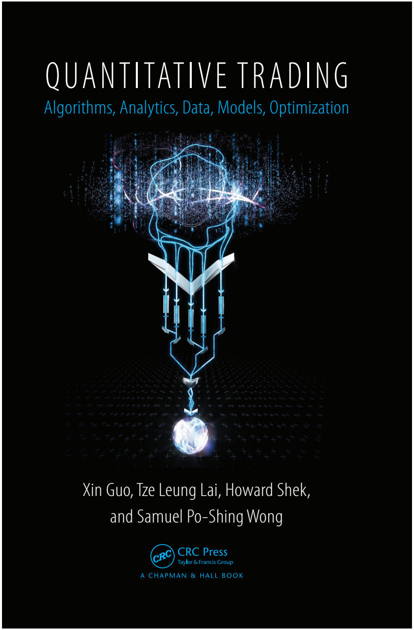
Quantitative Trading (I)

Tze Leung Lai

Stanford University



- **Data** cloud is first processed by an analytics machine.
- **Analytics** refers to both analysis of the data and the development of data-driven trading strategies which naturally make use of **Optimization**.
- **Models** provide the connection between the data and the trading strategies.
- **Algorithms** are step-by-step procedures for computing the solutions of not only optimization but also other mathematical and data analysis problems.

Outline

1. Efficient Market Hypothesis (EMH) and related methodologies
   1. Stylized Features of Returns
   2. Models that Capture the Features
   3. From Markowitz Pricing Theory (MPT) to Neo-MPT
   4. Active and Passive Portfolio
   5. Multi-period Dynamic Portfolio: Transaction Cost & Parameter Uncertainty.
2. Statistical Arbitrage Strategies
   1. Change-point Detection
   2. Moving Average & Generalized Likelihood Ratio Test
   3. Directional Trading
   4. Momentum & Pairs Trading
   5. Behavioral Finance & Contrarian Strategies
   6. Value Investing & Global Macro Strategies
   7. Evaluation of Investment Strategies and Multiple Testing

EMH

- The efficient market hypothesis (EMH): Eugene Fama, 1960s.
  - Utility-maximizing agents that have rational expectations.
  - Update their expectations whenever new information appears.
  - On average the agents are right even though individual agents can be wrong about the market so that the net effect of their reactions is random and cannot be reliably exploited to make an abnormal profit, especially in the presence of transaction costs.
- One of the simplest statistical models consistent with EMH is the random walk model, i.e. P_i denotes the closing price of a security on day i,
  - the price differences, $\Delta_i = P_i - P_{i-1}$, are i.i.d. $N(\mu, \sigma^2)$: Louise Bachelier's Ph.D. thesis *The Theory of Speculation* at the University of Paris (Sorbonne) in 1900.
  - the logarithmic returns, $r_i = \log P_i - \log P_{i-1}$, are i.i.d. $N(\mu_G, \sigma_G^2)$: Geometric Brownian Motion (GBM) proposed by Osborne (1959) and Samuelson (1965), who removes any possibility of negative price.

Infinite divisibility and Change-of-time-scale

Note that

$$\log(P_t) - \log(P_0) = \sum_{i=1}^{t} r_i.$$

Thus,

$$\begin{aligned} r^{(year)} &= \sum_{i=1}^{12} r_i^{(month)} \sim N(12 \times \mu_G^{(month)}, 12 \times (\sigma_G^{(month)})^2) \\ &= \sum_{i=1}^{252} r_i^{(day)} \sim N(252 \times \mu_G^{(day)}, 252 \times (\sigma_G^{(day)})^2) \end{aligned}$$

where $r_i^{(month)} \sim N(\mu_G^{(month)}, (\sigma_G^{(month)})^2)$ iid and $r_i^{(day)} \sim N(\mu_G^{(day)}, (\sigma_G^{(day)})^2)$ iid.

- Such "infinite divisibility" of normal distribution is critical in converting between returns of different time-scales: One of the most fundamental operations in quantitative finance.
- Empirical justification of the return's normality is crucial.

Empirical analysis of the random walk model

- Kendall (1953) analyzed 19 equity and 3 commodity price series that were sampled in a weekly and/or monthly manner and found the following stylized behaviors:
  (A) the serial correlation of all lags (the autocorrelation) for the price differences are all of very small magnitude, (which seems to support the random walk model),
  (B) the distribution of the price differences is symmetric about zero and leptokurtic, i.e. the tails of the distribution are thicker than the Gaussian tails (so-called the fat-tailedness) and the empirical peak is higher than the Gaussian peak.
- (A) was also found by Cowles and Jones (1937), Cowles (1960) and Working (1960).
- r_t and Δ_t behave quite similarly in terms of (A) and (B).

Paretian Stable Distribution (I)

- Mandelbrot (1963) suggests modelling the stylized features (A) & (B) of r_i by using the class of Stable Distribution such that similar change-in-time-scale formulas can be obtained.
- Self-decomposability (infinite divisibility): Let $Y_1, \ldots, Y_n$ be i.i.d. $Stable(\alpha, \beta, \gamma, \delta)$. Then there exist $c_n > 0$ and d_n such that $Y_1 + \cdots + Y_n$ has the same distribution as $c_n Y + d_n$, where $Y \sim Stable(\alpha, \beta, \gamma, \delta)$. In particular, for the case $\beta = \delta = 0$ associated with symmetric stable distributions, we can choose $c_n = n^{1/\alpha}$ and $d_n = 0$.
- Using the symmetric stable distribution to model return r_i, the conversion of scales can be easily performed by

$$\sum_{i=1}^{n} r_i = n^{1/\alpha} \times Y$$

  where $r_i \sim Stable(\alpha, 0, \gamma, 0)$ iid and $Y \stackrel{d}{=} r_1$.
- In particular, GBM corresponds to $\beta = \delta = 0$ and $\alpha = 2$.

Paretian Stable Distribution (II)

- By applying a graphical method on the daily (respectively, monthly) closing cotton price differences from 1900 to 1905 (respectively, from 1880 to 1940), Mandelbrot (1963) estimates the corresponding $\alpha = 1.7$ (< 2).
- Mandelbrot calls non-normal stable distributions "stable Paretian distributions" because they are the limiting distributions of normalized random walks with Paretian (a.k.a. power law) increments.
- In particular, Paretian refers to the distribution, introduced by Pareto, that has probability $cx^{-\alpha}$ of exceeding x as $x \to \infty$. The Pareto distribution has infinite mean for $\alpha \leq 1$, and infinite variance for $1 < \alpha \leq 2$.
- Since the density function of Stable$(\alpha, \beta, \gamma, \delta)$ does not have explicit formulas except for special cases such as normal, Cauchy, and the inverse Gaussian distribution (i.e., Lévy's distribution of the first passage time of Brownian motion), estimation of its parameters is far from being routine and has evolved during the past 50 years after Mandelbrot's seminal paper.

Subordinated Stochastic Processes (I)

- Mandelbrot and Taylor (1967) :Use $B_{T(t)}$ to model the price P_t where $\{B_s, s \geq 0\}$ is BM and $T(t)$ is a stable process with positive increments which is independent of the BM.
- $T(t)$ is called the *subordinator*, which is a "random clock".
- Thus, $B_{T(t)}$ has stationary independent increments and the increments have infinite variance because of $T(t)$.
- Clark (1973) removed the stable process assumption on $T(t)$. Instead he assumed $v_t = T(t) - T(t-1) > 0$ to be i.i.d. with mean μ and variance $\sigma^2 > 0$. Then, $\Delta_t = P_t - P_{t-1} | v(t) \sim N(0, \gamma^2 v_t)$. Under such model, the kurtosis of Δ_t is greater than 3, i.e., the stylized behaviors are captured within the finite variance framework.
- Clark further relaxed the assumption of stationarity for the independent increments of $T(t)$ and provided the important insight that $v_t = T(t) - T(t-1)$ might vary with the trading activity on day t.
- By using the trading volume V_t on day t as a proxy for v_t, he proposed to approximate $\gamma^2 v_t$ by AV_t^a or $\exp(B + bV_t)$.

Subordinated Stochastic Processes: Real Data Example (I)

- Pfizer daily closing prices between January 1, 2005 and December 31, 2014 together with the corresponding trading volumes.
- Following Clark (1973), 43 cases that have $\Delta_t = 0$ are removed and the following linear models are fitted:

$$\log \Delta_t^2 = \log A + a \log V_t + \varepsilon_t \quad \text{or} \quad \log \Delta_t^2 = B + bV_t + \varepsilon_t. \tag{1}$$

- The mean of V_t is of the order of 10^7, and the estimates are $\log(\hat{A}) = -25.2$, $\hat{a} = 1.23$; $\hat{B} = -4.83$, $\hat{b} = 2.47 \times 10^{-8}$.
- Since $\Delta_t / \sqrt{v_t} \sim N(0, \gamma^2)$ under Clark's model, normal QQ-plots of

$$z_t^{(a)} = \frac{\Delta_t}{\sqrt{\hat{A} \times V_t^{\hat{a}}}}, \qquad z_t^{(b)} = \frac{\Delta_t}{\sqrt{\exp\left\{\hat{B} + \hat{b} V_t\right\}}} \tag{2}$$

  are shown Figure 1. It is clear that $z_t^{(a)}$ and $z_t^{(b)}$ are much closer to the normal distribution than Δ_t. In facf, the kurtosis of $z_t^{(a)}$ and $z_t^{(b)}$ are 3.50 and 3.56, respectively, which are much closer to 3 than 9.9 (the kurtosis of Δ_t).

Subordinated Stochastic Processes: Real Data Example (II)

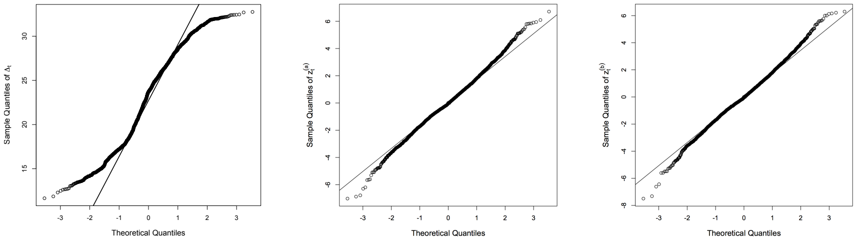


Figure 1: Normal QQ-plots of Δ_t, $z_t^{(a)}$ and of $z_t^{(b)}$ with $z_t^{(a)}$ and $z_t^{(b)}$ defined in (2).

ARMA, GARCH & Martingale Regression

Further developements of statistical models for r_t or Δ_t:

- Box and Jenkins (1970): ARMA (autoregressive, moving average) model built for daily Δ_t of IBM stock.
- Volatility clustering patterns were found in stock returns and other econometric time series and were captured by the ARCH model introduced by Engle (1982) and its generalization GARCH by Bollerslev: extensions of the ARMA model with martingale difference (instead of i.i.d.) innovations to r_t^2.
- Lai and Wei (1982) : Developed "stochastic regression" to analyze the behavior of stochastic input-output systems. In particular, the output y_t is related to $\mathbf{x}_t$ via $y_t = \boldsymbol{\beta}^\top \mathbf{x}_t + \varepsilon_t$, in which ε_t represents random noise that is assumed to be a martingale difference sequence. (Thus, it is also called the 'martingale regression" .)
  - $\mathbf{x}_t$ depends on the past outputs and inputs.
  - It includes ARMA and GARCH models as special cases (explained in the next 2 slides).

ARMA, GARCH & Martingale Regression (II)

- Specifically, letting $\mathscr{F}_t$ be the σ-field (information set) generated by $(\mathbf{x}_s, y_s)$, $s \le t$, it is assumed in stochastic regression that $E(\varepsilon_t | \mathscr{F}_{t-1}) = 0$ and $\mathbf{x}_t$ is $\mathscr{F}_{t-1}$-measurable. The AR(p) model is a special case with $\mathbf{x}_t = (y_{t-1}, \ldots, y_{t-p})^\top$.
- The ARMA(p, q) model is more complicated because the random disturbances come in the form of $\varepsilon_t + c_1\varepsilon_{t-1} + \cdots + c_q\varepsilon_{t-q}$.
- On the other hand, if $\varepsilon_{t-1}, \ldots, \varepsilon_{t-q}$ were observable, then this would reduce to a stochastic regression model with $\mathbf{x}_t = (y_{t-1}, \ldots, y_{t-p}, \varepsilon_{t-1}, \ldots, \varepsilon_{t-q})^\top$.
- Lai and Wei (1986): The extended least squares estimator of the ARMA parameters, i.e, if the parameters were known, then one could indeed retrieve $\varepsilon_{t-1}, \ldots, \varepsilon_{t-q}$ from $(\mathbf{x}_s, y_s)$, $1 \le s \le t$, by assuming $\mathbf{x}_s = \mathbf{0}$ and $\varepsilon_s = 0$ for $s \le 0$, which is the basis of computing the likelihood function.
- Lai and Ying (1991) subsequently applied martingale theory similar to that for extended least squares to analyze the recursive maximum likelihood estimator.

ARMA, GARCH & Martingale Regression (III)

- Note that the GARCH(h,k) model $\varepsilon_t = \sigma_t \zeta_t$, $\sigma_t^2 = \omega + \sum_{i=1}^h \beta_i \sigma_{t-i}^2 + \sum_{j=1}^k \alpha_j \varepsilon_{t-j}^2$ in which ζ_t are i.i.d. with mean 0 and variance 1, can be written as an ARMA model for ε_t^2:

$$\varepsilon_t^2 = \omega + \sum_{j=1}^{\max(h,k)} (\alpha_j + \beta_j)\varepsilon_{t-j}^2 + \eta_t - \sum_{i=1}^{h} \beta_i \eta_{t-i},$$

  in which $\eta_t = \varepsilon_t^2 - \sigma_t^2$ is a martingale difference sequence and $\alpha_j = 0$ for $j > k$, $\beta_i = 0$ for $i > h$.
- Thus, the assumption of martingale difference (instead of i.i.d. zero-mean) ε_t allows time series modeling of ε_t to incorporate dynamic changes in volatility. This volatility modeling also involves a martingale structure with martingale difference innovations η_t.
- Thus, $y_t = \boldsymbol{\beta}^\top \mathbf{x}_t + \varepsilon_t$ is called the "martingale regression"
- Martingale regression is a much more effective modeling tool (than the random walk and its cousins) in capturing the stylized features of asset returns.

From Single Asset Return Model to Multi-Asset Strategy

- Markowitz Portfolio Theory (MPT): m assets' return with mean vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)^\top$ and covariance matrix $\boldsymbol{\Sigma}$.
- Portfolio weight vector: $\mathbf{w} = (w_1, \ldots, w_m)^\top$ with $\mathbf{w}^\top \mathbf{1} = 1$.
- Markowitz's efficient portfolio for target mean return μ_* (short-selling is allowed):

$$\mathbf{w}_{\text{eff}} = \arg\min_{\mathbf{w}} \mathbf{w}^\top \boldsymbol{\Sigma} \mathbf{w} \qquad \text{subject to } \mathbf{w}^\top \boldsymbol{\mu} = \mu_*, \ \mathbf{w}^\top \mathbf{1} = 1.$$

- $\mathbf{w}_{\text{eff}} = \left\{ B\boldsymbol{\Sigma}^{-1}\mathbf{1} - A\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \mu_* \left(C\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - A\boldsymbol{\Sigma}^{-1}\mathbf{1} \right) \right\} / D$ where $A = \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}$, $B = \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$, $C = \mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}$, and $D = BC - A^2$.
- *Efficient frontier* is the collection of all possible $(\mu_*, \sqrt{\mathbf{w}_{\text{eff}}^\top \boldsymbol{\Sigma} \mathbf{w}_{\text{eff}}})$
- $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are actually unknown – Plug-in frontier: Replacing them by the sample mean vector $\widehat{\boldsymbol{\mu}}$ and covariance matrix $\widehat{\boldsymbol{\Sigma}}$ of a training sample of historical returns $\mathbf{r}_t = (r_{1t}, \ldots, r_{mt})^\top$, $1 \le t \le n$.
- However, this plug-in frontier is no longer optimal because $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\Sigma}}$ actually differ from $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, and portfolios associated with the plug-in frontier can perform worse than an equally weighted portfolio that is highly inefficient.

Three Approaches to Markowitz's Enigma

- Dimension reduction in estimating $\boldsymbol{\Sigma}$ via multifactor models, i.e., relating the ith asset returns r_i to k factors $f_1, \ldots, f_k$ by $r_i = \alpha_i + (f_1, \ldots, f_k)^\top \boldsymbol{\beta}_i + \varepsilon_i$, where α_i and $\boldsymbol{\beta}_i$ are unknown regression parameters and ε_i is an unobserved random disturbance that has mean 0 and is uncorrelated. Examples: CAPM, APT and Fama-French three-factor model.
- Use shrinkage estimates of $\boldsymbol{\Sigma}$ in the form of $\widehat{\boldsymbol{\Sigma}} = \widehat{\delta}\widehat{\mathbf{F}} + (1 - \widehat{\delta})\mathbf{S}$ where $\widehat{\delta}$ is an estimator of the *optimal shrinkage constant* and $\mathbf{S} = n^{-1} \sum_{i=1}^{n} (\mathbf{r}_i - \bar{\mathbf{r}})(\mathbf{r}_i - \bar{\mathbf{r}})^\top$. $\widehat{\mathbf{F}}$ is given by the mean of the prior distribution or a structured covariance matrix $\mathbf{F}$ with much fewer parameters than $m(m+1)/2$; see Ledoit and Wolf (2003, 2004). The estimate of $\boldsymbol{\mu}$ can also be handled by shrinkage similarly.
- To correct for the bias of $\widehat{\mathbf{w}}_{\text{eff}}$, use the average of the bootstrap weight vectors $\bar{\mathbf{w}} = B^{-1} \sum_{b=1}^{B} \widehat{\mathbf{w}}_b^*$, where $\widehat{\mathbf{w}}_b^*$ is the estimated optimal portfolio weight vector based on the bth bootstrap sample $\{\mathbf{r}_{b1}^*, \ldots, \mathbf{r}_{bn}^*\}$ drawn with replacement from the observed sample $\{\mathbf{r}_1, \ldots, \mathbf{r}_n\}$; see Michaud (1989).

New approach of Lai et al. (2011b) : Neo-MPT (I)

To handle the parameter uncertainty, Lai et al. (2011b) (abbreviated by LXC) proposed using the portfolio via the following optimization scheme (to replace the classical MPT which is derived under the assumption of known parameters)

$$\max \left\{ E(\mathbf{w}^\top \mathbf{r}_{n+1}) - \lambda \operatorname{Var}(\mathbf{w}^\top \mathbf{r}_{n+1}) \right\} \tag{3}$$

LXC solve (3) by rewriting it as the following maximization problem over η:

$$\max_{\eta} \left\{ E\left[\mathbf{w}^\top(\eta)\mathbf{r}_{n+1}\right] - \lambda \operatorname{Var}\left[\mathbf{w}^\top(\eta)\mathbf{r}_{n+1}\right] \right\}, \tag{4}$$

where $\mathbf{w}(\eta)$ is the solution of the stochastic optimization problem

$$\mathbf{w}(\eta) = \arg\min_{\mathbf{w}} \left\{ \lambda E\left[(\mathbf{w}^\top \mathbf{r}_{n+1})^2\right] - \eta E(\mathbf{w}^\top \mathbf{r}_{n+1}) \right\}. \tag{5}$$

New approach of LXC: Neo-MPT (II)

In particular, when there is no limit on short selling, $\mathbf{w}(\eta)$ in (5) is given explicitly by

$$\begin{aligned} \mathbf{w}(\eta) &= \arg\min_{\mathbf{w}:\mathbf{w}^\top \mathbf{1}=1} \left\{ \lambda \mathbf{w}^\top \mathbf{V}_n \mathbf{w} - \eta \mathbf{w}^\top \boldsymbol{\mu}_n \right\} \\ &= \frac{1}{C_n} \mathbf{V}_n^{-1} \mathbf{1} + \frac{\eta}{2\lambda} \mathbf{V}_n^{-1} \left(\boldsymbol{\mu}_n - \frac{A_n}{C_n} \mathbf{1} \right), \end{aligned} \tag{6}$$

where the second equality can be derived by using a Lagrange multiplier and

$$A_n = \boldsymbol{\mu}_n^\top \mathbf{V}_n^{-1} \mathbf{1} = \mathbf{1}^\top \mathbf{V}_n^{-1} \boldsymbol{\mu}_n, \quad B_n = \boldsymbol{\mu}_n^\top \mathbf{V}_n^{-1} \boldsymbol{\mu}_n, \quad C_n = \mathbf{1}^\top \mathbf{V}_n^{-1} \mathbf{1}.$$

with μ_n and $\mathbf{V}_n$ being estimated by the martingale regression models in LXC.

The above approach is called *Neo-MPT* to highlight their similarity with MPT in searching for optimal portfolio.

Remark: Quadratic programming can be used to compute $\mathbf{w}(\eta)$ for more general linear and quadratic constraints (to handle various form of short selling limits).

Combining Martingale Regression with Neo-MPT (I)

- To incorporate the factor structure to the model that captures the stylized properties, LXC proposed a martingale regression model for r_{it} (the return of the ith security on day t):

$$r_{it} = \boldsymbol{\beta}_i^\top \mathbf{x}_{i,t-1} + \varepsilon_{it}, \tag{7}$$

  where the components of $\mathbf{x}_{i,t-1}$ include 1, factor variables such as the return of a market portfolio like S&P 500 at time $t-1$, and lagged variables $r_{i,t-1}, r_{i,t-2}, \ldots$. (Note: $\boldsymbol{\beta}_i$ can be estimated by the method of moments.)
- LXC model the heteroskedasticity by assuming that $\varepsilon_{it} = s_{i,t-1}(\boldsymbol{\gamma}_i) z_{it}$, where z_{it} are i.i.d. with $E(z_{it}) = 0$ and $Var(z_{it}) = 1$, $\boldsymbol{\gamma}_i$ is a parameter vector which can be estimated by maximum likelihood or generalized method of moments, and $s_{i,t-1}$ is a given function that depends on $r_{i,t-1}, r_{i,t-2}, \ldots$.
- A well-known example is the GARCH(1,1) model

$$\varepsilon_{it} = s_{i,t-1} z_{it}, \qquad s_{i,t-1}^2 = \omega_i + a_i s_{i,t-2}^2 + b_i r_{i,t-1}^2, \tag{8}$$

  for which $\boldsymbol{\gamma}_i = (\omega_i, a_i, b_i)$.

Combining Martingale Regression with Neo-MPT (II)

- Note that (7)–(8) models the asset returns separately and the returns of different assets are linked only via the vector $\mathbf{z}_t$ which are assumed to be i.i.d. (but their components are not uncorrelated).
- Such model is substantially more parsimonious than a multivariate regression or a multivariate GARCH model.
- The conditional cross-sectional covariance between the returns of assets i and j given $\mathscr{R}_n = \{\mathbf{r}_1, \ldots, \mathbf{r}_n\}$ is given by

$$\mathrm{Cov}(r_{i,n+1}, r_{j,n+1} | \mathscr{R}_n) = s_{i,n}(\boldsymbol{\gamma}_i) s_{j,n}(\boldsymbol{\gamma}_j) \mathrm{Cov}(z_{i,n+1}, z_{j,n+1} | \mathscr{R}_n),$$

  for the model (7)–(8). Thus, the estimator of $E(\mathbf{r}_{n+1}|\mathscr{R}_n)$ and $E(\mathbf{r}_{n+1}\mathbf{r}_{n+1}^\top|\mathscr{R}_n)$ are $\boldsymbol{\mu}_n = (\widehat{\boldsymbol{\beta}}_1^\top \mathbf{x}_{1,n}, \ldots, \widehat{\boldsymbol{\beta}}_m^\top \mathbf{x}_{m,n})^\top$ and $\mathbf{V}_n = \boldsymbol{\mu}_n \boldsymbol{\mu}_n^\top + (\widehat{s}_{i,n}\widehat{s}_{j,n}\widehat{\sigma}_{ij})_{1 \le i,j \le n}$,, respectively, in which $\widehat{\boldsymbol{\beta}}_i$ is the least squares estimate of $\boldsymbol{\beta}_i$, and $\widehat{s}_{l,n}$ and $\widehat{\sigma}_{ij}$ are the usual estimates of $s_{l,n}$ and $\mathrm{Cov}(z_{i,1}, z_{j,1})$ based on $\mathscr{R}_n$.
- These estimators of $E(\mathbf{r}_{n+1}|\mathscr{R}_n)$ and $E(\mathbf{r}_{n+1}\mathbf{r}_{n+1}^\top|\mathscr{R}_n)$ are the required inputs of the LXC's Neo-MPT.

Choice of λ

- Theoretically, the Lagrange multiplier λ in (3) can be regarded as the investor's risk-aversion index when variance is used to measure risk.
- However, in practice, it may be difficult to specify an investor's risk aversion parameter λ.
- LXC suggests taking λ as a tuning parameter which is chosen by maximizing (over a grid of possible λ) the bootstrap estimate of the information ratio $E_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\mathbf{w}_\lambda \mathbf{r} - r_0) \Big/ \sqrt{\mathrm{Var}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(\mathbf{w}_\lambda^\top \mathbf{r} - r_0)}$.
- In LXC's real data study, the S&P 500 Index is taken as the benchmark portfolio.
- Also, λ is chosen by maximizing the information ratio over the grid $\lambda \in \{2^i : i = -3, -2, \ldots, 6\}$.

Real Data Study of Neo-MPT (I)

- LXC Data: CRSP monthly stock market data from January 1985 to December 2009.
- $m = 50$ stocks with the largest market values among those that have no missing monthly prices in the training period of the first 10 years.
- LXC evaluate out-of-sample performance for each month after the training period.
- They choose $m = 50$ stocks with the largest market values among those that have no missing monthly prices in the training period.
- Short selling is allowed, with the constraint $w_i \geq -0.05$ for all i.

Real Data Study of Neo-MPT (II)

- Two forms of martingale regression model 7 are considered. NPEB_{AR} sets $\mathbf{x}_{i,t-1} = (1, r_{i,t-1})$ and NPEB_{SRG} simply adds the first lag of S&P 500 Index return to $\mathbf{x}_{i,t-1}$.
- NPEB refers to the non-parametric empirical Bayes methodology which is the general approach for computing the optimal portfolio under Neo-MPT framework (by treating the unknown parameters as state variables).
- Figure 2: Time series plot of the cumulative realized excess returns over the S&P 500 Index (the benchmark) during the test period of 180 months, for NPEB_{SRG}, NPEB_{AR}, the plug-in portfolio of MPT as well as the Ledoit-Wolf portfolio and Michaud's bootstrap portfolio, using $\mu_* = 0.015$ as the target return.
- Figure 2: NPEB_{SRG} performs markedly better than NPEB_{AR}, which already greatly outperforms the other three procedures that perform similarly to each other.

Real Data Study of Neo-MPT (III)

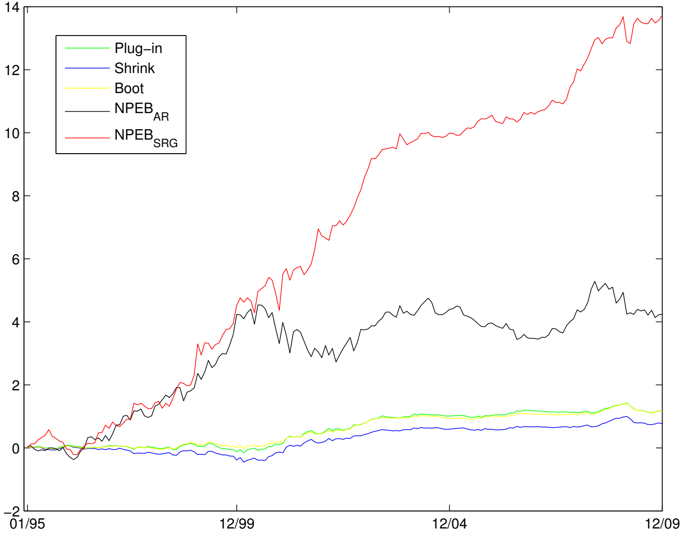


Figure 2: Realized cumulative excess returns over the S&P 500 Index.

Active Portfolio Management

- The cornerstones of quantitative portfolio management are prediction of asset returns from a large pool of investment possibilities, risk estimation, and portfolio optimization.
- There are two main styles of portfolio management – passive and active.
- Passive portfolio management constructs and administers portfolios that tracks some given index. Rationale: Index tracking incurs low cost as it does not require much information gathering on individual stocks. By reducing investment costs, the net return improves. Moreover, relatively infrequent trading results in fewer capital gains and therefore lower taxes.
- Goal of active portfolio management: Construct portfolios that aim to outperform some index or benchmark. The additional return that a portfolio generates relative to the benchmark is commonly known as the *alpha* of the portfolio. Performance is measured by the information ratio that expresses mean excess return in units of its standard deviation.

Active Alpha and Exotic Beta

- Two main sources of alpha are (a) superior information, and (b) efficient information processing.
- Active portfolio management outperforms the passive approach in the absence of transaction costs, but the advantage may be outweighed by the transaction costs.
- Let $\boldsymbol{r} \in \mathbb{R}^m$ be the return vector of m assets. The returns of the portfolio and benchmark are $r_P = \boldsymbol{w}_P^\top \boldsymbol{r}$ and $r_B = \boldsymbol{w}_B^\top \boldsymbol{r}$ where $\boldsymbol{w}_P$ and $\boldsymbol{w}_B$ are the corresponding portfolio weights. Consider the one-factor model $r_P - r_f = \alpha + \beta(r_B - r_f) + \varepsilon$ where r_f is the risk-free rate, subtracting $r_B - r_f$ from both sides of the equation defines the "active return" of the portfolio by $r_P - r_B = \alpha + \beta_P(r_B - r_f) + \varepsilon$, in which the "active" α denotes the additional return of the portfolio over that of the benchmark, $\beta_P = \beta - 1$ is known as the active beta of the portfolio.
- Since the benchmark portfolio has beta equal to 1 (and therefore zero active beta), a portfolio with positive alpha and small $|\beta| < 1$ can have a high information ratio. Such a portfolio is said to have an "exotic beta".

Active Portfolio Optimization via Neo-MPT

- An active portfolio seeks to get better risk-adjusted returns than the benchmark portfolio B to justify the fees of the portfolio manager.
- Since the difference $\widetilde{\boldsymbol{w}} = \boldsymbol{w}_P - \boldsymbol{w}_B$ satisfies $\widetilde{\boldsymbol{w}}^\top \mathbf{1} = 0$ for given portfolio P, Neo-MPT can be reformulated for the active portfolio weight vector $\widetilde{\boldsymbol{w}}$ as:

$$\begin{gathered} \max \left\{ E(\widetilde{\boldsymbol{w}}^\top \boldsymbol{r}_{n+1}) - \lambda \mathrm{Var}(\widetilde{\boldsymbol{w}}^\top \boldsymbol{r}_{n+1}) \right\}, \\ \text{subject to } \widetilde{\boldsymbol{w}}^\top \mathbf{1} = 0 \text{ and } \widetilde{\boldsymbol{w}} \in \mathscr{C}, \end{gathered} \tag{9}$$

  where $\mathscr{C}$ represents long-short and other constraints.
- λ is again chosen by maximizing the information ratio estimated from training data.
- This is different from the previous Neo-MPT because it tries to modify the benchmark portfolio weights, which are often updated daily and therefore already contains “free" market information.
- The formulation of (9) can conveniently incorporate domain knowledge and prior beliefs about future movements of the stocks and of their associated firms and financial sectors.

Neo-MPT Active Portfolio: Real Data Study (I)

- To illustrate Neo-MPT Active Portfolio, LXC use a value-weighted portfolio of 50 stocks as the benchmark portfolio and the constraint set $\mathscr{C} = \{\widetilde{\boldsymbol{w}} : -\boldsymbol{w}_B \leq \widetilde{\boldsymbol{w}} \leq c\mathbf{1} - \boldsymbol{w}_B\}$, with $c = 0.1$, i.e., the portfolio is long only and the total position in any stock cannot exceed an upper bound c.
- Similar to Neo-MPT with martingale regression, $m = 50$ stocks are chosen at the beginning of each month t so that they have a largest market value among those in the CRSP database that have no missing monthly pieces in the first 120 months, which are used as the training samples.
- Note that LXC does not incorporate domain knowledge in this active portfolio study so that the corresponding performance of the proposed methodology is comparable to the other procedures presented in Table 1.

Neo-MPT Active Portfolio: Real Data Study (II)

- The plug-in portfolio in Table 1 is obtained via solving the optimization problem

$$\min E(\widetilde{\boldsymbol{w}}^\top \boldsymbol{\Sigma} \widetilde{\boldsymbol{w}}), \quad \text{subject to } \widetilde{\boldsymbol{w}}^\top \boldsymbol{\mu} = \widetilde{\mu}_*, \ \widetilde{\boldsymbol{w}}^\top \mathbf{1} = 0 \text{ and } \widetilde{\boldsymbol{w}} \in \mathscr{C}; \quad (10)$$

  in which $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are replaced, for the plug-in active portfolio, by their sample estimates based on the training sample.
- The covariance-shrinkage (abbreviated "shrink" in Table 1) active portfolio uses a shrinkage estimator of $\boldsymbol{\Sigma}$ (shrink towards a patterned matrix that assumes all pairwise correlations to be equal (Ledoit and Wolf, 2003).
- Note that some levels of $\widetilde{\mu}_*$ may be vacuous for the plug-in, the "shrink" and resampled (abbreviated "boot" for bootstrapping) active portfolios in a given test period.
- In this real data study, for $\widetilde{\mu}_* = 0.01, 0.015, 0.02, 0.03$, there are 92, 91, 91 and 80 test periods, respectively, for which (10) has solutions when $\boldsymbol{\Sigma}$ is replaced by either the sample covariance matrix or the Ledoit-Wolf shrinkage estimator computed from the training sample.

Neo-MPT Active Portfolio: Real Data Study (III)

- Higher levels of target returns result in even fewer of the 180 test periods for which (10) has solutions.
- On the other hand, values of $\widetilde{\mu}_*$ that are lower than 1% may be of little practical interest to active portfolio managers.
- When (10) does not have a solution to provide a portfolio of a specified type for a test period, LXC uses the value-weighted benchmark as the portfolio for the test period.
- Table 1(a) gives the actual (annualized) mean realized excess returns to show the extent to which they match the target value $\widetilde{\mu}_*$, and also the corresponding annualized standard deviations, over the 180 test periods for the plug-in, covariance-shrinkage and resampled active portfolios constructed with the above modification.
- These numbers are very small, showing that the three portfolios differ little from the benchmark portfolio, so the realized information ratios that range from 0.24 to 0.83 for these active portfolios can be quite misleading if the actual mean excess returns are not taken into consideration.

Neo-MPT Active Portfolio: Real Data Study (IV)

Table 1: Means and standard deviations (in parentheses) of annualized realized excess returns over the value-based benchmark

$\widetilde{\mu}_*$	0.01	0.015	0.02	0.03
λ	2^2	2	2^{-1}	2^{-2}
(a) All test periods by re-defining portfolios in some periods				
Plug-in	0.001 (4.7e-3)	0.002 (7.3e-3)	0.003 (9.6e-3)	0.007 (1.4e-2)
Shrink	0.003 (4.3e-3)	0.004 (6.6e-3)	0.006 (8.8e-3)	0.011 (1.3e-2)
Boot	0.001 (2.5e-3)	0.001 (3.8e-3)	0.001 (5.1e-3)	0.003 (7.3e-3)
NPEB	0.029 (1.2e-1)	0.046 (1.3e-1)	0.053 (1.5e-1)	0.056 (1.6e-1)
(b) Test periods in which all portfolios are well defined				
Plug-in	0.002 (6.6e-3)	0.004 (1.0e-2)	0.006 (1.4e-2)	0.014 (1.9e-2)
Shrink	0.005 (5.9e-3)	0.008 (9.0e-3)	0.012 (1.2e-2)	0.021 (1.8e-2)
Boot	0.001 (3.5e-3)	0.003 (5.3e-3)	0.003 (7.1e-3)	0.006 (1.0e-2)
NPEB	0.282 (9.3e-2)	0.367 (1.1e-1)	0.438 (1.1e-1)	0.460 (1.1e-2)

From Single Period to Multiperiod (I)

- diBartolomeo (2012): MPT "frames the time dimension of investing as a single period over which the parameters of the probability distribution of asset returns are both known with certainty and unchanging", but that "neither assumption is true in the real world."
- Academic research in "full multi-period optimization" has seldom been used by investment professionals who instead focus on when it is really necessary to bear the costs and use single-period mean-variance optimization to rebalance their portfolio weights.

From Single Period to Multiperiod (II)

- Negative comments towards "all or nothing" rebalancing rules:
  - When active managers are inactive because the potential benefits of rebalancing are too small, this lack of trading is perceived by clients as the manager being neglectful rather than as an analytically-driven decision to reduce trading costs.
  - After a period of inactivity, the eventual rebalancing concentrates the required trading into a particular moment in time, resulting in market impact and higher transaction cost.
- Here we review some methods in the area of "academic full multi-period optimization" and share some practical ideas on the development of multiperiod strategies in when and how to rebalance with transaction cost and parameter uncertainty in mind.

Samuelson-Merton Theory of "Lifetime Portfolio Selection" (I)

- Merton (1969)'s continuous-time model: a bond paying a fixed risk-free rate $r > 0$ and a stock with price dynamic: $dS_t = S_t(\alpha\, dt + \sigma\, dB_t)$ with $\alpha > 0$ and $\sigma > 0$.
- The investor's position is denoted by (X_t, Y_t), where X_t and Y_t are the dollar value of investment in bond and in stock, respectively.
- For each $t \in (0, T]$, the investor consumes at rate C_t from the bond, and L_t (respectively, M_t) represents the cumulative dollar value of stock bought (respectively, sold) within the time interval $[0, t]$. Thus, (X_t, Y_t) satisfies

$$dX_t = (rX_t - C_t)\, dt - dL_t + dM_t, \tag{11}$$

$$dY_t = \alpha Y_t\, dt + \sigma Y_t\, dB_t + dL_t - dM_t. \tag{12}$$

Samuelson-Merton Theory of "Lifetime Portfolio Selection" (I)

Solution of (12) is obtained by

- Requiring C_t, L_t and M_t to be non-negative
- Maximizing the expected utility (with the consumption and terminal wealth both chosen from the class of constant relative risk aversion)

Solution: The optimal strategy is to devote a constant proportion (the Merton proportion) p of the investment to the stock and to consume at a rate proportional to wealth.

Proportional Transaction Cost in "Lifetime Portfolio Selection"

- Suppose the investor pays fractions $0 \leq \lambda < 1$ and $0 \leq \mu < 1$ of the dollar value transacted on purchase and sale of a stock, respectively. Then, the investor's position (X_t, Y_t) satisfies

$$dX_t = (rX_t - C_t)\,dt - (1+\lambda)\,dL_t + (1-\mu)\,dM_t, \tag{13}$$

$$dY_t = \alpha Y_t\,dt + \sigma Y_t\,dW_t + dL_t - dM_t. \tag{14}$$

- With these changes, the problem is still to find (C, L, M) within the feasible set that maximizes the expected utility.
- It turns out to be a *singular stochastic control* problem, characterized by a no-trade region that is bounded between a buy-boundary $\mathscr{B}_t$ and a sell boundary $\mathscr{S}_t$, buying stock immediately when the stock price S_t is at or below the buy-boundary $\mathscr{B}_t$, and selling stock immediately when $S_t \geq \mathscr{S}_t$.

Multiperiod Mean-Variance Portfolio Rebalancing (I)

- Even after incorporating proportional transaction costs and multiperiod expected utility maximization theory, we cannot ignore the fact that the active (or passive) portfolio fund's performance is evaluated periodically according to some risk-adjusted measure such as information ratio.
- That's the reason why most active portfolio managers still prefer to use Markowitz's framework of mean-variance optimization.
- Moreover, Levy and Markowitz (1979) and Kroll et al. (1984) have shown how expected utility can be approximated by a function of mean and variance in applications to portfolio selection.
- However, mean-variance portfolio rebalancing in a multiperiod dynamic framework has been a long-standing problem. Chapter 21 of Grinold and Kahn (2000) :"Active portfolio management is a dynamic problem,. . . . With a proper frame, managers should make decisions now, accounting for these dynamics and interactions now and in the future. One simple open question is, when should we trade (in the presence of transaction costs)?"

Multiperiod Mean-Variance Portfolio Rebalancing (II)

- Merton (1990) has introduced a continuous-time analog of mean-variance optimization and Zhou and Li (2000) provide a linear-quadratic (LQ) stochastic control framework for the problem. Pliska (1997) and Li and Ng (2000) consider the discrete-time version of this multiperiod problem. However, because of the “curse of dimensionality” in dynamic programming, this approach has to be limited to relatively few assets and is seldom used in practice; see Kritzman et al. (2007).
- Instead, heuristic procedures extending single-period mean-variance analysis to multiple periods in a changing world, which are scalable to higher dimensions, are often used.
- In particular, Markowitz and van Dijk (2003) (denoted by MvD) propose to use (a) a “mean-variance surrogate”, which is a linear combination of the mean and variance of the portfolio return, to substitute for the discounted sum of the conditional expected future utilities given the information up to the present in an infinite-horizon setting, and (b) approximate dynamic programming (ADP).

Multiperiod Mean-Variance Portfolio Rebalancing (III)

- Kritzman et al. (2007) report some simulation studies, using real world scenarios, comparing the Markowitz-van-Dijk (MvD) approach with the dynamic programming (DP) solution that is feasible in these cases and several standard industry heuristics.
- They find that the performance of MvD is "remarkably close" to that of DP in cases with DP can be computed with reliable numerical accuracy, and "far superior to solutions based on standard industry heuristics."
- They also highlight the issue of "a changing world" in MvD's title, saying: "Almost immediately upon implementation (of mean-variance analysis to determine optimal portfolio weights), however, the portfolio's weights become sub-optimal as changes in asset prices cause the portfolio to drift away from the optimal targets."

Dynamic Mean-Variance Portfolio Optimization in the Presence of Transaction Costs (I)

- Markowitz and van Dijk (2003) and Kritzman et al. (2007) incorporate linear transaction costs

$$C_t = \sum_{j=1}^{m} \kappa_j |w_t^j - w_{t-1}^j| \tag{15}$$

  in carrying out the MvD heuristic procedure, in which $\kappa_1, \ldots, \kappa_m$ are prescribed constants, and their portfolios do not include short selling.
- Gârleanu and Pedersen (2013) consider more tractable transaction costs that are quadratic functions of $\Delta \boldsymbol{u}_t = \boldsymbol{u}_t - \boldsymbol{u}_{t-1}$, where u_t^j is the number of shares of stock j in the portfolio at time t.

Dynamic Mean-Variance Portfolio Optimization in the Presence of Transaction Costs (II)

- They assume a multi-factor model for the vector $\tilde{\boldsymbol{R}}_{t+1}$ of next period's excess returns over the risk-free rate R_f, multiplied by current stock prices, so that the components of $\tilde{\boldsymbol{R}}_{t+1}$ are $\tilde{R}^j_{t+1} = P^j_t(R^j_{t+1} - R_f)$, $j = 1, \ldots, m$:

$$\tilde{\boldsymbol{R}}_{t+1} = \boldsymbol{B}\boldsymbol{f}_t + \boldsymbol{\varepsilon}_{t+1}, \qquad (16)$$

  with observed factor $\boldsymbol{f}_t$ at time t and known covariance matrix $\boldsymbol{\Sigma}$ for the i.i.d. random disturbances $\boldsymbol{\varepsilon}_{t+1}$.
- Assuming a vector autoregressive model $\boldsymbol{f}_{t+1} = (\boldsymbol{I} - \boldsymbol{\Phi})\boldsymbol{f}_t + \boldsymbol{w}_{t+1}$ for the factors, they consider the optimization problem of sequential choice of $\boldsymbol{u}_t$ to maximize

$$E\sum_{t=1}^{\infty}(1-\rho)^t\left[(1-\rho)(\boldsymbol{u}_t^\intercal\tilde{\boldsymbol{R}}_{t+1} - \gamma\boldsymbol{u}_t^\intercal\boldsymbol{\Sigma}\boldsymbol{u}_t) - \frac{1}{2}(\Delta\boldsymbol{u}_t)^\intercal\boldsymbol{\Lambda}\Delta\boldsymbol{u}_t\right], \qquad (17)$$

  in which the transaction cost matrix $\boldsymbol{\Lambda}$ is proportional to $\boldsymbol{\Sigma}$, with $\lambda > 0$ being the constant of proportionality.

Dynamic Mean-Variance Portfolio Optimization in the Presence of Transaction Costs (III)

- The parameter $0 < \rho < 1$ is a discount factor in the infinite-horizon problem (17), in which γ either a Lagrange multiplier when a risk constraint is imposed or a measure of the investor's risk aversion in the absence of risk constraints.
- Because of the linear dynamics and quadratic costs consisting of both $\boldsymbol{u}_t^\intercal \boldsymbol{\Sigma} \boldsymbol{u}_t$ and $(\Delta \boldsymbol{u}_t)^\intercal \boldsymbol{\Lambda} \Delta \boldsymbol{u}_t$, (17) (as a a Markov decision problem with state $\boldsymbol{f}_t$ and control $\Delta \boldsymbol{u}_t$) can be solved explicitly by dynamic programming.

Dynamic Mean-Variance Portfolio Optimization in the Presence of Transaction Costs (IV)

- They give an empirical application of this approach using 15 different commodity futures in the period January 1, 1996 – January 23, 2009, showing superior net returns relative to several benchmarks.
- The factors they choose are (a) mean divided by standard deviation of the past 5 days' price changes for each commodity future, stacked into a vector, (b) vector of an analogous quantities for past year's price changes, and (c) corresponding vector for the past 5 years' price changes.
- For the transition matrix $\mathbf{\Lambda} = \lambda \mathbf{\Sigma}$, they use an estimate of λ proposed by Engle et al. (2012). They also use a shrinkage estimator of $\mathbf{\Sigma}$ for their empirical study.

Dynamic Portfolio Selection in the Presence of Parameter Uncertainty (I)

- While the issue of parameter uncertainty under MPT framework is addressed by various researchers (including LXC), the role of sequential learning under the fully dynamic (continuous-time) portfolio optimization has also been recognized by Williams (1977), Detemple (1986), Dothan and Feldman (1986), Gennotte (1986), and Brennan (1998).
- In particular, Brennan (1998) treat the unknown drift parameters as state variables and estimate them sequentially by a filtering approach.
- These sequential estimates are used to substitute for the parameters in the procedures of Merton (1971) and others that assume known parameters in the asset price models.

Dynamic Portfolio Selection in the Presence of Parameter Uncertainty (II)

- Model of Brennan (1998): a risk-free asset with instantaneous rate of return r and a risky asset $dP_t/P_t = \mu dt + \sigma dB_t$ in which σ is known to the investor but μ is unknown and with a prior distribution $N(m_0, v_0)$.
- μ can be learnt via the posterior distribution $N(m_t, v_t)$, in which m_t is given by the Kalman-Bucy filter

$$dm_t = \frac{v_t}{\sigma^2}\left(\frac{dP_t}{P_t} - m_t dt\right), \quad \text{with } v_t = \frac{v_0 \sigma^2}{v_0 t + \sigma^2}; \tag{18}$$

- By maximizing the expected utility of the investor's terminal wealth Z_T, the optimal portfolio weight w_t, $0 \le t \le T$ can then be solved by using dynamic programming.
- He applies this approach to an empirical study with the S&P 500 index as the risky asset for the 69-year period 1926–1994 and shows that imperfect knowledge about the parameters gives rise to two quite distinct phenomena which he labels as “estimation risk" and “learning".

Dynamic Portfolio Selection in the Presence of Parameter Uncertainty (III)

- He points out that although increased risk due to uncertainty about the mean return tends to reduce the fraction of the portfolio allocated to the risky asset, "the prospect of learning more about the true value of the mean parameter, μ, as more returns are observed, induces an additional hedging demand for the risky asset."
- Xia (2001) examines the effects of estimation risk and learning in a more general framework and shows that "the hedge demand associated with the uncertain parameters plays a predominant role in the optimal strategy", which is "horizon dependent: the optimal stock allocation can increase, decrease or vary non-monotonically with the horizon, because parameter uncertainty induces a state-dependent hedge demand that may increase or decrease with the horizon."

Dynamic Portfolio Selection in the Presence of Parameter Uncertainty (IV)

- The price process of Xia (2001) is $dP_t/P_t = \mu_t dt + \sigma dB_t$, in which $\mu_t = \alpha + \boldsymbol{\beta}^\top \boldsymbol{f}_t$ and $\boldsymbol{f}_t$ is a $k \times 1$ vector of observable factors that undergo linear stochastic dynamics $d\boldsymbol{f}_t = (\boldsymbol{A}_{0,t} + \boldsymbol{A}_{1,t}\boldsymbol{\beta})dt + \boldsymbol{\Sigma}_t^{1/2} d\tilde{\boldsymbol{B}}_t$, where $\boldsymbol{A}_{1,t}$ and $\boldsymbol{\Sigma}_t$ are given $(k \times k)$ matrices and $\boldsymbol{A}_{0,t}$ is a $k \times 1$ vector that may depend on $(\boldsymbol{P}_t, \boldsymbol{f}_t)$, and $\tilde{\boldsymbol{B}}_t$ is a k-dimensional Brownian motion that is independent of B_t. Assuming a normal prior distribution for $(\alpha, \boldsymbol{\beta}^\top)^\top$, the posterior distribution given $(\boldsymbol{P}_s, \boldsymbol{f}_s)$, $0 \le s \le t$, is normal and is given by the Kalman-Bucy filter.
- Cvitanić et al. (2006) follow up on Brennan's work but extend to m risky assets plus a market or benchmark portfolio. They use the the Kalman-Bucy filter to show that the posterior distribution of the unknown drift vector given the observed prices up to time t is multivariate normal. Also, they derive an explicit formula for the optimal portfolio weight by adopting the power utility function.

Dynamic Portfolio Selection in the Presence of Parameter Uncertainty (V)

- Thus, filtering (specifically the Kalman-Bucy filter) has played a prominent role in dynamic portfolio optimization when there is uncertainty about the drift parameters of the price processes.
- Uncertainty about the volatility parameters, however, has not received much attention in the literature and previous works all assumed them to be known. The main reason for the lack of similar treatment of uncertainty about the volatility parameters is that it would involve nonlinear filtering methods.
- However, there are important recent advances in this problem by using sequential Monte Carlo methods called adaptive particle filters that can be used to handle volatility estimation; see Chan and Lai (2013, 2016), Lai and Bukkapatanam (2013), and Chapter 6 of Lai and Xing (2016).

Statistical Arbitrage

- Malkiel (2003) : "Obviously, I am biased against the chartist. . . . Technical analysis is anathema to the academic world. . . the method is patently false."
- Brock et al. (1992, p. 1732): " (it) has been enjoying a renaissance on Wall Street" and that "all major brokerage firms publish technical commentary on the market and individual securities. . . based on technical analysis."
- Let's put off EMH and consider a wide variety of quantitative trading strategies that attempt to generate expected returns exceeding a certain level by searching for and capitalizing on statistical arbitrage opportunities (SAOs).
- "Technical analysis" could link to nonparametric regression and change-point detection methods as discussed in Brock et al. (1992), Lo et al. (2000) and Section 11.1.1 of Lai and Xing (2008).
- The validation of investment strategies could be performed by using the multiple testing methodology developed by Lai, Gross, and Shen (2011a).

Filter Rule and equivalent CUSUM Change-point Detection Rule (I)

- Filter rule is built upon the concepts of: *drawdown* $D_t = \max_{0\le k\le t} P_k - P_t$, and *drawup* $U_t = P_t - \min_{0\le k\le t} P_k$ at time t.
- Buy signal if $D_t/(\max_{0\le k\le t} P_k) \ge c$ (suggesting a $100c\%$ drop from the running maximum price).
- Sell signal if $U_t/(\min_{0\le k\le t} P_k) \ge c$. Letting $r_k = \log(P_k/P_{k-1})$ and $S_t = \log P_0 + \sum_{k=1}^{t} r_k$.
- Note that

$$U_t \Big/ \left(\min_{0\le k\le t} P_k\right) \ge c \Longleftrightarrow S_t - \min_{0\le k\le t} S_k \ge \log(1+c) \qquad (19)$$

Filter Rule and equivalent CUSUM Change-point Detection Rule (II)

- Hence the sell signal of the filter rule corresponds to the CUSUM rule in quality control: take corrective action as soon as $S_t - \min_{0 \leq k \leq t} S_k \geq b$; see Lai and Xing (2008, p. 279) and Page (1954).
- Alexander (1961) and Fama and Blume (1966) showed empirically that after taking transaction costs into account, the filter rule did not outperform the buy-and-hold strategy that simply buys the stock and holds it throughout the time period under consideration.
- Lam and Yam (1997) consider more general CUSUM rules and convert them to the filter rule form that involves drawups and drawdowns, and provide empirical evidence of the improvement of these rules over classical filter rules.

Moving Average Rules and Window-Limited GLR Fault Detection Schemes

- Moving average rules: Popular because of their simplicity and intuitive appeal (see Lai and Xing (2008, p. 277), Brock et al. (1992), and Sullivan et al. (1999)).
- Lai (1995, Lemma 2) shows that the moving average rule with window size m can have detection delay that is asymptotically as efficient as the CUSUM rule if m is chosen appropriately.
- Variable-length moving averages are closely related to the window-limited GLR fault detection schemes (when the post-change parameter is not assumed known). See Willsky and Jones (1976) and Lai (1995, Section 3.3).

Directional Trading using Neural Networks (I)

- Gencay (1998): Directional trading strategies whose cumulative log-return for an investment horizon of n days is $S_n = \sum_{t=1}^{n} \hat{y}_t r_t$, where $\hat{y}_t = 1$ (or -1) for a long (or short) position of a stock/security/currency at time t.
- $\hat{y}_t$ chosen is an estimate of the Bernoulli variable $y_t = 2 \times (I_{\{r_t > 0\}} - 1)$ based on observations up to time $t-1$. Gencay (1998) uses a single-layer neural network estimate and cross-validation to determine the number of hidden units in the neural network.
- He also reports an empirical study that shows markedly better performance than the buy-and-hold strategy without considering of the impact of transaction costs and prediction errors $y_t - \hat{y}_t$.

Directional Trading using Neural Networks (II)

- Incorporating these considerations suggests a modified directional trading rule

$$\hat{y}_t = \begin{cases} 1 & \text{if } \hat{\pi}_t^1 \geq q \\ -1 & \text{if } \hat{\pi}_t^2 \geq q \\ 0 & \text{otherwise,} \end{cases}$$

  where $\hat{\pi}_t^1$, or $\hat{\pi}_t^2$, is the estimate of the conditional probability that $r_t \geq c_1$, or $r_t \leq c_2$, respectively, given the information set $\mathscr{F}_{t-1}$ to time $t-1$, and $2q < 1$.
- The conditional probabilities can be estimated by logistic regression; see Section 4.1 of Lai and Xing (2008).

Time Series, Momentum, and Pairs Trading Strategies (I)

- Whereas the (nonparametric) regression approach focuses on patterns of past data with the hope that such patterns can be extrapolated to the future, a suitably chosen time series model can address prediction more directly.
- Cross-sectional momentum strategies: Buying recent "winners" that outperform others and selling recent "losers" that underperform.
- Rationale: Continuation into the future of the relative performance of a group of stock securities.
- Jegadeesh and Titman (1993): Firms with relatively high returns over the past 3 to 12 months continue to outperform firms with relatively lower returns over the same sample period in their empirical study.
- Lewellen (2002): "... suggests that prices are not even weak-form efficient".

Time Series, Momentum, and Pairs Trading Strategies (II)

- Moskowitz et al. (2012): Time-series momentum strategies (which are based on each security's past absolute performance, not relative to those of other securities) performed well relative to cross-sectional momentum strategies.
  - Substantial abnormal returns in a diversified portfolio of time series momentum strategies across all asset classes that performed best during extreme markets.
  - After examining of the trading activities of speculators and hedgers, they conclude "speculators profit from time series momentum at the expense of hedgers".

Contrarian Strategies & Behavioral Finance (I)

- Contrarian strategies monitor markets and investor sentiments to detect cognitive biases prevailing in the market.
- Rationale: The psychological and interpretational aspects of human investment decisions often push prices away from their intrinsic values.
- Psychology, and in particular the herd behavior, is part of human nature.
- The interpretational difficulty of human beings in estimating future company value from annual reports, news and commentaries.
- All these lead to poor estimate of the stock's intrinsic value and may produce periods of undervaluation or overvaluation.

Contrarian Strategies & Behavioral Finance (II)

- Contrarian strategies invest against (contrary to) the market, buying the out-of-favor stocks or shorting the preferred ones, and waiting for price reversals when the market rediscovers value in the out-of-favor stocks or shuns the high-fliers.
- Contrarian strategies are therefore closely related to behavioral finance: a subfield of Finance and of Economics that brings psychological theory and human behavior into financial modeling, predictions and decisions, and economic analysis and policy.
- Behavioral finance provides an alternative view of the market to EMH under which prices should reflect their intrinsic values.

Prospect theory of Kahneman and Tversky (1979)

- Certainty effect
  - "People underweight outcomes that are merely probable in comparison with outcomes that are obtained with certainty.
  - "Risk aversion in choices involving sure gains and to risk-seeking in choices involving sure losses."
- Isolation effect
  - "People generally discard components that are shared by all prospects under consideration."
  - "Leads to inconsistent preferences when the same choice is presented in different forms."
- Prospect theory is "an alternative theory of choice, in which value is assigned to gains and losses rather than to final assets and in which probabilities are replaced by decision weights".
- "The value function is normally concave for gains, commonly convex for losses, and is generally steeper for losses than for gains. Decision weights are generally lower than the corresponding probabilities, except in the range of low probabilities. Overweighting of low probabilities may contribute to the attractiveness of both insurance and gambling."

From Value Investing to Global Macro Strategies (I)

- *Value investing*: Determining the "Intrinsic value" of a listed company via fundamental analysis of the company and its business sector. plays an important role in determining the company's intrinsic value,
- In particular, it involves predicting future cash flows of the company over an investment horizon.
- Such prediction is also related to the analysis of macroeconomic trends.
- The analysis and prediction of global macroeconomic developments feature prominently in *global macro* strategies that invest on a large scale around the world based on these predictions and geopolitical developments including government policies and inter-government relations.
- The analytics component of systematic global macro strategies used by hedge funds involves monitoring interest rate trends, business cycles, the global network of flow of funds, global imbalance patterns, and changing growth models of emerging economies.

From Value Investing to Global Macro Strategies (II)

- Hedge fund strategies can be broadly classified as "discretionary", relying on the skill and experience of the fund managers, and "systematic", relying on quantitative analysis of data and computer models and algorithms to implement the trading positions.
- Many global macro funds trade in the commodities and futures markets.
- Famous example: George Soros is best known for netting $1 billion profit by taking a short position of the pound sterling in 1992 when he correctly predicted that the British government would devalue the pound sterling at the time of the European Rate Mechanism debacle.

Evaluation of Investment Strategies & Multiple Testing (I)

- Data snooping issue in the evaluation of the profitability of trading strategies
  - Bootstrap methods of White (2000) and Hansen (2005)
  - Stepdown bootstrap tests to control the family-wise error rate in multiple testing; seeRomano and Wolf (2005).
  - These methods are useful for in-sample evaluation of trading strategies during their development.
- Out-of-sample (ex-post) evaluation is more definitive but may be too late because they have already been used and resulted in losses.
- An in-sample (ex-ante) surrogate of an ex-post performance measure can be implemented by k-fold cross-validation; see Hastie et al. (2009).

Evaluation of Investment Strategies & Multiple Testing (II)

- The aforementioned bootstrap methods can maintain the family-wise error rate and avoid the pitfalls of data snooping in empirical testing of the profitability of trading strategies. However, there is a lack of systematic simulation study and theoretical development of the power of these tests of a large number of hypotheses.
- Lai and Tsang (2016) have recently filled the gap by developing a comprehensive theory and methodology for efficient post-selection testing.

Evaluation of Investment Strategies & Multiple Testing (III)

- How to evaluate the performance of investment managers/strategies has been a long-standing problem in finance.
- Henriksson and Merton (1981) proposed statistical methods to test for the forecasting skills of market-timers.
- Pesaran and Timmermann (1992) developed a nonparametric test of the forecasting skill in predicting the direction of change of an economic variable under consideration, and applied it in two empirical studies in the British manufacturing sector.
- The above 2 tests are based on the assumption of i.i.d. pairs of forecasts and outcomes.

Evaluation of Investment Strategies & Multiple Testing (IV)

- From our previous discussion about the stylized features of returns, the test statistics should account for the time series effects in returns.
- In addition, more sophisticated forecasts would incorporate uncertainties in the forecasts by giving, for example, the probability of price increase rather than whether the price will increase.
- Such probability forecasts are in fact implicit in the directional trading strategies.
- Testing the skills in probability forecasts when there are also time series effects is considerably more difficult.
- Lai, Gross, and Shen (2011a) have recently made use of the martingale structure implicit in forecasting to resolve these difficulties.

Conclusion

- After the tumultuous period marked by the 2007-2008 Financial Crisis and the Great Recession of 2009, the financial industry has entered a new era.The onset of this era is marked by two "revolutions" that have transformed modern life and business.
- One is technological, dubbed "the FinTech revolution" for financial services by the May 9, 2015, issue of *The Economist* which says: "In the years since the crash of 2007-08, policymakers have concentrated on making finance safer. . . . Away from the regulator spotlight, another revolution is under way. . . . From payments to wealth management, from peer-to-peer lending to crowdfunding, a new generation of startups is taking aim at the heart of the industry – and a pot of revenues that Goldman Sachs estimates is worth $4.7 trillion. . . . fintech firms are growing fast." The other is called "big data revolution".
- Such new technological environment enables swift development in quantitative trading in terms of data analysis, modeling, optimization and strategy evaluation.

References I

S. S. Alexander. Price movements in speculative markets: Trends or random walks. *Industrial Management Review*, 2(2):7–26, 1961.

N. Barberis, M. Huang, and T. Santos. Prospect theory and asset prices. *Quarterly J. Economics*, 116:1–53, 2001.

S. Benartzi and R. H. Thaler. Myopic loss aversion and the equity premium puzzle. *Quarterly J. Economics*, 110(1):73–92, 1995.

G. E. P. Box and G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, 1970.

M. J. Brennan. The role of learning in dynamic portfolio decisions. *European Finance Review*, 1(3):295–306, 1998.

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

References II

H. P. Chan and T. L. Lai. A general theory of particle filters in hidden Markov models and some applications. *Ann. Stat.*, 41(6):2877–2904, 2013.

H. P. Chan and T. L. Lai. MCMC with sequential state substitutions: Theory and applications. Working paper, Department of Statistics, Stanford University, 2016.

P. K. Clark. A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 41(1):135–155, 1973.

A. Cowles. A revision of previous conclusions regarding stock price behavior. *Econometrica*, 28:909–915, 1960.

A. Cowles and H. E. Jones. Some a posteriori probabilities in stock market action. *Econometrica*, 5:280–294, 1937.

J. Cvitanić, A. Lazrak, L. Martellini, and F. Zapatero. Dynamic portfolio choice with parameter uncertainty and the economic value of analysts' recommendations. *Review of Financial Studies*, 19(4):1113–1156, 2006.

References III

J. B. Detemple. Asset pricing in a production economy with incomplete information. *J. Finance*, 41(2):383–391, 1986.

D. diBartolomeo. Smarter rebalancing: Using single period optimization in a multi-period world. Technical report, Northfield Research, 2012.

M. U. Dothan and D. Feldman. Equilibrium interest rates and multiperiod bonds in a partially observable economy. *J. Finance*, 41(2):369–382, 1986.

R. F. Engle. Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation. *Econometrica*, 45:987–1007, 1982.

R. F. Engle, R. Ferstenberg, and J. Russell. Measuring and modeling execution cost and risk. *J. Portfolio Management*, 38(2):14, 2012.

E. F. Fama and M. E. Blume. Filter rules and stock-market trading. *J. Business*, 39(1):226–241, 1966.

N. Gârleanu and L. H. Pedersen. Dynamic trading with predictable returns and transaction costs. *J. Finance*, 68(6):2309–2340, 2013.

References IV

R. Gencay. Optimization of technical trading strategies and the profitability in security markets. *Economics Letters*, 59(2):249–254, 1998.

G. Gennotte. Optimal portfolio choice under incomplete information. *J. Finance*, 41(3):733–746, 1986.

R. C. Grinold and R. N. Kahn. *Active Portfolio Management*. McGraw-Hill, New York, 2nd edition, 2000.

P. R. Hansen. A test for superior predictive ability. *J. Business & Economic Statistics*, 23(4):365–380, 2005.

T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer, New York, 2nd edition, 2009.

R. D. Henriksson and R. C. Merton. On market timing and investment performance. II. Statistical procedures for evaluating forecasting skills. *J. Business*, 54(4):513–533, 1981.

References V

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

H. Jin and X. Y. Zhou. Behavioral portfolio selection in continuous time. *Mathematical Finance*, 18(3):385–426, 2008.

D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, 1979.

M. Kendall. The analysis of economic time-series. Part I: Prices. *J. Roy. Statist. Soc. Ser. A*, 116(1):11–25, 1953.

M. Kritzman, S. Myrgren, and S. Page. Portfolio rebalancing: A test of the Markowitz-Van Dijk heuristic. MIT Sloan Research Paper, 2007.

Y. Kroll, H. Levy, and H. M. Markowitz. Mean-variance versus direct utility maximization. *J. Finance*, 39(1):47–61, 1984.

T. L. Lai. Sequential changepoint detection in quality control and dynamical systems. *J. Roy. Statist. Soc. Ser. B*, 57(4):613–658, 1995. With discussion and a reply by the author.

References VI

T. L. Lai and V. Bukkapatanam. Adaptive filtering, nonlinear state-space models, and applications to finance and econometrics. In Y. Zeng and S. Wu, editors, *State-Space Models: Applications in Economics and Finance*, Statistics and Econometrics for Finance, Vol. 1, pages 3–22. Springer, New York, 2013.

T. L. Lai and K. W. Tsang. Post-selection multiple testing and a new approach to test-based variable selection. Working paper, Department of Statistics, Stanford University, 2016.

T. L. Lai and C. Z. Wei. Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Ann. Stat.*, 10(1):154–166, 1982.

T. L. Lai and C. Z. Wei. Extended least squares and their applications to adaptive control and prediction in linear systems. *IEEE Trans. Automat. Control*, 31(10):898–906, 1986.

References VII

T. L. Lai and H. Xing. *Statistical Models and Methods for Financial Markets*. Springer Texts in Statistics. Springer, New York, 2008.

T. L. Lai and H. Xing. *Risk Analytics and Management in Finance and Insurance*. Chapman & Hall/CRC, 2016. forthcoming.

T. L. Lai and Z. Ying. Recursive identification and adaptive prediction in linear stochastic systems. *SIAM J. Control Optim.*, 29(5):1061–1090, 1991.

T. L. Lai, S. T. Gross, and D. B. Shen. Evaluating probability forecasts. *Ann. Statist.*, 39(5):2356–2382, 2011a.

T. L. Lai, H. Xing, and Z. Chen. Mean-variance portfolio optimization when means and covariances are unknown. *Ann. Appl. Statist.*, 5(2A): 798–823, 2011b.

K. Lam and H. Yam. Cusum techniques for technical trading in financial markets. *Financial Engineering and the Japanese Markets*, 4(3): 257–274, 1997.

References VIII

O. Ledoit and M. Wolf. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *J. Empirical Finance*, 10(5):603–621, 2003.

O. Ledoit and M. Wolf. Honey, I shrunk the sample covariance matrix. *J. Portfolio Management*, 30(4):110–119, 2004.

H. Levy and H. M. Markowitz. Approximating expected utility by a function of mean and variance. *The American Economic Review*, 69(3): 308–317, 1979.

J. Lewellen. Momentum and autocorrelation in stock returns. *Review of Financial Studies*, 15(2):533–564, 2002.

D. Li and W.-L. Ng. Optimal dynamic portfolio selection: Multiperiod mean-variance formulation. *Mathematical Finance*, 10(3):387–406, 2000.

References IX

A. W. Lo, H. Mamaysky, and J. Wang. Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *J. Finance*, 55(4):1705–1770, 2000.

B. G. Malkiel. *A Random Walk Down Wall Street: The Time-tested Strategy for Successful Investing*. W.W. Norton & Company, 2003.

B. Mandelbrot. The variation of certain speculative prices. *J. Business*, 36 (4):394–419, 1963.

B. Mandelbrot and H. M. Taylor. On the distribution of stock price differences. *Operations Research*, 15(6):1057–1062, 1967.

H. M. Markowitz and E. L. van Dijk. Single-period mean-variance analysis in a changing world (corrected). *Finan. Analysts J.*, 59(2):30–44, 2003.

R. C. Merton. Lifetime portfolio selection under uncertainty: The continuous-time case. *The Review of Economics and Statistics*, 51(3): 247–257, 1969.

References X

R. C. Merton. Optimum consumption and portfolio rules in a continuous-time model. *J. Economic Theory*, 3(4):373–413, 1971.

R. C. Merton. *Continuous-Time Finance*. Macroeconomics and Finance Series. Basil Blackwell, 1990.

R. O. Michaud. *Efficient Asset Management*. Harvard Business School Press, 1989.

T. J. Moskowitz, Y. H. Ooi, and L. H. Pedersen. Time series momentum. *J. Finan. Econom.*, 104(2):228–250, 2012.

M. F. M. Osborne. Brownian motion in the stock market. *Operations Research*, 7:145–173, 1959.

E. Page. Continuous inspection schemes. *Biometrika*, 41(1/2):100–115, 1954.

M. H. Pesaran and A. Timmermann. A simple nonparametric test of predictive performance. *J. Business & Economic Statistics*, 10(4): 461–465, 1992.

References XI

S. Pliska. *Introduction to Mathematical Finance*. Blackwell Publishers, Oxford, 1997.

J. P. Romano and M. Wolf. Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4):1237–1282, 2005.

P. A. Samuelson. Proof that properly anticipated prices fluctuate randomly. *Industrial Management Review*, 6(2):41–49, 1965.

R. Sullivan, A. Timmermann, and H. White. Data-snooping, technical trading rule performance, and the bootstrap. *J. Finance*, 54(5): 1647–1691, 1999.

H. White. A reality check for data snooping. *Econometrica*, 68(5): 1097–1126, 2000.

J. T. Williams. Capital asset prices with heterogeneous beliefs. *J. Finan. Econom.*, 5(2):219–239, 1977.

References XII

A. S. Willsky and H. L. Jones. A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems. *IEEE Trans. Automat. Contr.*, 21(1):108–112, 1976.

H. Working. A note on the correlation of first differences of averages in a random chain. *Econometrica*, 28:323–362, 1960.

Y. Xia. Learning about predictability: The effects of parameter uncertainty on dynamic asset allocation. *J. Finance*, 56(1):205–246, 2001.

X. Y. Zhou and D. Li. Continuous-time mean-variance portfolio selection: A stochastic LQ framework. *Appl. Math. Optimiz.*, 42(1):19–33, 2000.