Today’s Agenda

1. Empirical Portfolio Choice
2. Bayesian Estimation
3. Black-Litterman—Introduction
   (a) Fundamental Indexation
4. Parametric Portfolio Weights
5. Market Microstructure
6. Bootstrap
1 Empirical Portfolio Choice–Mean-Variance Implementation

- The solution to the mean-variance problem:
  \[
  \min_x \text{var} \left( r_{p,t+1} \right) = x'\Sigma x \\
  \text{s.t.} \quad E(r_p) = x'\mu = \bar{\mu}
  \]
  is
  \[
  x^* = \frac{\bar{\mu}}{\mu'\Sigma\mu} \times \Sigma^{-1}\mu = \lambda\Sigma^{-1}\mu
  \]

- Now, we have to rely on econometrics, to implement the solution.

- Two step approach:
  - Solve the economic model
  - Estimate the parameters and plug them in!
• **PLUG-IN APPROACH:**

• We continue with the assumption that returns are i.i.d.

• Then, we can estimate

\[ \hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} r_{t+1} \]

\[ \hat{\Sigma} = \frac{1}{T - N - 2} \sum_{t=1}^{T} (r_{t+1} - \hat{\mu})(r_{t+1} - \hat{\mu})' \]

• We plug in the estimates into the optimal solution

\[ \hat{x}^* = \frac{1}{\gamma} \hat{\Sigma}^{-1} \hat{\mu} \]

• Under the normality assumption, this estimator is unbiased, or

\[ E(\hat{x}^*) = \frac{1}{\gamma} E(\hat{\Sigma}^{-1}) E(\hat{\mu}) \]

• In the univariate case we can show by the delta method that

\[ Var(\hat{x}^*) = \frac{1}{\gamma^2} \left( \frac{\mu}{\sigma^2} \right)^2 \left( \frac{var(\hat{\mu})}{\mu^2} + \frac{var(\hat{\sigma}^2)}{\sigma^4} \right) \]
• Example

• Suppose we have 10 years of monthly data, or $T = 120$.

• Suppose we have a stock with $\mu = 0.06$ and $\sigma = 0.15$.

• Suppose that $\gamma = 5$.

• Note that

$$\hat{x}^* = \frac{1}{\gamma \sigma} \mu = \frac{0.06}{5 \times 0.15^2} = 0.533$$

• Very close to the usual 60/40 advice by financial advisors!

• With i.i.d. returns, the standard errors of the mean and variance are

$$\text{var}(\hat{\mu}) = \frac{\sigma}{\sqrt{T}} = \frac{0.15}{\sqrt{120}} = 0.014$$

$$\text{var}(\hat{\sigma}^2) = \sqrt{2} \frac{\sigma^2}{\sqrt{T}} = \sqrt{2} \frac{0.15^2}{\sqrt{120}} = 0.003$$

• Plugging all these in the formula for $\text{Var} (\hat{x}^*)$, we obtain

$$\text{Var} (\hat{x}^*) = 0.14$$

• We can test hypotheses as with every other parameter of interest.
• Estimating $\Sigma$ is very problematic

• Many parameters to estimate
  – Suppose we have 500 assets in the portfolio. We have 125,250 unique elements to estimate.
  – In general, for N assets, we have $N(N + 1)/2$ unique elements to estimate!

• We need $\Sigma^{-1}$. Small estimation errors $\hat{\Sigma}$ results in very different $\hat{\Sigma}^{-1}$.

• Solution: Shrink the matrix

\[
\hat{\Sigma}^s = \delta S + (1 - \delta) \hat{\Sigma}
\]

where

\[
\delta \approx \frac{1}{T} \frac{A - B}{C}
\]

\[
A = \sum_i \sum_j \text{asy var} \left( \sqrt{T} \hat{\sigma}_{i,j} \right)
\]

\[
B = \sum_i \sum_j \text{asy cov} \left( \sqrt{T} \hat{\sigma}_{i,j}, \sqrt{T} s_{i,j} \right)
\]

\[
C = \sum_i \sum_j \left( \hat{\sigma}_{i,j} - s_{i,j} \right)^2
\]

where $S$ is often taken to be $I$. For more discussions, see Ledoit and Wolf (2003)
• We can also shrink the weights directly

\[ x^s = \delta x_0 + (1 - \delta) x^* \]

• This approach is often used in applied work.

• Problem with shrinkage: Ad-hoc. No economic justification for it or for \( \delta \).

• Bayesian framework

• Economic constraints [Jagannathan and Ma (JF, 2003)]
• Another solution: Factor models for stock \( i \)

\[
r_{i,t} = \alpha_i + \beta_i f_m + \varepsilon_{i,t}
\]

• We can take variances to show that

\[
\Sigma_r = \sigma_m^2 \beta \beta' + \Sigma_r
\]

where \( \beta \) is a vector of the betas and \( \Sigma_r \) is a diagonal matrix with diagonal elements the variances of \( \varepsilon_{i,t} \).

• Now, the problem is reduced significantly!

• What about time variation in \( \mu \) and \( \Sigma \)!
2 Bayesian Estimation.

- Premise: We are not estimating parameter values, but rather always updating and sharpening our subjective beliefs about the state of the world.

- The centerpiece of the Bayesian methodology is Bayes’ theorem:

\[
P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}
\]

- In terms of the applications at hand, we are interested in the value of the parameters, given the data.

- We can write

\[
P(\text{Parameters}|\text{Data}) = \frac{P(\text{Data}|\text{Parameters})P(\text{Parameters})}{P(\text{Data})}
\]
We will view the data as not depending on the parameters, or we can further write

\[ P(\text{Parameters}|\text{Data}) \propto P(\text{Data}|\text{Parameters}) \cdot P(\text{Parameters}) \]

The terms are:
- \( P(\text{Data}|\text{Parameters}) \): Density of the data, given the parameters
- \( P(\text{Parameters}) \): Prior density of the parameters—Prior belief of the econometrician
- \( P(\text{Parameters}|\text{Data}) \): Posterior density of the parameters, given the data—It is a mixture of the prior and the “current information” from the data.

We must specify the prior and the form of the data in order to get the posterior.

Once we get more data, the posterior becomes the prior and we update again.
The calculations involved in Bayesian analysis might become quite burdensome, but here we will present a simple example.

Suppose that the data is
\[ f(y|\beta, \sigma, X) = N(X\beta, \sigma^2) \]

We also assume an “uninformative” prior
\[ f(\text{Parameters}) \propto \text{constant} \]

For simplicity, we assume for now that \( \sigma \) is known.

We can write
\[
\begin{align*}
y - X\beta &= y - Xb - X(\beta - b) \\
b &= (X'X)^{-1} X'y
\end{align*}
\]
Then, the posterior is

\[
    f(\beta | \beta, \sigma, X) = h(\sigma^2) \times (2\pi \sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta) \right\} \\
    = (2\pi \sigma^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma^2} (y - Xb)'(y - Xb) - \frac{1}{2\sigma^2} (\beta - b)'X'X(\beta - b) \right\} \\
    \propto \sigma^2 \left| (X'X)^{-1} \right|^{-1/2} \exp \left\{ -\frac{1}{2} (\beta - b)' \left( \sigma^2 (X'X)^{-1} \right)^{-1} (\beta - b) \right\}
\]

In other words, the density is normal with mean \( \beta \) and variance matrix \( \sigma^2 (X'X)^{-1} \).

We can do the similar calculations when we impose another prior on \( \sigma \).

The results would change. The distribution of \( \beta \) would be a multivariate-t which is another familiar result.

Note: We had to specify the priors and the distribution of the data. If we change any of these two assumptions, the results would change.

Note: We obtain exact results, because me make distributional assumptions on the data.

We did not need such assumptions in the frequentist world. Why?
3 Introduction to Black-Litterman Model


- Premise: In the Markowitz portfolio optimization results

\[ w_p = \frac{1}{\gamma} V^{-1} \mu \]

we need estimates of \( V^{-1} \) and \( \mu \).

- Suppose we have \( n \) assets. Therefore \( \mu \) is \( n \) by 1 vector and \( V \) is \( n \) by \( n \) matrix.

- We can estimate \( V \) with some degree of certainty, especially if we have high-frequency observations.

- The mean \( \mu \) is harder to estimate.

- We want to introduce economic prior beliefs about \( \mu \).

- For instance, suppose we believe that sector \( A \) will out-perform sector \( B \) by 1 percent. How can we incorporate that information with the rest of the model?

- Can you think of a few ways to specify \( \mu \) without using sample averages?
• Suppose we know $V$ (or can estimate it)
• Suppose we are not sure about $\mu$ but believe that it has a normal distribution
  \[ \mu \sim N(\pi, \Sigma) \]
• $\mu$ is the expected return that is uncertain!
• Note that $V$ and $\Sigma$ are different concepts.
  – $V$ is the estimated variance of realized returns.
  – $\Sigma$ is the posited variance of expected returns
    (specified by economist)
• We have views about expected returns, captured by
  \[ P\mu = Q - \varepsilon \]
  where
  – $P$ is a $k$ by $n$ matrix
  – $Q$ is an $n$ by 1 vector
  – $\varepsilon \sim N(0, \Omega)$
• We will need to specify the parameters $P$, $Q$, and $\Omega$.
• $P$ takes $k$ linear combinations of the expected returns.
• $Q$ expresses the new views
• $\Omega$ expresses uncertainty about the new views (if $\Omega = 0$, no uncertainty)
Black and Litterman (1990) show that
\[ E(\mu|views) = [\Sigma^{-1} + P\Omega^{-1}P]^{-1}[\Sigma^{-1}\pi + P\Omega^{-1}Q]. \]

Or equivalently
\[ E(\mu|views) = \pi + \Sigma P'[P\Sigma P' + \Omega]^{-1}[Q - P\pi] \]

If we are not uncertain about our views, then we have
\[ E(\mu|views) = \pi + \Sigma P'[P\Sigma P']^{-1}[Q - P\pi] \]

This is the posterior expected return of the mean, given the beliefs!

This formula can also be seen as a updating step in a KF exercise. Recall that the updating step in KF is
\[ z_{t|t} = z_{t|t-1} + P_{t|t-1}H (H'P_{t|t-1}H + R)^{-1} (y_t - A'x_t - H'z_{t|t-1}) \]

Similar setup: We have an unobservable quantity \( z_{t|t-1} \) or \( \pi \).

Insight: If we treat expected returns as unobservable then we are in situation similar to KF!
• Example:

\[ \mu = N \left( \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 9.1 & 3 & 6 \\ 3 & 1.1 & 2 \\ 6 & 2 & 4.1 \end{bmatrix} \right) \]

• The investor believes that the first asset will outperform the second by 2%. This view is purely subjective. Prior!

• This view can be captured by \( P = [1 \ -1 \ 0] \), \( Q = [2] \). Or

\[ P\mu = Q - \varepsilon \]
\[ \mu_1 - \mu_2 = 0.02 - \varepsilon \]

• The investor is not terribly certain about her prior, or \( \varepsilon \sim N(0,1) \). Hence, \( \Omega = 1 \).

• Evaluating using the above formulas, we obtain

\[ E(\mu|\text{views}) = \begin{bmatrix} 3.3462 \\ 1.7308 \\ 2.5385 \end{bmatrix} \]

• Notice that the difference in expected returns between the first two assets is not exactly 2 (because of the uncertainty).
• If \( \Omega = 0 \), then the difference is exactly 2

\[
E(\mu|\text{views}) = \begin{bmatrix}
3.9048 \\
1.9048 \\
2.9048
\end{bmatrix}
\]

• We can plot the expected returns as a function of \( \Omega \):

![Graph showing expected returns as a function of Omega](image)

• For \( \Omega \to \infty \), we have diffuse (uninformative) prior—back to the initial (prior) expected returns and allocations!

• This is the model for expected returns that we will use in evaluating the portfolio weights.

• The \( P \), \( Q \), and \( \Omega \) have to be specified by the investor, based on prior beliefs!

• In sum, this approach shrinks (or imposes priors on) expected returns!
Alternative approach: Shrinking the Covariance Matrix

Ledoit and Wolf (2003)

You have a sample covariance matrix $S$

We observe returns $R$ and factors, $f$

You have a model for returns

$$R_i = \beta F + \varepsilon$$

$$\Sigma_r = \beta' \Sigma_f \beta + \Sigma_\varepsilon$$

We can estimate $\beta$, $\Sigma_r$, $\Sigma_f$, and $\Sigma_\varepsilon$.

Goal: Shrink $\Sigma_r$ toward $\beta' \Sigma_f \beta$

We can use

$$\Sigma_{shrunken} = (1 - \delta) \hat{\Sigma}_r + \delta \left( \beta' \Sigma_f \beta \right)$$

The shrinkage parameter $\delta$ is quite arbitrary—see Ledoit and Wolf (2003)

Comments about shrinkages: Tradeoffs
- Robustness vs dynamics
- Priors vs data
4 Holding the Market vs Fundamental Indexation

- Under certain assumptions, holding the market yields the highest Sharpe ratio.
- Holding the market implies holding each asset with a weight proportional to its size (market capitalization) relative the market portfolio.
- Recently, a few people have proposed an alternative idea, called fundamental indexation: Research affiliates (Rob Arnott)
- We can create weights based on:
  - book value of assets
  - NOI
  - Revenues
  - Sales
  - Dividends
  - Employment
  - Combination
- This portfolio will have different properties from the market portfolio.
- Tilts the market in certain dimensions
- Results: Arnott, Hsu, and Moore (FAJ, 2005)
Criticising Fundamental Indexation
- Is it value creation—\(\alpha\)—or risk—\(\beta\)? Jun and Malkiel evidence
- Which weight categorization to choose?
- Under the CAPM assumptions, neither categorization is optimal.
- Turnover and transaction costs?
- What is the theoretical foundation for fundamental indexation

Note that fundamental indexation based on book to value and size is not the same as Fama-French, 3-factor model. Why?
- Hint: If expected returns are driven by a three factor model, what is the optimal portfolio to hold, in equilibrium?

A lot of interest in the profession.
5 Parameteric Portfolio Weights
Typical Portfolio Choice Problem

- Find portfolio weights \((w_{i,t})\) that maximize utility of representative investor (given conditional information)
  - Impose simplifying assumptions on utility, horizon, conditioning information, etc.

- Distribution of returns \((r_{i,t})\) is specified.

- Solve for the optimal portfolio weights:
  \[ w_t = w(\gamma, x_t, \Delta) \]
  - \(\gamma\) are the preference parameters (e.g., risk aversion parameter for CRRA utility)
  - \(x_t\) is a vector of state variables
  - \(\Delta\) are the parameters of the data generating process.
Traditional Approach to Portfolio Problem

- Model expected excess returns of stocks as function of characteristics:
  - cross-sectional regressions
  - characteristics-based sorts

- Model *unconditional* covariance matrix of stocks:
  - large number of \((N_t^2 + N_t)/2\) elements
  - hard to guarantee matrix is well-conditioned
  - very hard to incorporate conditioning information
  - higher moments and cross moments ignored

- Markowitz portfolio solution \(w \propto \Sigma^{-1}\mu\) with exploding portfolio weights:
  - error maximization (Michaud, 1989)

- Common fixes:
  - factor models (e.g., BARRA or Northfield)
  - shrinkage (e.g., Jorion, 1986, Black and Litterman, 1992)
  - portfolio constraints (e.g., Jagannathan and Ma 2003)
Main Idea

Specify the portfolio weight function as:

\[ w_{i,t} = f(x_{i,t}; \theta) \]

where

- \( f(\cdot; \cdot) \) is a known function
- \( x_{i,t} \) is a vector of conditioning information (possibly firm specific)
- \( \theta \) are parameters to be estimated
Parameterized Portfolio Weights

- Portfolio weight function:

\[
\begin{align*}
w_{i,t} & = f(x_{i,t}; \theta) \\
& = \tilde{w}_{i,t} + \theta^\top x_{i,t}/N_t \\
& = \tilde{w}_{i,t} + (\theta_{me} \times me_{i,t} + \theta_{btm} \times btm_{i,t} + \theta_{mom} \times mom_{i,t})/N_t
\end{align*}
\]

where:

- \(\tilde{w}_{i,t}\) = weight of stock \(i\) in the benchmark (or market) portfolio at time \(t\)
- \(N_t\) = number of stocks at time \(t\)
- \(\theta\) = vector of coefficients to be estimated

- Vector of characteristics for each firm \(x_{i,t}\) normalized across stocks (mean zero and standard deviation one) at each time \(t\). Example:

  - market capitalization, \(me\)
  - the book-to-market ratio, \(btm\)
  - lagged twelve-month return, \(mom\)
Given the portfolio weight function, we compute portfolio returns:

\[ r_{p,t+1} = \sum_{i=1}^{N_i} w_{i,t} r_{i,t+1} = \sum_{i=1}^{N_i} (\bar{w}_{i,t} + \theta^\top x_{i,t}/N_t) r_{i,t+1} \]

\[ = r_{m,t+1} + r_{a,t+1} \]

where

- \( w_{i,t} \) is a function of predetermined variables
- \( r_{m,t+1} \) is the market (or some benchmark) portfolio return
- \( r_{a,t+1} \) is the return of an actively managed (hedge) portfolio
Example

• MSFT circa 2004:
  – MSFT has $me = 4$, $btm = -1$, and $mom = 1$
  – The weight of MSFT in the market portfolio is 0.02
  – The coefficients are $\theta_{me} = -1.2$, $\theta_{btm} = 3.5$, and $\theta_{mom} = 2.0$
  – There are currently 6,000 stocks in the investable universe

• Then, the weight on MSFT would be:

$$w_{MSFT,04} = 0.02 + ((-1.2) \times 4 + 3.5 \times (-1) + 2.0 \times 1)/6,000 = 0.0189$$

• Decrease our weight on MSFT by 0.0011

• Going through the data, we can calculate the weight on MSFT at the beginning of each month using only information available at that time
Comments on the Parameterization

• Recall:

\[ w_{i,t} = \bar{w}_{i,t} + \theta^\top x_{i,t}/N_t \]

\[ = \bar{w}_{i,t} + (\theta_{me} \times me_{i,t} + \theta_{btm} \times btm_{i,t} + \theta_{mom} \times mom_{i,t}) / N_t \]

• Coefficients \( \theta \) are constant across stocks and through time:
  
  – portfolio policy only cares about characteristics, not stocks per se
  
  – unconditional portfolio policy instead of conditional portfolio weights

• Standardized characteristics \( \Rightarrow \theta^\top x_{i,t}/N_t \) sum to zero across stocks:
  
  – deviations from benchmark (active portfolio management)
  
  – portfolio weights are guaranteed to sum to one

• Normalization \( 1/N_t \) allows number of firms to change over time
Estimation of $\theta$’s in Investment Function

• Find $\theta$ that maximizes expected utility:
  \[
  \max_\theta \mathbb{E} \left[ u \left( r_{p,t+1} \right) \right]
  \]

• where the portfolio returns are
  \[
  r_{p,t+1} = \sum_{i=1}^{N_t} w_{i,t} r_{i,t+1} = \sum_{i=1}^{N_t} \left( \bar{w}_{i,t} + \theta^\top x_{i,t}/N_t \right) r_{i,t+1}
  \]

• Estimated by maximizing average realized utility in sample:
  \[
  \max_\theta \frac{1}{T} \sum_{t=1}^{T} u \left( r_{p,t+1} \right)
  \]

• In the case of power utility with relative risk aversion $\gamma$:
  \[
  \max_\theta \frac{1}{T} \sum_{t=1}^{T} \frac{(1 + r_{p,t+1})^{1-\gamma}}{1 - \gamma}
  \]
Comments on the Estimation

• The estimator

\[ \hat{\theta} = \arg \min \frac{1}{T} \sum_{t=1}^{T} u(r_{p,t+1}) \]

is an extremum estimator (Amemiya (1985))

• We can use the standard GMM toolbox

• Portfolio optimization framed as statistical estimation problem:
  – “maximum expected utility estimator”
  – statistical and economic objective functions are aligned

• Optimization takes into account all moments of portfolio returns:

\[ \mathbb{E}[u(r_{p,t+1})] \approx u(\mathbb{E}[r_{p,t+1}]) + u''(\mathbb{E}[r_{p,t+1}]) \text{Var}[r_{p,t+1}]/2 + u'''(\mathbb{E}[r_{p,t+1}]) \text{Skew}[r_{p,t+1}]/6 + ... \]

which depend implicitly on all the moments of the stock returns
Comments on the Estimation (Cont’d)

• Analytic derivatives are straight forward to derive

• Optimization is easy and fast:
  – small number of parameters (3 parameters ⇒ a few seconds on a laptop)
  – Bootstrapping is feasible

• Optimization is numerically robust and stable:
  – if you pick a stock, you pick all other stocks with similar characteristics
  – optimizer cannot game “wiggles” in the data

• Weakest link:
  – Have to specify utility function
  – Have to specify the investment policy function
Inference

• The GMM estimator $\hat{\theta}$ satisfies the first-order conditions:

$$
H \equiv \frac{1}{T} \sum_{t=0}^{T-1} h(r_{t+1}, x_t; \hat{\theta}) \equiv \frac{1}{T} \sum_{t=0}^{T-1} u'(r_{p,t+1})(x_t^T r_{t+1}) / N_t = 0
$$

• GMM estimator is consistent, asymptotically normal, with asymptotic covariance matrix:

$$
\text{AsyVar}[\hat{\theta}] = [G^T V^{-1} G]^{-1}
$$

$$
G \equiv \frac{1}{T} \sum_{t=0}^{T-1} \frac{\partial h(r_{t+1}, x_t; \theta)}{\partial \theta} = \frac{1}{T} \sum_{t=0}^{T-1} u''(r_{p,t+1})(x_t^T r_{t+1})(x_t^T r_{t+1})^T / N_t^2
$$

and $V$ is an estimator of the covariance matrix of $h(r, x; \theta)$ like:

$$
\frac{1}{T} \sum_{t=0}^{T-1} h(r_{t+1}, x_t; \hat{\theta}) h(r_{t+1}, x_t; \hat{\theta})^T
$$

• Standard errors for $\hat{\theta}$, hypothesis tests, selection of conditioning variables
Test of Equilibrium in the Stock Market

- In equilibrium, the representative investor chooses to hold the market portfolio with weights $\bar{w}$ (this is true in CAPM, equilibrium APT, CCAPM)

- In our formulation, when $\theta = 0$, the market portfolio is optimal

- Test equilibrium conditions for a given representative investor by:
  - testing whether $\theta = 0$
  - testing whether the expected utility of the optimum portfolio is equal to the expected utility of the market (“expected utility ratio” test)

- Search over different preference specifications (e.g. different levels of $\gamma$) to see if $\theta = 0$ can be satisfied for any representative investor

- Since the optimal portfolio policy is likely to be misspecified (missing variables, wrong functional form), the test is conservative

- The test does not require modeling the data generating process
Empirical Application

- CRSP/COMPSTAT merged database 1962:7 to 2002:12
- Eliminate smallest 20% of companies
- Eliminate companies with negative book value of equity
- Number of stocks varies from about 2,000 to more than 6,000
Mean and Standard Deviation of Characteristics

![Graphs showing the mean and standard deviation of various characteristics over years from 1970 to 2000.](image-url)
## Base Case

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value Weight</th>
<th>Optimum</th>
<th>Value Weight</th>
<th>Optimum</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\theta_{me}$</td>
<td>(0.547)</td>
<td>1.220</td>
<td></td>
<td>1.389</td>
</tr>
<tr>
<td>$\theta_{btm}$</td>
<td>(0.922)</td>
<td>3.466</td>
<td></td>
<td>1.557</td>
</tr>
<tr>
<td>$\theta_{mom}$</td>
<td>(0.742)</td>
<td>2.000</td>
<td></td>
<td>1.709</td>
</tr>
<tr>
<td>$me$</td>
<td>2.095</td>
<td>-0.118</td>
<td>1.527</td>
<td>-0.136</td>
</tr>
<tr>
<td>$btm$</td>
<td>-0.451</td>
<td>3.296</td>
<td>-0.240</td>
<td>1.430</td>
</tr>
<tr>
<td>$mom$</td>
<td>0.012</td>
<td>1.839</td>
<td>-0.026</td>
<td>1.622</td>
</tr>
</tbody>
</table>

- Tilt toward small, value, momentum stocks
Portfolio Characteristics \( \left( \sum_i w_{i,t} x_{i,t} \right) \)
## Base Case

<table>
<thead>
<tr>
<th>Variable</th>
<th>All Stocks</th>
<th>Top 500 Stocks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Value Weight</td>
<td>Optimum</td>
</tr>
<tr>
<td>$</td>
<td>w_i</td>
<td>\times 100$</td>
</tr>
<tr>
<td>$\max w_i \times 100$</td>
<td>4.625</td>
<td>4.363</td>
</tr>
<tr>
<td>$\min w_i \times 100$</td>
<td>0.000</td>
<td>-0.323</td>
</tr>
<tr>
<td>$\sum I(w_i &lt; 0)/N_T$</td>
<td>0.000</td>
<td>0.468</td>
</tr>
<tr>
<td>$\bar{r}$</td>
<td>0.120</td>
<td>0.244</td>
</tr>
<tr>
<td>$\sigma(r)$</td>
<td>0.161</td>
<td>0.190</td>
</tr>
<tr>
<td>$\text{CE}(r)$</td>
<td>0.052</td>
<td>0.153</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>—</td>
<td>0.157</td>
</tr>
<tr>
<td>$\beta$</td>
<td>—</td>
<td>0.440</td>
</tr>
<tr>
<td>$\sigma(\epsilon)$</td>
<td>—</td>
<td>0.177</td>
</tr>
<tr>
<td>$\text{IR}$</td>
<td>—</td>
<td>0.890</td>
</tr>
</tbody>
</table>
Decomposing the Portfolio Return

• Recall that

\[ r_{p,t+1} = r_{m,t+1} + r_{a,t+1} \]
\[ = 0.120 + 0.124 \]

• The return \( r_{a,t+1} \) is due to two effects: (i) Taking short positions; (ii) Using the leverage from short positions to invest in long positions

• We can decompose the hedge portfolio as

\[ r_{a,t+1} = q(r_{a,t+1}^+ - r_{a,t+1}^-) \]

- \( r_{a,t+1}^+ \) is the return on the long part of the portfolio
- \( r_{a,t+1}^- \) is the return on the short part of the portfolio
- \( q \) is the leverage of the long-short portfolio

\[ r_{a,t+1} = 1.7(.190 - .118) \]
Other Specifications of the Portfolio Policy Function

• Portfolio weight constraints, e.g., no short-sales:
  
  \[
  w_{i,t}^+ = \frac{\max[0, w_{i,t}]}{\sum_{j=1}^{N_t} \max[0, w_{j,t}]}
  \]

• Interaction of characteristics (e.g., “growth stocks with momentum”)

• Time-varying coefficients:
  
  \[
  \theta_t = \theta^\top z_t,
  \]
  
  where \(z_t\) are business cycle indicators (e.g., interest rate spreads)

• Transactions costs

• Screens, e.g., sort stocks on \(\theta^\top x_{i,t}\) and equal weight top quantile
### No Short Sales

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>Weight</th>
<th>Unconstrained</th>
<th>Constrained</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_{me}$</td>
<td>—</td>
<td>—</td>
<td>-1.220</td>
<td>-0.182</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.547)</td>
<td>(1.037)</td>
</tr>
<tr>
<td>$\theta_{btm}$</td>
<td>—</td>
<td>—</td>
<td>3.466</td>
<td>2.043</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.922)</td>
<td>(1.113)</td>
</tr>
<tr>
<td>$\theta_{mom}$</td>
<td>—</td>
<td>—</td>
<td>2.000</td>
<td>2.358</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.742)</td>
<td>(1.199)</td>
</tr>
<tr>
<td>$me$</td>
<td>2.095</td>
<td></td>
<td>-0.118</td>
<td>0.632</td>
</tr>
<tr>
<td>$btm$</td>
<td>-0.451</td>
<td></td>
<td>3.296</td>
<td>0.451</td>
</tr>
<tr>
<td>$mom$</td>
<td>0.012</td>
<td></td>
<td>1.839</td>
<td>0.886</td>
</tr>
</tbody>
</table>

- With short-sales constraint, no $me$, $btm$, and $mom$ effects!

- These anomalies may be due to short-sales contraints
## No Short Sales

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>Weight</th>
<th>Unconstrained</th>
<th>Constrained</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>w_i</td>
<td>\times 100$</td>
<td>0.035</td>
<td>0.121</td>
</tr>
<tr>
<td>max $w_i \times 100$</td>
<td>4.625</td>
<td>4.363</td>
<td>2.455</td>
<td></td>
</tr>
<tr>
<td>min $w_i \times 100$</td>
<td>0.000</td>
<td>-0.323</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>$\sum I(w_i \leq 0)/N_T$</td>
<td>0.000</td>
<td>0.468</td>
<td>0.439</td>
<td></td>
</tr>
<tr>
<td>$\bar{r}$</td>
<td>0.120</td>
<td>0.244</td>
<td>0.173</td>
<td></td>
</tr>
<tr>
<td>$\sigma(r)$</td>
<td>0.161</td>
<td>0.190</td>
<td>0.180</td>
<td></td>
</tr>
<tr>
<td>CE($r$)</td>
<td>0.052</td>
<td>0.153</td>
<td>0.085</td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>—</td>
<td>0.157</td>
<td>0.054</td>
<td></td>
</tr>
<tr>
<td>$\beta$</td>
<td>—</td>
<td>0.440</td>
<td>0.977</td>
<td></td>
</tr>
<tr>
<td>$\sigma(\epsilon)$</td>
<td>—</td>
<td>0.177</td>
<td>0.088</td>
<td></td>
</tr>
<tr>
<td>IR</td>
<td>—</td>
<td>0.890</td>
<td>0.620</td>
<td></td>
</tr>
</tbody>
</table>

- Constrained can exploit only positive positions
- Unconstrained portfolio can benefit from leverage from short position
## Interaction Between Characteristics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>Weight</th>
<th>Unconstrained</th>
<th>Constrained</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_{me}$</td>
<td>—</td>
<td>-0.744</td>
<td>-0.836</td>
<td></td>
</tr>
<tr>
<td>$\theta_{btm}$</td>
<td>—</td>
<td>4.444</td>
<td>5.596</td>
<td></td>
</tr>
<tr>
<td>$\theta_{mom}$</td>
<td>—</td>
<td>4.631</td>
<td>4.565</td>
<td></td>
</tr>
<tr>
<td>$\theta_{me\times btm}$</td>
<td>—</td>
<td>-1.727</td>
<td>3.325</td>
<td></td>
</tr>
<tr>
<td>$\theta_{me\times mom}$</td>
<td>—</td>
<td>4.718</td>
<td>5.346</td>
<td></td>
</tr>
<tr>
<td>$\theta_{btm\times mom}$</td>
<td>—</td>
<td>4.249</td>
<td>4.000</td>
<td></td>
</tr>
<tr>
<td>me</td>
<td>2.095</td>
<td>-0.003</td>
<td>0.369</td>
<td></td>
</tr>
<tr>
<td>btm</td>
<td>-0.451</td>
<td>4.714</td>
<td>0.684</td>
<td></td>
</tr>
<tr>
<td>mom</td>
<td>0.012</td>
<td>1.446</td>
<td>0.620</td>
<td></td>
</tr>
<tr>
<td>me $\times$ btm</td>
<td>-1.142</td>
<td>-5.398</td>
<td>-0.254</td>
<td></td>
</tr>
<tr>
<td>me $\times$ mom</td>
<td>0.017</td>
<td>1.864</td>
<td>0.181</td>
<td></td>
</tr>
<tr>
<td>btm $\times$ mom</td>
<td>-0.014</td>
<td>3.281</td>
<td>0.718</td>
<td></td>
</tr>
</tbody>
</table>
## Interaction Between Characteristics

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>Weight</th>
<th>Unconstrained</th>
<th>Constrained</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>w_i</td>
<td>\times 100$</td>
<td>0.035</td>
<td>0.186</td>
</tr>
<tr>
<td>$\max w_i \times 100$</td>
<td>4.625</td>
<td>4.926</td>
<td>1.684</td>
<td></td>
</tr>
<tr>
<td>$\min w_i \times 100$</td>
<td>0.000</td>
<td>-0.242</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>$\sum I(w_i &lt; 0)/N_T$</td>
<td>0.000</td>
<td>0.503</td>
<td>0.511</td>
<td></td>
</tr>
<tr>
<td>$\bar{r}$</td>
<td>0.120</td>
<td>0.428</td>
<td>0.200</td>
<td></td>
</tr>
<tr>
<td>$\sigma(r)$</td>
<td>0.161</td>
<td>0.278</td>
<td>0.186</td>
<td></td>
</tr>
<tr>
<td>$\text{CE}(r)$</td>
<td>0.052</td>
<td>0.231</td>
<td>0.106</td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>—</td>
<td>0.324</td>
<td>0.083</td>
<td></td>
</tr>
<tr>
<td>$\beta$</td>
<td>—</td>
<td>0.742</td>
<td>0.962</td>
<td></td>
</tr>
<tr>
<td>$\sigma(\epsilon)$</td>
<td>—</td>
<td>0.252</td>
<td>0.104</td>
<td></td>
</tr>
<tr>
<td>IR</td>
<td>—</td>
<td>1.283</td>
<td>0.795</td>
<td></td>
</tr>
</tbody>
</table>
### Conditioning on Slope of Yield Curve

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>Weight</th>
<th>Optimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_{me} \times I(tsp \geq 0)$</td>
<td>—</td>
<td>—</td>
<td>-1.918</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>—</td>
<td>(0.707)</td>
</tr>
<tr>
<td>$\theta_{me} \times I(tsp &lt; 0)$</td>
<td>—</td>
<td>—</td>
<td>0.952</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>—</td>
<td>(1.197)</td>
</tr>
<tr>
<td>$\theta_{btm} \times I(tsp \geq 0)$</td>
<td>—</td>
<td>—</td>
<td>3.270</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>—</td>
<td>(1.101)</td>
</tr>
<tr>
<td>$\theta_{btm} \times I(tsp &lt; 0)$</td>
<td>—</td>
<td>—</td>
<td>4.965</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>—</td>
<td>(2.064)</td>
</tr>
<tr>
<td>$\theta_{mom} \times I(tsp \geq 0)$</td>
<td>—</td>
<td>—</td>
<td>2.128</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>—</td>
<td>(0.911)</td>
</tr>
<tr>
<td>$\theta_{mom} \times I(tsp &lt; 0)$</td>
<td>—</td>
<td>—</td>
<td>3.499</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>—</td>
<td>(1.613)</td>
</tr>
<tr>
<td>$me \times I(tsp \geq 0)$</td>
<td>1.616</td>
<td>—</td>
<td>-0.556</td>
</tr>
<tr>
<td>$me \times I(tsp \leq 0)$</td>
<td>0.479</td>
<td>—</td>
<td>0.340</td>
</tr>
<tr>
<td>$btm \times I(tsp \geq 0)$</td>
<td>-0.346</td>
<td>—</td>
<td>2.565</td>
</tr>
<tr>
<td>$btm \times I(tsp \leq 0)$</td>
<td>-0.105</td>
<td>—</td>
<td>0.909</td>
</tr>
<tr>
<td>$mom \times I(tsp \geq 0)$</td>
<td>0.027</td>
<td>—</td>
<td>1.534</td>
</tr>
<tr>
<td>$mom \times I(tsp &lt; 0)$</td>
<td>-0.015</td>
<td>—</td>
<td>0.693</td>
</tr>
</tbody>
</table>
## Conditioning on Slope of Yield Curve

| Variable                        | Value $|w_i| \times 100$ | Weight 0.035 | Optimum 0.136 |
|---------------------------------|----------|----------------|--------------|
| $\max w_i \times 100$          | 4.625    | 4.370          |
| $\min w_i \times 100$          | 0.000    | -0.389         |
| $\sum I(w_i < 0)/N_T$          | 0.000    | 0.475          |
| $\bar{r}$                      | 0.120    | 0.273          |
| $\sigma(r)$                    | 0.161    | 0.203          |
| $\text{CE}(r)$                 | 0.052    | 0.171          |
| $\alpha$                       | —        | 0.191          |
| $\beta$                        | —        | 0.357          |
| $\sigma(\epsilon)$            | —        | 0.196          |
| $\text{IR}$                    | —        | 0.977          |
### Different Utility Functions

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\gamma = 1$</th>
<th>$\gamma = 5$</th>
<th>$\gamma = 100$</th>
<th>min var</th>
<th>max Sharpe</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_{me}$</td>
<td>-6.109</td>
<td>-1.220</td>
<td>0.146</td>
<td>0.006</td>
<td>-1.117</td>
</tr>
<tr>
<td></td>
<td>(2.881)</td>
<td>(0.547)</td>
<td>(0.227)</td>
<td>(0.196)</td>
<td>(0.587)</td>
</tr>
<tr>
<td>$\theta_{btm}$</td>
<td>6.970</td>
<td>3.466</td>
<td>5.025</td>
<td>1.887</td>
<td>3.303</td>
</tr>
<tr>
<td></td>
<td>(3.548)</td>
<td>(0.922)</td>
<td>(0.946)</td>
<td>(0.311)</td>
<td>(1.092)</td>
</tr>
<tr>
<td>$\theta_{mom}$</td>
<td>7.380</td>
<td>2.000</td>
<td>0.483</td>
<td>0.455</td>
<td>2.415</td>
</tr>
<tr>
<td></td>
<td>(2.915)</td>
<td>(0.742)</td>
<td>(0.195)</td>
<td>(0.297)</td>
<td>(1.037)</td>
</tr>
<tr>
<td>$me$</td>
<td>-5.932</td>
<td>-0.118</td>
<td>0.744</td>
<td>1.545</td>
<td>0.043</td>
</tr>
<tr>
<td>$btm$</td>
<td>8.043</td>
<td>3.296</td>
<td>4.508</td>
<td>1.414</td>
<td>3.085</td>
</tr>
<tr>
<td>$mom$</td>
<td>6.958</td>
<td>1.839</td>
<td>0.290</td>
<td>0.388</td>
<td>2.263</td>
</tr>
</tbody>
</table>

- max SR and $\gamma = 5$ results are similar
## Different Utility Functions

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\gamma = 1$</th>
<th>$\gamma = 5$</th>
<th>$\gamma = 100$</th>
<th>min var</th>
<th>max Sharpe</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>w_i</td>
<td>\times 100$</td>
<td>0.320</td>
<td>0.121</td>
<td>0.146</td>
</tr>
<tr>
<td>max $w_i \times 100$</td>
<td>4.433</td>
<td>4.363</td>
<td>4.497</td>
<td>4.568</td>
<td>4.382</td>
</tr>
<tr>
<td>min $w_i \times 100$</td>
<td>-0.960</td>
<td>-0.323</td>
<td>-0.350</td>
<td>-0.142</td>
<td>-0.336</td>
</tr>
<tr>
<td>$\sum I(w_i &lt; 0)/N_T$</td>
<td>0.520</td>
<td>0.468</td>
<td>0.477</td>
<td>0.418</td>
<td>0.466</td>
</tr>
<tr>
<td>$\bar{r}$</td>
<td>0.546</td>
<td>0.244</td>
<td>0.190</td>
<td>0.155</td>
<td>0.251</td>
</tr>
<tr>
<td>$\sigma(r)$</td>
<td>0.598</td>
<td>0.190</td>
<td>0.187</td>
<td>0.137</td>
<td>0.195</td>
</tr>
<tr>
<td>CE($r$)</td>
<td>0.363</td>
<td>0.153</td>
<td>-0.976</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.507</td>
<td>0.157</td>
<td>0.109</td>
<td>0.050</td>
<td>0.164</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.375</td>
<td>0.440</td>
<td>0.345</td>
<td>0.742</td>
<td>0.447</td>
</tr>
<tr>
<td>$\sigma(\epsilon)$</td>
<td>0.595</td>
<td>0.177</td>
<td>0.179</td>
<td>0.069</td>
<td>0.182</td>
</tr>
<tr>
<td>IR</td>
<td>0.851</td>
<td>0.890</td>
<td>0.609</td>
<td>0.731</td>
<td>0.900</td>
</tr>
</tbody>
</table>
Transcript of page:

**Transaction Costs: Non-linear portfolio weights**

- First way:

\[ r_{p,t+1} = \sum_{i=1}^{N_t} w_{i,t} r_{i,t+1} - c_{i,t} |w_{i,t} - w_{i,t-1}| \]

where \( c_{i,t} \) is the one way transaction cost and \( T_t = |w_{i,t} - w_{i,t-1}| \) is the turnover.

- \( c_{i,t} \) can be constant, say 0.005 or
- \( c_{i,t} \) can vary across stocks and time: \( c_{i,t} = z_{i,t} \times Q_t \) where \( z_{i,t} = 0.006 - 0.0025 \times m_{e_{i,t}} \) and \( Q_t \) is a time trend, (1,4).

- This policy function is used by practitioners and is simple to understand.

- However, it is not optimal in the presence of proportional transactions costs (Magill and Constantinides (1976), Taksar, Klass, and Assaf (1988), and Davis and Norman (1990)).
Transaction Costs: Non-linear portfolio weights

- Second way: In the presence of proportional transaction costs, the optimal policy is to trade only if we are outside some boundaries around the target weight. The optimal weight will be defined as:

\[ w_{i,t} = \alpha_t w^h_{i,t} + (1 - \alpha_t) w^t_{i,t}, \text{ if } \frac{1}{N_t} \sum_{i=1}^{N_t} (w^t_{i,t} - w^h_{i,t})^2 > k^2 \]

\[ w_{i,t} = w^h_{i,t}, \text{ if } \frac{1}{N_t} \sum_{i=1}^{N_t} (w^t_{i,t} - w^h_{i,t})^2 < k^2 \]

where \( w^h_{i,t} \) is the buy and hold portfolio between time \( t-1 \) and \( t \) and

\[ w^t_{i,t} = \bar{w}_{i,t} + \theta^\top x_{i,t}/N_t \text{ and} \]

\[ \alpha_t = \frac{k \sqrt{N_t}}{(\sum_{i=1}^{N_t} (w^t_{i,t} - w^h_{i,t})^2)^{1/2}} \]

so that \( w_t \) is exactly at the boundary \( k \).
Table 6: Simple Portfolio Policy with Transactions Costs

This table shows estimates of the portfolio policy with three characteristics: size (me), book-to-market ratio (btm), and momentum (mom), specified in equation (3) and optimized for a power utility function with relative risk aversion of 5. The utility function is maximized for returns after transaction costs. In the first specification, the proportional transactions costs are 0.5%, constant across stocks and over time. In the second specification, transaction costs vary across stocks and over time as shown in Figure 3. For comparison, we also present results with zero transaction costs. We use data from the merged CRSP-Compustat database from January 1964 through December 2002. In the “Out-of-Sample” results, we use data until December 1973 to estimate the coefficients of the portfolio policy and then form out-of-sample monthly portfolios using those coefficients in the next year. Every subsequent year, we re-estimate the portfolio policy by enlarging the sample. All statistics are reported for the period January, 1974 to December, 2002. The first set of rows shows the estimated coefficients of the portfolio policy with bootstrapped standard errors in parentheses. The bootstrapped p-value of the Wald test under the null hypothesis that the parameter estimates are jointly equal to zero is also displayed. The “Out-of-Sample” results display time-series averages of coefficients, standard errors, and p-value. The second set of rows shows statistics of the portfolio weights, averaged across time. These statistics include the average absolute portfolio weight, the average minimum and maximum portfolio weights, the average sum of negative weights in the portfolio, the average fraction of negative weights in the portfolio, and the turnover in the portfolio. The third set of rows displays average portfolio return statistics: certainty equivalent returns, average return, standard deviation, and Sharpe ratio of returns, the alpha, beta, and volatility of idiosyncratic shocks of a market model regression, and the information ratio. We compute the certainty equivalent return for the policy with and without adjustment for transaction costs. The final set of rows displays the average normalized characteristics of the portfolio. The average risk-free rate in the sample is 0.061 (annualized).

<table>
<thead>
<tr>
<th>Variable</th>
<th>In-Sample PPP</th>
<th>Out-of-Sample PPP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$c_{i,t} = f(me_{i,t})$</td>
<td>$c_{i,t} = f(me_{i,t})$</td>
</tr>
<tr>
<td>$\theta_{me}$</td>
<td>-1.451</td>
<td>-1.391</td>
</tr>
<tr>
<td>Std.Err.</td>
<td>(0.548)</td>
<td>(0.547)</td>
</tr>
<tr>
<td>$\theta_{btm}$</td>
<td>3.606</td>
<td>3.557</td>
</tr>
<tr>
<td>Std.Err.</td>
<td>(0.921)</td>
<td>(0.922)</td>
</tr>
<tr>
<td>$\theta_{mom}$</td>
<td>1.772</td>
<td>1.651</td>
</tr>
<tr>
<td>Std.Err.</td>
<td>(0.743)</td>
<td>(0.741)</td>
</tr>
<tr>
<td>LRT p-value</td>
<td></td>
<td>0.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>VW</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.000</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$</td>
<td>w_i</td>
<td>\times 100$</td>
</tr>
<tr>
<td>max $w_i \times 100$</td>
<td>3.678</td>
<td>3.485</td>
</tr>
<tr>
<td>min $w_i \times 100$</td>
<td>0.000</td>
<td>-0.216</td>
</tr>
<tr>
<td>$\sum w_i I(w_i &lt; 0)$</td>
<td>0.000</td>
<td>-1.279</td>
</tr>
<tr>
<td>$\sum f(w_i \leq 0)/N_t$</td>
<td>0.000</td>
<td>0.472</td>
</tr>
<tr>
<td>$\sum</td>
<td>w_i - w^h_i</td>
<td>$</td>
</tr>
<tr>
<td>CE$</td>
<td>c_{i,t} = 0.000$</td>
<td>0.064</td>
</tr>
<tr>
<td>CE$</td>
<td>c_{i,t} = 0.005$</td>
<td>0.169</td>
</tr>
<tr>
<td>CE$</td>
<td>c_{i,t} = f(me_{i,t})$</td>
<td>0.162</td>
</tr>
<tr>
<td>$\bar{r}$</td>
<td>0.139</td>
<td>0.262</td>
</tr>
<tr>
<td>$\sigma(r)$</td>
<td>0.169</td>
<td>0.188</td>
</tr>
<tr>
<td>SR</td>
<td>0.438</td>
<td>1.048</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>-0.174</td>
<td>0.183</td>
</tr>
<tr>
<td>$\beta$</td>
<td>-0.311</td>
<td>0.328</td>
</tr>
<tr>
<td>$\sigma(\epsilon)$</td>
<td>-0.181</td>
<td>0.175</td>
</tr>
<tr>
<td>IR</td>
<td>-0.960</td>
<td>0.929</td>
</tr>
<tr>
<td>me</td>
<td>2.118</td>
<td>-0.337</td>
</tr>
<tr>
<td>btm</td>
<td>-0.418</td>
<td>3.553</td>
</tr>
<tr>
<td>mom</td>
<td>0.016</td>
<td>1.623</td>
</tr>
</tbody>
</table>

Table 6: Simple Portfolio Policy with Transactions Costs

This table shows estimates of the portfolio policy with three characteristics: size (me), book-to-market ratio (btm), and momentum (mom), specified in equation (3) and optimized for a power utility function with relative risk aversion of 5. The utility function is maximized for returns after transaction costs. In the first specification, the proportional transactions costs are 0.5%, constant across stocks and over time. In the second specification, transaction costs vary across stocks and over time as shown in Figure 3. For comparison, we also present results with zero transaction costs. We use data from the merged CRSP-Compustat database from January 1964 through December 2002. In the “Out-of-Sample” results, we use data until December 1973 to estimate the coefficients of the portfolio policy and then form out-of-sample monthly portfolios using those coefficients in the next year. Every subsequent year, we re-estimate the portfolio policy by enlarging the sample. All statistics are reported for the period January, 1974 to December, 2002. The first set of rows shows the estimated coefficients of the portfolio policy with bootstrapped standard errors in parentheses. The bootstrapped p-value of the Wald test under the null hypothesis that the parameter estimates are jointly equal to zero is also displayed. The “Out-of-Sample” results display time-series averages of coefficients, standard errors, and p-value. The second set of rows shows statistics of the portfolio weights, averaged across time. These statistics include the average absolute portfolio weight, the average minimum and maximum portfolio weights, the average sum of negative weights in the portfolio, the average fraction of negative weights in the portfolio, and the turnover in the portfolio. The third set of rows displays average portfolio return statistics: certainty equivalent returns, average return, standard deviation, and Sharpe ratio of returns, the alpha, beta, and volatility of idiosyncratic shocks of a market model regression, and the information ratio. We compute the certainty equivalent return for the policy with and without adjustment for transaction costs. The final set of rows displays the average normalized characteristics of the portfolio. The average risk-free rate in the sample is 0.061 (annualized).
Table 7: Boundary Portfolio Policy with Transactions Costs

This table shows estimates of the portfolio policy with three characteristics: size (me), book-to-market ratio (btm), and momentum (mom), specified in section 2.4 and optimized for a power utility function with relative risk aversion of 5. The utility function is maximized for returns after transaction costs. In the first specification, the proportional transactions costs are 0.5%, constant across stocks and over time. In the second specification, transaction costs vary across stocks and over time as shown in Figure 3. For comparison, we also present results with zero transaction costs. We use data from the merged CRSP-Compustat database from January 1964 through December 2002. In the “Out-of-Sample” results, we use data until December 1973 to estimate the coefficients of the portfolio policy and then form out-of-sample monthly portfolios using those coefficients in the next year. Every subsequent year, we re-estimate the portfolio policy by enlarging the sample. All statistics are reported for the period January, 1974 to December, 2002. The first set of rows shows the estimated coefficients of the portfolio policy with bootstrapped standard errors in parentheses. The bootstrapped $p$-value of the Wald test under the null hypothesis that the parameter estimates are jointly equal to zero is also displayed. The “Out-of-Sample” results display time-series averages of coefficients, standard errors, and $p$-value. The second set of rows shows statistics of the portfolio weights, averaged across time. These statistics include the average absolute portfolio weight, the average minimum and maximum portfolio weights, the average sum of negative weights in the portfolio, the average fraction of negative weights in the portfolio, and the turnover in the portfolio. The third set of rows displays average portfolio return statistics: certainty equivalent returns, average return, standard deviation, and Sharpe ratio of returns, the alpha, beta, and volatility of idiosyncratic shocks of a market model regression, and the information ratio. We compute the certainty equivalent return for the policy with and without adjustment for transaction costs. The final set of rows displays the average normalized characteristics of the portfolio. The average risk-free rate in the sample is 0.061 (annualized).

<table>
<thead>
<tr>
<th>Variable</th>
<th>In-Sample PPP</th>
<th>Out-of-Sample PPP</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$c_{it} = f(me_{i,t})$</td>
<td>$c_{it} = f(me_{i,t})$</td>
</tr>
<tr>
<td>$\theta_{me}$</td>
<td>-1.147</td>
<td>-0.979</td>
</tr>
<tr>
<td>Std.Err.</td>
<td>(0.561)</td>
<td>(0.577)</td>
</tr>
<tr>
<td>$\theta_{btm}$</td>
<td>4.432</td>
<td>4.364</td>
</tr>
<tr>
<td>Std.Err.</td>
<td>(1.137)</td>
<td>(1.153)</td>
</tr>
<tr>
<td>$\theta_{mom}$</td>
<td>2.366</td>
<td>3.587</td>
</tr>
<tr>
<td>Std.Err.</td>
<td>(0.964)</td>
<td>(1.077)</td>
</tr>
<tr>
<td>$\kappa \times 10^3$</td>
<td>0.273</td>
<td>0.289</td>
</tr>
<tr>
<td>Std.Err.</td>
<td>(0.078)</td>
<td>(0.097)</td>
</tr>
<tr>
<td>LRT p-value</td>
<td>0.023</td>
<td>0.084</td>
</tr>
<tr>
<td>$</td>
<td>w_i</td>
<td>\times 100$</td>
</tr>
<tr>
<td>max $w_i \times 100$</td>
<td>0.00</td>
<td>-0.401</td>
</tr>
<tr>
<td>min $w_i \times 100$</td>
<td>0.00</td>
<td>4.021</td>
</tr>
<tr>
<td>$\sum w_i I(w_i &lt; 0)$</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$\sum I(w_i \leq 0)/N_t$</td>
<td>0.00</td>
<td>4.021</td>
</tr>
<tr>
<td>$\sum</td>
<td>w_i - w^h_i</td>
<td>/N_t$</td>
</tr>
<tr>
<td>CE</td>
<td>$c_{it} = 0.000$</td>
<td>0.064</td>
</tr>
<tr>
<td>CE</td>
<td>$c_{it} = 0.005$</td>
<td>-0.139</td>
</tr>
<tr>
<td>CE</td>
<td>$c_{it} = f(me_{i,t})$</td>
<td>0.139</td>
</tr>
<tr>
<td>$\bar{r}$</td>
<td>0.169</td>
<td>0.293</td>
</tr>
<tr>
<td>SR</td>
<td>0.438</td>
<td>1.021</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>-0.196</td>
<td>0.208</td>
</tr>
<tr>
<td>$\beta$</td>
<td>-0.266</td>
<td>0.272</td>
</tr>
<tr>
<td>$\sigma(e)$</td>
<td>-0.190</td>
<td>0.219</td>
</tr>
<tr>
<td>IR</td>
<td>-1.028</td>
<td>0.950</td>
</tr>
<tr>
<td>me</td>
<td>2.118</td>
<td>0.057</td>
</tr>
<tr>
<td>btm</td>
<td>-0.418</td>
<td>0.497</td>
</tr>
<tr>
<td>mom</td>
<td>0.016</td>
<td>0.497</td>
</tr>
<tr>
<td>avg $\alpha$</td>
<td>-0.556</td>
<td>0.549</td>
</tr>
<tr>
<td>min $\alpha$</td>
<td>-0.260</td>
<td>0.247</td>
</tr>
<tr>
<td>max $\alpha$</td>
<td>-0.828</td>
<td>0.740</td>
</tr>
</tbody>
</table>
Utility Relative to Benchmark $u(r_{p,t} - r_{bench,t})$

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>Weight</th>
<th>Optimum</th>
<th>Value</th>
<th>Weight</th>
<th>Optimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta_{me}$</td>
<td>—</td>
<td>—</td>
<td>-3.717 (1.104)</td>
<td>—</td>
<td>—</td>
<td>-1.252 (1.238)</td>
</tr>
<tr>
<td>$\theta_{btm}$</td>
<td>—</td>
<td>—</td>
<td>0.314 (0.648)</td>
<td>—</td>
<td>—</td>
<td>1.685 (0.680)</td>
</tr>
<tr>
<td>$\theta_{mom}$</td>
<td>—</td>
<td>—</td>
<td>1.924 (0.478)</td>
<td>—</td>
<td>—</td>
<td>1.670 (0.513)</td>
</tr>
<tr>
<td>me</td>
<td>1.527</td>
<td></td>
<td>-2.273</td>
<td>1.527</td>
<td></td>
<td>-0.019</td>
</tr>
<tr>
<td>btm</td>
<td>-0.240</td>
<td>0.535</td>
<td>-0.240</td>
<td>1.539</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mom</td>
<td>-0.026</td>
<td>1.959</td>
<td>-0.026</td>
<td>1.573</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Utility Relative to Benchmark \( u(r_{p,t} - r_{bench,t}) \)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value Weight</th>
<th>Optimum</th>
<th>Value Weight</th>
<th>Optimum</th>
</tr>
</thead>
<tbody>
<tr>
<td>(</td>
<td>w_i</td>
<td>\times 100)</td>
<td>0.200</td>
<td>0.557</td>
</tr>
<tr>
<td>(\max w_i \times 100)</td>
<td>5.343</td>
<td>3.537</td>
<td>5.343</td>
<td>4.295</td>
</tr>
<tr>
<td>(\min w_i \times 100)</td>
<td>0.037</td>
<td>-1.400</td>
<td>0.037</td>
<td>-1.146</td>
</tr>
<tr>
<td>(\sum I(w_i &lt; 0)/N_T)</td>
<td>0.000</td>
<td>0.378</td>
<td>0.000</td>
<td>0.361</td>
</tr>
<tr>
<td>(\bar{r})</td>
<td>0.070</td>
<td>0.182</td>
<td>0.070</td>
<td>0.180</td>
</tr>
<tr>
<td>(\sigma(r))</td>
<td>0.035</td>
<td>0.136</td>
<td>0.035</td>
<td>0.173</td>
</tr>
<tr>
<td>(\text{CE}(r))</td>
<td>0.067</td>
<td>0.135</td>
<td>0.067</td>
<td>0.092</td>
</tr>
<tr>
<td>(\alpha)</td>
<td>0.009</td>
<td>0.115</td>
<td>0.009</td>
<td>0.127</td>
</tr>
<tr>
<td>(\beta)</td>
<td>0.986</td>
<td>1.125</td>
<td>0.986</td>
<td>0.507</td>
</tr>
<tr>
<td>(\sigma(\epsilon))</td>
<td>0.034</td>
<td>0.135</td>
<td>0.034</td>
<td>0.166</td>
</tr>
<tr>
<td>(\text{IR})</td>
<td>0.852</td>
<td>0.852</td>
<td>0.852</td>
<td>0.765</td>
</tr>
</tbody>
</table>
### Out-of-Sample Performance

<table>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
<th>Weight</th>
<th>Optimum</th>
<th>Out-of-Sample</th>
<th>Value</th>
<th>Weight</th>
<th>Optimum</th>
<th>Out-of-Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\theta_{me}$</td>
<td>—</td>
<td>—</td>
<td>-1.707</td>
<td>-0.702</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>-0.702</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>(0.822)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>(0.726)</td>
</tr>
<tr>
<td>$\theta_{btm}$</td>
<td>—</td>
<td>—</td>
<td>2.891</td>
<td>4.166</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>4.166</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>(1.357)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>(1.084)</td>
</tr>
<tr>
<td>$\theta_{mom}$</td>
<td>—</td>
<td>—</td>
<td>3.111</td>
<td>1.346</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>1.346</td>
</tr>
<tr>
<td></td>
<td>—</td>
<td>(1.056)</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>(0.704)</td>
</tr>
<tr>
<td>me</td>
<td>2.078</td>
<td>-0.586</td>
<td>0.026</td>
<td>2.111</td>
<td>0.322</td>
<td>-0.228</td>
<td>2.913</td>
<td></td>
</tr>
<tr>
<td>btm</td>
<td>-0.528</td>
<td>2.732</td>
<td>3.784</td>
<td>-0.375</td>
<td>3.952</td>
<td>2.913</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mom</td>
<td>-0.070</td>
<td>2.890</td>
<td>1.044</td>
<td>0.094</td>
<td>1.293</td>
<td>3.034</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Out-of-Sample Performance

| Variable | First Sample (62-82) | | | Second Sample (82-02) | | |
|----------|----------------------|----------------------|----------------------|----------------------|----------------------|
|          | Value | Weight | Optimum | Out-of-Sample | Value | Weight | Optimum | Out-of-Sample |
| $|w_i| \times 100$ | 0.049 | 0.176 | 0.180 | 0.021 | 0.079 | 0.075 |
| $\text{max } w_i \times 100$ | 6.119 | 5.661 | 5.802 | 3.130 | 3.019 | 2.990 |
| $\text{min } w_i \times 100$ | 0.000 | -0.516 | -0.472 | 0.000 | -0.181 | -0.205 |
| $\sum I(w_i < 0)/N_T$ | 0.000 | 0.470 | 0.464 | 0.000 | 0.480 | 0.478 |
| $\bar{r}$ | 0.103 | 0.293 | 0.201 | 0.136 | 0.245 | 0.265 |
| $\sigma(r)$ | 0.155 | 0.233 | 0.194 | 0.166 | 0.163 | 0.224 |
| $\text{CE}(r)$ | 0.043 | 0.151 | 0.113 | 0.061 | 0.177 | 0.121 |
| $\alpha$ | — | 0.197 | 0.110 | — | 0.177 | 0.197 |
| $\beta$ | — | 0.795 | 0.660 | — | 0.154 | 0.153 |
| $\sigma(\epsilon)$ | — | 0.198 | 0.165 | — | 0.162 | 0.223 |
| IR | — | 0.994 | 0.668 | — | 1.095 | 0.884 |
Multiperiod Portfolio Optimization

- Multiperiod portfolio returns:

\[ r_{p,t+1\rightarrow t+K} = \prod_{k=0}^{K} (1+r_{p,t+k+1}) - 1 = \prod_{k=0}^{K} \left[ 1 + \sum_{i=1}^{N_{t+k}} (\bar{w}_{i,t+k} + \theta_k^T x_{i,t+k}/N_{t+k}) r_{i,t+k+1} \right] - 1 \]

- Maximize sample average realized utility of multiperiod returns:

\[
\max_{\{\theta_k\}_{k=0}^{K-1}} \frac{1}{T} \sum_{t=0}^{T-1} u \left( r_{p,t+1\rightarrow t+K} \right)
\]

- To avoid proliferation of parameters, the coefficients \( \theta_k \) can be modeled as constants or simple functions of time-to-horizon \((K - k)\).
Conclusions

- New approach to portfolio optimization ⇒ parameterize optimal portfolio weights as functions of the assets’ characteristics instead of modeling the joint distribution of returns

- Many advantages relative to the traditional approach
  - simple and feasible even for a large number of assets
  - numerically robust
  - easily modified and extended
  - expected utility estimation and inference

- New test of the equilibrium implication of asset pricing models that does not rely on distributional assumptions
6 Market Microstructure

- Liquidity is like...love. Everybody knows what it is, but it is hard to explain.
- Ability to buy or sell significant quantities of a security quickly, anonymously, and with minimal to no price impact.
- Liquidity is very important for a well-functioning financial market.
- Market-makers: provide liquidity by taking the opposite side of a transaction.
  - If an investor wants to buy, the market-maker sells and vice versa.
- In exchange for this service, market-makers buy at a low bid price $P^b$ and sell at a higher ask price $P^a$.
- This ability to buy low and sell high insures that the market-makers will make some profits.
- The difference $P^a - P^b$ is called the bid-ask spread.
● The bid-ask spread complicates things a bit, since we don’t observe the true price.
  – We have three prices: The bid, the ask, and the true price.
  – The true price is often between the bid and the ask, although it need not be.
  – How do we define returns: From bid to bid, from ask to ask, from bid to ask...
  – How is the bid-ask spread determined?

● It is fairly intuitive that the bid-ask spread has an effect on returns.

● Roll (1984) provides a particularly simple and appealing model of how the bid-ask spread might impact the time-series properties of returns.

● This model provides most of the intuition and the framework on how financial economists think about the bid-ask spread.
- The observed market price is
  \[ P_t = P_t^* + I_t \frac{s}{2} \]
  - \( P_t^* \) is the fundamental price in a frictionless economy
  - \( s \) is the bid-ask spread
  - \( I_t \) is an iid index variable that takes values of 1 with probability 0.5 (if the trade is initiated by a buyer) and -1 with probability 0.5 (if the trade is initiated by a seller).

- Note that \( E (I_t) = 0 \) and \( Var(I_t) = 1 \).
- For simplicity assume that \( P_t^* \) does not change.
- The change in price is
  \[ \Delta P_t = \Delta P_t^* + \frac{s}{2} I_t - \frac{s}{2} I_{t-1} \]
- Its variance, covariance, and correlation are:
  \[ Var(\Delta P_t) = 2 \frac{s^2}{4} = \frac{s^2}{2} \]
  \[ Cov(\Delta P_t, \Delta P_{t-1}) = -\frac{s^2}{4} \]
  \[ Cov(\Delta P_t, \Delta P_{t-k}) = 0, k > 1 \]
  \[ Corr(\Delta P_t, \Delta P_{t-1}) = -\frac{1}{2} \]
• The fundamental value is fixed, but there is variation from $s$.
• The bid-ask spread induces negative correlation in returns even in the absence of other fluctuations!
• This is quite intuitive, right?
• The variance and serial correlation depend on the magnitude of the bid-ask spread.
• This particular example induces a first-order serial correlation.
• We can also express the spread as a function of the covariance, or
  \[ s = 2 \sqrt{-\text{Cov}(\Delta P_t, \Delta P_{t-1})} \]
• However, in practice $\text{Cov}(\Delta P_t, \Delta P_{t-1}) > 0$ is not uncommon.
• Roll (1984) defines the spread as
  \[ s = -2 \sqrt{|\text{Cov}(\Delta P_t, \Delta P_{t-1})|} \]
• One can also take $s = 2 \sqrt{-\text{Cov}(\Delta P_t, \Delta P_{t-1})}$ to be a testable implication. The fact that we observe $\text{Cov}(\Delta P_t, \Delta P_{t-1}) > 0$ implies that the model is misspecified (Glosten and Harris (1988) and Stoll (1989)).
• Roll proposes to estimate the effective spread as
  \[ s = -2\sqrt{|Cov(\Delta P_t, \Delta P_{t-1})|} \]

• But why estimate a quantity that is observable. After all, we know the bid and the ask prices, and the spread?

• Roll argues that the quoted (or observed) spread is different from the effective spread. Sometimes transactions occur at prices within the bid-ask spread because
  – market-makers do not update bid-ask quotes in a timely manner.
  – provide discount to customers that are trading for reasons other than private info (Glosten and Milgrom (1985), Goldstein (1993)).
  – inventory rebalancing.

• The assumption that \( P_t^* \) is fixed might be relaxed. As long as it is independent of \( I_t \), we can express
  \[
  Corr(\Delta P_t, \Delta P_{t-1}) = -\frac{s^2}{4} \frac{s^2}{\frac{s^2}{2} + Var(\Delta P_t^*)}
  \]
• Roll’s (1984) model is designed to illustrate how the spread can induce negative serial correlation in returns.
• However, the serial correlation is a function of the spread.
• But the spread is set exogenously.
• The question remains: What determines the bid-ask spread?
• The spread is very important for market-makers.
• The spread is unlikely to be independent of the price $P_t^*$. 
What are the fundamental forces that might have an impact on the bid-ask spread?

- Order-processing costs—basic setup and operation costs.
- Inventory costs—holding an undesired security that is subject to risk.
- Adverse selection costs—some investors are better informed than the market maker about the stock. The market-maker has no way of distinguishing between informed and uninformed traders and must be compensated for the added risk.

Glosten (1987) has a nice model where the adverse selection costs are modelled explicitly.

- Premise: People trade because of new information. Then larger (measured by volume) trades—trades that reflect the revelation of ‘lots’ of new information—must have a larger impact on prices than smaller trades.

- Hasbrouck (1991) conducts a VAR analysis and finds that there is such a price impact and that in fact it is quite large.
6.2 Transactions Data

- Recent datasets such as the TAQ (Trades and Quotes) or the Optionmetrics databases have given economists access to a lot of new information.
- The possibilities and the challenges are great.
- Such datasets are often tick-by-tick, all transactions of every stock are recorded.
- The transactions are not evenly spaced.
- The IID assumption fails.
- Most economists are only now starting to develop tools to handle such datasets.
- Discreteness must be taken into account.
- In many instances, people aggregate or smooth the data.
6.3 Discreteness

- Equities used to be quoted in ticks—1/8, 2/8, ...(equity options are quoted in 1/16’s)
- A priori, one would think that prices are uniformly distributed across ticks.

- Finding:
  - Prices fall more often on whole-dollar multiples than on half-dollar multiples
  - Prices fall more often on half-dollar multiples than on quarter-dollar multiples
  - Prices fall more often on quarter-dollar multiples than on eighth-dollar multiples
  - Prices fall more often on even eighths than on odd eighths.

- The last finding drew a lot of attention.

- Christie and Schultz (1994) and Christie, Harris, and Schultz (1994) show that Nasdaq quotes tend to cluster more frequently on even eighths than on odd eighths.
• They argue that this clustering is an indication of tacit collusion among Nasdaq dealers to maintain wider spreads.

• Their study and allegations prompted the SEC to launch an investigation.
7 Bootstrap

- Sometimes:
  - The normality assumption is not adequate. But we need an adequate assumption in order to have accurate tests.
  - There are peculiarities of the data that cannot be handled with asymptotic theory
  - The CLT might not provide good approximation (it is only an asymptotic result).

- Example: Short rate might be bi-modal.
- Example: Extreme overlap in data
- Example: Missing values
- Q: What to do?
- IDEA: Treat the sample as if it were the population.
- Resample WITH replacement in order to create many pseudo-samples.
- Treat the pseudo-samples as if they are realizations from the true population.
- Those samples will give you an idea about the true distribution (provided the sample is representative)
- CAREFUL: This works only if we have a representative sample. Sample selection problems abound.....
• Example: Suppose we have a sample \( \{x_1, x_2, \ldots, x_n\} \). We believe that there is heteroskedasticity in the sample, but we don’t know how to model it.

  – From the sample, we compute the mean \( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \).

  – Now the question is to find the dispersion around the true value of \( \mu \) in order to conduct tests.

  – Draw a sample of \( n \) observations WITH REPLACEMENT from \( \{x_1, x_2, \ldots, x_n\} \). Name this sample \( \{x_1^1, x_2^1, \ldots, x_n^1\} \).

  – Draw a second sample of \( n \) observations WITH REPLACEMENT from \( \{x_1, x_2, \ldots, x_n\} \). Name this sample \( \{x_1^2, x_2^2, \ldots, x_n^2\} \).

  – Draw \( J \) more samples, where the \( j \)th sample is denoted by \( \{x_1^j, x_2^j, \ldots, x_n^j\}_{j=1}^{J} \).

  – For each pseudo-sample, \( j \), we can compute its mean, \( \bar{x}^j = \frac{1}{n} \sum_{i=1}^{n} x_i^j \).

  – We can plot the histogram of the means \( \{\bar{x}^j\}_{j=1}^{J} \). This is the empirical distribution of \( \bar{x} \).

  – We can form confidence intervals and tests based on the above distribution.
• The same principle can be applied to any statistic. In the previous example, we could have focused on the t-statistic, instead of the mean.

• From the sample, we compute the t statistic \( t = \frac{\overline{x} - 0}{se} \).
  – Now the question is to find the dispersion around the true value of \( t \) in order to conduct tests.
  – Draw a sample of \( n \) observations WITH REPLACEMENT from \( \{x_1, x_2, \ldots, x_n\} \). Name this sample \( \{x^1_1, x^1_2, \ldots, x^1_n\} \).
  – Draw a second sample of \( n \) observations WITH REPLACEMENT from \( \{x_1, x_2, \ldots, x_n\} \). Name this sample \( \{x^2_1, x^2_2, \ldots, x^2_n\} \).
  – Draw \( J \) more samples, where the \( j \)th sample is denoted by \( \{x^j_1, x^j_2, \ldots, x^j_n\} \).
  – For each pseudo-sample, \( j \), we can compute its t statistic, \( t^j = \frac{1}{n} \sum_{i=1}^{n} x^j_i \).
  – We can plot the histogram of the t statistics \( \{t^j\}_{j=1}^{J} \). This is the empirical distribution of \( t \).
  – We can form confidence intervals and tests based on the above distribution.
7.1 Bootstrap in OLS—Conditional vs Unconditional Bootstrap

7.1.1 Conditional Bootstrap

- We have the regression
  \[ y_i = \beta x_i + \varepsilon_i \]
- Estimate it using OLS, to obtain \( \hat{\beta} \), its t-statistic, and \( \{\hat{\varepsilon}_1, \hat{\varepsilon}_2, ..., \hat{\varepsilon}_n\} \)
- Conditional on the \( x_i \)'s, we can simulate the \( y_i \)'s as follows:
  - Draw with replacement from \( \{\hat{\varepsilon}_1, \hat{\varepsilon}_2, ..., \hat{\varepsilon}_n\} \) \( J \) samples of \( n \) observations each.
  - Construct the samples as: \( y_i^j = \hat{\beta} x_i + \varepsilon_i^j \)
  - Note that the randomness comes only from resampling the residuals.
  - For each sample compute \( \hat{\beta}^j \) and/or \( t^j \)
  - Form tests as before.
7.1.2 Unconditional Bootstrap

- In the conditional bootstrap, we held the $x's$ fixed. In other words, we were really bootstrapping the marginal distribution of $y|x$.

- Again, suppose we have
  \[ y_i = \beta x_i + \varepsilon_i \]

- Estimate it using OLS, to obtain $\hat{\beta}$, its t-statistic, and $\{\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n\}$

- We also have $\{x_1, x_2, \ldots, x_n\}$ which are independent of $\{\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n\}$ by construction

- Therefore, we can simulate the $y's$ as follows:
  - Draw with replacement from $\{\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n\}$ $J$ samples of $n$ observations each.
  - Draw with replacement from $\{x_1, x_2, \ldots, x_n\}$ $J$ samples of $n$ observations each.
  - Construct the samples as: $y_i^j = \hat{\beta} x_i^j + \varepsilon_i^j$
  - Note that the randomness comes not only from the residuals but also from the explanatory variables.
  - For each sample compute $\hat{\beta}^j$ and/or $t^j$
• The unconditional bootstrap test is more conservative.

• Test for spurious correlation.

• CAUTION: The bootstrap does not work if the data is serially dependent. Why?
  – Block bootstrap: Works pretty well, depending how dependent the data is.
7.2 Bootstrap in GMM context

7.2.1 Conditional

- Recall the moment condition
  \[ E \left[ \left( 1 - \beta \left( \frac{c_{t+1}}{c_t} \right)^{-\gamma} (1 + R_{t+1}) \right) x_t \right] = 0 \]

- Draw with replacement from \( \{c_1, c_2, \ldots\} \) and \( \{R_1, R_2, \ldots\} \) \( J \) samples of \( T \) observations each. Do not sample independently.

- Condition on the \( x_t \).

- Run the GMM for each sample, producing estimates \( \{\hat{\gamma}^1, \hat{\gamma}^2, \ldots, \hat{\gamma}^J\} \).

- Plot the empirical distribution and test if the estimate \( \hat{\gamma} \) is in the tails.
7.2.2 Unconditional

- Follow the same procedure, but also sample $\{x_t\}^T_{t=1}$ independently from the other two series.
- Form the same tests.
- This test can have the interpretation to test for spurious correlation.
7.3 Simulations

- Not to be confused with bootstrap.
- Neither better nor worse, just different.
- We have the regression
  \[ y_i = \beta x_i + \varepsilon_i \]
- Estimate it using OLS, to obtain \( \hat{\beta} \) and its t-statistic.
- Assume that the errors \( \varepsilon_i \) have a certain distribution, say \( \varepsilon_i \sim T(15) \)
- Conditional on the \( x_i \)'s, we can simulate the \( y_i \)'s as follows:
  - Simulate from, say, \( T(15) \), \( J \) samples of \( n \) observations each.
  - Construct the samples as: \( y_i^j = \hat{\beta} x_i + \varepsilon_i^j \)
  - Note that the randomness comes only from the simulated residuals.
  - For each sample compute \( \hat{\beta}^j \) and/or \( t^j \)
  - Form tests as before.
• We can also assume dependence between the residuals, as in \( \varepsilon_i = \phi \varepsilon_{i-1} + \nu_t \), where \( \nu_t \) is NID(0,1).
• We can also simulate the \( x' \)'s.
• IMPORTANT: In the simulations, we have to make crucial assumptions about the underlying distribution of the \( \varepsilon' \)'s, the very assumption we want to circumvent in the bootstrap exercise.
• IMPORTANT: The simulations exercise can handle dependent data.