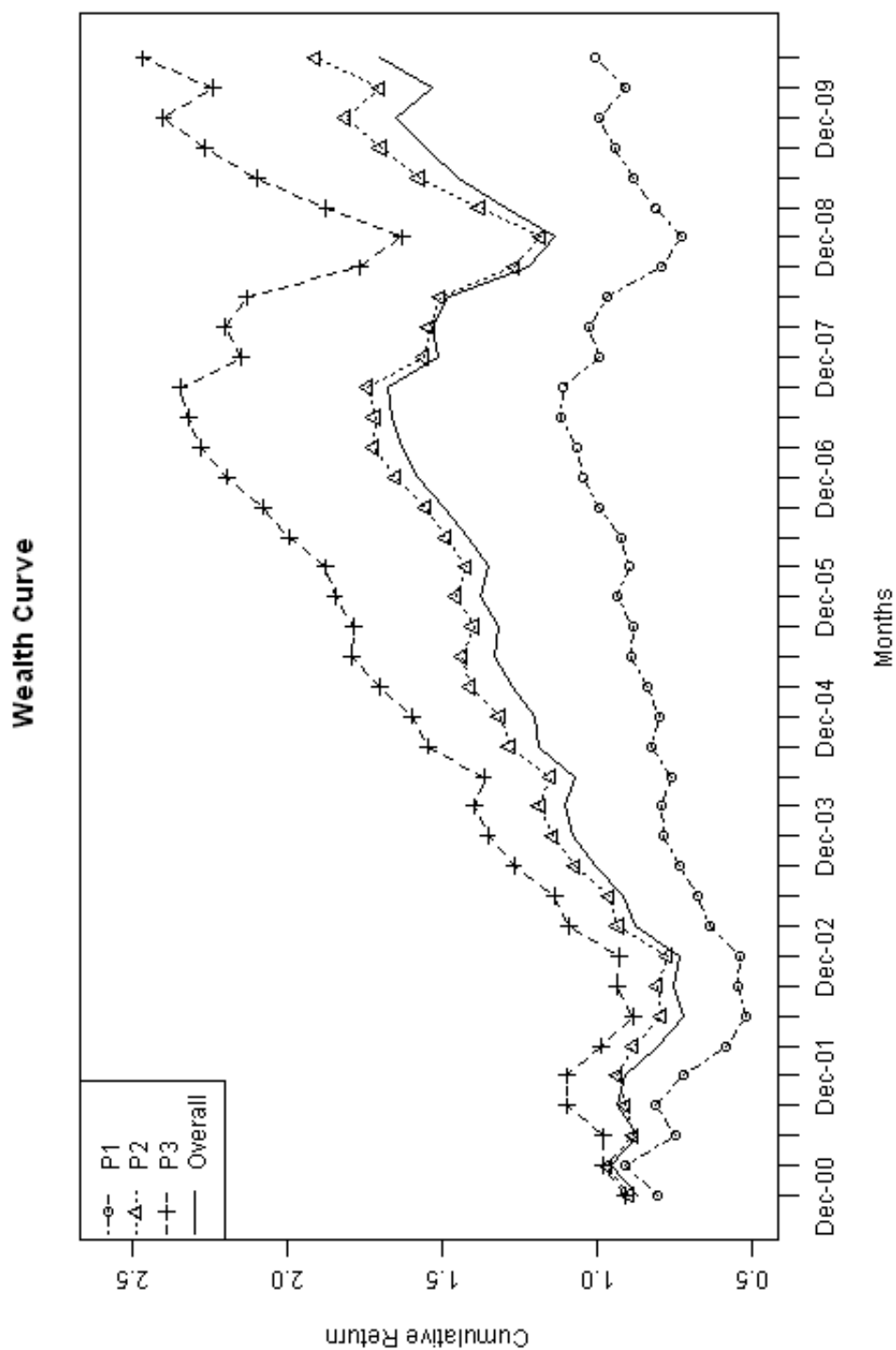
Figure 6.3: Wealth curves by the three portfolios formed by the hybrid method, from September 2000 to June 2010. The benchmark (solid line) is chosen to be the overall equally weighted returns of the companies in the sample.

133

6.3 diagrammatically depicts the wealth curves associated with the portfolios formed by the hybrid approach compared to the benchmark during the out-of-sample period. It is encouraging that the forecasts derived from the hybrid model generate relatively consistent performance during a period that is characterised by significant swings in market sentiment, including the Global Financial Crisis and the subsequent recovery.

6.6   Results: Model Comparison

Whilst the application of the hybrid approach on defensive stocks is encouraging, we now compare its performance with a number of alternatives, including separately estimated CART and logistic models as well as the random forest approach. As logistic regression and the random forest approach are potentially troublesome in the presence of missing values, such observations are replaced with the corresponding quarterly medians for these two models. An AIC-variable selection procedure is once again used to select the important variables in the logistic regression analysis.

Table 6.3 summaries the out-of-sample panel hit rates, time series hit rates, excess returns, information ratios and holding period information for the portfolios formed using the four different approaches. We can see that whilst the standalone CART model outperforms both a simple logistic and the random forest approach, the performance of the long and short portfolios formed by the hybrid method is superior in this particular sample period. This is true regardless of the performance diagnostic used but it is particularly encouraging that the improved returns of the hybrid model are also associated with a high degree of

Table 6.3: This table reports the backtest results for the four different models. RF stands for random forest. The performance metrics reported here are panel hit rate, time-series hit rates (TS Hit Rate), annualized excess returns, information ratio (IR) and mean and median of stock holding periods. The details of these portfolio performance measures are given in Section 5.6 and Section 6.4.

Model	Port-folio	Panel Hit Rate	TS Hit Rate	Excess Return (%)	IR	Holding Period	
						Mean	Median
RF	P1	0.495	0.513	-0.64	-0.14	1.64	1.50
	P2	0.495	0.436	-0.73	-0.29	1.46	1.38
	P3	0.506	0.577	1.26	0.38	1.73	1.50
Logistic	P1	0.477	0.384	-3.46	-0.48	3.67	2.92
	P2	0.513	0.614	1.46	0.44	2.72	2.18
	P3	0.508	0.538	1.98	0.37	3.61	3.01
CART	P1	0.464	0.385	-4.05	-0.52	2.10	1.70
	P2	0.516	0.564	1.37	0.46	2.04	1.80
	P3	0.510	0.667	2.15	0.50	4.13	2.75
Hybrid	P1	0.414	0.317	-4.86	-0.87	3.26	2.78
	P2	0.524	0.549	1.13	0.27	2.17	2.05
	P3	0.565	0.692	3.71	0.93	3.96	2.95

consistency as measured by the time series hit rate.

In terms of stock holding period, stocks in P1 and P3 tend to be traded less often than stocks in P2. Overall, the portfolios based on the random forest method display the highest turnover (*i.e.*, they have the shortest holding period). As measured by the median stock holding period, the logistic regression has the lowest turnover out of the four approaches, which is attributed to the smooth and slower changing probability surface it produces. The hybrid approach has a similar length of holding period as the logistic regression. Therefore, the enhanced performance of the portfolios formed by the hybrid approach is not lost in transaction costs relative to the alternative approaches.

6.7 Summary

This chapter proposes a hybrid approach for stock ranking that combines the benefits of the CART and logistic regression. We apply the hybrid model to the task of building a stock selection model for the North American defensive companies over the past decade with some success. Moreover, it offers enhanced performance compared to either a standalone CART or a logistic regression model and also compares favourably with the random forest method.

The primary advantages of the proposed approach are listed below.

Firstly, the hybrid approach overcomes the less sensitive response of CART to continuous variables whilst it can easily explore both linear and non-linear patterns in stock data. The CART model is used to capture nonlinearities and high-order interactions among stock characteristics whilst the logistic regression procedure can be regarded as a further refinement to capture and approximate the remaining linear effects.

Secondly, the hybrid model provides an interpretable explanation. The clear visualization of various components of the model is the key to a good understanding of the model.

Thirdly, it minimizes the ad-hoc nature of handing missing data values which is relatively common in financial data. Traditional parametric models usually provide challenges in dealing with missing information. The hybrid approach overcomes it by using surrogate splitting in the CART algorithm.

Whilst further empirical testing is warranted in a wider array of data sets, we believe that the hybrid approach has the potential to offer the best of both worlds and therefore potentially a useful addition to the range of techniques available.

Chapter 7

Conclusion and Discussion

7.1 Summary

Focusing on return predictability and its implications, the research in this thesis inquires into a range of issues: (i) it investigates the estimation bias issue in predictive regressions and also proposes new methodology for bias removal; (ii) it examines predictability of return distributions; (iii) it explores the implications of return distribution predictability for portfolio selection and (vi) it evaluates non-linear dynamics of cross-sectional stock returns using CART. More specifically, the key contributions and findings of the thesis are listed below.

- The thesis provides insightful explanations to the estimation bias issues in predictive regressions and also develops a jackknife-based approach for bias reduction. Extensive simulations show that, compared with existing bias-reduction methods in the literature, the proposed approach is more stable, robust and flexible. More importantly, the proposed approach can successfully reduce the estimation bias in long-horizon regressions, whereas

traditional bias-reduction methods do not work efficiently.

- The research stresses the importance of bias reduction in predictive regressions in practical applications through both simulations and empirical applications. Along with other studies in the literature (see, for example, Stambaugh, 1999; Goyal and Welch, 2003; Amihud and Hurvich, 2004), the research identifies the difficulties in establishing the predictability of the mean of the stock market returns after accounting for finite-sample biases.

- The thesis concludes that it is insufficient to use predictive regressions to investigate return predictability. Instead, it promotes the use of quantile regression to incorporate "tail" information in return prediction. The empirical studies in the thesis report strong evidence of distribution predictability for both stock and bond returns. It is also demonstrated that a wide range of economic state factors have significant and heterogenous effects on different return quantiles.

- The thesis establishes a quantile-copula framework for modelling conditional joint return distributions. This framework hinges on quantile regression for marginal return distributions and a copula for the return dependence structure. The framework is shown to be flexible and general enough to model a joint distribution while, at the same time, capturing any non-Gaussian characteristics in both marginal and joint returns.

- To exploit predicted return distributions, the thesis develops a distribution-based portfolio selection framework which uses the generalised Omega as the objective function and the threshold accepting optimisation for obtain-

ing solutions. An empirical application to asset allocation between stocks and bonds demonstrates the efficiency of the proposed portfolio selection approach.

- The thesis evaluates the use of CART for cross-sectional return prediction. It finds that CART offers an alternative means of analysing complex stock data as it sidesteps many of the known issues associated with traditional linear regression based approaches. CART is more flexible, robust and capable of capturing nonlinear return dynamics. In practice, it also provides a high degree of diversification to the risk of using similar models in the investment community.

- To overcome weaknesses of CART, such as local optimum, solution instability and insensitivity to continuous variables, and in the meanwhile maintaining its flexibility, the thesis proposes a novel hybrid approach combining CART and logistic regression. An empirical application to cross-sectional stock return prediction shows that the hybrid approach offers enhanced performance compared to either a standalone CART or a logistic model. It also compares favourably with the random forest method.

7.2   Limitations and Future Research

While this thesis has investigated a number of fundamental issues of return predictability and has provided some new frameworks that are shown to improve prediction, there are a number of issues that need further research.

For predictive regressions, the thesis focuses on the finite-sample bias in the

class of predictive regressions as analysed by Stambaugh (1999). It would be interesting to investigate other frameworks. In particular, Ferson *et al.* (2003) and Ferson *et al.* (2008) study another class of predictive regressions which involves a latent variable. They assume that the returns are driven by a latent variable. However, an instrumental variable which may or may not correlate with the latent variable is used to predict the returns. In their setup, Ferson *et al.* (2003) find substantial spurious regression bias under certain conditions. The literature on bias issues under the setup of Ferson *et al.* (2003) and Ferson *et al.* (2008) is sparse. Given their setup is more general and closely mimics what an econometrician faces in forecasting returns, it would be of great interest to evaluate the effectiveness of the proposed jackknife procedure or develop new appropriate methodology for bias reduction within their setup.

Furthermore, predictive regressions can be generalised to allow for time-varying coefficients. Time-variation in coefficients (*i.e.*, parameter instability) is well studied in the literature (see for example, Lettau and Ludvigson, 2001; Goyal and Welch, 2003; Paye and Timmermann, 2006; Ang and Bekaert, 2007; Dangl and Halling, 2008). However, estimation bias, coupled with parameter instability, imposes substantial challenges in finance modelling and predictability testing. The generalised framework established in this thesis offers great opportunities for future research in this direction.

The study of return distribution predictability reports strong empirical evidence of predictability in-sample. It is not clear whether, and to what extent, quantile regression estimation is affected by finite-sample biases when using highly persistent predictor variables. Given the fact that model assumptions on quantile regression are less stringent than those of predictive regressions, the problem of

140

the finite-sample bias may not be as severe as it is for the OLS estimator. This is also an interesting question worth future research.

The examination of the out-of-sample predicability of the full return distribution in this thesis is based on short time series. When longer time series are available, more accurate evaluation of the out-of-sample predicability should be carried out.

Regarding implications of return distribution predictability for portfolio selection, this thesis has developed a framework for distribution-based portfolio selection which uses a generalised version of the Omega measure as the objective function. It also generates several interesting directions for future research: (i) an investigation of other utility functions which can also capture the higher-order moments for portfolio selection; (ii) an assessment of ex-post portfolio performance using the proposed framework and a comparison with traditional moment-based methods (*i.e.*, the mean-variance analysis) and (iii) as only two assets are considered in the study, a natural extension, therefore, is to allocate resources among three or more assets using the proposed framework. Although modelling joint return distribution of multiple assets and carrying out the corresponding asset allocation are theoretically feasible within the framework proposed, the computational issues that arise from high dimensionality would be much more challenging.

Lastly, the research in the thesis represents only a preliminary study of capturing the nonlinear dynamics in the cross-sectional stock returns. Nonlinear dynamics of returns are not well understood in the literature and are not the focus of mainstream research in finance. The implications of the probabilistic return prediction by CART and other similar approaches would be worth further exploration.

Appendix A: Proof of Theorem

This appendix provides the proof of Theorem 2.4.1 in Chapter 2.

Theorem 2.4.1. *Suppose $\hat{\theta}_{JK}$ is the estimator of β obtained from jackknifing the predictive regression specified by (2.1) to (2.3), we have $E(\hat{\beta}_{JK} - \beta) = O(T^{-1})$.*

Proof. By denoting $\bar{\beta}_{-1} = \sum_{i=1}^{T} \hat{\beta}_{(-i)}/T$, the ordinary delete-one jackknife estimator is $\hat{\beta}_{JK} = T\hat{\beta} - (T-1)\bar{\beta}_{-1}$. Its expectation can be written as

$$
\begin{aligned}
E(\hat{\beta}_{JK} - \beta) &= E\{T\hat{\beta} - (T-1)\bar{\beta}_{-1}\} - \hat{\beta}) + E(\hat{\beta} - \beta) \\
&= (T-1)\{E(\hat{\beta} - \bar{\beta}_{-1})\} + E(\hat{\beta} - \beta). \quad\quad (A.1)
\end{aligned}
$$

From (2.6), the expectation $E(\hat{\beta} - \bar{\beta}_{-1})$ can be written as

$$
E(\hat{\beta} - \bar{\beta}_{-1}) = \xi E\left\{ \frac{\sum_{t=1}^{T}(x_{t-1} - \bar{x})x_t}{\sum_{t=1}^{T}(x_{t-1} - \bar{x})^2} - \sum_{i=1}^{T} \frac{\sum_{t=1,t\neq i}^{T}(x_{t-1} - \bar{x}_{-i})x_t}{T\sum_{t=1,t\neq i}^{T}(x_{t-1} - \bar{x}_{-i})^2} \right\}.
$$

The right-hand side of the above expression can be decomposed as the sum of the following four differences:

$$
h_1 = \frac{\sum_{t=1}^{T}(x_{t-1} - \bar{x})x_t}{\sum_{t=1}^{T}(x_{t-1} - \bar{x})^2} - \frac{1}{T-1}\sum_{i=1}^{T}\left\{ \frac{\sum_{t=1,t\neq i}^{T}(x_{t-1} - \bar{x})x_t}{\sum_{t=1}^{T}(x_{t-1} - \bar{x})^2} \right\},
$$

$$
h_2 = \frac{1}{T-1} \sum_{i=1}^{T} \left\{ \frac{\sum_{t=1,t\neq i}^{T}(x_{t-1}-\bar{x})x_t}{\sum_{t=1}^{T}(x_{t-1}-\bar{x})^2} \right\} - \frac{1}{T} \sum_{i=1}^{T} \left\{ \frac{\sum_{t=1,t\neq i}^{T}(x_{t-1}-\bar{x})x_t}{\sum_{t=1,t\neq i}^{T}(x_{t-1}-\bar{x})^2} \right\},
$$

$$
h_3 = \frac{1}{T} \sum_{i=1}^{T} \left\{ \frac{\sum_{t=1,t\neq i}^{T}(x_{t-1}-\bar{x})x_t}{\sum_{t=1,t\neq i}^{T}(x_{t-1}-\bar{x})^2} - \frac{\sum_{t=1,t\neq i}^{T}(x_{t-1}-\bar{x}_{-i})x_t}{\sum_{t=1,t\neq i}^{T}(x_{t-1}-\bar{x})^2} \right\},
$$

$$
h_4 = \frac{1}{T} \sum_{i=1}^{T} \left\{ \frac{\sum_{t=1,t\neq i}^{T}(x_{t-1}-\bar{x}_{-i})x_t}{\sum_{t=1,t\neq i}^{T}(x_{t-1}-\bar{x})^2} - \frac{\sum_{t=1,t\neq i}^{T}(x_{t-1}-\bar{x}_{-i})x_t}{\sum_{t=1,t\neq i}^{T}(x_{t-1}-\bar{x}_{-i})^2} \right\}.
$$

The theorem can be proved by showing $\sum_{k=1}^{4} \mathrm{E}(h_k) = O(T^{-3})$. For this purpose, the geometric series formula is used in the proof,

$$
\frac{1}{1-a} = \sum_{n=0}^{\infty} a^n, \text{ for } |a| < 1.
$$

Define the following notations: $d_i = (x_{i-1} - \bar{x})^2$, $D = \sum_{i=1}^{T} d_i$, $\delta_i = D - (x_{i-1} - \bar{x})^2$, $S_0 = \sum_{t=1}^{T} x_{t-1}$, $S_1 = \sum_{t=1}^{T} x_t$ and $\omega_i = \sum_{t=1,t\neq i}^{T}(\bar{x}_{-i} - \bar{x})x_t$. Without loss of generality, let $\phi = 0$, i.e., x series is centralised. Note that $v_i = x_i - \rho x_{i-1}$ and for any $j \geq 0$, $k \geq 0$ and $l \geq 0$,

$$
\mathrm{E}(x_i x_{i+j} x_{i+j+k} x_{i+j+k+l}) = \frac{\rho^{j+l}(1 + 2\rho^{2k})}{(1 - \rho^2)^2}\sigma_v^4 = \rho^{j+l}(1 + 2\rho^{2k})\sigma_x^4.
$$

According to Marriott and Pope (1956), we have

$$
\mathrm{var}(\bar{x}) = \frac{\sigma_v^2}{T^2}\{T + (T-1)\rho + (T-2)\rho^2 + ... + \rho^{T-1}\}
$$

$$
\mathrm{E}(D) = T\left\{\frac{1}{1-\rho^2} - \frac{1}{T(1-\rho)^2} + O\left(\frac{1}{T^2}\right)\right\}\sigma_v^2
$$

$$
= T\left\{1 - \frac{1-\rho^2}{T(1-\rho)^2} + O\left(\frac{1}{T^2}\right)\right\}\sigma_x^2
$$

$$
\mathrm{E}(D^2) = T^2\left\{\frac{1}{(1-\rho^2)^2} + \frac{2(1+\rho)^2}{T(1-\rho^2)^3} + O\left(\frac{1}{T^2}\right)\right\}\sigma_v^4
$$

$$= T^2 \left\{ 1 + \frac{2(1+\rho)^2}{T(1-\rho^2)} + O\left(\frac{1}{T^2}\right) \right\} \sigma_x^4$$

$$\mathrm{E}(\delta_i) = (T-1) \left\{ \frac{1}{1-\rho^2} - \frac{1}{T(1-\rho)^2} + O\left(\frac{1}{T^2}\right) \right\} \sigma_v^2.$$

Furthermore, the following equations hold,

$$\mathrm{E}\frac{S_0^2}{T} = O(1), \quad \mathrm{E}\frac{S_0 S_1}{T} = O(1), \quad \mathrm{E}\frac{S_0^2 \sum_{i=1}^{T} x_i x_{i-1}}{T^2} = O(1),$$

$$\mathrm{E}\frac{S_1 S_0^3}{T^2} = O(1), \quad \mathrm{E}\frac{S_0^2 \sum_{i=1}^{T} x_i x_{i-1}}{T^2} = O(1), \quad \mathrm{E}\frac{\sum_{i=1}^{T} x_i x_{i-1}^3}{T} = O(1)$$

$$\mathrm{E}\frac{1}{D^2} = O\left(\frac{1}{T^2}\right), \quad \mathrm{E}\frac{1}{D} = O\left(\frac{1}{T}\right), \quad \mathrm{E}\frac{S_0 S_1}{D} = O(1), \quad \mathrm{E}\frac{S_0^2}{D} = O(1).$$

Now we move on to compute the expectations of h_1 to h_4.

For h_1, we have

$$
\begin{aligned}
h_1 &= \frac{\sum_{t=1}^{T}(x_{t-1}-\bar{x})x_t}{\sum_{t=1}^{T}(x_{t-1}-\bar{x})^2} - \frac{\sum_{i=1}^{T}\sum_{t=1,t\neq i}^{T}(x_{t-1}-\bar{x})x_t}{(T-1)\sum_{t=1}^{T}(x_{t-1}-\bar{x})^2} \\
&= \frac{\sum_{t=1}^{T}(x_{t-1}-\bar{x})x_t}{\sum_{t=1}^{T}(x_{t-1}-\bar{x})^2} - \left\{ \frac{T\sum_{t=1}^{T}(x_{t-1}-\bar{x})x_t - \sum_{i=1}^{T}(x_{i-1}-\bar{x})x_i}{(T-1)\sum_{t=1}^{T}(x_{t-1}-\bar{x})^2} \right\} \\
&= \frac{\sum_{t=1}^{T}(x_{t-1}-\bar{x})x_t}{\sum_{t=1}^{T}(x_{t-1}-\bar{x})^2} - \frac{\sum_{t=1}^{T}(x_{t-1}-\bar{x})x_t}{\sum_{t=1}^{T}(x_{t-1}-\bar{x})^2} \\
&= 0.
\end{aligned}
$$

Hence the expectation of h_1 is

$$\mathrm{E}(h_1) = 0. \tag{A.2}$$

For h_2, we have

$$h_2 = \frac{1}{T-1}\sum_{i=1}^{T}\left\{ \frac{\sum_{t=1,t\neq i}^{T}(x_{t-1}-\bar{x})x_t}{D} \right\} - \frac{1}{T}\sum_{i=1}^{T}\left\{ \frac{\sum_{t=1,t\neq i}^{T}(x_{t-1}-\bar{x})x_t}{\delta_i} \right\}$$

144

$$
= \frac{1}{T}\sum_{i=1}^{T}\left\{\sum_{t=1,t\neq i}^{T}(x_{t-1}-\bar{x})x_t\right\}\left\{\frac{T}{(T-1)D}-\frac{1}{D-(x_{i-1}-\bar{x})^2}\right\}
$$

$$
= \frac{1}{T}\sum_{i=1}^{T}\left\{\sum_{t=1,t\neq i}^{T}(x_{t-1}-\bar{x})x_t\right\}\left\{\frac{T}{(T-1)D}-\frac{1}{D}-\frac{(x_{i-1}-\bar{x})^2}{D^2}-\frac{(x_{i-1}-\bar{x})^4}{D^3}-\cdots\right\}
$$

$$
= \frac{1}{T}\sum_{i=1}^{T}\left\{\sum_{t=1}^{T}(x_{t-1}-\bar{x})x_t\right\}\left\{\frac{1}{(T-1)D}-\frac{(x_{i-1}-\bar{x})^2}{D^2}-\frac{(x_{i-1}-\bar{x})^4}{D^3}-\cdots\right\}
$$

$$
-\frac{1}{T}\sum_{i=1}^{T}(x_{i-1}-\bar{x})x_i\left\{\frac{1}{(T-1)D}-\frac{(x_{i-1}-\bar{x})^2}{D^2}-\cdots\right\}\bar{x}^4\sum_{t=1}^{T}(x_{t-1}-\bar{x})x_t.
$$

Taking expectation of h_2 gives,

$$
\begin{aligned}
\mathrm{E}(h_2) &= \frac{1}{T}\sum_{i=1}^{T}\mathrm{E}\left[\left\{\sum_{t=1}^{T}(x_{t-1}-\bar{x})x_t\right\}\left\{\frac{1}{(T-1)D}-\frac{(x_{i-1}-\bar{x})^2}{D^2}-\frac{(x_{i-1}-\bar{x})^4}{D^3}\right\}\right]\\
&\quad -\frac{1}{T}\sum_{i=1}^{T}\mathrm{E}\left[(x_{i-1}-\bar{x})x_i\left\{\frac{1}{(T-1)D}-\frac{(x_{i-1}-\bar{x})^2}{D^2}\right\}\right]+O\left(\frac{1}{T^3}\right)\\
&= \frac{1}{T}\mathrm{E}\left\{\frac{\sum_{i=1}^{T}x_i(x_{i-1}-\bar{x})^3}{D^2}\right\}\\
&\quad -\frac{1}{T}\mathrm{E}\left\{\frac{\sum_{i=1}^{T}(x_{i-1}-\bar{x})^4\sum_{t=1}^{T}(x_{t-1}-\bar{x})x_t}{D^3}\right\}+O\left(\frac{1}{T^3}\right)\\
&= \frac{1}{T}\mathrm{E}\left\{\frac{\sum_{i=1}^{T}x_i x_{i-1}^3}{D^2}\right\}-\frac{3}{T^2}\mathrm{E}\left\{\frac{3S_0\sum_{i=1}^{T}x_i x_{i-1}^2}{D^2}\right\}+\frac{3}{T^3}\mathrm{E}\left\{\frac{S_0^2\sum_{i=1}^{T}x_i x_{i-1}}{D^2}\right\}\\
&\quad -\frac{1}{T}\mathrm{E}\left\{\frac{\sum_{i=1}^{T}(x_{i-1}-\bar{x})^4\sum_{t=1}^{T}(x_{t-1}-\bar{x})x_t}{D^2}\frac{}{D}\right\}+O\left(\frac{1}{T^3}\right)\\
&= \frac{1}{T}\mathrm{E}\left\{\frac{\sum_{i=1}^{T}x_i x_{i-1}^3}{D^2}\right\}-\frac{3\rho}{T^2}+O\left(\frac{1}{T^3}\right)\\
&= \frac{1}{T}\frac{\sum_{i=1}^{T}3\rho}{T^2}-\frac{3\rho}{T^2}+O\left(\frac{1}{T^3}\right)\\
&= O\left(\frac{1}{T^3}\right). \tag{A.3}
\end{aligned}
$$

For h_3, note that

$$
\begin{aligned}
\omega_i &= \sum_{t=1,t\neq i}^{T} (\bar{x}_{-i} - \bar{x}) x_t \\
&= \sum_{t=1,t\neq i}^{T} \left(\frac{T\bar{x} - x_{i-1}}{T-1} - \bar{x} \right) x_t \\
&= \frac{\bar{x} - x_{i-1}}{T-1} \sum_{t=1,t\neq i}^{T} x_t \\
&= \frac{1}{T(T-1)} (S_0 - Tx_{i-1})(S_1 - x_i),
\end{aligned}
$$

and

$$
\begin{aligned}
\sum_{i=1}^{T} \omega_i &= \frac{1}{T(T-1)} \sum_{i=1}^{T} \{ (S_0 - Tx_{i-1})(S_1 - x_i) \} \\
&= \frac{1}{(T-1)} \left\{ \sum_{i=1}^{T} x_i x_{i-1} - \frac{S_0 S_1}{T} \right\}.
\end{aligned}
$$

Recall that $\delta_i = D - (x_{i-1} - \bar{x})^2$ and $d_i = (x_{i-1} - \bar{x})^2$. Then

$$
\begin{aligned}
h_3 &= \frac{1}{T} \sum_{i=1}^{T} \frac{\omega_i}{\delta_i} \\
&= \frac{1}{T} \sum_{i=1}^{T} \frac{\omega_i}{D - (x_{i-1} - \bar{x})^2} \\
&= \frac{1}{T} \sum_{i=1}^{T} \frac{\omega_i}{D} \left(1 + \frac{(x_{i-1} - \bar{x})^2}{D} + \frac{(x_{i-1} - \bar{x})^4}{D^2} + \dots \right).
\end{aligned}
$$

Upon taking expectation, we have

$$
\mathrm{E}(h_3) = \mathrm{E} \left\{ \frac{1}{T} \sum_{i=1}^{T} \frac{\omega_i}{D} \left(1 + \frac{(x_{i-1} - \bar{x})^2}{D} + \frac{(x_{i-1} - \bar{x})^4}{D^2} + \dots \right) \right\}
$$

$$
= \mathrm{E}\left\{\frac{1}{T}\sum_{i=1}^{T}\frac{\omega_i}{D}\left(1+\frac{(x_{i-1}-\bar{x})^2}{D}\right)\right\}+O\left(\frac{1}{T^3}\right)
$$

$$
= \mathrm{E}\left\{\frac{1}{TD}\sum_{i=1}^{T}\omega_i+\frac{1}{TD^2}\sum_{i=1}^{T}\omega_i(x_{i-1}-\bar{x})^2\right\}+O\left(\frac{1}{T^3}\right)
$$

$$
= \mathrm{E}\left\{\frac{1}{T(T-1)D}\left(\sum_{i=1}^{T}x_i x_{i-1}-\frac{S_0 S_1}{T}\right)\right\}
$$

$$
+\mathrm{E}\left\{\frac{1}{T^2(T-1)D^2}\sum_{i=1}^{T}(S_0-Tx_{i-1})(S_1-x_i)d_i\right\}+O\left(\frac{1}{T^3}\right)
$$

$$
= \frac{1}{T(T-1)}\left\{\mathrm{E}\left(\frac{\sum_{i=1}^{T}x_i x_{i-1}}{D}\right)\right\}+\frac{1}{T^2(T-1)}\left\{3\mathrm{E}\left(\frac{S_0 S_1\sum_{i=1}^{T}x_{i-1}^2}{D^2}\right)\right.
$$

$$
-\frac{2T+1}{T^2}\mathrm{E}\left(\frac{S_0^3 S_1}{D^2}\right)-3\mathrm{E}\left(\frac{S_0\sum_{i=1}^{T}x_i x_{i-1}^2}{D^2}\right)+3\mathrm{E}\left(\frac{S_0^2\sum_{i=1}^{T}x_i x_{i-1}}{TD^2}\right)
$$

$$
\left.-T\mathrm{E}\left(\frac{S_1\sum_{i=1}^{T}x_{i-1}^3}{D^2}\right)+T\mathrm{E}\left(\frac{\sum_{i=1}^{T}x_i x_{i-1}^3}{D^2}\right)\right\}+O\left(\frac{1}{T^3}\right)
$$

$$
= \frac{1}{T(T-1)}\left\{\mathrm{E}\left(\frac{\sum_{i=1}^{T}x_i x_{i-1}}{D}\right)\right\}+\frac{1}{T^2(T-1)}\left\{O(1)-O\left(\frac{1}{T}\right)-O(1)\right.
$$

$$
\left.+O\left(\frac{1}{T}\right)-O(1)+O(1)\right\}+O\left(\frac{1}{T^3}\right)
$$

$$
= \frac{1}{T(T-1)}\mathrm{E}\left(\frac{\sum_{i=1}^{T}x_i x_{i-1}}{D}\right)+O\left(\frac{1}{T^3}\right). \tag{A.4}
$$

Before moving to h_4, it is useful to write

$$
\sum_{t=1,t\neq i}^{T}(x_{t-1}-\bar{x}_{-i})^2 = \sum_{t=1,t\neq i}^{T}(x_{t-1}-\bar{x})^2-(T-1)(\bar{x}-\bar{x}_{-i})^2
$$

$$
\bar{x}_{-i} = (S_0-x_{i-1})/(T-1)
$$

and

$$
(\bar{x}-\bar{x}_{-i})=x_{i-1}/(T-1)-S_0/\{T(T-1)\}.
$$

For h_4, we have

$$
\begin{aligned}
h_4 &= \frac{1}{T}\sum_{i=1}^{T}\left\{\frac{\sum_{t=1,t\neq i}^{T}(x_{t-1}-\bar{x}_{-i})x_t}{\sum_{t=1,t\neq i}^{T}(x_{t-1}-\bar{x})^2} - \frac{\sum_{t=1,t\neq i}^{T}(x_{t-1}-\bar{x}_{-i})x_t}{\sum_{t=1,t\neq i}^{T}(x_{t-1}-\bar{x}_{-i})^2}\right\} \\
&= \frac{1}{T}\sum_{i=1}^{T}\left\{\sum_{t=1,t\neq i}^{T}(x_{t-1}-\bar{x}_{-i})x_t\right\}\left\{\frac{1}{D-(x_{i-1}-\bar{x})^2}\right. \\
&\qquad \left. -\frac{1}{D-(x_{i-1}-\bar{x})^2-(T-1)(\bar{x}-\bar{x}_{-i})^2}\right\} \\
&= \frac{1}{T}\sum_{i=1}^{T}\left\{\sum_{t=1,t\neq i}^{T}(x_{t-1}-\bar{x}_{-i})x_t\right\}\left[\frac{1}{D}\left\{1+\frac{(x_{i-1}-\bar{x})^2}{D}+...\right\}\right. \\
&\qquad \left. -\left\{1+\frac{(x_{i-1}-\bar{x})^2+(T-1)(\bar{x}-\bar{x}_{-i})^2}{D}+...\right\}\right],
\end{aligned}
$$

Taking expectation of h_4 gives

$$
\begin{aligned}
\mathrm{E}(h_4) &= \frac{1}{T}\sum_{i=1}^{T}\mathrm{E}\left(\left\{\sum_{t=1,t\neq i}^{T}(x_{t-1}-\bar{x}_{-i})x_t\right\}\left[\frac{1}{D}\left\{1+\frac{(x_{i-1}-\bar{x})^2}{D}\right\}\right.\right. \\
&\qquad \left.\left. -\left\{1+\frac{(x_{i-1}-\bar{x})^2+(T-1)(\bar{x}-\bar{x}_{-i})^2}{D}\right\}\right]\right) + O\left(\frac{1}{T^3}\right) \\
&= \mathrm{E}\left[-\frac{T-1}{TD^2}\sum_{i=1}^{T}(\bar{x}-\bar{x}_{-i})^2\sum_{t=1,t\neq i}^{T}(x_{t-1}-\bar{x}_{-i})x_t\right] + O\left(\frac{1}{T^3}\right) \\
&= \mathrm{E}\left[-\frac{1}{TD^2}\sum_{i=1}^{T}(\bar{x}-\bar{x}_{-i})^2\left\{(T-1)\sum_{t=1}^{T}x_tx_{t-1}-S_0S_1\right.\right. \\
&\qquad \left.\left. +S_1x_{i-1}-S_0x_i+Tx_ix_{i-1}\right\}\right] + O\left(\frac{1}{T^3}\right) \\
&= \mathrm{E}\left[-\frac{1}{T(T-1)^2D^2}\sum_{i=1}^{T}\left(x_{i-1}-\frac{S_0}{T}\right)^2\right. \\
&\qquad \left.\left\{(T-1)\sum_{t=1}^{T}x_tx_{t-1}-S_0S_1+S_1x_{i-1}-S_0x_i+Tx_ix_{i-1}\right\}\right] + O\left(\frac{1}{T^3}\right)
\end{aligned}
$$

$$
\begin{aligned}
&= \text{E}\left[-\frac{1}{T(T-1)^2 D^2}\left\{(T-1)D\sum_{t=1}^{T} x_t x_{t-1} - S_0 S_1 D \right.\right. \\
&\quad \left.\left. +S_1 \sum_{i=1}^{T} x_{i-1} d_i - S_0 \sum_{i=1}^{T} x_i d_i + T\sum_{i=1}^{T} x_i x_{i-1} d_i \right\}\right] + O\left(\frac{1}{T^3}\right) \\
&= -\frac{1}{T(T-1)}\text{E}\left(\frac{\sum_{i=1}^{T} x_{t-1} x_t}{D}\right) + O\left(\frac{1}{T^3}\right). \quad\quad\quad\text{(A.5)}
\end{aligned}
$$

Note that $\sum_{i=1}^{T}\left(x_{i-1} - \frac{S_0}{T}\right)^2 = D$. It follows from (A.2) to (A.5) that

$$
\begin{aligned}
\sum_{k=1}^{4}\text{E}(h_k) &= \frac{1}{T(T-1)}\text{E}\left(\frac{\sum_{i=1}^{T} x_{t-1} x_t}{D}\right) - \frac{1}{T(T-1)}\text{E}\left(\frac{\sum_{i=1}^{T} x_{t-1} x_t}{D}\right) + O\left(\frac{1}{T^3}\right) \\
&= O\left(\frac{1}{T^3}\right),
\end{aligned}
$$

and hence

$$
\text{E}(\hat{\beta} - \bar{\beta}_{-1}) = \xi\sum_{k=1}^{4}\text{E}(h_k) = O\left(\frac{1}{T^3}\right). \quad\quad\quad\text{(A.6)}
$$

From (A.1) and (A.6), the bias expression for the ordinary jackknife estimator
is

$$
\begin{aligned}
\text{E}(\hat{\beta}_{JK} - \beta) &= (T-1)\{\text{E}(\hat{\beta} - \bar{\beta}_{-1})\} + \text{E}(\hat{\beta} - \beta) \\
&= -\frac{(1+3\rho)}{T}\xi + O\left(\frac{1}{T^2}\right) \\
&= O\left(\frac{1}{T}\right).
\end{aligned}
$$

$\square$

Appendix B: Standard Errors of Quantile Estimator

This appendix provides details on the method introduced by Wang *et al.* (2009) for obtaining the standard errors of the quantile regression coefficients in Chapter 3.

To overcome the problems due to the unsmoothness in the estimating function (3.2), we first express $\hat{\boldsymbol{\beta}}_\tau$ as $\boldsymbol{\beta}_\tau^* + \Lambda^{1/2}Z$, where $\boldsymbol{\beta}_\tau^*$ is the true value of the parameter $\boldsymbol{\beta}_\tau$ and Z follows the multivariate standard normal distribution $N(0, \mathrm{I})$. The smoothed objective function can be naturally defined as $\tilde{L}(\boldsymbol{\beta}_\tau) = \mathrm{E}_Z\{L(\boldsymbol{\beta}_\tau + \Lambda^{1/2}Z)\}$, where expectation is over Z. However, Λ is unknown and hence this expectation cannot be evaluated. To this end, we nominate a known matrix Γ for Λ, and then update Γ as an estimator for Λ. The interpretation is also simple, $\Gamma^{1/2}Z$ can be regarded as a perturbation to $\boldsymbol{\beta}_\tau$.

Let $\sigma_t^2 = X_t^T\Gamma X_t$, $a_t = r_{t+1} - X_t^T\boldsymbol{\beta}_\tau$ and $b_t = a_t/\sigma_t$. We have

$$
\begin{aligned}
\tilde{L}(\boldsymbol{\beta}_\tau) &= \mathrm{E}_Z L(\boldsymbol{\beta}_\tau + \Gamma^{1/2}Z) = T^{-1}\sum_{t=1}^{T}\mathrm{E}_Z\rho\{r_{t+1} - X_t^T(\boldsymbol{\beta}_\tau + \Gamma^{1/2}Z)\} \\
&= T^{-1}\sum_{t=1}^{T}(1-\tau)\int_{b_t}^{\infty}(\sigma_t z - a_t)\phi(z)dz + \tau\int_{-\infty}^{b_t}(a_t - \sigma_t z)\phi(z)dz
\end{aligned}
$$

$$= T^{-1} \sum_{t=1}^{T} \left[a_t \{ \Phi(b_t) - 1 + \tau \} + \sigma_t \phi(b_t) \right], \tag{B.1}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the distribution and density functions of the standard normal variable, respectively. Note that the score function of $L(\boldsymbol{\beta}_\tau)$ is $U(\boldsymbol{\beta}_\tau) = \partial L(\boldsymbol{\beta}_\tau)/\partial \boldsymbol{\beta}_\tau$. We have the smoothed version of $U(\boldsymbol{\beta}_\tau)$ as

$$\begin{aligned}
\tilde{U}(\boldsymbol{\beta}_\tau) &= \partial \tilde{L}(\boldsymbol{\beta}_\tau)/\partial \boldsymbol{\beta}_\tau = \mathrm{E}_Z \{ U(\boldsymbol{\beta}_\tau + \Gamma^{1/2} Z) \} \\
&= T^{-1} \sum_{t=1}^{T} X_t \{ \Phi(b_t) + \tau - 1 \},
\end{aligned} \tag{B.2}$$

which is a smooth function of $\boldsymbol{\beta}_\tau$. As a result, we can easily calculate $\partial \tilde{U}(\boldsymbol{\beta}_\tau)/\partial \boldsymbol{\beta}_\tau$ and use it as a smoothing estimator of A, that is,

$$\tilde{A} = T^{-1} \sum_{t=1}^{T} \frac{\phi(b_t)}{\sigma_t} X_t X_t^T, \tag{B.3}$$

for given values of $\boldsymbol{\beta}_\tau$ and Γ.

The smoothed estimator $\tilde{\boldsymbol{\beta}}_\tau$ for $\boldsymbol{\beta}_\tau$ can be obtained from the smoothed score function $\mathrm{E}\{ U(\boldsymbol{\beta}_\tau + \Gamma^{1/2} Z) \} = 0$. It is easy to show that $\tilde{L}(\boldsymbol{\beta}_\tau)$ is a strictly convex function of $\boldsymbol{\beta}_\tau$, and hence there is a unique minimizer of $\boldsymbol{\beta}_\tau$ for each given positive definite matrix Γ. Furthermore, the characteristics of the resulting estimator $\tilde{\boldsymbol{\beta}}_\tau$ are not changed much by such a smoothing method but the function $\tilde{U}$ is smoother than U in the cases of interest to us.

The covariance matrix of $\tilde{\boldsymbol{\beta}}_\tau$ can be estimated by

$$\tilde{\Lambda} = \tilde{A}^{-1} \mathrm{cov}\{ \tilde{U}(\boldsymbol{\beta}_\tau) \} (\tilde{A}^{-1})^T. \tag{B.4}$$

In general, iteration is needed to find the final estimates $\tilde{\boldsymbol{\beta}}_\tau$ and the corresponding asymptotic covariance matrix $\tilde{\Lambda}$. The iteration is executed by alternately updating estimates of Γ and $\boldsymbol{\beta}_\tau$. Note that there is no need to update B because it is free from $\boldsymbol{\beta}_\tau$ and Λ values (see (3.5)). The iteration procedure to find the smoothed estimates $\tilde{\boldsymbol{\beta}}_\tau$ and $\tilde{\Lambda}$ can be summarized by the following stepwise procedure.

(i) The initial value for Γ is taken as $\Gamma^{(0)} = T^{-1}I_p$, where p is the number of the regression parameters.

(ii) In the jth iteration ($j = 1, 2, \cdots$), we update $\boldsymbol{\beta}_\tau^{(j)}$ by minimizing $\tilde{L}(\boldsymbol{\beta})_\tau$ or solving the smoothed estimating function, $\tilde{U}(\boldsymbol{\beta}_\tau) = 0$, as given by (B.2).

(iii) We use $\boldsymbol{\beta}_\tau^{(j)}$ and $\Gamma^{(j-1)}$ to update $\tilde{A}$ (see eqn (B.3)), and then obtain an updated Γ as

$$\Gamma^{(j)} = T\tilde{A}^{-1}B(\tilde{A}^{-1})^T.$$

(iv) Repeat the above iteration steps (ii) and (iii) until a selected stopping criterion is reached, e.g., $\max |\Gamma^{(j+1)} - \Gamma^{(j)}| < 10^{-4}$.

The final values of $\boldsymbol{\beta}_\tau^{(j)}$ and $\Gamma^{(j)}$ will be taken as the smoothed estimate of $\boldsymbol{\beta}_\tau$ and the covariance matrix Λ. As we have shown that the smoothed $\tilde{U}$ after the first iteration is already very close to the original U, and the smoothed version can easily provide updated A and Λ matrices. Any further iteration (although not necessary from asymptotic viewpoint) is therefore to fine-tuning these quantities.

Appendix C: Random Forests

This appendix provides details on the random forest approach proposed by Breiman (2001). This approach is employed for cross-sectional stock return prediction in Chapter 6.

Random forest is an ensemble classifier that consists of many decision trees and outputs the class that is the mode of the classes output by individual trees. It is proposed to overcome the local optimum and instability of a single decision tree solution. Suppose that there are N training samples and M potential predictor variables. Random forest is constructed using the following algorithm:

- Draw k bootstrap samples with replacement from the training data.

- For each of the bootstrap samples, grow an unpruned decision tree, with the following modification: at each node, rather than choosing the best split among all predictors, randomly sample m predictors (m should be much less than M) and choose the best split from among those variables.

- Predict new data by aggregating the predictions of the trees, *i.e.*, majority votes for classification, average for regression.

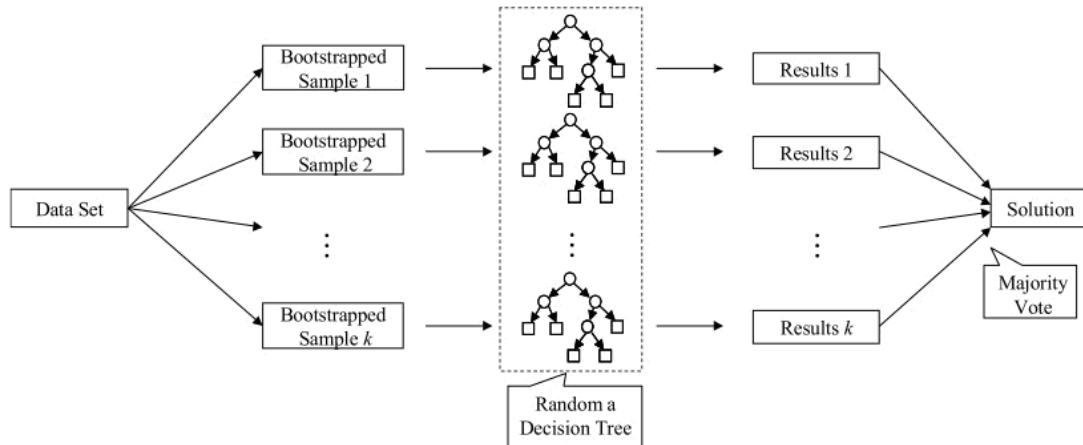Intuitively, the above process can be graphically illustrated as in Figure C.1.

Figure C.1: A graphic illustration of random forest.

An estimate of the error rate can be obtained, based on the training data, by the following:

- At each bootstrap iteration, predict the data using the sample outside the bootstrap sample (what Breiman calls out-of-bag data) using the tree grown with the bootstrap sample.

- Aggregate the OOB predictions and calculate the error rate, the so-called out-of-bag estimate of error rate.

Considered as one of the best off-the-shelf classifiers currently available (Satten *et al.*, 2004), random forest is fast becoming the starting point for modern tree-based analysis. However, the random forest approach also has a number of disadvantages. Firstly, the improved stability in random forests is at the cost

of interpretability of the model output. Secondly, random forests are prone to over fitting for some data sets. This is even more pronounced in noisy classification/regression tasks (Segal, 2003). Thirdly, the OOB estimate of error rate can bias upward Bylander (2002) which leads to problematic statistical inference. Finally, random forests cannot handle large numbers of irrelevant features as well as ensembles of entropy-reducing decision trees (Gashler *et al.*, 2008).

Bibliography

ADCOCK, C. (2010). Asset pricing and portfolio selection based on the multi-variate extended skew-student-t distribution. *Annals of Operations Research*, **176** (1), 221–234. 81

ADRIAN, T. and BRUNNERMEIER, M. (2011). Covar. *Available online at NBER http://www.nber.org/papers/w17454.ack.* 52

AMIHUD, Y., HURVICH, C. and WANG, Y. (2009). Multiple-predictor regressions: hypothesis testing. *Review of Financial Studies*, **22** (1), 413–434. 21, 23

— and HURVICH, C. M. (2004). Predictive regressions: a reduced-bias estimation method. *Journal of Financial and Quantitaive Analysis*, **39**, 813–841. 4, 18, 21, 22, 23, 138

AMMANN, M. and SÜSS, S. (2009). Asymmetric dependence patterns in financial time series. *European Journal of Finance*, **15** (7), 703–719. 78

ANG, A. (2008). *The Quant Meltdown: August 2007*. Columbia CaseWorks ID 080317. 10

— and BEKAERT, G. (2002). International asset allocation with regime shifts. *Review of Financial Studies*, **15** (4), 1137–1187. 81

— and — (2007). Stock return predictability: is it there? *Review of Financial Studies*, **20**, 651 – 707. 2, 4, 39, 43, 44, 46, 140

BANSAL, R., HSIEH, D. and VISWANATHAN, S. (1993). A new approach to international arbitrage pricing. *Journal of Finance*, **48**, 1719–1747. 9

— and VISWANATHAN, S. (1993). No arbitrage and arbitrage pricing: A new approach. *Journal of Finance*, **48**, 1231–1262. 9

BARBEE JR, W., MUKHERJI, S. and RAINES, G. (1996). Do sales-price and debt-equity explain stock returns better than book-market and firm size? *Financial Analysts Journal*, **52** (2), 56–60. 104

BARBERIS, N. and HUANG, M. (2008). Stocks as lotteries: the implications of probability weighting for security prices. *American Economic Review*, **98** (5), 2066–2100. 6

— and THALER, R. (2003). A survey of behavioral finance. In G.M. Constantinides, M. Harris and R.M. Stulz (eds), *Handbook of the Economics of Finance.* Amsterdam: Elevier, pp. 1053–1128. 3

BEASLEY, J., MEADE, N. and CHANG, T.-J. (2003). An evolutionary heuristic for the index tracking problem. *European Journal of Operational Research*, **148** (3), 621–643. 88

BHANDARI, L. (1988). Debt/equity ratio and expected common stock returns: Empirical evidence. *Journal of Finance*, **43** (2), 507–528. 104

BINSBERGEN, J. and KOIJEN, R. (2010). Predictive regressions: a present-value approach. *Journal of Finance*, **65** (4), 1439–1471. 2

BOUDOUKH, J., RICHARDSON, M. and WHITELAW, R. (2008). The myth of long-horizon predictability. *Review of Financial Studies*, **21** (4), 1577–1605. 4, 20, 46

BOUYÉ, E., DURRLEMAN, V., NIKEGHBALI, A., RIBOULET, G. and RONCALLI, T. (2000). Copulas for finance: a reading guide and some applications. *Unpublished Manuscript, Financial Econometrics Research Center, City University Business School.* 78

BOYER, B., MITTON, T. and VORKINK, K. (2010). Expected idiosyncratic skewness. *Review of Financial Studies*, **23** (1), 169–202. 6

BREEN, W., GLOSTEN, L. and JAGANNATHAN, R. (1989). Economic significance of predictable variations in stock index returns. *Journal of Finance*, **44** (5), 1177–1189. 2

BREIMAN, L. (2001). Random forests. *Machine Learning*, **45** (1), 5–32. 121, 128, 153

—, FRIEDMAN, J. H., OLSHEN, R. and STONE, C. J. (1984). *Classification and Regression Trees*. Belmont, California: Wadsworth. 11, 97, 100

BRIEC, W., KERSTENS, K. and JOKUNG, O. (2007). Mean-variance-skewness portfolio performance gauging: a general shortage function and dual approach. *Management Science*, **53** (1), 135–149. 8

BROCKETT, P. and KAHANE, Y. (1992). Risk, return, skewness and preference. *Management Science*, **38** (6), 851–866. 8

BRUNNERMEIER, M., GOLLIER, C. and PARKER, J. (2007). Optimal beliefs, asset prices, and the preference for skewed returns. *American Economic Review: Papers and Proceedings*, **97**, 159–165. 6

BYLANDER, T. (2002). Estimating generalization error on two-class datasets using out-of-bag estimates. *Machine Learning*, **48** (1), 287–297. 155

CAMPBELL, J. (1987). Stock returns and the term structure. *Journal of Financial Economics*, **18**, 373–399. 2

— (2001). Why long horizons? a study of power against persistent alternatives. *Journal of Empirical Finance*, **8** (5), 459–491. 3

—, HILSCHER, J. and SZILAGYI, J. (2008). In search of distress risk. *Journal of Finance*, **63** (6), 2899–2939. 104

— and SHILLER, R. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, **1**, 195–228. 2

— and THOMPSON, S. (2008). Predicting the equity premium out of sample: Can anything beat the historical average? *Review of Financial Studies*, **21**, 1509–1531. 104

— and VUOLTEENAHO, T. (2004). Inflation illusion and stock prices. *American Economic Review*, **94** (2), 19–23. 2

CAMPBELL, J. Y. and YOGO, M. (2006). Efficient tests of stock return predictability. *Journal of Financial Economics*, **81**, 27–60. 3, 4, 5, 18, 21, 29, 43

CARHART, M. (1997). On persistence in mutual fund performance. *Journal of Finance*, **52** (1), 57–82. 9, 127

CENESIZOGLU, T. and TIMMERMANN, A. (2008). Is the distribution of stock returns predictable? *Available online at SSRN http://ssrn.com/abstract=1107185.* 7, 13, 70

CHANG, T.-J., MEADE, N., BEASLEY, J. and SHARAIHA, Y. (2000). Heuristics for cardinality constrained portfolio optimisation. *Computers and Operations Research*, **27** (13), 1271–1302. 88

CHEN, J., HONG, H. and STEIN, J. (2001). Forecasting crashes: trading volume, past returns, and conditional skewness in stock prices. *Journal of Financial Economics*, **61** (3), 345–381. 6

CHEN, L., NOVY-MARX, R. and ZHANG, L. (2011). An alternative three-factor model. *Available online at SSRN http://ssrn.com/abstract=1418117.* 104

CHERNOZHUKOV, V. and UMANTSEV, L. (2001). Conditional value-at-risk: Aspects of modeling and estimation. *Empirical Economics*, **26** (1), 271–292. 52

CHERUBINI, U., LUCIANO, E. and VECCHIATO, W. (2004). *Copula methods in finance.* New York: John Wiley. 78

CHIANG, T. and LI, J. (2007). Empirical analysis of asset returns with skewness,

kurtosis, and outliers: Evidence from 30 Dow-Jones industrial stocks. *Working paper. Drexel University.* 6

CHOW, E., LEE, W. and SOLT, M. (1997). The exchange-rate risk exposure of asset returns. *Journal of Business*, **70** (1), 105–123. 57

COCHRANE, J. (1999). New facts in finance. *Economic Perspectives*, **23**, 36–58. 3

— (2011). Presidential address: Discount rates. *Journal of Finance*, **66** (4), 1047–1108. 1, 11

COCHRANE, J. H. (2008). The dog that did not bark: A defense of return predictability. *Review of Financial Studies*, **21** (4), 1533–1575. 2, 69

CONT, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative finance*, **1** (2), 223–236. 6

DANGL, T. and HALLING, M. (2008). Predictive regressions with time-varying coefficients. *Available online at SSRN http://ssrn.com/abstract=1571670.* 140

DANIEL, K. and TITMAN, S. (2006). Market reactions to tangible and intangible information. *Journal of Finance*, **61** (4), 1605–1643. 3

DARSINOS, T. and SATCHELL, S. (2004). Generalising universal performance measures. *Risk Magazine*, **17** (6), 80–84. 84

DE ATHAYDE, G. and FLÔRES, R. (2004). Finding a maximum skewness portfolio–a general solution to three-moments portfolio choice. *Journal of Economic Dynamics and Control*, **28** (7), 1335–1352. 8

DE BONDT, W. and THALER, R. (1985). Does the stock market overreact? *Journal of Finance*, **40** (3), 793–805. 104

— and — (1987). Further evidence on investor overreaction and stock market seasonality. *Journal of Finance*, **42** (3), 557–581. 104

DE SANTIS, M. (2007). Movements in the equity premium: Evidence from a time-varying VaR. *Studies in Nonlinear Dynamics and Econometrics*, **11** (4), 1–39. 33

DECHOW, P., HUTTON, A., MEULBROEK, L. and SLOAN, R. (2001). Short-sellers, fundamental analysis, and stock returns. *Journal of Financial Economics*, **61** (1), 77–106. 104

DEMBO, R. and MAUSSER, H. (2000). The put/call efficient frontier. *Algo Research Quarterly*, **3** (1), 13–25. 86

— and ROSEN, D. (1999). The practice of portfolio replication. a practical overview of forward and inverse problems. *Annals of Operations Research*, **85**, 267–284. 86

DITTMAR, R. (2002). Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns. *Journal of Finance*, **57** (1), 369–403. 6

DUECK, G. and WINKER, P. (1992). New concepts and algorithms for portfolio choice. *Applied Stochastic Models and Data Analysis*, **8** (3), 159–178. 88

EDGEWORTH, F. (1888). On a new method of reducing observations relating to several quantities. *Philosophical Magazine*, **25**, 184–191. 7

ELLIOTT, G. and STOCK, J. (1994). Inference in time series regression when the order of integration of a regressor is unknown. *Econometric theory*, **10**, 672–700. 4

EMBRECHTS, P., MCNEIL, A. and STRAUMANN, D. (1999). Correlation: pitfalls and alternatives. *Risk Magazine*, **12**, 69–71. 78

ENGLE, R. and MANGANELLI, S. (2004). CAViaR. *Journal of Business and Economic Statistics*, **22** (4), 367–381. 52

FABOZZI, F., FOCARDI, S. and JONAS, C. (2008). On the challenges in quantitative equity management. *Quantitative Finance*, **8** (7), 649–665. 10

—, — and KOLM, P. (2006). Trends in Quantitative Finance (Digest Summary). *CFA Digest*, **36** (3), 69–72. 10

FAMA, E. (1965). The behavior of stock-market prices. *Journal of Business*, **38** (1), 34–105. 6

— (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, **25** (2), 383–417. 2

— (1981). Stock returns, real activity, inflation, and money. *American Economic Review*, **71**, 545–565. 2

— (1998). Market efficiency, long-term returns, and behavioral finance1. *Journal of financial economics*, **49** (3), 283–306. 3

— and FRENCH, K. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics*, **22** (1), 3–25. 2, 3

— and — (2008). Dissecting anomalies. *Journal of Finance*, **63** (4), 1653–1678. 2

— and SCHWERT, G. (1977). Asset returns and inflation. *Journal of Financial Economics*, **5**, 115–146. 2

FAMA, E. F. and FRENCH, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, **33** (1), 3–56. 9, 127

— and — (1996). Multifactor explanations of asset pricing anomalies. *Journal of Finance*, **51** (1), 55–84. 108

FANG, H. and LAI, T. (1997). Co-kurtosis and capital asset pricing. *Financial Review*, **32** (2), 293–307. 6

FERSON, W., SARKISSIAN, S. and SIMIN, T. (2003). Spurious regressions in financial economics? *Journal of Finance*, **58** (4), 1393–1414. 3, 140

—, — and — (2008). Asset pricing models with conditional betas and alphas: The effects of data snooping and spurious regression. *Journal of Financial and Quantitative Analysis*, **43** (2), 331–354. 3, 140

FISCHER, M., KÖCK, C., SCHLÜTER, S. and WEIGERT, F. (2009). An empirical analysis of multivariate copula models. *Quantitative Finance*, **9** (7), 839–854. 78

FRYDMAN, H., ALTMAN, E. I. and KAO, D.-L. (1985). Introducing recursive partitioning for financial classification: the case of financial distress. *Journal of Finance*, **40** (1), 269–291. 11

GASHLER, M., GIRAUD-CARRIER, C. and MARTINEZ, T. (2008). Decision tree ensemble: small heterogeneous is better than large homogeneous. In *The Seventh International Conference on Machine Learning and Applications*, pp. 900–905. 155

GIACOMINI, R. and WHITE, H. (2006). Tests of conditional predictive ability. *Econometrica*, **74** (6), 1545–1578. 52

GILLI, M., KËLLEZI, E. and HYSI, H. (2006). A data-driven optimization heuristic for downside risk minimization. *Journal of Risk*, **8** (3), 1–18. 88, 92

— and SCHUMANN, E. (2010). Distributed optimisation of a portfolio's Omega. *Parallel Computing*, **36** (7), 381–389. 88, 92

GIVOLY, D. and LAKONISHOK, J. (1980). Financial analysts' forecasts of earnings: Their value to investors. *Journal of Banking and Finance*, **4** (3), 221–233. 105

GOWLLAND, C., XIAO, Z. and ZENG, Q. (2009). Beyond the central tendency: Quantile regression as a tool in quantitative investing. *Journal of Portfolio Management*, **35** (3), 106–119. 7

GOYAL, A. and WELCH, I. (2003). Predicting the equity premium with dividend ratios. *Management Science*, **49** (5), 639–654. 4, 44, 69, 138, 140

GRANGER, C. and NEWBOLD, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics*, **2** (2), 111–120. 21

GUIDOLIN, M. and TIMMERMANN, A. (2007). Asset allocation under multivari-

ate regime switching. *Journal of Economic Dynamics and Control*, **31** (11), 3503–3544. 81

— and — (2008). International asset allocation under regime switching, skew, and kurtosis preferences. *Review of Financial Studies*, **21** (2), 889–935. 6, 8

Harvey, C., Liechty, J., Liechty, M. and Müller, P. (2010). Portfolio selection with higher moments. *Quantitative Finance*, **10** (5), 469–485. 81

— and Siddique, A. (2000). Conditional skewness in asset pricing tests. *Journal of Finance*, **55** (3), 1263–1295. 6

— and Zhou, G. (1993). International asset pricing with alternative distributional specifications. *Journal of Empirical Finance*, **1** (1), 107–131. 6

Haugen, R. (2009). *The new finance: overreaction, complexity, and their consequences*. Pearson Prentice Hall, 4th edn. 109

Hiemstra, C. and Jones, J. (1994). Testing for linear and nonlinear granger causality in the stock price-volume relation. *Journal of Finance*, **49**, 1639–1664. 9

Hodrick, R. (1992). Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *Review of Financial studies*, **5** (3), 357. 2

Hoeffding, W. (1940). Scale-invariant correlation theory. In N. Fisher and P. Sen (eds.), *The Collected Works of Wassily Hoeffding*, New York: Springer-Verlag, pp. 57–107. 77

— (1941). Scale-invariant correlation measures for discontinuous distributions. In N. Fisher and P. Sen (eds.), *The Collected Works of Wassily Hoeffding*, New York: Springer-Verlag, pp. 109–133. 77

HSIEH, D. (1991). Chaos and nonlinear dynamics: application to financial markets. *Journal of Finance*, **46**, 1839–1877. 9

HUENG, C. and McDONALD, J. (2005). Forecasting asymmetries in aggregate stock market returns: Evidence from conditional skewness. *Journal of Empirical Finance*, **12** (5), 666–685. 6

IKENBERRY, D., LAKONISHOK, J. and VERMAELEN, T. (1995). Market underreaction to open market share repurchases. *Journal of Financial Economics*, **39**, 181–208. 2

INOUE, A. and KILIAN, L. (2005). In-sample or out-of-sample tests of predictability: Which one should we use? *Econometric Reviews*, **23** (4), 371–402. 69

JEGADEESH, N. and TITMAN, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, **48** (1), 65–91. 104

— and — (2001). Profitability of momentum strategies: an evaluation of alternative explanations. *Journal of Finance*, **56** (2), 699–720. 104

JONDEAU, E. and ROCKINGER, M. (2006). Optimal portfolio allocation under higher moments. *European Financial Management*, **12** (1), 29–55. 81

JURCZENKO, E., MAILLET, B. and MERLIN, P. (2006). Hedge funds portfolio selection with higher-order moments: a non-parametric mean-variance-skewness-

kurtosis efficient frontier. In E. Jurczenko and B. Maillet (eds.), *Multi-moment asset allocation and pricing models*, John Wiley & Sons, pp. 51–66. 8

KAHNEMAN, D. and TVERSKY, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, **47**, 263–291. 84

KAO, D. and SHUMAKER, D. (1999). Equity style timing. *Financial Analysts Journal*, **55**, 37–48. 11

KENDALL, M. (1954). Note on bias in the estimation of autocorrelations. *Biometrika*, **41**, 403–404. 20, 23

KHANDANI, A. and LO, A. (2011). What happened to the quants in August 2007? Evidence from factors and transactions data. *Journal of Financial Markets*, **14**, 1–46. 10

KOENKER, R. (2005). *Quantile regression*. Cambridge University Press. 51, 53

— and BASSETT JR, G. (1978). Regression quantiles. *Econometrica*, **46** (1), 33–50. 7, 51, 52

— and — (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica*, **50** (1), 43–61. 7, 51, 52

— and D'OREY, V. (1987). Computing regression quantiles. *Applied Statistics*, **36** (3), 383–393. 53

KOTHARI, S. and SHANKEN, J. (1997). Book-to-market, dividend yield, and expected market returns: a time-series analysis. *Journal of Financial Economics*, **44** (2), 169–203. 2

LAKONISHOK, J., SHLEIFER, A. and VISHNY, R. W. (1994). Contrarian investment, extrapolation, and risk. *Journal of Finance*, **49** (5), 1541–1578. 104, 108

LAMONT, O. (1998). Earnings and expected returns. *Journal of Finance*, **53**, 1563–1587. 2

LANNE, M. (2002). Testing the predictability of stock returns. *Review of Economics and Statistics*, **84** (3), 407–415. 4

LEITCH, G. and TANNER, J. (1991). Economic forecast evaluation: profits versus the conventional error measures. *American Economic Review*, **81**, 580–590. 73

LETTAU, M. and LUDVIGSON, S. (2001). Consumption, aggregate wealth, and expected stock returns. *Journal of Finance*, **56** (3), 815–849. 2, 140

— and — (2005). Expected returns and expected dividend growth. *Journal of Financial Economics*, **76** (3), 583–626. 2

— and VAN NIEUWERBURGH, S. (2008). Reconciling the return predictability evidence. *Review of Financial Studies*, **21** (4), 1607–1652. 2

LEWELLEN, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics*, **74** (2), 209–235. 4, 5, 43

LI, D. (2000). On default correlation: a copula function approach. *Journal of Fixed Income*, **9** (4), 43–54. 78

LIN, D. and TU, W. (1995). Dual response surface optimization. *Journal of Quality Technology*, **27** (1), 34–34. 43

LINTNER, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, **47** (1), 13–37. 9, 101

LOUGHRAN, T. and RITTER, J. (1995). The new issues puzzle. *Journal of Finance*, **50** (1), 23–51. 3

LYS, T. and SOHN, S. (1990). The association between revisions of financial analysts' earnings forecasts and security-price changes. *Journal of Accounting and Economics*, **13** (4), 341–363. 105

MA, L. and POHLMAN, L. (2008). Return forecasts and optimal portfolio construction: a quantile regression approach. *European Journal of Finance*, **14** (5), 409–425. 7

MARINGER, D. (2005). *Portfolio management with heuristic optimization*. Springer Verlag. 88

MARKOWITZ, H. (1952). Portfolio selection. *Journal of Finance*, **7** (1), 77–91. 8

MARRIOTT, F. and POPE, J. (1954). Bias in the estimation of autocorrelations. *Biometrika*, **41**, 390–402. 20

MENCÍA, J. and SENTANA, E. (2009). Multivariate location-scale mixtures of normals and mean-variance-skewness portfolio allocation. *Journal of Econometrics*, **153** (2), 105–121. 8, 81

MERTON, R. C. (1982). On the mathematics and economics assumptions of continuous-time models. In W.F. Sharpe and C.M. Cootner (eds.), *Financial*

Economics: Essays in honor of Paul Cootner, Englewood Cliffs: Prentice-Hall. 6

MITTON, T. and VORKINK, K. (2007). Equilibrium underdiversification and the preference for skewness. *Review of Financial Studies,* **20** (4), 1255–1288. 6

MOSSIN, J. (1966). Equilibrium in a capital asset market. *Econometrica,* **34** (4), 768–783. 9

NELSEN, R. (2006). *An Introduction to Copulas.* Springer, 2nd edn. 77, 78

NELSON, C. and KIM, M. (1993). Predictable stock returns: the role of small sample bias. *Journal of Finance,* **48** (2), 641–661. 3, 4, 5, 41

NEWEY, W. and WEST, K. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica,* **55** (3), 703–708. 46

OHLSON, J. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research,* **18** (1), 109–131. 122

PATTON, A. (2006). Modelling asymmetric exchange rate dependence. *International Economic Review,* **47** (2), 527–556. 78

PAYE, B. and TIMMERMANN, A. (2006). Instability of return prediction models. *Journal of Empirical Finance,* **13** (3), 274–315. 44, 140

PEDERSEN, T. Q. (2010). Predictable return distributions. *Available online at SSRN http://ssrn.com/abstract=1658394.* 7, 13, 81

PHILLIPS, P. and YU, J. (2009). Simulation-based estimation of contingent-claims prices. *Review of Financial Studies*, **22** (9), 3669–3705. 31

PONTIFF, J. and SCHALL, L. D. (1998). Book-to-market ratios as predictors of market returns. *Journal of Financial Economics*, **49**, 141–160. 2

— and WOODGATE, A. (2008). Share issuance and cross-sectional returns. *Journal of Finance*, **63** (2), 921–945. 3

POON, S. and GRANGER, C. (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, **41** (2), 478–539. 5

POST, T., VAN VLIET, P. and LEVY, H. (2008). Risk aversion and skewness preference. *Journal of Banking and Finance*, **32** (7), 1178–1187. 6

QUENOUILLE, M. (1949). Approximate tests of correlation in time-series. *Journal of the Royal Statistical Society*, **11**, 68–84. 23

— (1956). Notes on bias in estimation. *Biometrika*, **43**, 353–360. 23

RAPACH, D., STRAUSS, J. and ZHOU, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy. *Review of Financial Studies*, **23** (2), 821 – 862. 65

ROSENBERG, B. and GUY, J. (1995). Prediction of beta from investment fundamentals. *Financial Analysts Journal*, pp. 101–112. 31

ROSS, S. A. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, **13** (3), 341–360. 9, 101

ROZEFF, M. (1984). Dividend yields are equity risk premia. *Journal of Portfolio Managment*, **49**, 141160. 2

SATTEN, G., DATTA, S., MOURA, H., WOOLFITT, A., CARVALHO, M., CARLONE, G., DE, B., PAVLOPOULOS, A. and BARR, J. (2004). Standardization and denoising algorithms for mass spectra to classify whole-organism bacterial specimens. *Bioinformatics*, **20** (17), 3128–3136. 154

SCHUCANY, W., GRAY, H. and OWEN, D. (1971). On bias reduction in estimation. *Journal of the American Statistical Association*, **66**, 524–533. 23

SEGAL, M. (2003). Machine learning benchmarks and random forest regression. *Available online at http://escholarship.org/uc/item/35x3v9t4*. 155

SHADWICK, W. and KEATING, C. (2002). A universal performance measure. *Journal of Performance Measurement*, **6** (3), 59–84. 13, 15, 83

SHARPE, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, **19** (3), 425–442. 9, 101

SIMONOFF, J. (1993). The relative importance of bias and variability in the estimation of the variance of a statistic. *The Statistician*, pp. 3–7. 31

SKLAR, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, **8**, 229–231. 77

SLOAN, R. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings? *Accounting Review*, **71** (3), 289–315. 2

SORENSEN, E., MEZRICH, J. and MILLER, K. (1998). A new technique for tactical asset allocation. In F. Fabozzi (ed.), *Active Equity Portfolio Management*, New Hope, Pennsylvania, pp. 195–202. 11

SORENSEN, E. H., MILLER, K. L. and OOI, C. K. (2000). The decision tree approach to stock selection. *Journal of Portfolio Management*, **27** (1), 42–52. 11

STAMBAUGH, R. (1999). Predictive regressions. *Journal of Financial Economics*, **54** (3), 375–421. 3, 4, 18, 20, 21, 22, 29, 138, 140

TAYLOR, J. (1999). A quantile regression approach to estimating the distribution of multiperiod returns. *Journal of Derivatives*, **7** (1), 64–78. 52

TIMMERMANN, A. (2006). Forecast combinations. In G. Elliott, C. Granger and A. Timmermann (eds.), *Handbook of economic forecasting*, Amsterdam, the Netherlands: Elsevier, pp. 135–196. 65

TOROUS, W., VALKANOV, R. and YAN, S. (2004). On predicting stock returns with nearly integrated explanatory variables. *Journal of Business*, **77** (4), 937–966. 3, 4

TRUEMAN, B. (1994). Analyst forecasts and herding behavior. *Review of financial studies*, **7** (1), 97–124. 105

VALKANOV, R. (2003). Long-horizon regressions: theoretical results and applications. *Journal of Financial Economics*, **68** (2), 201–232. 4, 5, 36

VAN DER SMAGT, T. and LUCARDIE, L. (1991). Decision making under not-

well-defined conditions: from data processing to logical modelling. *Tijdschrift voor economische en sociale geografie*, **82** (4), 295–304. 102

WANG, Y.-G., SHAO, Q. and ZHU, M. (2009). Quantile regression without the curse of unsmoothness. *Computational Statistics and Data Analysis*, **53** (10), 3696–3705. 54, 62, 150

WELCH, I. and GOYAL, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, **21**, 1455–1508. 44, 70

WHITE, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, **50**, 1–25. 53

YU, J. (2012). Bias in the estimation of the mean reversion parameter in continuous time models. *Journal of Econometrics. Available online 20 January 2012.* 31

YULE, G. (1926). Why do we sometimes get nonsense-correlations between time-series? – a study in sampling and the nature of time-series. *Journal of the Royal Statistical Society*, **89** (1), 1–63. 21