ALTERNATIVE PORTFOLIO METHODS

by

RUANMIN CAO

A thesis submitted to the University of Birmingham for the degree of
DOCTOR OF PHILOSOPHY

Department of Economics

Business School

College of Social Science

University of Birmingham

April 2015

Abstract

Portfolio optimization in an uncertain environment has great practical value in investment decision process. But this area is highly fragmented due to fast evolution of market structure and changing investor behavior. In this dissertation, four methods are investigated/designed to explore their efficiency under different circumstances.

Parametric portfolio decomposes weights by a set of factors whose coefficients are uniquely determined via maximizing utility function. A robust bootstrap method is proposed to assist factor selection. If investors exhibit asymmetric aversion of tail risk, pessimistic models on Choquet utility maximization and coherent risk measures acquire superiority. A new hybrid method that inherits advantage of parameterization and tail risk minimization is designed. Mean-variance, which is optimal with elliptical return distribution, should be employed in the case of capital allocation to trading strategies. Nonparametric classifiers may enhance homogeneity of inputs before feeding the optimizer. Traditional factor portfolio can be extended to functional settings by applying FPCA to return curves sorted by factors. Diversification is always achieved by mixing with detected nonlinear components.

This research contributes to existing literature on portfolio choice in three-folds: strength and weakness of each method is clarified; new models that outperform traditional approaches are developed; empirical studies are used to facilitate comparison.

Contents

List of Figures

List of Tables

1 Introduction

Portfolio selection is probably one of the most dynamic area in modern financial theory. It has broad connection with preference under uncertainty, forecasting techniques of stationary/non-stationary time series and stochastic price behavior. Optimizations of asset allocation are designed as a set of methodologies that assemble aforementioned components to obtain a weighting policy, rather than integrated stand-alone systems. Consequently portfolio performance is not uniquely determined by the framework adopted, but bounded jointly by accuracy of inputs and model validity. This entangling relation, together with unavoidable heterogeneous nature of data, causes assessment of any approach excessively challenging.

Another difficulty in this field is absence of a general paradigm or principle that characterizes current portfolio optimizers. The chaos may arise from the dilemma between theoretical robustness and practical efficiency. After over six decades of academic exploration since Markowitz first paper, there is seldom optimization scheme with general applicability. Unfortunately this conflict is likely to persist in the foreseeable future.

With regard to insufficiency of existing literature, this dissertation might be considered as a three-dimensional attempt to reach the high frontier: developing new approaches in the framework of utility maximization; conducting robust statistical tests; simulating with real data. Although it is not as ambitious as covering every piece of the highly fragmented area, fresh interpretations of influential methods might contribute to better understanding of market dynamics and complexity we are facing. In this section, I intend to necessitate the study by discussing weakness and strength of several portfolio schemes.

1.1 Mean-Variance

Modern portfolio theory was pioneered by Markowitz (1952). The fundamental relation between expected return and its risk measure plays a critical role in multi-dimensional choice under uncertainty. The proposed optimization was later solved as quadratic programming (Markowitz (1959)). Despite of theoretical dilemma and calibration problem, it is the first model that explicitly characterizes portfolio choice under conditions of risk as a dual optimization paradigm. It essentially provides Pareto-style welfare maximization in that investors at the efficient frontier are prohibited from increasing their return in exchange of decreasing risk. Another contribution, which also leaves a Achilles heel, is disentangling return and risk that are traditionally nested in expected utility. The separation advantageously makes asset allocation a structured selection where forecast techniques might be employed.

Following researches that proliferated on similar groud extended to asset allocation in capital market (Sharpe, 1964, 1965, 1970), which is subsequently known as CAPM model, and life-time portfolio choice (Merton, 1973). The former uses equilibrium approach to analyz capital asset price behavior and construct capital market line on which investors are able to gain higher return in compensation of assuming additional risk. It introduces well-known systematic risk that cant be diversified, beta that measures exposure on systematic risk for individual stocks and Sharpe ratio that describes risk-reward profile. The latter is an intertemporal portfolio choice, aiming at maximizing expected summation utility of consumption and terminal wealth subject to stochastic budge equation. Utility functional is chosen to be parabolic which is shown to hold Lie symmetry and reserves economic rationality.

However literature has never ceased criticizing mean-variance for the lack of realistic assumption. Quiggin (1981, 1993) argued that quadratic utility function is counter-intuitive and volatility as risk measure is not consistent with existing preference. Despite recent defense of mean-variance (Markowitz (2012)) as a quadratic utility approximation, it is still sensitive to distributional condition of asset return. More specifically, only the family of elliptic distribution is applicable to MV utility functions (Chamberlain, 1983).

Michaud (1989) discussed the unavoidable enigma of MV that it intends to magnify inaccuracies of inputs. Estimation errors in mean and variance lead to significant deviation from efficient frontier and severe out-of-sample degeneration. Additionally, it has a tendency of overconcentration as expected return is targeted higher. This risk-taking arrangement worsens performance.

To alleviate delicacy of calibration and enhance mean-variance feasibility, literatures focused on designing robust techniques bettering forecast of return and its risk. Michard (1989) proposed Stern shrinkage to smooth out temporary disturbance of return. Another popular approach is bootstrapping original data set to construct a resampled frontier. Black and Litterman (1992) introduced a Bayesian model by incorporating investors perspective. The method is that equilibrium expected return in CAPM is adjusted to reflect some particular information one has. Alternatively, we may first obtain prevailing forecast of return and variance by reversely engineering market weights and modify the benchmark portfolio according to additional information. It is advantageous in that diversification achieved by market portfolio is not harmed compared to plug-in method. Performance is no longer mechanically dependent on techniques employed. To capture time-varying property of covariance matrix, Engle (2002, 2009) designed dynamic conditional correlation to estimate correlations of large system of assets. The procedure basically is using volatility adjusted return after ′DE-GARCHING′ to measure decomposed correlation and the standardization (rescaling) extracts the dynamic matrix from data. Both simulation and real data record an improved performance. But it restricts applicability in normality or t distribution assumption which is often rejected in financial data. Pesaran and Pesaran (2007) extended the model in multivariate case and applied the method to futures market. Laloux et al. (2000) investigated correlation matrix with limited samples. They showed that small eigenvalues of the matrix are likely to be tortured and contain more observational noise. Thus stable relations are expected to reserve in principal components. Lai et al. (2011) converted dual problem into a stochastic optimal control with specific risk aversion so that mean-variance framework can be conducted even without pre-specified location and risk values. This approach tactically avoids direct estimation of inputs that suffers heavily from errors thus providing rewarding risk return profile close to true frontier.

Besides polishing techniques of calibration, recent advancement concentrates on application in continuous scenarios. The typical setting is assuming stock price behavior is shaped by geometric Brownian motions. By demonstrating the connection between volatility matrix, which is the square root of covariance matrix, and return vector, Lindberg (2009) analytically solved Markowitz problem in continuous time. It was applied in industry sector data set and a significant boost of Sharpe ratio indicates better volatility adjusted return relative to naive strategy. This model was then generalized by Alp and Korn (2011) to introduce jump process in stock price behavior. Their contribution is a restatement of optimality condition and an accurate interpretation of optimal strategies.

Inefficiency of plug-in MV using a fixed window length can be caused by heterogeneity of data sample. Kernel classifiers, concerning the inconsistency, offer purification approach to modeling inputs of mean-variance portfolio. Nonparametric methods such as k-nearest neighbors (knn) quantify similarities by features and calculate best estimator based on most similar observations. Their efficiency was first investigated by Cover (1968). Later researches was led by Short, R. and Fukanaga, K. (1980,1981). Concerning bias in high dimensions, Hastie and Tibshirani (1996) combines linear discriminant analysis with nearest neighborhood classification (DANN). It shrinks orthogonal to local decision boundaries determined in last round. Iterative convergence is expected to be achieved with locally best behavior at center. Delannay et al. (2006) improved DANN with automatic adjustment to hyper-parameters, avoiding training sample over-fitting. In Section 4, I apply conventional mean-variance to strategy selection, where technical features are used to gathering similar observations.

1.2 Minimizing Tail Risk

Paralleling to burgeoning literature on both critiques and improvement on mean-variance method, dual optimization with alternative risk measure was actively explored. Accounting for the failure to explain Ellsberg paradox, Schmedler (1986, 1989) and Quiggin (1981) cast doubt on additive expected utility theory established by von Neumann and Morgenstern (1944) and Salvage (1954). Comonotonicity was then proposed to replace additivity. Together with non-degeneracy, continuity, state independence, non-additive utility leads to subjective probability that is perceived by investors. Mathematically, capacity should be used to distort original probability and Choquet integral is applicable. Gilboa and Schmeidler (1992) further discussed the properties of set function and properties of Choquet integrals. Some results are the Choquet integral accumulates on the lower boundaries of the integrand, Radon-Nikodym derivative is definable and formulating Bayesian update.

Inspired by early studies on preferences, Artzner et al. (1999) axiomatized the celebrated coherent risk measure by four properties: monotonicity, sub-additivity, positive homogeneity and translation invariance. In this framework, traditional volatility and value-at-risk are no longer valid. Instead, alpha risk, known also as expected shortfall, conditional value-at-risk, serves as an important instance of coherent risk measure that has subsequently been elaborately investigated.

There is comparable literature on convexity, spectral risk measures, and their connection with coherence (Follmer, 2008, 2010). A recent advance by Follmer (2013) is an attempt to make analogy between statistical mechanics and risk structure. The so-called spatial risk measure considers topology of financial institutions. Consistency is defined as a mapping from original information set to its subset. This property, with law-invariant, strongly sensitive and Labesgue property, restricts risk measure to be in entropic form. Indication from the major theorem is that aggregation of local risk measure is possible with delicate settings. But if uniqueness of global risk is violated, a transition phase is present which offers another aspect of general systematic risk.

With the spirit of retaining dual optimization paradigm but replacing volatility with coherent risk measures, Rockafellar and Uryasev (1999) designed a portfolio optimization by minimizing VaR and CVaR simultaneously. The basic idea is to translate portfolio construction into minimizing an objective function where stochastic programming applies. The trick is that first order condition of the function respect to hyper parameter gives exact expression of alpha risk. So duality is successfully avoided. Solvability is further discussed by discretization. Since the sample equation is linear and convex, linear programming is naturally applicable. Another solution is approximated by mean-variance approach but accuracy restricted. Krokhmal at al. (2001) extended this model to generalized class of problems within the structure of CVaR constrained maximization. Stacked objective functions with set of confidence levels are jointly optimized and transaction cost is tackled in linear programing. The classical objective function was even generalized by Krokhmal (2007) by interpreting it as a special case of another measure. The underlying measure can be chosen to be expectation degenerating to Rockafellars model. With higher moments, optimization of coherent measure is decomposed to large set of linear inequalities. It records better performance on S&P 500 than mean-variance and optima using CVaR.

Koenker (2005) formulated optimization from the perspective of quantile regression. Pessimistic risk is first defined if it is Choquet Integrable by some alpha risk. It is essentially a reinterpretation of coherent risk measure by Artzner et al. (1999). Central theorem is that Choquet utility distorted by CVaR is equivalent to a classical quantile loss function. As coherence can always be framed by pessimism, stacked quantile regression has the power of optimizing portfolio using CRM. This conclusion is in accordance with Krokhmals (2007) assertion. Albeit tail risk measure is theoretically robust, its practicability is even questionable since the distribution of extreme events is considerably volatile.

In Section 3, I discuss the set of existing pessimistic models which shows equivalence of Koenker (2005) and Rockafellar & Uryasev (1999). A parameterized method is proposed to tackle instability of tail risk. Its robustness is empirically evidenced in country indics investment.

1.3 Portfolio Parameterization

Since portfolio selection discussed above sticks to maximizing rewards and minimizing risk jointly, disadvantages, inherited from mean-variance, are unavoidable. Over exposure on risk with increasing target return erodes the benefit of diversification and results in poor out-of-sample performance. Choosing risk measure seems to fail in resolving the problem.

Obtaining portfolio policy directly by utility maximization, in hope of bypassing deficiencies in dual paradigm, is proposed by some researchers. By introducing efficient predictors, Ait-Sahalia and Brandt (2001) proposed linear model with risk drivers that determines first two moments of returns. The idea of mimicing dynamically rebalanced strategy with static portfolios leads to parametric portfolio by Brandt and Santa-Clara (2006). Two aspects are considered in a conditional portfolio. One is the variable chosen to make capital tilting toward stocks with specific feature. Another is a mixed multi-period arrangement which invests risk assets for one period and risk-free one otherwise. Static Markowitz problem is then typically solved numerically as a linear function of state variables.

Brandt, Santa-Clara and Valkanov (2009) applied similar methodology to all listed stocks in US from January 1964 to December 2002 to infer best allocation. The mechanism is simple: it parameterizes policy instead of return, and optimizes it by maximizing utility function. The difference from earlier studies is that weights are linearly decomposed by firm-specific features. It has similarity with traditional cross-sectional factor portfolio in that characteristics are standardized to ensure overall neutrality. This consistency builds asymmetric connection between the two since factor model is the first order Taylor expansion of parametric method. Experiment results showed that parameterization outperforms naive strategy and market portfolio. Another value is that it functions well out of sample without losing significant efficiency. However nonlinear functions are not necessarily numerically insolvable. The method is also inconvenient when we need to pick useful factors from a large set of candidates.

White (2000) designed reality check to test data-snooping problem. Null hypothesis is constructed as profit of the strategy does not exceed that of benchmark. Since normality holds asymptotically, p-value is acquired by stationary bootstrap (Romano and Politis, 1994). Sullivan et al.(1999) applied the methodology to test profitability of a broad set of trading strategies including filter rules, moving average, support and resistance, channel breakout, on-balance volume averages. They used both return and Sharpe ratio as loss function in statistics. Benchmark is buy-and-hold strategy. Their experiment documented remarkable outperformance in some subsets. Giacomini and White (2006) explored a forecast technique of predictive ability based on kernel of functions. Since data generating process is unknown, they construct out-of-sample statistic with rolling window to capture heterogeneity. Parameterization of portfolio weights, equipped with a robust factor selection, as I present in Section 2, can deliver decent risk-adjusted return on FT100.

1.4 Pricing Risk Factors

Besides being rooted to utility maximization, a popular portfolio among practioners is to gain exposure on one risk factor and remove systemic risk by ensuring market neutrality. Sufficient condition of profitability is that market behaviorally prices this anomaly. Traditionally, efficiency of firm size, book-to-market ratio and momentum have been widely tested in various stock markets (Fama and French, 1992; Carhart 1997). More recent studies can be roughly categorized either as an attempt to exploring new factors (Asness et al. (2013)) or test on existing ones with different data sets (Fama-French, 2000)

The popular techniques employed in analysis are Fama-Macbeth regression (cross-sectional) and time-serial approach. The former is to conduct time-series analysis in a rolling sample and run cross-sectional OLS on de-correlated data. Both of them implicitly assume that cross-section of returns is linearly correlated with factors that can be readily violated. As the final piece of this research, a functional dependence test is designed to capture nonlinear dynamics of predictability. The idea is simply applying functional principal component analysis on return curves that are sorted by the investigated factor.

The advantage is two-fold. It is an extensive framework that includes linear relation as a special component. Quadratic basis function, as we can see in Section 5, remarkably contributes variation of functional returns. A mixed portfolio is proved to achieve better risk-adjusted return than traditional factor ones. Secondly, profitability of specific functionals of factors can be quantified by the corresponding coefficients. It is then possible to analyse the dynamics and even ambitiously predict portfolio performance.

2 Portfolio Parameterization

2.1 Background

Portfolio management is primarily as well as ultimately a mechanism of selecting one or one set of optimal portfolios from candidate assets for the purpose of obtaining the most desirable risk-reward profile. In other words, the selected portfolio is supposed to maximize assets expected return under risk constraints, or minimize portfolio risk for given return, see Markowitz (1952, 1959). Although subsequent researches endeavored to enhance dual problem paradigm by using asymmetric risk measure (Schmeidler (1989), Koenker (2005)) or more accurate calibration(Michaud (1989), Jobson and Korkie (1980)), these modifications cant circumvent problems arisen from making any presumptions about potential return and risk. Some researchers developed alternative methods to MV optimization, skipping the return prediction procedure and focusing directly on the portfolio weights. (See Ait-Sahalia and Brandt (1994); Nigmatullin, 2003; Brandt and Santa-Clara (2006); Santa-Clara and Saretto, 2006; Brandt et al. (2009))

By introducing predictors and Akaike Information Criterion (AIC), Ait-Sahalia and Brandt (1994) had identified linear combinations of risk drivers that potentially contributes to shaping first two moments of returns. Brandt and Santa-Clara (2006) designed parametric portfolio policy in dynamic selection setting. They, with the belief that a static managed portfolio selection approximated dynamic portfolio strategy, expanded the asset space to include mechanically managed portfolios, namely conditional portfolios that invest in each basis asset an amount proportional to conditioning variables and timing portfolios which invest in each basis asset for a single period and in the risk-free asset for all other periods. They then solved a static Markowitz problem under expanded asset space by parameterizing the portfolio policy as a linear function of state variables.

Brandt et al. (2009) applied similar methodology to all listed stocks in US from January 1964 to December 2002 to infer best allocation. The difference from earlier studies is that weights are linearly decomposed by firm-specific features. Unlike factor model by Fama-French (1992), it parameterizes policy instead of return, and optimizes it by maximizing utility function. Experiment results showed that the method outperforms naive strategy and market one with major measures. Another value is that it functions well out of sample without losing significant efficiency.

A robust selection of factors that can capture dynamics of return play an important role before applying parametric portfolio policy. Most existing literature are focused on seeking for source of returns. Fama and French (1992) investigated cross-sectional interpretability of size, value and excess market return. The finding is empirically supported by the profitability of extreme spread portfolios (which is subsequently known as factor portfolio). This framework was then extended by Fama and French (1993, 1998). Carhart (1997) incorporate momentum effect remaining unexplained by FF factors. It indicates that price movement inertia can be employed in portfolio construction. In Section 5 I will discuss functional dependence detection on predictability of factors, including the simple idea a special case. Subsequent researches are either on testing popular risk factors with different data sets or modification of indicators aiming at enhancing efficiency.

Persistence in price movement is validated by profitability of momentum factors. It bring additional conjecture on the value of technical trading rules which has received little attention in literature, since developed market is generally considered to be at least weakly efficient. However due to its nonlinear nature and simplicity in constructing feasible testing sample, popular technical indicators can serve as proper inputs of parameterized optimization. Early researches by Brock et al. (1992) simulated trading strategy performance of moving average and channel breakout on Dow Jones Index from 1897 to 1986. A larger sets including MACD, RSI and KDJ were then investigated by Ye (2011), whose aim is to identify the difference between two empirical samples with/without conditioning on the rules. One study on UK FTSE30 was conducted by Hudson et al. (1996). They report implementation difficulty because of low marginal profits.

Lack of nonparametric test on predictability of factors remains the major obstacle in applying factors. One popular method was proposed by White (2000). He designed a statistic on conditional excess return with mild assumption of underlying distribution. In consideration of unobserved process return follows and potential data-snooping, stationary bootstrap is equipped to generate p-value.

The method was employed by Sullivan et al. (1999) in testing validity of technical rules on daily return of Dow Jones Industrial Average (DJIA) from 1987 through 1996. This approach is can be used in factor selection if profitability of a trading strategy is equivalent to predictability of an indicator. It facilitates a mechanical selection process which is robust against in-sample over fitting.

The left part of this paper is organized as follows: Basic model is first introduced and discussed with candidates of extensions; we then discussed how to convert predictability into profitability where reality check applies; Finally the methodology is evidenced by a real-data example followed by conclusion.

2.2 Model Configuration

2.2.1 Basic Model

Following Brandt et al. (2009), Brandt and Santa-Clara (2006), lets assume the number of investable asset at time t is N_t which is usually large and complete.[1] A portfolio policy is a vector $w = [w_i]'$ of capital invested in each asset.Then rational investors behavior is thus to choose the optimal policy at time t that maximizes expected utility of portfolio return at time $t+1$:

$$\max_w E_t u(r_{p,t+1}) = \max_w E_t u(\sum_{i=1}^{N_t} w_{i,t} r_{i,t+1}) \tag{2.1}$$

If the optimal weight vector is decomposed by benchmark vector and linearized combination of stock characteristics, it is able to construct a parameterized

[1] Completenesss ensures that all investable assets forms a σ-algebra

policy in the form of

$$w_{i,t} = \overline{w}_{i,t} + \frac{1}{N_t}\beta^{'} f_{i,t} \tag{2.2}$$

Where $\overline{w}_{i,t}$ represents weight assigned to asset i, $f_{i,t}$ is cross-sectionally standardized factors with zero mean and unit standard deviation to guarantee condition $\sum_{i=1}^{N_t} w_{i,t} = 1$. β is contribution parameter to be estimated. Replace equation (2.1) with $w_{i,t}$ gives

$$\max_{w} E_t u(r_{p,t+1}) = \max_{\beta} E_t u(\sum_{i=1}^{N_t}(\overline{w}_{i,t} + \frac{1}{N_t}\beta^{'} f_{i,t}) r_{i,t+1}) \tag{2.3}$$

This model does not consider multi-period portfolio optimization as the one studied by Merton (1973). In equation (2.1),distributional stationarity should be satisfied for robust forward forecast. Scaling factor $\frac{1}{N_t}$ facilitates its application in time-varying asset sets. One should note that each factor have the same marginal effect on deviation from $\overline{w}_{i,t}$ accross assets.

Practically objective function (2.1) is discretized with a sample from 0 to T as

$$\max_{w} \frac{1}{T}\sum_{t=0}^{T} u(r_{p,t+1}) = \max_{\beta} \frac{1}{T}\sum_{t=0}^{T} u(\sum_{i=1}^{N_t}(\overline{w}_{i,t} + \frac{1}{N_t}\beta^{'} f_{i,t}) r_{i,t+1}) \tag{2.4}$$

Since asset return, factors and benchmark weights are known in training period, β is estimated via unconstrained convex optimization.

A special case of (2.3) is to choose utility function $u(x) = x$. It degrades to factor model, to see this point,

$$E_t \sum_{i=1}^{N_t}(\overline{w}_{i,t} + \frac{1}{N_t}\beta^{'} f_{i,t}) r_{i,t+1} = E_t r_{b,t+1} + \frac{1}{N_t}\beta^{'} E_t r_{t+1} \tag{2.5}$$

Where $F_t = [f_{1,t}, ..., f_{N_t,t}] = [f^{'}_{t,1}, ..., f^{'}_{t,K}]^{'}$,$f_{1,t}$ is factor of stock i, while $f^{'}_{t,j}$ is the cross-sectional vector of factor j,$r_{t+1} = [r_{1,t+1}, ..., r_{N_t,t+1}]^{'}$ and $E_t r_{b,t+1}$ are returns on individual assets and benchmark portfolio respectively.And $E_t F_t r_{t+1}$ is factor portfolio returns.If factors are source of individual asset return or $r_{t+1} = F_t^{'}\beta$, (2.5) becomes

$$E_t r_{b,t+1} + \frac{1}{N_t}\beta^{'} E_t V_t \beta \tag{2.6}$$

V_t is covariance matrix or risk measure of factors. (2.6) is thus essentially a factor model with risk exposure determined by β.

Parameterization also leads to separation between passive portfolio and long-short strategic allocation:

$$\max_w E_t u(r_{p,t+1}) = E_t r_{b,t+1} + \max_\beta \frac{1}{N_t}\beta^{'} E_t r_{t+1} \tag{2.7}$$

Parameterization therefore estimates expected factor covariance instead of asset covariance. It avoids misperceived correlation when a set of common risk-drivers is shared (Roll, 2013).

2.2.2 Some Utility Functionals

Now consider an investor has different utility functions of returns. It is typical in commercial banks, structured products, leveraged investment, and exotic securities. Changing in risk tolerance or reward expectation can lead to utility functional shift. One scenario is that an investor needs additional maintenance margin once loss exceeds beyond some level. This may significantly increase prudence after margin call is hit. To capture the piecewise characteristics, the following lemma applies.

Specific functional form as a reflection of investors risk/reward profile must be chosen. The simplest is (2.5) and a more complicated one is given by constant relative risk aversion which is popular in both economic analysis and Mertons model (1973).

$$u_{CRRA}(r_{p,t+1}) = \frac{(1 + r_{p,t+1})^{1-\gamma}}{1-\gamma} \tag{2.8}$$

Thus objective function (2.3) becomes

$$\max_{\beta} E_t(\sum_{i=1}^{N_t}(\overline{w}_{i,t} + \frac{1}{N_t}\beta' f_{i,t})r_{i,t+1})^{1-\gamma}/(1-\gamma) \tag{2.9}$$

Which is a classical unconstrained convex optimization that can be solved by Newtons method. Perets and Yashiv (2012) discussed fundamental characteristics of HARA functional to economic analysis which ensures optima scale invariant. This property is crucial in reserving linearity of the optimal solution set and independence of portfolio policy from initial wealth.

Lemma 2.1 (Piecewise Utility) *For any piecewise utility function in the form of*

$$E_t u(r_{p,t+1}) = E_t \sum_{i=1}^{N} 1_{\{r_{p,t+1}\in\Omega_i\}} u_i(r_{p,t+1}) \tag{2.10}$$

Joint maximization using (2.3) is also piecewise combination of separate utility maximization

$$\max_{\beta} E_t u(r_{p,t+1}) = \sum_{i=1}^{N} 1_{\{r_{p,t+1}\in\Omega_i\}} \max_{\beta} E_t u_i(r_{p,t+1}) \tag{2.11}$$

if $E_t u_i(r_{p,t+1}) < \infty, \forall i$

Intuitively, Lemma (2.1) can be explained to choose the function with highest value. Thus we are able to construct a parametrically solvable utility function.

Lemma 2.2 (Convolutional Utility) *For any utility function in the form of*

$$E_t u(r_{p,t+1}) = \max_i E_t u_i(r_{p,t+1}) \tag{2.12}$$

Where $i = 1, ..., N_u$, maximization using (2.3) is

$$\max_\beta E_t u(r_{p,t+1}) = \max_\beta \max_i E_t u_i(r_{p,t+1}) = \max_i \max_\beta E_t u_i(r_{p,t+1}) \tag{2.13}$$

Because we can always find an array of Ω_i,so that utility (2.12) can be converted to piecewise utility (2.10). Both Lemma (2.1) and (2.2) states that we can separately estimate coefficients β rather than optimizing objective function simultaneously, which is much more convenient in implementation.

2.2.3 Extensions

If some distortion function v is introduced, Choquet integral replaces expectation in utility (2.3):

$$E_{t,\nu} u(r_{p,t+1}) = \int_{-\infty}^{+\infty} u(r_{p,t+1}) d\nu[F(r_{p,t+1})] \tag{2.14}$$

Where F is cumulative distribution function of portfolio. Denote $F(r_{p,t+1}) = y \in [0, 1]$. As F is monotonic, $r_{p,t+1} = F^{-1}y$. (2.14) becomes

$$\int_{-\infty}^{+\infty} u(r_{p,t+1}) d\nu[F(r_{p,t+1})] = \int_0^1 u(F^{-1}(y)) d\nu(y) \tag{2.15}$$

Let $\nu_\alpha(x) = \min(1, \frac{x}{\alpha})$, maximizing (2.3) is converted to classical quantile regression.This approach is called pessimistic portfolio (Rockafellar et al., 2000, Koenker 2005).

Following Yacine and Brandt (2000), we may incorporate ambiguity as Choquet integral. The difference is that capacity v is not pre-specified. Min-max problem in this framework is

$$\min_{\nu} \max_{w} \int_{-\infty}^{+\infty} u(r_{p,t+1}) d\nu[F(r_{p,t+1})] \tag{2.16}$$

if ν is also parameterized with functional form of alpha-risk, minimization in (2.16) shrinks to estimate α instead of choosing functional of ν

So the objective function is in the case of $\nu_\alpha(x)$

$$\int_0^1 u(F^{-1}(y)) d\nu_\alpha(y) = \int_0^\alpha u(F^{-1}(y)) dy \tag{2.17}$$

Thus the right side is tractable pessimistic portfolio optimization. Denote optimal weights

$$w^* = w^*(r, \alpha) \tag{2.18}$$

By varying target return $w^{'} r = \overline{r}_{p,t+1}$ it constructs efficient frontier. With different α (2.18) indicates efficient area. Since $\Gamma(\alpha) = \int_0^\alpha u(F^{-1}(y))dy$ s continuous, its first order condition with respect to α is

$$\frac{\partial \Gamma}{\partial \alpha} = u(F^{-1}(\alpha)) = G(\alpha) = 0 \tag{2.19}$$

As both u and F^{-1} re non-decreasing, with $xu(x) \geq 0, \forall x$, we know that one solution is

$$\alpha = F(\sup(w^{'} r_{p,t+1})) \tag{2.20}$$

Given the condition that $w^{'} r = \overline{r}_{p,t+1} \geq 0$ This condition is valid in that any unprofitable portfolios wont be chosen as they are strictly inferior to holding cash. The intuition of (2.20) is that investors are more likely to select optimal portfolios with lowest expected return in compensation of uncertainty in risk perception. Ambiguity in probability measure of asset return may help explaining underperformance targeting high return. Additionally, it implicitly penalizes model uncertainty as one type of investors misperception.

Combining Condition (2.18) and (2.20), it is able to obtain global optima by solving them simultaneously. This procedure can be shown in the following figure. Solutions are convolution of efficient area.

Interestingly, by evaluating risk of misperception in utility maximization, investors always hold extreme pessimism. And it seems likely that much more risks have to be assumed to gain higher return.

In reality, market friction is hardly negligible while most portfolio schemes are constructed without taking it into consideration. Thus to enhance practicability, we develop a tractable model that is applicable in high-friction environment. It can be easily extended to any market that is not immunized from trading cost.

The vector of discrete weights w_t at time t satisfies standardization constraint $w^{'}I = 1$ which implies that only $N - 1$ degrees of freedom for the space Δ^n spanned by w.It also indicates zero total increment cross-sectionally $\sum_{i=1}^{N}(w_t - w_{t-1}) = 0$

Turnover at time t accordingly is defined as the sum of absolute value of differentials:

$$T(t) = \sum_{i=1}^{N} |w_t - w_{t-1}| \tag{2.21}$$

This is the capital subject to friction. Denote $c_i(t)$ be cost rate of asset i at time t and total cost of complete rebalancing can be expressed as

$$C(t) = \sum_{i=1}^{N} c_i(t)|w_t - w_{t-1}| \tag{2.22}$$

With respect to definition (2.22), current literature discussed a revised version of portfolio optimization by maximizing utility of net portfolio return or maximizing net return given upper boundary of risk measure. Deviating from the integrated procedure, it might be useful to set up generally applicable back-end component. Basically, the method is to adapt weights derived from some policy to cost, aiming at maximizing net return.

It is advantageous in that original portfolio optimization is not tortured by cost penalty. It isolates adjustment to friction that is presumably uncorrelated with investors preference.

For simplicity but without losing too much generality, we restrict our discussion on a special case that a proportion $\beta \in R^1$ of capital is rebalanced and the rest is left untouched. Additionally, we allow a time-varying shrinkage $p = [p_1, ..., pN]' \in R^N$ for each rebalancing. Thus adjusted weights

$$\tilde{w}_t = w_t + (1 - \beta)w_{t-1} - p \tag{2.23}$$

which become path-dependent. Thus capital subject to friction is

$$C_\beta(t) = \sum_{i=1}^{N} c_i(t)|w_{i,t} - \beta w_{i,t-1} - p_i| \tag{2.24}$$

This quantile problem is linked to utility maximization, mean-variance and mean-variance using VaR and CVaR. Notice that no restrictions are placed on value of β and α. A negative proportion is valid when portfolio at previous period is leveraged up. $\beta > 1$ is implementable in a way that one essentially goes short the former policy.

Technical shrinkage p_i also has its economic explanation. It may be interpreted as cash draw periodically from previous base. The total amount reduced is $D(t) = p^{'}1$ in each period. At this point, our method has inherited some constant consumption plan (Merton, 1973) in linear case. We leave stochastic discount factor over multi-period utility maximization for further research.

To gain some asymptotic properties of $\tilde{w}_t$, some lemmas are introduced.

Lemma 2.3 *$\tilde{w}_i = \lim_{t\to\infty} E\tilde{w}_{i,t}$ exists if $Ew_{i,t} = \overline{w}_i$ and p_i are bounded.*

Computing some terminal weights at t using the rebalancing policy sequentially

$$\tilde{w}_{i,t} = \sum_{t=1}^{t}(1-\beta)^{t-s}(w_{i,s} - p_i) \tag{2.25}$$

Split $w_{i,s}$ by its expectation and deviation

$$\tilde{w}_{i,t} = \sum_{s=0}^{t}(1-\beta)^{(t-s)}\overline{w}_i + \sum_{s=0}^{t}(1-\beta)^{(t-s)}(\overline{w}_i - w_{i,t}) - \sum_{s=0}^{t}(1-\beta)^{(t-s)}p_i \tag{2.26}$$

Taking unconditional expectation and let T go to infinity, we have

$$\tilde{w}_i = \lim_{t\to\infty} E\tilde{w}_{i,t} = \frac{1}{\beta}(\overline{w}_i - p_i) \tag{2.27}$$

And arrive at Lemma (2.27).

It tells that long-term adjusted weights are proportional to difference between unadjusted weights and constant cash draw. If no increments is allowed asymptotically which is $\tilde{w}_i = \overline{w}_i$, $p_i = \frac{1-\beta}{\beta}\overline{w}_i$ It means that when positive capital is allocated in stock i, partially adjusted portfolio guarantees a positive cash draw.

We may relax individual relation by aggregating (3.23). If overall capitalization is fixed, $\sum_{i=1}^{N} p_i = \frac{1-\beta}{\beta}$. The intuition is that positive cash draw is present when $\beta \in (0, 1)$. It is reasonable because equity change is only associated with capital inflow and outflow. However this restriction is not placed in this framework or leverage is also a flexible dimension.

Up to now, it can be summarized as two basic scenarios we are facing. The first one is to stick to previous weights, whose net portfolio return at $t + 1$ is given as

$$r^0_{p,t+1} = \sum_{i=0}^{N} w_{i,t-1} r_{i,t+1} \tag{2.28}$$

The other is the stated partially rebalancing scheme

$$r^\beta_{p,t+1} = \sum_{i=0}^{N} \tilde{w}_{i,t} r_{i,t+1} \tag{2.29}$$

Thus our portfolio optimization in high friction environment can be interpreted as maximizing excess return over unbalanced one, which is

$$\max_{p,\beta} E\Delta r_{p,t+1} = \max_{p,\beta} E[r^\beta_{p,t+1} - r^0_{p,t+1}] \tag{2.30}$$

Notice that objective function (2.30) does not account for higher moments or utility profile. It is a simply discretion. It is able to further explore optimization with some loss function at the sacrifice of tractability. Reality check is also applicable for robustness.

Proposition 2.4 *Assuming fixed time-sequential return on each asset* $r_{i,t} = r_i$ *and constant transaction cost on each asset* $c_i(t) = c_i$*, solving (2.30) is a stacked quantile problem.*

Proof.

Lets expand (2.29) with (2.23) and (2.24).

$$r^{\beta}_{p,t+1} = \sum_{i=0}^{N} \tilde{w}_{i,t} r_{i,t+1} = \sum_{i=1}^{N} r_{i,t+1}(w_{i,t}+(1-\beta)w_{i,t-1}-p_i) - \sum_{i=1}^{N} c_i(t+1)|w_{i,t}-\beta w_{i,t-1}-p_i|$$

So objective function becomes

$$\max_{p,\beta} E\Delta r_{p,t+1} = E\sum_{i=1}^{N}(r_{i,t+1}\psi_{i,t} - c_i(t)|\psi_{i,t}|) \tag{2.31}$$

Here denote $\psi_{i,t} = w_{i,t} - \beta w_{i,t-1} - p_i$ Defining $\tau_i = -\frac{r_i}{2c_i} + \frac{1}{2}$, (2.30) is

$$\max_{p,\beta} E\Delta r_{p,t+1} = \max_{p,\beta} E\sum_{i=1}^{N} \rho_{\tau_i}(\psi_{i,t}) \tag{2.32}$$

rho_{τ_i} is quantile loss function. Therefore time-sequential estimation of p, β is simply solving a stacked quantile regression in the form of

$$\max_{p,\beta} \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N} \rho_{\tau_i}(\psi_{i,t}) \tag{2.33}$$

Here we allow parameter varies cross-sectionally. To ensure its solvability, $\tau_i \in (0,1)$ indicates that $|r_i| < c_i$. This inequality bounds applicability of our proposal in highly viscous market. Another situation takes place when magnitude of return decreases as frequency increase. For example, daily portfolio maintenance may suffer more compared to monthly counterpart since daily return is usually smaller.

Because process (2.23) is truncated by including one period lag, we may characterize this model as quantile AR(1).

2.3 Measuring Predictive/Profitable Ability

One difficulty in practicing parametric portfolio is to select factors that have forecasting accuracy. Specifically, if predictive ability exists in some set of factors $f \in \Sigma$ for some probability measure P, expectation of return conditioning on f should be biased. Here Σ is the mapping from information set of observations to real space which is a filtration in stochastic settings.

Definition 2.5 (Relative Predicative Ability) *RPA is*

$$\xi_f = E_P[r|f] - E_P r \tag{2.34}$$

r is a real-valued random variable and $E_P[r|f] - E_P r \neq 0$ indicates existence of predictability. One should also notice that definition (2.5) is naturally linked with a stronger expectation conditioning on one factor f_i which is given by

$$E_P[r|f_i, i = 1, ..., K] - E_P r \tag{2.35}$$

More generally, return can be partially biased by factors for some subspace $G \in \sigma(\Sigma)$

$$E_{P,G}[r|f] - E_P(r1_{\{f \in G\}}) = E_P[r|f \in G] - E_P(r1_{\{f \in G\}}) \tag{2.36}$$

If f is continuous in R^K and Σ is open, absolute predicative ability can be stated in the form of Cauchy limit:

Definition 2.6 (Relative Predicative Ability) *For any f that centers at the open set $B_r(f)$ of radius r,*

$$\lim_{r \to 0^+} E_P[r|B_r(f)] - E_P r \tag{2.37}$$

Definition 2.7 (Trading Strategy) *A trading strategy is a mapping $s \in S : R^K \rightarrow R$ of factors f and $s(f)$ is capital invested by following rule s.Specifically, a profitable strategy is then the one such that $E_P[rs(f)] - E[rs(f)]E_P r \neq 0$. S is the set of all possible strategies.*

Here we didnt assume that the difference is positive since a profitable strategy is always possible in the case of $E_P[rs^{'}(f)] > E[rs^{'}(f)]E_P r$ by letting $s^{'}(f) = -s(f)$ and $E_P[rs(f)] < E[rs(f)]E_P r$. Additionally, a trading strategy does not necessarily hold continuity. It is also possible to place restrictions on strategy configuration. For example, a long-only strategy is a mapping $s^+ : R^K \rightarrow R$. Similarly, market neutral strategy will guarantee $s^0 : R^K \rightarrow 0$. One special case is buy-and-hold strategy which is expressed as $s^b : R^K \rightarrow 1$. Validity of buy-and-hold strategy is therefore rejection of any excess return by designing rules on factors.

Another interesting aspect is the connection between trading strategy set and relative predicative ability, which is described as:

Proposition 2.8 (Equivalence between Predictability and Profitability) *Assuming probability measure P and conditional measure $P|\Sigma$ is differentiable, $E_P[r|f] - E_P r$ (Inequality A) holds for true if and only if there exists a trading strategy $s \in S$ which gives $E_P[rs(f)] - E[rs(f)]E_P r \neq 0$ (Inequality B).*

Remark (2.8) essentially finds a method that is able to exploit predictability. To see it, expand conditional pdf $p_{f|r}[r|f]$ we obtain:

$$s(f) = \frac{p_{f|r}[r|f]}{p_f(f)} \tag{2.38}$$

Where maximum likelihood applies.

In reality however it is more general to evaluate effectiveness via loss function.

Definition 2.9 (Generalized Strategy Profitability) *GSP is defined by*

$$E_P L(rs(f), E[s(f)]r) \tag{2.39}$$

With $\inf E_P L(rs(f), E[s(f)]r)$ *is reached when* $E_P[rs(f)] = E[rs(f)]E_P r$*. In order to restrict loss function behavior, we assume that*

$$\frac{\partial E_P L(rs(f), E[s(f)]r)}{\partial s(f)} \geq 0, \frac{\partial E_P L(rs(f), E[s(f)]r)}{\partial r} \leq 0 \tag{2.40}$$

Consider linear case $L(rs(f), E[s(f)]r)$. Higher moment may need concern, Let loss function $L(u) = \| u \|^p$. More specifically, when $p = 2$, linearization yields classical OLS. If loss operator is chosen as $L_\tau(u) = u(\tau - 1_{\{-\infty,0\}}(u))$ which essentially becomes quantile regression. Another linear scenario is separable loss function

$$L(rs(f), E[s(f)]r) = L(rs(f)) - L(E[s(f)]r) \tag{2.41}$$

If L is second-order differentiable, Taylor expansion of the right side in equation (2.41) offers a quadratic approximation and indicates that $L(rs(f), E[s(f)]r)$ does not need to be zero when $E_P[rs(f)] = E[rs(f)]E_P r$. An interesting aspect is that concavity of loss function determines direction of bias. In other words it is more conservative to concave loss functions in testing profitability. Based on discussion above, testing predictability of factors is equivalent to assessing profitability of a set of trading strategies instead of traditional regression. The obvious advantage is that it never presumes linearity in $E[r|f]$. As we will see, this framework can also be easily extended to analyzing parametric portfolio efficiency.

For researches that typically parameterize indicators to establish relations in the form of , White (2000) proposed a statistic of potential outperformance which is defined by:

$$f_{k,t+1} = ln[1 + y_{t+1}s_k(\chi_t; \beta_k)] - ln[1 + y_{t+1}s_0(\chi_t; \beta_0)], k = 1, ..., l \tag{2.42}$$

Where $f_{k,t+1}$ is the statistic k-th strategy at time $t+1$, y and s are asset return and trading signal respectively. As its defined,s depends on original price chi_t and parameter β. Ultimately, the following null hypothesis is expected to be tested:

$$H_0 : \max_{k=1,...,l} E(f_k) \leq 0 \tag{2.43}$$

As the maximum of statistic is chosen to compare with benchmark, it is not testing profitability of trading rule with specific parameter setting, but the effectiveness of its configuration. We average f_k along timeline to construct estimator of expectation.

In order to obtain distribution which is not analytically solvable, we use stationary bootstrap method (Politis and Romano, 1994) to randomly resample return series with total number of samples B=1000. The following statistics are then constructed:

$$\overline{V}_l = \max_{k=1,...,l} \sqrt{n}(\overline{f}_k) \tag{2.44}$$

$$\overline{V}_{l,i} = \max_{k=1,...,l} \sqrt{n}(\overline{f}^*_{k,i} - \overline{f}_k) \tag{2.45}$$

The order statistics of $\overline{V}_{l,i}$ offers p-value of $\overline{V}_l$.

The essence of stationary bootstrap is to introduce auxiliary binomial distribution in deciding whether next data point is chosen sequentially or randomly. The details of the method is also briefly discussed in Sullivan, Timmermann (1999) and White (2000)s paper.

2.4 Performance on FTSE100

In order to verify that technical analysis could also be applied to parametric portfolio policy. We consider forming optimal portfolio from all constituents of FTSE 100 during January 1985 and December 2008. We use technical indicators to substitute for fundamental factors in study of Brandt et al. (2009). Since the technical trading rules family is very huge and the best ones among them are not invariable with markets and time, first of all we use Whites reality check bootstrap method to recognize the best rules. Quantifying these rules and introducing them into parametric portfolio policy model, we then get our best portfolio policy (indicator coefficients and stock weights) to compare with the benchmark portfolio.

The stock list under consideration consists of all stocks that have been historical constituents of FTSE 100 index, a total of 283 companies. The rest 24 out of the original 307 constituents were deleted from our list for events such as rename, merger and acquisition, default and data unavailability. One serious problem with this processing is that the number of firms in my sample is various over time, rather than constant on 100. To handle this, add a term $\frac{1}{N_t}$ to the portfolio weight function. I trace dynamics of the index: companies data are exploited only when e they are on the FTSE 100 list, otherwise their data are removed from the sample. Weekly stock closing data are recorded from Data Stream to calculate trading rules as well as stock past returns.

The full sample data is divided into two periods: 1985-1994 and 1995-2008. Data of the first period is used to pin down optimal trading rules. We borrowed a complete set of trading rules from Sullivan et al. (1999) on which we conduct reality check. (Table 1)

Since stationary bootstrap offers a return distribution of the trading rules under weakly dependent settings, we are able to choose optimal parameters using preference a functional of the distribution. If quadratic utility is employed, first and second moment with properly defined risk aversion can rank performance with various parameters. Here for simplicity, the one with highest mean return is selected. The criterion then gives Table 2.

And sharpe ratio distribution of EMA, ROC and RSI is compiled in Figure 1

The corresponding technical indicators of the best rules are then calculated from original data to be characteristics applied to parametric portfolio model. The three characteristics are calculated as:

$$\Delta M_t(5,10|P_t) = M_t(5|P_t) - M_t(10|P_t)$$

$$R_t(2|P_t) = \log P_t - \log P_{t-2}$$

$$S_t(14|P_t) = 100 - \frac{100}{1 + RS_t(14|P_t)}$$

Figure 2.1: Parametric Portfolio Performance ON FT100

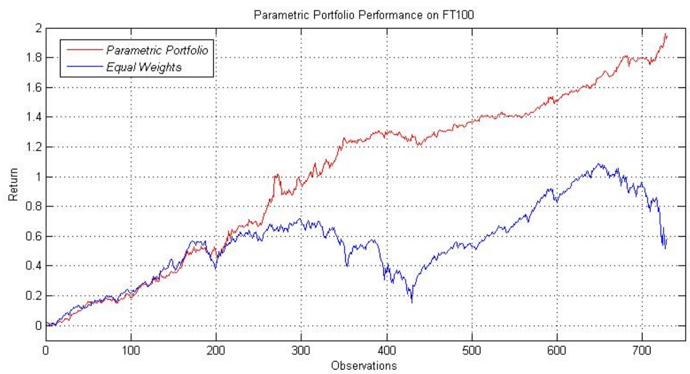


Table 2.1: White p-value of Five Technical Trading Rule Systems

This table presents Whites reality check for five technical trading systems considered: exponential moving average, rate of change, relative strength index, support and resistance and Alexanders filter. The Whites p-value under mean return and Sharpe Ratio criterion are both reported. EMA consists of 832 rules, ROC 12 rules, RSI 156 rules, S-R 24 rules, and ALF 120 rules.

		EMA	ROC	RSI	S_R	ALF
Whites p-value	Mean return	0.002	0.004	0.062	0.8	0.214
	Sharpe Ratio	0	0	0.02	0.578	0.124

Table 2.2: Optimal Technical Trading Rule under Sharpe Ratio Criterion

This table represents parameters of best trading rule for each trading system, based on Sharpe Ratio comparison. For EMA, parameters are time span of short-term moving average line, time span of long-term moving average line, percentage band around intersection point of the two lines, length of holding period. For ROC, parameter is the number of periods between current closing prices and past closing price. For RSI, parameters are considered periods and oversold threshold.

EMA			Parameter	
	n_1	n_2	n	h
	2	10	0.01	20
EMA			Parameter	
	n			
	2			
RSI			Parameter	
	n	LR		
	14	6		

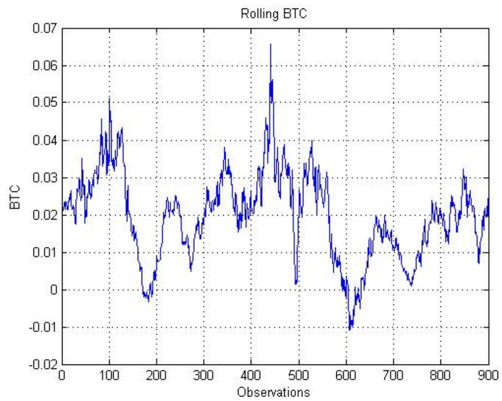
Table 2.3: Parametric Portfolio Performance Statistics

Statistics	Parametric Portfolio	Benchmark
Mean	0.142	0.049
Standard Deviation	0.101	0.135
Skewness	0.27	-0.26
Kurtosis	6.56	6.03
Sharpe Ratio	1.40	0.36
$VaR_{0.05}$	-0.0158	-0.027
$CVaR_{0.05}$	-0.0275	-0.044
CEG	0.121	0.012
Turnover	120.48	0
T_0	0.0215	–

Table 2.4: MOMENTS OF DEGENERATION DISTRIBUTION

Moments	Mean	SD	Skewness	Kurtosis
	-0.0118	0.0143	0.254	9.00

Figure 2.2: Rolling Breakeven Transation Cost

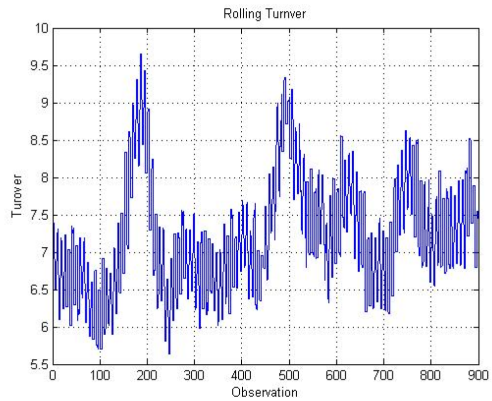


These characteristics data are standardized cross-sectionally to subject to standard normal distribution $N(0,1)$ before applying to parametric portfolio model. The standardization brings about several benefits. First, standardization unifies the scale of all characteristics so that the possibility of overestimating coefficient effect just because of its larger calibration. Second, the original cross-section characteristics data may be nonstationary, while the standardized ones are definitely stationary. Furthermore, standardization makes the total deviations from benchmark portfolio equal to zero and hence the sum of optimal portfolio weights to be one.

As for portfolio optimization, the first period of data is used for in-sample estimation for initial optimal weights while the second period for out-of-sample experiment.

Certainty equivalence gain (CEG) in Table (**??**) is calculated using CRRA function which indicates investors indifference between a risky gain and its CEG.

Figure 2.3: ROLLING TURNOVER



The formula adopted here is

$$CEG = [\frac{1}{T}\sum_{t=1}^{T}(1 + r_{p,t})^{1-\gamma}]^{\frac{1}{1-\gamma}}$$

Note also that weekly mean, standard deviation and CEG are annualized. Breakeven transaction cost (BTC) is a threshold rate beyond which active portfolio is no longer profitable. In discrete case, capital subject to market friction is adopting Scenario (3.24) with $\beta = 1, p = 0$. We also assume no risky asset holds in initial period, or $w_{i,0} = 0$. Then BTC should satisfy the following equation

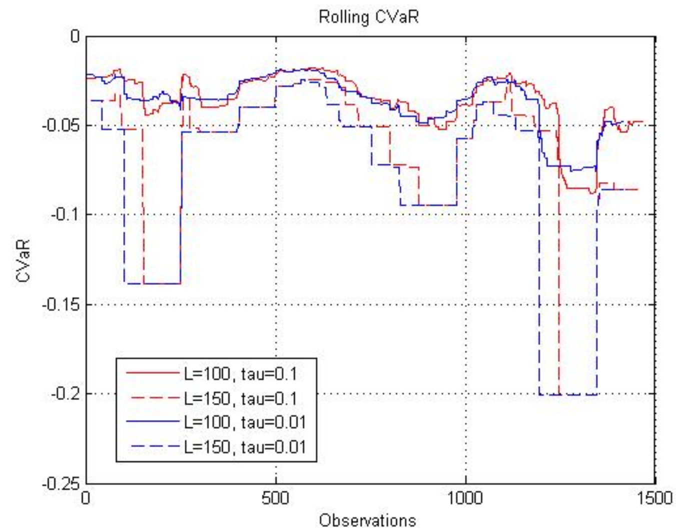
$$r_{net} = r_{gross} - T_c C \tag{2.46}$$

Where r_{net} and r_{gross} are net return and return in frictionless environment respectively. T_c is transaction rate. Let $r_{net} = 0$, T_0 is reversely defined as $T_0 = \frac{r_{gross}}{C}$

In order to capture time-varying profitability resistant to market friction, we may analyze breakeven transaction cost in rolling basis. The Figure 4 shows

Figure 2.4: ROLLING $CVaR_{0.05}$ (WINDOW LENGTH = 100)

Solid line: Parametric Portfolio Dash Line: Equal Weights



the dynamics of BTC with T=52.[1]

Long-term average breakeven transaction cost is approximately 2% which is consistent with the statistic on whole sample. But temporary small or even negative BTC is also present. A check on rolling turnover indicates that variation of BTC is largely due to instability of return.

It is also possible to calculate risk measure dynamically. Figure 4 compares rolling $CVaR_{0.05}$ of equal weights and parametric portfolio.

Overall parametric portfolio has a much better risk profile in most of the time. Discretion widens in financial crisis when pp profits from persistent down-trend. And tail risk is effectively managed to be anchored around 0.03 level.

Empirical return distributions are qq-plotted (Figure 5,6,7) to detect existence

[1]This is the average number of weeks in a year.

of non-nomarlity. Both parametric portfolio and equal weights significantly deviate from normal distribution. PP offers no help in alleviating ‘fat tail. Notice that parametric portfolio return is positively leptokurtic to equal weights scheme while avoiding introducing more negative extreme events. This property can achieve better risk/reward profile because more favorable outcomes are acquired without additional risks.

Finally, out-of-sample degeneration defined as the difference between realized return and target return has a negatively-skewed distribution. Moments are summarized in Table 4

As expected, degeneracy exists in parametric portfolio. Its high kurtosis is problematic since more extremes are anticipated. It is also evidence of difficulty in forecasting return.

2.5 Concluding Remarks

In this paper, we systematically discussed parameterization in portfolio optimization. The basic model is split policy into benchmark and add-on term that is factorized. It can be extended considering trading constraints, market friction, multi-utility and multi-frequency. Its validity is proved in stochastic settings. Then in order to mechanically select efficient predictors, we first construct a link between predictability and profitability so that reality check applies. Finally, performance on FTSE 100 is well documented to demonstrate robustness of the method. Its flexibility and rich potential advancement is highly appreciated.

Figure 2.5: QQPLOT: Parametric Portfolio vs Normality

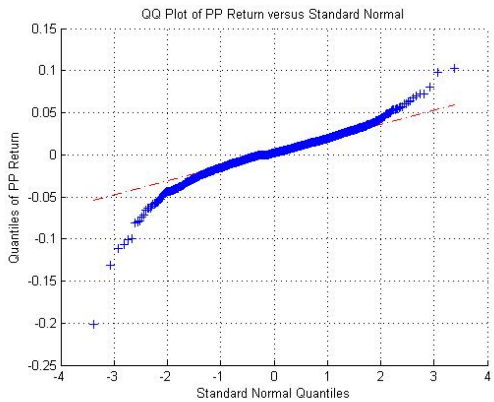


Figure 2.6: QQPLOT: Equal Weights vs Normality

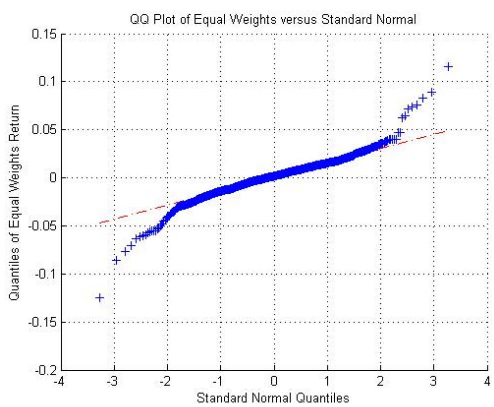


Figure 2.7: QQPLOT: Parametric Portfolio vs Equal Weights

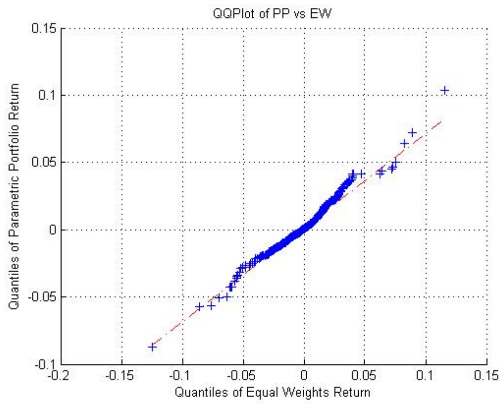
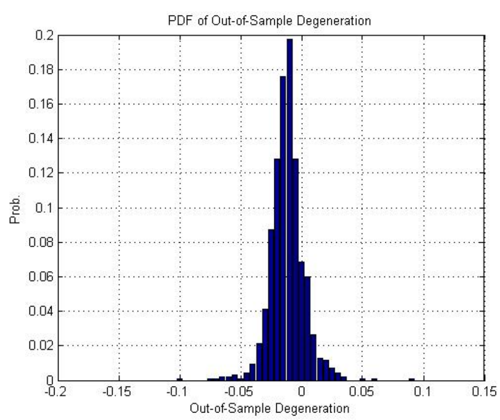


Figure 2.8: Out-of-Sample Degeneracy



2.6 Appendix

2.6.1 Technical Indicators

Diverse technical trading indicators have been developed to forecast stock price so far. Modern automated trading system combines multiple rules to generate accurate trading signals based on genetic algorithm or an artificial neural network. Here in our model, we only chose the rules with best predictive power for FTSE 100 constituents from huge universe of technical indicators. Below is a brief view of trading rules under consideration, prominently featured by Murphy (1999) and Kaufman (2005).

EMA (exponential moving average rules)

Exponential moving average is a trend following indicator screening fluctuations, thus identifying major trend of price efficiently. It entails variable weights to each price according to length of its history. In other words, it assigns heavier weights to data points the more recent they are and the weights decays exponentially, as is represented by formula (2.47)

$$M_t(n|P_t) = (1 - \frac{2}{n+1})M_{t-1}(n) + \frac{2}{n+1}P_t \tag{2.47}$$

Where M_t denotes EMA value at time t, P_t denotes stock prices, n represents the length of time periods considered. Common rule of EMA is to generate buy or sell signal by crossover of a fast (s) and slow (l) EMA line. Or

$$\Delta M_t(n_1, n_2|P_t) = M_t(n_1|P_t) - M_t(n_2|P_t) \tag{2.48}$$

In order to remove weak and false signals, filters should be added to the rule. Fixed band (b) that requires crossover big enough to exceed a minimum range is imposed. In addition, holding period for a given position (h), ignoring any signals triggered in subsequent periods, should also be considered in the rules. EMA rules parameters are designed as below:

n_1=1, 2, 3, 4, 5, 7, 10 (7 values)

n_2=5, 10, 15, 20, 25, 30, 35, 40 (8 values)

b=0.001, 0.005, 0.01, 0.05 (4 values)

h=5, 10, 15, 20 (4 values)

Note that s must be smaller than l.

ROC (rate of change rules)

ROC is the percentage difference between current closing price and the price several time periods ago. It measures the speed at which price is moving, hence generating trading signals (See equation (2.49))

$$R_t(n|P_t) = \log P_t - \log P_{t-n} \tag{2.49}$$

(2.49), P_t and P_{t-n} represent stock closing price now and n periods ago respectively. The midpoint of ROC is zero. When it crosses above (below) zero, a buy (sell) signal is sent out. Sometimes we may filter trading signal by some smoothers like Parameters of this rule:

$$\overline{R}_t(n, L|P_t) = M_t(L|R_t(n; P_t)) \tag{2.50}$$

n=1, 2, 3, 4, 5, 7, 10, 12 (8 values)

RSI (relative strength index rules)

The relative strength index is among famous momentum oscillators, whose value volatile around a midpoint line and function well in predicting price trend reversals. RSI is defined as the relative value between stock recent gains and losses. It is calculated as:

$$S_t(n|P_t) = 100 - \frac{100}{1 + RS_t(n|P_t)} \tag{2.51}$$

where RS denotes the ratio of average of n periods up closing prices to the average of n periods down closes. (Murphy, 1999)

Its value fluctuates between 0-100, with 50 as its midpoint. When RSI line crosses over the upper boundary $(100 - LR)$, there is high possibility that prices fall in the following periods. Conversely, if it breaks below the oversold threshold (LR), investors can expect a strong rise in price in the future. Parameters include

n=3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 (12 values)

LR=6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30 (13 values)

Support and Resistance Rules

This system produce buy or sell signals according to whether closing price exceeds the minimum or maximum level over the past n periods. As with the moving average rules, fixed band filters (b), and holding period requirements (h) can be imposed.

Parameters:

n=5, 10, 15, 20, 25, 50 (6 values)

h=5, 10, 15, 20 (4 values)

ALR (Alexanders Filter Rules)

Unlike rules above, ALR generate initiating and liquidating position signals. Investors should initiate a long (short) position when current closing price rises by x% above (below) its recent extreme low (high), defined as the lowest (highest) closing price obtained during a short (long) position trading period, and liquidate long (short) position held when todays closing price falls (rises) below recent extreme high (low).

Parameters are set as:

x=0.005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04, 0.045, 0.05 (10 values)

y=0.005, 0.01, 0.015, 0.02, 0.025, 0.03, 0.035, 0.04, 0.045, 0.05, 0.06, 0.08 (12 values)

All the tested technical trading rules from the five systems above add up to 1140.

3 On Investor's Pessmism

3.1 Background

As Von Neumann and Morgenstern (1944) and Salvage (1954) axiomatized expected utility under uncertainty, capital allocation among a set of risky assets becomes a classical decision making problem. Pioneered by Markowitz (1952, 1959), mean-variance framework has been dominant as a standard model in most textbooks. Just as controversies on expected utility, MV has never been accepted with irrefutable evidence. Critiques are mainly categorized into two disadvantages: paradoxical assumptions on preferences or return distribution (Quiggin (1981, 1993)), and accurate calibration of expected return and covariance matrix (Bawa et al. (1979); Michaud (1989)). Markowitz (2012) opposed the assertion that neither quadratic utility function nor Gaussian distribution is required. These conditions are sufficient rather than necessary. There are subsequently numerous efforts in literature on dealing with the second issue. Estimating expected mean has been constructively discussed by Jobson and Korkie (1980), Michaud (1989). Stability of correlations has attracted concerns as well (like Engle (2002); Bouchaud and Potters (2009); Laloux et al. (2000)). Black and Litterman (1992) introduced Bayesian method and updating belief in optimization procedure. Continuous-time mean-variance approach is developed recently. (Lindberg (2009))

Paralleling to expected utility, Schmeidler (1986); Quiggin (1993), Quiggin (1981) introduced Choquet utility function for incorporating subjectivity. Artzner et al.

(1999) lay axiomatic foundation for coherent risk measure. Several papers contribute to developing portfolio schemes using VaR and CVaR (Rockafellar and Uryasev (2000, 2002); Jaschke and Kuchler (2001)). The basic idea is to introduce an objective function with optimal parameters equal to the ones in VaR and CVaR. The stochastic programming problem can then be discretely solved. Another significant advancement is proposed by Bassett et al. (2004),Koenker (2005). Portfolios with a family of regular coherent risk measures are translated to classical quantile regression which is, however, essentially equivalent to Rockafeller approach in alpha-risk case. Despite of economic attractiveness of those procedures, the Achilles heel is inefficient estimation of tail risk. Since events are rarely observed in sample, risk of calibration might be severely high due to sampling variation (Jorison, 1996).

In regards with instability of inputs to optimizers, Brandt et al. (2009) design a parametric scheme that avoids predicting return and risk via maximizing utility function directly. It is especially superior with large asset base where estimation error is exponentially magnified. Following the spirit of parameterization, it is possible to factorize pessimistic portfolio weights to form parametric pessimistic portfolio. It inherits both the higher tolerance to data-snooping that most optimizers suffer and better flexibility which pessimistic portfolio delivers.

In sum, this paper is focused on discussing extensive framework of pessimistic investment strategies and empirically comparing the approaches with MV approach and benchmark using global equity index data. The left part is organized as: general models starting from utility maximization are developed in section 2; experiment details are described in data and performance; evaluation of efficiency is then discussed in quantile characteristics, degeneracy, robustness followed by a bootstrap that potentially alleviates overconcentration and conclusion.

3.2 Pessimistic Models

3.2.1 Quadratic Utility Function and Coherent Risk Measure

Consider a set of lotteries $X = [X_1, ..., X_n]^{'}$ and weight vector $w = [w_1, ...w_n]^{'}$. A portfolio consisting of them can be expressed as $X_p = w^{'}X$. Denote $F(X)$ probability distribution of the random vector X, expected utility of compounded lottery is

$$E_F u(X_p) = \int_{-\infty}^{\infty} u(X_p) dF(X) \tag{3.1}$$

Using Taylor expansion approximating utility around expectation (Markowitz 2012), which gives

$$u(X_p) \approx u(w^{'}EX) + \frac{\partial u(w^{'}EX)}{\partial X} w^{'}(X - EX) - \frac{1}{2}\frac{\partial^2 u(w^{'})}{(\partial X)^2} w^{'}Vw \tag{3.2}$$

We assume utility function is second-order differentiable. Expectation of lotteries and covariance matrix are defined by $EX = \int_{-\infty}^{\infty} X dF(X)$ and $V = (X - EX)(X - EX)^{'}$, respectively. Substitute utility with approximation (3.2), we attain

$$E_F u(X_p) = u(w^{'}EX) - \frac{1}{2}\frac{\partial^2 u(w^{'}EX)}{w^{'}\Sigma w} \tag{3.3}$$

Σ is the covariance matrix on probability space of $F(X)$. Equation (3.3) transforms maximization of expected utility into quadratic programming problem. Denote $\delta = \frac{1}{2}\frac{\partial^2 u(w^{'}EX)}{w^{'}\Sigma w}$ risk aversion coefficient, classical mean variance analysis becomes

$$\min_w \delta w^{'}\Sigma w$$

Given utility level, which is equivalent to a linear constraint $w^{'}EX = \mu_0$. Note that we did not pre-assume either probability distribution of random variables or quadratic utility function, which is emphasized to verify the validity of mean-variance framework. Therefore, Markowitz (1959, 2012) argued that Gaussian distribution and quadratic utility form are sufficient but not necessary conditions. As mentioned earlier, a set of distributions is always compatible with this approach. But we still need the assertion that expectation of return and covariance matrix can be reasonably calibrated.

The formulation above relies heavily on the accuracy of the quadratic approximation. Sometimes ineligible effect from higher order moment may significantly deteriorate applicability of MV approach.

Back to expected utility, let $X = F^{-1}(t)$, definition (3.1) becomes:

$$E_F u(F^{-1}(t)) = \int_{-\infty}^{\infty} u(F^{-1}(t))dt$$

It shows that expected utility is essentially a uniform integral on R. Alternatively, if distortion function is introduced to capture asymmetry of preference, we may obtain

$$E_F u(F^{-1}(t)) = \int_{-\infty}^{\infty} u(F^{-1}(t)) d\gamma(t) \tag{3.4}$$

This is expected Choquet utility. Specifically, conditional VaR $\phi_\alpha(X)$ or $\alpha-$risk measure under wide investigation, has a capacity of $\gamma(\alpha, X) = \min\{\frac{X}{\alpha}, 1\}$. Bassett et al. (2004) proved that a portfolio selection that minimizes the function of $\phi_\alpha(X) - \lambda EX$ is a quantile regression problem $\min_{\epsilon \in R} E\rho_\alpha(X - \epsilon)$ by the following theorem.

Theorem 3.1 (Bassett et al. (2004)) *If $EX < \infty$, then*

$$\min_{\epsilon \in R} E\rho_\alpha(X - \epsilon) = \alpha(\phi_\alpha(X) + EX) \tag{3.5}$$

Their work also shows that any pessimistic risk measure, by definition, is a Choquet integral of $\alpha-$risk measure for some probability measure:

$$\phi(X) = \int_{-\infty}^{\infty} \gamma(\alpha, x) d\varphi(\alpha)$$

The equivalence between coherence and pessimism makes quantile solution a tractable paradigm to all coherent risk measure. Despite of its universality, we will restrict our investigation within $\alpha-$risk measure in this paper.

3.2.2 Optimization of Conditional Value-at-Risk and Pessimistic Portfolio

Not surprisingly, optimization of conditional value-at-risk developed by Rockafellar and Uryasev (2000) is connected to pessimistic portfolio, since both adopt

α-risk in bi-criteria optimization problem. The former approach, however, first constructs loss function $\rho(z), z \in R^n$ associated with the outcome measured by loss that could be subsequently discretized and solved as a classical linear programming optimization. The latter analogously translate the problem in quantile regression which is essentially employing the same technique. Thus paralleling two methods is considered as an attempt to setting up a consistent CVaR optimization procedure.

Start from general expression of $\rho(z)$ with respect to α, η is (Pflug, 2000)

$$\rho_\alpha(z, \eta) = \eta + \frac{1}{1-\alpha} E[z-\eta]^+ \tag{3.6}$$

Where expectation takes the form of $E[z-\eta]^+ = \int_\eta^\infty (z-\eta)dF(z)$. To gain intuitive understanding, a piecewise loss function defined on real-valued random variables is given as:

$$L_\alpha = \left\{ \begin{array}{ll} z & z > \eta \\ \eta[1 - \frac{\alpha}{F(\eta)}] & z \leq \eta \end{array} \right.$$

Without much algebra, we prove the following proposition

Proposition 3.2

$$\rho_\alpha(z, \eta) = \frac{1}{1-\alpha} EL_\alpha(z, \eta)$$

One remarkable property of $\rho_\alpha(z, \eta)$ is the constant penalty rate imposed on loss lower than threshold level α, albeit linear relation is reserved in upper side. This asymmetry indicates pessimism consistent with coherent risk measure (Koenker, 2005) and leads to the difference from volatility. Also notice that $\rho_\alpha(z, \eta)$ may not be continuous at $X = \eta$. More accurately, outcome below some breakpoint always delivers no lower satisfaction. And continuity is achieved if and only if $\alpha = 0$. Another scenario worth noting is $\alpha = 1$, loss function $\rho(z)$ is no longer bounded. $\rho_\alpha(z, \eta)$ has abrupt shift but is still finite.

In Choquet utility setting, such proposition applies:

Proposition 3.3 *Consider distortion function*

$$\gamma_\alpha(X) = \max\{1, \frac{X}{\eta}\}$$

and linear utility

$$u(X) = \frac{1}{1-\alpha}(X + \frac{\eta(F(\eta) - \alpha)}{1 - F(\eta)})$$

$\rho(z) = \int u(X) d\gamma_\alpha(X)$ *is a Choquet utility. It is coherent and pessimistic. Specifically, if* $F(\eta)$ *is chosen to be* α, $\rho(z)$ *degrades to* $\alpha-$*risk.*

Theorem 3.4 (Rockafellar and Uraysev, 2000) *As function of* $\alpha, \rho_\alpha(z, \eta)$ *is convex and continuously differentiable.* $CVaR_\alpha$ *associated with any* $z \in R^n$ *is determined by*

$$\min_{\eta \in R} \rho_\alpha(z, \eta) \tag{3.7}$$

And VaR_α *is jointly derived by the value of* η *attaining the minimum*

$$VaR_\alpha = \arg\min_{\eta \in R} \rho_\alpha(z, \eta) \tag{3.8}$$

Proof. By expressing expectation with integral, equation (3.6) reads

$$\rho_\alpha(z, \eta) = \eta + \frac{1}{1-\alpha}\int_\eta^\infty zF(z) - \frac{\eta}{1-\alpha}[1 - F(\eta)]$$

Taking first-order condition with respect to η, as it is continuously differentiable

$$\frac{-\alpha + F(\eta)}{1-\alpha} = 0$$

Which gives $F(\eta^*) = \alpha$ or $\eta^* = F^{-1}(\alpha)$. We then plug the η^* into equation (3.6) to obtain

$$\begin{aligned}\rho_\alpha(z,\eta) &= F^{-1} + \frac{1}{1-\alpha}\int_{F^{-1}}^{\infty}[z - F^{-1}(\alpha)]^+ dF(z) \\ &= F^{-1} + \frac{1}{1-\alpha} zdF(z) - F^{-1} = \frac{1}{1-\alpha} zdF(z)\end{aligned}$$

Doing further expansion

$$\begin{aligned}\rho_\alpha(z,\eta) = \frac{1}{1-\alpha}\int_{F^{-1}}^{\infty} zdF(z) &= \frac{1}{1-\alpha}\int_{-\infty}^{\infty} zdF(z) - \frac{1}{1-\alpha}\int_{-\infty}^{F^{-1}} zdF(z) \\ &= \frac{1}{1-\alpha}E(z) - \frac{1}{1-\alpha}\int_{-\infty}^{F^{-1}} zdF(z)\end{aligned}$$

Comparing it with equation (3.4) and let $\gamma(\alpha, X) = \min\{\frac{X}{\alpha}, 1\}$ completes proof. One should notice that Theorem (3.4) does not explicitly assume bounded expectation of z, which is necessary for existence of minimum. The required condition has caused fragility in the case of tail distribution that first moment is not finite.

Most importantly, optimal loss function under Rockafellar?s setting combines an expectation and risk assessment which is essentially the right side of equation in Theorem (3.1). Thus we restate the last expansion by borrowing definition of $\phi_\alpha(z)$.

$$\rho_\alpha(z,\eta^*) = \frac{1}{1-\alpha}[E(z) + \phi_\alpha(z)]$$

Albeit formulating differently, risk measure and loss function share the same nature. If we apply $\rho_\alpha(z,\eta)$ in the context of asset allocation, optimizing CVaR is to obtain decision vector that minimize $\rho_\alpha(z,\eta^*)$

$$\min_{w,\eta} \rho_\alpha(X^{'}w, \eta) = \min_{w,\eta} \eta + E[X^{'}w - \eta]^+ \tag{3.9}$$

As $CVaR$ captures expected tail risk, objective function (3.6) incorporates first order moment. A more general version may consider a stochastic programing problem

Theorem 3.5 (Krokhmal,2006) *Let a mapping $\phi : \chi \to R$ satisfies monotonicity, sub-additivity and positive homogeneity defined by Arzner (1999), and $\phi(\eta) > \eta$ for all real η, optima of*

$$\rho(z) = \inf_{\eta} \eta + \phi(z - \eta)$$

Is coherent.

The contribution of Theorem (3.5) is bridging between a mapping with relaxed properties and coherent risk measure. Solving high moment coherent risk measure requires linearization with restrictions on number of observation which is far beyond our discussion. It is believed that its computational cost might prevent wide application.

Further exploring link between pessimistic portfolio and optimization method by Rockafellar and Uryasev (2000), we establish such proposition:

Proposition 3.6 *w, η that minimize loss function in the form of (3.9) is also solution to quantile regression problem*

$$\min_{w,\eta} E\rho_\alpha(X^{'}w - \eta) \tag{3.10}$$

Proof. Expand quantile loss function in continuous case

$$\begin{aligned} E\rho_\alpha(X^{'}w-\eta) &= \alpha\int_\eta^\infty (X^{'}w-\eta)dF(X) + (\alpha-1)\int_{-\infty}^\eta (X^{'}w-\eta)dF(X) \\ &= (\alpha-1)\int_{-\infty}^\infty (X^{'}w-\eta)dF(X) + \int_\eta^\infty (X^{'}w-\eta)dF(X) \\ &= (\alpha-1)EX^{'}w + (1-\alpha)\eta + \int_\eta^\infty (X^{'}w-\eta)dF(X) \end{aligned}$$

Since EX only contributes to optimal value, divided by scalar factor, w^*, η^* achieve minima simultaneously at

$$\min_{w,\eta} E\rho_\alpha(X^{'}w-\eta) \Leftrightarrow \min_{w,\eta} \eta + \frac{1}{1-\alpha}\int_\eta^\infty (X^{'}w-\eta)dF(X)$$

Which is exactly the loss function (3.10). In discrete case, quantile regression applies as follows.

$$E\rho_\alpha(X^{'}w-\eta) = (\alpha-1)\sum_{k=1}^{q} X^{'}w + \frac{1}{q}\sum_{k=1}^{q}[X^{'}w-\eta]^+ + (1-\alpha)\eta$$

in which q periods of observations replace integral except some notational changes. This is consistent with approximate solution proposed by Rockafellar and Uryasev (2000).

To be brief, under both continuous and discrete circumstances, pessimistic portfolio and optimization of CVaR by Rockafellar share the same risk measure and structure of solution. The former takes the root from Choquet utility maximization and can be easily extended to a wide range of regular coherent risk measures. Its elegant tractability, offering additional convenience in bootstrap analysis, is favored in this paper.

3.2.3 Parametric Pessimistic Portfolio

We may wish to decompose portfolio policy by a benchmark and deviation that seeks for excess return. This seperation is popular since ability of over-performance could be evaluated based on some statistics constructed by the two components. They include, to name a few, Sharpe ratio, information ratio and Treynor ratio. Generally, active deviation could be motivated by a set of characteristics, further linearization yields a parameterized expansion.

$$w(t) = w_b(t) + \frac{1}{N(t)} F^{'}(t)\beta \tag{3.11}$$

Where $w(t) = (w_1, ..., w_N)^{'}$ and $w_b(t) = (w_{1,b}, ..., w_{N,b})^{'}$ are portfolio and benchmark weights, respectively. Denote $f_i(t) = (f_{i,1}, ..., f_{i,K})^{'}$ K-dimensional factor/characteristics vector of the ith asset at time t, and $\psi_j(t) = (\psi_{1,j}, ..., \psi_{N,j})^{'}$ cross-sectional vector of jth factor with dimension N at time t. As a result $F(t) = (f_1, ..., f_N) = (\psi_1, ..., \psi_K)^{'}$ a $K \times N$ factor matrix and β their loadings to deviation. Here we allow time varying asset base. The convenience of flexibility is frequently exploited when constituents of benchmark change. The selected factors are firm-specific. In other words, common factors are not applicable in equation (3.11).

Multiplying both sides of equation (3.11) by $N \times 1$ unit vector, we obtain $\mathbf{1}^{'} w(t) = \mathbf{1}^{'} w_b(t) + \frac{1}{N(t)} \mathbf{1}^{'} F^{'}(t)\beta$. To faciliate comparison, it is reasonable to ensure that the portfolio is geared by the same leverage with benchmark, which gives $\mathbf{1}^{'} w(t) = \mathbf{1}^{'} w_b(t)$ irrespective of factors. Thus we should standardize $F(t)$

to have zero mean cross-sectionally so that $\mathbf{1}^{'}F^{'}(t) = 0$. Now if we want to minimize risk given target return μ. With asset return $X = (X_1, ..., X_N)^{'}$, it is, in the context of pessimistic portfolio

$$\min_{w} \phi_\alpha(X^{'}w) \qquad \text{subject to} \qquad X^{'}w = \mu \tag{3.12}$$

Replace w in optimization (3.12) with equation (3.11), objective function becomes

$$\min_{w} \phi_\alpha(X^{'}w) = \min_{\beta} \phi_\alpha(X_b + \frac{1}{N(t)} X^{'}F^{'}(t)\beta) \tag{3.13}$$

Where X_b is return on benchmark portfolio. The second term is weighted average of unit factor risk exposure. To see this, let $\xi_j = \langle \psi_j, X \rangle$ be inner product of ψ_j and r, which is the contribution per unit factor j to return. So $FX = \xi = (\xi_1, ..., \xi_K)^{'}$ in equation (3.13) indicates our assertion. It offers another interpretation of the optimization: investors assumes β_j amount of factor j that could offset benchmark downside risk and achieve better risk/reward profile. It also states that only those factors that systematically reduce benchmark risk should be chosen.

Unlike Brandt (2009), Newton's method can not be applied to the optimization problem (3.12), because loss function $\phi_\alpha(x)$ is not everywhere differentiable (Koenker, 2005). The following proposition is valid in the framework of pessimistic portfolio.

Proposition 3.7 *For bounded benchmark return $EX_b < \infty$, assume finite expectation $\bar{\xi} = E\tilde{\xi} < \infty$ of $\tilde{\xi} = \frac{1}{N}\xi = \frac{1}{N}FX$ which is the normalized projections with expectation. Let excess return $\gamma = \mu - EX_b$ Solving parametric pessimistic* *portfolio optimization problem*

$$\min_{\beta} \phi_\alpha(X_b + \tilde{\xi}' \beta) \qquad \textit{subject to} \qquad EX_b + \bar{\xi}' \beta = \mu \;\; \textit{and} \;\; \mathbf{1}' w(t) = 1 \quad (3.14)$$

is equivalent to constrained quantile regression problem

$$\beta^* = -\arg\min_{\beta,\nu} E\rho_\alpha(X_b - \tilde{\xi}' \beta - \nu) \qquad \textit{subject to} \qquad \bar{\xi}' \beta^* = -\gamma \quad (3.15)$$

Proof. In previous discussion, since we standardize factors cross-sectionally, constraint $\mathbf{1}' w(t) = 1$ is always satisfied if benchmark weights $\mathbf{1}' w_b(t) = 1$. By Theorem 2.1, we have

$$\alpha(EX_b + \bar{\xi}' \beta + \phi_\alpha(X_b + \tilde{\xi}' \beta)) = \min_{\nu} E\rho_\alpha(X_b + \tilde{\xi}' \beta - \nu) \quad (3.16)$$

Given target return $EX_b + \bar{\xi}' \beta = \mu$ or $\bar{\xi}' \beta^* = \gamma$, optimizing portfolio (3.14) is to minimize left side of equation (3.16). The right side formulates solution by letting $\beta^* = -\beta$.

Remark. We may restate problem (3.15) as a quantile factor model. If benchmark return can be interpreted as linear model of factors, it is possible to eliminate those effect on tail risk by opposite positioning. Alternatively, systemic risk is factorized and hedged. However, this scheme permits under-hedge or over-hedge as long as target return can be achieved. The trade-off spirit is essentially constructing efficient frontier on which each portfolio delivers best return/α-risk profile.

Practically, plug-in method applies after discretizing objective function (3.15)

$$\beta^* = -\arg\min_{\beta,\nu} \sum_{t=1}^{T} \rho_\alpha(X_{t,b} - \tilde{\xi}_t' \beta - \nu)$$

In more general settings, we may extend α-risk to all pessimistic risk measures that are able to be represented by α-risk.

Proposition 3.8 *Given the assumptions in proposition (3.7), minimizing any regular coherent risk measure given target return is a weighted quantile loss function.*

By definition, any risk measure that can be expressed as $\phi(r_p) = \int_0^1 \phi(r_p) d\varphi(\alpha)$ is pessimistic. Here $\varphi(\alpha)$ is a probability measure. Accordingly, objective function of parametric pessimistic portfolio becomes a weighted sum of loss functions.

$$\beta^* = -\min_{\beta} \int_0^1 \min_{\nu_\alpha} \frac{1}{\alpha} E\rho_\alpha(r_b - \tilde{\xi}^{'}\beta - \nu_\alpha) d\varphi(\alpha)$$

where ν is α-specific, but slopes do not change. Bassett et al. (2004) proves that pessimism is equivalent to regular coherence, which justifies our assertion. Consider $\varphi = \sum_{i=1}^{N} c_i I(\alpha_i)$ to be a mass function. it offers a more explicit form.

$$\beta^* = -\min_{\beta,\nu} \sum_{t=1}^{T} \sum_{i=1}^{N} \pi_i \rho_\alpha(r_{t,b} - \tilde{\xi}_t^{'}\beta - \nu_i)$$

Where weights $\pi_i = \frac{c_i}{\alpha_i}$.

A natural extension is to place some restrictions on weights. They are typically inequalities of equation (3.11) in the form of

$$Aw = A\bar{w}_b + \frac{1}{N} AF^{'}\beta \in T^*$$

where A is a $p \times N$ matrix with number of restrictions p and T^* is an real-valued set. For instance, long-only portfolio is constructed by letting $A = I_N$ a $N \times N$ identity matrix and $T^* = R_+^N$. The additional constraints on β hardly complicate our optimization since it is a standard inequality constrained quantile regression: see Koenker and Ng (2005) for interior solution procedure.

If cost is incorporated, a trade-off between risk reduction (or return enhancing) and transaction cost incurred by rebalancing should be considered. In simple linear scenario, denote c $1 \times N$ cost rate vector. Here the case that cost of one security might have effect on that of other assets is excluded, thus c can be viewed as an uncorrelated cost structure and retains less flexibility than the correlated one. It is allowed to have time-varying c. Our net return that meets required return as with α-risk becomes:

$$Er_b + \bar{\xi}'\beta - cE|\Delta w_b + \Delta F'\beta| = \mu$$

where $\Delta w_b = w_b(t) - w_b(t-1)$ is the difference of benchmark and $\Delta F = \frac{1}{N(t)}F(t) - \frac{1}{N(t-1)}F(t-1)$ the difference of deviation. As cost is risk-free, the objective function associated with coherent risk measure should remain unchanged. This problem is also

3.2.4 Efficient Frontier and Maximizing Reward/Risk Ratio

The model discussed does not infer a unique solution. As investors vary in risk tolerance, optimal risk reward pairs extend the efficient frontier. Just as all feasible portfolios are bounded by efficient frontier in mean-variance settings, we can characterize pessimistic models by risk-reward profile. The central problem is what risk measure should be adopted to match expected return. A natural choice, analogously, is the measure consistent with its min-max problem. Figure 1 shows how risk increases as higher return is targeted. As anticipated, we indeed demonstrate a concave frontier and a hump reward/risk ratio.

More generally, maximum drawdown (popular among practitioners), VaR expected shortfalls and traditional standard deviation are also candidates. Yet whatever measure is used, trade-off relationship between risk and return is commonly expected.

In practice, we have to fix portfolio selection for comparison. A common idea is to set the same target return μ_0. The problem associated with this simple procedure is that instable performance at different return level usually leads to contradictory conclusions. Another empirical difficulty is that we always have to set $\mu_0 < \sup\{EX_t\}$ in order to make the problem feasible.

Alternatively, we may consider optimization without constraints like:

$$\min_w \frac{\delta(s)}{s}$$

Here $s = \frac{\sigma(w,X)}{w'X}$. $\sigma(w,X)$ is risk measure concerned while $w'X$ is expected return accordingly.

It can be easily shown that for any $w \in W$, there is a w^f whose portfolio performance is on efficient frontier, such that $\frac{\delta(s^f)}{s^f} \leq \frac{\delta(s)}{s}$. Thus practically, searching on global minimum is restricted within boundary. Lai et al.(2011) shows that mean-variance optimization, under general setting that mean and variance are unobserved, is maximizing information ratio. The assertion also holds when we adopt different risk measures. The following proposition is a simplified justification of our procedure.

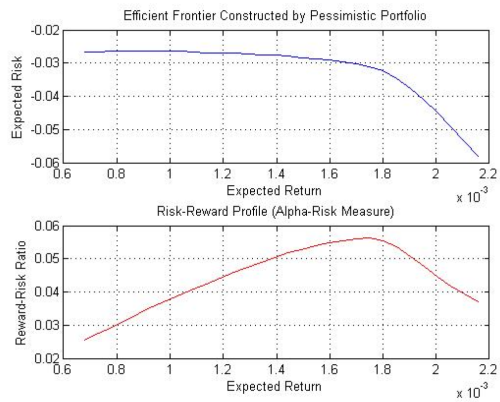


Figure 3.1: The upper panel shows concave efficient frontier of pessimistic portfolio, indicating a finite optima can be attained. Conditional value-at-Risk is chosen to be the risk measure. The Lower panel exhibits hump shape functional of risk/reward ratio on α-risk. We construct the figure using weekly indices data across 24 countries. Expected return is calculated historically without annualizing.

Proposition 3.9 *For any concave risk measure $\rho : w \to R$, concave reward measure $\gamma : w \to R$, convex decision space χ and bounded constant μ there exists unique vector $w^* \in W$ such that*

$$\frac{\gamma(w^*) - \mu}{\rho(w^*)} \geq \frac{\gamma(w) - \mu}{\rho(w)}$$

For all $w \in \chi, w \neq w^*$

3.3 Empirical Study on Global Indices Portfolio

3.3.1 Data

We choose major indices of 24 developed economies as representatives to their stock markets. All weekly price data denominated in local currency are collected from DataStream and logarithmic differenced. Table 1 compiles names of indices, geographical features and prevailing currencies. Total number of observations amounts to 624 ranging from 26 January 2001 to 28 December 2012. In the case that some countries have more than one index available, the one with highest trading volume is selected. The benefit of the criterion is two-fold. Liquidity is sufficient for practical implementation of our strategies. A wider popularity generally indicates enhanced market efficiency which might serve as a conservatism in our simulation.

Table 3 is summary of descriptive statistics. As observed in return on most financial instruments, all return distributions are negatively skewed and leptokurtic. VaR and CVaR vary little across countries, indicating similar historical tail risk. Only Italy experienced average loss in the sample period.

In order to construct parametric pessimistic portfolio, we choose interest rate, exchange rate and GDP as three sovereign fundamental inputs, among which the latter two indicators are log differenced to maintain stationarity. Since these factors are not weekly observed, we use the value most recently observed for each round of optimization.

One problem arises from international investment is currency risk hedging because foreign equity values have to be translated to aggregate performance. Some empirical evidence has shown convenience of currency hedging in global asset allocation. (Eun and Resnick, 1988; Perold and Schulman, 1988; Glen and Jorison 1993). In reality, constructing optimal portfolio cannot neglect the effect of exchange rate fluctuation. We can employ the techniques developed by

Black (1990), Adler and Prasad (1992), Walker (2008) to partially alleviate, if not fully eliminate, the risk. Yet our research should not deviate to detailing implementation issue and concentrate purely on return on speculative capital gains.

3.3.2 Simulation Results

In order to compare policy efficiency, we conduct simulation with two pessimistic models (pessimistic portfolio and parametric pessimistic portfolio), mean-variance approach. At each time t of rebalancing, a segment of sample S_t up to t with total number of L observations is collected. It trains optimizers to obtain portfolio weights $w(t)$. This allocation does not change until rebalancing is again called. For simplicity, the time interval Δt is fixed which is the holding period during which we stick to $w(t)$.

The procedure described may affect performance in two ways. By deliberately selecting rolling window length, return on certain policy can be boosted. Empirically, it is usually necessary to compromise between the risk of introducing additional noise and being less adaptive to fresh information which may significantly deteriorate efficiency. In our case, in hope of avoiding unintentional data-snooping, we offer two L for each portfolio. As one generally needs more data than dimension of assets in pessimistic portfolio and mean-variance approach, their L_1 is much larger than that L_2 of parametric pessimistic portfolio. Another aspect is that policies might behave differently associated with varying

Table 3.1: Details of Indices Data Set

24 indices are selected to construct international equity portfolio, among which 17 are European indices, 2 North American, 4 Asian and 2 Oceanian, respectively, to gain the geographical diversification. Each indice is a representative of one country equity market. Thus the most actively traded one is chosen.

Country	Main index	Region	Local Currency
Australia	S&P/ASX 200	Oceania	Australian Dollar
Austria	ATX 200	Europe	Euro
Belgium	BEL 20	Europe	Euro
Canada	S&P/TSX	North America	Canadian Dollar
Denmark	OMXC20	Europe	Danish Krone
France	CAC 40	Europe	Euro
Finland	OMXH	Europe	Euro
Germany	DAX 30 PERFORMANCE	Europe	Euro
Greece	ATHEX COMPOSITE	Europe	Euro
Hong Kong	HANG SENG	Asia	Hong Kong Dollar
Ireland	ISEQ	Europe	Euro
Israel	ISRAEL TA 100	Asia	New Shekel
Italy	FTSE MIB	Europe	Euro
Japan	NIKKEI 225	Asia	Yen
Netherlands	AEX	Europe	Euro
New Zealand	NZX 50	Oceania	New Zealand Dollar
Norway	OSLO	Europe	Norwegian Krone
Portugal	PSI-20	Europe	Euro
Singapore	STRAITS TIMES INDEX L	Asia	Singapore Dollar
Spain	IBEX 35	Europe	Euro
Sweden	OMXS30	Europe	Swedish Krona
Switzerland	SMI	Europe	Swiss franc
UK	FTSE 100	Europe	British Pound
USA	S&P 500	North America	US Dollar

holding period. To enhance robustness, both weekly and monthly rebalancing are simulated ($\Delta t = 1$ and $\Delta t = 4$ given weekly sampling frequency). Finally, we also vary significance τ in quantile loss functions, so that sensitivity to tail risk calibration can be assessed. Our benchmark is allocating equal capital to each equity index. Table 2 compiles annual return, Sharpe ratio calculated using annualized return and 2 risk measures both with significance level 0.01.

In all scenarios, active portfolio outperforms benchmark, which indicates existence of alphas. Parametric portfolio delivers the highest Sharpe ratio irrespective to $L, \tau, \Delta t$. Moreover, its exposure on tail risk is much smaller than other portfolios, an indication of effective risk control. Despite of mediocre annual return, it has the best reward/risk profile among all three portfolios. Mean-variance approach gives the highest return and is especially competitive alternative in low rebalancing frequency. Pessimistic portfolio overall records the lowest return and Sharpe ratio. Another feature worth noting is that performance on PPP seems to be insensitive to parameters. Comparatively, return of mean-variance increases with lower Δt. This could be caused by mismatching between in-sample (long-term perspective) and out-of-sample (short-term fluctuations) return/risk frames.

Table 3.2: Simulation Result From 2001 to 2012

The table shows out-of-sample performance from January 2001 to December 2012 using pessimistic portfolio (PP), mean-variance (MV) and parametric pessimistic portfolio (PPP). Each method applies to a rolling sample to obtain allocation for the next holding period. Simulation is conducted with two alternative window lengths L. Because one generally needs $L \gg N$, PP and MV have L of 150/100 which are much larger than that of PPP ($L = 50/25$). Loss significance τ is taken value of 0.1/0.01, to gain difference of pricing tail risk by investors. Portfolio is rebalanced for each additional observation if holding period is one week. Otherwise, we allow weights frozen for next one month. Weekly return is annualized to compute Sharpe ratio, assuming zeros risk-free rate. Tail risk measure, $\text{VaR}_{0.01}$ and $\text{CVaR}_{0.01}$, is calculated based on empirical distribution of portfolio return.

Rebalance weekly			Annual Return	Sharpe Ratio	$\text{VaR}_{0.01}$	$\text{CVaR}_{0.01}$
$L = 100$	PP	$\tau = 0.1$	0.0206	0.143	-0.0668	-0.0864
		$\tau = 0.01$	0.0147	0.091	-0.0756	-0.112
	MV	–	0.0455	0.240	-0.0730	-0.1235
$L = 50$	PPP	$\tau = 0.1$	0.0327	0.458	-0.0290	-0.0379
		$\tau = 0.01$	0.0282	0.396	-0.0290	-0.0379
$L = 150$	PP	$\tau = 0.1$	0.0232	0.191	-0.0635	-0.0950
		$\tau = 0.01$	0.0105	0.049	-0.0818	-0.1454
	MV	–	0.0496	0.245	-0.0812	-0.1373
$L = 25$	PPP	$\tau = 0.1$	0.0387	0.567	-0.0315	-0.0420
		$\tau = 0.01$	0.0362	0.537	-0.0315	-0.0394
Rebalance monthly			Annual Return	Sharpe Ratio	$\text{VaR}_{0.01}$	$\text{CVaR}_{0.01}$
$L = 100$	PP	$\tau = 0.1$	0.0184	0.091	-0.0677	-0.1014
		$\tau = 0.01$	0.0066	0.036	-0.0782	-0.1219
	MV	–	0.0672	0.345	-0.0720	-0.1177
$L = 50$	PPP	$\tau = 0.1$	0.0301	0.500	-0.0268	-0.0366
		$\tau = 0.01$	0.0289	0.473	-0.0268	-0.0376
$L = 150$	PP	$\tau = 0.1$	0.0127	0.061	-0.0681	-0.1050
		$\tau = 0.01$	0.0028	0.007	-0.0758	-0.1365
	MV	–	0.0615	0.301	-0.0823	-0.1351
$L = 25$	PPP	$\tau = 0.1$	0.0288	0.489	-0.0233	-0.0344
		$\tau = 0.01$	0.0277	0.463	-0.0246	-0.0354
Benchmark			-0.0049	-0.027	-0.0805	-0.114

3.4 Discussion

3.4.1 Tail Distribution Stationarity, Overconcentration and Bootstrap

The central assumption that justifies minimizing expected adverse tail events using historical data is stability of tail distribution. Although there is rich literature discussing the effect of leptokurtic return distributions on portfolio construction (as early as Mandelbrot, 1963), rarely has any study been conducted on consistency of tail behavior until recently (Straetmans and Candelon, 2013). To capture the dynamics of tail distribution, several researches, such as Embrechts (1997), Quintos (2001) contribute to designing statistics to test structural change of extreme returns. Jansen (2000), Danielsson and de Vries (1997) assess practical constraints placed on portfolio selection. To be concrete, we introduce rolling α-risk CVaR_α to demonstrate its time-varying characteristics.

In figure 2, CVaR_α can persistently deviate from its long-term mean, indicating ?random-walk? nature. Also notice that pessimistic portfolio hardly adapt to abrupt sizable movement occurring at 2008 financial crisis. It is therefore showing the fragile nature of extreme negative risk measure.

In multi-variant context, things get worse since concurrence of risk across assets plays the key role. Similar to covariance matrix in mean-variance approach, joint risk measure can be remarkably volatile and potentially endanger accuracy of forecast.

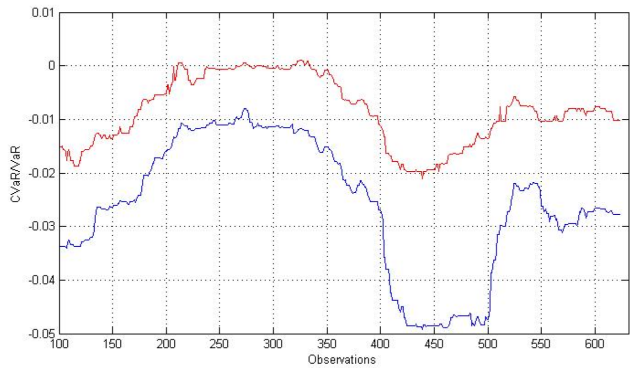


Figure 3.2: 100 weeks rolling CVaR_α on benchmark return. $\tau = 0.1$

Another problem that partially cohabited with non-stationary tail is caused by local risk-reward profile. Assets exhibiting superior temporary risk adjusted return have a strong gravitational effect on capital allocation in the framework of duality structure (see Section 2). This problem can be remarkably damaging to pessimistic portfolio. Figure 2 shows an example of such phenomena.

It plots the relation of weights with in-sample target return[2]. As minimum-acceptable return level climbs, diversification is diminishing. Extreme case takes place at the highest achievable target, where only Ireland index is selected. It ultimately becomes a momentum mechanism.

In figure 6, we offer dynamics of pessimistic portfolio weights with different target return level. It shows that with higher expected return investor requires, optimizer tilts to those assets historically performs better. This goes to extreme as the highest return attainable is selected. As Michaud (1989) criticized, overconcentration has inevitable disadvantages that could lead to poor our-of-

[2]Return is annualized.$L = 100, \tau = 0.1$

sample performance and over-exposure of risk that is presumed to be reduced through diversification. More risk is assumed due to absence of diversification. Local behavior dominates resource allocation, incurring additional transaction cost. Most importantly, it heavily relies on effectiveness of forecasting expected return using buy-winner strategy, which in most case has lost validity.

In combination with inefficient tail modeling and inappropriate portfolio optimization, resample technique can be employed to properly extract information from finite sample while retaining stochastic nature of return and risk measure. We call it bootstrap pessimistic portfolio.

In 1979, Efron first stated bootstrap algorithm, as a generalized resampling method, to empirically obtain confidence intervals and significance level in the case of violating normality assumption. Basic algorithm is to randomly select elements from the distribution of independently, identically distributed observations to approximate underlying distribution. Some early studies on its asymptotical properties and connections with other nonparametric approaches include Efron (1983), Efron and Tibshirani (1986), Bickel and Freedman (1981), and Singh (1981).

Some modified bootstrap methods were then proposed to resample from weakly dependent stationary observations, pioneered by Kunsch (1989) and Liu and Singh (1992) and later generalized by Politis and Romano (1992a, 1992b) and Politis, Romano, and Lai (1992). The central idea instead of select single data point is to conduct ?block resampling?. This philosophy was further improved by Politis and Romano (1994). It is essentially a weighted average of the block resampling distribution, but the main difference is that the pseudo-time series obtained by the stationary bootstrap is still stationary and the length of a block is geometrically distributed. By introducing smoothing parameter q, it is gaining relative stability compared with aforementioned moving block bootstrap. Recent applications include, not exhaustingly, Boles et al. (2005) and Ledoit and Wolf (2008).

Our bootstrap pessimistic portfolio essentially consists of two stages:

- Applying stationary bootstrap technique to randomly generate price paths.
- Optimize portfolio on the expanded data set.

The first advantage of this procedure is that diversification benefit is restored. Consider previous example, bootstrap portfolio no longer overweigh single asset[3]. (Figure 7) And weights vary little with different smoothing parameters. Its efficiency can be further examined by out-of-sample performance. Let 6 years up to December 2006 be training set and subsequent 6 years testing period, performance with or without bootstrap is able to be compared. Significance level was chosen $\tau = 0.1$.

3.4.2 Quantile Analysis

As it is well known, normal distribution maximizes differential entropy, dictates that the most probable outcome should follow the distribution whose entropy is

[3] For simplicity, we average weights on each sample.

a local maximum. Thus deviation from normality is an indication of information loss in portfolio construction, the extent of which might serve as a qualitative measure of information efficiency.

If two return distribution share the same moment generating function, they should deliver the same reward/risk ratio. Smirnov (1948) stated that the supreme of the difference between two distributions should asymptotically converges to the supreme of the absolute value of a Brownian bridge. Thus a quantile analysis would help identifying distributional consistency. Here, we use qq plot to illustrate the characteristics. Basically, leptokurtic return bias left tail downward and right tail upward.

In Figure 2, we construct qq plot for three portfolios and benchmark. As observed, all active schemes fail to control fat tail at both sides. Mean-variance deviate from the central Gaussian area $[\mu - \epsilon_L, \mu + \epsilon_U]$ which the others share similar local feature. To further illustrate relations between two portfolio returns, we plot mean-variance and pessimistic portfolio quantiles over parametric pessimistic portfolio quantiles. It shows that return on PPP is less leptokurtic, an evidence of enhancing tail risk control.

3.4.3 Out-of-sample Degeneration

As Davies and Servigny (2012) states, the discretion between realized return and target return usually exhibits a negatively distorted distribution. The extent of degeneracy, when extending in-sample allocation scheme beyond, is a critical part of evaluation. It is also worthy of investigation from the perspective of historical information extraction.

The figure 5 and table 5 show estimation error, which is defined by the difference between realized return and target return:

$$e_t = W_{t-1}(X_t - EX_{t-1})$$

Here denote W_{t-1}, X_t, EX_{t-1} portfolio weights at $t-1$, asset return at t and expected asset return at $t-1$ respectively. Ideally,$Ee_t = 0$ indicates maximal efficiency that active portfolio can reach. As long as EX_{t-1} is an unbiased estimator of EX_t, e_t should be symmetric. Thus negative mean indicates inefficient modeling of return. More severe degeneracy is present in pessimistic portfolio as both magnitude and uncertainty are higher. One explanation is that tail distribution can be relatively mobile and instable over time and thus causes unfavorable misallocation. The phenomenon is well documented by several researchers.

Another aspect we observe is that pessimistic portfolio has remarkable higher-order advantage in higher order moments (skewness and kurtosis). This is in connection with risk measure: pessimistic portfolio minimizes tail risk while mean-variance approach neglects higher order effects in approximation.

3.4.4 Profitability and Resistence to Market Friction

In reality, market friction is hardly negligible but portfolio schemes discussed previouly are constructed without taking it into consideration. The simplest measure of resistance to market friction is breakeven transaction cost c_b, which is defined as the average return over rebalancing turnover. Table 5 compiles c_b in all simulation scenarios. However under high friction environment, investors behavior might be distorted to maximize net return. As a result, breakeven transaction is not an accurate description.

To enhance practicability, we develop a tractable model that is applicable in high-friction environment. It can be easily extended to any market that is not immunized from trading cost.

The vector of discrete weights w_t at time t satisfies standardization constraint $w^{'}I = 1$ which implies that only $N - 1$ degrees of freedom for the space Δ^n spanned by w.It also indicates zero total increment cross-sectionally $\sum_{i=1}^{N}(w_t - w_{t-1}) = 0$

Turnover at time t accordingly is defined as the sum of absolute value of differentials:

$$T(t) = \sum_{i=1}^{N} |w_t - w_{t-1}| \tag{3.17}$$

This is the capital subject to friction. Denote $c_i(t)$ cost rate of asset i at time t and total cost of complete rebalancing can be expressed as

$$C(t) = \sum_{i=1}^{N} c_i(t)|w_t - w_{t-1}| \tag{3.18}$$

With respect to Definition (3.18), current literature discussed a revised version of portfolio optimization by maximizing utility of net portfolio return or maximizing net return given upper boundary of risk measure. Deviating from the integrated procedure, it might be useful to set up generally applicable ?back-end component?. Basically, the method is to adapt weights derived from some policy to cost, aiming at maximizing net return.

It is advantageous in that original portfolio optimization is not tortured by cost penalty. It isolates adjustment to friction that is presumably uncorrelated with investor?s preference.

For simplicity but without losing too much generality, we restrict our discussion on a special case that a proportion $\beta \in R^1$ of capital is rebalanced and the rest is left untouched. Additionally, we allow a time-varying shrinkage $p = [p_1, ..., pN]^{'} \in R^N$ for each rebalancing. Thus adjusted weights

$$\tilde{w}_t = w_t + (1 - \beta)w_{t-1} - p \tag{3.19}$$

which become path-dependent. Thus capital subject to friction is

$$C_\beta(t) = \sum_{i=1}^{N} c_i(t)|w_{i,t} - \beta w_{i,t-1} - p_i| \tag{3.20}$$

This quantile problem is linked to utility maximization, mean-variance and mean-variance using VaR and CVaR. Notice that no restrictions are placed on value of β and α. A negative proportion is valid when portfolio at previous period is leveraged up. $\beta > 1$ is implementable in a way that one essentially goes short the former policy.

Technical shrinkage p_i also has its economic explanation. It may be interpreted as cash draw periodically from previous base. The total amount reduced is $D(t) = p^{'}1$ in each period. At this point, our method has inherited some constant consumption plan (Merton, 1973) in linear case. We leave stochastic discount factor over multi-period utility maximization for further research.

To gain some asymptotic properties of $\tilde{w}_t$, some lemmas are introduced.

Lemma 3.10 *$\tilde{w}_i = \lim_{t\to\infty} E\tilde{w}_{i,t}$ exists if $Ew_{i,t} = \overline{w}_i$ and p_i are bounded.*

Computing some terminal weights at t using the rebalancing policy sequentially

$$\tilde{w}_{i,t} = \sum_{t=1}^{t} (1-\beta)^{t-s}(w_{i,s} - p_i) \tag{3.21}$$

Split $w_{i,s}$ by its expectation and deviation

$$\tilde{w}_{i,t} = \sum_{s=0}^{t} (1-\beta)^{(t-s)}\overline{w}_i + \sum_{s=0}^{t} (1-\beta)^{(t-s)}(\overline{w}_i - w_{i,t}) - \sum_{s=0}^{t} (1-\beta)^{(t-s)} p_i \tag{3.22}$$

Taking unconditional expectation and let T go to infinity, we have

$$\tilde{w}_i = \lim_{t\to\infty} E\tilde{w}_{i,t} = \frac{1}{\beta}(\overline{w}_i - p_i) \tag{3.23}$$

And arrive at Lemma (3.23).

It tells that long-term adjusted weights are proportional to difference between unadjusted weights and constant cash draw. If no increments is allowed asymptotically which is $\tilde{w}_i = \overline{w}_i$, $p_i = \frac{1-\beta}{\beta}\overline{w}_i$ It means that when positive capital is allocated in stock i, partially adjusted portfolio guarantees a positive cash draw. We may relax individual relation by aggregating (3.23). If overall capitalization is fixed, $\sum_{i=1}^{N} p_i = \frac{1-\beta}{\beta}$. The intuition is that positive cash draw is present when $\beta \in (0, 1)$. It is reasonable because equity change is only associated with capital inflow and outflow. However this restriction is not placed in this framework or leverage is also a flexible dimension.

Up to now, it can be summarized as two basic scenarios we are facing. The first one is to stick to previous weights, whose net portfolio return at $t + 1$ is given as

$$r^0_{p,t+1} = \sum_{i=0}^{N} w_{i,t-1} r_{i,t+1} \tag{3.24}$$

The other is the stated partially rebalancing scheme

$$r^{\beta}_{p,t+1} = \sum_{i=0}^{N} \tilde{w}_{i,t} r_{i,t+1} \tag{3.25}$$

Thus our portfolio optimization in high friction environment can be interpreted as maximizing excess return over unbalanced one, which is

$$\max_{p,\beta} E\Delta r_{p,t+1} = \max_{p,\beta} E[r^{\beta}_{p,t+1} - r^{0}_{p,t+1}] \tag{3.26}$$

Notice that objective function (3.26) does not account for higher moments or utility profile. It is a simple discretion. It is able to further explore optimization with some loss function at the sacrifice of tractability. Reality check is also applicable for robustness.

Proposition 3.11 *Assuming fixed time-sequential return on each asset $r_{i,t} = r_i$ and constant transaction cost on each asset $c_i(t) = c_i$, (3.26) is a stacked quantile problem.*

Proof. Let us expand (3.25) with (3.19) and (3.20).

$$r^{\beta}_{p,t+1} = \sum_{i=0}^{N} \tilde{w}_{i,t} r_{i,t+1} = \sum_{i=1}^{N} r_{i,t+1}(w_{i,t} + (1-\beta) w_{i,t-1} - p_i) - \sum_{i=1}^{N} c_i(t+1) |w_{i,t} - \beta w_{i,t-1} - p_i|$$

So objective function becomes

$$\max_{p,\beta} E\Delta r_{p,t+1} = E \sum_{i=1}^{N} (r_{i,t+1} \psi_{i,t} - c_i(t) |\psi_{i,t}|) \tag{3.27}$$

Here denote $\psi_{i,t} = w_{i,t} - \beta w_{i,t-1} - p_i$. Defining $\tau_i = -\frac{r_i}{2c_i} + \frac{1}{2}$, equation (3.26) is

$$\max_{p,\beta} E\Delta r_{p,t+1} = \max_{p,\beta} E \sum_{i=1}^{N} \rho_{\tau_i}(\psi_{i,t}) \tag{3.28}$$

rho_{τ_i} is quantile loss function. Therefore time-sequential estimation of p, β is simply solving a stacked quantile regression in the form of

$$\max_{p,\beta} \frac{1}{T} \sum_{t=1}^{T} \sum_{i=1}^{N} \rho_{\tau_i}(\psi_{i,t}) \tag{3.29}$$

Here we allow parameter τ varying cross-sectionally. To ensure its solvability, $\tau_i \in (0, 1)$ indicates that $|r_i| < c_i$. This inequality bounds applicability of our proposal in highly viscous market. Another situation takes place when magnitude of return decreases as frequency increase. For example, daily portfolio maintenance may suffer more compared to monthly counterpart since daily return is usually smaller.

Because process (3.19) is truncated by including one period lag, we may characterize this model as quantile AR(1). In the situation that condition $|r_i| < c_i$ is violated, it is possible to split equity base with one part suffering higher friction and the other is free of transaction cost. In the situation that condition $|r_i| < c_i$ is violated, it is possible to split equity base with one part suffering higher friction and the other is free of transaction cost. We plot the dynamics of each portfolio scheme with varying c in figure 8. c_b is higher than the non-adaptive case. Parametric portfolio is much more resistant to friction while mean-variance approach loses advantage when $c > 0.0003$.

3.5 Concluding Remarks

In hope of gaining broader acceptance on conservatism on extreme cases, this paper might be interpreted as an attempt to generalizing pessimism and empirically testing the efficiency among those models. Difficulty in accurate calibration of risk measures impedes profitable application of pessimistic models. In the case that tail risk is more volatile than volatility measure, mean-variance optimization could be more preferred. This phenomenon is documented by our empirical study in global indices investment.

To avoiding modeling tail distribution of return, we attempt to decompose portfolio weights and propose a framework of parameterized pessimistic models. This factorized approach is advantageous in high dimensional asset pool which most optimizers suffer from. It retains more flexibility in additional constraints on weights and is less affected by in-sample data-snooping. Overall, the approach delivers highest risk-adjusted return that is robust to market friction.

As Liu and Cao (2014) suggest, profitability of parametric pessimistic portfolio is sourced from serial dependence of factors and their explanatory ability on asset return. Thus a mechanism of assessing factor selection needs further investigation.

3.6 Appendix

3.6.1 Computational aspects of pessimistic models

For pessimistic portfolio with α-risk measure, analogous quantile problem is given as:

$$\min_{(\beta,\xi)\in R^p} \sum_{i=1}^{n} \rho_\alpha(x_{i1} - \sum_{i=2}^{p}(x_{i1} - x_{ij}) - \xi) \qquad \text{s.t.} \bar{x}^{'}\pi = \mu_0 \tag{3.30}$$

We packaged portfolio coefficients and location parameter in one set under estimation, which is $\beta = (\beta_2, ..., \beta_p, \xi)^{'}$. Slack variables u, v convert inequality into equality constraints.

In order to rearrange it into classical linear programming structure:

$$\min_x \{f^{'}y | Ay = b, y \in [\gamma_l, \gamma_u]\} \tag{3.31}$$

γ_l, γ_u are lower and upper boundary respectively. And corresponding matrixes are

$$y = (u^{'}, v^{'}, \beta^{'})^{'} \tag{3.32}$$

$$f = (\tau e^{'}, (1-\tau)e^{'}, 0)^{'} \tag{3.33}$$

$$b = (x_1^{'}, \mu_0 - \bar{x}_1)^{'} \tag{3.34}$$

Where x_1 is the observation vector $x_1 = (x_{11}, ..., x_{n1})^{'}$. Similarly, observation matrix is $X = (x_2, ..., x_p)$. A is an extended matrix, such that target return constraint is incorporated:

$$A = \begin{pmatrix} I & -I & X \\ O_{1\times n} & O_{1\times n} & \Delta\bar{x}_j \end{pmatrix} \tag{3.35}$$

Here $\Delta\bar{x}_j$ is the difference between mean of returns $\Delta\bar{x}_j = (\bar{x}_2 - \bar{x}_1, ..., \bar{x}_p - \bar{x}_1)$ Traditionally, the problem is solved by simplex method. It searches optimal vertex of the polyhedral expanded by constraints, from an initial feasible point. Usually complexity is increasing as it has more dimensions. In worst cast, Klee-Minty shows that time consumption goes up exponentially.

Interior point method was considered as the most computationally efficient algorithm for linear programming. Basically, nonlinear terms are added to objective function. The continuous Lagrangian expression is then differentiated to obtain first order conditions (known as Karush-Kuhn-Tucker KKT conditions). Affine scaling step is calculated recursively by Newton method.

For unconstrained quantile regression, Koenker and Portnoy presented primal and dual functions as

$$\min_{u,v,b}\{\tau e^{'}u + (1-\tau)e^{'}v | Xb + u - v = y, (u^{'}, v^{'}, b^{'}) \in R^{2n} \times R^{p}\} \tag{3.36}$$

$$\max_{d}\{y^{'}d | X^{'}d = (1-\tau)X^{'}e, d \in [0,1]^{n}\} \tag{3.37}$$

Then Frisch-Newton algorithm can be employed to search for optimal point following the central path toward boundary.

For quantile regression with linear constraints, primal and dual problems are similar:

$$\min_{u,v,b}\{\tau e^{'}u + (1-\tau)e^{'}v | Xb + u - v = y, Rb \geq r, (u^{'}, v^{'}, b^{'}) \in R^{2n} \times R^{p}\} \tag{3.38}$$

$$\max_{k_1,k_2}\{y^{'}k_1 + r^{'}k_2 | X^{'}k_1 + R^{'}k_2 = (1-\tau)X^{'}e, k_1 \in [0,1]^{n}, k_2 \geq 0\} \tag{3.39}$$

This log-barrier function is formulated to penalize feasible solutions as they approach the boundary of the constraints set:

$$L = c^{'}k - y^{'}(X^{'}k_1 + R^{'}k_2 - b) - w^{'}(u - k_1 - s) - u(\sum \log k_1 + \sum \log k_2 + \sum \log s) \tag{3.40}$$

c is the temporary variable to be estimated, while $k = (k_1^{'}, k_2^{'})^{'}$

First order condition provides classical KKT system, which the affine scaling step is set by Newton?s Method:

$$dy = (AQ^{-1}A^{'})^{-1}(\tilde{r}_3 + A_1 Q_1^{-1}\tilde{r}_1 + A_2 Q_2^{-1}\tilde{r}_2) \tag{3.41}$$

$$dk_1 = Q_1^{-1}(A_1^{'}dy - \tilde{r}_1) \tag{3.42}$$

$$dk_2 = Q_2^{-1}(A_2^{'}dy - \tilde{r}_2) \tag{3.43}$$

$$ds = -dk_1 \tag{3.44}$$

$$dz = -z - K^{-1}zdk \tag{3.45}$$

$$dw = -w - S^{-1}Wdk \tag{3.46}$$

$$Q_1 = K_1^{-1}Z_1 + S^{-1}W \tag{3.47}$$

$$Q_2 = K_2^{-1}Z_2 \tag{3.48}$$

Here all notations follow linear programming conventions, with main diagonal of each matrix the elements of corresponding vectors. As a result, reverse is no longer costing, and only effort is needed for Cholesky decomposition for the first equation in iterative algorithm.

This tentative affine scaling step should be further modified, to properly set moving speed toward boundary. The scaling factor is determined by feasibility checking. If the full affine scaling step is infeasible, Mehrotro predictor-corrector step should be brought into, which is actually taking ignored bilinear terms into account. It is essentially anther Newton step starting from previously proposed point. Formulas are quite similar in this setting:

$$\delta y = (AQ^{-1}A^{'})^{-1}(A_1 Q_1^{-1}\hat{r}_1 + A_2 Q_2^{-1}\hat{r}_2) \tag{3.49}$$

$$\delta k_1 = Q_1^{-1}(A_1^{'}\delta y - \hat{r}_1) \tag{3.50}$$

$$\delta k_2 = Q_2^{-1}(A_2^{'}\delta y - \hat{r}_2) \tag{3.51}$$

$$\delta s = -\delta k_1 \tag{3.52}$$

$$\delta z = -z - K^{-1} z \delta k \tag{3.53}$$

$$\delta w = -w - S^{-1} W \delta k \tag{3.54}$$

$$\hat{r}_1 = S_{-1}(\mu e - dSdWe) - K_1^{-1}(\mu e - dK_1 dZ_1 e) \tag{3.55}$$

$$\hat{r}_2 = -K_2^{-1}(\mu e - dK_2 dZ_2 e) \tag{3.56}$$

Finally, the duality gap is calculated iteratively until it reduces below tolerance level. The convergence is achieved through trade-off between stickiness to central path and gap reducing speed (Gonzago).

Practically, it is difficult for balance, since the process is path-dependent and fairly sensitive to initial point. Koenker and Pin states that $d_0 = (1-\tau)e$ is a natural initial feasible point for inequality constrained quantile regression. However the variable z is also flexible and has a significant impact on convergence rate. $z_2 \neq 0$ is an additional constraint, yet it can be shown that any value assigned to z_2 would significantly deteriorates efficiency of interior point method.

Therefore, more robust initial point is essential in the algorithm.

3.7 Tables and Figures

Table 3.3: Descriptive Statistics of Country Indices

Country	Mean	SD	Skewness	Kurtosis	$VaR_{0.01}$	$CVaR_{0.01}$	$VaR_{0.05}$	$CVaR_{0.05}$
Australia	0.00093	0.0204	-0.8476	8.815	-0.0604	-0.0807	-0.0339	-0.0500
Austria	0.00092	0.0327	-1.1861	14.540	-0.0988	-0.1425	-0.0491	-0.0818
Belgium	0.00046	0.0268	-1.1633	12.653	-0.0814	-0.1130	-0.0451	-0.0670
Canada	0.00117	0.0222	-0.9475	10.932	-0.0670	-0.0998	-0.0339	-0.0553
Denmark	0.00128	0.0267	-0.9791	10.214	-0.0712	-0.1152	-0.0407	-0.0641
France	0.00066	0.0299	-0.6744	7.747	-0.0860	-0.1143	-0.0475	-0.0690
Finland	0.00127	0.0372	-0.5399	6.263	-0.1087	-0.1446	-0.0578	-0.0912
Germany	0.00167	0.0303	-0.6586	7.743	-0.0838	-0.1171	-0.0486	-0.0719
Greece	0.00086	0.0421	0.0983	6.370	-0.1256	-0.1526	-0.0656	-0.0958
Hong Kong	0.00214	0.0366	-1.2987	14.752	-0.0975	-0.1575	-0.0541	-0.0857
Ireland	0.00150	0.0295	-1.5688	15.735	-0.0928	-0.1406	-0.0447	-0.0739
Israel	0.00243	0.0360	-0.6029	5.678	-0.1205	-0.1392	-0.0617	-0.0873
Italy	-0.0005	0.0348	-0.7602	9.022	-0.1325	-0.1581	-0.0576	-0.0875
Japan	0.00016	0.0285	-0.8494	10.184	-0.0769	-0.1121	-0.0460	-0.0672
Netherlands	0.00125	0.0287	-1.1806	12.248	-0.0959	-0.1255	-0.0451	-0.0721
New Zealand	0.00029	0.0169	-0.8382	7.912	-0.0551	-0.0706	-0.0259	-0.0428
Norway	0.00225	0.0293	-1.0754	10.067	-0.0874	-0.1318	-0.0467	-0.0730
Portugal	0.00060	0.0268	-0.8859	9.995	-0.0917	-0.1209	-0.0416	-0.0650
Singapore	0.00056	0.0286	-0.4028	7.809	-0.0732	-0.1175	-0.0469	-0.0674
Spain	0.00094	0.0315	-0.7408	8.308	-0.0856	-0.1295	-0.0496	-0.0745
Sweden	0.00177	0.0315	-0.4523	7.142	-0.0861	-0.1178	-0.0507	-0.0733
Switzerland	0.00118	0.0259	-0.8744	14.034	-0.0721	-0.1122	-0.0398	-0.0612
UK	0.00125	0.0242	-1.3656	17.338	-0.0633	-0.1004	-0.0351	-0.0565
USA	0.00146	0.0232	-0.8104	9.393	-0.0707	0.0963	-0.0354	-0.0558

Table 3.4: Out-of-Sample Performance of Boostrap Pessimistic Portfolio, January 2007 - December 2012

Under real data option, we directly use the sample of first 6 years to train pessimistic portfolio, which is classical plug-in method. In other scenarios, stationary bootrap method is employed to compute average weights. q is the smooth parameter.

	Quantile	Annual Return	Sharpe Ratio
Real Data	1	0.0285	0.204
Real Data	10	0.0346	0.205
Real Data	20	-0.0179	-0.070
$q = 0.5$	1	0.0269	0.203
$q = 0.5$	10	0.0343	0.205
$q = 0.5$	20	0.0424	0.222
$q = 0.1$	1	0.0249	0.191
$q = 0.1$	10	0.0345	0.207
$q = 0.1$	20	0.0394	0.209
$q = 0.01$	1	0.0201	0.150
$q = 0.01$	10	0.0277	0.163
$q = 0.01$	20	0.0222	0.112

Table 3.5: Out-of-Sample Degeneracy Moments

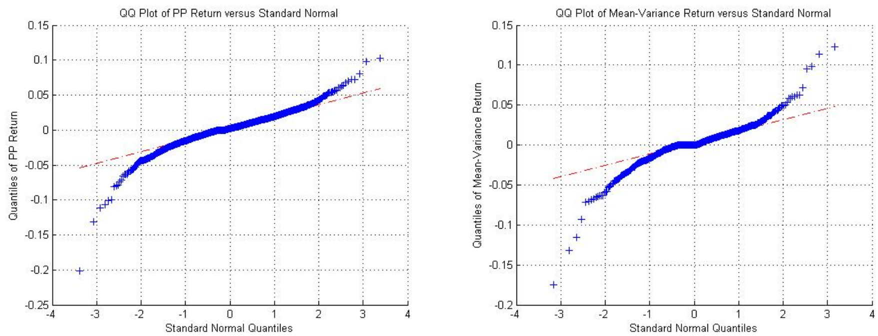
Each moment is based on difference between realized return and in-sample target return. Parameters of MV and PP are $L_1 = 100, \tau = 0.1, \Delta t = 4$. PPP1 colume has a rolling window length $L_2 = 25$ while the last colume $L_2 = 50$.

	PP	MV	PPP1	PPP2
Mean	-0.0015	-0.0013	-0.0013	-0.0017
Standard Deviation	0.0218	0.0194	0.0094	0.0089
Skewness	-1.190	-1.581	-0.8562	-0.743
Kurtosis	11.68	21.67	7.0996	6.624

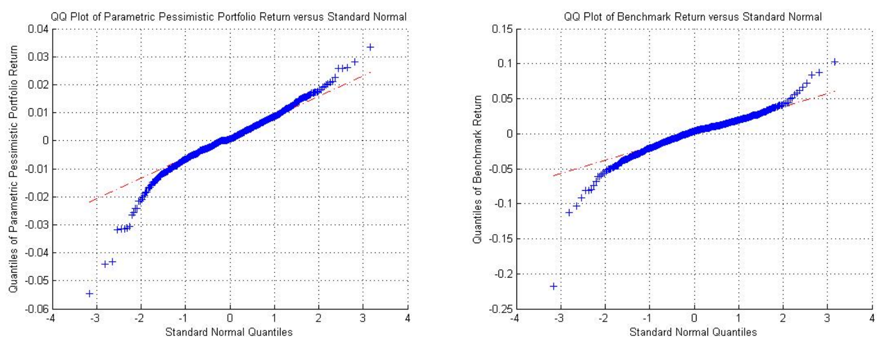
Table 3.6: Breakeven Transaction Cost

Breakeven transaction cost is ratio between annual return and average turnover, measuring vulnerability of portfolio profitability to market friction. Parametric pessimistic portfolio has overall the highest c_b due to significant lower turnover. Mean-variance suffers from volatile capital re-allocation. Thus its return deteriorates with increasing transaction cost.

Rebalance weekly			Total Turnover	c_b
$L = 100$	PP	$\tau = 0.1$	598.99	0.0014
		$\tau = 0.01$	556.19	0.0008
	MV	–	789.70	0.0026
$L = 50$	PPP	$\tau = 0.1$	254.35	0.0068
		$\tau = 0.01$	277.66	0.0069
$L = 150$	PP	$\tau = 0.1$	516.91	0.0013
		$\tau = 0.01$	511.12	0.0012
	MV	–	600.30	0.0036
$L = 25$	PPP	$\tau = 0.1$	303.95	0.0059
		$\tau = 0.01$	283.53	0.0057
Rebalance monthly			Total Turnover	b
$L = 100$	PP	$\tau = 0.1$	223.48	0.0025
		$\tau = 0.01$	224.89	0.0019
	MV	–	354.82	0.0046
$L = 50$	PPP	$\tau = 0.1$	207.69	0.010
		$\tau = 0.01$	238.38	0.0090
$L = 150$	PP	$\tau = 0.1$	190.91	0.0044
		$\tau = 0.01$	211.04	0.0023
	MV	–	283.51	0.0055
$L = 25$	PPP	$\tau = 0.1$	224.83	0.0099
		$\tau = 0.01$	250.85	0.0085

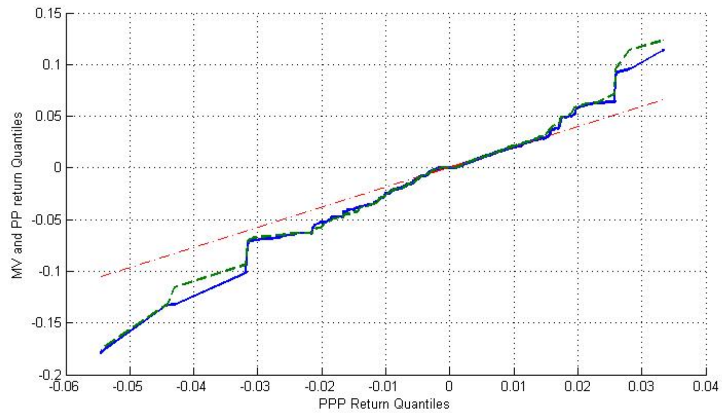


(a) Pessimistic Portfolio and Mean Variance

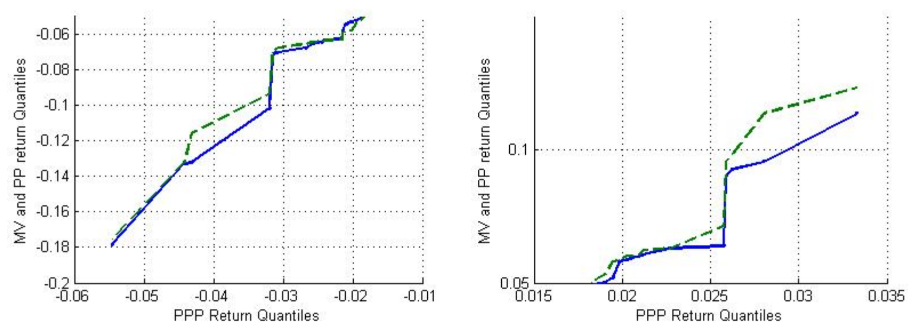


(b) Parametric Pessimistic Portfolio and Benchmark

Figure 3.3: QQ Plot of empirical return distribution vs normal distribution. As x-axis is quantile of normal distribution, adeviation from the dash fitted line is evidence of non-normality. Specifically, leptokurtice portfolio returns have downward-biased left tail and upward-biased right tail. Pessimistic and mean-variance return are simulated with rolling window length $L_1 = 100$, parametric pessimistic return $L_2 = 25$. $\tau = 0.1$ in both pessimistic models. All portfolios are rebalanced monthly ($\Delta t = 4$).



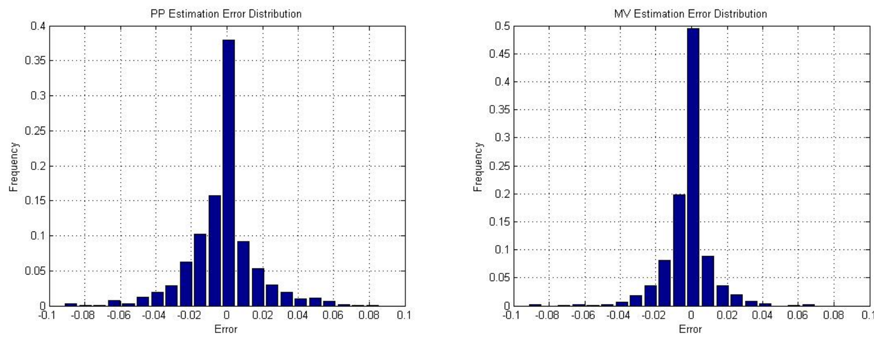
(a) QQ Plot of MV and PP Return on PPP return



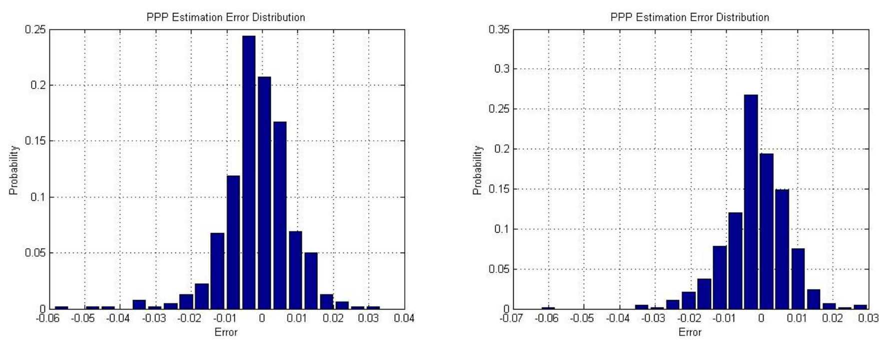
(b) Left Tail and Right Tail

Figure 3.4: QQ Plot Comparison between two simulated returns.

In upper figure, dashed curve shows distributional characteristics between MV and PPP while solid curve is that of PP. Lower panel plots details at left and right tail. Parameters: $L_1 = 100, L_2 = 25, \tau = 0.1, \Delta t = 4$.



(a) PP and MV



(b) PPP with $L_2 = 25$ and $L_2 = 50$

Figure 3.5: Out-of-Sample Degeneracy of three active portfolio schemes.

It is calculated by the difference between training sample target return and realized return in the holding period. Parameters of upper panel are $L_1 = 100, \tau = 0.1, \Delta t = 4$. The left figure at lower panel has a rolling window length $L_2 = 25$ while the right one $L_2 = 50$.

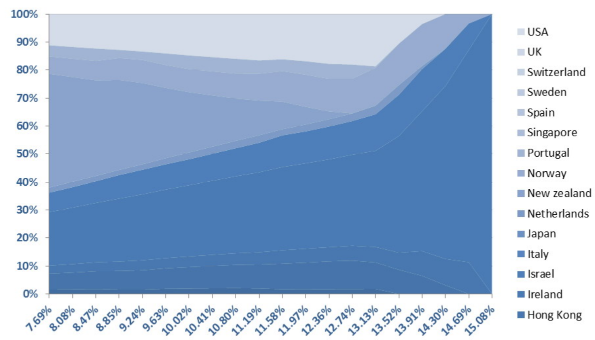
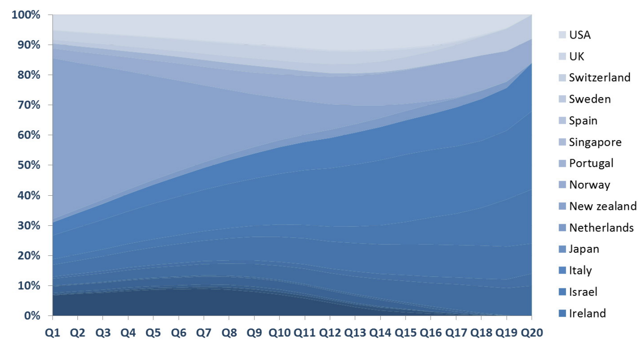
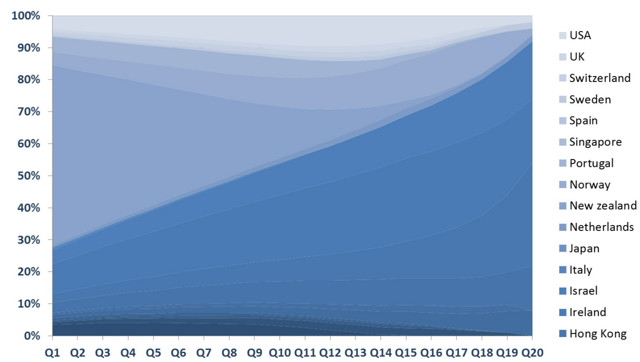


Figure 3.6: Portfolio Composition with Various Target Return on December. 2006.

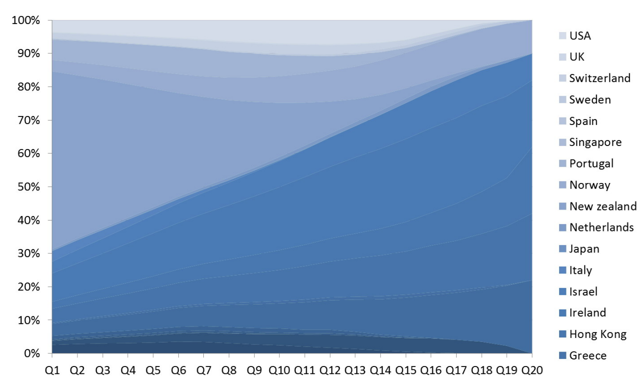
This is an example of pessimistic portfolio weights varying with target return. As μ increases, allocation tilts to the assets with high in-sample return. $\tau = 0.1$



(a) $q = 0.5$



(b) $q = 0.1$



(c) $q = 0.01$

Figure 3.7: Bootstrap Portfolio Composition with Various Target Return on December. 2006.

Various average resample block lengths q are selected to check stability of bootstrap pessimistic portfolio. $\tau = 0.1$.

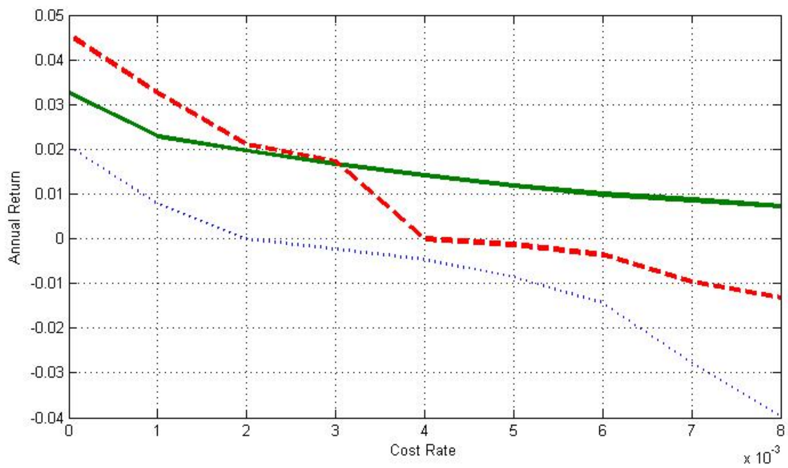


Figure 3.8: Portfolio profitability resistence to market friction.

Dash line: PP, bold dash line: MV, bold solid line: PPP. Parameters of MV and PP are $L_1 = 100, \tau = 0.1, \Delta t = 4$. PPP1 colume has a rolling window length $L_2 = 25$ while the last colume $L_2 = 50$.

4 Mean-Variance on Heterogeneous Sample

4.1 Background

The traditional Mean-variance approach has first been investigated by Harry Markowitz (1952, 1959). Mean-variance model was later carefully researched and developed to be a framework of strategic asset management (CAPM). Portfolio optimization is formulated, in this paradigm, as a process of utility maximization problem by investors (Markowitz, 1956) and analogous optimizers with constraints such as stochastic dominance by Kuosmann (2004), Giorgi (2005), Dentcheva (2006) and tail risk by Schmeidler (1989) are invoked. The spirit of the approach is to minimize risk with given a level of expected return, based on mean and variance to estimate the diversify risk. Early studies as early as Markowitz, adopted the expected mean return and standard deviation as the risk of asset to do the asset portfolio. Fisher Black and Robert Litterman (1992) did the special study for global portfolio optimization. They used an approach that combines two established tenets of modern portfolio theory, which are the Mean-Variance Model of Markowitz and the Capital Asset Pricing Model (CPAM) of Sharpe (1964). This approach allows investors to construct portfolio across global equities, bonds and currencies. Their data set covered 17 developed countries from January 1975 to August 1991. Through rolling-window simulation, investors are able to compare ongoing performance with two strategies: one is investing in the bonds market with high yields, and the other is in the equities market with high dividend yield relative to bond yield. Their research could be empirical evidence that investors are indeed in informational advantage of forming perspective on global portfolio behavior. Choueifaty and Coignard (2008) find mean-variance portfolio outperforms both market capitalization weighted portfolio and equity weighted portfolio in long term investment. And mean-variance portfolio has higher sharp ratio and lower volatility. Nicholas George Baccash (2010) argues that mean-variance is quite better for global portfolio. He researches G7, MSCI DM and MSCI EM, find that emerging markets offer to American investor the best opportunities to reap the benefits of global equity diversification.

However, mean-variance builds its validity on a set of unrealistic assumptions. Follmer and Leukert (2000) describe a problem of partial hedging with quadratic loss function. Actually it is maximizing expected utility with non-negative marginal utility. Richard (2004) explains why mean-variance optimization is not useful for investment management. Investors are never perfectly informationally efficient, and facing the risk of estimating portfolio risk and return (ambiguity or model uncertainty), and biasedness is unfortunately magnified by MV optimizers. Michaud (1989) discuss the practical problems of mean-variance model such as generating suboptimal portfolios far from efficient frontier. Others studies also prove that traditional mean-variance could be problematic when distributional conditions are violated. The main problem of Markowitz model is that the optimization procedure often results in concentrated portfolios, requirement of too much sample data, model lack of robustness and corner solutions. (see Michaud (1989), Best and Grauer (1991) and Black and Litterman (1992))

As more critiques focused on unrealistic quadratic utility function assumption, forecasting error magnifier, some researchers (Engle 2002, Timmermann 2006) propose several techniques enhancing estimation accuracy in expected return and covariance matrix. In this chapter, we intend to design a hierarchical mean-variance portfolio scheme consisting of trading strategy, state classifier and classical optimizer. The aim is to offering a fresh application of traditional portfolio technique. The following section is organized as: 3.2 establishes an explicit mean-variance interpretation both in discrete and continuous case. 3.3 attempts to set up a momentum strategy from path-dependent PDEs. 3.4 introduces data pre-processer as a classifier for various market state. 3.5-3.8 are an empirical example followed by conclusion.

4.2 Model Configuration

Since risk/reward paradigm has been well established and restated by different frameworks, interpretation can be varied from a special case of stochastic dominance portfolio choice to continuous stochastic multi-period utility maximization under certain constraints. In this section, we are not so ambitious to synthesize a comprehensive model applicable to all scenarios, but intending to demonstrate its error-magnifier nature from econometric perspective.

Consider N risky assets with random returns vector $r(t)$ and a risk free asset with known return $r^f(t)$. Define the excess returns by the difference of the two, with conditional means and covariance matrix $\mu(t)$ and $V(t)$ respectively.

Assume, for now, that the excess returns i.i.d. with constant moment.

Suppose the investor can only allocate wealth to the N risky securities. In the absence of a risk-free asset, the mean-variance problem is to choose the vector of portfolio weights $x(t)$, which represent the investors relative allocations of wealth to each of the N risky assets, to minimize the variance of the resulting portfolio return $r_p(t) = x^{'}r$ for a predetermined target expected return of the portfolio $r^f + \mu$:

$$min_x Var(r_p) = x^{'}Vx \tag{4.1}$$

Subject to

$$Er_p = x^{'}(r^f + \mu) \ \ and \ \ \sum_{i=1}^{N} x_i = 1 \tag{4.2}$$

The first constraint fixes the expected return of the portfolio to its target and the second one is a standardized condition to ensure constant leverage in the risky assets. Applying the Lagrangian and solving the corresponding first-order conditions, we have the following optimal portfolio weights:

$$x^* = \Gamma + \Psi x$$

With

$$\Gamma = \frac{1}{C_4}[C_2(V^{-1}\lambda) - C_1(V^{-1}\lambda)] \ \ and \ \ \Psi = \frac{1}{C_4}[C_3(V^{-1}\mu) - C_1(V^{-1}\mu)]$$

Where λ a scalar factor measuring risk aversion and $C_1 = \lambda^{'}V^{-1}\mu, C_2 = \mu^{'}V^{-1}\mu, C_3 = \lambda^{'}V^{-1}\lambda, C_4 = C_2C_3 - C_1^2$. The minimized portfolio variance is equal $x^{*'}Vx^*$.

The Markowitz paradigm yields two important economic insights. First, it illustrates the effect of diversification. Imperfectly correlated assets can be combined into portfolios with preferred expected return-risk characteristics. Second, it shows that, once a portfolio is fully diversified, higher expected returns can only be achieved through more extreme allocations (notice x^* is linear in μ) and therefore by taking on more risk. If we have non-zeros risk-free rate, efficient concave frontier by CAPM is constructed as figure

The problem is also evident as the paradigm relies on accuracy of forecast on expected return and covariance. To understand mechanism of error magnification, derivative of optimal weights respect to expected return is given by

$$dx = \alpha X dt + \sigma X dW \tag{4.3}$$

Which has an explicit variance as

$$VarX(t) = [X(0)]^2 exp(2\alpha t)[exp(\sigma^2 t) - 1]$$

We define equivalence of two stocks by the same parameter set of $\{\alpha, \sigma, W\}$ in equation (4.3). Here essentially both characteristics of stochastic process and driven force of randomness are shared. Path consistency is a strong condition for stochastic equivalence in this setting. A weak similarity may be defined as

Definition 4.1 *Two risky assets are said to be weakly stochastically equivalent if and only if they shared* $\{\alpha, \sigma\}$.

One interesting aspect concerning the equivalence is that there is no possibility of arbitrage. Path equivalence guarantees $X_1 - X_2 = 0$. Weak equivalence, contrarily, indicates non-stationary process of difference that mean-reverting style strategy does not work. A portfolio P is a positive valued combination of risky assets which satisfies the following conditions. Consider the identity:

$$\frac{dP}{P} = w' \frac{dX}{X} \tag{4.4}$$

Where $w = w(t)$ is weight variable vector, reflecting implementing certain portfolio policy and $\sum w = 1$. X is vector of stock process $\{X_1, X_N\}$. A passive portfolio is a portfolio in which the number of shares remains constant which reads:

$$dP = v^{'} dX \tag{4.5}$$

Where v is the fraction of shares in portfolio P. Market portfolio is a passive portfolio by nave summation of all stocks. A balanced portfolio is a portfolio of the form (4.4) where all weights w are constant. It has consequently two parameters $\{\alpha_P, \sigma_P\}$ describing its lognormal process.

With concave utility function, followed by Markowitz (1952, 1956, 1959) and Merton (1973), CAPM provides efficient portfolios that maximizes investors utility.

4.3 Predictability and Trading Strategy Design

In this section, we intend to develop a predictor of return from path-dependent PDEs and time-series momentum, in hope of enhancing mean-variance portfolio performance.

Theorem 4.2 (Dupire, B. 2009) *Consider the functional running maximum of the path:*

$$M_t = \max_{0 \le \mu \le t} X_t(\mu)$$

Here $X \in \Lambda_t \in \Lambda$ *is a path of* x_t *up to time* t. x_t *is a real-valued random variable with standard Brownian motion* $dx_t = dw$. *A functional is a mapping* f:$X_t \to R$ *Then following equation holds:*

$$M_t(u) = m_u = x_0 + \frac{1}{2}\bar{L}_u^0 \tag{4.6}$$

Where $u \in [0, t]$ *and* m_u *is maximum value at time* u. $\bar{L}_u^0$ *denotes local time when* $m_u - x_u = 0$ *is satisfied.*

Remark: In classical Brownian motion, maximum is proportional to time spent to beat the maximum. Equation (4.6) however shows momentum effect that stronger upside potential is associated with the time spent on beating maximum. It justifies profitability of popular momentum strategy. Note that it doesnt assume fractional Brownian motion which means that even path-standard Brownian motion somewhat exhibit inertia and long-term memory.

We may also define the functional running minimum of the path and obtain the following results.

Lemma 4.3 *Assuming the same condition as Theorem 3.1, consider the functional running minimum of the path:*

$$K_t = f(X_t) = \min_{0 \le u \le t} X_t(u)$$

Then following equation holds:

$$K_t(u) = k_u = x_0 - \frac{1}{2}\underline{L}_u^0 \tag{4.7}$$

Lemma 4.4 *With conditions and notations the same as Theorem (4.2) and Lemma (4.3) except $dx_t = \sigma dw$ and σ is constant. We have analogous relations:*

$$M_t(u) = m_u = x_0 + \frac{\sigma}{2}\bar{L}_u^0 \tag{4.8}$$

$$K_t(u) = k_u = x_0 - \frac{\sigma}{2}\underline{L}_u^0 \tag{4.9}$$

Remark: (4.8) and (4.9) are consistent with intuition that higher volatility indicates wider deviation. We may rearrange those equations to demonstrate relative profit of directional movement:

$$ED_t = (m_t - x_0) - (x_0 - k_t) = \frac{\sigma}{2}(\bar{L}_u^0 - \underline{L}_u^0)$$

And define sharpe ratio consequently assuming zeros risk-free rate

$$sr_t = \frac{D_t}{\sigma} = \frac{1}{2}(\bar{L}_u^0 - \underline{L}_u^0)$$

It suggests that contribution of sharpe ratio is the discretion of the time that price stays at highest level or lowest level.

A profitable strategy thus is designed to gain an exposure on $\bar{L}_u^0 - \underline{L}_u^0$. Or trading signal is generated by $sign(\bar{L}_u^0 - \underline{L}_u^0)$. To avoid the situation of stagnate m_t or k_t when price temporarily moves in adverse direction after a new m_t or k_t was achieved, we may allow flexible origin. One practical method is to fix rolling window length in calculating $(\bar{L}_u^0, -\underline{L}_u^0)$, and base decision on most recent fluctuations.

4.4 Nonparametric Classifier Filtering

4.4.1 K Nearest Neighbors (KNN)

Suppose a real-valued random variable $X \in R^q$ and we have a sample of realizations $\{X_1, , X_n\}$.

In general setting, consider a point in $q-$dimensional real space $X \in R^q$, and random variable x_i, some distance metric $D_i = \|x_i - X\|$ is defined to measure similarity. Here $\{x_i, i = 1, ...n\}$ may be chosen to be feature vector. The order statistics follow definition by Kendall (1956), i.e. $R_i = \sum_{j=1}^{n} \mathbf{1}(D_j \leq D_i)$. The observations that satisfy $\{x_i | R_i \geq k\}$ are the k nearest neighbors the most similar and comparable elements. We therefore are able to construct knn density estimate and knn regression like kernel method. Note that the classifier truncates sample to reduce contamination of outliers

Definition 4.5 *Suppose $X \in R^q$ has multivariate density $f(x)$. A multivariate uniform kernel to estimate local $f(x)$ at x can be designed inversely proportional to the rank*

$$\tilde{f}(i) = \frac{k}{nc} R_i^{-q}$$

Where $c = \pi^{q/2}/(\Sigma((q+2)/2))$ represents space of unit ball expanded in R^q.

Using second-order taylor expansion, it is easy to show that unconditional expectation of density is true value biased by the square of ranks multiplied by gradient of $f(x)$. Variance is affected by the expectation of ranks and $f(x)$.

Definition 4.6 *Given dependent variable is functionally correlated with ex-*

planatory variables:

$$y_i = \Psi(x_i) + e_i \;\; and \;\; E(e_i|x_i) = 0$$

Simple knn estimator is the arithmetic mean of nearest neighbors:

$$\tilde{\Psi}(x,k) = E\mathbf{1}(R_i \geq k)y_i = \frac{1}{k}\sum_{i=1}^{n}\mathbf{1}(R_i \geq k)y_i$$

If some weighting function w is introduced, smoothed k-nn estimator is

$$\tilde{\Psi}(x) = \frac{\sum_{i=1}^{n} w(R_i \geq k)y_i}{\sum_{i=1}^{n} w(R_i \geq k)}$$

Cover (1968) first discussed convergence rate for knn algorithm. Short, R. and Fukanaga, K. (1980,1981) further discussed measures applicable in defining optimal distance.

4.4.2 Discriminant Adaptive Nearest Neighbor Classification (DANN)

Nearest neighbor classifier expects a locally constant conditional probability and spread a sphere to boundarize similar points. This method suffers severely as dimension increases. It also performs poorly compared to Bayesian technique when boundaries exhibit distorted nature. Hastie and Tibshirani (1996), Delannay, N. et. al. (2006) proposed alternative classification DANN, which is flexibly adaptable to determine local decision boundaries efficiently. Its procedure is summarized as:

- Set prior metric matrix $\Gamma_0 = I$ (typically identity one) or start from updated Γ_i

- Obtain k nearest neighbors round point concerned x_b with some specific distance measure
- Compute the weighted loss matrics either within the included sample W or interactive one across two subsets B.
- Conduct bayesian modification by $\Gamma_{i+1} = W^{-1/2}[W^{1/2}BW^{-1/2}+\varepsilon I]W^{1/2}$
- Stop iteration until the maximum number of loops is reached, and use the revised Γ^*

4.5 Data Input and Implementation

Up to now, several blocks have been fragmentally discussed. We are able to introduce a coherent framework of combining these techniques. The basic idea is hierarchical structure aiming to maximize portfolio value.

- Use strategies in Section 4.3 as preliminary filter biasing asset return. Positively skewed distribution tends to enhance absolute overall performance regardless of portfolio scheme. It offers opportunity to test applicability of knn mean-variance optimization.
- Cluster similar historical data using knn or DANN. The pre-processing procedure is acting as a filter purifying local behavior. This is the major difference that our approach deviates from traditional mean-variance analysis.
- c) Apply mean-variance optimizing portfolio.

Practically indicators characterizing similarity should be defined in advance. In this research, we focus on a set of technical indicators whose effectiveness can be properly assessed (White 2000, Timmermann and White 1999). It turns out that only elements with statistically significant predictive power contribute positively to purifying samples.

To test efficiency of the approach, we collect 24 global commodity future daily data including metals, energy, live stocks, agriculture and soft dating back to 1st Jan. 1999. They are traded in Chicago Board of Trade (CBOT), London Metal Exchange (LME), Inter-Continental Exchange (ICE) and Chicago Metal Exchange (CME). For each future contract, a joint index that combines the most actively traded contracts is selected. If one commodity is traded in various exchanges, only the most liquid contract is chosen. Rolling return is ignored in this study. Table 4.1 shows contract details and the date of listing. Descriptive Statistics are summarized in table 4.2.

All commodities during the past 15 years have experienced price appreciation, reflected by positive mean returns. Ex Ante Volatility is fairly stable cross section. Unlike stocks, futures differ remarkably in 3rd-order moment, an indication of diverse directional long-term movement. Applicability of momentum strategy is anticipated evidenced by much higher kurtosis.

In addition, both $VaR_{0.01}$ and $CVaR_{0.01}$ varies significantly relative to $VaR_{0.05}$ and $CVaR_{0.05}$, which is more volatile than standard deviation. The phenomena may verify employing mean-variance is preferable to pessimistic portfolio, whose performance is heavily influenced by tail risk stationarity. Section 3.4.1 discussed this issue in details.

4.6 Performance

Figure 3.2 plots out-of-sample performance of three portfolio scheme: Mean-variance approach, nave portfolio that equally invests in each contract and buy-and-hold benchmark.

Albeit MV approach is able to achieve higher return, the larger volatility can significantly erod its superiority. Another remarkable advantage is MVs return is positively skewed, indicating higher proposition of profitable period.

4.7 Analysis

4.7.1 Reality Check on Strategy Profitability

As we discussed in Section 4.5, the revised mean-variance scheme also depends on effectiveness of trading strategy developed in Section 4.3. Table 4.4 compiles manipulated return after it is loaded on each contract.

Compared with Table 4.2, mean return shrinks associated with trading strategy. This is partially because neutral position occupies large portion of trading time in contrast to full long position in buy-and-long. Standard deviation and tail risk, on the contrary, is effectively controlled. Skewness and Kurtosis shows its little improvement in higher order normality. It seems that mixed evaluation results cast doubt on strategy profitability. In order to further investigate its validity, we adopt reality check proposed by White (2000). Basically it experimentally tests the null hypothesis:

$$H_0 = E[ln(1 + S_k r) - ln(1 + S_0 r)] = 0$$

Where S_k is trading signal generated by strategy k. r is asset return. S_0 represents some benchmark signal. In our case, buy-and-hold serves as basic policy which means $S_0 1$. Further details of the technique are referred to Section 6.3. Table 4.5 summarizes p-value with different smooth parameter q, or average block length of $1/q$.

Null hypothesis H_0 is fail to reject only in live cattle case. Wheat and cotton has some mixed conclusion but the trading strategy reserves ability of profiting in 21 out of 24 contracts. One may note that various smooth parameters provide consistent p-value which is a major advantage of stationary bootstrap (Romano, 1994)

4.7.2 Robustness

Out-of-sample degeneration is defined by the difference between target return and realized return. Distribution of mean-variance degeneration is shown in Figure 4.3.

As Michaud.R (2008) states, active portfolio suffers more or less from the inferior performance relative to training period. It is evoked by inaccuracy in forecasting techniques or improper modeling of risk pricing as well as data generating process. Since both asset return and strategy return is leptokurtic, we may use qq-plot detecting absolute and relative normality

4.8 Concluding Remarks

Mean-variance approach, deserving appraisal mixed with critiques, remains controversial in literature after 60 years debate. Yet however severely it is restricted by unrealistic assumptions and practical inoperability, dual problem paradigm or constrained optimization, in the spirit of utility maximization, is still motivating portfolio research. As the starting point, we attempt to construct a hierarchical mean-variance to capitalize on directional moment and strategically allocate resource on state-dependent risk factor. The framework has a higher profitability than both equal weights scheme and benchmark, but suffers from out-of-sample degeneration and non-normality. It reveals some basic disadvantages that most active portfolio management cannot avoid.

4.9 Appendix

Table 4.1: Contracts Covered in MSCI

Commodity	Exchange	Category	Start Date
Aluminum	LME	Metal	1st Jan 1999
Copper	LME	Metal	28th May 2002
Lead	LME	Metal	1st Jan 1999
Nickel	LME	Metal	1st Jan 1999
Zinc	LME	Metal	1st Jan 1999
Gold	CME	Metal	1st Jan 1999
Silver	CME	Metal	1st Jan 1999
Light Crude Oil	CME	Energy	1st Jan 1999
Brent Crude Oil	CME	Energy	1st Jan 1999
Heating Oil	CME	Energy	1st Jan 1999
Gasoline	CME	Energy	20th Oct. 2005
Gas Oil	CME	Energy	8th Sep. 2003
Natural Gas	CME	Energy	1st Jan 1999
Lean Hogs	CBOT	Live Stocks	1st Jan 1999
Live Cattle	CBOT	Live Stocks	1st Jan 1999
Feeder Cattle	CBOT	Live Stocks	1st Jan 1999
Wheat	CBOT	Agriculture	1st Jan 1999
Corn	CBOT	Agriculture	1st Jan 1999
Soybean	CBOT	Agriculture	30th Mar. 2006
Coffee	CBOT	Agriculture	1st Jan 1999
Sugar	CBOT	Agriculture	1st Jan 1999
Cocoa	CBOT	Agriculture	1st Jan 1999
Cotton	CBOT	Agriculture	1st Jan 1999

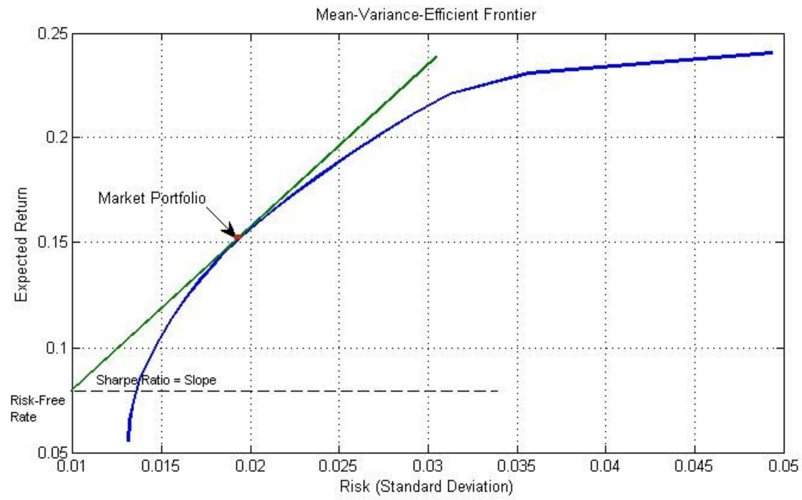


Figure 4.1: Mean-Variance Efficient Frontier

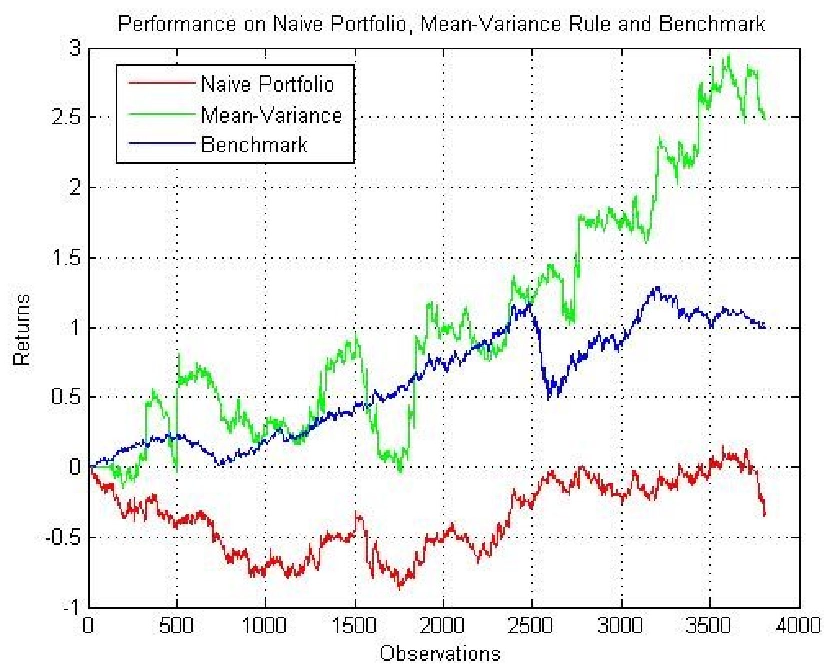


Figure 4.2: Portfolio Performance

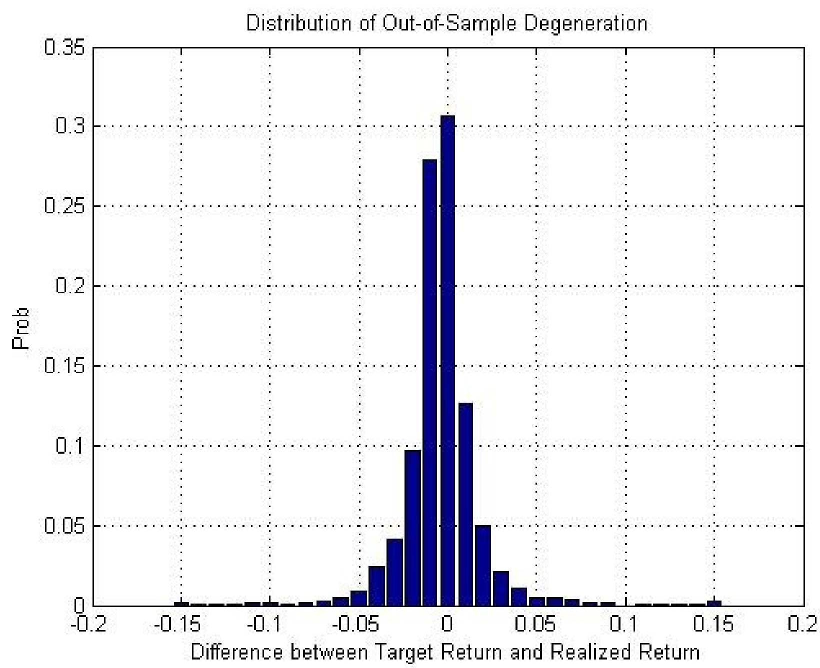
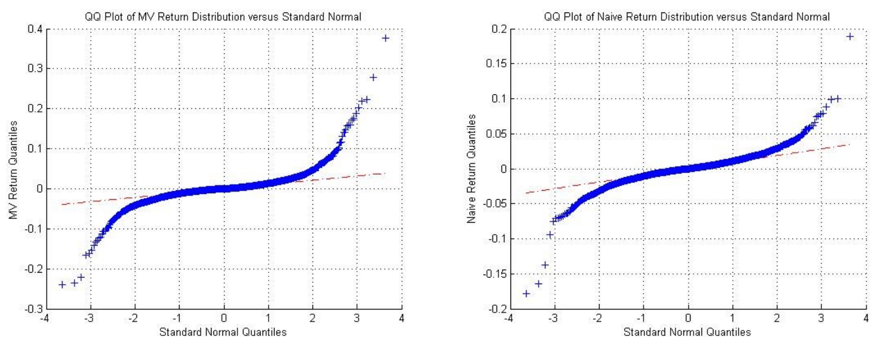
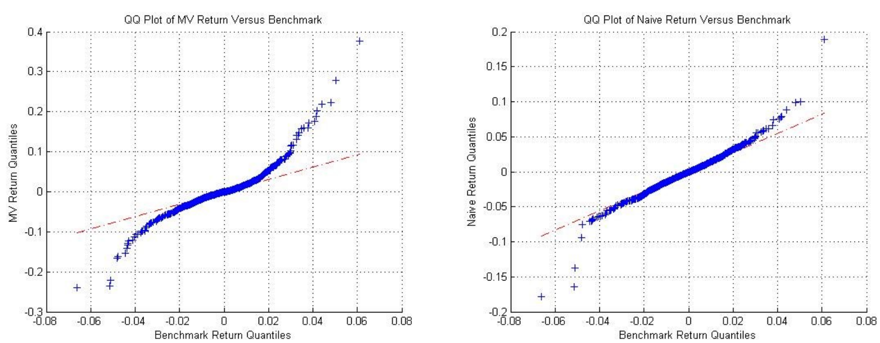


Figure 4.3: Out-of-Sample Degeneracy



(a) Pessimistic Portfolio and Mean Variance



(b) Pessimistic Portfolio and Mean Variance

Figure 4.4: Out-of-Sample Degeneracy

Table 4.2: Descriptive Statistics of Continuous Futures Index

Mean is multiplied by 10,000. Empirical Value-at-Risk and Conditional VaR are quantiles and average values beyoud quantiles after sorting returns.

Commodity	Mean	SD	Skewness	Kurtosis	$VaR_{0.01}$	$CVaR_{0.01}$	$VaR_{0.005}$	$CVaR_{0.005}$
Aluminum	0.94	0.0138	-0.2479	5.4812	-0.0404	-0.0515	-0.0214	-0.0320
Copper	4.9	0.0190	-0.1278	6.8561	-0.0558	-0.0714	-0.0300	-0.0450
Lead	3.8	0.0211	-0.2543	6.3788	-0.0627	-0.0777	-0.0346	-0.0512
Nickel	3.2	0.0240	-0.1460	6.5721	-0.0673	-0.0867	-0.0382	-0.0551
Zinc	1.8	0.0192	-0.1984	5.8767	-0.0541	-0.0716	-0.0307	-0.0459
Gold	3.9	0.0117	-0.1609	9.6894	-0.0334	-0.0441	-0.0184	-0.0281
Silver	3.6	0.0200	-1.0049	11.240	-0.0615	-0.0900	-0.0320	-0.0517
Light Crude	5.7	0.0238	-0.2171	7.4242	-0.0669	-0.0917	-0.0373	-0.0558
Brent Crude	6.1	0.0219	-0.2390	5.9506	-0.0599	-0.0796	-0.0348	-0.0512
Heating Oil	5.7	0.0231	-0.5126	7.9880	-0.0601	-0.0879	-0.0357	-0.0530
Gas Oil	5.3	0.0194	0.0619	4.7888	-0.0504	-0.0601	-0.0326	-0.0429
Gasoline	3.0	0.0241	-0.1707	6.8677	-0.0730	-0.0905	-0.0408	-0.0581
Natural Gas	1.4	0.0349	0.5931	8.7527	-0.0847	-0.1103	-0.0533	-0.0746
Lean Hogs	2.5	0.0207	0.4378	28.11	-0.0432	-0.0814	-0.0281	-0.0435
Live Cattle	1.9	0.0101	0.1934	9.1447	-0.0265	-0.0367	-0.0155	-0.0228
Feeder Cattle	2.1	0.0089	-0.1718	7.0558	-0.0239	-0.0320	-0.0144	-0.0205
Wheat	2.2	0.0203	0.2460	5.4825	-0.0511	-0.0670	-0.0300	-0.0436
Wheat	2.1	0.0179	0.0969	5.1622	-0.0492	-0.0600	-0.0272	-0.0397
Corn	2.1	0.0187	-0.3942	12.846	-0.0506	-0.0663	-0.0285	-0.0419
Soybean	3.7	0.0171	-0.7544	7.9480	-0.0530	-0.0679	-0.0290	-0.0437
Coffee	0.042	0.0223	0.3285	9.2241	-0.0582	-0.0792	-0.0345	-0.0500
Sugar	0.14	0.0234	-0.5249	7.9498	-0.0696	-0.0958	-0.0354	-0.0567
Cocoa	1.4	0.0204	-0.1611	5.7570	-0.0554	-0.0739	-0.0324	-0.0476
Cotton	0.91	0.0200	-0.4898	15.181	-0.0513	-0.0682	-0.0321	-0.0457

Table 4.3: Momentum Strategy Performance

Mean is multiplied by 10^5 while standard deviation and all tail risk measures by 10^4.

Commodity	Mean	SD	Skewness	Kurtosis	$VaR_{0.01}$	$CVaR_{0.01}$	$VaR_{0.005}$	$CVaR_{0.005}$
Aluminum	3.8	20	0.2491	33.652	-56	-93	-23	-46
Copper	6.0	25	-0.1971	27.153	-43	-73	-17	-35
Lead	2.6	16	5.0006	171.77	-42	-72	-16	-34
Nickel	4.5	10	0.4184	36.346	-26	-40	-13	-22
Zinc	5.3	17	3.4549	58.328	-50	-69	-18	-36
Gold	5.3	27	-0.3716	53.511	-82	-139	-29	-64
Silver	4.9	16	0.9518	32.878	-49	-77	-19	-39
Light Crude	-1.4	11	-0.1709	26.643	-35	-53	-16	-29
Brent Crude	3.5	12	0.1814	18.874	-40	-57	-17	-31
Heating Oil	-0.75	11	-0.3096	21.061	-37	-55	-16	-29
Gasoline	1.5	11	0.8728	26.939	-33	-51	-14	-26
Gas Oil	0.22	15	-0.5590	31.879	-48	-76	-20	-38
Natural Gas	-0.39	7	-0.6946	23.280	-27	-39	-10	-20
Lean Hogs	-3.7	22	-4.8581	171.60	-51	-125	-18	-47
Live Cattle	8.5	26	3.1357	80.278	-71	-115	-26	-56
Feeder Cattle	8.8	26	-0.3764	19.773	-82	-124	-35	-65
Wheat	-3.5	12	-3.9207	64.937	-34	-69	-15	-31
Wheat	-1.9	14	-1.8019	33.321	-47	-72	-18	-36
Corn	-3.7	18	-7.8238	227.23	-54	-99	-20	-44
Soybean	3.6	18	0.8528	23.101	-57	-78	-23	-43
Coffee	-2.7	11	0.8992	26.409	-37	-58	-14	-29
Sugar	2.4	13	-2.5954	89.374	-38	-65	-14	-29
Cocoa	-2.9	14	-4.9292	150.01	-46	-76	-16	-35
Cotton	-3.3	13	-1.2842	60.423	-41	-66	-17	-33

Table 4.4: Reality Check p-Value

r is the reality check statistics on return. sr is the reality check statistics on sharpe ratio.

	$q = 0.5$		$q = 0.1$		$q = 0.02$	
	r	sr	r	sr	r	sr
Aluminum	0	0	0	0	0	0
Copper	0	0	0	0	0	0
Lead	0	0	0	0	0	0
Nickel	0	0	0	0	0	0
Zinc	0	0	0	0	0	0
Gold	0	0	0	0	0	0
Silver	0	0	0	0	0	0
Light Crude	0	0	0	0	0	0
Brent Crude	0	0	0	0	0	0
Heating Oil	0	0	0	0	0	0
Gasoline	0	0	0	0	0	0
Gas Oil	0	0	0	0	0	0
Natural Gas	0	0	0	0	0	0
Lean Hogs	0	0	0	0	0	0
Live Cattle	1	0.52	0.05	0.06	0.42	0.08
Feeder Cattle	0	0	0	0	0	0
Wheat	0	0.12	0	0.28	0	0.2
Wheat	0	0.02	0	0.04	0	0
Corn	0	0	0	0	0	0
Soybean	3.6	0	0	0	0	0
Coffee	0	0	0	0	0	0
Sugar	0	0	0	0	0	0
Cocoa	0	0	0	0	0	0
Cotton	0	0	0.54	0	0.38	0

Table 4.5: Statistics of Performance

Mean and standard deviation are multiplied by 10^2.

	Mean-Variance	Naive	Benchmark
Mean	16.35	-2.14	6.54
SD	38.48	23.87	14.72
Skewness	1.57	-0.42	-0.32
Kurtosis	41.71	23.73	7.52

Table 4.6: Out-of-Sample Degeneration Moments

Mean and standard deviation are multiplied by 10^2.

	Mean	SD	Skewness	Kurtosis
D	-3.7	2.47	1.48	39.95

5 Functional Predictability of Factors and Basis Portfolios

5.1 Background

5.1.1 Risk Factors

One theoretically critical and practically valuable topic in financial research is whether and how asset return can be predicted from a set of risk factors. Rich literature associated with the issue can be roughly categorized by three motives: exploring new factors that are unexplained by existing ones; measuring efficiency with alternative statistical approaches; the mechanism that assets are behaviorally priced. Early attempts to the first dimension might be dated back to Sharpe (1964) and Black (1972). The failure of well-known CAPM suggests that cross-sectional return distribution should not be solely dependent on sensitivity to market return. This leads to a generalized version by introducing more macro indicators (Cox and Ross (1976); Roll and Ross (1980)). Firm-specific factors, including size (Banz (1981)), leverage effect (Bhandari (1988)) and B/M ratio (also known as HML or B/P ratio) (Chan et al. (1991)) to name a few, offer interpretation different from unobserved loadings like beta. Their joint effect on US stocks is then investigated by Fama and French (Fama and French (1992); Fama and French (1993)). Leverage is abandoned because it is absorbed by B/M. Endogeneity has to be tackled due to interaction of Beta and firm size. It is addressed by constructing 100 Size-beta portfolios. In methodology session, we will show that there exists an elegant way in the context of functional dependence.

One effort of subsequent research is disentangling effect of F-F factors and assessing efficiency with new observations. Berk (1995) pointed out that size-related measure is an evidence of model misspecification rather than a traditional anomaly. Ali et al. (2001) confirmed that the B/P ratio is a representation of market mispricing due to high idiosyncratic risk, transaction costs and lower investor sophistication. Concerning empirical contradiction to F-F result, Penman et al. (2007) found that B/P ratio can be decomposed into enterprise B/M ratio and leverage component, where the latter, as a proxy of financial distress, is negatively related to expected return conditioning on B/M ratio. Frazzini and Pedersen (2014) proposed a tilting strategy of market betas to capitalize on the compensation for liquidity premium. Apergis and Payne (2014) revisited size effect in G7 stock markets from 1991 to 2012, using panel threshold cointegration. Their finding is its asymmetric impact after controlling B/M and P/E ratio.

Another widely-documented anomaly is strong inertia of price movement, pioneered by Jegadeesh and Titman (1993). Thay showed that there is persistent outperformance of buy-winners-sell-losers in US stock market. Carhart (1997) extended F-F model with momentum factor and generated less pricing error in contrast with other comparable models. Asness (1997) argued that negative relation between value and momentum is present. Grinblatt and Moskowitz (2004) studied return dynamics with different time scale and observed that one-year winners gain but one-month and three-years ones suffer. The research by Shieh et al. (2012) suggested successful application in NYSE but not AMEX and NASDAQ. Recent advances include the phenomenon that momentum occasionally crashes during recovery period (Daniel et al. (2012) and Daniel and Moskowitz (2013)) and remarkable forecastability in residual term in F-M regression (Blitz et al. (2011)).

Other factors, with continual exploration on behavioral characteristics, are also burgeoning. Ang et al. (2006) and Ang et al. (2009) reported significant contribution of idiosyncratic volatility. There are also researches on net stock issues effect (Pontiff and Woodgate (2008)) and financial fundamentals, for instance, asset growth effect (Cooper et al. (2008)) and F-score (Piotroski (2000)). Whether investors require a premium for higher moment risk is investigated by Kraus and Litzenberger (1976), Amaya et al. (2011). Asness et al. (2013) designed a comprehensive firm quality factor (QMJ) to evaluate firms strength in four dimensions. Strong profitability of market-neutral strategy on the quality indicator, which is the average of ranking scores, is empirically evidenced in US market. However, predictability of QMJ is not statistically supported after excluding four factors. This inconsistency cast doubt on widely-adopted testing method.

In regard with growing family of factors and their potential interpretability, Chen et al. (2010) proposed an alternative three-factor model (CNZ) consisting of a market risk, an investment factor and ROA. The main intuition is that the price fluctuation is significantly affected by its discount rate, of which investment factor and ROA can be good proxies. The empirical results document an enhancement in U.S. stock market compared to F-F model. Ammann et al. (2012) applied CNZ in 10 EU countries with data ranging from 1990 to 2006. They report comparable, if not better, performance.

To summarize, risk factors have three sources: the sensitivity of individual stock with market index, fundamental elements as indicators of firm′s prospect, and past performance. The efficiency is from investor behavior and affected by their perceptions of those risks. Thus as long as factors are fully priced, excess return is merely a reward for excessive risk exposure.

5.1.2 Empirical Studies on Emerging Stock Market

Due to different regulations, cultures, matching mechanisms, data availability and investor behavior, emerging markets may vary in risk structure, on which the researches may offer a test of factors′ universality. In this section, we will apply cross-sectional analysis for US and Chinese market, in hope of gaining robustness from comparison.

Claessens et al. (1995), Fama and French (1998) conducted nascent investigations with the International Finance Cooperation data from 19 and 13 emerging stock markets, respectively. In spite of slight difference in data sample, they concluded opposite impacts on market risk, size and value factors. The contradiction may be attributed by the difference in modeling technique. More recent research by Cakici et al. (2013) is examination on momentum and value effect in 18 emerging market from 1990 to 2011. Price correction for low P/B ratio is widely documented but buy-winners-sell-losers did not succeed in Eastern EU. In global context, domestic factors have higher significant interpretability. Lischewski and Voronkova (2012) investigated existence of liquidity premium in Warsaw Stock Exchange. The failure of its significant contribution may be caused by the fact that a large proportion of shares is controlled by a small number of institutions and compensation of illiquidity is not effectively reflected. On the contrary, Narayan and Zheng (2010) showed that the risk is priced in Chinese stock market between 1993 and 2003.

In case for Chinese market, size and B/M are considered important, while insignificance of market beta is caused by individual investors' irrational behavior and government intervention (Eun and Huang (2007); Wang and Di Iorio (2007)). With respect to momentum, Naughton et al. (2008) examined alternative trading strategies exploiting the anomaly in Shanghai Stock Exchange, whose profitability explains short-term dynamics in this market. Consistent with Ang et al. (2006) and Ang et al. (2009), idiosyncratic volatility factor also has remarkable influence (Eun and Huang (2007); Nartea et al. (2013)).

Our research on emerging market has two-fold benefit. It offers data set of a counterpart with different legitimate system and market structure, therefore, a way of ′robustification′. A further study in global market is also rewarded.

5.1.3 Functional Data Analysis

Since stocks exhibit similar behavior cross-sectionally, estimated standard error would be correlated when applying pooled OLS. One approach, proposed by Fama and MacBeth (1973), is first to conduct time-series analysis in a rolling sample and run cross-sectional OLS on de-correlated cross-sectional data. It implicitly assumes linearity of prediction, which is not necessarily satisfied (Hiemstra and Jones (1994); Qi and Maddala (1999)). One attempt to tackling nonlinearity is neural network, which typically has several hidden layers trained to forecast multi outputs via a set of factors (Nicholas Refenes et al. (1994)). But this method fails to deliver a measure of predictability, therefore, can be easily contaminated by irrelevant variables. Worse still, it suffers over-fitting in training sample as information to noise ratio is low. A flexible approach that is able to offer statistical inference as well as modeling nonlinear dependence is needed. Another concern of F-M is the difficulty in implementation accounting for missing observations. This problem is especially severe in emerging markets. Finally, multivariate analysis is computationally costing accounting for large data set.

With regard to inefficiencies of current methods, we propose a FDA framework for testing factor predictability. It is advantageous in dimension reduction. For example, NYSE has more than 3,600 stocks; similar case is in Chinese A share market with approximately 2,600 stocks. In this case, idea of continuous curve, rather than vectors in multivariate analysis, is more appropriate for a large system. The technique commonly used in smoothing (reducing dimension) is functional principle component analysis (FPCA), which extracts a set of basis function that maximally explains variation of the curves (Ramsay and Silverman (2005)). By projecting functional objects on orthonormal basis, we are able to represent them as a vector of scores, where classical analysis applies. Kokoszka and Reimherr (2013) proved that dependence of the sign of scores is asymptotically normal and report insignificant simulation results in predicting intraday shapes of curve. Hyndman and Ullah (2007) proposed a two-step algorithm followed by the principle of FPCA to predict mortality and fertility rates in France and Australia. Other applications have been further discussed in many aspects (Ramsay and Silverman (2005);Horvath and Kokoszka (2012); Mestekemper et al. (2010)).

One successful application of functional forecasting is in vehicle insurance pricing Segovia-Gonzalez et al. (2009). They argued that even though two methods have the same objective in re-building an entire process within sampling from different time period, multivariate analysis still bears shortcomings, such as the difficulty in interpretation of variance-covariance and correlation structure; while, FDA has its merit in de-noising observations through local information. Cai and Hall (2006) showed the existence of functional forecasting method, as well as the difference in properties of functional regression estimation with prediction from an estimator of functional slopes. Sood et al. (2009) reported that a better prediction of penetration of new products in using augmented functional model than that of traditional forecasting model. Goia et al. (2010) collected heating demand curves to forecast intra-daily electricity loading demand. As samples in the same group exhibit clustered variance but differs massively from other groups, classification of observations is recommended. The spirit of the procedure is to enhance prediction accuracy via categorization. Antoch et al. (2010) argued that electricity consumption are intrinsically different in two periods, thus grouping objects by weekdays and weekends is valid. A functional clustering forecast model was proposed by Abraham et al. (2003). James and Hastie (2001) constructed a functional linear discriminate analysis to classify new curves.

Asymptotics associated with PCA is dated back to Anderson (1963). Assuming observations are from a multivariate normal distribution, the function of eigenvalues is asymptotically normally distributed. Hall and Hosseini-Nasab (2006) showed that the efficiency of eigenvalues in FPCA is dominantly affected by their spacings. Bootstrap is recommended for constructing confidence interval. Yao et al. (2005) proposed a method of functional linear analysis whose estimator is consistent and asymptotically normal. Mas (2007) discussed the conditions of weak convergence in functional autoregressive model.

At the best of our knowledge, there are few researches employing FDA to analyze factorized return. One close study by Zhang (2011) is to estimate market beta in functional CAPM. It differs from our aim, however, in that intraday return curve is used instead of constructing a functional object of cross-sectional returns. The difficulty in the latter application is absence of natural topological structure of a market. Our methodology part is then responsible for designing an algorithm of functionalizing cross-sectional observations.

5.2 Methodology

5.2.1 Inference for functional linear dependence

The primary concern of this part is to design a robust estimator and make inference on cross-sectional correlation between stock return $r = (r_1, \cdots, r_N)^T$ and factor $f = (f_1, \cdots, f_N)^T$. For simplicity, it is assumed that both r and f are symmetrically distributed and standardized with zero mean and unit standard deviation. The null hypothesis is

$$H_0 : \rho = \frac{E\pi(r)^T \pi(f)}{N} = 0 \tag{5.1}$$

Where π is some kennel function reserving monotonicity. It can take either simple linear form or nonlinear ones that incorporate local information and are robust to outliers (Devlin et al. (1975)). One common version of the former is the sample correlation coefficient:

$$\rho_c = \frac{r^T f}{N} \tag{5.2}$$

Which is proved to the maximum likelihood estimator under normal distribution. However, it loses efficiency in case that normality is violated and leads to contaminated estimation of correlation. One can penalize observations with large deviations from the mean (Pitas and Venetsanopoulos (1990)) or even trim tails by introducing Huber bounded function $H(z, k) = max\{-k, min(z, k)\}$. Other two popular methods are based on rank statistics. One is Spearman rank correlation (Spearman (1904)) coefficient

$$\rho_c = 1 - \frac{6\sum_{N}^{i=1}(R(r_i) - R(f_i))^2}{N^3 - N} \tag{5.3}$$

For a set of observations $\{x_i, i = 1, \cdots, N\}$, rank operator is,

$$R(x_i) = \sum_{j=1}^{N} \mathbb{1}(x_j \leq x_i) \tag{5.4}$$

where $\mathbb{1}$ is the identity function. Another is Kendall coefficient τ (Kendall and Gibbons (1990)), serving as a measure of concordance. Where sgn is sign function. Asymptotics of them can refer to Kendall and Gibbons (1990), Hjek et al. (1967).

In the context of forecasting, the interest is whether the effect of lagged factors on current returns is significant, which has important implication for efficiency of the predictors. It is easy to show that any market-neutral[4] portfolio w, perfectly concordant with the ranks of factors, delivers positive expected return.

$$\{x_i, i = 1, \cdots, N\} \;\; if H_0 \; is \; rejected \; and \; \tau(w, f) = 1 \tag{5.5}$$

The intuition is that market anomalies with predictability can always be exploited by a family of allocation schemes different from nave strategy (equal weights). But assertion of the inverse relation is not necessarily true. Even if a profitable portfolio, with functional $w(f)$ is present, predictability of f cannot be assured.

[4] $1^T w = 0$

The correlation measures discussed previously implicitly require that the number of stocks equal to that of factors. This assumption may not hold when all stocks in a market are accounted for, due to frequent listing and delisting. This problem also exists in the scenario that only constituents of an index are investigated. The index is rebalanced frequently to maintain its representativeness, but this additional adjustment exaggerates difficulty in correlation estimation. Generally speaking, because traditional methods require homogeneous data generating process, they suffer from a highly changing sample.

Another disadvantage is the poor performance with time-varying correlation. Consider a modified gross error model Tukey (1960) in which the joint distribution of (r_t, f_t)at time t conditioning on an unobserved process ρ_t is given by

$$\psi(r_t, f_t|\rho_t) = N(r_t, f_t; 0, 1 + k\mathbb{1}(\rho_t > 0), \rho_t) \tag{5.6}$$

Here let $k \gg 1$. The probability density function is a bivariate normal distribution with zeros mean, asymmetric volatility $1 + k(\mathbb{1}\rho_t > 0)$ and correlation structure ρ_t. We also assume $\rho_t = -\rho_{t-1}$ so that r_t and f_t behave exactly the opposite from the previous period. It is clear to verify that H_0 cannot be rejected for any of the traditional measures, albeit an evolution of ρ_t is indeed present. In consideration of two problems, we propose a functional method aiming at testing existence of dependence and filtering the path of ρ_t simultaneously. The procedure for univariate case is simple.

U.1 Locate stocks in descendent/ascendant order based on their fac-

tors and form a cross-sectional return accordingly.

The purpose of the step is to construct a return series that is functionalizable. In the next section, we will show an alternative that uses time-serial correlation. As long as a stock has an observation of factor, it always has an unambiguous location. It can even go delisted when return is no longer available because smoothing fills the missing data with local information (Ramsay and Silverman (2005)). One remarkable feature is that stocks are not geographically static but dependent on the value of factor. If price behaves similarly with specific rankings, this sorting algorithm offers best arrangement of data.

U.2 Apply functional principle component analysis (FPCA) to the smoothed return curve and obtain all basis functions $\Psi = \{\psi_i | E(\xi_i) \neq 0\}$**. Null hypothesis of functional linear test**

$$H_0 : \Psi = \emptyset \tag{5.7}$$

Specifically, define trend functions $\Psi_0 = \{\psi_0 | \, |\tau(\psi_0, f)| = 1\}$**. Null hypothesis of linear functional dependence test:**

$$H_0{}' : E(\xi_0) = 0 \tag{5.8}$$

Because of orthonomality of basis functions, there are only two trend components with difference up to sign. By definition, they are in either perfect concordance or complete disorder. Rejection of null is equivalent to assert $\Psi_0 \subseteq \Psi$. It is obvious that expected score is not zero even in case of gross error model since both scenarios give positive contributions.

The advantages of this procedure are two folds. It extracts continuous (nth order differentiable) relations with a set of pre-specified continuous functions, thus retains higher persistence to cross-sectional noise. Moreover, the time series of scores can be utilized for forecasting predictability of factors. This topic will be covered in section 5.5. As for multi factor model, the strategy needs additional dimension reduction.

M.1 Use PCA to find a set of linear combinations $\{\tilde{f}^1, \cdots, \tilde{f}^K\}$ of those factors that accounts for most variation.

M.2 Employ univariate analysis for each principal component $\tilde{f}^i$. Since $\{\tilde{f}^1, \cdots, \tilde{f}^K\}$ are orthogonal, inference in step U.2 for $\tilde{f}^i$ is not affected by other factors. It avoids infeasible multivariate test that has to design sorting algorithm for all factors. Its additional elegance is the capability of processing large set of factors exhibiting highly correlated structure.

5.2.2 Functional dependence forecast model

As discussed in 5.2.1, cross-sectional return curve has at least one significant component. After applying FPCA, $PC1 = \psi_1$; $PC2 = \psi_2$; $PC3 = \psi_3$. As long as functional linear dependence test implies that expected values of components ψ_i^* deviate from zero (Equation (5.7)), we assert that it has predictability for return at next period in the form of

$$\hat{r}(t+1) = \sum_{i=1}^{K} \xi_{i,t+1} \psi_i^*(t+1) \tag{5.9}$$

However, the above equation is not feasible unless estimates of scores $\xi_{i,t+1}$ can be attained. Dependence of scores can be weakly tested by adopting Kokoszka and Reimherr (2013) procedure. The limiting distribution of signs is normal with null hypothesis:

$$H_0 : \Lambda^{(k)} = (N-1)^{\frac{1}{2}} \sum_{i-1}^{N-1} I_{N,n}^{(k)} I_{N,n+1}^{k} \xrightarrow{d} N(0,1) \tag{5.10}$$

where $I_{N,n}^{(k)} = sign(\hat{\xi})$. Positive (negative) curve shape persistence is anticipated if H_0 is rejected in right (left) tail. Assuming scores follow AR(L) process, which reads

$$\hat{\xi}_{i,t+1} = \sum_{j=0}^{L} \gamma_i \xi_{i,t-j} \tag{5.11}$$

Then, equation (5.9) can be estimated as,

$$\hat{r}(t+1) = \sum_{i=1}^{K} \hat{\xi}_{i,t+1} \psi_i^*(t+1) \tag{5.12}$$

Nevertheless, it is still necessary to prove the $\hat{r}(t+1)$ is an unbiased estimator of $r(t+1)$. In Hilbert space H, we know that $r(t) = \sum_i^\infty \langle r_t, \psi_i \rangle \psi_i(t)$, where the operator $\langle\rangle$ is the inner product so that $\langle r_t, \psi_i \rangle$ is the projection of r_t and ψ_i, i.e. score of basis function i at time t. Then, we have $\langle r_t, \psi_i \rangle = \int_0^1 r(t)\psi_i(t) \cdot dt$, for $t \in [0,1]$. Note that because $r(t)$ is observed discretely in the real world,

$\langle r_t, \psi_i \rangle$ may be approximated as,

$$\begin{aligned}\langle r_t, \psi_i \rangle &= \int_0^1 r(t)\psi_i(t) \cdot dt \\ &\approx \sum_{j=1}^{K} \frac{1}{t_j - t_{j-1}} \cdot \frac{1}{2} \cdot (x(t_j)\psi_i(t_j) + x(t_{j-1}\psi_i(t_{j-1}))\end{aligned} \tag{5.13}$$

Which is simply inner product of two vectors. Now assuming $\{r_t\}$ is a stationary functional process, functional curve $r_{(t+1)}$ can be written as following,

$$r(t+1) = \sum_{i=1}^{\infty} \langle r_{t+1}, \psi_i \rangle \psi_i(t+1) \tag{5.14}$$

If finite number of basis is sufficient for explaining it (where FPCA applies) with expected residual vanishing, rearranging equation (5.14) leads to

$$\begin{aligned}r(t+1) &= \sum_{i=1}^{K} \langle r_{t+1}, \psi_i \rangle \psi_i + \sum_{i=K+1}^{\infty} \langle r_{t+1}, \psi_i \rangle \psi_i \\ &= \tilde{r}(t+1) + \varepsilon(t+1)\end{aligned} \tag{5.15}$$

where $E(\varepsilon_t) = 0$. We can have the expression below from Equation (5.12),

$$\begin{aligned}\hat{r}(t+1) &= \sum_{i=1}^{K} \hat{\xi}_{i,t+1} \psi_i(t+1) \\ &= \sum_{i=1}^{K} \sum_{j=0}^{L} \beta_j \xi_{i,t-j} \psi_i(t+1) \\ &= \sum_{i=1}^{K} (\sum_{j=0}^{L} \beta_j \langle r_{t-j}, \psi_i \rangle) \psi_i(t+1)\end{aligned} \tag{5.16}$$

The difference between equation (5.15) and equation (5.12) is given by,

$$\begin{aligned}r(t+1) - \hat{r}(t+1) &= \tilde{r}(t+1) + \varepsilon_t - \tilde{r}(t+1) \\ &= \sum_{i=1}^{K} \langle r_{t+1}, \psi_i \rangle \psi_i - \sum_{i=1}^{K} (\sum_{j=0}^{L} \beta_j \langle r_{t-j}, \psi_i \rangle) \psi_i + \varepsilon_t \\ &= \sum_{i=1}^{k} [\langle r_{t+1}, \psi_i \rangle - \sum_{j=0}^{L} \beta_j \langle r_{t-j}, \psi_i \rangle] \psi_i + \varepsilon_t\end{aligned} \tag{5.17}$$

Thus, it is required that the condition below should be satisfied for all i,

$$\langle r_{t+1}, \psi_i \rangle = \sum_{j=0}^{L} \beta_j \langle r_{t-j}, \psi_i \rangle \tag{5.18}$$

Note that the equation (5.11) merely demonstrates linear relation without basis functions. Therefore, the estimator $\hat{r}_{t+1}$ will be an unbiased estimator for r_{t+1}, which means any unbiased time-serial estimator of score is valid.

5.2.3 Constructing Portfolio with Basis Functions

A portfolio policy allocates capital across return object aiming at gaining positive overall performance, thus return on a policy r_t^p is essentially the inner product of portfolio weights $w(t)$ and cross-sectional return $r(t)$.

$$r_t^p = \langle w(t), r(t) \rangle \tag{5.19}$$

Practically, investment decision is made before return is realized. Following expansion (5.15)

$$r_t^p = \langle w(t), \mu_t + \sum_{i=1}^{K} \xi_i \psi_i + \varepsilon(t) \rangle \tag{5.20}$$

Randomness is stem from the risk of contributions from each basis function. Thus functional principal component analysis is regarded as a calibration of pattern uncertainty. The error term $\varepsilon(t)$ has zero correlation with basis function i.e. $E\langle \psi_i, \varepsilon(t) \rangle = 0$.

One should notice that there exists calibration error as $\varepsilon(t)$ is not determinstic. By standard context of functional PCA, it assumes that

$$Var(\langle \psi_i, r_t \rangle) = \lambda_i \tag{5.21}$$

Consider static basis portfolio linear with some basis function, which is

$$w_i = k\psi_i = sign(E\langle \psi_i, \mu_t \rangle)\psi \tag{5.22}$$

If $\int \psi_i ds = 0$, it ensures that our portfolio is a market-neutral strategy . By letting $r_t = \mu_t + \sum_{j=1}^{K} \langle r_t, \psi_j \rangle \psi_j + \varepsilon(t)$ and $\mu_t \neq 0$, its expected return is

$$\begin{aligned} Er_t^i &= E\langle k\psi_i, \mu_t + \sum_{j=1}^{K} \xi_i \psi_j + \varepsilon(t) \rangle \\ &= k(E\langle \psi_i, \mu_t \rangle + E\sum_{j=1}^{K} \langle \psi_i, \xi_i \psi_j \rangle + E\langle \psi_i, \varepsilon(t) \rangle) \\ &= k(E\langle \psi_i, \mu_t \rangle + E\xi_i) \\ &= kE\langle \psi_i, \mu_t \rangle = |E\langle \psi_i, \mu_t \rangle| \end{aligned} \tag{5.23}$$

Here we use a general assumption that expected score is zero. We now know that the porfolio with respect to some basis function delivers strictly positve return as long as it is not orthogonal to the functional mean.

So it is correct that portfolio performance can be boosted if some estimator on the sign of score can be designed. Further more, the risk associated with it is

$$Var(r_t^i) = Var(\langle k\psi_i, r_i \rangle) = \lambda_i \tag{5.24}$$

And Sharpe ratio should be

$$SR^i = \frac{(\theta_i - \gamma)}{\lambda_i} \tag{5.25}$$

where θ_i is projection of basis function on functional mean and γ is risk-free ratio. Equation (5.25) tells that one threshold the score needs exceeding to guarantee positive Sharpe ratio is γ. It is also affected by the covariance structure of basis fucntion and the noise remaining unexplained. If those return curves are not independent, it allows an unbiased estimator $\hat{\xi}_i$ and cofficient $\hat{k}_t = sign(E\langle\psi_i, \mu_t\rangle) + \hat{\xi}_t$, equation (5.23) becomes

$$E\tilde{r}_t^i = (E\hat{k}_t\langle\psi_i, \mu_t\rangle + E\hat{k}_t\xi_i) = |E\langle\psi_i, \mu_t\rangle| + E\xi_i^2 > Er_t^i \qquad (5.26)$$

One may concern a mixed approaches consisted of basis portfolios. By equation (5.23), excess return $Er_t^b = (|E\langle\psi_1, \mu_t\rangle|, ..., |E\langle\psi_K, \mu_t\rangle|)^{'} - \gamma\mathbf{1}$. Without losing generality, we assume $k_i = sign(|E\langle\psi_i, \mu_t\rangle|) = 1$. Since $E\langle\psi_i, \mu_t\rangle$ doesn't contribute to variability of return curve, covariance structure is

$$C_b = (r_t^b - Er_t^b)^{'}(r_t^b - Er_t^b) = E\xi^{'}\xi = C_\xi \qquad (5.27)$$

Where $\xi = (\xi_1, ..., \xi_K)^{'}$ is score vector. It is thus clear that two basis portfolios are intercorrelated with scores. In the framework of mean-variance optimization, consider maximizing a linear utility function with loss aversion χ and weights β,

$$L = \beta' Er_t^b - \chi\beta' C_b\beta \qquad (5.28)$$

First-order differential with repect to β gives

$$\beta = \frac{1}{\chi}C_b^{-1}Er_t^b \qquad (5.29)$$

When building up a fractional Kelly system on the set of basis portfolios, the above equation can also be derived with χ the proportion invested in Kelly portfolio. It degenerates to time-series momentum in case of zeros correlations among all return series. To obtain (5.29), We may express equity dynamics of basis portfolio using stochastic process drive by M wiener processes $W =$

$(W_1, ..., W_M)'$, then exhibit

$$\frac{dS_t}{S_t} = \theta dt + \Pi dW \tag{5.30}$$

Where $S_t = (S_{1,t}, ..., S'_{K,t})$ is equity vector of K basis portfolios at time t. θ is drift vector defined in equation (5.25) which has the same dimension of K. Π is a $K \times M$ matrix, capturing the effect of different sources of uncertainty. considering a portfolio policy by investing β in the risky asset pool and left for earning risk-free rate γ, equity of our portfolio thus follows the SDE

$$\begin{aligned} \frac{dA_t}{A_t} &= (1 - \beta'\mathbf{1})\gamma dt + \beta'(\theta dt + \Pi dW) \\ &= (\gamma - \beta'\mathbf{1}\gamma + \beta'\theta)dt + \beta'\Pi dW \end{aligned} \tag{5.31}$$

Apply Ito formula and it is easy to see that $K \times K$ covariance matrix $\Pi'\Pi = C_\xi$

$$dlog(A_t) = (\gamma - \beta'\mathbf{1}\gamma + \beta'\theta - \frac{1}{2}\beta' C_\xi \beta)dt + \beta'\Pi dW \tag{5.32}$$

The spirit of Kelly system is then maximizing the drift term, whose first order condition offers (5.29)

$$\frac{\partial D}{\partial \beta} = -\gamma\mathbf{1} + \theta - C_\xi\beta = Er_t^b - C_\xi\beta = 0 \tag{5.33}$$

5.3 Experiment

5.3.1 Implementation Procedure

To apply the method of functional data, we conduct an experiment on the predictability of momentum effect on return in Chinese stock market. The returns in each period are sorted by the factor. The number of stocks varies, thus those cross-sectional objects cannot be directly fed to FPCA without additional restrictions being placed. Our advice in this case is to categorize sorted returns and average the ones within the same group. In this simulation, the number of groups is always 100, so dimension of aggregated ones does not change.

Denote n_i the number of returns in group i and $\bar{n}$ the maximum of $\{n_i, i = 1, ..., 100\}$. The categorization can be uniquely determined for each return using the following rule:

1. $n_i \geq n_j$ if $i \leq j$

2. $\bar{n} - n_i \leq 1, \forall i$

The additional advantage of this approach is reducing noise potentially harmful to functionalization while retaining local behavior. It is also consistent with FF methods but more information is reserved.

The functionalized object is smoothed by conventional B-splines. FPCA is then employed to obtain major components and their time-varying scores. Analysis on functional mean, predictability of scores and their implications on momentum portfolio can be followed.

5.3.2 Data Sample

Daily prices of all stocks listed in Shanghai and Shenzhen Stock Exchanges are collected from Jan. 2006 to May 2014. They are adjusted for stock split, which is frequent, and cash dividends. Return is the difference of log prices.

The data are sufficiently representative in the sense of including a booming economy (2006-2007), a recession (2008), slow recovery (2009-2010) and reconstructure (2011-2014). In anticipation of excessive irrational behavior and faster price correction in emerging market, weekly and monthly frequencies are chosen in Chinese stock market.

Another feature in Chinese market is that preferred stocks were not permitted until recently. The amount of issuance is still negligible. Due to complicated shareholder structure of state-owned company, free float market capitalization is preferable to market value. The stock prices are then adjusted to the sequence of share arrangements since listing.

5.3.3 Data Reformat

In consideration of the 6 million records stored in relational database, we use a *C#* programmable platform to retrieve stock prices and financial data from SQL engine and adapt them in a format for feeding FDA package. It not only serves as interaction between database and analytics module, but also offers a simulator for simple portfolio strategies.

The first component consists of stored procedures and functions that repackage those enquiries. Data can be distributed to either a sorting algorithm or convertor. Structure for original records inherits *C#* comparer, thus is sortable in generic classes. Convertor is a virtual template that is inherited by smoother, aggregator. The former sends a signal to its successor whenever new observation is available and the length of data array exceeds pre-specified parameter. Aggregator can convert daily prices to low-frequency data with consistent structure. The output is formatted in text.

As for performance simulation, weighting schemes based on ranks are introduced after sorting. If market-neutral strategy is adopted, a rank-dependent method is constructed. We also allow quintile portfolios which assign equal weights to stocks in specific quantile (10%-20% for instance) and zero to others. The asset allocation policy is then passed to simulator for performance calculation.

Since data flow is a sequence of records instead of matrix, the framework is streaming processing architecture that is also extendable in paralleling context and complex event processing (CEP). The figure 2 illustrates class dependencies of the project.

5.3.4 Cross-sectional Momentum

Our primary interest, in this section, is to test predictability of return using historical performance. However unlike time-series momentum capturing time-serial persistence in individual futures, the portfolio to be constructed aims at exploiting a functional anomaly. More specifically, momentum factor which is the deviation from average growth.

$$M_{t+1} = R_t - \bar{R}_t$$

could be functionally correlated with return at time $t + 1$, therefore generate a significant nonzero return with basis portfolios. R_t is the average return over a pre-specified look-back period h. In our simulation, it is chosen to be one month (4 weeks), two months (8 weeks), one quarter (12 weeks), half year (26 weeks) and one yer (12 months).

5.4 Empirical Results

5.4.1 FPC of Return Curve

Figure 3-7 illustrate principal functionals with various look-back period. The first component, accounting for on average 60% of the total variation, is linear. It indicates that behavior of the return object is dominated by either conventional momentum or mean-reversion. This result can also serve as an advocate of Fama-French factor portfolio method. However there is room for improvement as long as other components are not negligible. Additionally, profitability of the linear exposure on momentum is affected by contribution (score) of the first component of FPCA to return functional, as suggested by equation (5.23).

The quadratic component gives roughly 15% of variation. It suggests a pattern that stocks at two tails diverge from ones with average past performance. Not as obvious as the linear component, we have an absence of behavioral explanation of second component in FPCA. It states that it is profitable to construct a spread between 'mediocre' stocks and extreme ones (both exceptional and poor). If buy-winners-sell-losers is favor for first moment of past performance, we may conjecture that the second strategy is simply a reflection of risk attitude. To explore a possible relation, it might be good to start from descriptive statistics of each group sorted by momentum factor. The Figure 1 offers volatility with different look-back periods.

X-axis is the index of groups from the highest momentum to the lowest. We found that the shape is consistent regardless of look-back period. It seems to be quite similar with second basis function. Thus we may assert that the second basis function, is a measure of investor's risk attitude. If market is currently highly risk-averse, the score of the basis should be negative and vice versa. This is an important implication on selection of leading indicators for score.

A natural interpretation of the scores on other basis functions, following the previous path, is risk measure of higher moments. Since the left has a contribution of typically less then 5% and is almost orthogonal to functional mean, our discussion is restricted within the first two components.

One worthnoting feature is that the means of all basis functions are not zeros. It tilts toward groups with low momentum. This phenomena, albeit marginal, leads to a profitable mean-reversion strategy. Our assertion is consistent with previous reseaches.

5.5 Portfolio Performance

We propose three portfolio methods in regard with feasibility. The rank-dependent portfolio assigns weights perfectly concordant with one basis function. If capital invested in each group is proportional to the relative location in the curve, it reflects the first basis function. This method can be discretized and formulated

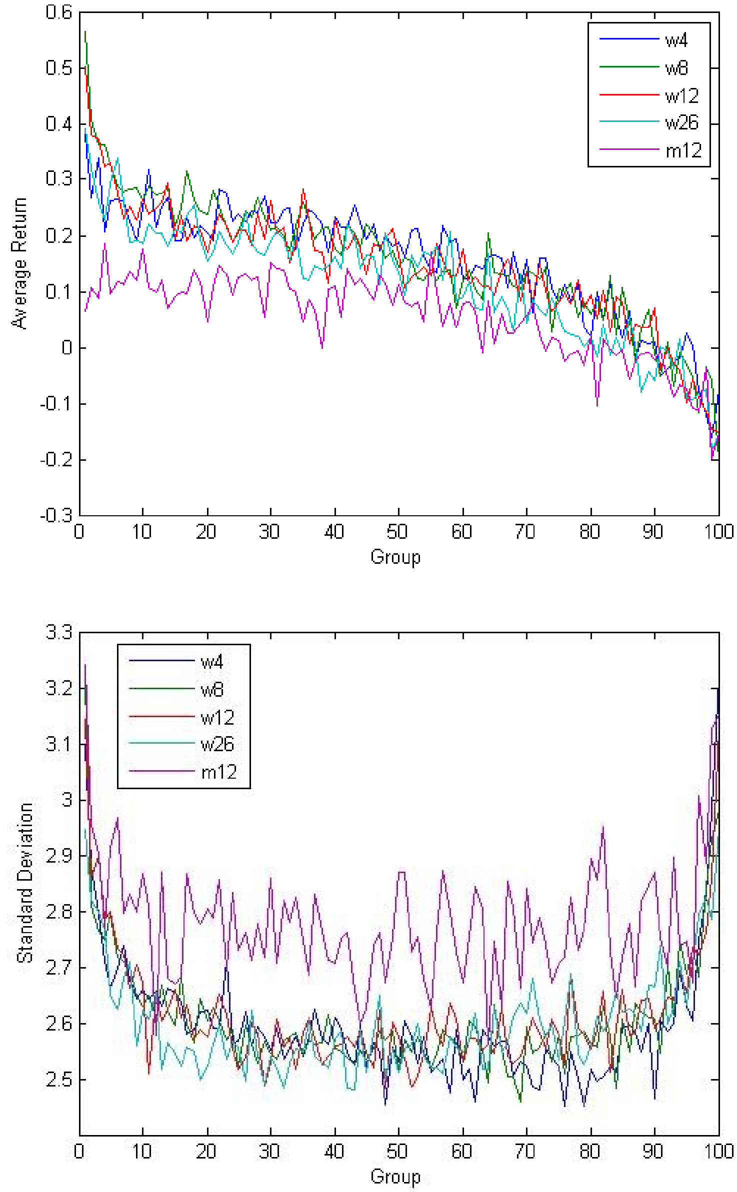


Figure 5.1: The First and Second Moments of Returns on Each Group Sorted by Momentum Factor

as

$$w_i = \nu(R_i - \bar{R}) = \nu(\sum_{j=1}^{100} \mathbf{1}(M_j \leq Mi) - 50.5) \tag{5.34}$$

which is essentially inherited from Fama-French. R_i is the rank of momentum factor with average of $\bar{R} = 50.5$. ν is a scaling factor to ensure constant leverage.

The second component exhibits symmetry in two tails. We may approximate it with a quadratic function. Again define ν for standardization.

$$w_i = \nu(R_i - \bar{R})^2 \tag{5.35}$$

Mixtures of them are also considered to investigate diversification of benefit. In our experiment, the proportion of each basis portfolio depends on its explanatory power. [5]

The left panel in Table 10 summarizes performance on those portfolios. Return on the first component is uniformly higher because it has a larger projection over the mean. Mean-reversion strategy has positive profit for all 5 look-back period, suggesting a prevailing pattern in Chinese market. Consistent with theory, mixed portfolio delivers higher Sharpe ratio than any of two basis portfolios. Maximum drawdown is defined as the largest spread between current equity and its historical maximum

Practically, to avoid excessive turnover, extreme spread approach approximates basis portfolio by retaining those groups that exhibit highest deviations. Specifically only 10 groups with highest predicted returns are used to construct long

[5] It is optimal when scores between two basis are not correlated.

portfolio and those 10 with poorest performance for short one. In case of weights (5.34), 10 groups at two tails are selected. Similarly, we choose 5 groups on two sides and 10 in the middle to replace equation (5.35).

The results in middle panel shows higher return and maximum drawdown compared with rank-dependent portfolios. Risk adjusted returns on the second basis portfolios deteriorate with shrinking underlying assets, but not the first basis. Maximum drawdown can be reduced by mixing two basis portfolios. Some extent of diversification is still present with longer look-back period.

In consideration of short constraints in China, we sell index futures instead of individual stocks in index-hedged method. Since systemic risk is removed by short the same amount of index, it is essentially a market-neutral strategy exploiting alphas of long portfolio. Lower risk-adjusted returns are documented with significant increase of maximum drawdown. The second basis portfolio no longer offers diversification.

5.5.1 Predicting Scores

If return on momentum or mean-reversion exhibits persistence, the autocorrelation can be exploited to enhance performance. In our framework, it is to say that scores are time serially dependent and follow some AR(p) processes. Following Kokoszka and Reimherr (2013), statistics for detecting weak dependency of sign of scores are summarized in Table 1.

We cannot reject null hypothesis at 10% in scores on the first two components.

Table 5.1: $p-$value of Dependency Test

	C1	C2	C3	C4
w4	0.3798	0.4945	0.0637	0.1432
w8	0.4328	0.3775	0.0499	0.4328
w12	0.4305	0.5544	0.6222	0.0612
w26	0.4828	0.3161	0.6884	0.1087
m12	0.7449	0.3290	0.5876	0.4477

Although some are significant in third and fourth basis, the pattern is still sensitive to look-back period. The test suggests that cross-sectional return are independent functional curve.

Daniel and Moskowitz (2012) found that momentum crashes are related with the sign of index return for past one year. The mechanism is those firms experiencing large value depreciation are more likely to outperform due to their high betas. Its counterpart is that market state has predictability on the scores of the first component, because the variation of basis portfolio is always sourced from the dynamics of scores.

To explore this relation, we conduct a set of simple regression of scores on the sign and value of CSI300 index return, which is smoothed either annually or semi-annually. Results are compiled in Table 2-5. We found that the indicator is no longer effficient except for 4 weeks mean-reversion. It means that Chinese investors do not alter their strategy respect to overall market performance.

A natural extension is to test the dependence on higher order moments of market return. Table 6 and 7 summarize regressions on ex ante volatility which is defined as the exponential average of squared deviation from mean.

$$\sigma_t^2 = \alpha(r_t - \bar{r}_t)^2 + (1 - \alpha)\sigma^2 \tag{5.36}$$

where $\bar{r}_t = \alpha r_t + (1 - \alpha) r_t$ is the running exponential average return.

Table 8 and 9 summarize regression on realized kurtosis with similar definition

$$\kappa_t^4 = \alpha(r_t - \bar{r}_t)^4 + (1 - \alpha)\kappa^4 \tag{5.37}$$

It seems that higher order moment has a better prediction than simple average. Interestingly, the score of linear basis functions is always signicantly dependent on kurtosis. Tail risk is, in this sense, priced by those participants whose investment decisions are made on relative historical performance of individual stocks.

5.6 Concluding Remarks

In this section, we propose a nonparametric approach to exploring the functional dependence on factors and its portfolio implications. It can capture nonlinear relations which is believed to be important elements of risks. We can always benefit from the portfolio on nonlinear basis functions, as long as they have nonzero projection on functional mean.

The experiment on Chinese stock market shows that linear component accounts for 60% of total variation. Quadratic component plays a significant role in reducing risk of conventional mean-reversion portfolio. But it loses power with shrinking sample and short constraints.

Finally in hope to enhance basis portfolios, several indicators designed are tested in predicting scores. Kurtosis of market return, according to our regression, has a remarkable influence on the linear component, as a result, the performance of mean-reversion strategies.

5.7 Appendix

Table 5.2: Summary of Test on Predicting Scores (Sign of Index Return)

We run a simple regression of sign of index return on score of each component. Index return is the averaging logarithmic return semiannually, which is a sample of 26 weeks in case of weekly observations and 6 months of monthly data. 'wn' column contains the scores filtered from the return functional sorted by n-weeks momentum while 'm12' is for 12-months momentum.

		w4	w8	w12	w26	m12
C1	Intercept	-0.0001	-0.0927	-0.0353	0.066	0.827
		(-0.0003)	(-0.2202)	(-0.0827)	(0.1562)	(0.4650)
	Slope	0.6855	0.4844	0.3893	0.4309	2.1924
		(1.8097)*	(1.1505)	(0.9131)	(1.0191)	(1.2327)
C2	Intercept	-0.0221	0.0503	0.0508	-0.0198	0.0019
		(-0.0996)	(0.2428)	(0.2524)	(-0.1156)	(0.0027)
	Slope	0.4819	0.3041	0.1307	-0.2279	-0.2166
		(2.1751)**	(1.4666)	(0.6493)	(-1.3306)	(-0.3101)
C3	Intercept	-0.0122	-0.006	-0.0041	0.0061	-0.2836
		(-0.0906)	(-0.0545)	(-0.0360)	(0.0561)	(-0.6798)
	Slope	0.152	0.0094	-0.1117	-0.0697	-0.4795
		(1.1292)	(0.0852)	(-0.9779)	(-0.6379)	(-1.1494)
C4	Intercept	0.0381	0.0585	-0.0349	-0.0109	-0.0214
		(0.3656)	(0.5800)	(-0.3538)	(-0.1125)	(-0.0520)
	Slope	-0.2371	-0.0903	0.1042	-0.0429	0.4302
		(-2.2768)**	(-0.8949)	(1.0573)	(-0.4415)	(1.0463)

Table 5.3: Summary of Test on Predicting Scores (Index Return)

We run a simple regression of index return on score of each component. Index return is the averaging logarithmic return semiannually, which is a sample of 26 weeks in case of weekly observations and 6 months of monthly data. 'wn' column contains the scores filtered from the return functional sorted by n-weeks momentum while 'm12' is for 12-months momentum.

		w4	w8	w12	w26	m12
C1	Intercept	-0.1147	-0.1524	-0.0594	0.021	0.6876
		(-0.3014)	(-0.3599)	(-0.1386)	(0.0497)	(0.3876)
	Slope	59.86	26.6	2.2194	5.0182	36.65
		(2.0724)**	(0.8297)	(0.0685)	(0.1567)	(1.0123)
C2	Intercept	-0.1006	0.0249	0.0254	0.0092	0.0113
		(-0.4518)	(0.1192)	(0.1259)	(0.0539)	(0.0162)
	Slope	40.7421	8.4078	13.51	-8.3737	-4.4753
		(2.4115)**	(0.5316)	(0.8848)	(-0.6454)	(-0.3157)
C3	Intercept	-0.02574	-0.0006	0.005	0.0124	-0.2458
		(-0.1898)	(-0.0053)	(0.0437)	(0.1137)	(-0.5899)
	Slope	5.558	-4.0213	-2.2733	0.2953	-6.5993
		(0.5400)	(-0.4781)	(-0.2619)	(0.0357)	(-0.7759)
C4	Intercept	0.0529	0.0525	-0.066	-0.0082	-0.052
		(0.5016)	(0.5183)	(-0.6712)	(-0.0847)	(-0.1268)
	Slope	-4.555	6.7909	18.79	1.4252	-0.3765
		(-0.5692)	(0.8848)	(2.5283)**	(0.1938)	(-0.8751)

Table 5.4: Summary of Test on Predicting Scores (Sign of Index Return)

We run a simple regression of sign of index return on score of each component. Index return is the averaging logarithmic return annually, which is a sample of 52 weeks in case of weekly observations and 12 months of monthly data. 'wn' column contains the scores filtered from the return functional sorted by n-weeks momentum while 'm12' is for 12-months momentum.

		w4	w8	w12	w26	m12
C1	Intercept	-0.0625	-0.2182	-0.165	0.3413	1.5778
		(-0.1577)	(-0.5066)	(-0.3839)	(0.8021)	(0.9208)
	Slope	-0.1392	-0.4905	-0.4541	0.1218	2.7686
		(-0.3510)	(-1.1385)	(-1.0562)	(0.2864)	(1.6157)
C2	Intercept	-0.0733	0.0793	0.162	-0.0213	-0.0176
		(-0.3217)	(0.3680)	(0.7897)	(-0.1232)	(-0.0239)
	Slope	0.4109	0.2719	0.2989	-0.0546	-0.8622
		(1.8037)*	(1.2622)	(1.4571)	(-0.3153)	(-1.1720)
C3	Intercept	-0.0266	0.0187	0.1099	0.0509	-0.4921
		(-0.1963)	(0.1695)	(0.9363)	(0.4477)	(-1.1643)
	Slope	0.1963	-0.0081	0.1183	0.0546	-0.5923
		(1.4487)	(-0.0735)	(1.0081)	(0.4802)	(-1.4012)
C4	Intercept	0.0903	0.08	-0.0501	-0.0146	-0.0409
		(0.8452)	(0.7733)	(-0.5028)	(-0.1480)	(-0.0951)
	Slope	-0.1044	-0.0519	0.133	-0.0801	-0.3765
		(-0.9767)	(-0.5014)	(1.3343)	(-0.8140)	(-0.8751)

Table 5.5: Summary of Test on Predicting Scores (Index Return)

We run a simple regression of index return on score of each component. Index return is the averaging logarithmic return annually, which is a sample of 52 weeks in case of weekly observations and 12 months of monthly data. 'wn' column contains the scores filtered from the return functional sorted by n-weeks momentum while 'm12' is for 12-months momentum.

		w4	w8	w12	w26	m12
C1	Intercept	-0.1046	-0.1524	-0.0785	0.3188	1.9776
		(-0.2644)	(-0.3545)	(-0.1838)	(0.7659)	(1.1614)
	Slope	48.33	0.0033	-24.62	66.55	122.76
		(1.2651)	(0.0001)	(-0.5786)	(1.4883)	(2.2736)**
C2	Intercept	-0.1791	0.0215	0.0814	-0.011	-0.019
		(-0.7877)	(0.1002)	(0.4017)	(-0.0646)	(-0.0255)
	Slope	45.63	20.58	45.46	-12.27	-23.1
		(2.0777)**	(0.9804)	(2.2505)**	(-0.6725)	(-0.9772)
C3	Intercept	-0.0626	0.0196	-0.0801	0.0404	-0.4887
		(-0.4617)	(0.1781)	(0.6883)	(0.3620)	(-1.1387)
	Slope	9.924	0.2029	15.42	2.1716	-15.32
		(0.7572)	(0.0189)	(1.3296)	(0.1811)	(-1.1259)
C4	Intercept	0.1023	0.0872	-0.0805	0.0008	0.0863
		(0.9568)	(0.8466)	(-0.8137)	(0.0085)	(0.1976)
	Slope	0.5902	-0.2891	13.43	-5.4809	5.6312
		(0.0571)	(-0.0287)	(1.3628)	(-0.5278)	(0.4068)

Table 5.6: Summary of Test on Predicting Scores (Volatility)

We run a simple regression of volatility of index return on score of each component. Ex ante volatility is the exponential average of squared exess return on its mean: $\sigma_t^2 = \alpha(r_t - \bar{r}_t)^2 + (1-\alpha)\sigma^2$ where $\alpha = 1/26$. 'wn' column contains the scores filtered from the return functional sorted by n-weeks momentum while 'm12' is for 12-months momentum.

		w4	w8	w12	w26	m12
C1	Intercept	2.813	2.2512	2.1935	2.4059	5.7939
		(2.7147)**	(1.9119)*	(1.7811)	(1.9000)*	(1.2771)
	Slope	-80.28	-66.44	-62.57	-65.39	-77.34
		(-2.9367)**	(-2.1491)**	(-1.9459)*	(-1.9922)*	(-1.2662)
C2	Intercept	1.1881	0.0538	0.7528	0.2476	-0.5947
		(1.9468)*	(0.0921)	(1.2914)	(0.4798)	(-0.3337)
	Slope	-34.71	-0.4675	-19.72	-6.7573	9.187
		(-2.1559)**	(-0.0305)	(-1.2957)	(-0.5052)	(0.3829)
C3	Intercept	-0.8569	-0.3341	-0.165	0.1016	-0.6325
		(-2.3249)**	(-1.0777)	(-0.4975)	(0.3089)	(-0.5898)
	Slope	23.77	9.2044	4.6425	-2.4418	6.1457
		(2.4450)**	(1.1309)	(0.5361)	(-0.2865)	(0.4256)
C4	Intercept	0.2422	0.1291	0.5445	0.2191	-0.5461
		(0.8393)	(0.4552)	(1.9139)*	(0.7499)	(-0.5176)
	Slope	-5.5568	-1.872	-16.27	-6.2087	5.6312
		(-0.7230)	(-0.2515)	(-2.1903)**	(-0.8198)	(0.4068)

Table 5.7: Summary of Test on Predicting Scores (Volatility)

We run a simple regression of volatility of index return on score of each component. Ex ante volatility is the exponential average of squared exess return on its mean: $\sigma_t^2 = \alpha(r_t - \bar{r}_t)^2 + (1-\alpha)\sigma^2$ where $\alpha = 1/52$. 'wn' column contains the scores filtered from the return functional sorted by n-weeks momentum while 'm12' is for 12-months momentum.

		w4	w8	w12	w26	m12
C1	Intercept	3.5667	2.0735	0.9903	2.427	9.0826
		(3.0768)***	(1.5895)	(0.7676)	(1.9011)*	(1.7783)
	Slope	-98.52	-60.22	-29.51	-57.17	-103.18
		(-3.3067)***	(-1.8052)*	(-0.8938)	(-1.7474)	(-1.6790)
C2	Intercept	0.7575	-0.0996	1.3029	0.3155	0.0908
		(1.1189)	(-0.1520)	(2.1232)**	(0.6052)	(0.0410)
	Slope	-24.02	3.853	-32.11	-8.8452	0.9935
		(-1.3808)	(0.2299)	(-2.0451)**	(-0.6620)	(0.0373)
C3	Intercept	-1.3215	-0.5831	0.2674	-0.0898	-0.0132
		(-3.3308)***	(-1.7462)	(0.7587)	(-0.2621)	(-0.0103)
	Slope	34.66	16.31	-4.7412	3.5298	-4.4683
		(3.3994)***	(1.9102)*	(-0.5256)	(0.4018)	(-0.2906)
C4	Intercept	0.6992	0.4237	0.4734	0.2028	-0.4194
		(2.2152)**	(1.3509)	(1.5874)	(0.6836)	(-0.3252)
	Slope	-16.26	-9.1108	-14.72	-5.4736	5.8555
		(-2.0043)**	(-1.1358)	(-1.9288)*	(-0.7197)	(0.3774)

Table 5.8: Summary of Test on Predicting Scores (Kurtosis)

We run a simple regression of volatility of index return on score of each component. Ex ante volatility is the exponential average of squared exess return on its mean: $\kappa_t^4 = \alpha(r_t - \bar{r}_t)^4 + (1-\alpha)\kappa^4$ where $\alpha = 1/26$. 'wn' column contains the scores filtered from the return functional sorted by n-weeks momentum while 'm12' is for 12-months momentum.

		w4	w8	w12	w26	m12
C1	Intercept	1.3499	1.6406	2.121	2.2814	9.5638
		(1.8827)*	(2.0699)**	(2.6543)**	(2.7651)**	(2.9028)***
	Slope	-1.1433	-1.458	-1.82	-2.0102	-1.8813
		(-2.2521)**	(-2.6043)**	(-3.2047)***	(-3.1659)***	(-3.1918)***
C2	Intercept	-0.0354	-0.1658	0.196	0.1162	-0.5686
		(-0.0838)	(-0.4208)	(0.5129)	(0.3428)	(-0.4177)
	Slope	-0.0021	0.1686	-0.1272	-0.1021	0.1251
		(-0.0069)	(0.6060)	(-0.4686)	(-0.3916)	(0.5139)
C3	Intercept	-0.5059	-0.1334	-0.0552	0.1063	-1.0936
		(-1.9905)*	(-0.6370)	(-0.2543)	(0.4925)	(-1.3459)
	Slope	0.4071	0.1055	0.0477	-0.0835	0.183
		(2.2621)**	(0.7133)	(0.3096)	(-0.5023)	(1.2590)
C4	Intercept	0.1978	-0.0009	0.1315	0.243	0.4479
		(0.9953)	(-0.0047)	(0.7031)	(1.2691)	(0.5567)
	Slope	-0.1265	0.0526	-0.1438	-0.2227	5.8555
		(-0.8992)	(0.3893)	(-1.0818)	(-1.5113)	(0.3774)

Table 5.9: Summary of Test on Predicting Scores (Kurtosis)

We run a simple regression of volatility of index return on score of each component. Ex ante volatility is the exponential average of squared exess return on its mean: $\kappa_t^4 = \alpha(r_t - \bar{r}_t)^4 + (1-\alpha)\kappa^4$ where $\alpha = 1/52$. 'wn' column contains the scores filtered from the return functional sorted by n-weeks momentum while 'm12' is for 12-months momentum.

		w4	w8	w12	w26	m12
C1	Intercept	1.5266	1.6058	1.8695	1.8775	11.38
		(2.0389)**	(2.0031)**	(2.2484)**	(2.2892)**	(3.0614)***
	Slope	-1.7083	-1.895	-2.2499	-1.8698	-3.6385
		(-2.4609)**	(-2.5839)**	(-2.7456)**	(-2.2041)**	(-3.0996)***
C2	Intercept	-0.2634	-0.3992	0.1379	0.1029	-1.7773
		(-0.6049)	(-0.9887)	(0.3436)	(0.3064)	(-1.0670)
	Slope	0.1521	0.4764	-0.0227	-0.1363	0.6807
		(0.3768)	(1.2896)	(-0.0573)	(-0.3923)	(1.2943)
C3	Intercept	-0.5536	-0.2084	-0.0656	0.0742	-0.1972
		(-2.1560)**	(-1.0081)	(-0.2863)	(0.3360)	(-0.2029)
	Slope	0.5466	0.246	0.1808	-0.0405	-0.0584
		(2.2961)**	(1.3007)	(0.8005)	(-0.1775)	(-0.1903)
C4	Intercept	0.2657	0.1537	0.207	0.2789	0.4938
		(1.3064)	(0.7927)	(1.0655)	(1.4639)	(0.5041)
	Slope	-0.1768	-0.072	-0.3163	-0.3332	-0.1586
		(-0.9374)	(-0.4056)	(-1.6521)	(-1.6912)	(-0.5127)

Table 5.10: Portfolio Performance on Momentum Factor

The rank-dependent portfolio assigns weights perfectly concordant with one basis function (C1 and C2). Mixtures of them are also considered. Extreme spread approach approximates basis portfolio by retaining those groups that exhibit highest deviations. In consideration of short constraints in China, we sell index futures instead of individual stocks in index-hedged method. Annual return r, sharpe ratio sr and maximum drawdown d_{max} are simulated for each portfolio scheme. No transaction cost is incorporated. r and d_{max} are in percentage terms.

		Rank Dependent			Extreme Spread			Index Hedged		
		C1	C2	Mixed	C1	C2	Mixed	C1	C2	Mixed
r	w4	10.57	3.66	7.11	11.92	4.13	10.02	17.17	9.27	15.19
	w8	12.49	0.77	6.63	20.16	-0.14	10.01	24.86	6.22	20.20
	w12	11.38	1.54	6.46	19.63	1.04	10.34	23.28	7.41	19.31
	w26	11.25	2.06	6.66	16.84	2.99	9.91	20.34	8.83	17.46
	m12	7.03	3.11	5.07	10.55	5.31	7.93	14.30	11.90	13.70
sr	w4	1.60	0.96	1.79	1.60	0.70	1.63	0.75	0.44	0.70
	w8	1.73	0.22	1.62	1.87	-0.02	1.60	1.06	0.30	0.92
	w12	1.56	0.45	1.56	1.78	0.20	1.63	1.00	0.36	0.88
	w26	1.63	0.72	1.75	1.63	0.64	1.72	0.89	0.42	0.80
	m12	1.13	1.20	1.38	1.10	1.26	1.40	0.72	0.68	0.73
d_{max}	w4	-7.29	-7.03	-4.40	-13.88	-12.49	-7.71	-29.90	-48.36	-34.52
	w8	-8.47	-9.55	-4.50	-14.65	-17.73	-7.89	-29.59	-46.73	-29.94
	w12	-8.65	-8.88	-5.40	-13.96	-14.58	-8.22	-31.53	-47.10	-31.43
	w26	-8.37	-5.87	-4.37	-11.94	-9.65	-5.57	-31.93	-44.32	-32.24
	m12	-11.02	-4.82	-6.27	-18.80	-8.10	-10.31	-32.11	-36.22	-33.13

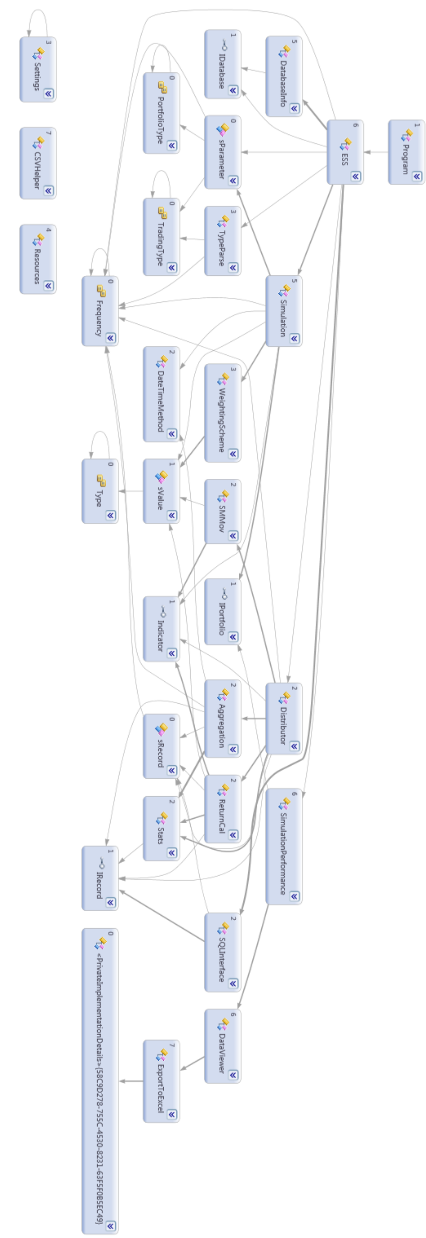


Figure 5.2: Simulation Platform designed for retrieving data from database, sorting stocks based on factors and grouping returns

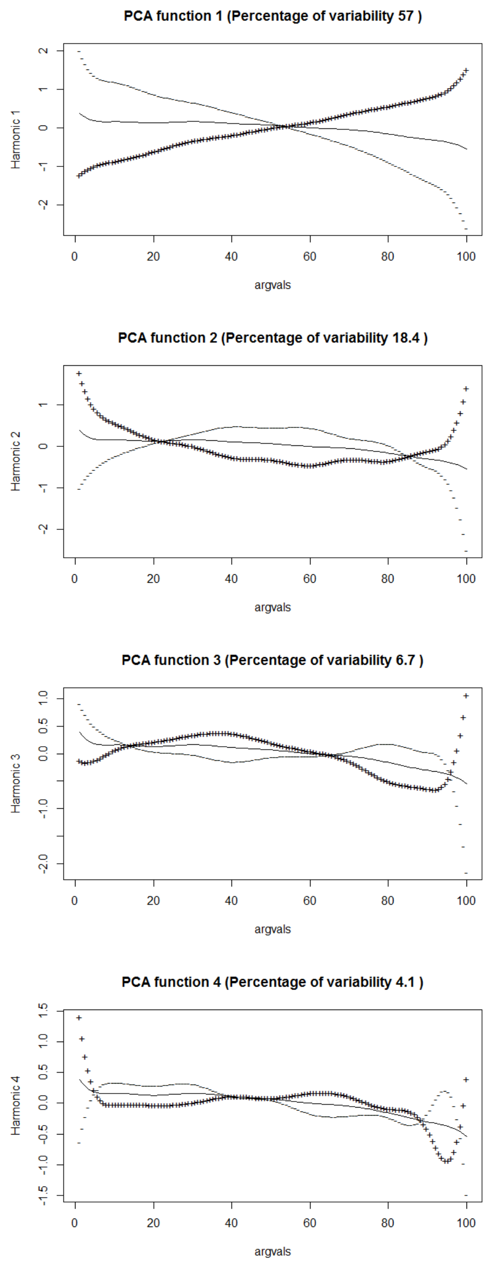


Figure 5.3: Basis Functions derived from FPCA on cross-sectional return curve. Look-back period is one month $h = 4$.

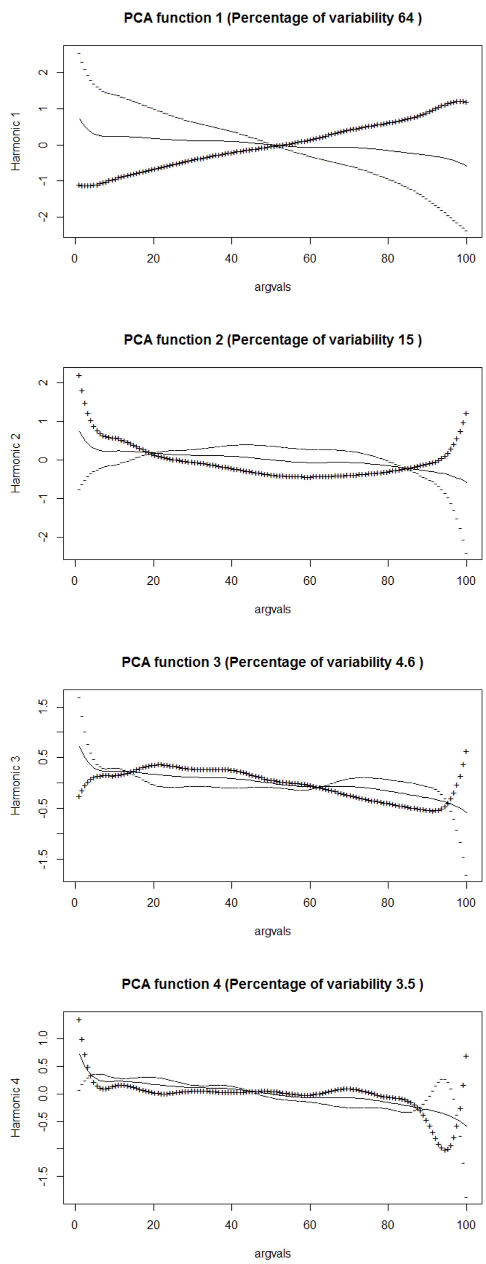


Figure 5.4: Basis Functions derived from FPCA on cross-sectional return curve. Look-back period is two months $h = 8$.

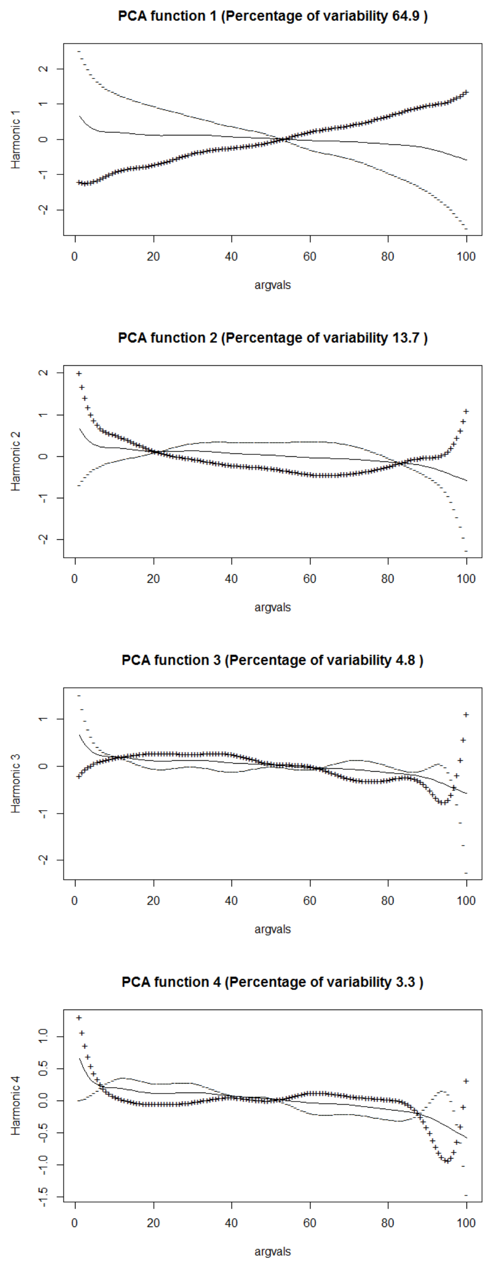


Figure 5.5: Basis Functions derived from FPCA on cross-sectional return curve. Look-back period is one quarter $h = 12$.

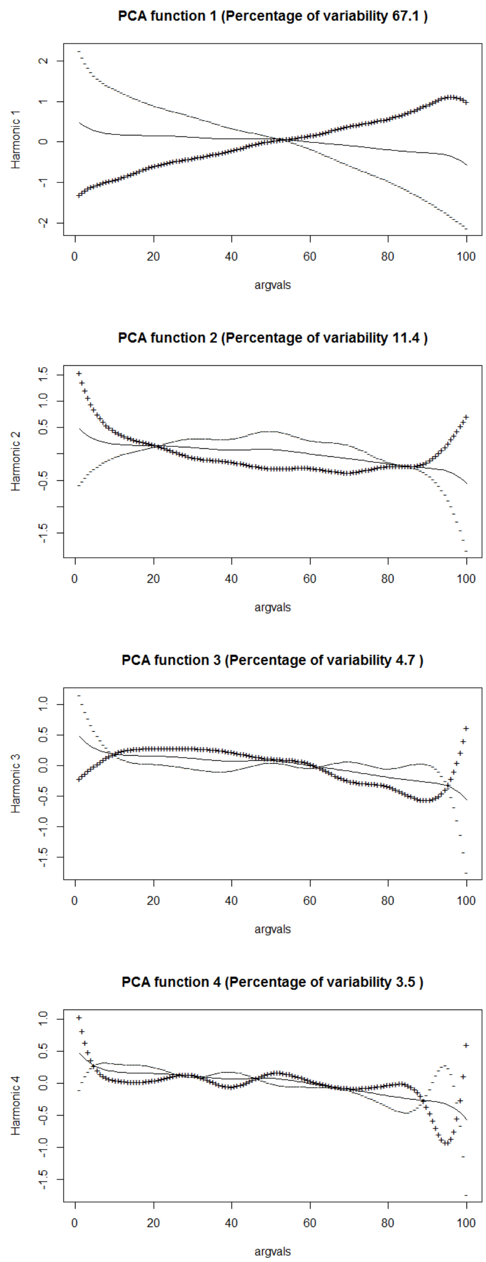


Figure 5.6: Basis Functions derived from FPCA on cross-sectional return curve. Look-back period is half year $h = 26$.

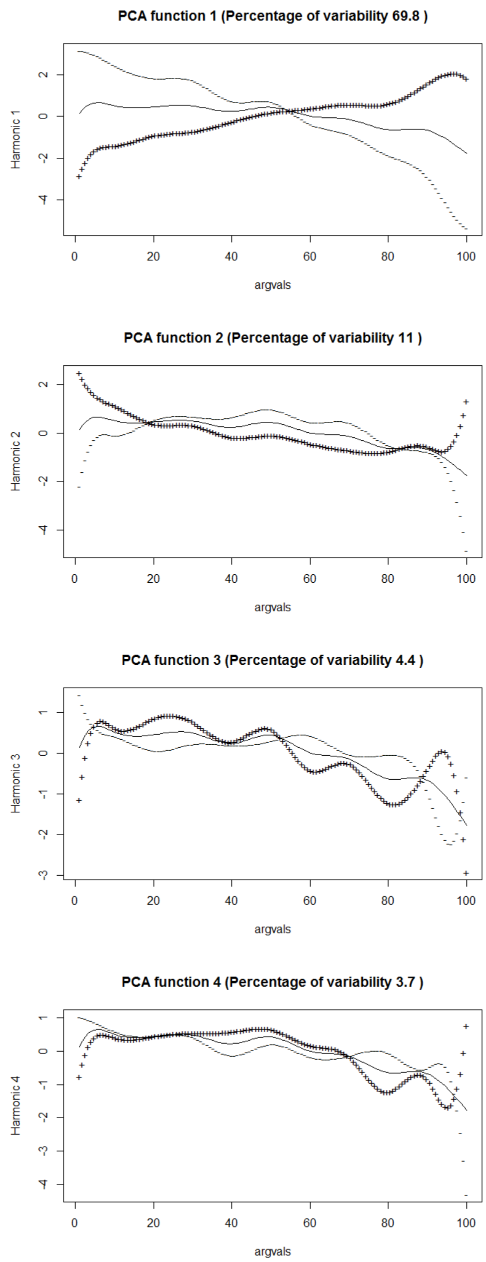


Figure 5.7: Basis Functions derived from FPCA on cross-sectional return curve. Look-back period is one year $h = 52$.

6 Conclusion

The portfolio methods under investigation may be acclaimed in three circumstances. If market participants always make rational investment decisions, modelling expected utility function should be of primary concern. Assumption of CRRA functional leads to a favor for parametric portfolio, which gains additional out-of-sample stability from decomposition of weights. However, selecting factors with predictability requires a robust technique resistant to data-snooping. I offer functional equivalence between predictive ability and profitability of trading strategies. Thus, the bootstrap method enhances efficiency of parameterized optimization by filtering out efficient indicators. An experiment on FT100 shows promising performance of the mechanism.

Alternatively, mean-variance is recommended when input variables follow elliptical distribution. Although the condition is often violated on asset return, many nonlinear trading strategies are able to produce valid training sample before feeding into optimizers. Another problem of plug-in method arises from heterogeneous data generating process. Classifiers can then be applied in gathering similar observations. My study shows that mean-variance can beat naive strategy in constructing a portfolio of 24 commodity futures. But this outperformance is marginal after incorporating transaction cost.

In case that investors exhibit asymmetric aversion of tail risk, Choquet utility maximization and coherent risk measures accquire superiority. It can be shown that the scheme proposed by Koenker (2005) is equivalent to optimization using conditional value-at-risk (Rockafellar and Uraysev, 2000) both in continuous and discrete cases. By decomposing quantile loss function, a new parametric approach is proposed aiming at minimizing weighted downside risk exposure on a set of factors given fixed target return. An empirical study on comparing portfolio efficiency among mean-variance, pessimistic portfolio and its parameterized version in global equity indices investment is then conducted. Pessimistic portfolio offers the lowest return in frictionless environment, it also suffers from overconcentration and non-stationarity of tail risk. Deterioration of out-of-sample robustness can be possibly mitigated by stationary bootstrap method but averaged performance barely beats benchmark. Parametric pessimistic portfolio effectively reduces tail risk exposure and data-snooping bias in training samples and thus delivering a better risk-adjusted return.

Instead of utility maximization, an allocation policy can be formed by gaining exposures on risk factors that are behaviorally priced. Trading factor portfolio implicitly relies on the assumption that cross-sectional stock returns are linearly correlated with market anomalies. I propose a test on functional dependence using FPCA with linearity special basis function. It can be proved that diversification benefit is always present as long as nonlinear components are negligible. An experiment shows that momentum/mean-reversion accounts for approximately 60% of return variation while quadratic function contributes to 15% of that. Time series of scores, however, does not follow AR(p) process. Finally, it is found that profitability of buy-winners-sell-losers is affected by kurtosis of index return because investors require premium for extreme events.

The bottom line of this research is that a portfolio choice only succeeds within specific market conditions. Sustainable performance can be achieved by an appropriate model on investor behavior and robust in-sample optimization techniques.

References

Abraham, C., Cornillon, P. A., Matzner-L?ber, E. R. I. C., and Molinari, N. (2003). Unsupervised curve clustering using b-splines. *Scandinavian journal of statistics*, 30(3):581–595.

Ait-Sahalia, Y. and Brandt, M. (1994). Variable selection for portfolio choice. *Journal of Finance*, 56(4):1297–1351.

Ali, A., Hwang, L. S., and Trombley, M. A. (2001). Arbitrage risk and the book-to-market anomaly. *Journal of Financial Economics*, 69(2):355–373.

Amaya, D., Christoffersen, P., Jacobs, K., and Vasquez, A. (2011). Do realized skewness and kurtosis predict the cross-section of equity returns. *School of Economics and Management, University of Aarhus. No. 2011-44.*

Ammann, M., S, O., and D., O. (2012). An alternative three-factor model for international markets: Evidence from the european monetary union. *Journal of Banking & Finance*, 36(7):1857–1864.

Anderson, T. W. (1963). Asymptotic theory for principal component analysis. *Annals of Mathematical Statistics*, pages 122–148.

Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2006). The cross-section of volatility and expected returns. *The Journal of Finance*, 61(1):259–299.

Ang, A., Hodrick, R. J., Xing, Y., and Zhang, X. (2009). High idiosyncratic volatility and low returns: International and further us evidence. *Journal of Financial Economics*, 91(1):1–23.

Antoch, J., Prchal, L., Rosaria De Rosa, M., and Sarda, P. (2010). Electricity consumption prediction with functional linear regression using spline estimators. *Journal of Applied Statistics*, 37(12):2027–2041.

Apergis, N. and Payne, J. E. (2014). Resurrecting the size effect: Evidence from a panel nonlinear cointegration model for the g7 stock markets. *Review of Financial Economics*, 23(1):46–53.

Artzner, P., Delbaen, J., Eber, M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9:203–228.

Asness, C., Frazzini, A., and Pedersen, L. H. (2013). Quality minus junk. Available at SSRN.

Asness, C. S. (1997). The interaction of value and momentum strategies. *Financial Analysts Journal*, pages 29–36.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of financial economics*, 9(1):3–18.

Bassett, G., Koenker, R., and Kordas, G. (2004). Pessimistic portfolio allocation and choquet expected utility. Available at http://cemmap.ifs.org.uk/wps/cwp0409.pdf.

Bawa, V., Brown, S., and Klein, R. (1979). *Estimation Risk and Optimal Portfolio Choice*. North Holland.

Berk, J. B. (1995). A critique of size-related anomalies. *Review of Financial Studies*, 8(2):275–286.

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

Black, F. (1972). Capital market equilibrium with restricted borrowing. *Journal of business*, pages 444–455.

Black, F. and Litterman, R. (1992). Global portfolio optimization. *Financial Analyst Journal*, 48(5):28–43.

Blitz, D., Huij, J., and Martens, M. (2011). Residual momentum. *Journal of Empirical Finance*, 18(3):506–521.

Bouchaud, J. and Potters, M. (2009). Financial applications of random matrix theory: A short review. working paper.

Brandt, M. and Santa-Clara, P. (2006). Dynamic portfolio selection by augmenting the asset space. *Journal of Finance*, 61:2187–2217.

Brandt, M., Santa-Clara, P., and Valkanov, R. (2009). Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *Society of Financial Studies*, 22(9):3411–3447.

Brock, W., Lakonishok, J., and Lebaron, B. (1992). Simple technical trading rules and the stochastic properties of stock returns. *Journal of Finance*, 47(5):1731–1764.

Cai, T. T. and Hall, P. (2006). Prediction in functional linear regression. *The Annals of Statistics*, 34(5):2159–2179.

Cakici, N., Fabozzi, F. J., and Tan, S. (2013). Size, value, and momentum in emerging market stock returns. *Emerging Markets Review*, 16:46–65.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance*, 52(1):57–82.

Chan, L. K., Hamao, Y., and Lakonishok, J. (1991). Fundamentals and stock returns in japan. *The Journal of Finance*, 46(5):1739–1764.

Chen, L., Novy-Marx, R., and Zhang, L. (2010). An alternative three-factor model. Washington University in St. Louis, University of Chicago, and University of Michigan.

Claessens, S., Dasgupta, S., and Glen, J. (1995). *The cross-section of stock returns: Evidence from the emerging markets.* World Bank Publications.

Cooper, M. J., Gulen, H., and Schill, M. J. (2008). Asset growth and the cross-section of stock returns. *The Journal of Finance*, 63(4):1609–1651.

Cox, J. C. and Ross, S. A. (1976). The valuation of options for alternative stochastic processes. *Journal of financial economics*, 3(1):145–166.

Daniel, K., Jagannathan, R., and Kim, S. (2012). Tail risk in momentum strategy returns. *National Bureau of Economic Research, No. w18169.*

Daniel, K. and Moskowitz, T. J. (2013). *Momentum Crashes.* University of Geneva.

Devlin, S. J., Gnanadesikan, R., and Kettenring, J. R. (1975). Robust estimation and outlier detection with correlation coefficients. *Biometrika*, 62(3):531–545.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics*, 20(3):339–350.

Eun, C. S. and Huang, W. (2007). Asset pricing in china's domestic stock markets: Is there a logic? *Pacific-Basin Finance Journal*, 15(5):452–480.

Fama, E. F. and French, K. R. (1992). The cross-section of expected stock returns. *the Journal of Finance*, 47(2):427–465.

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.

Fama, E. F. and French, K. R. (1998). Value versus growth: The international evidence. *The Journal of Finance*, 53(6):1975–1999.

Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *The Journal of Political Economy*, pages 607–636.

Frazzini, A. and Pedersen, L. H. (2014). Betting against beta. *Journal of Financial Economics*, 111(1):1–25.

Goia, A., May, C., and Fusai, G. (2010). Functional clustering and linear regression for peak load forecasting. *International Journal of Forecasting*, 26(4):700–711.

Grinblatt, M. and Moskowitz, T. J. (2004). Predicting stock price movements from past returns: The role of consistency and tax-loss selling. *Journal of Financial Economics*, 71(3):541–579.

Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):109–126.

Hiemstra, C. and Jones, J. D. (1994). Testing for linear and nonlinear granger causality in the stock price-volume relation. *The Journal of Finance*, 49(5):1639–1664.

Hjek, J., ?idk, Z., and Sen, P. K. (1967). *Theory of rank tests*. New York: Academic press.

Horvath, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer Series in Statistics.

Hudson, R., Dempsey, M., and Keasey, K. (1996). A note on the weak form efficiency of capital markets: the application of simple technical trading rules to uk stock prices c 1935-1994. *Journal of Banking and Finance*, 20:1121–1132.

Hyndman, R. J. and Ullah, S. (2007). Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis*, 51(10):4942–4956.

James, G. M. and Hastie, T. J. (2001). Functional linear discriminant analysis for irregularly sampled curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):533–550.

Jaschke, S. and Kuchler, U. (2001). Coherent risk measure and good deal bounds. *Finance and Stochastics*, 5:181–200.

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

Jobson, J. and Korkie, B. (1980). Estimation of markowitz efficient portfolios. *Journal of American Statistical Association*, 75(371):544–554.

Kendall, M. and Gibbons, J. (1990). *Rank Correlation Methods*. Oxford University Press.

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press.

Kokoszka, P. and Reimherr, M. (2013). Predictability of shapes of intraday price curves. *The Econometrics Journal*, 16(3):285–308.

Kraus, A. and Litzenberger, R. H. (1976). Skewness preference and the valuation of risk assets. *The Journal of Finance*, 31(4):1085–1100.

Laloux, L., Cizeau, P., Potters, M., and Bouchaud, J. (2000). Random matrix theory and financial correlations. *International Journal of Theoretical and Applied Finance*, 3(3):391–307.

Lindberg, C. (2009). Portfolio optimization when expected stock returns are determined by exposure to risk. *Bernoulli*, 15(2):464–474.

Lischewski, J. and Voronkova, S. (2012). Size, value and liquidity. do they really matter on an emerging stock market? *Emerging Markets Review*, 13(1):8–25.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, 7(1):77–91.

Markowitz, H. (1959). *Portfolio Selection: Efficient Diversification of Investments*. Yale University Press.

Markowitz, H. (2012). Mean-variance approximations to expected utility. *European Journal of Operational Research*.

Mas, A. (2007). Weak convergence in the functional autoregressive model. *Journal of Multivariate Analysis*, 98(6):1231–1261.

Mestekemper, T., Windmann, M., and Kauermann, G. (2010). Functional hourly forecasting of water temperature. *International Journal of Forecasting*, 26(4):684–699.

Michaud, R. (1989). The markowitz optimization enigma: Is optimized optimal? *Financial Analyst Journal*, pages 31–42.

Narayan, P. K. and Zheng, X. (2010). Market liquidity risk factor and financial market anomalies: Evidence from the chinese stock market. *Pacific-Basin finance journal*, 18(5):509–520.

Nartea, G. V., Wu, J., and Liu, Z. (2013). Does idiosyncratic volatility matter in emerging markets? evidence from china. *Journal of International Financial Markets, Institutions and Money*, 27:137–160.

Naughton, T., Truong, C., and Veeraraghavan, M. (2008). Momentum strategies and stock returns: Chinese evidence. *Pacific-Basin Finance Journal*, 16(4):476–492.

Nicholas Refenes, A., Zapranis, A., and Francis, G. (1994). Stock performance modeling using neural networks: a comparative study with regression models. *Neural Networks*, 7(2):375–388.

Penman, S. H., Richardson, S. A., and Tuna, I. (2007). The book-to-price effect in stock returns: Accounting for leverage. *Journal of Accounting research*, 45(2):427–467.

Piotroski, J. D. (2000). Value investing: The use of historical financial statement information to separate winners from losers. *Journal of Accounting Research*, pages 1–41.

Pitas, I. and Venetsanopoulos, A. N. (1990). *Nonlinear digital filters*, volume 84. Springer.

Pontiff, J. and Woodgate, A. (2008). Share issuance and cross-sectional returns. *The Journal of Finance*, 63(2):921–945.

Qi, M. and Maddala, G. S. (1999). Economic factors and the stock market: a new perspective. *Journal of Forecasting*, 18(3):151–166.

Quiggin, J. (1981). Perception and risk aversion among australian farmers. *Australian Journal of Agriculture Economics*, 25:160–169.

Quiggin, J. (1993). *Generalized Expected Utility Theory.* Kluwer Academic Publishers, Massachusetts.

Ramsay, J. and Silverman, B. W. (2005). *Functional Data Analysis.* Springer Series in Statistics.

Rockafellar, R. and Uryasev, S. (2000). Optimization of conditional var. *Journal of Risk*, 2:21–41.

Rockafellar, R. and Uryasev, S. (2002). Conditional value-at-risk for general loss distribution. *Journal of Banking and Finance*, 26:1443–1471.

Roll, R. and Ross, S. A. (1980). An empirical investigation of the arbitrage pricing theory. *The Journal of Finance*, 35(5):1073–1103.

Salvage, L. (1954). *Foundations of Statistics.* Wiley.

Schmeidler, D. (1986). Integral representation without additivity. In *Proceedings of the American Mathematical Society*, volume 97, pages 255–261.

Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Econometrica*, 57:571–587.

Segovia-Gonzalez, M. M., Guerrero, F. M., and Herranz, P. (2009). Explaining functional principal component analysis to actuarial science with an example on vehicle insurance. *Insurance: Mathematics and Economics*, 45(2):278–285.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance*, 19(3):425–442.

Shieh, S. J., Lin, C. Y., and Ho, P. H. (2012). Large changes in stock prices: Market, liquidity, and momentum effect. *The Quarterly Review of Economics and Finance*, 52(2):183–197.

Sood, A., James, G. M., and Tellis, G. J. (2009). Functional regression: A new model for predicting market penetration of new products. *Marketing Science*, 28(1):36–51.

Spearman, C. (1904). The proof and measurement of association between two things. *The American journal of psychology*, 15(1):72–101.

Sullivan, R., Timmermann, A., and White, H. (1999). Data-snooping, technical analysis trading rule performance, and the bootstrap. *Journal of Finance*, 54(5):1647–1691.

Tukey, J. W. (1960). A survey of sampling from contaminated distributions. *Contributions to probability and statistics*, 39:448–485.

Von Neumann, J. and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton.

Wang, Y. and Di Iorio, A. (2007). The cross section of expected stock returns in the chinese a-share market. *Global Finance Journal*, 17(3):335–349.

White, H. (2000). A reality check for data-snooping. *Econometrica*, 68:1097–1127.

Yao, F., Mller, H. G., and Wang, J. L. (2005). Functional linear regression analysis for longitudinal data. *The Annals of Statistics*, 33(6):2873–2903.

Zhang, Y. (2011). Estimation of beta in a simple functional capital asset pricing model for high frequency us stock data. Master Dissertation.