Course Materials


• Asset Pricing, by John H. Cochrane, Princeton University Press, 2005

• Class notes as well as published and working papers in finance and economics as listed in the reference list
From Rational Asset pricing to Market Anomalies
Expected Return

• Statistically, an expected asset return (in excess of the risk free rate) can be formulated as

$$\mathbb{E}(r_{i,t}) = \alpha_i + \beta_i' \mathbb{E}(f_t)$$  \hspace{1cm} (1.1/1)

where $f$ denotes a set of $K$ factor mimicking portfolios, $\beta_i$ is a $K$ vector of factor loadings, and $\alpha_i$ reflects the expected return component unexplained by factors, or model mispricing.

If factors are not return spreads (e.g., consumption growth) then the intercept does no longer represent asset mispricing.
For example, the market model is a statistical model with $f$ being represented by excess return on the market portfolio.

An asset pricing model intends to identify economic [(I)CAPM] or statistical [APT] common factors (which are portfolio spreads) that eliminate $\alpha$.

In the absence of alpha, expected return differential across assets is triggered by factor loadings only.

The presence of mispricing (nonzero alpha) could give rise to additional cross sectional differences in expected returns.
Empirical evidence shows that:

- $\alpha$ is significant in time-series asset pricing tests (e.g., Gibbons, Ross, and Shanken (1987)) based on various factor specifications;
- $\alpha$ varies cross-sectionally with firm level variables (e.g., size, book-to-market, turnover, past return);
- $\alpha$ varies over time with economy-wide business conditions (proxied by aggregate macro variables such as investor sentiment, the default spread, the dividend yield, and the term spread).

Black, Jensen, and Scholes (1972) is the first time-series test of factor models; Fama and MacBeth (1973) is the first cross-sectional test.
1. Asset pricing models > 1.2 CAPM

**CAPM**

- The CAPM of Sharpe (1964), Lintner (1965), and Mossin (1996) has originated the literature on asset pricing models.
- The CAPM is a general equilibrium model in a single-period economy.
- The model imposes an economic restriction on the statistical structure in (1.1/1).
- The unconditional version of the CAPM says that

\[
\mathbb{E}(r_{i,t}) = \text{cov}(r_{i,t}, r_{m,t}) \frac{\mathbb{E}(r_{m,t})}{\text{var}(r_{m,t})},
\]

\[
= \beta_{i,m} \mathbb{E}(r_{m,t}),
\]

where \( r_{m,t} \) is excess return on the market portfolio at time \( t \).
• There are conditional versions of the CAPM wherein the expected market premium or risk loadings or both are time varying.

• The CAPM measures an asset risk as the covariance of the asset return with the market portfolio return.

• The higher the covariance the less desirable the asset is; thus the asset price is lower which amounts to higher expected return.

• The market price of risk, common to all assets, is determined in equilibrium by the risk aversion of investors.
1. Asset pricing models > 1.2 CAPM

- For practical purposes such as
  - estimating the cost of equity capital in corporate finance applications
  - testing the CAPM
  - evaluating performance of mutual funds, hedge funds, and investment newsletters
  - testing the robustness of market anomalies

  the market portfolio is commonly proxied by the value weighted NYSE portfolio, S&P500 index, or CRSP index (AMEX, NYSE, and NASDAQ combined).

- Overall, the CAPM is nice, simple, and intuitive.

- But there are just too many empirical and theoretical drawbacks.
Market Anomalies

• The CAPM is at odds with anomalous patterns in the cross section of stock returns – market anomalies.

• Market anomalies include the size, book-to-market, past return (short term reversals and intermediate term momentum), earnings momentum, dispersion, net equity issuance, accruals, credit risk (level and changes), asset growth, capital investment, profitability, and idiosyncratic volatility.

• There are many other cross sectional effects. A good list is in Harvey, Liu, and Zhu (2013). “... and the Cross-Section of Expected Returns”. They document 316 factors discovered by academia. Some of those factors are highly correlated – hence, the ultimate number of cross sectional effects is lower than 316 – but it is still large enough.
Market Anomalies: Different Views

• Scholars like Fama would claim that the presence of anomalies merely indicates the inadequacy of the CAPM. An alternative risk based model would capture all anomalous patterns in stock returns. Markets are in general efficient and the risk-return tradeoff applies.

• Scholars like Shiller would claim that asset prices are subject to behavioral biases. Markets are inefficient. Higher risk need not imply higher return.

• Both Fama and Shiller won the Nobel Prize in Economics in 2013.

• While they represent different school of thoughts, they both argue against the pricing abilities of the CAPM.
The size and value effects:

- Basu (1977), Banz (1981), and Fama and French (FF) (1992) suggest that the CAPM cannot explain the effect of size, the dividend yield, the earnings yield, and the book-to-market on average returns.
Size effect: higher average returns on small stocks than large stocks. Beta cannot explain the difference.

Table I
Average Returns, Post-Ranking $\beta$s and Average Size For Portfolios Formed on Size and then $\beta$: Stocks Sorted on ME (Down) then Pre-Ranking $\beta$ (Across): July 1963 to December 1990

Portfolios are formed yearly. The breakpoints for the size (ME, price times shares outstanding) deciles are determined in June of year $t$ ($t = 1963-1990$) using all NYSE stocks on CRSP. All NYSE, AMEX, and NASDAQ stocks that meet the CRSP-COMPUSTAT data requirements are allocated to the 10 size portfolios using the NYSE breakpoints. Each size decile is subdivided into 10 $\beta$ portfolios using pre-ranking $\beta$s of individual stocks, estimated with 2 to 5 years of monthly returns (as available) ending in June of year $t$. We use only NYSE stocks that meet the CRSP-COMPUSTAT data requirements to establish the $\beta$ breakpoints. The equal-weighted monthly returns on the resulting 100 portfolios are then calculated for July of year $t$ to June of year $t + 1$.

The post-ranking $\beta$s use the full (July 1963 to December 1990) sample of post-ranking returns for each portfolio. The pre- and post-ranking $\beta$s (there and in all other tables) are the sum of the slopes from a regression of monthly returns on the current and prior month’s returns on the value-weighted portfolio of NYSE, AMEX, and (after 1972) NASDAQ stocks. The average return is the time-series average of the monthly equal-weighted portfolio returns, in percent. The average size of a portfolio is the time-series average of monthly averages of ln(ME) for stocks in the portfolio at the end of June of each year, with ME denominated in millions of dollars.

The average number of stocks per month for the size-$\beta$ portfolios in the smallest size decile varies from 70 to 177. The average number of stocks for the size-$\beta$ portfolios in size deciles 2 and 3 is between 15 and 41, and the average number for the largest 7 size deciles is between 11 and 22.

The All column shows statistics for equal-weighted size-decile (ME) portfolios. The All row shows statistics for equal-weighted portfolios of the stocks in each $\beta$ group.

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Low-$\beta$</th>
<th>$\beta$-2</th>
<th>$\beta$-3</th>
<th>$\beta$-4</th>
<th>$\beta$-5</th>
<th>$\beta$-6</th>
<th>$\beta$-7</th>
<th>$\beta$-8</th>
<th>$\beta$-9</th>
<th>High-$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>1.25</td>
<td>1.34</td>
<td>1.29</td>
<td>1.36</td>
<td>1.31</td>
<td>1.33</td>
<td>1.28</td>
<td>1.24</td>
<td>1.21</td>
<td>1.25</td>
<td>1.14</td>
</tr>
<tr>
<td>Small ME</td>
<td>1.52</td>
<td>1.71</td>
<td>1.57</td>
<td>1.79</td>
<td>1.61</td>
<td>1.50</td>
<td>1.50</td>
<td>1.37</td>
<td>1.63</td>
<td>1.50</td>
<td>1.42</td>
</tr>
<tr>
<td>ME-2</td>
<td>1.29</td>
<td>1.25</td>
<td>1.42</td>
<td>1.36</td>
<td>1.39</td>
<td>1.65</td>
<td>1.61</td>
<td>1.37</td>
<td>1.31</td>
<td>1.34</td>
<td>1.11</td>
</tr>
<tr>
<td>ME-3</td>
<td>1.24</td>
<td>1.12</td>
<td>1.31</td>
<td>1.17</td>
<td>1.70</td>
<td>1.29</td>
<td>1.10</td>
<td>1.31</td>
<td>1.36</td>
<td>1.26</td>
<td>0.76</td>
</tr>
<tr>
<td>ME-4</td>
<td>1.25</td>
<td>1.27</td>
<td>1.13</td>
<td>1.54</td>
<td>1.06</td>
<td>1.34</td>
<td>1.05</td>
<td>1.41</td>
<td>1.17</td>
<td>1.35</td>
<td>0.98</td>
</tr>
<tr>
<td>ME-5</td>
<td>1.29</td>
<td>1.34</td>
<td>1.42</td>
<td>1.39</td>
<td>1.48</td>
<td>1.42</td>
<td>1.18</td>
<td>1.27</td>
<td>1.18</td>
<td>1.04</td>
<td>1.02</td>
</tr>
<tr>
<td>ME-6</td>
<td>1.17</td>
<td>1.08</td>
<td>1.53</td>
<td>1.27</td>
<td>1.15</td>
<td>1.20</td>
<td>1.21</td>
<td>1.18</td>
<td>1.04</td>
<td>1.07</td>
<td>1.02</td>
</tr>
<tr>
<td>ME-7</td>
<td>1.07</td>
<td>0.95</td>
<td>1.21</td>
<td>1.26</td>
<td>1.09</td>
<td>1.18</td>
<td>1.11</td>
<td>1.24</td>
<td>0.62</td>
<td>1.32</td>
<td>0.76</td>
</tr>
<tr>
<td>ME-8</td>
<td>1.10</td>
<td>1.09</td>
<td>1.05</td>
<td>1.37</td>
<td>1.20</td>
<td>1.27</td>
<td>0.98</td>
<td>1.18</td>
<td>1.02</td>
<td>1.01</td>
<td>0.94</td>
</tr>
<tr>
<td>ME-9</td>
<td>0.95</td>
<td>0.98</td>
<td>0.88</td>
<td>1.02</td>
<td>1.14</td>
<td>1.14</td>
<td>1.07</td>
<td>1.23</td>
<td>0.94</td>
<td>0.82</td>
<td>0.88</td>
</tr>
<tr>
<td>Large ME</td>
<td>0.89</td>
<td>1.01</td>
<td>0.93</td>
<td>1.10</td>
<td>0.94</td>
<td>0.93</td>
<td>0.89</td>
<td>1.03</td>
<td>0.71</td>
<td>0.74</td>
<td>0.56</td>
</tr>
</tbody>
</table>
1. Asset pricing models > 1.2 CAPM > 1.2.1 Empirical drawbacks

Value Effect

- Value effect: higher average returns on value stocks than growth stocks. Beta cannot explain the difference.

- Value firms: Firms with high E/P, B/P, D/P, or CF/P. The notion of value is that physical assets can be purchased at low prices.

- Growth firms: Firms with low ratios. The notion is that high price relative to fundamentals reflects capitalized growth opportunities.

### Table I

|---|

| $R_f$ | is the one-month Treasury bill rate observed at the beginning of the month (from CRSP). The explanatory returns $R_m$, SMB, and HML are formed as follows. At the end of June of each year $t$ (1963–1993), NYSE, AMEX, and Nasdaq stocks are allocated to two groups (small or big, S or B) based on whether their June market equity (ME, stock price times shares outstanding) is below or above the median ME for NYSE stocks. NYSE, AMEX, and Nasdaq stocks are allocated in an independent sort to three book-to-market equity (BE/ME) groups (low, medium, or high; L, M, or H) based on the breakpoints for the bottom 30 percent, middle 40 percent, and top 30 percent of the values of BE/ME for NYSE stocks. Six size-BE/ME portfolios (S/L, S/M, S/H, B/L, B/M, B/H) are defined as the intersections of the two ME and the three BE/ME groups. Value-weight monthly returns on the portfolios are calculated from July to the following June. SMB is the difference, each month, between the average of the returns on the three small-stock portfolios (S/L, S/M, and S/H) and the average of the returns on the three big-stock portfolios (B/L, B/M, and B/H). HML is the difference between the average of the returns on the two high-BE/ME portfolios (S/H and B/H) and the average of the returns on the two low-BE/ME portfolios (S/L and B/L). The 25 size-BE/ME portfolios are formed like the six size-BE/ME portfolios used to construct SMB and HML, except that quintile breakpoints for ME and BE/ME for NYSE stocks are used to allocate NYSE, AMEX, and Nasdaq stocks to the portfolios.

BE is the COMPUSTAT book value of stockholders’ equity, plus balance sheet deferred taxes and investment tax credit (if available), minus the book value of preferred stock. Depending on availability, we use redemption, liquidation, or par value (in that order) to estimate the book value of preferred stock. The BE/ME ratio used to form portfolios in June of year $t$ is then book common equity for the fiscal year ending in calendar year $t – 1$, divided by market equity at the end of December of $t – 1$. We do not use negative BE firms, which are rare prior to 1980, when calculating the breakpoints for BE/ME or when forming the size-BE/ME portfolios. Also, only firms with ordinary common equity (as classified by CRSP) are included in the tests. This means that ADR’s, REIT’s, and units of beneficial interest are excluded.

The market return $R_m$ is the value-weight return on all stocks in the size-BE/ME portfolios, plus the negative BE stocks excluded from the portfolios.

### Book-to-Market Equity (BE/ME) Quintiles

<table>
<thead>
<tr>
<th>Size</th>
<th>Low</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>High</th>
<th>Low</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>0.31</td>
<td>0.70</td>
<td>0.82</td>
<td>0.95</td>
<td>1.08</td>
<td>7.67</td>
<td>6.74</td>
<td>6.14</td>
<td>5.85</td>
<td>6.14</td>
</tr>
<tr>
<td>2</td>
<td>0.48</td>
<td>0.71</td>
<td>0.91</td>
<td>0.93</td>
<td>1.09</td>
<td>7.13</td>
<td>6.25</td>
<td>5.71</td>
<td>5.23</td>
<td>5.94</td>
</tr>
<tr>
<td>3</td>
<td>0.44</td>
<td>0.68</td>
<td>0.75</td>
<td>0.86</td>
<td>1.05</td>
<td>6.52</td>
<td>5.53</td>
<td>5.11</td>
<td>4.79</td>
<td>5.48</td>
</tr>
<tr>
<td>4</td>
<td>0.51</td>
<td>0.39</td>
<td>0.64</td>
<td>0.80</td>
<td>1.04</td>
<td>5.86</td>
<td>5.28</td>
<td>4.97</td>
<td>4.81</td>
<td>5.67</td>
</tr>
<tr>
<td>Big</td>
<td>0.37</td>
<td>0.39</td>
<td>0.36</td>
<td>0.68</td>
<td>0.71</td>
<td>4.84</td>
<td>4.61</td>
<td>4.28</td>
<td>4.18</td>
<td>4.89</td>
</tr>
</tbody>
</table>

Panel A: Summary Statistics

<table>
<thead>
<tr>
<th>Means</th>
<th>Standard Deviations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>7.67</td>
</tr>
<tr>
<td>2</td>
<td>7.13</td>
</tr>
<tr>
<td>3</td>
<td>6.52</td>
</tr>
<tr>
<td>4</td>
<td>5.86</td>
</tr>
<tr>
<td>Big</td>
<td>4.84</td>
</tr>
</tbody>
</table>
1. Asset pricing models > 1.2 CAPM > 1.2.1 Empirical drawbacks

The International Value Effect

Table III

<table>
<thead>
<tr>
<th></th>
<th>Market</th>
<th>HB/M</th>
<th>LB/M</th>
<th>H-LB/M</th>
<th>HE/P</th>
<th>LE/P</th>
<th>H-LE/P</th>
<th>HC/P</th>
<th>LC/P</th>
<th>H-LC/P</th>
<th>HD/P</th>
<th>LD/P</th>
<th>H-LD/P</th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S.</td>
<td>9.57</td>
<td>(14.64)</td>
<td>7.75</td>
<td>6.79</td>
<td>14.09</td>
<td>7.38</td>
<td>(2.81)</td>
<td>13.74</td>
<td>7.08</td>
<td>6.66</td>
<td>11.79</td>
<td>8.01</td>
<td>3.73</td>
</tr>
<tr>
<td>France</td>
<td>11.26</td>
<td>(32.35)</td>
<td>9.46</td>
<td>7.64</td>
<td>15.68</td>
<td>8.70</td>
<td>6.98</td>
<td>16.17</td>
<td>9.30</td>
<td>8.66</td>
<td>15.12</td>
<td>6.25</td>
<td>8.88</td>
</tr>
<tr>
<td>Germany</td>
<td>9.88</td>
<td>(31.36)</td>
<td>10.01</td>
<td>2.75</td>
<td>11.13</td>
<td>10.58</td>
<td>0.55</td>
<td>13.28</td>
<td>5.14</td>
<td>8.13</td>
<td>9.99</td>
<td>10.42</td>
<td>-0.43</td>
</tr>
<tr>
<td>Italy</td>
<td>8.11</td>
<td>(43.77)</td>
<td>5.45</td>
<td>11.44</td>
<td>-5.99</td>
<td>7.62</td>
<td>12.99</td>
<td>-5.37</td>
<td>11.05</td>
<td>0.37</td>
<td>10.69</td>
<td>10.07</td>
<td>12.68</td>
</tr>
<tr>
<td>Netherlands</td>
<td>13.30</td>
<td>(38.55)</td>
<td>14.47</td>
<td>2.30</td>
<td>14.17</td>
<td>9.26</td>
<td>5.11</td>
<td>11.66</td>
<td>11.84</td>
<td>-0.19</td>
<td>13.47</td>
<td>13.05</td>
<td>0.41</td>
</tr>
<tr>
<td>Belgium</td>
<td>12.62</td>
<td>(38.07)</td>
<td>21.01</td>
<td>0.44</td>
<td>21.07</td>
<td>20.48</td>
<td>1.04</td>
<td>33.02</td>
<td>25.26</td>
<td>-0.63</td>
<td>21.38</td>
<td>30.81</td>
<td>0.07</td>
</tr>
<tr>
<td>Switzerland</td>
<td>11.07</td>
<td>(30.50)</td>
<td>27.63</td>
<td>3.99</td>
<td>30.47</td>
<td>27.88</td>
<td>0.78</td>
<td>28.84</td>
<td>25.57</td>
<td>1.27</td>
<td>26.47</td>
<td>29.26</td>
<td>1.29</td>
</tr>
<tr>
<td>Sweden</td>
<td>12.44</td>
<td>(27.21)</td>
<td>28.57</td>
<td>3.49</td>
<td>23.59</td>
<td>11.04</td>
<td>1.54</td>
<td>12.32</td>
<td>9.78</td>
<td>2.53</td>
<td>12.62</td>
<td>10.44</td>
<td>2.18</td>
</tr>
<tr>
<td>Australia</td>
<td>8.92</td>
<td>(24.91)</td>
<td>26.26</td>
<td>5.30</td>
<td>12.32</td>
<td>15.64</td>
<td>5.97</td>
<td>18.32</td>
<td>4.03</td>
<td>14.29</td>
<td>14.92</td>
<td>6.83</td>
<td>7.79</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>22.52</td>
<td>(41.96)</td>
<td>40.21</td>
<td>1.35</td>
<td>44.85</td>
<td>38.11</td>
<td>0.82</td>
<td>42.47</td>
<td>42.72</td>
<td>1.37</td>
<td>38.76</td>
<td>42.05</td>
<td>0.09</td>
</tr>
<tr>
<td>Singapore</td>
<td>13.13</td>
<td>(27.29)</td>
<td>27.71</td>
<td>9.67</td>
<td>15.21</td>
<td>3.12</td>
<td>2.09</td>
<td>13.42</td>
<td>8.05</td>
<td>5.39</td>
<td>10.64</td>
<td>13.10</td>
<td>-2.46</td>
</tr>
</tbody>
</table>

The Journal of Finance
Past Return Anomalies

• Lehmann (1990) and Jegadeesh (1990) show that contrarian strategies that exploit the short-run return reversals in individual stocks generate abnormal returns of about 1.7% per week and 2.5% per month, respectively.

• Jegadeesh and Titman (1993) and a great body of subsequent work uncover abnormal returns to momentum-based strategies focusing on investment horizons of 3, 6, 9, and 12 months.


• In sum, the literature has documented short term reversals, intermediate term momentum, and long term reversals.

• Given its prominence among academic scholars as well as practitioners, I will elaborate more on momentum.
1. Asset pricing models > 1.2 CAPM > 1.2.1 Empirical drawbacks

Momentum: Cumulative Return

• Momentum: Cumulative Return 1927 Jan-2013 Aug

<table>
<thead>
<tr>
<th>Year</th>
<th>Mkt-RF</th>
<th>SMB</th>
<th>HML</th>
<th>WML</th>
</tr>
</thead>
<tbody>
<tr>
<td>1927</td>
<td>$154.82</td>
<td>$7.08</td>
<td>$33.34</td>
<td>$312.38</td>
</tr>
<tr>
<td>1928</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1929</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1930</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1931</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1932</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1933</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1934</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1935</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1936</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1937</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1938</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1939</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1940</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1941</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1942</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1943</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1944</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1945</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1946</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1947</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1948</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1949</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1950</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1951</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1952</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1953</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1954</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1955</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1956</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1957</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1958</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1959</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1960</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1961</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1962</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1963</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1964</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1965</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1966</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1967</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1968</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1969</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1970</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1971</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1972</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1973</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1974</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1975</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1976</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1977</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1978</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1979</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1980</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1981</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1982</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1983</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1984</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1985</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1986</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1987</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1988</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1989</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1990</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1991</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1992</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1993</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1994</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
1. Asset pricing models > 1.2 CAPM > 1.2.1 Empirical drawbacks

### Momentum: Cumulative Return

- Last 2 decades: 1993 – Aug 2013

![Graph showing Momentum: Cumulative Return](image)

<table>
<thead>
<tr>
<th>Year</th>
<th>Mkt-RF</th>
<th>SMB</th>
<th>HML</th>
<th>WML</th>
</tr>
</thead>
<tbody>
<tr>
<td>1993</td>
<td>$3.14</td>
<td>$1.49</td>
<td>$1.51</td>
<td>$2.06</td>
</tr>
<tr>
<td>1994</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1995</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1996</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1997</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1999</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2004</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2007</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2009</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2011</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2015</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2016</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2017</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2018</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2019</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2020</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2021</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2022</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2023</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2024</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2025</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2026</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2027</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2028</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2029</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2030</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ASSET PRICING, Professor Doron Avramov, Finance Department, Hebrew University of Jerusalem, Israel
From Momentum Robustness to Momentum Crush

• Fama and French (1996) show that momentum profitability is the only CAPM-related anomaly unexplained by the Fama and French (1993) three-factor model.

• In fact, regressing gross momentum payoffs on the Fama-French factors tend to strengthen, rather than explain, momentum profitability.

• Because momentum loads negatively on the market, size, and value factors.

• Momentum also seems to appear in bonds, currencies, commodities, as well as mutual funds and hedge funds (both cold hands and hot hands).

• As Asness, Moskowitz, and Pedersen (2013) note: Momentum (and value) are everywhere.
2. Momentum > 2.1 Momentum robustness

- Schwert (2003) demonstrates that the size and value effects in the cross section of returns, as well as the ability of the aggregate dividend yield to forecast the equity premium disappear, reverse, or attenuate following their discovery.


Momentum > 2.1 Momentum robustness

- Korajczyk and Sadka (2004) find that momentum survives trading costs, whereas Avramov, Chordia, and Goyal (2006) show that the profitability of the other past-return anomaly, namely reversal, disappears in the presence of trading costs.

- Fama and French (2008) show that momentum is among the few robust anomalies – it works also among large cap stocks.

- A very recent paper by Geczy and Samonov (2013) examine momentum during the period 1801 through 1926 – probably the world’s longest back-test.

- Momentum had been fairly robust in a cross-industry analysis, cross-country analysis, and cross-style analysis.
The prominence of the momentum profitability has generated both behavioral and rational theories.

Momentum Interactions

Momentum interactions have been documented at the stock, industry, and aggregate level.

Stock level interactions:

- Hon, Lim, and Stein (2000) show that momentum profitability concentrates in small stocks.
- Lee and Swaminathan (2000) show that momentum payoffs increase with trading volume.
- Zhang (2006) finds that momentum concentrates in high information uncertainty stocks (stocks with high return volatility, cash flow volatility, or analysts’ forecast dispersion) and provides behavioral interpretation.
2. Momentum > 2.2 Momentum interactions

- Avramov, Chordia, Jostova, and Philipov (2007) document that momentum concentrates in low rated stocks. Moreover, the credit risk effect seems to dominate the information uncertainty and size effects.

**Potential industry level interactions:**

Moskowitz and Grinblatt (1999) show that industry momentum subsumes stock level momentum. That is, buy the past winning industries and sell the past loosing industries.

Grundy and Martin (2001) find no industry effects in momentum.
2. Momentum > 2.2 Momentum interactions

**Interactions at the aggregate level:**

- Chordia and Shavikumar (2002) show that momentum is captured by business cycle variables.

- Avramov and Chordia (2006a) demonstrate that momentum is captured by the component in model mispricing that varies with business conditions.

- Illiquidity – see my own paper on Time Series Momentum Payoffs and Illiquidity.

- Momentum also interacts with market states, market volatility, and market sentiment.
2. Momentum > 2.2 Momentum interactions

**Market States**

- Cooper, Gutierrez, and Hameed (2008) show that momentum profitability heavily depends on the state of the market.

- In particular, from 1929 to 1995, the mean monthly momentum profit following positive market returns is 0.93%, whereas the mean profit following negative market return is -0.37%.

- The study is based on the market performance over three years prior to the implementation of the momentum strategy.
Market sentiment

- Antoniou, Doukas, and Subrahmanyam (2010) and Stambaugh, Yu, and Yuan (2012) find that the momentum effect is stronger when market sentiment is high.

- The former paper suggests that this result is consistent with the slow spread of bad news during high-sentiment periods.

- Stambaugh, Yu, and Yuan (2013) use momentum along with ten other anomalies to form a stock level composite overpricing measure. For instance, loser stocks are likely to be overpriced due to impediments on short selling.
Momentum Crush

• In 2009 momentum delivers a negative 85% payoff.

• The negative payoff is especially attributable to the short side of the trade.

• Loser stocks had forcefully bounced back.

• There have been other episodes of momentum crush.

• The down side risk of momentum can be immense.
Momentum Spillover

- Gebhardt, Hvidkjaer, and Swaminathan examine the interaction between momentum in the returns of equities and corporate bonds.

- They find significant evidence of a momentum spillover from equities to investment grade corporate bonds of the same firm.

- In particular, firms earning high (low) equity returns over the previous year earn high (low) bond returns the following year.

- The spillover results are stronger among firms with lower-grade debt and higher equity trading volume.
2. Momentum > 2.1 Momentum robustness

**Bond Price Momentum**

- Momentum is also prominent among corporate bonds.
- Bond momentum profits are significant in the second half of the sample period, 1991 to 2008, and amount to 64 basis points per month.
- Momentum strategies are only profitable among non-investment grade bonds.
Earnings Momentum

• Ball and Brown (1968) document the post-earnings-announcement drift, also known as earnings momentum.

• This anomaly refers to the fact that firms reporting unexpectedly high earnings subsequently outperform firms reporting unexpectedly low earnings.

• The superior performance lasts for about nine months after the earnings announcements.
1. Asset pricing models > 1.2 CAPM > 1.2.1 Empirical drawbacks

Earnings Momentum

![Graph showing earnings momentum](image)
Accruals: Sloan (1996) shows that firms with high accruals earn abnormal lower returns on average than firms with low accruals. Sloan suggests that investors overestimate the persistence of the accrual component of earnings when forming earnings expectations. Total accruals are calculated as changes in noncash working capital minus depreciation expense scaled by average total assets for the previous two fiscal years.

Information uncertainty: Diether, Malloy, and Scherbina (2002) suggest that firms with high dispersion in analysts’ earnings forecasts earn less than firms with low dispersion. Other measures of information uncertainty: firm age, cash flow volatility, etc.
Anomalous patterns related to credit conditions:


- Campbell, Hilscher, and Szilagyi suggest that their finding is a challenge to standard models of rational asset pricing. They use failure probability estimated by a dynamic logit model with both accounting and equity market variables as explanatory variables. Using Ohlson (1980) O-score as the distress measure yields similar results.


- Interestingly, price response to credit rating changes is asymmetric – while stock and bond prices sharply fall following credit rating downgrades – there is only a mild response to rating upgrades.

- Given their high risk low return profile – an essential question is who holds distressed stocks?

- Coelho, John, and Tafller (2012) argue that some retail investors hold distress stocks due to their lottery-like characteristics – or they have high skewness.

- Stocks as lotteries is a concept nicely illustrated by Barberis and Huang (2008).
1. Asset pricing models > 1.2 CAPM > 1.2.1 Empirical drawbacks

- **Asset Growth:**
  - Cooper, Gulen, and Schill (2008) find companies that grow their total asset more earn lower subsequent returns.
  - They suggest that this phenomenon is due to investor initial overreaction to changes in future business prospects implied by asset expansions.
  - Asset growth can be measured as the annual percentage change in total assets.
• **Capital Investment:**


  – Capital investment to assets is the annual change in gross property, plant, and equipment plus the annual change in inventories divided by lagged book value of assets.

  – Changes in property, plants, and equipment capture capital investment in long-lived assets used in operations many years such as buildings, machinery, furniture, and other equipment.

  – Changes in inventories capture working capital investment in short-lived assets used in a normal business cycle.
1. Asset pricing models > 1.2 CAPM > 1.2.1 Empirical drawbacks

• Return on Assets (ROA):
  
  – Fama and French (2006) find that more profitable firms (ROA) have higher expected returns than less profitable firms.
  
  – ROA is typically measured as income before extraordinary items divided by one quarter lagged total assets.
Gross Profitability Premium:

Novy-Marx (2010) discovers that sorting on gross-profit-to-assets creates abnormal benchmark-adjusted returns, with more profitable firms having higher returns than less profitable ones.

Novy-Marx argues that gross profits scaled by assets is the cleanest accounting measure of true economic profitability. The farther down the income statement one goes, the more polluted profitability measures become, and the less related they are to true economic profitability.
Idiosyncratic volatility (IVOL):

• Ang, Hodrik, Xing, and Zhang (2006, 2009) show negative cross section relation between IVOL and average return in both US and global markets.

• In contrast, Fu (2009) documents an equally large positive relation.

• The AHXZ proxy for IVOL is the standard deviation of residuals from time-series regressions of excess stock returns on the Fama-French factors.

• The Fu measure is the expected IVOL based on EGARCH applied to such residuals.

• Avramov and Cederburg (2014) reconcile that conflicting evidence in a simple dividend discount model.
• They show that IVOL by itself is unrelated to the cross section of average returns.

• Nevertheless, it does predict the cross section of expected returns since it is related to dividend size and expected dividend growth – both of which impact the risk premiums.

• The AHXZ measure of IVOL is related to expected growth and thus it produces a negative relation.

• The Fu measure is related to dividend size – and thus it produces a positive relation.
Counter intuitive relations

• The forecast dispersion, credit risk, and IV effects apparently violate the risk-return tradeoff.

• Investors seem to pay premiums for purchasing higher risk stocks.

• Intuition may suggest it should be the other way around.
The turnover effect:

• Higher turnover is followed by lower future return. See, for example, Avramov and Chordia (RFS 2006).

• It should be noted that Swaminathan and Lee (2000) find that high turnover stocks exhibit features of high growth stocks.

• Turnover can be constructed using various methods. For instance, for any trading day within a particular month, compute the daily volume in either $ or the number of traded stocks or the number of transactions. Then divide the volume by the market capitalization or by the number of outstanding stocks. Finally, use the daily average, within a trading month, of the volume/market capitalization ratio as the monthly turnover.
F-Score

• The F-Score is due to Piotroski (2000).

• The FSOCRE is designed to identify firms with the strongest improvement in their overall financial conditions while meeting a minimum level of financial performance.

• High F-score firms demonstrate distinct improvement along a variety of financial dimensions, while low FSCORE firms exhibit poor fundamentals along these same dimensions.

• The FSCORE is computed as the sum of nine components which are either zero or one.

• The overall FSCORE thus ranges between zero and nine, where a low (high) score represents a firm with very few (many) good signals about its financial conditions.
Economic links and predictable returns:

- Cohen and Frazzini (2008) show that stocks do not promptly incorporate news about economically related firms.
- A long short strategy that capitalizes on this effect generates abnormal return of about 1.5% per month.
Are anomalies pervasive? There have been lots of doubts as explained below.

- Lo and MacKinlay (1990) claim that the size effect may very well be the result of unconscious, exhaustive search for a portfolio formation creating with the aim of rejecting the CAPM. Today they would have probably provided the same insight about many other documented anomalies.

- Schwert (2003) shows that anomalies (time-series and cross section) disappear or get attenuated following their discovery. Momentum is an exception.

- Avramov, Chordia, and Goyal (2006) show that implementing short term reversal strategies yields profits that do not survive direct transactions costs and costs related to market impact.

- Wang and Yu (2010) find that the ROA anomaly exists primarily among firms with high arbitrage costs and high information uncertainty, suggesting that mispricing is a culprit.
Avramov, Chordia, Jostova, and Philipov (2007a,b) (2013) show that the momentum, dispersion, and credit risk effects, among others, concentrate in a very small portion of high credit risk high illiquid stocks. Moreover, such strategies are profitable especially due to the short side of the trades.

Chordia, Subrahmanyam, and Tong (2014) find that several anomalies have attenuated significantly over time, particularly in liquid NYSE/AMEX stocks, and virtually none have significantly accentuated. One caveat: their analysis is essentially unconditional.

McLean and Pontiff (2014) study the post publication return predictability of 82 characteristics. They show that most anomalies are based on mispricing not risk.

Cochrane (1999) explains it: if predictability represents risk it is likely to persist even after its discover no matter how many people know about it.

Following Miller (1977), they claim that due to costly short selling there might be overpriced stocks.

Overpricing is especially prominent during high sentiment periods.

Thus, anomalies are profitable during high sentiment periods and are attributable to the short-leg of the trade.
Avramov, Chordia, Jostova, and Philipov (2013), Stambaugh, Yu, and Yuan (2012), and Drechsler and Drechsler (2014) all seem to agree that anomalies represent an un-exploitable stock overvaluation attributable to costly (even infeasible) short selling.

In the absence of undervalued investments, long-only positions cannot generate positive performance.

In my recent work in progress, I show that even the top decile of actively managed mutual funds that invest in supposedly the “right” stocks – or stocks that are the winners of those 11 anomalies considered by Stambaugh, Yu, and Yuan - cannot beat the Fama-French momentum benchmarks.

That is they generate insignificant average alpha in time series regressions of excess fund returns on the market, size, value, and momentum factors.

In contrast, mutual funds investing in the most overpriced stocks do underperform.
• The findings in Avramov, Chordia, Jostova, and Philipov (2013), Stambaugh, Yu, and Yuan (2012), and Drechsler and Drechsler (2014) are consistent not only with Miller (1977) - but with other economic theories that allow for overpricing but preclude underpricing.

• For instance, the Harrison and Kreps (1978) basic insight is that when agents agree to disagree and short selling is impossible, asset prices may exceed their fundamental value.

• The positive feedback economy of De Long, Shleifer, Summers, and Waldmann (1990) also recognizes the possibility of overpricing – arbitrageurs do not sell or short an overvalued asset, but rather buy it, knowing that the price rise will attract more feedback traders.
• But like many other topical issues in financial economics we also have research controversy here from both theory and empirical perspectives. That is, there might be underpricing after all.

• In Diamond and Verrecchia (1987), investors are aware that, due to short sale constraints, negative information is withheld, so individual stock prices reflect an expected quantity of bad news. Prices are correct, on average, in the model, but individual stocks can be overvalued or undervalued.

• Empirically, Boehmer, Huszar, and Jordan (2010) show that heavily trades stocks with low short interest exhibit positive abnormal returns. Short sellers avoid those apparently underpriced stocks.
Corporate Anomalies

• The corporate finance literature has documented a host of other interesting anomalies.

• Fama (1998) for one studies several corporate anomalies.

• Some anomalies are suspected to be under-reaction to news some over reaction.

• Here is a list of prominent anomalies:
  – Stock Split
  – Dividend initiation and omission
  – Stock repurchase
  – Spinoff
  – Merger arbitrage
  – The long horizon performance of IPO and SEO firms.
• External Financing is something noticeable to talk about.

• Finance research has documented negative relation between transactions of external financing and future stock returns.

• In particular, future returns are typically low following IPOs (initial public offerings), SEOs (seasoned public offerings), debt offerings, and bank borrowings.

• Conversely, future stock returns are typically high following stock repurchases.
• Richardson and Sloan (2003) nicely summarize all external financing transactions in one measure.

• They show that their comprehensive measure of external financing has a stronger relation with future returns relative to measures based on individual transactions.

• The external financing measure, denoted by $\Delta XFIN$ is the total cash received from issuance of new debt and equity offerings minus cash used for retirement of existing debt and equity. All components are normalized by the average value of total assets.
1. Asset pricing models > 1.2 CAPM > 1.2.1 Empirical drawbacks

- This measure considers all sorts of equity offerings including common and preferred stocks as well as all sorts of debt offerings including straight bonds, convertible bonds, bank loans, notes, etc. Interest payments on debt as well as dividend payments on preferred stocks are not considered as retiring debt or equity. However, dividend payments on common stocks are considered as retiring equity. In essence, dividends on common stocks are treated as stock repurchases.
1. Asset pricing models > 1.2 CAPM > 1.2.1 Empirical drawbacks

• The $\Delta XFIN$ measure can be decomposed as

$$\Delta XFIN = \Delta CEquity + \Delta PEquity + \Delta Debt$$  \hspace{1cm} (1.2/3)

Where:

• $\Delta CEquity$ is the common equity issuance minus common equity repurchase minus dividend
• $\Delta PEquity$ is the preferred equity issuance minus retirement and repurchase of preferred stocks
• $\Delta Debt$ is the debt issuance minus debt retirement and repurchase.
Multifactor Models:

Multifactor extensions of the CAPM, such as Chen, Roll, and Ross (1986), Fama and French (1993), and Pastor and Stambaugh (2003), associate expected returns with a tendency to move with multiple risk factors.

Recently, Drechsler and Drechsler (2014) propose to augment the Fama-French factors with a factor capturing the impact of impediments on short selling.

On one hand, multifactor models dominate the explanation of the cross section dispersion in expected returns.

At the same time, however,

Multifactor models need not perform better out-of-sample.

The economic appeal of the additional factors is often questionable.
Mutual funds and hedge fund based trading strategies seem to outperform benchmark indexes, even after controlling for market risk. However, multifactor extensions explain most fund persistence.


Before we conclude this section, let me note that according to Roll (1977), the CAPM is untestable since the market return is unobservable and it cannot be measured accurately. Using proxy for the market factor could deliver unreliable inference.
A. The CAPM assumes that the average investor cares only about the performance of his investment portfolio.

- But eventual wealth could come from both investment, labor, and entrepreneurial incomes.

- An additional factor is therefore needed.

- The CAPM says that two stocks that are equally sensitive to market movements must have the same expected return.

- But if one stock performs better in recessions it would be more desirable for most investors who may actually lose their jobs or get lower salaries in recessions.
• The investors will therefore bid up the price of that stock, thereby lowering expected return.

• Thus, pro-cyclical stocks should offer higher average returns than countercyclical stocks, even if both stocks have the same market beta.

• Put another way, co-variation with recessions seems to matter in determining expected returns.

• You may correctly argue that the market tends to go down in recessions.

• But recessions tend to be unusually severe or mild for a given level of market returns.
B. The CAPM applies to single-period — myopic — investors.

Merton (1973) introduces a multi-period version of the CAPM - that is the inter-temporal CAPM (ICAPM).

- In the ICAPM setup, the demand for risky assets is attributed not only to the mean variance component, as in the CAPM, but also to hedging against unfavorable shifts in the investment opportunity set. The hedging demand is something extra relative to the CAPM.

- Merton points out that asset’s risk should be measured by its covariance with the marginal utility of investors, which need not be the same as the covariance with the market return.
Merton shows that multiple state variables that are sources of priced risk are required to explain the cross section variation in expected returns.

In such inter-temporal models, equilibrium expected returns on risky assets may differ from the riskless rate even when they have no systematic (market) risk.

But Merton does not say which other state variables are priced - this gives license to fish factors that work well in explaining the data but at the same time they void any economic content.
Most prominent in the class of inter-temporal models is the consumption CAPM of Rubinstein (1976), Lucas (1978), Breeden (1979), and Grossman and Shiller (1981).

The greatness of the CCAPM is that it is to identify an economically based asset pricing factor.

Indeed, the CCAPM is a single state variable model; real consumption growth is the single factor. See details below.
From Consumption CAPM to Long Run Risk Models

- The CCAPM is a theoretically motivated model, one of the few cases wherein theory invokes empirical studies - not the other way around.
- The CCAPM replaces the market factor by consumption growth.
- Let us derive a simple discrete time version of the CCAPM.
The derivation is based on the textbook treatment of Cochrane (2007).

We use the power utility to describe investor preferences:

\[
    u(c_t) = \frac{1}{1-\gamma} c_t^{1-\gamma} \quad \text{for } \gamma \neq 1, \tag{1.3/1} 
\]

\[
    u(c_t) = \log(c_t) \quad \text{for } \gamma = 1. \tag{1.3/2} 
\]

Notation:

- \(c_t\) denotes consumption at time \(t\)
- \(\gamma\) is the relative risk aversion parameter.
1. Asset pricing models > 1.3 The Consumption CAPM — CCAPM

• The investor must decide how to allocate her wealth between consumption and saving.

• The investor can freely buy or sell any amount of a security whose current price is \( p_t \) and next-period payoff is \( x_{t+1} \) \( (x_{t+1} = p_{t+1} + d_{t+1}) \).

• How much will she buy or sell?

• To find the answer, consider (w.l.o.g.) a two-period investor whose income at time \( s \) is \( e_s \), and let \( y \) be the amount of security she chooses to buy.
The investor’s problem is

\[
\max_y u(c_t) + \mathbb{E}_t[\rho u(c_{t+1})] \quad s.t. \quad c_t = e_t - p_t y, \quad (1.3/3)
\]

\[
c_{t+1} = e_{t+1} + x_{t+1} y, \quad (1.3/5)
\]

where \(\rho\) denotes the impatience parameter, also called the subjective discount factor.
Substituting the constraints into the objective, and setting the derivative with respect to \( y \) equal to zero, we obtain the first order condition for an optimal consumption and portfolio choice,

\[
p_t u'(c_t) = \mathbb{E}_t [\rho u'(c_{t+1}) x_{t+1}]. \quad (1.3/6)
\]

The left hand side in (1.3/6) reflects the loss in utility from consuming less as the investor buys an additional unit of the asset.

The right hand side describes the expected increase in utility obtained from the extra payoff at \( t + 1 \) attributed to this additional unit of the asset.
• A well-known representation for the first order condition is obtained by dividing both sides of \((1.3/6)\) by \(p_t u'(c_t)\),

\[
1 = \mathbb{E}_t(\xi_{t+1} R_{t+1}), \tag{1.3/7}
\]

where \(\xi_{t+1} = \frac{\rho u'(c_{t+1})}{u'(c_t)}\) stands for the pricing kernel, also known as the marginal rate of substitution or the stochastic discount factor, and \(R_{t+1} = \frac{x_{t+1}}{p_t}\) denotes the gross return on the security.

• The relation in equation \((1.3/7)\) is the fundamental discount factor view of asset pricing theories.
Several Insights about the First Order Condition

- Observe from (1.3/7) that the gross risk-free rate, the rate known at time \( t \), which is uncorrelated with the discount factor, is given by \( R_{t,t+1}^f = 1 / \mathbb{E}_t(\xi_{t+1}) \).

- When investor preferences are described by the power utility function, as in equation (1.3/1), the pricing kernel takes the form \( \xi_{t+1} = \rho(c_{t+1}/c_t)^{-\gamma} \).
1. Asset pricing models > 1.3 The Consumption CAPM — CCAPM
   > 1.3.1 Several Insights about the First Order Condition

- Assuming lognormal consumption growth one can show that the continuously compounded risk-free rate is

\[ r_{t,t+1}^f = -\ln(\rho) - \ln \mathbb{E}_t [\exp(-\gamma \Delta \ln c_{t+1})], \]

\[ = -\ln(\rho) + \gamma \mathbb{E}_t (\Delta \ln c_{t+1}) - \frac{\gamma^2}{2} \sigma_t^2 (\Delta \ln c_{t+1}). \] (1.3/8)

- To derive (1.3.8) we have used the useful relation that if \( x \) is normally distributed then

\[ \mathbb{E}(e^{ax}) = e^{\mathbb{E}(ax)} e^{\frac{a^2}{2} \sigma(x)}. \] (1.3/9)

- We can see from (1.3/8) that the EIS (elasticity of inter-temporal substitution) is \( 1/\gamma \) which creates some problems, as discussed below.
Then, we obtain a beta pricing model

\[ E_t(r_{i,t+1}) = r^f_{t,t+1} + \left( \frac{\text{cov}_t(r_{i,t+1}, \xi_{t+1})}{\text{var}_t(\xi_{t+1})} \right) \left( - \frac{\text{var}_t(\xi_{t+1})}{E_t(\xi_{t+1})} \right) \]  

This model is used to estimate the expected return on each security, stock, bond, or option, which should be proportional to the coefficient in the regression of that return on the discount factor.

In words, expected excess return on each security, stock, bond, or option, should be proportional to the coefficient in the regression of that return on the discount factor.

The constant of proportionality, common to all assets, is the risk premium.
• Focusing on the power utility function and using a first order Taylor series expansion, we obtain

\[ \mathbb{E}(r_{i,t+1}) \approx r^f + \beta_{i,\Delta c} \lambda_{\Delta c}, \]  

(1.3/11)

where

\[ \beta_{i,\Delta c} = \frac{\text{cov}(r_{i,t+1}, \Delta c)}{\text{var}(\Delta c)}, \]  

(1.3/12)

\[ \lambda_{\Delta c} = \gamma \text{var}(\Delta c). \]  

(1.3/13)

• This is the discrete time version of the consumption CAPM.
• The relation is exact in continuous time.
1. Asset pricing models > 1.3 The Consumption CAPM — CCAPM
   > 1.3.1 Several Insights about the First Order Condition

• The asset’s risk is defined as the covariance between the asset return and consumption growth.

• The risk premium is obtained as the product of the investor risk aversion and the volatility of consumption growth.

• Notice from equation (1.3/11) that the expected return on an asset is larger the larger the covariance between the return on that asset with consumption growth.
1. Asset pricing models > 1.3 The Consumption CAPM — CCAPM
   > 1.3.1 Several Insights about the First Order Condition

- Intuition: an asset doing badly in recessions (positive covariance) when the investor consumes little, is less desirable than an asset doing badly in expansions (negative covariance) when the investor feels wealthy and is consuming a great deal.

The former asset will be sold for a lower price, thereby having a higher expected return.
Theoretically, The CCAPM Appears Preferable to the Traditional CAPM

• It takes into account the dynamic nature of portfolio decisions

• It integrates the many forms of wealth beyond financial asset wealth

• Consumption should deliver the purest measure of good and bad times as investors consume less when their income prospects are low or if they think future returns will be bad.

• Empirically, however, The CCAPM has been unsuccessful even relative to the CAPM.
CCAPM: The Empirical Evidence

- Hansen and Singleton (1982, 1983) formulate a consumption-based model in which a representative agent has time-separable power utility of consumption.

- They reject the model on U.S. data, finding that it cannot simultaneously explain the time-variation of interest rates and the cross-sectional pattern of average returns on stocks and bonds.
Wheatley (1988) rejects the model based on international data.

Mankiw and Shapiro (1986) show that the CCAPM performs no better, and in many respects even worse, than the CAPM.

They regress the average returns of the 464 NYSE stocks that were continuously traded from 1959 to 1982 on their market betas, on consumption growth betas, and on both betas.
• They find that the market betas are more strongly and robustly associated with the cross section of average returns, and that market beta drives out consumption beta in multiple regressions.

• Breeden, Gibbons, and Litzenberger (1989) find comparable performance of the CAPM and a model that uses a mimicking portfolio for consumption growth as the single factor.
• Cochrane (1996) finds that the static CAPM substantially outperforms the consumption CAPM in pricing size portfolios.

• Academics and practitioners typically use portfolio-based models, such as the CAPM or the Fama and French extension to correct for risk.

• So the CCAPM has not been a great success.

• Recently, Amir and Bansal (2004) suggest the Long Run Risk Model as a replacement to the CCAPM

• Consumption Strikes Back!.
• Indeed, Parker and Julliard (2005), Bansal, Dittmar, and Lundblad (2005), Hansen, Heaton, and Li (2008), and Avramov, Cederburg, and Hore (2013) suggest that if systematic consumption risk is measured over long horizons the CCAPM is able to explain the cross section dispersion in average stock returns and several market anomalies.
The Consumption CAPM and the Equity Premium Puzzle

• From a cross section perspective, the CCAPM fails if consumption beta is unable to explain why average returns differ across stocks, which is indeed the case.

• At the aggregate level (time-series perspective) the CCAPM leads to the so-called equity premium puzzle documented by Mehra and Prescott (1985) as well as the risk-free puzzle.
• To illustrate, let us manipulate the first order condition (1.3/7) (for notational clarity I suppress the time dependence)

\[ 1 = E(\xi R), \quad (1.4/1) \]

\[ = E(\xi)E(R) + \text{cov}(\xi, R), \quad (1.4/2) \]

\[ = E(\xi)E(R) + \rho_{\xi,R} \sigma(\xi) \sigma(R). \quad (1.4/3) \]

• Dividing both sides of (1.4/3) by \( E(\xi) \sigma(R) \) leads to

\[ \frac{E(R) - R^f}{\sigma(R)} = -\rho_{\xi,R} \frac{\sigma(\xi)}{E(\xi)}, \quad (1.4/4) \]
Which implies that

$$\left| \frac{E(R) - R_f}{\sigma(R)} \right| \leq \frac{\sigma(\xi)}{E(\xi)} (= \sigma(\xi) R_f). \quad (1.4/5)$$

- The left hand side in (1.4/5) is known as the Sharpe ratio.
- The highest Sharpe ratio is associated with portfolios lying on the mean-variance efficient frontier.
- Notice that the slope of the frontier is governed by the volatility of the discount factor.
1. Asset pricing models > 1.4 The Consumption CAPM and the Equity Premium Puzzle

• Under the CCAPM it follows that

\[
\left| \frac{E(R_{m,v}) - R_f}{\sigma(R_{m,v})} \right| = \frac{\sigma[(c_{t+1}/c_t)^{-\gamma}]}{E[(c_{t+1}/c_t)^{-\gamma}]}.
\]  (1.4/6)

• When log consumption growth is normally distributed the right hand side of (1.4/6) can be shown to be equal to

\[
\sqrt{e^{\gamma^2 \sigma^2 (\Delta \ln c_{t+1})}},
\]  (1.4/7)

• which can be approximated by

\[
\gamma \sigma (\Delta \ln c).
\]  (1.4/8)
• In words, the slope of the mean-variance efficient frontier is higher if the economy is riskier, i.e., if consumption growth is more volatile, or if investors are more risk averse.

• Over the past several decades in the US, real stock returns have averaged 9% with a std. of about 20%, while the real return on T-Bills has been about 1%.

• Thus, the historical annual market Sharpe ratio has been about 0.4.
Moreover, aggregate nondurable and services consumption growth had a std. of 1%.

This fact can only be reconciled with $\gamma=50$.

But the empirical estimates are between 2 and 10.

This is the “equity premium puzzle.” The historical Sharpe ratio is simply too large than the one obtained with reasonable risk aversion and consumption volatility estimates.
The Risk-Free Rate Puzzle

• Using the standard CCAPM framework also gives rise to the risk-free rate puzzle.

• Recall, we have shown that

\[ r_{t,t+1}^f = -\ln(\rho) + \gamma \mathbb{E}_t(\Delta \ln c_{t+1}) - \frac{\gamma^2}{2} \sigma_t^2(\Delta \ln c_{t+1}) \]  

  \[ (1.5/1) \]

• Using \( \gamma = 2 \) the risk-free rate should be around 5% to 6% per year.

• The actually observed rate is less than 1%
How Shall We Resolve The Equity Premium and Risk-Free Puzzles?

• Perhaps investors are much more risk averse than we may have thought.
  – This indeed resolves the equity premium puzzle.
  – But higher risk aversion parameter implies higher risk-free rate. So higher risk aversion reinforces the risk-free puzzle.

• Perhaps the stock returns over the last 50 years are good luck rather than an equilibrium compensation for risk.
Perhaps something is deeply wrong with the utility function specification and/or the use of aggregate consumption data.

- Indeed, the CCAPM assumes that agents’ preferences are time additive VNM representation (e.g., power).
- Standard power utility preferences impose tight restrictions on the relation between the equity premium and the risk free rate.
- As shown earlier, EIS and the relative risk aversion parameter are reciprocals of each other.
1. Asset pricing models > 1.6 How shall we resolve the equity premium and riskfree puzzles?

- Economically they should not be tightly related.
- EIS is about deterministic consumption paths - it measures the willingness to exchange consumption today with consumption tomorrow for a given risk-free rate, whereas risk aversion is about preferences over random variables.
- Epstein and Zin (1989) and Weil (1990) entertain recursive inter-temporal utility functions that separate risk aversion from elasticity of inter-temporal substitution, thereby separating the equity premium from the risk-free rate.
Duffie and Epstein (1992) introduces the Stochastic Differential Utility which is the continuous time version of Epstein-Zin-Weil.

They show that under certain parameter restrictions, the risk-free rate actually diminishes with higher risk aversion.

Empirically, recursive preferences perform better in matching the data.

Reitz (1988) comes up with an interesting idea: he brings the possibility of low probability states of economic disaster and is able to explain the observed equity premium.

Weitzman (2007) proposes an elegant solution using a Bayesian framework to characterize the ex ante uncertainty about consumption growth.

The asset pricing literature typically assumes that the growth rate is normally distributed.

\[
g \sim N(\mu_g, \sigma_g^2). \tag{1.6/1}\]

The literature also assumes that \(\mu_g\) and \(\sigma_g\) are known to the agents in the economy.
1. Asset pricing models > 1.6 How shall we resolve the equity premium and riskfree puzzles?

- What if you assume that $\mu_g$ is known and $\sigma_g$ is unknown.
- Moreover, you model $\sigma_g$ as inverted gamma distributed random variable.
- Then $g$ has the Student-t distribution.
- The student t distribution captures both the high equity premium and low risk-free rate.
• It should also be noted that there is some recent literature proposing alternative measures of consumption which purportedly help with measurement errors in consumption growth and work better empirically - see e.g., Savov (2011) and Da and Yun (2010).
Long Run Risk

• The LRR of Bansal and Yaron (2004) has been one of the most successful asset pricing theory over the last decade.

• LRR models feature a small but highly persistent component in consumption growth that is hard to capture directly using consumption data.

• Never-the-less that small component is important for asset pricing.
The persistent component is modeled either as stationary (Bansal and Yaron (2004)) or as co-integrated (Bansal, Dittmar, and Kiku (2009)) stochastic process.

The model has been found useful in explaining the equity premium puzzle, size and book to market effects, momentum, long term reversals, risk premiums in bond markets, real exchange rate movements, among others (see a review paper by Bansal (2007)).
• However, all the evidence is based on calibration experiments or in-sample data fitting.

• Ferson, Nallareddy, and Xie (2011) examine the out of sample performance of the LRR paradigm.

• They examine both the stationary and the co-integrated versions of the model, noted earlier.

• They find that the model performs comparably overall to the simple CAPM as well as a co-integrated version outperforms the stationary version.
• Beeler and Campbell (2012) display several weaknesses of LRR models:
  — Too strong predictability of consumption growth by the price-dividend ratio.
  — Too small variance risk premium.
  — Too strong predictive power of return volatility by the price-dividend ratio.
  — Too small discount rate risk versus cash flow risk.
  — In response, Zhou and Zhu (2013) propose an extra volatility factor that helps resolve all these weaknesses.
Types of Multifactor Models

- The poor empirical performance of the single factor (C)CAPM motivates a search for alternative asset pricing models.
- Either include more factors or go along with conditional models or both.
- Multiple factors have been inspired along the spirit of
  The Arbitrage Pricing Theory — APT — (1976) of Ross
  The inter-temporal CAPM (ICAPM) of Merton (1973).
Distinguishing between the APT and ICAPM is often confusing.

Cochrane (2001) argues that the biggest difference between them for empirical work is in the inspiration of factors:

— The APT suggests a statistical analysis of the covariance matrix of returns to find factors that characterize common movements.

— The ICAPM puts some economic meaning to the selected factors.

• Later we will study a Bayesian approach for extracting factors using a panel of stock returns.

• These approaches to extracting factors remain silent about the economic forces that determine factor risk prices.

• It is also difficult to use the extracted factors as benchmarks in studies of fund performance.
The Fama-French Model

• The three-factor FF model is attractive from an empirical perspective.

• FF (1992, 1993) have shown that the cross-sectional variation in expected returns can be captured using the following factors:

1. the return on the market portfolio in excess of the risk free rate of return

2. a zero net investment (spread) portfolio that is long in small firm stocks and short in large firm stocks (SMB)

3. a zero net investment (spread) portfolio that is long in high book-to-market stocks and short in low book-to-market stocks (HML)
• The factors as well as description about constructions are available at Kenneth French’s web site.

• FF (1996) have shown that their model is able to explain many of the market anomalies known back then - excluding the momentum effect.

• Meanwhile we also know the FF model is unable to explain the IVOL effect, the credit risk effect, the dispersion effect, earnings momentum, as well as the net equity issues (net equity issued less the amount of seasoned equity retired).
The six portfolios are constructed at the end of each June as the intersections of 2 portfolios formed on market equity and 3 portfolios formed on the ratio of book equity to market equity.

- The size breakpoint for year t is the median NYSE market equity at the end of June of year t.
- BE/ME for June of year t is the book equity for the last fiscal year end in t-1 divided by ME for December of t-1. The BE/ME breakpoints are the 30th and 70th NYSE percentiles.
• SMB is the average return on the three small portfolios minus the average return on the three big portfolios.

• HML is the average return on the two value portfolios minus the average return on the two growth portfolios.

• FF (1993) argue that their factors are state variables in an ICAPM sense.

• Liew and Vassalou (2000) make a good case for the FF assertion.
• They find that the FF factors forecast GDP growth.

• But in the end, the Fama-French findings are empirically based, lack any theoretical underpinning.

• The model does poorly out of sample.

• And it does not explain a few market anomalies, noted earlier, that had been discovered more recently.
• The statistical tests promoting the Fama-French model are based on equity portfolios (25 size book to market portfolios) not individual securities.

• Indeed, if you run asset pricing tests using individual stocks you will see that the Fama-French performance is rather poor.
Other Multifactor Models

- Carhart (1997) proposes a four-factor model to evaluate performance of equity mutual funds — MKT, SMB, HML, and WML, where WML is a momentum factor.

- He shows that profitability of “hot hands” based trading strategies (documented by Hendricks, Patel, and Zeckhauser (1993)) disappears when investment payoffs are adjusted by WML.

- The profitability of “smart money” based trading strategies in mutual funds (documented by Zheng (1999)) also disappears in the presence of WML.
• Pastor and Stambaugh (2003) propose adding a liquidity factor.

• So we have five major factors to explain equity returns

1. market
2. SMB
3. HML
4. WML
5. Liquidity
• Often bond portfolios serve as common factors (see Ferson and Harvey (1999) and the references therein).

• The collection of factors explaining HF performance is different (see Avramov, Kosowski, Naik, and Teo (2011) and the references therein) to account for non linear payoffs generated by hedge funds.
Two recent apparently competing papers:

- Fama and French (2014) propose a five-factor model which is based on the original market, size, and book-to-market factors and adds investment and profitability factors.

1. Asset pricing models > 1.9 Do multifactor models really perform well?

Do Multifactor Models Really Perform Well?

• The unfortunate news is that SMB, HML, WML, and a liquidity factor do not really capture the entire size, book-to-market, past return, and liquidity effects in individual stock returns.

• See, in particular, Avrmaov and Chordia (2006a).

• They examine pricing abilities of conditional and unconditional models.
1. Asset pricing models > 1.9 Do multifactor models really perform well?

- They first consider a time series specification of the form

\[ r_{it} = \alpha_i + \beta_i (z_{t-1}, z_{it-1})' f_t + \epsilon_{it} \]  

(1.9/1)

for any security in the sample, where \( z_t \) is a macro level conditioning information and \( z_{it} \) denotes firm level conditioning information such as size and book to market.

- Note that the risk adjusted return is obtained as the sum of intercept and residual

\[ r_{i,t}^* = \alpha_i + \epsilon_{it} \]  

(1.9/2)
1. Asset pricing models > 1.9 Do multifactor models really perform well?

• They then consider the cross section regression of risk adjusted returns on stock level size, book to market, past return variables, and turnover.

• If indeed the pricing model does a good job in pricing individual stocks then the slope coefficients in the cross section regressions should be statistically indistinguishable from zero.

• The slope coefficients, however, are highly significant.
1. Asset pricing models > 1.9 Do multifactor models really perform well?

- That is, firm level size, book to market, past return, and turnover are still important even when stock returns are adjusted by the SMB, HML, WML, and liquidity factors.

- We will revisit the Avramov-Chordia methodology in the section dealing with asset pricing test methodologies.

- Ferson and Harvey (1999) show that alpha in multi factor models varies with business conditions.

- Whereas the Ferson and Harvey (1999) metric is statistical, Avramov and Chordia (2006b) demonstrate that mispricing variation with macro variables is economically significant.
Quick Summary

• So the empirical evidence on model pricing abilities is disappointing whether you take single or multi factor models whether you take conditional or unconditional models.

• It is still nice to develop new or improve existing asset pricing theories and test methodologies.

• Special focus should be paid to testing models among individual stocks rather than industry or characteristics sorted portfolios.
• Moreover, we will entertain some Bayesian methods that take the view that asset pricing models, albeit non perfect, are still useful for estimating cost of equity estimation, evaluating fund performance, and selecting optimal portfolios.
Estimating and Evaluating Asset Pricing Models
Asset Pricing Tests Applied to Single Stocks

- Below we describe an asset pricing test that applies to both portfolios as well as single securities.

- The ability to test pricing models using single stocks is sound.

- For one it avoids the data-snooping biases that are inherent in portfolio based approaches, as noted by Lo and MacKinlay (1990).
4. Estimating and Evaluating Asset Pricing Models > 4.1 Another useful cross sectional approach

- It is also robust to the sensitivity of asset pricing tests to the portfolio grouping procedure.


- Avramov and Chordia (2006a) extend the BCS methodology to test pricing models with potential time varying factor loadings.

- The first stage is a time series regression of excess returns on asset pricing factors with time varying factor loadings.
The second stage is a regression of risk adjusted returns on equity characteristics.

Under the null of exact pricing, equity characteristics should be statistically insignificant in the cross-section.

The practice of using risk adjusted returns, rather than gross or excess returns, is intended to address the finite sample bias attributable to errors in estimating factor loadings in the first-pass time series regressions.
Let us formalize the Avramov-Chordia test.

Assume that returns are generated by a conditional version of a $K$-factor model

$$R_{jt} = E_{t-1}(R_{jt}) + \sum_{k=1}^{K} \beta_{jkt-1} f_{kt} + e_{jt},$$  \hspace{1cm} (4.1/1)

where $E_{t-1}$ is the conditional expectations operator, $R_{jt}$ is the return on security $j$ at time $t$, $f_{kt}$ is the unanticipated (with respect to information available at $t - 1$) time $t$ return on the $k'$th factor, and $\beta_{jkt-1}$ is the conditional beta.
• $E_{t-1}(R_{jt})$ is modeled using the exact pricing specification

$$E_{t-1}(R_{jt}) - R_{Ft} = \sum_{k=1}^{K} \lambda_{kt-1} \beta_{jkt-1}, \tag{4.1/2}$$

• where $R_{Ft}$ is the risk-free rate and $\lambda_{kt}$ is the risk premium for factor $k$ at time $t$.

• The estimated risk-adjusted return on each security for month $t$ is then calculated as:

$$R_{jt}^* \equiv R_{jt} - R_{Ft} - \sum_{k=1}^{K} \hat{\beta}_{jkt-1} F_{kt}, \tag{4.1/3}$$

• where $F_{kt} \equiv f_{kt} + \lambda_{kt-1}$ is the sum of the factor innovation and its corresponding risk premium and $\hat{\beta}_{jkt}$ is the conditional beta estimated by a first-pass time-series regression over the entire sample period as per the specification given below.
• Our risk adjustment procedure assumes that the conditional zero-beta return equals the conditional risk-free rate, and that the factor premium is equal to the excess return on the factor, as is the case when factors are return spreads.

• Next, we run the cross-sectional regression

\[ R_{jt}^* = c_{0t} + \sum_{m=1}^{M} c_{mt} Z_{mjt-1} + e_{jt}, \]  

(4.1/4)

• where \( Z_{mjt-1} \) is the value of characteristic \( m \) for security \( j \) at time \( t - 1 \), and \( M \) is the total number of characteristics.
Under exact pricing, equity characteristics do not explain risk-adjusted return, and are thus insignificant in the specification (4.1/4).

To examine significance, we estimate the vector of characteristics rewards each month as

\[ \hat{c}_t = (Z_{t-1}'Z_{t-1})^{-1}Z_{t-1}'R^*_t, \quad (4.1/5) \]

where \( Z_{t-1} \) is a matrix including the \( M \) firm characteristics for \( N_t \) test assets and \( R^*_t \) is the vector of risk-adjusted returns on all test assets.
• To formalize the conditional beta framework developed here let us rewrite the specification in (6.4/4) using the generic form

\[ R_{jt} - [R_{Ft} + \beta(\theta, z_{t-1}, X_{jt-1})'F_t] = c_{0t} + c_tZ_{jt-1} + e_{jt}, \]  

(4.1/6)

• where \( X_{jt-1} \) and \( Z_{jt-1} \) are vectors of firm characteristics, \( z_{t-1} \) denotes a vector of macroeconomic variables, and \( \theta \) represents the parameters that capture the dependence of \( \beta \) on the macroeconomic variables and the firm characteristics.
Ultimately, the null to test is $c_t = 0$.

While we have checked the robustness of our results for the general case where $X_{jt-1} = Z_{jt-1}$, the paper focuses on the case where the factor loadings depend upon firm-level size, book-to-market, and business-cycle variables.

That is, the vector $X_{jt-1}$ stands for size and book-to-market and the vector $Z_{jt-1}$ stands for size, book-to-market, turnover, and various lagged return variables.
• The dependence on size and book-to-market is motivated by the general equilibrium model of Gomes, Kogan, and Zhang (2003), which justifies separate roles for size and book-to-market as determinants of beta.

• In particular, firm size captures the component of a firm’s systematic risk attributable to its growth option, and the book-to-market ratio serves as a proxy for risk of existing projects.
• Incorporating business-cycle variables follows the extensive evidence on time series predictability (see, e.g., Keim and Stambaugh (1986), Fama and French (1989), and Chen (1991)).

• In the first pass, the conditional beta of security $j$ is modeled as

$$\beta_{jt-1} = \beta_1 + \beta_2 z_{t-1} + (\beta_3 + \beta_4 z_{t-1}) \text{Size}_{jt-1} + (\beta_5 + \beta_6 z_{t-1}) \text{BM}_{jt-1},$$  \hspace{1cm} (4.1/7)

• where $\text{Size}_{jt-1}$ and $\text{BM}_{jt-1}$ are the market capitalization and the book-to-market ratio at time $t - 1$. 

4. Estimating and Evaluating Asset Pricing Models > 4.1 Another useful cross sectional approach
The first pass time series regression for the very last specification is

\[ r_{jt} = \alpha_j + \beta_{j1} r_{mt} + \beta_{j2} z_{t-1} r_{mt} + \beta_{j3} Size_{jt-1} r_{mt} \]

\[ + \beta_{j4} z_{t-1} Size_{jt-1} r_{mt} \]

\[ + \beta_{j5} BM_{jt-1} r_{mt} + \beta_{j6} z_{t-1} BM_{jt-1} r_{mt} + u_{jt}, \]  \hspace{1cm} (4.1/8)

where \( r_{jt} = R_{jt} - R_{Ft} \) and \( r_{mt} \) is excess return on the value-weighted market index.

Then, \( R_{jt}^* \) in (6.4/4), the dependent variable in the cross-section regression, is given by \( \alpha_j + u_{jt} \).
• The time series regression (4.1/8) is run over the entire sample.

• While this entails the use of future data in calculating the factor loadings, Fama and French (1992) indicate that this forward looking does not impact any of the results.

• For perspective, it is useful to compare our approach to earlier studies.

• Fama and French (1992) estimate beta by assigning the firm to one of 100 size-beta sorted portfolios. Firm’s beta (proxied by the portfolio’s beta) is allowed to evolve over time when the firm changes its portfolio classification.
• Fama and French (1993) focus on 25 size and book-to-market sorted portfolios, which allow firms’ beta to change over time as they move between portfolios.

• Brennan, Chordia, and Subrahmanyam (1998) estimate beta each year in a first-pass regression using 60 months of past returns. They do not explicitly model how beta changes as a function of size and book-to-market, as we do, but their rolling regressions do allow beta to evolve over time.
• We should also distinguish the beta-scaling procedure in Avramov and Chordia from those proposed by Shanken (1990) and Ferson and Harvey (1999) as well as Lettau and Ludvigson (2001).

• Shanken and Ferson and Harvey use predetermined variables to scale factor loadings in asset pricing tests.

• Lettau and Ludvigson use information variables to scale the pricing kernel parameters.

• In both procedures, a one-factor conditional CAPM can be interpreted as an unconditional multifactor model.
4. Estimating and Evaluating Asset Pricing Models > 4.1 Another useful cross sectional approach

- The beta pricing specification of Avramov and Chordia does not have that unconditional multifactor interpretation since the firm-level $Size_j$ and $BM_j$ are asset specific – that is, they are uncommon across all test assets.
Stock Return Predictability
5. Stock Return Predictability > 5.1 Predictability based on observable macro variables

Predictability Based on Observable Macro Variables

• A question of long-standing interest to both academics and practitioners is whether returns on risky assets are predictable.

• Fama and Schwert (1977), Keim and Stambaugh (1986), Fama and French (1989), and Lettau and Ludvigson (2001), among others, identify ex ante observable variables that predict future returns on stocks and bonds.
The evidence on predictability is typically based upon the joint system

\[ r_t = a + \beta' z_{t-1} + u_t, \]  
\[ z_t = c + \rho z_{t-1} + v_t. \]

• Statistically, predictability means that at least one of the \( \beta \) coefficients is significant at conventional levels.

• Economically, predictability means that you can properly time the market, switching between an equity fund and a money market fund.
5. Stock Return Predictability > 5.2 Is the evidence on predictability robust?

Is The Evidence on Predictability Robust?

• Predictability based on macro variables is still a research controversy:
  
  — Asset pricing theories do not specify which ex ante variables predict asset returns;
  
  — Menzly, Santos, and Veronesi (2006) among others provide some theoretical validity for the predictive power of variables like the dividend yield – but that is an ex post justification.
  
  — Statistical biases in slope coefficients of a predictive regression;
  
  — Potential data mining in choosing the macro variables;
  
  — Poor out-of-sample performance of predictive regressions;
5. Stock Return Predictability > 5.2 Is the evidence on predictability robust?

• Schwert (2003) shows that time-series predictability based on the dividend yield tends to attenuate and even disappears after its discovery.

• Indeed, the power of macro variables to predict the equity premium substantially deteriorates during the post-discovery period.
Potential Data Mining

- Repeated visits of the same database lead to a problem that statisticians refer to as data mining (also model over-fitting or data snooping).

- It reflects the tendency to discover spurious relationships by applying tests inspired by evidence coming up from prior visits to the same database.
Merton (1987) and Lo and MacKinlay (1990), among others, discuss the problems of over-fitting data in tests of financial models.

In the context of predictability, data mining has been investigated by Foster, Smith, and Whaley (FSW - 1997).
Foster, Smith, and Whaley (1997)

- FSW adjust the test size for potential over-fitting, using theoretical approximations as well as simulation studies.
- They assume that
  1. M potential predictors are available.
  2. All possible regression combinations are tried.
  3. Only $m < M$ predictors with the highest $R^2$ are reported.
Their evidence shows:

1. Inference about predictability could be erroneous when potential specification search is not accounted for in the test statistic.

2. Using other industry, size, or country data as a control to guard against variable-selection biases can be misleading.
5. Stock Return Predictability > 5.2 Is the evidence on predictability robust?
   > 5.2.4 The poor out-of-sample performance of predictive regressions

The Poor Out-of-Sample Performance of Predictive Regressions

- A good reference here is Bossaerts and Hillion (1999) and a follow up work by Goyal and Welch (2006).

- The adjustment of the test size merely help in correctly rejecting the null of no relationship.

- It would, however, provide little information if the empiricist is asked to discriminate between competing models under the alternative of existing relation.
BH and GW propose examining predictability using model selection criteria.
- Suppose there are $M$ potential predictors.
- There are $2^M$ competing specifications.
- Select one winning specification based on the adjusted $R^2$, the AIC, the SIC, and other criteria.

The selected model (regardless of the criterion used) always retains predictors – not a big surprise – indicating in-sample predictability.

Implicitly you assign a $\frac{1}{2^M}$ probability that the IID model is correct - so you are biased in favor of detecting predictability.

> 5.2.4 The poor out-of-sample performance of predictive regressions
5. Stock Return Predictability > 5.2 Is the evidence on predictability robust? > 5.2.4 The poor out-of-sample performance of predictive regressions

- The out-of-sample performance of the selected model is always a disaster.
- The Bayesian approach of model averaging improves out-of-sample performance. Coming soon!
- We will also study asset allocation with predictability.
- Before we move on let us note at least two papers responding to the apparently nonexistent out of sample predictability of the equity premium.
5. Stock Return Predictability > 5.2 Is the evidence on predictability robust?
   > 5.2.4 The poor out-of-sample performance of predictive regressions

- Cochrane (2008) points out that we should detect predictability either in returns or dividends.
- Campbell and Thompson (2008) document predictability after restricting the equity premium to be positive.
- Rapach, Strauss, and Zhou (2010) combine model forecast, similar to the Bayesian Model Averaging concept, but using equal weights and considering a small subset of models.
Finite sample bias in the slope coefficient

• Recall, we work with (assuming one predictor for simplicity) the predictive system

\[ r_t = \alpha + \beta z_{t-1} + u_t, \quad (5.2/1) \]

\[ z_t = c + \rho z_{t-1} + v_t. \quad (5.2/2) \]

• Now, let \( \sigma_v^2 \) denote the variance of \( v_t \), and let \( \sigma_{uv} \) denote the covariance between \( u_t \) and \( v_t \).
5. Stock Return Predictability > 5.2 Is the evidence on predictability robust?

>5.2.1 Finite sample bias in the slope coefficient

- We know from Kandall (1954) that the OLS estimate of the persistence parameter $\rho$ is biased, and that the bias is $-1(1 + 3\rho)/T$.

- Stambaugh (1999) shows that under the normality assumption, the finite sample bias in $\hat{\beta}$, the slope coefficient in a predictive regression, is

$$
Bias = \mathbb{E}(\hat{\beta} - \beta) = -\frac{\sigma_{uv}}{\sigma_v^2} \left( \frac{1 + 3\rho}{T} \right).
$$

- The bias can easily be derived.
Note that the OLS estimates of $\beta$ and $\rho$ are

$$\hat{\beta} = (X'X)^{-1}X'R = \beta + (X'X)^{-1}X'U,$$  \hspace{1cm} (5.2/4)

$$\hat{\rho} = (X'X)^{-1}X'Z = \rho + (X'X)^{-1}X'V,$$  \hspace{1cm} (5.2/5)

Where:

$R = [r_1, r_2, ..., r_T]'$, $Z = [z_1, z_2, ..., z_T]'$, 
$U = [u_1, u_2, ..., u_T]'$, $V = [v_1, v_2, ..., v_T]'$, 
$X = [\iota_T, Z_{-1}]$, $\iota_T$ is a $T$-dimension vector of ones 
and $Z_{-1} = [z_0, z_1, ..., z_{T-1}]'$. 

5. Stock Return Predictability

>5.2 Is the evidence on predictability robust?

>5.2.1 Finite sample bias in the slope coefficient
• Also note that $u_t$ can be decomposed into two orthogonal components

$$u_t = \frac{\sigma_{uv}}{\sigma_v^2} v_t + e_t,$$

(5.2/6)

where $e_t$ is uncorrelated with $z_{t-1}$ and $v_t$.

• Hence, the predictive regression slope can be rewritten as

$$\hat{\beta} = \beta + \frac{\sigma_{uv}}{\sigma_v^2} (\hat{\rho} - \rho) + (X'X)^{-1}X'E,$$

(5.2/7)

where $E = [e_1, e_2, ..., e_T]'$.

• Amihud and Hurvich (2004, 2009) advocate a nice approach to deal with statistical inference in the presence of a small sample bias.
Return Distribution and Investment Horizons
Multi Period Returns and Continuous Compounding

- Let $R_t$ be the time $t$ rate of return including capital gain and dividend.

- The continuously compounded (cc) return is

\[ r_t = \ln (1 + R_t) \]  

(3.1/1)

- The holding period return (HPR) over an investment horizon of $K$ periods is

\[ R(t + 1, t + K) = \prod_{k=1}^{K}(1 + R_{t+k}), \]

(3.1/2)

\[ = \exp \left( \sum_{k=1}^{K} r_{t+k} \right). \]

(3.1/3)
• Thus, the continuously compounded HPR is

\[ r_{t+1,t+K} = \sum_{k=1}^{K} r_{t+k} \]  

\[ (3.1/4) \]

• We often assume that the cc return is iid normally distributed with mean \( \mu \) and variance \( \sigma^2 \). Then

\[ r_{t+1,t+K} \sim N(K\mu, K\sigma^2) \]  

\[ (3.1/5) \]

• What is the distribution of HPR?

• Lognormal with mean and variance given by

\[ E[HPR] = \exp\{K(\mu + \sigma^2/2)\}, \]  

\[ (3.1/6) \]

\[ V[HPR] = \exp\{2K(\mu + \sigma^2/2)\} [\exp(K\sigma^2) - 1]. \]  

\[ (3.1/7) \]
• Note that the lognormal distribution is not symmetric. The median of HPR is

$$\text{median}[HPR] = \exp \{ K\mu \},$$

suggesting that the mean to median ratio is \( \exp \left( \frac{K}{2} \sigma^2 \right) \).

• This, in turn, implies that the pdf of the HPR becomes more positively skewed the longer the investment horizon.
Partitioning the Return Interval

Let $r_y$ denote the cc return for period $y$ - say one year. Partition a year into $H$ intervals, and let $r_{y,h} \ h = 1, \ldots, H$ denote the cc return over interval $h$ in period $y$.

Note that

$$r_y = \sum_{h=1}^{H} r_{y,h}.$$  \hspace{1cm} (3.2/1)

If $r_y$ has mean and variance $\mu$ and $\sigma^2$, the moments of $r_{y,h}$ are

$$E\{r_{y,h}\} = \mu_H = \frac{\mu}{H},$$ \hspace{1cm} (3.2/2)

$$var\{r_{y,h}\} = \sigma^2_H = \frac{\sigma^2}{H}.$$ \hspace{1cm} (3.2/3)
• Assume we have a sample of $Y$ years so that there are $n = YH$ observations of the short interval returns.

• Define the short-interval sample moments by $\bar{r}$ and $s^2$ - then the estimators of $\mu$ and $\sigma$ are

\[
\hat{\mu} = H\bar{r}, \quad (3.2/4)
\]

\[
\hat{\sigma}^2 = HS^2. \quad (3.2/5)
\]
• Clearly these estimators are unbiased. Moreover,

\[ \text{var}\{H\bar{r}\} = H^2 \text{var}(\bar{r}), \]

\[ = H^2 \frac{\sigma_H^2}{n}, \]

\[ = \frac{\sigma^2}{Y}, \]

which is the same as if the annual returns were used.

• That is, partitioning the return interval does not improve the estimator of \( \mu \).

• However, as explained below, it does improve the estimator for \( \sigma^2 \).
• In particular,

\[ \text{var}\{Hs^2\} = H^2 \frac{2}{n-1} \sigma_H^4, \quad (3.2/9) \]

\[ = H^2 \frac{2}{HY-1} \frac{\sigma^4}{H^2}, \quad (3.2/10) \]

\[ = \frac{2\sigma^4}{HY-1}. \quad (3.2/11) \]

• Notice that the variance approaches 0 as \( H \) becomes large.

• Andersen et al (2001) among many others exploit this property to estimate the realized variance.
3. Return Distribution and Investment Horizons > 3.3 Variance and investment horizon when returns are predictable

Variance and Investment Horizon when Returns are Predictable

• Let $r_t$ be the cc return in time $t$ and $r_{t+1}$ be the cc return in time $t + 1$.

• Assume that $\text{var}(r_t) = \text{var}(r_{t+1})$.

• Then the cumulative two period return is $R(t, t + 1) = r_t + r_{t+1}$.

• The question of interest: is $\text{var}(R(t, t + 1))$ greater than equal to or smaller than $2 \text{var}(r_t)$. 
• Of course if stock returns are iid the variance grows linearly with the horizon as noted earlier.

• However, let us assume that stock returns can be predictable by the dividend yield.

• That is, we consider the predictive system

\[ r_t = \alpha + \beta div_{t-1} + \epsilon_t \]  \hfill (3.3/1)

\[ div_t = \phi + \delta div_{t-1} + \eta_t \]  \hfill (3.3/2)

where \( var(\epsilon_t) = \sigma_1^2 \), \( var(\eta_t) = \sigma_2^2 \), and \( cov(\epsilon_t, \eta_t) = \sigma_{12} \) and the residuals are uncorrelated in all leads and lags.
It follows that

\[ \text{var}(r_t + r_{t+1}) = 2\sigma_1^2 + \beta^2\sigma_2^2 + 2\beta\sigma_{12} \]  

Thus, if \( \beta^2\sigma_2^2 + 2\beta\sigma_{12} < 0 \) the conditional variance of two period return is less than the twice conditional variance of one period return, which is indeed the case based on empirical evidence.

This property is called mean-reversion.

Then, the longer the investment horizon the less risky equities appear if one is making investment decisions based on the current value of the dividend yield.

Thus one would be willing to invest more in equities over longer investment horizons.
Pastor and Stambaugh (2012) claim that stocks are substantially more volatile over long horizons from an investor’s perspective.

This perspective recognizes that parameters are uncertain, even with two centuries of data, and that observable predictors imperfectly deliver the conditional expected return.

Mean reversion contributes strongly to reducing long-horizon variance but is more than offset by various uncertainties faced by the investor.

The same uncertainties reduce desired stock allocations of long-horizon investors contemplating target-date funds.
Bayesian Econometrics & Asset Pricing
6. Bayesian econometrics and asset pricing

Overview

• To those of you interested in Bayesian methods in economics and finance I would highly recommend using the textbook of Zellner (1971) along with the references below.

• It all starts with the well known Bayes Theorem that deals with conditional probabilities

\[
P(A|B) = \frac{P(A,B)}{P(B)}, \quad (6.1/1)
\]

\[
= \frac{P(B|A)P(A)}{P(B)} \quad (6.1/2)
\]

where \( A \) and \( B \) are events and \( P(A|B) \) is a conditional probability that event \( A \) occurs based on the realization of event \( B \).
Let us take Bayes Theorem into finance.

In particular, consider the time series predictive regression

\[ r_t = a + b z_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2) \]  (6.1/3)

where \( r_t \) is excess stock return realized at time \( t \) and \( z_{t-1} \) is a variable observed at time \( t - 1 \).

Let us denote the set of parameters by \( \Theta \) where \( \Theta = [a, b, \sigma^2]' \).

For a frequentist (classical) econometrician \( \Theta \) is fixed yet unknown.

The frequentist estimates \( \Theta \) using parametric (ML) or non parametric (GMM) methods.
Then the ML/GMM estimate $\hat{\Theta}$ is asymptotically normally distributed.

The Bayesian philosophy is quite different.

$\Theta$ is stochastic, not fixed!!!!!

You can learn about $\Theta$ based on its posterior distribution.

That is, given prior views about $\Theta$ as well as evidence coming up from data - the posterior summarizes everything we know about $\Theta$. 
For simplicity, let us assume the predictor is non-stochastic and let 
$r = [r_1, r_2, ..., r_T]'$ and $x = [x_0, x_1, ... x_{T-1}]'$ such that $x_t = [1, z_t']'$. Then the Bayes Theorem becomes

$$P(\Theta|r) = \frac{P(r|\Theta)p(\Theta)}{P(r)} \quad (6.1/4)$$

where $P(\Theta|r)$ is the posterior density of $\Theta$ - the density of interest, $P(r|\Theta)$ is the likelihood function, and $P(\Theta)$ is the prior distribution of $\Theta$. 

6. Bayesian econometrics and asset pricing> 6.1 Overview
• Notice that \( P(r) \) does not tell us anything about \( \Theta \). So we can write

\[
P(\Theta|r) \propto P(r|\Theta)p(\Theta).
\] (6.1/5)

• Of course, since a probability distribution function integrates to unity it follows that

\[
P(\Theta|r) = \frac{P(r|\Theta)p(\Theta)}{\int_{\Theta} P(r|\Theta)p(\Theta) \, d\Theta}.
\] (6.1/6)

• The posterior density could have an analytic reduced form expression if prior beliefs are of the conjugate form or if the prior density is diffuse or improper, as illustrated below.

• But there are many cases in which the exact form of the posterior is unknown.
Then we can often employ very powerful MCMC techniques to simulate the posterior density.

The implementation of MCMC is major advantage of Bayesian econometrics.

Indeed, there are many applications in finance (mostly in asset pricing but also in corporate finance) that use MCMC methods including Gibbs Sampling, Metropolis Hastings, and Importance Sampling.
6. Bayesian econometrics and asset pricing > 6.1 Overview

• Other advantages to using Bayesian econometrics include:
  — Eliciting economically motivated prior beliefs
  — Incorporating model uncertainty
  — Accounting for estimation risk

• We will discuss all these issues in the subsections that follow.
Let’s start with the tractable case of conjugate priors.

Consider the multivariate form of the predictive regression (6.1/3)

\[ R = XB + U , \]  

(6.2/1)

where

\[ vec(U) \sim N(0, \Sigma \otimes I_T). \]  

(6.2/2)

Moreover, \( R \) is a \( T \times N \) matrix of excess returns, \( X \) is a \( T \times (m + 1) \) matrix with \( m \) being the number of predictors, and \( U \) is a \( T \times N \) matrix of residuals.
6. Bayesian econometrics and asset pricing > 6.2 The Case of Conjugate Priors

- The priors for $B$ and $\Sigma$ are the normal inverted Wishart

\[
P(b|\Sigma) \propto |\Sigma|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} (b - b_0)'[\Sigma^{-1} \otimes \Psi_0](b - b_0) \right), \tag{6.2/3} \]

\[
P(\Sigma) \propto |\Sigma|^{-\frac{\nu_0+N+1}{2}} \exp \left( -\frac{1}{2} \text{tr}[S_0\Sigma^{-1}] \right), \tag{6.2/4} \]

where

\[
b = vec(B) \tag{6.2/5} \]

and $b_0$, $\Psi_0$, and $S_0$ are prior parameters to be specified by the researcher.
The likelihood function of normally distributed data constituting the actual sample obeys the form

\[
P(R|B, \Sigma, X) \propto |\Sigma|^{-\frac{T}{2}} \exp \left( -\frac{1}{2} \text{tr}[(R - XB)'(R - XB)]\Sigma^{-1} \right), \tag{6.2/6}
\]

which can be rewritten in a very convenient form

\[
P(R|B, \Sigma, X) \propto |\Sigma|^{-\frac{T}{2}} \exp \left( -\frac{1}{2} \text{tr}[\hat{S} + (B - \hat{B})'X'X(B - \hat{B})]\Sigma^{-1} \right) \tag{6.2/7}
\]

where

\[
\hat{S} = (R - X\hat{B})'(R - X\hat{B}), \tag{6.2/8}
\]

\[
\hat{B} = (X'X)^{-1}X'R. \tag{6.2/9}
\]
• An equivalent representation for the likelihood function (8.2/7) is given by

\[ P(R|b, \Sigma, X) \propto |\Sigma|^{-\frac{T}{2}} \exp \left( -\frac{1}{2} (b - \hat{b})'[\Sigma^{-1} \otimes (X'X)](b - \hat{b}) \right), \]

\[ \times \exp \left( -\frac{1}{2} \text{tr}[\hat{S}\Sigma^{-1}] \right), \]  

where

\[ \hat{b} = \text{vec}(\hat{B}). \]  

(6.2/10)

• Combining the likelihood (8.2/10) with the prior and completing the square on \( b \) yield

\[ P(b|\Sigma, R, X) \propto |\Sigma|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} (b - \tilde{b})'[\Sigma^{-1} \otimes F](b - \tilde{b}) \right) \]  

(6.2/12)
6. Bayesian econometrics and asset pricing > 6.2 The Case of Conjugate Priors

\[ P(\Sigma|R, X) \propto |\Sigma|^{-\frac{\nu+N+1}{2}} \exp \left( -\frac{1}{2} \text{tr}[\tilde{S}\Sigma^{-1}] \right), \quad (6.2/13) \]

- where

\[ F = \Psi_0 + X'X, \quad (6.2/14) \]
\[ \tilde{b} = \text{vec}(\tilde{B}), \quad (6.2/15) \]
\[ \tilde{B} = F^{-1}X'X\hat{B} + F^{-1}\Psi_0B_0, \quad (6.2/16) \]
\[ \tilde{S} = \hat{S} + S_0 + \hat{B}'X'X\hat{B} + B_0'\Psi_0B_0 - \tilde{B}'F\tilde{B}, \quad (6.2/17) \]
\[ \nu = T_0 + T. \quad (6.2/18) \]

- So the posterior for \( B \) is normal and for \( \Sigma \) is inverted Wishart.
• That is the conjugate prior idea - the prior and posterior have the same distributions but with different parameters.

• Interestingly, $\tilde{B}$ is a weighted average of $B_0$ and $\hat{B}$. That is,

$$\tilde{B} = WB_0 + (I - W)\hat{B},$$

(6.2/19)

where $W = I - F^{-1}X'X$. 

What if the posterior does not obey a reduced form expression?

- Let us demonstrate the Gibbs Sampling - an appealing Bayesian numerical method - in the context of measuring pricing errors.

- The analysis below is based on “Measuring the Pricing Error of the Arbitrage Pricing Theory” by Geweke and Zhou (RFS 1996) who test the pricing abilities of the APT (Ross (1976)).
6. Bayesian econometrics and asset pricing
   > 6.3 What if the posterior does not obey a reduced form expression?

• The basic APT model assumes that returns on $N$ risky portfolios are related to $K$ pervasive unknown factors.

• The relation is described by the $K$ factor model

\[ r_t = \mu + \beta f_t + \epsilon_t, \]  

(6.3/1)

• where $r_t$ denotes returns (not excess returns) on $N$ assets and $f_t$ is a set of $K$ factor innovations (factors are not pre-specified by are latent).
Specifically,

\[ E\{f_t\} = 0, \]
\[ E\{f_t f'_t\} = I_K, \]
\[ E\{\epsilon_t|f_t\} = 0, \]
\[ E\{\epsilon_t \epsilon'_t|f_t\} = \Sigma = \text{diag}(\sigma_1^2, \ldots, \sigma_N^2), \]
\[ \beta = [\beta_1, \ldots, \beta_K]. \]

Moreover, under exact APT, the \( \mu \) vector satisfies the restriction

\[ \mu = \lambda_0 + \beta_1 \lambda_1 + \cdots + \beta_K \lambda_K. \]  

(6.3/2)
• The original APT model is about an approximated relation.

• The objective throughout is to explore a measure that summarizes the deviation from exact pricing.

• That measure is denoted by $Q^2$ and is given by

$$Q^2 = \frac{1}{N} \mu' [I_N - \beta^* (\beta'^* \beta^*)^{-1} \beta'^*] \mu,$$

where

$$\beta^* = [I_N, \beta].$$
6. Bayesian econometrics and asset pricing
   > 6.3 What if the posterior does not obey a reduced form expression?

- Recovering the sampling distribution of $Q^2$ is hopeless.
- However, using the Gibbs sampling technique, we can simulate the posterior distribution of $Q^2$.
- Specifically, we assume that observed returns and latent factors are jointly normally distributed.
- That is, we have

$$\begin{bmatrix} f_t \\ r_t \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ \mu \end{bmatrix}, \begin{bmatrix} I_K & \beta' \\ \beta & \beta' + \Sigma \end{bmatrix} \right).$$ (6.3/5)
6. Bayesian econometrics and asset pricing

> 6.3 What if the posterior does not obey a reduced form expression?

- Here are some additional notations:
  - Data: $R = [r'_1, ..., r'_T]'$
  - Parameters: $\Theta = [\mu', vec(\beta)', vech(\Sigma)']'$
  - Latent factors: $f = [f'_1, ..., f'_T]'$

- To evaluate the pricing error we need draws from the posterior distribution $P(\Theta|R)$.

- We draw from the joint posterior in a slightly different manner than that suggested in the paper.
First, we employ a multivariate regression setting. Moreover, the well-known identification problem is not accounted for to simplify the analysis.

The prior is improper

\[ P_0(\Theta) \propto |\Sigma|^{-\frac{1}{2}} = (\sigma_1 \ldots \sigma_N)^{-1}. \] (6.3/6)

Re-expressing the arbitrage pricing equation, we obtain:

\[ r_t' = F_t' B' + \epsilon_t', \] (6.3/7)

where

\[ F_t' = [1, f_t'], \] (6.3/8)

\[ B = [\mu, \beta]. \] (6.3/9)
Rewriting the system in a matrix notation, we get

\[ R = F B' + E \]  \hspace{1cm} (6.3/10)

Why do we need to use the Gibbs sampling technique?

Because the likelihood function \( P(R|\theta) \) (and therefore the posterior density) cannot be expressed analytically.
• However, $P(R|\Theta, F)$ does obey an analytical form:

$$P(R|\Theta, F) \propto |\Sigma|^{-\frac{T}{2}} \exp \left[ -\frac{1}{2} \text{tr}\{[R - FB']'[R - FB']\Sigma^{-1}\} \right] \quad (6.3/11)$$

• Therefore, we can compute the full conditional posterior densities:

$$P(B|\Sigma, F, R), \quad (6.3/12)$$
$$P(\Sigma|B, F, R), \quad (6.3/13)$$
$$P(F|B, \Sigma, R). \quad (6.3/14)$$
• The Gibbs sampling chain is formed as follows:

1. Specify starting values $B^{(0)}, \Sigma^{(0)},$ and $F^{(0)}$ and set $i = 1.$

2. Draw from the full conditional distributions:
   - Draw $B^{(i)}$ from $P(B | \Sigma^{(i-1)}, F^{(i-1)}, R)$
   - Draw $\Sigma^{(i)}$ from $P(\Sigma | B^{(i)}, F^{(i-1)}, R)$
   - Draw $F^{(i)}$ from $P(F | B^{(i)}, \Sigma^{(i)}, R)$

3. Set $i = i + 1$ and go to step 2.

> 6.3 What if the posterior does not obey a reduced form expression?
6. Bayesian econometrics and asset pricing
   > 6.3 What if the posterior does not obey a reduced form expression?

- After \( m \) iterations the sample \( B^{(m)}, \Sigma^{(m)}, F^{(m)} \) is obtained.

- Under mild regularity conditions (see, for example, Tierney, 1994),
  \((B^{(m)}, \Sigma^{(m)}, F^{(m)})\) converges in distribution to the relevant marginal and joint distributions:

\[
P(B^{(m)}|R) \to P(B|R), \quad (6.3/15)
\]
\[
P(\Sigma^{(m)}|R) \to P(\Sigma|R), \quad (6.3/16)
\]
\[
P(F^{(m)}|R) \to P(F|R), \quad (6.3/17)
\]
\[
P(B^{(m)}, \Sigma^{(m)}, F^{(m)}|R) \to P(B, \Sigma, F|R). \quad (6.3/18)
\]
What if the posterior does not obey a reduced form expression?

For \( m \) large enough, the \( G \) values

\[
(B^{(g)}, \Sigma^{(g)}, F^{(g)})_{g=m+1}^{m+G}
\]

are a sample from the joint posterior.

A remaining task is to find the full conditional posterior densities.

Note:

\[
P(B|\Sigma, F, R) \propto \exp \left\{ -\frac{1}{2} \text{tr}[R - FB']^T[R - FB']\Sigma^{-1} \right\}.
\] (6.3/19)
6. Bayesian econometrics and asset pricing

> 6.3 What if the posterior does not obey a reduced form expression?

- Let $b = \text{vec}(B')$. Then

$$P(b|\Sigma, F, R) \propto \exp \left\{ -\frac{1}{2} [b - \hat{b}]' (\Sigma^{-1} \otimes (F'F)) [b - \hat{b}] \right\},$$

(6.3/20)

where

$$\hat{b} = \text{vec}[(F'F)^{-1} F'R].$$

(6.3/21)
Therefore,

\[ b|\Sigma, F, R \sim N(\hat{b}, \Sigma \otimes (F'F)^{-1}) \] (6.3/22)

Also note:

\[ P(\sigma_i|B, F, R) \propto \sigma_i^{-(T+1)} \exp \left(-\frac{TS_i^2}{2\sigma_i^2}\right), \] (6.3/23)

where \( TS_i^2 \) is the i-th diagonal element of the \( N \times N \) matrix

\[ [R - FB']' [R - FB'], \] (6.3/24)

suggesting that:

\[ \frac{TS_i^2}{\sigma_i^2} \sim \chi^2(T). \] (6.3/25)
Finally,

\[ f_t | \mu, \beta, \Sigma, r_t \sim N(M_t, H_t), \]  \hspace{1cm} (6.3/26)

where

\[ M_t = \beta' (\beta \beta' + \Sigma)^{-1} (r_t - \mu), \]  \hspace{1cm} (6.3/27)

\[ H_t = I_K - \beta' (\beta \beta' + \Sigma)^{-1} \beta. \]  \hspace{1cm} (6.3/28)
Incorporating Model Uncertainty - Bayesian Model Averaging

- We noted earlier that the sample evidence about return predictability had been quite disappointing.

- So should we dismiss predictability altogether due to the absence of out of sample predictability? Not so fast!
Avramov (2002) concludes that Bossaerts and Hillion fail to discover out-of-sample predictability because they ignore model uncertainty.


Indeed, in the heart of the model selection approach, one uses a specific criterion to select a single model and then operates as if the model is correct with a unit probability.
• What about the other $2^M - 1$ competing specifications?

• Avramov proposes examining the sample evidence on predictability using Bayesian Model Averaging (BMA) that considers all models.

• Specifically, BMA computes posterior probabilities for all $2^M$ models and uses the probabilities as weights on the individual models to obtain a composite weighted model.
• Statistical and economic inferences on predictability are made based on the weighted model.

• We can also use the “Bayesian t ratio" to examine predictability.

• Let $p(M_j)$ be the posterior probability of model $j$, let $\beta_j$ be the vector (of size $M$) of the slope coefficients, and let $Var(\beta_j)$ be the $(M \times M)$ variance covariance matrix of the errors in estimating $\beta$. 
Recall, the ‘usual’ $t$ ratio is obtained by dividing $\beta_j$ by its standard error.

Let us now compute the $t$-ratio when model uncertainty is accounted for.

The numerator would be

$$\beta = \sum_{j=1}^{2^M} p(\mathcal{M}_j) \beta_j$$  \hspace{0.5cm} (6.4/1)

The denominator would be the square root of the diagonal elements in

$$\text{Var}(\beta) = \sum_{j=1}^{2^M} p(\mathcal{M}_j) \left[ \text{Var}(\beta_j) + (\beta_j - \beta)(\beta_j - \beta)' \right]$$  \hspace{0.5cm} (6.4/2)
• Another metric: cumulative probability

• The idea here is to compute the cumulative probability for every predictor.

• Let $A$ be a $2^M \times M$ matrix representing all forecasting models by zeros (exclusion of predictors) and ones (inclusion).

• Let $\mathcal{P}$ be the $2^M \times 1$ vector of posterior probabilities.

• The cumulative probability of each predictor is given by $A'\mathcal{P}$. 
## 6. Bayesian econometrics and asset pricing

> 6.4 Incorporating model uncertainty - Avramov (2002) - Bayesian Model Averaging

The empirical evidence: Table 2 in Avramov (2002)

<table>
<thead>
<tr>
<th>Predictive Variables</th>
<th>Div</th>
<th>BM</th>
<th>EY</th>
<th>WML</th>
<th>Def</th>
<th>Tbill</th>
<th>RET</th>
<th>DEF</th>
<th>TERM</th>
<th>Jan</th>
<th>Inf</th>
<th>SMB</th>
<th>HML</th>
<th>Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>Portfolio:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SL</td>
<td>0.38</td>
<td>0.08</td>
<td>0.20</td>
<td>0.02</td>
<td>0.14</td>
<td>0.28</td>
<td>0.48</td>
<td>0.04</td>
<td>0.12</td>
<td>0.21</td>
<td>0.31</td>
<td>0.16</td>
<td>0.05</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SM</td>
<td>0.16</td>
<td>0.06</td>
<td>0.12</td>
<td>0.02</td>
<td>0.10</td>
<td>0.09</td>
<td>0.40</td>
<td>0.06</td>
<td>0.54</td>
<td>0.77</td>
<td>0.15</td>
<td>0.12</td>
<td>0.02</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SH</td>
<td>0.07</td>
<td>0.05</td>
<td>0.06</td>
<td>0.03</td>
<td>0.06</td>
<td>0.04</td>
<td>0.49</td>
<td>0.03</td>
<td>0.35</td>
<td>1.00</td>
<td>0.06</td>
<td>0.08</td>
<td>0.02</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BL</td>
<td>0.14</td>
<td>0.05</td>
<td>0.12</td>
<td>0.04</td>
<td>0.14</td>
<td>0.13</td>
<td>0.05</td>
<td>0.07</td>
<td>0.20</td>
<td>0.03</td>
<td>0.69</td>
<td>0.05</td>
<td>0.05</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BM</td>
<td>0.15</td>
<td>0.06</td>
<td>0.15</td>
<td>0.03</td>
<td>0.20</td>
<td>0.34</td>
<td>0.03</td>
<td>0.09</td>
<td>0.54</td>
<td>0.07</td>
<td>0.23</td>
<td>0.04</td>
<td>0.03</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>BH</td>
<td>0.06</td>
<td>0.06</td>
<td>0.07</td>
<td>0.03</td>
<td>0.09</td>
<td>0.09</td>
<td>0.02</td>
<td>0.03</td>
<td>0.17</td>
<td>0.92</td>
<td>0.21</td>
<td>0.02</td>
<td>0.02</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>
Out-of-sample performance: Table 6 in Avramov (2002) – (part 1 of 3)

### The Rolling Scheme – Monthly Sample

<table>
<thead>
<tr>
<th></th>
<th>T0 = 50</th>
<th>T0 = 100</th>
<th>T0 = 25</th>
<th>All</th>
<th>iid</th>
<th>Adj</th>
<th>$R^2$</th>
<th>AIC</th>
<th>FIC</th>
<th>PIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPE</td>
<td>0.0006</td>
<td>0.0007</td>
<td>0.0003</td>
<td>-0.0006</td>
<td>0.0007</td>
<td>-0.0002</td>
<td>0.0000</td>
<td>-0.0023</td>
<td>0.0001</td>
<td>-0.0003</td>
</tr>
<tr>
<td>t-statistic</td>
<td>0.4225</td>
<td>0.4944</td>
<td>0.2368</td>
<td>-0.3874</td>
<td>0.5126</td>
<td>-0.1551</td>
<td>0.0176</td>
<td>-1.5588</td>
<td>0.0365</td>
<td>-0.2117</td>
</tr>
<tr>
<td>Efficiency</td>
<td>-0.0563</td>
<td>-0.0287</td>
<td>-0.2335</td>
<td>-0.7874</td>
<td>-0.4371</td>
<td>-0.7642</td>
<td>-0.7919</td>
<td>-0.9454</td>
<td>-0.8709</td>
<td>-0.7926</td>
</tr>
<tr>
<td>t-statistic</td>
<td>-0.1788</td>
<td>-0.0863</td>
<td>-0.8557</td>
<td>-7.8065</td>
<td>-1.3761</td>
<td>-7.4691</td>
<td>-6.9512</td>
<td>-7.9715</td>
<td>-7.4193</td>
<td>-7.2763</td>
</tr>
<tr>
<td>Serial Correlation</td>
<td>0.0397</td>
<td>0.0499</td>
<td>0.0323</td>
<td>-0.0284</td>
<td>0.0684</td>
<td>-0.0043</td>
<td>-0.0051</td>
<td>0.0274</td>
<td>-0.0185</td>
<td>-0.0269</td>
</tr>
<tr>
<td>t-statistic</td>
<td>0.6676</td>
<td>0.8326</td>
<td>0.5494</td>
<td>-0.4856</td>
<td>1.1288</td>
<td>-0.0738</td>
<td>-0.0895</td>
<td>0.5024</td>
<td>-0.3167</td>
<td>-0.4659</td>
</tr>
<tr>
<td>MSE</td>
<td>0.2137</td>
<td>0.2141</td>
<td>0.2139</td>
<td>0.2333</td>
<td>0.2155</td>
<td>0.2309</td>
<td>0.2298</td>
<td>0.2319</td>
<td>0.2339</td>
<td>0.2312</td>
</tr>
</tbody>
</table>
### The Recursive Scheme – Monthly Sample

<table>
<thead>
<tr>
<th></th>
<th>T0 = 50</th>
<th>T0 = 100</th>
<th>T0 = 25</th>
<th>All</th>
<th>iid</th>
<th>Adj</th>
<th>$R^2$</th>
<th>AIC</th>
<th>FIC</th>
<th>PIC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MPE</strong></td>
<td>-0.0003</td>
<td>-0.0004</td>
<td>-0.0003</td>
<td>0.0005</td>
<td>-0.0010</td>
<td>0.0007</td>
<td>0.0012</td>
<td>0.0013</td>
<td>0.0028</td>
<td>0.0020</td>
</tr>
<tr>
<td><strong>t-statistic</strong></td>
<td>-0.1049</td>
<td>-0.1659</td>
<td>-0.1421</td>
<td>0.1847</td>
<td>-0.3924</td>
<td>0.3018</td>
<td>0.5103</td>
<td>0.5099</td>
<td>1.1455</td>
<td>0.8152</td>
</tr>
<tr>
<td><strong>Efficiency</strong></td>
<td>-0.2357</td>
<td>-0.1047</td>
<td>-0.4018</td>
<td>-0.6708</td>
<td>-0.4500</td>
<td>-0.5959</td>
<td>-0.5804</td>
<td>-0.7953</td>
<td>-0.7319</td>
<td>-0.7300</td>
</tr>
<tr>
<td><strong>t-statistic</strong></td>
<td>-0.6675</td>
<td>-0.2572</td>
<td>-1.3455</td>
<td>-3.0407</td>
<td>-0.9504</td>
<td>-2.8158</td>
<td>-2.7292</td>
<td>-3.7994</td>
<td>-3.0175</td>
<td>-2.9242</td>
</tr>
<tr>
<td><strong>Serial Correlation</strong></td>
<td>0.0401</td>
<td>0.0489</td>
<td>0.0372</td>
<td>0.0036</td>
<td>0.0706</td>
<td>0.0144</td>
<td>0.0143</td>
<td>0.0417</td>
<td>0.0144</td>
<td>0.0120</td>
</tr>
<tr>
<td><strong>t-statistic</strong></td>
<td>0.6728</td>
<td>0.8079</td>
<td>0.6320</td>
<td>0.0655</td>
<td>1.1414</td>
<td>0.2597</td>
<td>0.2572</td>
<td>0.7386</td>
<td>0.2663</td>
<td>0.2194</td>
</tr>
<tr>
<td><strong>MSE</strong></td>
<td>0.2133</td>
<td>0.2133</td>
<td>0.2143</td>
<td>0.2231</td>
<td>0.2155</td>
<td>0.2197</td>
<td>0.2189</td>
<td>0.2260</td>
<td>0.2237</td>
<td>0.2239</td>
</tr>
</tbody>
</table>
6. Bayesian econometrics and asset pricing
   > 6.4 Incorporating model uncertainty - Avramov (2002) - Bayesian Model Averaging

### Out-of-sample performance: Table 6 in Avramov (2002) – (part 3 of 3)

**MSE’s for the Quarterly Sample**

<table>
<thead>
<tr>
<th></th>
<th>$T_0 = 50$</th>
<th>$T_0 = 100$</th>
<th>$T_0 = 25$</th>
<th>All</th>
<th>iid</th>
<th>Adj $R^2$</th>
<th>AIC</th>
<th>FIC</th>
<th>PIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rolling</td>
<td>0.7546</td>
<td>0.7577</td>
<td>0.7651</td>
<td>0.9333</td>
<td>0.7777</td>
<td>0.9041</td>
<td>0.8629</td>
<td>0.9286</td>
<td>0.9347</td>
</tr>
<tr>
<td>Recursive</td>
<td>0.7757</td>
<td>0.7678</td>
<td>0.7930</td>
<td>0.8312</td>
<td>0.7781</td>
<td>0.8163</td>
<td>0.8233</td>
<td>0.8952</td>
<td>0.8170</td>
</tr>
</tbody>
</table>
Bayesian Asset Allocation
7. Bayesian Asset Allocation

- This section is based upon a review paper of Avramov and Zhou (2010) published in the *Annual Review of Financial Economics*.
- We first study asset allocation when stock returns are assumed to be unpredictable or IID.
- We then incorporate potential return predictability.
Asset Allocation
When Returns Are IID
The Mean Variance Framework and Estimation Risk

- Assume there are $N + 1$ assets, one of which is riskless and others are risky.

- Denote by $r_{ft}$ and $r_t$ the rates of returns on the riskless asset and the risky assets at time $t$, respectively.

- Then:

$$R_t \equiv r_t - r_{ft}1_N$$  \hspace{1cm} (7.1.1/1)

are excess returns on the $N$ risky assets.
• Assume that the joint distribution of $R_t$ is iid over time, with mean $\mu$ and covariance matrix $V$.

• In the static mean-variance framework an investor at time $T$ chooses his portfolio weights $w$, so as to maximize his quadratic objective function

$$U(w) = E[R_p] - \frac{\gamma}{2} \text{Var}[R_p] = w'\mu - \frac{\gamma}{2}w'Vw,$$

(7.1.1/2)
where $R_p = w' R_{T+1}$ is the future uncertain portfolio return at time $T + 1$ and $\gamma$ is the coefficient of relative risk aversion.

- When both $\mu$ and $V$ are assumed to be known, the optimal portfolio weights are
  \[ w^* = \frac{1}{\gamma} V^{-1} \mu, \]  
  \[ (7.1.1/3) \]

- and the maximized expected utility is
  \[ U(w^*) = \frac{1}{2\gamma} \mu' V^{-1} \mu = \frac{\theta^2}{2\gamma}, \]  
  \[ (7.1.1/4) \]
• where $\theta^2 = \mu'V^{-1}\mu$ is the squared Sharpe ratio of the ex ante tangency portfolio of the risky assets.

• This is the well known mean-variance theory pioneered by Markowitz (1952).

• In practice, the problem is that $w^*$ is not computable because $\mu$ and $V$ are unknown. As a result, the above mean-variance theory is usually applied in two steps.

• In the first step, the mean and covariance matrix of the asset returns are estimated based on the observed data.
Given a sample size of $T$, the standard maximum likelihood estimators are

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} R_t ,$$  \hspace{1cm} (7.1.1/5)

$$\hat{\mathbf{V}} = \frac{1}{T} \sum_{t=1}^{T} (R_t - \hat{\mu})(R_t - \hat{\mu})'.$$ \hspace{1cm} (7.1.1/6)

Then, in the second step, these sample estimates are treated as if they were the true parameters, and are simply plugged in to compute the estimated optimal portfolio weights,

$$\hat{\mathbf{w}}^{\text{ML}} = \frac{1}{\gamma} \hat{\mathbf{V}}^{-1} \hat{\mu}.$$ \hspace{1cm} (7.1.1/7)
The two-step procedure gives rise to a parameter uncertainty problem because it is the estimated parameters, not the true ones, that are used to compute the optimal portfolio weights.

Consequently, the utility associated with the plug-in portfolio weights can be substantially different from \( U(w^*) \).

Denote by \( \theta \) a vector of all the parameters (both \( \mu \) and \( V \)).

Mathematically, the two-step procedure maximizes the expected utility conditional on the estimated parameters, denoted by \( \hat{\theta} \), being equal to the true ones,

\[
\max_{w} \left[ U(w) \mid \theta = \hat{\theta} \right].
\]  

(7.1.1/8)

and the uncertainty or estimation errors are ignored.
Intuitively, two risky assets with the same estimated expected returns of 10% can be entirely different if the first one has an estimation error of 20%, while the second has 1%. However, both are treated as the same in the above two-step procedure.

The Bayesian approach does account for estimation risk.

Under the normality assumption, and under the standard diffuse prior,

\[ p_0(\mu, V) \propto |V|^{-\frac{N+1}{2}}, \quad (7.1.1/9) \]
• The posterior distribution is (see, e.g., Zellner (1971)),

\[ p(\mu, V | \Phi_T) = p(\mu | V, \Phi_T) \times p(V | \Phi_T) \quad (7.1.1/10) \]

• with

\[ p(\mu | V, \Phi_T) \propto |V|^{-1/2} \exp\{-\frac{1}{2} \text{tr}[T(\mu - \hat{\mu})(\mu - \hat{\mu})'V^{-1}]\}, \quad (7.1.1/11) \]

\[ P(V) \propto |V|^{-\nu/2} \exp\{-\frac{1}{2} \text{tr} V^{-1}(T\hat{V})\}, \quad (7.1.1/12) \]

• where \( R \) is a \( T \times N \) matrix formed by the returns and \( \nu = T + N \).
• The predictive distribution is:

\[ p(R_{T+1} | \Phi_T) \propto |V + (R_{T+1} - \hat{\mu}) \left(R_{T+1} - \frac{\hat{\mu}}{T+1}\right)|\frac{T}{2} \]

which is a multivariate \( t \)-distribution with \( T - N \) degrees of freedom.

• While the problem of estimation error is recognized by Markowitz (1952), it is only in the 70s that this problem receives serious attention.

• Winkler (1973) and Winkle and Barry (1975) are earlier examples of Bayesian studies on predicting prices and portfolio choice.
• Brown (1976, 1978) and Klein and Bawa (1976) lay out independently and clearly the Bayesian predictive density approach, especially Brown (1976) who explains thoroughly the estimation error problem and the associated Bayesian approach.

• Later, Bawa, Brown, and Klein (1979) provide an excellent review of the early literature.

• Under the diffuse prior, (7.1.1/9), it is known that the Bayesian optimal portfolio weights are

\[
\hat{W}_{\text{Bayes}} = \frac{1}{\gamma} \left( \frac{T-N-2}{T+1} \right) \hat{V}^{-1} \hat{\mu}.
\]  

(7.1.1/14)
• In contrast with the classical weights $\hat{w}^{ML}$, the Bayesian holds the portfolio of proportion to $\hat{V}^{-1}\hat{\mu}$, but the proportional coefficient is $(T - N - 2)/(T + 1)$ instead of 1.

• The coefficient can be substantially smaller when $N$ is large relative to $T$. Intuitively, the assets are riskier in the Bayesian framework since parameter uncertainty is an additional source of risk and this risk is now factored into the portfolio decision.

• As a result, the overall position in the risky assets are generally smaller than before.
• However, in the classical framework, $\hat{w}^{\text{ML}}$ is a biased estimator of the true weights since, under the normality assumption,

$$E\hat{w}^{\text{ML}} = \frac{T-N-2}{T} w^* \neq w^*. \quad (7.1.1/15)$$

• Let

$$\tilde{V}^{-1} = \frac{T}{T-N-2} V^{-1}, \quad (7.1.1/16)$$

then $\tilde{V}^{-1}$ is an unbiased estimator of $V^{-1}$.

• The unbiased estimator of $w^*$ is

$$\tilde{w} = \frac{1}{\gamma} \frac{T-N-2}{T} \tilde{V}^{-1} \hat{\mu}, \quad (7.1.1/17)$$

• which is a scalar adjustment of $\hat{w}^{\text{ML}}$. 

7. Bayesian Asset Allocation > 7.1 Asset allocation when returns are IID > 7.1.1 The Mean Variance Framework and Estimation Risk
• The unbiased classical weights differ from their Bayesian counterparts by a scalar $T/(T + 1)$.

• The difference is independent of $N$, and is negligible for all practical sample sizes $T$.

• Hence, parameter uncertainty makes little difference between Bayesian and classical approaches if the diffuse prior is used.

• Therefore, to provide new insights, it is important for a Bayesian to use informative priors, which is a decisive advantage of the Bayesian approach that can incorporate useful information easily into portfolio analysis.
Conjugate Prior

• The conjugate prior, which retains the same class of distributions, is a natural and common information prior on any problem.

• This prior in our context assumes a normal prior for the mean and inverted prior for $V$,

$$\mu \mid V \sim N(\mu_0, \frac{1}{\tau} V), \quad (7.1.2/1)$$

$$V \sim IW(V_0, \nu_0), \quad (7.1.2/2)$$
where $\mu_0$ is the prior mean, and $\tau$ is a prior parameter reflecting the prior precision of $\mu_0$, and $\nu_0$ is a similar parameter on $V$.

Under this prior, the posterior distribution of $\mu$ and $V$ are of the same form as the case for the diffuse prior, except that now the posterior mean of $\mu$ is given by

$$\tilde{\mu} = \frac{\tau}{T+\tau} \mu_0 + \frac{T}{T+\tau} \hat{\mu}.$$ (7.1.2/3)

This says that the posterior mean is simply a weighted average of the prior and sample.
Similarly, $V_0$ can be updated by

$$\tilde{V} = \frac{T+1}{T+(\nu_0+N-1)} \left( V_0 + T\hat{V} + \frac{T\tau}{T+\tau} (\mu_0 - \hat{\mu})(\mu_0 - \hat{\mu})' \right), \quad (7.1.2/4)$$

which is a weighted average of the prior, sample and deviations of $\hat{\mu}$ from $\mu_0$.

Frost and Savarino (1986) provide an interesting case of the conjugate prior by assuming the assets are with identical means, variances, and patterned covariances, a priori.

They find that such a prior improves performance.

This prior is related the well known $1/N$ rule that invests equally across $N$ assets.
Hyper-parameter Prior

• By introducing hyper-parameters $\eta$ and $\lambda$, Jorion (1986) uses the following prior on $\mu$,

$$ p_0(\mu | \eta, \lambda) \propto |V|^{-1} \exp\left\{ -\frac{1}{2} (\mu - \eta 1_N)' (\lambda V)^{-1} (\mu - \eta 1_N) \right\} \quad (7.1.3/1) $$

• in which $\eta$ and $\lambda$ govern the prior distribution of $\mu$. 

7. Bayesian Asset Allocation > 7.1 Asset allocation when returns are IID 
> 7.1.3 Hyperparameter Prior
Using diffuse priors on both $\eta$ and $\lambda$, and integrating them out from a suitable distribution, the predictive distribution of the future asset return can be obtained as usual.

Based on this, Jorion (1986) obtains

$$w_{\text{PJ}} = \frac{1}{\gamma} (\hat{V}_{\text{PJ}})^{-1} \hat{\mu}_{\text{PJ}},$$

where

$$\hat{\mu}_{\text{PJ}} = (1 - \hat{\nu})\hat{\mu} + \hat{\nu}\hat{\mu}_g 1_N,$$

$$\hat{V}_{\text{PJ}} = \left( 1 + \frac{1}{T+\hat{\lambda}} \right) \tilde{V} + \frac{\hat{\lambda}}{T(T+1+\hat{\lambda})} \frac{1_N 1_N'}{1_N' \tilde{V}^{-1} 1_N},$$

(7.1.3/2)

(7.1.3/3)

(7.1.3/4)
\[ \hat{\nu} = \frac{N+2}{(N+2)+T(\hat{\mu} - \hat{\mu}_g 1_N)'\tilde{V}^{-1}(\hat{\mu} - \hat{\mu}_g 1_N)}, \] 

(7.1.3/5)

\[ \hat{\lambda} = \frac{(N + 2)/[(\hat{\mu} - \hat{\mu}_g 1_N)'\tilde{V}^{-1}(\hat{\mu} - \hat{\mu}_g 1_N)]} \]

(7.1.3/6)

• with \( \tilde{V} = T\hat{\nu} / (T - N - 2) \) an adjusted sample covariance matrix, and 
\( \hat{\mu}_g = 1_N'\tilde{V}^{-1}\hat{\mu}/1_N'\tilde{V}^{-1}1_N \) the average excess return on the sample global minimum-variance portfolio.
• An alternative motivation of Jorion (1986) portfolio rule is from a shrinkage perspective by considering the Bayes-Stein estimator of the expected return, \( \mu \), with

\[
\hat{\mu}_{BS} = (1 - v)\hat{\mu} + v\mu_g 1_N, \tag{7.1.3/7}
\]

• where \( \mu_g 1_N \) is the shrinkage target with

\[
\mu_g = 1_N'V^{-1}\mu / 1_N'V^{-1}1_N,
\]

and \( v \) is the weight given to the target.

• Jorion (1986) as well as subsequent studies find that \( w^{PJ} \) improves \( w^{ML} \) substantially, implying that it does so also for the Bayesian strategy under the diffuse prior.
The Black-Litterman Model

The Black-Litterman (BL) Approach for Estimating Mean Returns:

Basics, Extensions, and Incorporating Market Anomalies
• It is notoriously difficult to propose good estimates for mean returns.

• The sample means are quite noisy.

• Perhaps asset pricing models - even if misspecified - could give a good guidance.

• To illustrate, you consider a K-factor model (factors are portfolio spreads) and you run the time series regression

\[ r_t^e = \alpha + \beta_1 f_{1t} + \beta_2 f_{2t} + \cdots + \beta_K f_{Kt} + e_t \]
Then the estimated excess mean return is given by

\[ \hat{\mu}^e = \hat{\beta}_1 \hat{\mu}_{f_1} + \hat{\beta}_2 \hat{\mu}_{f_2} + \cdots + \hat{\beta}_K \hat{\mu}_{f_K} \]

where \( \hat{\beta}_1, \hat{\beta}_2 \ldots \hat{\beta}_K \) are the sample estimates of factor loadings, and \( \hat{\mu}_{f_1}, \hat{\mu}_{f_2} \ldots \hat{\mu}_{f_K} \) are the sample estimates of the factor mean returns.
The BL Mean Returns

- The BL approach combines a model (CAPM) with some views, either relative or absolute, about expected returns.

- The BL vector of mean returns is given by

\[
\mu_{BL} = \left[ \left( \tau_{1 \times 1} \Sigma_{N \times N} \right)^{-1} + P' \Omega^{-1} P \right]^{-1} \left[ \left( \tau_{1 \times 1} \Sigma_{N \times N} \right)^{-1} \mu_{eq} + P' \Omega^{-1} \mu^v \right] \tag{7.1.4/1}
\]
Understanding the BL Formulation

• We need to understand the essence of the following parameters, which characterize the mean return vector: $\Sigma, \mu^{eq}, P, \tau, \Omega, \mu^v$

• Starting from the $\Sigma$ matrix - you can choose any feasible specification either the sample covariance matrix, or the equal correlation, or an asset pricing based covariance.
Constructing Equilibrium Expected Returns

- The \( \mu^{eq} \), which is the equilibrium vector of expected return, is constructed as follows.

- Generate \( \omega_{MKT} \), the N ×1 vector denoting the weights of any of the N securities in the market portfolio based on market capitalization.

- Of course, the sum of weights must be unity.
Then, the price of risk is \( \gamma = \frac{\mu_m - R_f}{\sigma_m^2} \) where \( \mu_m \) and \( \sigma_m^2 \) are the expected return and variance of the market portfolio.

Later, we will justify this choice for the price of risk.

One could pick a range of values for \( \gamma \) and examine performance of each choice.
• If you work with monthly observations, then switching to the annual frequency does not change \( \gamma \) as both the numerator and denominator are multiplied by 12.

• Having at hand both \( \omega_{MKT} \) and \( \gamma \), the equilibrium return vector is given by

\[
\mu^{eq} = \gamma \Sigma \omega_{MKT} \quad (7.1.4/2)
\]

• This vector is called neutral mean or equilibrium expected return.
To understand why, notice that if you have a utility function that generates the tangency portfolio of the form

\[ w_{TP} = \frac{\Sigma^{-1} \mu^e}{\Sigma^{-1} \mu^e} \quad (7.1.4/3) \]

then using \( \mu^{eq} \) as the vector of excess returns on the N assets would deliver \( \omega_{MKT} \) as the tangency portfolio.
What if you Directly Apply the CAPM?

• The question being – would you get the same vector of equilibrium mean return if you directly use the CAPM?

• Yes, if...
The CAPM Based Expected Returns

- Under the CAPM the vector of excess returns is given by

$$\mu^e = \beta_{N \times 1} \mu^e_m \quad (7.1.4/4)$$

$$\beta = \frac{\text{cov} (r^e, r^e_m)}{\sigma^2_m} = \frac{\text{cov} (r^e, (r^e)'^w_{MKT})}{\sigma^2_m} = \frac{\Sigma w_{MKT}}{\sigma^2_m} \quad (7.1.4/5)$$

**CAPM:**

$$\mu^e_{N \times 1} = \frac{\Sigma w_{MKT}}{\sigma^2_m} \mu^e_m = \gamma \Sigma w_{MKT} \quad (7.1.4/6)$$
• Since \( \mu^e_m = (\mu^e)'w_{MKT} \) and \( r^e_m = (r^e)'w_{MKT} \) \((7.1.4/7)\)
then \( \mu^e = \frac{\mu^e_m}{\sigma^2_m} \sum w_{MKT} = \mu^{eq} \) \((7.1.4/8)\)

• So indeed, if you use (i) the sample covariance matrix, rather than any other specification, as well as (ii)
\[
\gamma = \frac{\mu_m - R_f}{\sigma^2_m} \quad \text{(7.1.4/9)}
\]
then the BL equilibrium expected returns and expected returns based on the CAPM are identical.
The P Matrix: Absolute Views

• In the BL approach the investor/econometrician forms some views about expected returns as described below.

• P is defined as that matrix which identifies the assets involved in the views.

• To illustrate, consider two "absolute" views only.

• The first view says that stock 3 has an expected return of 5% while the second says that stock 5 will deliver 12%.
In general the number of views is $K$.

In our case $K=2$.

Then $P$ is a $2 \times N$ matrix.

The first row is all zero except for the fifth entry which is one.

Likewise, the second row is all zero except for the fifth entry which is one.
Let us consider now two "relative views".

Here we could incorporate market anomalies into the BL paradigm.

As noted earlier, anomalies are cross sectional patterns in stock returns unexplained by the CAPM.

Example: price momentum, earnings momentum, value, size, accruals, credit risk, dispersion, and volatility.
Black-Litterman: Momentum and Value Effects

- Let us focus on price momentum and the value effects.
- Assume that both momentum and value investing outperform.
- The first row of \( P \) corresponds to momentum investing.
- The second row corresponds to value investing.
- Both the first and second rows contain \( N \) elements.
Winner, Loser, Value, and Growth Stocks:

- Winner stocks are the top 10% performers during the past six months.
- Loser stocks are the bottom 10% performers during the past six months.
- Value stocks are 10% of the stocks having the highest book-to-market ratio.
- Growth stocks are 10% of the stocks having the lowest book-to-market ratios.
Momentum and Value Payoffs

- The momentum payoff is a return spread – return on an equal weighted portfolio of winner stocks minus return on equal weighted portfolio of loser stocks.

- The value payoff is also a return spread – the return differential between equal weighted portfolios of value and growth stocks.
Back to the P Matrix

• Suppose that the investment universe consists of 100 stocks

• The first row gets the value 0.1 if the corresponding stock is a winner (there are 10 winners in a universe of 100 stocks).

• It gets the value -0.1 if the corresponding stock is a loser (there are 10 losers).

• Otherwise it gets the value zero.
• The same idea applies to value investing.

• Of course, since we have relative views here (e.g., return on winners minus return on losers) then the sums of the first row and the sum of the second row are both zero.

• More generally, if N stocks establish the investment universe and moreover momentum and value are based on deciles (the return difference between the top and bottom deciles) then the winner stock is getting $10/N$ while the loser stock gets $-10/N$. 

• The same applies to value versus growth stocks.

• Rule: the sum of the row corresponding to absolute views is one, while the sum of the row corresponding to relative views is zero.
Computing the $\mu^v$ Vector

- It is the $K \times 1$ vector of $K$ views on expected returns.

- Using the absolute views above
  
  \[ \mu^v = [0.05, 0.12]' \]

- Using the relative views above, the first element is the payoff to momentum trading strategy (sample mean); the second element is the payoff to value investing (sample mean).
The $\Omega$ Matrix

- $\Omega$ is a $K \times K$ covariance matrix expressing uncertainty about views.
- It is typically assumed to be diagonal.
- In the absolute views case described above $\Omega(1,1)$ denotes uncertainty about the first view while $\Omega(2,2)$ denotes uncertainty about the second view – both are at the discretion of the econometrician/investor.
In the relative views described above: $\Omega(1,1)$ denotes uncertainty about momentum. This could be the sample variance of the momentum payoff.

$\Omega(2,2)$ denotes uncertainty about the value payoff. This is the could be the sample variance of the value payoff.
Deciding Upon $\tau$

• There are many debates among professionals about the right value of $\tau$.

• From a conceptual perspective it should be $1/T$ where $T$ denotes the sample size.

• You can pick $\tau = 0.1$

• You can also use other values and examine how they perform in real-time investment decisions.
Maximizing Sharpe Ratio

• The remaining task is to run the maximization program

\[
\max_w \frac{w'\mu_{BL}}{\sqrt{w'\Sigma w}} \tag{7.1.4/10}
\]

• such that each of the \(w\) elements is bounded below by 0 and subject to some agreed upon upper bound, as well as the sum of the \(w\) elements is equal to one.
Consider a sample of size $T$, e.g., $T=60$ monthly observations.

Let us estimate the mean and covariance ($V$) of our $N$ assets based on the sample.
Then the vector of expected return that serves as an input for asset allocation is given by

\[ \mu = \left[ \Delta^{-1} + \left( \frac{V_{\text{sample}}}{T} \right)^{-1} \right]^{-1} \cdot \left[ \Delta^{-1} \mu_{\text{BL}} + \left( \frac{V_{\text{sample}}}{T} \right)^{-1} \mu_{\text{sample}} \right] \]  

(7.1.4/11)

where

\[ \Delta = \left[ (\tau \Sigma)^{-1} + P' \Omega^{-1} P \right]^{-1} \]
Asset Pricing Prior

• Pástor (2000) and Pástor and Stambaugh (2000) introduce interesting priors that reflect an investor’s degree of belief in an asset pricing model.

• To see how this class of priors is formed, assume \( R_t = (y_t, x_t) \), where \( y_t \) contains the excess returns of \( m \) non-benchmark positions and \( x_t \) contains the excess returns of \( K \) (\( = N - m \)) benchmark positions.
Consider a factor model multivariate regression

\[ y_t = \alpha + Bx_t + u_t, \]  

(7.1.5/1)

where \( u_t \) is an \( m \times 1 \) vector of residuals with zero means and a non-singular covariance matrix \( \Sigma = V_{11} - BV_{22}B', \) and \( \alpha \) and \( B \) are related to \( \mu \) and \( V \) through

\[ \alpha = \mu_1 - B\mu_2, \quad B = V_{12}V_{22}^{-1}, \]  

(7.1.5/2)

where \( \mu_i \) and \( V_{ij} \) \((i, j = 1,2)\) are the corresponding partition of \( \mu \) and \( V \),

\[ \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad V = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}. \]  

(7.1.5/3)
For a factor-based asset pricing model, such as the three-factor model of Fama and French (1993), the restriction is $\alpha = 0$.

To allow for mispricing uncertainty, Pástor (2000), and Pástor and Stambaugh (2000) specify the prior distribution of $\alpha$ as a normal distribution conditional on $\Sigma$,

$$\alpha | \Sigma \sim N \left[ 0, \sigma_\alpha^2 \left( \frac{1}{s_\Sigma^2} \Sigma \right) \right], \quad (7.1.5/4)$$

where $s_\Sigma^2$ is a suitable prior estimate for the average diagonal elements of $\Sigma$. The above alpha-Sigma link is also explored by MacKinlay and Pástor (2000) in the classical framework.
• The magnitude of $\sigma_\alpha$ represents an investor’s level of uncertainty about the pricing ability of a given model.

• When $\sigma_\alpha = 0$, the investor believes dogmatically in the model and there is no mispricing uncertainty.

• On the other hand, when $\sigma_\alpha = \infty$, the investor believes that the pricing model is entirely useless.

• Although the above prior is motivated from asset pricing consideration, it also has a shrinkage interpretation.
• The prior on $\alpha$ implies a prior mean on $\mu$, say $\mu_0$.

• It can be shown that the predictive mean is

$$\mu_p = \tau \mu_0 + (1 - \tau) \hat{\mu}, \quad (7.1.5/5)$$

• Where $\tau$ depends upon the sample size and the confidence of belief in the pricing model.

• Pástor (2000) and Pástor and Stambaugh (2000) find that the asset pricing priors make a substantial CER difference in portfolio decisions.
Objective Prior

• Previous priors are placed on $\mu$ and $V$, not on the solution to the problem, the portfolio weights.

• In many applications, supposedly innocuous diffuse priors on some basic model parameters can actually imply rather strong prior convictions about particular economic dimensions of the problem.
• For examples, in the context of testing portfolio efficiency, Kandel, McCulloch, and Stambaugh (1995) find that the diffuse prior in fact implies a strong prior on inefficiency of a given portfolio.

• Tu and Zhou (2009) show that the diffuse prior implies a large prior differences in portfolios weights across assets.

• In short, diffuse priors can be unreasonable in an economic sense in some applications.

• As a result, it is important to use informative priors on the model parameters that can imply reasonable priors on functions of interest.
Tu and Zhou (2009) advocate a method of constructing priors based on a prior on the solution of an economic objective.

In maximizing an economic objective, even before the Bayesian investor observes any data, he is likely to have some idea about the range of the solution.

This will allow him to form a prior on the solution, from which the prior on the parameters can be backed out.
In other words, to maximizing the mean-variance utility here, an investor may have a prior on the portfolio weights, such as an equal- or value-weighted portfolios of the underlying assets.

This prior can then be transformed into a prior on $\mu$ and $V$.

This prior on $\mu$ and $V$ implies a reasonable prior on the portfolio weights by construction.

Because of the way such priors on the primitive parameters are motivated, they are called objective-based priors.
Formally, for the current portfolio choice problem, the objective-based prior starts from a prior on $\mathbf{w}$,

$$w \sim N(w_0, V_0 V^{-1}/\gamma). \quad (7.1.6/1)$$

where $w_0$ and $V_0$ are suitable prior constants with known values, and then back out a prior on $\mu$,

$$\mu \sim N \left[ \gamma V w_0, \sigma_\rho^2 \left( \frac{1}{s^2} V \right) \right], \quad (7.1.6/2)$$

where $s^2$ is the average of the diagonal elements of $V$.

The prior on $V$ can be taken as the usual inverted Wishart distribution.
Investing in Mutual Funds

- Baks, Metrik, and Wachter (2001) (henceforth BMW) and Pastor and Stambaugh (2002a b) have explored the role of prior information about fund performance in making investment decisions.

- BMW consider a mean variance optimizing investor who is quite skeptical about the ability of a fund manager to pick stocks and time the market.
They find that even with a high degree of skepticism about fund performance, the investor would allocate considerable amounts to actively managed funds.

Pastor and Stambaugh nicely extend the BMW methodology.

In BMW as well as past papers studying mutual fund performance, performance is typically defined as the intercept in the regression of the fund’s excess returns on excess return of one or more benchmark assets.
Pastor and Stambaugh recognize the possibility that the intercept in such regressions could be a mix of fund performance as well as model mispricing.

In particular, consider the case wherein benchmark assets used to define fund performance are unable to explain the cross section dispersion of passive assets, that is, the sample alpha in the regression of non benchmark passive assets on benchmarks assets is nonzero.

Then model mispricing emerges in the performance regression.
• Thus, Pastor and Stambaugh formulate prior beliefs on both performance and mispricing.

• Geczy, Stambaugh, and Levin (2005) apply the Pastor Stambaugh methodology to study the cost of investing in socially responsible mutual funds.

• Comparing portfolios of these funds to those constructed from the broader fund universe reveals the cost of imposing the socially responsible investment (SRI) constraint on investors seeking the highest Sharpe ratio.
7. Bayesian Asset Allocation > 7.1 Asset allocation when returns are IID
   > 7.1.7 Investing in Mutual Funds

- This SRI cost depends crucially on the investor’s views about asset pricing models and stock-picking skill by fund managers.
- BMW and Pastor and Stambaugh assume that the prior on alpha is independent across funds.
- Jones and Shanken (2002) show that under such an independence assumption, the maximum posterior mean alpha increases without bound as the number of funds increases and "extremely large" estimates could randomly be generated.
• This is true even when fund managers have no skill.

• Thus they propose incorporating prior dependence across funds.

• Then, investors aggregate information across funds to form a general belief about the potential for abnormal performance.

• Each fund’s alpha estimate is shrunk toward the aggregate estimate, mitigating extreme views.
Let us extend the asset allocation framework to the case where future returns are potentially predictable by macro-wide variables such as the dividend yield.
One-period Models

- Consider a one-period optimizing investor who must allocate at time $T$ funds between the value-weighted NYSE index and one-month Treasury bills.
The investor makes portfolio decisions based on estimating the predictive system

\[ r_t = a + b'z_{t-1} + u_t, \tag{7.3/1} \]
\[ z_t = \theta + \rho z_{t-1} + v_t, \tag{7.3/2} \]

where \( r_t \) is the continuously compounded NYSE return in month \( t \) in excess of the continuously compounded T-bill rate for that month, \( z_{t-1} \) is a vector of \( M \) predictive variables observed at the end of month \( t - 1 \), \( b \) is a vector of slope coefficients, and \( u_t \) is the regression disturbance in month \( t \).
The evolution of the predictive variables is essentially stochastic.

Typically a first order vector auto-regression is employed to model that evolution.

The regression residuals are assumed to obey the normal distribution.

In particular, let \( \eta_t = [u_t, v_t']' \) then \( \eta_t \sim N(0, \Sigma) \) where

\[
\Sigma = \begin{bmatrix}
\sigma_u^2 & \sigma_{uv} \\
\sigma_{vu} & \Sigma_v
\end{bmatrix}.
\]
7. Bayesian Asset Allocation > 7.3 One-period Models

- The distribution of $r_{T+1}$, the time $T + 1$ NYSE excess return, conditional on data and model parameters is $N(a + b' z_T, \sigma_u^2)$.

- Assuming the inverted Wishart prior distribution for $\Sigma$ and multivariate normal prior for the intercept and slope coefficients in the predictive system, the Bayesian predictive distribution $P(r_{T+1} | \Phi_T)$ obeys the Student t density.
Then, considering a power utility investor with parameter of relative risk aversion denoted by $\gamma$ the optimization formulation is

$$\omega^* = \arg \max_{\omega} \int_{r_{T+1}} \left[ (1-\omega) \exp(r_f) + \omega \exp(r_f + r_{T+1}) \right]^{1-\gamma} P(r_{T+1}|\Phi_T) dr_{T+1} \quad (7.3/4)$$

subject to $\omega$ being nonnegative.

It is infeasible to have analytic solution for the optimal portfolio.
• However, it can easily be solved numerically. In particular, given $G$ independent draws for $R_{T+1}$ from the predictive distribution, the optimal portfolio is found by implementing a constrained optimization code to maximize the quantity

$$\frac{1}{G} \sum_{g=1}^{G} \left\{ (1-\omega) \exp (r_f) + \omega \exp (r_f) + R_{T+1}^{(g)} \right\}^{1-\gamma}$$

(7.3/5)

• subject to $\omega$ being nonnegative.

• Kandel and Stambaugh (1996) show that even when the statistical evidence on predictability, as reflected through the $R^2$ is the predictive regression, is weak, the current values of the predictive variables, $z_T$, can exert a substantial influence on the optimal portfolio.
Multi-period Models

• Implementing long horizon asset allocation in a buy-and-hold setup is quite straightforward.

• In particular, let $K$ denote the investment horizon, and let $R_{T+K} = \sum_{k=1}^{K} r_{T+k}$ be the cumulative (continuously compounded) return over the investment horizon.
Avramov (2002) shows that the distribution for $R_{T+K}$ conditional on the data (denoted $\Phi_T$) and set of parameters (denoted $\Theta$) is given by

$$R_{T+K}, \Theta, \Phi_T \sim N(\lambda, \gamma),$$  \hspace{1cm} (7.4/1)

where

$$\lambda = Ka + b'\left[(\rho^K - I_M)(\rho - I_M)^{-1}\right]z_T$$

$$+ b'[(\rho^{k-1} - I_M)(\rho - I_M)^{-1} - (K - 1)I_M](\rho - I_M)^{-1}\theta \hspace{1cm} (7.4/2)$$

$$\gamma = K\sigma_u^2 + \sum_{k=1}^K \delta(k)\Sigma_v\delta(k)' + \sum_{k=1}^K \sigma_{uv}\delta(k)' + \sum_{k=1}^K \delta(k)\sigma_{vu} \hspace{1cm} (7.4/3)$$

$$\delta(k) = b'\left[(\rho^{k-1} - I_M)(\rho - I_M)^{-1}\right]. \hspace{1cm} (7.4/4)$$
• Drawing from the Bayesian predictive distribution is done in two steps.
• First, draw the model parameters $\Theta$ from their posterior distribution.
• Second, conditional on model parameters, draw $R_{T+K}$ from the normal distribution formulated in (7.4/1) - (7.1/4).
• The optimal portfolio can then be found using (7.3/5) where $R_{T+K}$ replaces $R_{T+1}$ and $Kr_f$ replaces $r_f$. 
• It should be noted that incorporating dynamic rebalancing, intermediate consumption, and learning could establish a non trivial challenge for recovering the optimal portfolio choice.

• Brandt, Goyal, Santa Clara, and Stroud (2005) address the challenge using a tractable simulation based method.

• Notice that the iid set-up corresponds to $b = 0$ in the predictive regression (7.3/1), which yields $\lambda_{iid} = Ka$ and $\gamma_{iid} = K\sigma_u^2$ in (7.4/2) and (7.4/3).
• The conditional mean and variance in an iid world increase linearly with the investment horizon.

• Thus, there is no horizon effect when (i) returns are iid and (ii) estimation risk is not accounted for, as indeed shown by Samuelson (1969) and Merton (1969) in an equilibrium framework.

• Incorporating estimation risk, Barberis (2000) shows that the allocation to equity diminishes with the investment horizon, as stocks appear to be riskier in longer horizons.
Incorporating return predictability and estimation risk, Barberis (2000) shows that investors allocate considerably more heavily to equity the longer their horizon.

Recently, Pastor and Stambaugh (2012) implement a predictive system to show that stocks may not be less risky over long horizons from an investment perspective.
An essential question is what are the benefits of using the Bayesian approach in studying asset allocation with predictability?

There are at least three important features of the Bayesian approach including (i) the ability to account for estimation risk and model uncertainty, (ii) the feasibility of powerful and tractable simulation methods, and (iii) the ability to elicit economically based prior beliefs.
• Indeed, unlike in the single period case wherein estimation risk plays virtually no role, estimation risk does play an important role in long horizon investments with predictability.

• Barberis shows that a long horizon investor who ignores it may over-allocate to stocks by a sizeable amount.

• Further, advances in computational Bayesian methods facilitates tractable solutions of fairly complex portfolio choice problems.
To illustrate, even when the predictors evolve stochastically, both Kandel and Stambaugh (1996) and Barberis (2000) assume that the initial value of the predictive variables $z_0$ is non-stochastic.

With stochastic initial value the distribution of future returns conditioned on model parameters does not longer obey a well known distributional form.

There are other several powerful numerical Bayesian algorithms such as the Gibbs Sampler and data augmentation (see a review paper by Chib and Greenberg (1996)) which make the Bayesian approach broadly applicable.
Model Uncertainly and Asset Allocation

• Financial economists have been identified variables that predict future stock returns, as noted earlier.

• However, the “correct” predictive regression specification has remained an open issue for several reasons.
• For one, existing equilibrium pricing theories are not explicit about which variables should enter the predictive regression.

• This aspect is undesirable, as it renders the empirical evidence subject to data over-fitting concerns.

• Indeed, as noted earlier, Bossaerts and Hillion (1999) confirm in-sample return predictability, but fail to demonstrate out-of-sample predictability.

• Moreover, the multiplicity of potential predictors also makes the empirical evidence difficult to interpret.
• For example, one may find an economic variable statistically significant based on a particular collection of explanatory variables, but often not based on a competing specification.

• Given that the true set of predictive variables is virtually unknown, the Bayesian methodology of model averaging, described below, is attractive, as it explicitly incorporates model uncertainty in asset allocation decisions.
The Bayesian weighted predictive distribution of cumulative excess continuously compounded returns averages over the model space, and integrates over the posterior distribution that summarizes the within-model uncertainty about $\Theta_j$ where $j$ is the model identifier. It is given by

$$P(R_{T+K}|\Phi_T) = \sum_{j=1}^{2^M} P(\mathcal{M}_j|\Phi_T) \int_{\Theta_j} P(\Theta_j|\mathcal{M}_j, \Phi_T)P(R_{T+K}|\mathcal{M}_j, \Theta_j, \Phi_T)d\Theta_j \quad (7.5/1)$$

where $P(\mathcal{M}_j|\Phi_T)$ is the posterior probability that model $\mathcal{M}_j$ is the correct one.
• Drawing from the weighted predictive distribution is done in three steps.

• First draw the correct model from the distribution of models.

• Then conditional upon the model implement the two steps, noted above, of drawing future stock returns from the model specific Bayesian predictive distribution.
Stock return predictability and asset pricing models - informative priors

- The classical approach has examined whether return predictability is explained by rational pricing or whether it is due to asset pricing misspecification [see, e.g., Campbell (1987), Ferson and Korajczyk (1995), and Kirby (1998)].

- Studies such as these approach finance theory by focusing on two polar viewpoints: rejecting or not rejecting a pricing model based on hypothesis tests.
The Bayesian approach incorporates pricing restrictions on predictive regression parameters as a reference point for a hypothetical investor’s prior belief.

The investor uses the sample evidence about the extent of predictability to update various degrees of belief in a pricing model and then allocates funds across cash and stocks.
• Pricing models are expected to exert stronger influence on asset allocation when the prior confidence in their validity is stronger and when they explain much of the sample evidence on predictability.

• In particular, Avramov (2004) models excess returns on N investable assets as

\[ r_t = \alpha(z_{t-1}) + \beta f_t + u_{rt}, \]  \hspace{1cm} (7.6.1/1)

\[ \alpha(z_{t-1}) = \alpha_0 + \alpha_1 z_{t-1}, \]  \hspace{1cm} (7.6.1/2)

\[ f_t = \lambda(z_{t-1}) + u_{ft}, \]  \hspace{1cm} (7.6.1/3)
\[ \lambda(z_{t-1}) = \lambda_0 + \lambda_1 z_{t-1}, \quad (7.6.1/4) \]

- where \( f_t \) is a set of \( K \) monthly excess returns on portfolio based factors, \( \alpha_0 \) stands for an \( N \)-vector of the fixed component of asset mispricing, \( \alpha_1 \) is an \( N \times M \) matrix of the time varying component, and \( \beta \) is an \( N \times K \) matrix of factor loadings.

- Now, a conditional version of an asset pricing model (with fixed beta) implies the relation

\[ \mathbb{E}(r_t \mid z_{t-1}) = \beta \lambda(z_{t-1}) \quad (7.6.1/5) \]

- for all \( t \), where \( \mathbb{E} \) stands for the expected value operator.
The model (9.5.1/5) imposes restrictions on parameters and goodness of fit in the multivariate predictive regression

\[ r_t = \mu_0 + \mu_1 z_{t-1} + v_t, \]

(7.6.1/6)

where \( \mu_0 \) is an \( N \)-vector and \( \mu_1 \) is an \( N \times M \) matrix of slope coefficients.

In particular, note that by adding to the right hand side of (9.5.1/6) the quantity \( \beta(f_t - \lambda_0 - \lambda_1 z_{t-1}) \), subtracting the (same) quantity \( \beta u_{ft} \), and decomposing the residual.
in (7.6.1/6) into two orthogonal components \( v_t = \beta u_{ft} + u_{rt} \), we reparameterize the return-generating process (7.6.1/6) as

\[
{r}_t = (\mu_0 - \beta \lambda_0) + (\mu_1 - \beta \lambda_1)z_{t-1} + \beta f_t + u_{rt}. \tag{7.6.1/7}
\]

- Matching the right-hand side coefficients in (7.6.1/7) with those in (7.6.1/1) yields

\[
\mu_0 = \alpha_0 + \beta \lambda_0, \tag{7.6.1/8}
\]
\[
\mu_1 = \alpha_1 + \beta \lambda_1. \tag{7.6.1/9}
\]
That is, under pricing model restrictions where $\alpha_0 = \alpha_1 = 0$ it follows that:

$$\mu_0 = \beta \lambda_0, \quad (7.6.1/10)$$

$$\mu_1 = \beta \lambda_1. \quad (7.6.1/11)$$

This means that stock returns are predictable ($\mu_1 \neq 0$) iff factors are predictable.

Makes sense; after all, under pricing restrictions the systematic component of returns on $N$ stocks is captured by $K$ common factors.
• Of course, if we relax the fixed beta assumption - time varying beta could also be a source of predictability. More later!

• Is return predictability explained by asset pricing models? Probably not!

• Kirby (1998) shows that returns are too predictable to be explained by asset pricing models.

• Ferson and Harvey (1999) show that $\alpha_1 \neq 0$. 
Avramov and Chordia (2006b) show that strategies that invest in individual stocks conditioning on time varying alpha perform extremely well. More later!

So, should we disregard asset pricing restrictions? Not necessarily!

The notion of rejecting or not rejecting pricing restrictions on predictability reflects extreme polar views.

What if you are a Bayesian investor who believes pricing models could be useful albeit not perfect?
As discussed earlier, such an approach has been formalized by Black and Litterman (1992) and Pastor (2000) in the context of iid returns and by Avramov (2004) who accounts for predictability.

The idea is to mix the model and data.

This is shrinkage approach to asset allocation.

Let $\mu_d$ and $\Sigma_d$ ($\mu_m$ and $\Sigma_m$) be the expected return vector and variance covariance matrix based on the data (model).
• Simplistically speaking, moments used for asset allocation are

\[ \mu = \omega \mu_d + (1 - \omega) \mu_m, \]  \hspace{1cm} \text{(7.6.1/12)}

\[ \Sigma = \omega \Sigma_d + (1 - \omega) \Sigma_m, \]  \hspace{1cm} \text{(7.6.1/13)}

• Where \( \omega \) is the shrinkage factor

• The shrinkage of \( \Sigma \) is quite meaningless in this context.

• There are other quite useful shrinkage methods of \( \Sigma \) - see, for example, Jagannathan and Ma (2005).
• In particular,
  – If you completely believe in the model you set $\omega = 0$.
  – If you completely disregard the model you set $\omega = 1$.
  – Going with the shrinkage approach means that $0 < \omega < 1$.

• Avramov (2004) derives asset allocation under the pricing restrictions alone, the data alone, and pricing restrictions and data combined.
• He shows that
  — Optimal portfolios based on the pricing restrictions deliver the lowest Sharpe ratios.
  — Completely disregarding pricing restrictions results in the second lowest Sharpe ratios.
  — Much higher Sharpe ratios are obtained when asset allocation is based on the shrinkage approach.
7. Bayesian Asset Allocation > 7.7 Could you exploit predictability to design outperforming trading strategies?

Could you exploit predictability to design outperforming trading strategies?

- Avramov and Chordia (2006b), Avramov and Wermers (2006), and Avramov, Kosowski, Naik, and Teo (2009) are good references here.

- Let us start with Avramov and Chordia (2006b).
7. Bayesian Asset Allocation > 7.7 Could you exploit predictability to design outperforming trading strategies?

- They study predictability through the out of sample performance of trading strategies that invest in individual stocks conditioning on macro variables.
- They focus on the largest NYSE-AMEX firms by excluding the smallest quartile of firms from the sample.
- They capture 3123 such firms during the July 1972 through November 2003 investment period.
- The investment universe contains 973 stocks, on average, per month.
7. Bayesian Asset Allocation

> 7.7 Could you exploit predictability to design outperforming trading strategies?

> 7.7.1 The evolution of stock returns

The Evolution of Stock Returns

- The underlying statistical models for excess stock returns, the market premium, and macro variables are

\[ r_t = \alpha(z_{t-1}) + \beta(z_{t-1}) \text{mkt}_t + v_t, \]  

(7.7.1/1)

\[ \alpha(z_{t-1}) = \alpha_0 + \alpha_1 z_{t-1}, \]  

(7.7.1/2)

\[ \beta(z_{t-1}) = \beta_0 + \beta_1 z_{t-1}, \]  

(7.7.1/3)
7. Bayesian Asset Allocation

> 7.7 Could you exploit predictability to design outperforming trading strategies?

> 7.7.1 The evolution of stock returns

\[ mkt_t = a + b'z_{t-1} + \eta_t, \quad (7.7.1/4) \]

\[ z_t = c + dz_{t-1} + e_t. \quad (7.7.1/5) \]

- Stock level predictability could come up from:
  1. Model mispricing that varies with changing economic conditions \((\alpha_1 \neq 0)\);
  2. Factor sensitivities are predictable \((\beta_1 \neq 0)\);
  3. The equity premium is predictable \((b \neq 0)\).
7. Bayesian Asset Allocation
   > 7.7 Could you exploit predictability to design outperforming trading strategies?
     > 7.7.1 The evolution of stock returns

- In the end, time varying model alpha is the major source of predictability and investment profitability focusing on individual stocks, portfolios, mutual funds, and hedge funds.

- Time varying alpha is also the major source of predictability in mutual fund and hedge fund investing.

- In the mutual fund and hedge fund context alpha reflects skill. Be careful here! Alpha reflects skill only if the benchmarks used to measure performance are able to price all passive payoffs.
The Proposed Trading Strategy

- Form optimal portfolios from the universe of 3123 AMEX-NYSE stocks over the period 1972 through 2003 with monthly rebalancing on the basis of various models for stock returns.
7. Bayesian Asset Allocation

> 7.7 Could you exploit predictability to design outperforming trading strategies?

> 7.7.2 The proposed trading strategy

• For instance, when predictability in alpha, beta, and the equity premium is permissible, the mean and variance used to form optimal portfolios are

\[
\mu_{t-1} = \hat{\alpha}_0 + \hat{\alpha}_1 z_{t-1} + \hat{\beta}(z_{t-1}) [\hat{\alpha} + \hat{b} z_{t-1}],
\]

\[
\Sigma_{t-1} = \hat{\beta}(z_{t-1}) \hat{\beta}(z_{t-1})' \hat{\sigma}^2_{mkt} + \Psi
\]

\[
+ \delta_1 \hat{\beta}(z_{t-1}) \hat{\beta}(z_{t-1})' \hat{\sigma}^2_{mkt} + \delta_2 \Psi. \quad (7.7.2/2)
\]

Estimation risk
The trading strategy is obtained by maximizing

\[ w_t = \arg \max_{w_t} \left\{ w_t' \mu_t - \frac{1}{2(1/\gamma_t - r_f)} w_t' [\Sigma_t + \mu_t \mu_t'] w_t \right\}, \]

where \( \gamma_t \) is the risk aversion level.

- We do not permit short selling of stocks but we do allow buying on margin.
Performance evaluation

• We implement a recursive scheme:

  – The first optimal portfolio is based on the first 120 months of data on excess returns, market premium, and predictors. (That is, the first estimation window is July 1962 through June 1972.)

  – The second optimal portfolio is based on the first 121 months of data.
7. Bayesian Asset Allocation
   > 7.7 Could you exploit predictability to design outperforming trading strategies?
   > 7.7.3 Performance evaluation

- Altogether, we form 377 optimal portfolios on a monthly basis for each model under consideration.

- We record the realized excess return on any strategy

\[ r_{p,t+1} = \omega_t' r_{t+1}. \quad (7.7.3/1) \]

- We evaluate the ex-post out-of-sample performance of the trading strategies based on the realized returns.
7. Bayesian Asset Allocation
   > 7.7 Could you exploit predictability to design outperforming trading strategies?
     > 7.7.3 Performance evaluation

• Ultimately, we are able to assess the (quite large) economic value of predictability as well as show that our strategies successfully rotate across the size, value, and momentum styles during changing business conditions.
So far, we have shown

- Over the 1972-2003 investment period, portfolio strategies that condition on macro variables outperform the market by about 2% per month.

- Such strategies generate positive performance even when adjusted by the size, value, and momentum factors as well as by the size, book-to-market, and past return characteristics.

- In the period prior to the discovery of the macro variables, investment profitability is primarily attributable to the predictability in the equity premium.
So far, we have shown

- In the post-discovery period, the relation between the macro variables and the equity premium is attenuated considerably.
- Nevertheless, incorporating macro variables is beneficial because such variables drive stock-level alpha and beta variations.
- Predictability based strategies hold small, growth, and momentum stocks and load less (more) heavily on momentum (small) stocks during recessions.
- Such style rotation has turned out to be successful ex post.
Exploiting Predictability in Mutual Fund Returns

• Can we use our methodology to generate positive performance based on the universe of actively managed no-load equity mutual funds?

• What do we know about equity mutual funds?
  - In 2009 about $5 trillion is currently invested in U.S. equity mutual funds, making them a fundamental part of the portfolio of a domestic investor.
  - Previous work has shown that active fund management underperforms, on average, passive benchmarks.
— Strategies that attempt to identify subsets of successful funds using information variables such as past returns or new money inflows (“hot hands” or “smart money” strategies) have underperformed when investment payoffs are adjusted the Fama-French and momentum benchmarks.

• Avramov and Wermers (2006) show that strategies that invest in no-load equity funds conditioning on macro variables (especially time varying alpha) generate substantial positive performance.
<table>
<thead>
<tr>
<th></th>
<th>The Dogmatist</th>
<th></th>
<th>The Skeptic</th>
<th></th>
<th>The Agnostic</th>
<th></th>
<th>H-H</th>
<th>CAR</th>
<th>SM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ND</td>
<td>PD-1</td>
<td>PD-2</td>
<td>NS</td>
<td>PS-1</td>
<td>PS-2</td>
<td>PS-3</td>
<td>PS-4</td>
<td>NA</td>
</tr>
<tr>
<td>$\mu$</td>
<td>6.39</td>
<td>6.58</td>
<td>6.74</td>
<td>6.92</td>
<td>5.56</td>
<td>6.06</td>
<td>15.14</td>
<td>11.5</td>
<td>7.07</td>
</tr>
<tr>
<td>$\text{std}$</td>
<td>0.16</td>
<td>0.16</td>
<td>0.16</td>
<td>0.21</td>
<td>0.21</td>
<td>0.18</td>
<td>0.28</td>
<td>0.21</td>
<td>0.21</td>
</tr>
<tr>
<td>$SR$</td>
<td>0.39</td>
<td>0.41</td>
<td>0.43</td>
<td>0.33</td>
<td>0.27</td>
<td>0.34</td>
<td>0.55</td>
<td>0.54</td>
<td>0.33</td>
</tr>
<tr>
<td>skew</td>
<td>-0.71</td>
<td>-0.68</td>
<td>0.03</td>
<td>-0.2</td>
<td>-0.38</td>
<td>-0.28</td>
<td>1.13</td>
<td>0.32</td>
<td>-0.18</td>
</tr>
<tr>
<td>$\alpha_{cpm}$</td>
<td>-0.23</td>
<td>0</td>
<td>1.92</td>
<td>-0.38</td>
<td>-1.89</td>
<td>0.71</td>
<td>8.12</td>
<td>6.48</td>
<td>-0.23</td>
</tr>
<tr>
<td>$P(\alpha_{cpm})$</td>
<td>0.66</td>
<td>1</td>
<td>0.38</td>
<td>0.88</td>
<td>0.36</td>
<td>0.79</td>
<td>0.07</td>
<td>0.08</td>
<td>0.93</td>
</tr>
<tr>
<td>$\alpha_{ff}$</td>
<td>0.6</td>
<td>1.03</td>
<td>0.91</td>
<td>2.33</td>
<td>0.29</td>
<td>-0.51</td>
<td>11.14</td>
<td>6.86</td>
<td>2.49</td>
</tr>
<tr>
<td>$P(\alpha_{ff})$</td>
<td>0.22</td>
<td>0.04</td>
<td>0.68</td>
<td>0.22</td>
<td>0.87</td>
<td>0.85</td>
<td>0.01</td>
<td>0.05</td>
<td>0.2</td>
</tr>
<tr>
<td>$\alpha_{wml}$</td>
<td>0.37</td>
<td>0.62</td>
<td>3.92</td>
<td>-1.3</td>
<td>-2.49</td>
<td>0.83</td>
<td>5.98</td>
<td>4.95</td>
<td>-1.29</td>
</tr>
<tr>
<td>$P(\alpha_{wml})$</td>
<td>0.46</td>
<td>0.22</td>
<td>0.07</td>
<td>0.45</td>
<td>0.13</td>
<td>0.76</td>
<td>0.14</td>
<td>0.17</td>
<td>0.47</td>
</tr>
<tr>
<td>$\bar{\alpha}_{cpm}$</td>
<td>-0.09</td>
<td>0.14</td>
<td>1.85</td>
<td>0.14</td>
<td>-1.75</td>
<td>0.66</td>
<td>9.01</td>
<td>6.36</td>
<td>0.25</td>
</tr>
<tr>
<td>$P(\bar{\alpha}_{cpm})$</td>
<td>0.87</td>
<td>0.8</td>
<td>0.4</td>
<td>0.96</td>
<td>0.4</td>
<td>0.8</td>
<td>0.05</td>
<td>0.08</td>
<td>0.92</td>
</tr>
<tr>
<td>$\bar{\alpha}_{ff}$</td>
<td>0.2</td>
<td>0.84</td>
<td>0.91</td>
<td>2.96</td>
<td>1.02</td>
<td>0.29</td>
<td>13.3</td>
<td>9.28</td>
<td>3.1</td>
</tr>
<tr>
<td>$P(\bar{\alpha}_{ff})$</td>
<td>0.64</td>
<td>0.05</td>
<td>0.62</td>
<td>0.12</td>
<td>0.55</td>
<td>0.9</td>
<td>0</td>
<td>0.01</td>
<td>0.11</td>
</tr>
<tr>
<td>$\bar{\alpha}_{wml}$</td>
<td>0.24</td>
<td>0.69</td>
<td>1.9</td>
<td>-0.84</td>
<td>-1.99</td>
<td>-0.37</td>
<td>9.44</td>
<td>7.88</td>
<td>-0.78</td>
</tr>
<tr>
<td>$P(\bar{\alpha}_{wml})$</td>
<td>0.59</td>
<td>0.12</td>
<td>0.3</td>
<td>0.61</td>
<td>0.23</td>
<td>0.88</td>
<td>0.02</td>
<td>0.03</td>
<td>0.65</td>
</tr>
</tbody>
</table>
Downside Risk
8. Downside Risk Measures

Downside Risk

- Downside risk is the financial risk associated with losses.
- Downside risk measures quantify the risk of losses, whereas volatility measures are both about the upside and downside outcomes.
- That is, volatility treats symmetrically up and down moves (relative to the mean).
- Or volatility is about the entire distribution while down side risk concentrates on the left tail.
8. Downside Risk Measures

Downside Risk

- Example of downside risk measures
  - Value at Risk (VaR)
  - Expected Shortfall
  - Semi-variance
  - Maximum drawdown
  - Downside Beta
  - Shortfall probability

- We will discuss below all these measures.
8. Downside Risk Measures

Value at Risk (VaR)

• The $VaR_{95\%}$ says that there is a 5% chance that the realized return, denoted by $R$, will be less than $-VaR_{95\%}$.

• More generally,

$$\Pr(R \leq -VaR_{1-\alpha}) = \alpha$$
8. Downside Risk Measures

Value at Risk (VaR)

\[- \frac{VaR_{1-\alpha} - \mu}{\sigma} = \Phi^{-1}(\alpha) \Rightarrow VaR_{1-\alpha} = -\left(\mu + \sigma \Phi^{-1}(\alpha)\right)\]

where

\[\Phi^{-1}(\alpha),\] the critical value, is the inverse cumulative distribution function of the standard normal evaluated at \(\alpha\).

• Let \(\alpha=5\%\) and assume that

\[R \sim N(\mu, \sigma^2)\]

• The critical value is \(\Phi^{-1}(0.05) = -1.64\)
8. Downside Risk Measures

Value at Risk (VaR)

Therefore

\[ \text{VaR}_{1-\alpha} = - \left( \mu + \Phi^{-1}(\alpha) \sigma \right) = -\mu + 1.64\sigma \]

Check:

If \( R \sim N(\mu, \sigma^2) \)

Then \( \text{Pr}(R \leq -\text{VaR}_{1-\alpha}) = \text{Pr} \left( \frac{R - \mu}{\sigma} \leq \frac{-\text{VaR}_{\alpha} - \mu}{\sigma} \right) \)

\[ = \text{Pr} \left( \frac{R - \mu}{\sigma} < \frac{\mu - 1.64\sigma - \mu}{\sigma} \right) \]

\[ = \text{Pr}(z < -1.64) = \Phi(-1.64) = 0.05 \]
8. Downside Risk Measures

Example: The US Equity Premium

- Suppose: \( R \sim N(0.08, 0.20^2) \)
  \[ \Rightarrow VaR_{0.95} = -(0.08 - 1.64 \cdot 0.20) = 0.25 \]

- That is to say that we are 95% sure that the future equity premium won’t decline more than 25%.
- If we would like to be 97.5% sure – the price is that the threshold loss is higher.
- To illustrate,

\[ VaR_{0.975} = -(0.08 - 1.96 \cdot 0.20) = 0.31 \]
Evidently, the VaR of a portfolio is not necessarily lower than the combination of individual VaR-s – which is apparently at odds with the notion of diversification.

However, recall that VaR is a downside risk measure while volatility which diminishes at the portfolio level is a symmetric measure.
Backtesting The VaR

- The VaR requires the specification of the exact distribution and its parameters (e.g., mean and variance).
- Typically the normal distribution is chosen.
- Mean could be the sample average.
- Volatility estimates could follow ARCH, GARCH, EGARCH, stochastic volatility, and realized volatility, all of which are described later in this course.
- We can examine the validity of VaR using backtesting.
Backtesting The VaR

• Assume that stock returns are normally distributed with mean and variance that vary over time

\[ r_t \sim N(\mu_t, \sigma_t^2) \quad \forall t = 1,2,\ldots,T \]

• The sample is of length \( T \).

• Receipt for backtesting is as follows.

• Use the first, say, sixty monthly observations to estimate the mean and volatility and compute the VaR.

• If the return in month 61 is below the VaR set an indicator function \( I \) to be equal to one; otherwise, it is zero.
8. Downside Risk Measures

Backtesting The VaR

• Repeat this process using either a rolling or recursive schemes and compute the fraction of time when the next period return is below the VaR.

• If $\alpha=5\%$ - only 5% of the returns should be below the computed VaR.

• Suppose we get 5.5% of the time – is it a bad model or just a bad luck?
Model Verification Based on Failure Rates

- To answer that question – let us discuss another example which requires a similar test statistic.
- Suppose that $Y$ analysts are making predictions about the market direction for the upcoming year. The analysts forecast whether market is going to be up or down.
- After the year passes you count the number of wrong analysts. An analyst is wrong if he/she predicts up move when the market is down, or predict down move when the market is up.
Model Verification Based on Failure Rates

- Suppose that the number of wrong analysts is $x$.
- Thus, the fraction of wrong analysts is $P = x/Y$ – this is the failure rate.
The hypothesis to be tested is

\[ H_0 : P = P_0 \]

\[ H_1 : \text{Otherwise} \]

Under the null hypothesis it follows that

\[ f(x) = \binom{y}{x} P_0^x (1 - P_0)^{y-x} \]
The Test Statistic

• Notice that

\[ E(x) = P_0 y \]
\[ \text{VaR}(x) = P_0 (1 - P_0) y \]

Thus

\[ Z = \frac{x - P_0 y}{\sqrt{P_0 (1 - P_0) y}} \sim N(0,1) \]
8. Downside Risk Measures

Back to Backtesting VaR: A Real Life Example

• In its 1998 annual report, JP Morgan explains: In 1998, daily revenue fell short of the downside (95%VaR) band on 20 trading days (out of 252) or more than 5% of the time (252×5%=12.6).

• Is the difference just a bad luck or something more systematic? We can test the hypothesis that it is a bad luck.

\[ H_0 : x = 12.6 \]
\[ H_1 : \text{Otherwise} \]

\[ Z = \frac{20 - 12.6}{\sqrt{0.05 \cdot 0.95 \cdot 252}} = 2.14 \]
Back to Backtesting VaR: A Real Life Example

- Notice that you reject the null since 2.14 is higher than the critical value of 1.96.
- That suggests that JPM should search for a better model.
- They did find out that the problem was that the actual revenue departed from the normal distribution.
Expected Shortfall (ES): Truncated Distribution

- ES is the expected value of the loss conditional upon the event that the actual return is below the VaR.

- The ES is formulated as

\[ ES_{1-\alpha} = -E[R \mid R \leq -VaR_{1-\alpha}] \]
8. Downside Risk Measures

Expected Shortfall and the Truncated Normal Distribution

- Assume that returns are normally distributed:

\[ R \sim N(\mu, \sigma^2) \]

\[ \Rightarrow ES_{1-\alpha} = -E[R | R \leq \mu + \Phi^{-1}(\alpha)\sigma] \]

\[ \Rightarrow ES_{1-\alpha} = -\mu + \sigma \frac{\phi(\Phi^{-1}(\alpha))}{\alpha} \]
8. Downside Risk Measures

Expected Shortfall and the Truncated Normal Distribution

where $\phi(.)$ is the pdf of the standard normal density

e.g

$\phi(-1.64) = 0.103961$

This formula for ES is about the expected value of a truncated normally distributed random variable.
Expected Shortfall and the Truncated Normal Distribution

Proof:

\[ x \sim N(\mu, \sigma^2) \]

\[ E(x \mid x \leq -VaR_{1-\alpha}) = \mu - \frac{\sigma \phi(k_0)}{\Phi(k_0)} = \mu - \sigma \frac{\phi(\Phi^{-1}(\alpha))}{\alpha} \]

since

\[ k_0 = \frac{-VaR_{1-\alpha} - \mu}{\sigma} = \Phi^{-1}(\alpha) \]
Expected Shortfall: Example

Example

\( \mu = 8\% \)
\( \sigma = 20\% \)

\[
ES_{95\%} = -0.08 + 0.20 \frac{\phi(-1.64)}{0.05} \approx -0.08 + 0.20 \frac{0.10}{0.05} = 0.32
\]

\[
ES_{97.5\%} = -0.08 + 0.20 \frac{\phi(-1.96)}{0.025} \approx -0.08 + 0.20 \frac{0.06}{0.25} = 0.40
\]
8. Downside Risk Measures

Expected Shortfall and the Truncated Normal Distribution

- Previously, we got that the VaRs corresponding to those parameters are 25% and 31%.
- Now the expected losses are higher, 32% and 40%.
- Why?
- The first lower figures (VaR) are unconditional in nature relying on the entire distribution.
- In contrast, the higher ES figures are conditional on the existence of shortfall – realized return is below the VaR.
8. Downside Risk Measures

**Expected Shortfall in Decision Making**

- The mean variance paradigm minimizes portfolio volatility subject to an expected return target.
- Suppose you attempt to minimize ES instead subject to expected return target.
Expected Shortfall with Normal Returns

- If stock returns are normally distributed then the ES chosen portfolio would be identical to that based on the mean variance paradigm.

- No need to go through optimization to prove that assertion.

- Just look at the expression for ES under normality to quickly realize that you need to minimize the volatility of the portfolio subject to an expected return target.
Target Semi Variance

• Variance treats equally downside risk and upside potential.

• The semi-variance, just like the VaR, looks at the downside.

• The target semi-variance is defined as:

\[ \lambda(h) = E\left[\min(r - h, 0)^2\right] \]

where \( h \) is some target level.

• For instance, \( h = R_f \)

• Unlike the variance, \( \sigma^2 = E(r - \mu)^2 \)
8. Downside Risk Measures

Target Semi Variance

The target semi-variance:

• Picks a target level as a reference point instead of the mean.

• Gives weight only to negative deviations from a reference point.
8. Downside Risk Measures

**Target Semi Variance**

- Notice that if \( r \sim N(\mu, \sigma^2) \)

\[
\lambda(h) = \sigma^2 \frac{h - \mu}{\sigma} \phi\left(\frac{h - \mu}{\sigma}\right) + \sigma^2 \left[ \left(\frac{h - \mu}{\sigma}\right)^2 + 1 \right] \Phi\left(\frac{h - \mu}{\sigma}\right)
\]

where

\( \phi \) and \( \Phi \) are the PDF and CDF of a \( N(0,1) \) variable, respectively

- Of course if \( h = \mu \)

then \( \lambda(h) = \frac{\sigma^2}{2} \)
8. Downside Risk Measures

Maximum Drawdown (MD)

- The MD (M) over a given investment horizon is the largest M-month loss of all possible M-month continuous periods over the entire horizon.
- Useful for an investor who does not know the entry/exit point and is concerned about the worst outcome.
- It helps determine the investment risk.
Downside Beta

- I will introduce three distinct measures of downside beta – each of which is valid and captures the down side of investment payoffs.

- Displayed are the population betas.

- Taking the formulations into the sample – simply replace the expected value by the sample mean.
Downside Risk Measures

Downside Beta

\[
\beta_{im}^{(1)} = \frac{E[(R_i - R_f) \min((R_m - R_f), 0)]}{E[\min((R_m - R_f), 0)]^2}
\]

- The numerator in the equation is referred to as the co-semi-variance of returns and is the covariance of returns below \( R_f \) on the market portfolio with return in excess of \( R_f \) on security \( i \).
- It is argued that risk is often perceived as downside deviations below a target level by market participants and the risk-free rate is a replacement for average equity market returns.
8. Downside Risk Measures

**Downside Beta**

\[
\beta_{im}^{(2)} = \frac{E[(R_i - \mu_i)\min(R_m - \mu_m, 0)]}{E[\min(R_m - \mu_m, 0)^2]}
\]

- where \(\mu_i\) and \(\mu_m\) are security i and market average return respectively.

- One can modify the down side beta as follows:

\[
\beta_{im}^{(3)} = \frac{E[\min(R_i - \mu_i, 0)\min(R_m - \mu_m, 0)]}{E[\min(R_m - \mu_m, 0)^2]}
\]
Estimating and Evaluating Asset Pricing Models
Why Caring About Asset Pricing Models

• An essential question that arises is why would both academics and practitioners invest huge resources attempting to develop and test asset pricing models.

• It turns out that pricing models have crucial roles in various applications in financial economics – both asset pricing as well as corporate finance.

• In the following, I list five major applications.
1 – Common Risk Factors

• Pricing models characterize the risk profile of a firm.

• In particular, systematic risk is no longer stock return volatility – rather it is the loadings on risk factors.

• For instance, in the single factor CAPM the market beta – or the co-variation with the market – characterizes the systematic risk of the firm.
1 – Common Risk Factors

• Likewise, in the single factor (C)CAPM the consumption growth beta – or the co-variation with consumption growth – characterizes the systematic risk of the firm.

• In the multi-factor Fama-French (FF) model there are three sources of risk – the market beta, the SMB beta, and the HML beta.
1 – Common Risk Factors

- Under FF, other things being equal (ceteris paribus), a firm is riskier if its loading on SMB beta is higher.
- Under FF, other things being equal (ceteris paribus), a firm is riskier if its loading on HML beta is higher.
2 – Moments for Asset Allocation

- Pricing models deliver moments for asset allocation.
- For instance, the tangency portfolio takes on the form

\[ w_{TP} = \frac{V^{-1}\mu^e}{\mu^e - \mu} \]  

(9.1/1)
2 – Moments for Asset Allocation

• Under the CAPM, the vector of expected returns and the covariance matrix are given by:

\[ \mu^e = \beta \mu_m \] (9.1/2)

\[ V = \beta \beta' \sigma_m^2 + \Sigma \] (9.1/3)

• where \( \Sigma \) is the covariance matrix of the residuals in the time-series asset pricing regression.

• We denoted by \( \Psi \) the residual covariance matrix in the case wherein the off diagonal elements are zeroed out.
The corresponding quantities under the FF model are

\[ \mu^e = \beta_{MKT}\mu_m^e + \beta_{SML}\mu_{SML} + \beta_{HML}\mu_{HML} \]

\[ V = \beta \sum_F \beta' + \Sigma \]

where \( \Sigma_F \) is the covariance matrix of the factors.
3 – Discount Factor

• Expected return is the discount factor, commonly denoted by \( k \), in present value formulas in general and firm evaluation in particular:

\[
PV = \sum_{t=1}^{T} \frac{CF_t}{(1+k)^t}
\]

(9.1/5)

• In practical applications, expected returns are typically assumed to be constant over time, an unrealistic assumption.
3 – Discount Factor

• Indeed, thus far we have examined models with constant beta and constant risk premiums

$$\mu^e = \beta' \lambda \ (9.1/6)$$

where $\lambda$ is a K-vector of risk premiums.

• When factors are return spreads the risk premium is the mean of the factor.

• Later we will consider models with time varying factor loadings.
4 - Benchmarks

• Factors in asset pricing models serve as benchmarks for evaluating performance of active investments.

• In particular, performance is the intercept (alpha) in the time series regression of excess fund returns on a set of benchmarks (typically four benchmarks in mutual funds and more so in hedge funds):

\[ r_t^e = \alpha + \beta_{MKT} \times r_{MKT,t}^e + \beta_{SMB} \times SMB_t + \beta_{HML} \times HML_t + \beta_{WML} \times WML_t + \epsilon_t \]
There is a plethora of studies in corporate finance that use asset pricing models to risk adjust asset returns.

Here are several examples:

- Examining the long run performance of IPO firm.
- Examining the long run performance of SEO firms.
- Analyzing abnormal performance of stocks going through splits and reverse splits.
5 - Corporate Finance

- Analyzing mergers and acquisitions
- Analyzing the impact of change in board of directors.
- Studying the impact of corporate governance on the cross section of average returns.
- Studying the long run impact of stock/bond repurchase.

Time Series Tests

- Time series tests are designated to examine the validity of models in which factors are portfolio based, or factors that are return spreads.

- Example: the market factor is the return difference between the market portfolio and the risk-free asset.

- Consumption growth is not a return spread.

- Thus, the consumption CAPM cannot be tested using time series regressions, unless you form a factor mimicking portfolio (FMP) for consumption growth.
• FMP is a convex combination of asset returns having the maximal correlation with consumption growth.

• The statistical time series tests have an appealing economic interpretation.

• Testing the CAPM amounts to testing whether the market portfolio is the tangency portfolio.

• Testing multi-factor models amounts to testing whether some optimal combination of the factors is the tangency portfolio.
Testing the CAPM

• Run the time series regression:

\[ r_{1t}^e = \alpha_1 + \beta_1 r_{mt}^e + \varepsilon_{1t} \]

\[ \vdots \]

\[ r_{Nt}^e = \alpha_N + \beta_N r_{mt}^e + \varepsilon_{Nt} \]

• The null hypothesis is:

\[ H_0: \alpha_1 = \alpha_2 = \cdots = \alpha_N = 0 \]
In the following, four times series test statistics will be described:

- WALD;
- Likelihood Ratio;
- GRS (Gibbons, Ross, and Shanken (1989));
- GMM.
The Distribution of $\alpha$

- Recall, $a$ is asset mispricing.

- The time series regressions can be rewritten using a vector form as:

$$r_t^e = \alpha + \beta \cdot r_{mt}^e + \varepsilon_t$$

- Let us assume that

$$\varepsilon_t \sim iid N \left(0, \sum_{NN} \right)$$

for $t = 1, 2, 3, ..., T$

- Let $\Theta = (\alpha', \beta', vech(\varepsilon)')'$ be the set of all parameters.
Under normality, the likelihood function for $\varepsilon_t$ is

$$L(\varepsilon_t | \theta) = c \sum^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (r_t^e - \alpha - \beta r_{mt}^e)' \Sigma^{-1}(r_t^e - \alpha - \beta r_{mt}^e) \right]$$

where $c$ is the constant of integration (recall the integral of a probability distribution function is unity).
Moreover, the IID assumption suggests that

\[ L(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N | \theta) = c^T \Sigma^{-\frac{T}{2}} \]

\[ \times \exp \left[ -\frac{1}{2} \sum_{t=1}^{T} (r_t^e - \alpha - \beta r_{mt}^e)' \sum^{-1} (r_t^e - \alpha - \beta r_{mt}^e) \right] \]

Taking the natural log from both sides yields

\[ \ln (L) \propto -\frac{T}{2} \ln \left( \sum \right) - \frac{1}{2} \sum_{t=1}^{T} (r_t^e - \alpha - \beta r_{mt}^e)' \sum^{-1} (r_t^e - \alpha - \beta r_{mt}^e) \]
Asymptotically, we have $\theta - \hat{\theta} \sim N(0, \Sigma(\theta))$

where

$$\Sigma(\theta) = \left[-E\left[\frac{\ln 2 \alpha}{\partial \theta \partial \theta'}\right]\right]^{-1}$$
Let us estimate the parameters

\[
\frac{\partial \ln (L)}{\partial \alpha} = \sum \left[ \sum_{t=1}^{T} (r_t^e - \alpha - \beta r_{mt}) \right]^{-1}
\]

\[
\frac{\partial \ln (L)}{\partial \beta} = \sum \left[ \sum_{t=1}^{T} (r_t^e - \alpha - \beta r_{mt}) \times r_{mt} \right]^{-1}
\]

\[
\frac{\partial \ln (L)}{\partial \Sigma} = -\frac{T}{2} \sum \left[ \sum_{t=1}^{T} \varepsilon_t \varepsilon_t' \right]^{-1} + \frac{1}{2} \sum \left[ \sum_{t=1}^{T} \varepsilon_t \varepsilon_t' \right]^{-1} \sum
\]
Solving for the first order conditions yields

\[ \hat{\alpha} = \hat{\mu}^e - \hat{\beta} \cdot \hat{\mu}_m^e \]

\[ \hat{\beta} = \frac{\sum_{t=1}^{T} (r_t^e - \hat{\mu}^e)(r_{mt}^e - \hat{\mu}_m^e)}{\sum_{t=1}^{T} (r_{mt}^e - \hat{\mu}_m^e)^2} \]
Moreover,

\[
\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}'_t
\]

\[
\hat{\mu}^e = \frac{1}{T} \sum_{t=1}^{T} r^e_t
\]

\[
\mu^e_m = \frac{1}{T} \sum_{t=1}^{T} r^e_{mt}
\]
• Recall our objective is to find the variance-covariance matrix of $\hat{\alpha}$.

• Standard errors could be found using the information matrix.

• The information matrix is constructed as follows

\[ I(\theta) = -E \begin{bmatrix}
\frac{\partial^2 \ln (L)}{\partial \alpha \partial \alpha'} & \frac{\partial^2 \ln (L)}{\partial \alpha \partial \beta'} & \frac{\partial^2 \ln (L)}{\partial \alpha \partial \Sigma'} \\
\frac{\partial^2 \ln (L)}{\partial \beta \partial \alpha'} & \frac{\partial^2 \ln (L)}{\partial \beta \partial \beta'} & \frac{\partial^2 \ln (L)}{\partial \beta \partial \Sigma'} \\
\frac{\partial^2 \ln (L)}{\partial \Sigma \partial \alpha'} & \frac{\partial^2 \ln (L)}{\partial \Sigma \partial \beta'} & \frac{\partial^2 \ln (L)}{\partial \Sigma \partial \Sigma'}
\end{bmatrix} \]
The Distribution of the Parameters

• Try to establish yourself the information matrix.

• Notice that \( \hat{\alpha} \) and \( \hat{\beta} \) are independent of \( \hat{\Sigma} \) - thus, you can ignore the second derivatives with respect to \( \Sigma \) in the information matrix if your objective is to find the distribution of \( \hat{\alpha} \) and \( \hat{\beta} \).

• If you aim to derive the distribution of \( \hat{\Sigma} \) then focus on the bottom right block of the information matrix.
The Distribution of $\alpha$

- We get:

$$\hat{\alpha} \sim N\left(\alpha, \frac{1}{T} \left[1 + \left(\frac{\hat{\mu}_m}{\hat{\sigma}_m}\right)^2\right] \Sigma\right)$$

- Moreover,

$$\hat{\beta} \sim N\left(\beta, \frac{1}{T} \cdot \frac{1}{\hat{\sigma}_m^2} \Sigma\right)$$

$$T\hat{\Sigma} \sim (T - 2, \Sigma)$$
• Notice that $W(x, y)$ stands for the Wishart distribution with $x = T - 2$ degrees of freedom and a parameter matrix $y = \Sigma$. 
The Wald Test

• Recall, if

\[ X \sim N(\mu, \Sigma) \text{ then } (X - \mu)\Sigma^{-1}(X - \mu) \sim \chi^2(N) \]

• Here we test

\[ H_0: \hat{\alpha} = 0 \]
\[ H_1: \hat{\alpha} \neq 0 \]

where

\[ \hat{\alpha} \overset{H_0}{\sim} N(0, \Sigma_\alpha) \]

• The Wald statistic is \[ \hat{\alpha}'\Sigma_\alpha^{-1}\hat{\alpha} \sim \chi^2(N) \]
which becomes:

\[ J_1 = T \left[ 1 + \left( \frac{\hat{\mu}_m}{\hat{\sigma}_m} \right)^2 \right]^{-1} \hat{\alpha}'\hat{\Sigma}\hat{\alpha} = T \frac{\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}}{1 + S\hat{R}_m^2} \]

where \( S\hat{R}_m \) is the Sharpe ratio of the market factor.
Algorithm for Implementation

- The algorithm for implementing the statistic is as follows:

- Run separate regressions for the test assets on the common factor:

\[
\begin{align*}
r_t^e &= X_{Tx2} \theta_1 + \varepsilon_1 \\
\vdots \\
r_N^e &= X_{Tx2} \theta_N + \varepsilon_N
\end{align*}
\]

where

\[
X_{Tx2} = \begin{bmatrix} 1, r_{m1}^e \\ \vdots \\ 1, r_{mT}^e \end{bmatrix}
\]

\[
\theta_i = [\alpha_i, \beta_i]'
\]
• Retain the estimated regression intercepts

\[ \hat{\alpha} = [\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_N]' \quad \text{and} \quad \hat{\varepsilon} = [\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_N]_{T \times N} \]

• Compute the residual covariance matrix

\[ \hat{\Sigma} = \frac{1}{T} \hat{\varepsilon}' \hat{\varepsilon} \]

• Compute the sample mean and the sample variance of the factor.

• Compute \( J_1 \).
The Likelihood Ratio Test

• We run the unrestricted and restricted specifications:

un: \( r_t^e = \alpha + \beta r_{mt}^e + \varepsilon_t \quad \varepsilon_t \sim N(0, \Sigma) \)

res: \( r_t^e = \beta^* r_{mt}^e + \varepsilon_t^* \quad \varepsilon_t^* \sim N(0, \Sigma^*) \)
Using MLE, we get:

\[
\hat{\beta}^* = \frac{\sum_{t=1}^{T} r_t^e r_{mt}^e}{\sum_{t=1}^{T} (r_{mt}^e)^2}
\]

\[
\hat{\Sigma}^* = \frac{1}{T} \sum_{t=1}^{T} \hat{\epsilon}_t^* \hat{\epsilon}_t^{*\prime}
\]

\[
\hat{\beta}^* \sim N \left( \beta, \frac{1}{T} \left[ \frac{1}{\hat{\mu}_m^2 + \hat{\sigma}_m^2} \Sigma \right] \right)
\]

\[
T \hat{\Sigma}^* \sim W (T - 1, \Sigma)
\]
The LR Test

\[ LR = \ln (L^*) - \ln (L) = - \frac{T}{2} [\ln |\Sigma^*| - \ln |\Sigma|] \]

\[ J_2 = -2LR = T [\ln |\Sigma^*| - \ln |\Sigma|] \sim \chi^2(N) \]

• Using some algebra, one can show that

\[ J_1 = T \left( \exp \left( \frac{J_2}{T} \right) - 1 \right) \]

• Thus,

\[ J_2 = T \cdot \ln \left( \frac{J_1}{T} + 1 \right) \]
GRS (1989)

Theorem: let

\[
X_{nx1} \sim N(0, \Sigma)
\]

let

\[
A_{nx1} \sim W(\tau, \Sigma)
\]

where \( \tau \geq N \)

and let A and X be independent than

\[
\frac{\tau - N + 1}{N} X' A^{-1} X \sim F_{N, \tau - N + 1}
\]
• In our context:

\[ X = \sqrt{T} \left[ 1 + \left( \frac{\hat{\mu}_m}{\hat{\sigma}_m} \right)^2 \right]^{-\frac{1}{2}} \text{\(H^0 \sim N(0, \Sigma)\)} \\
A = T\hat{\Sigma} \sim W(\tau, \Sigma) \\
\text{where} \\
\tau = T - 2 \\

• Then:

\[ J_3 = \left( \frac{T - N - 1}{N} \right) \left[ 1 + \left( \frac{\hat{\mu}_m}{\hat{\mu}\sigma_m} \right)^2 \right]^{-1} \hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha} \sim F(N, T - N - 1) \]

• This is a finite-sample test.
I will directly give the statistic (derivation comes up later in the notes)

\[ J_4 = T \hat{\alpha}' (R (D_T' S_T^{-1} D_T)^{-1} R')^{-1} \cdot \hat{\alpha} \sim \chi^2 (N) \]

where

\[
 R_{N \times 2N} = \begin{bmatrix} I_N, & 0 \\ N \times N, & N \times N \end{bmatrix}
\]

\[
 D_T = - \begin{bmatrix} 1, & \hat{\mu}_m^e \\ \hat{\mu}_m, & (\hat{\mu}_m^e)^2 + \hat{\sigma}_m^2 \end{bmatrix} \otimes I_N
\]

- Assume no serial correlation but heteroskedasticity:

\[ S_T = \frac{1}{T} \sum_{t=1}^{T} (x_t x_t' \otimes \hat{\epsilon}_t \hat{\epsilon}_t') \]

where

\[ x_t = [1, r_{mt}^e]' \]

- Under homoskedasticity and serially uncorrelated moment conditions: \( J_4 = J_1 \).
- That is, the GMM statistic boils down to the WALD.
The Multi-Factor Version of Asset Pricing Tests

\[ r_t^e = \alpha_{N \times 1} + \beta_{N \times K} \cdot F_t + \epsilon_t_{N \times 1} \]

\[ J_1 = T \left( 1 + \hat{\mu}'_F \hat{\Sigma}_F^{-1} \hat{\mu}_F \right)^{-1} \hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha} \sim \chi(N) \]

\[ J_2 \text{ follows as described earlier.} \]

\[ J_3 = \frac{T - N - K}{N} \left( 1 + \hat{\mu}'_F \hat{\Sigma}_F^{-1} \hat{\mu}_F \right)^{-1} \hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha} \sim F_{(N,T-N-K)} \]

where \( \hat{\mu}_F \) is the mean vector of the factor based return spreads.
• \( \hat{\Sigma}_F \) is the variance covariance matrix of the factors.

• For instance, considering the Fama-French model:

\[
\hat{\mu}_F = \begin{bmatrix}
\hat{\mu}_m \\
\hat{\mu}_{SMB} \\
\hat{\mu}_{HML}
\end{bmatrix} \quad \hat{\Sigma}_F = \begin{bmatrix}
\hat{\sigma}_m^2, \hat{\sigma}_m, \hat{\sigma}_{SMB}, \hat{\sigma}_m, \hat{\sigma}_{HML} \\
\hat{\sigma}_{SMB,m}, \hat{\sigma}_{SMB}^2, \hat{\sigma}_{SMB,HML} \\
\hat{\sigma}_{HML,m}, \hat{\sigma}_{HML}, \hat{\sigma}_{HML,SMB}, \hat{\sigma}_{HML}^2
\end{bmatrix}
\]
The Economics of Time Series Test Statistics

Let us summarize the first three test statistics:

\[ J_1 = T \frac{\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}}{1 + S\hat{R}_m^2} \]

\[ J_2 = T \cdot \ln \left( \frac{J_1}{T} + 1 \right) \]

\[ J_3 = \frac{T - N - 1}{N} \frac{\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}}{1 + S\hat{R}_m^2} \]

The Economics of Time Series Test Statistics

• The $J_4$ statistic, the GMM based asset pricing test, is actually a Wald test, just like $J_1$, except that the covariance matrix of asset mispricing takes account of heteroskedasticity and often even potential serial correlation.

• Notice that all test statistics depend on the quantity

$$\hat{\alpha} \hat{\Sigma}^{-1} \hat{\alpha}$$
The Economics of Time Series Test Statistics

• GRS show that this quantity has a very insightful representation.

• Let us provide the steps.
Understanding the Quantity $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$

- Consider an investment universe that consists of $N + 1$ assets - the $N$ test assets as well as the market portfolio.

- The expected return vector of the $N + 1$ assets is given by

$$\hat{\lambda}_{(N+1)\times 1} = \begin{bmatrix} \hat{\mu}_m & \hat{\mu}_e' \end{bmatrix}'$$

where $\hat{\mu}_m$ is the estimated expected excess return on the market portfolio and $\hat{\mu}_e$ is the estimated expected excess return on the $N$ test assets.

Understanding the Quantity $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$

- The variance covariance matrix of the N+1 assets is given by

$$\hat{\Phi}_{(N+1)\times(N+1)} = \begin{bmatrix} \hat{\sigma}_m^2, & \hat{\beta}' \hat{\sigma}_m^2 \\ \hat{\beta} \hat{\sigma}_m^2, & \hat{V} \end{bmatrix}$$

where

$\hat{\sigma}_m^2$ is the estimated variance of the market factor.

$\hat{\beta}$ is the $N$-vector of market loadings and $\hat{V}$ is the covariance matrix of the $N$ test assets.

Understanding the Quantity $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$

- Notice that the covariance matrix of the $N$ test assets is $
\hat{\Sigma} = \hat{\beta}\hat{\beta}'\hat{\sigma}_m^2 + \hat{\Sigma}$

- The squared tangency portfolio of the $N + 1$ assets is
$S\hat{R}_{TP}^2 = \hat{\lambda}'\hat{\Phi}^{-1}\hat{\lambda}$
Understanding the Quantity $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$

• Notice also that the inverse of the covariance matrix is

$$\hat{\Phi}^{-1} = \begin{bmatrix}
(\hat{\sigma}_m^2)^{-1} + \hat{\beta}'\hat{\Sigma}^{-1}\hat{\beta}, & -\hat{\beta}'\hat{\Sigma}^{-1} \\
-\hat{\Sigma}^{-1}\hat{\beta}, & \hat{\Sigma}^{-1}
\end{bmatrix}$$

• Thus, the squared Sharpe ratio of the tangency portfolio could be represented as
Understanding the Quantity $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$

$$S\hat{R}_{TP}^2 = \left(\frac{\hat{\mu}_m^e}{\hat{\sigma}_m}\right)^2 + \left[(\hat{\mu}^e - \hat{\beta}\hat{\mu}_m^e)'\hat{\Sigma}^{-1}(\hat{\mu}^e - \hat{\beta}\hat{\mu}_m^e)\right]$$

$$S\hat{R}_{TP}^2 = S\hat{R}_m^2 + \hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$$

or

$$\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha} = S\hat{R}_{TP}^2 - S\hat{R}_m^2$$
Understanding the Quantity $\hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha}$

- In words, the $\hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha}$ quantity is the difference between the squared Sharpe ratio based on the $N + 1$ assets and the squared Sharpe ratio of the market portfolio.

- If the CAPM is correct then these two Sharpe ratios are identical in population, but not identical in sample due to estimation errors.
Understanding the Quantity $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$

• The test statistic examines how close the two sample Sharpe ratios are.

• Under the CAPM, the extra N test assets do not add anything to improving the risk return tradeoff.

• The geometric description of $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$ is given in the next slide.

Understanding the Quantity $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$

\[
\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha} = \Phi_1^2 - \Phi_2^2
\]

**Understanding the Quantity $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$**

- So we can rewrite the previously derived test statistics as

$$J_1 = T \frac{S\hat{R}_{TP}^2 - S\hat{R}_m^2}{1 + S\hat{R}_m^2} \sim \chi^2(N)$$

$$J_3 = \frac{T - N - 1}{N} \times \frac{S\hat{R}_{TP}^2 - S\hat{R}_m^2}{1 + S\hat{R}_m^2} \sim F(N, T - N - 1)$$
Cross Sectional Regressions

- The time series procedures are designed primarily to test asset pricing models based on factors that are asset returns.

- The cross-sectional technique can be implemented whether or not the factor is a return spread.

- Consumption growth is a good example of a non portfolio based factor.

- The central question in the cross section framework is why average returns vary across assets.
• So plot the sample average excess returns on the estimated betas.

• But even if the model is correct, this plot will not work out perfectly well because of sampling errors.

• The idea is to run a cross-sectional regression to fit a line through the scatterplot of average returns on estimated betas.

• Then examine the deviations from a linear relation.

• In the cross section approach you can also examine whether a factor is indeed priced.
Let us formalize the concepts.

Two regression steps are at the heart of the cross-sectional approach:

First, estimate betas from the time-series regression of excess returns on some pre-specified factors

\[ r_{i,t} = \alpha_i + \beta_i f_t + \epsilon_{i,t}. \quad (9.3/1) \]

Then run the cross-section regression of average returns on the betas

\[ \bar{r}_i = \beta_i \lambda + \nu_i. \quad (9.3/2) \]
• Notation: \( \lambda \) – the regression coefficient – is the risk premium, and \( \nu_i \) — the regression disturbance – is the pricing error.

• Assume for analytic tractability that there is a single factor, let \( \tilde{r} = [\tilde{r}_1, \tilde{r}_2, \ldots, \tilde{r}_N]' \), and let \( \beta = [\beta_1, \beta_2, \ldots, \beta_N]' \).

• The OLS cross-sectional estimates are

\[
\hat{\lambda} = (\beta'\beta)^{-1}\beta'\tilde{r}, \quad (9.3/3)
\]

\[
\hat{\nu} = \tilde{r} - \hat{\lambda}\beta. \quad (9.3/4)
\]
• Furthermore, let $\Sigma$ be the covariance matrix of asset returns, then it follows that

$$\sigma^2(\hat{\lambda}) = \frac{1}{T} (\beta' \beta)^{-1} \beta' \Sigma \beta (\beta' \beta)^{-1}, \quad (9.3/5)$$

$$\text{cov}(\hat{\nu}) = \frac{1}{T} (I - \beta (\beta' \beta)^{-1} \beta') \Sigma (I - \beta (\beta' \beta)^{-1} \beta'). \quad (9.3/6)$$

• We could test whether all pricing errors are zero with the statistic

$$\hat{\nu}' \text{cov}(\hat{\nu})^{-1} \hat{\nu} \sim \chi^2_{N-1}. \quad (9.3/7)$$

• We could also test whether a factor is priced

$$\frac{\hat{\lambda}}{\sigma(\hat{\lambda})} \sim t_{N-1} \quad (9.3/8)$$
• Thus far, we assume that $\beta$ in (9.3/2) is known.

• However, $\beta$ is estimated in the time-series regression and therefore is not a known parameter.

• So we have the EIV problem.

• Shanken (1992) corrects the cross-sectional estimates to account for the errors in estimating betas.

• Shanken assumes homoscedasticity in the variance of asset returns conditional upon the realization of factors.
• Under this assumption he shows that the standard errors based on the cross sectional procedure overstate the precision of the estimated parameters.

• The EIV corrected estimates are

\[
\sigma_{eiv}^2(\hat{\lambda}) = \sigma^2(\hat{\lambda}) \Upsilon + \frac{1}{T} \Omega_f
\]

\[
cov_{eiv}(\hat{\nu}) = cov(\hat{\nu}) \Upsilon,
\]

(9.3/9)

(9.3/10)

• where \( \Omega_f \) is the variance-covariance matrix of the factors and \( \Upsilon = 1 + \lambda' \Omega_f^{-1} \lambda \).

• Of course, if factors are return spreads then \( \lambda' \Omega_f^{-1} \lambda \) is the squared Sharpe ratio attributable to a mean-variance efficient investment in the factors.
Fama and MacBeth (FM) Procedure

- FM (1973) propose an alternative procedure for running cross-sectional regressions, and for producing standard errors and test statistics.

- The FM approach involves two steps as well.

- The first step is identical to the one described above. Specifically, estimate beta from a time series regression.

- The second step is different.
• In particular, instead of estimating a single cross-sectional regression with the sample averages on the estimated betas, FM run a cross-sectional regression at each time period

\[ r_{i,t} = \delta_{0,t} + \beta_i' \delta_{1,t} + \epsilon_{i,t}. \] (9.4/1)

• Let \( r_t = [r_{1,t}, r_{2,t}, ..., r_{N,t}]' \), let \( \delta_t = [\delta_{0,t}, \delta_{1,t}]' \), let \( X_i = [1, \beta_i]' \), and let \( X = [X_1, X_2, ..., X_N]' \) then the cross sectional estimates for \( \delta_t \) and \( \epsilon_{i,t} \) are given by

\[ \hat{\delta}_t = (X'X)^{-1}X'r_t, \] (9.4/2)

\[ \hat{\epsilon}_{i,t} = r_{i,t} - X_i' \hat{\delta}_t. \] (9.4/3)
• FM suggest that we estimate $\delta$ and $\epsilon_i$ as the averages of the cross-sectional estimates

$$\hat{\delta} = \frac{1}{T} \sum_{t=1}^{T} \delta_t ,$$

(9.4/4)

$$\hat{\epsilon}_i = \frac{1}{T} \sum_{t=1}^{T} \hat{\epsilon}_{i,t} .$$

(9.4/5)

• They suggest that we use the cross-sectional regression estimates to generate the sampling error for these estimates

$$\sigma^2(\hat{\delta}) = \frac{1}{T^2} \sum_{t=1}^{T} (\delta_t - \hat{\delta})^2 ,$$

(9.4/6)

$$\sigma^2(\hat{\epsilon}_i) = \frac{1}{T^2} \sum_{t=1}^{T} (\hat{\epsilon}_{i,t} - \hat{\epsilon}_i)^2 .$$

(9.4/7)
• In particular, let \( \hat{\varepsilon} = [\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_N]' \), then the variance-covariance matrix of the sample pricing errors is

\[
\text{cov}(\hat{\varepsilon}) = \frac{1}{T^2} \sum_{t=1}^{T} (\hat{\varepsilon}_t - \bar{\varepsilon})(\hat{\varepsilon}_t - \bar{\varepsilon})'.
\] (9.4/8)

• Then we can test whether all pricing errors are zero using the test statistic (6.2/7).
The SDF Approach

• The absence of arbitrage in a dynamic economy guarantees the existence of a strictly positive discount factor that prices all traded assets [see Harisson and Kreps (1979)].

• Asset prices are set by the investors’ first order condition:

\[ \mathbb{E}[\xi_{t+1} R_{i,t+1} | J_t] = 1, \quad (9.5/1) \]

• where \( \mathbb{E}[\cdot | J_t] \) is the expectation operator conditioned on \( J_t \), the full set of information available to investors at time \( t \).
The fundamental pricing equation (9.5/1) holds for any asset either stock, bond, option, or real investment opportunity.

It holds for any two subsequent periods $t$ and $t + 1$ of a multi-period model.

It does not assume complete markets

It does not assume the existence of a representative investor

It does not assume equilibrium in financial markets.

It imposes no distributional assumptions about asset returns nor any particular class of preferences.
• Let us now replace the consumption-based expression for marginal utility growth with a linear model obeying the form

\[ \xi_{t+1} = a_t + b'_t f_{t+1}. \] (9.5/2)

• Notation: \( a_t \) and \( b_t \) are fixed or time-varying parameters and \( f_{t+1} \) denotes \( K \times 1 \) vector of fundamental factors that are proxies for marginal utility growth.

• Theoretically, the pricing kernel representation is equivalent to the beta pricing specification.

• See equations (14) and (15) in Avramov (2004) and the references therein.
The CAPM, for one, says that
\[ \xi_{t+1} = a_t + b_t r_{w,t+1}, \]  
where \( r_{w,t+1} \) is the time \( t + 1 \) return on a claim to total wealth.
Is the pricing Kernel linear or nonlinear in the factors?

- In a single-period economy the pricing kernel is given by

\[ \xi_{t+1} = \frac{U'(W_{t+1})}{U'(W_t)}. \]  

(9.5.2/1)
• The Taylor’s series expansion of the pricing kernel around $U'(W_t)$ is

$$\xi_{t+1} = 1 + \frac{W_t U''(W_t)}{U'(W_t)} r_{w,t+1} + o(W_t), \quad (9.5.2/2)$$

$$= a + b r_{w,t+1}, \quad (9.5.2/3)$$

where $a = 1 + o(W_t)$ and $b$ is the negative relative risk aversion coefficient.

• This first order approximation results in the traditional CAPM.
• The second order approximation is given by

\[ \xi_{t+1} = 1 + \frac{W_t U''(W_t)}{U'(W_t)} r_{w,t+1} + \frac{W_t^2 U'''(W_t)}{2U'(W_t)} r_{w,t+1}^2 + o(W_t), \]

\[ = a + br_{w,t+1} + cr_{w,t+1}^2. \] (9.5.2/4)

• This additional factor is related to co-skewness in asset returns.

• Harvey and Siddique (2000) exhibit the relevance of this factor in explaining the cross-sectional variation in expected returns.
Let us start with preferences represented by $U(c_t) = \frac{c_t^{1-\gamma}}{1-\gamma}$.

Take Taylor’s series expansion of the pricing kernel $\rho \frac{U'(c_{t+1})}{U'(c_t)}$ around $U'(c_t)$ and obtain

$$\xi_{t+1} = 1 - \gamma \Delta c_{t+1} + o(c_t),$$

$$= a + b \Delta c_{t+1}.$$
• Under the power preferences, pricing kernel parameters are time invariate (time varying) if $\gamma$ is time invariate (time varying).

• Next, consider the habit-formation economy of Campbell and Cochrane (1999).

• The utility function under habit formation

$$U(c_t, x_t) = \frac{(c_t-x_t)^{1-\gamma}}{1-\gamma}, \quad (9.5.3/3)$$

where $x_t$ is the external consumption habit.

• Then the pricing kernel parameters are time-varying even when the risk aversion parameter is constant.
9. Estimating and Evaluating Asset Pricing Models > 9.5 Beta Pricing versus the Discount Factor Representation > 9.5.3 Are pricing kernel parameters fixed or time-varying?

- See, e.g., Lettau and Ludvigson (2000).

- Modeling time variation:
  - Assume that \( a_t \) and \( b_t \) are linear functions of \( z_t \) in a conditional single-factor model:
    \[
    \xi_{t+1} = a(z_t) + b(z_t)f_{t+1},
    \]
    \[
    a_t = a_0 + a_1 z_t, \quad (9.5.3/5)
    \]
    \[
    b_t = b_0 + b_1 z_t. \quad (9.5.3/6)
    \]

- Then a conditional single-factor model becomes an unconditional multifactor model
  \[
  \xi_{t+1} = a_0 + a_1 z_t + b_0 f_{t+1} + b_1 f_{t+1} z_t. \quad (9.5.3/7)
  \]
The set of factors is \( [z_t, f_{t+1}, f_{t+1}z_t]' \).

The multi-factor representation of (5.3/7) is

\[
\xi_{t+1} = a_0 + a_1'z_t + b_0'f_{t+1} + b_1'[f_{t+1} \otimes z_t].
\]  

Later we will formulate asset pricing tests based on the pricing kernel representation.

But we first focus on the beta pricing representation.

Even when both methods are equivalent theoretically - the empirical tests and their statistical properties are quite different.

Below we present test statistics based on both methods.
The Current State of Asset Pricing Models

• The CAPM has been rejected in asset pricing tests.

• The Fama-French model is not a big success.

• Conditional versions of the CAPM and CCAPM display some improvement.

• Should decision-makers abandon a rejected CAPM?
Should a Rejected CAPM be Abandoned?

- Not necessarily!

- Assume that expected stock return is given by

\[ \mu_i = \alpha_i + R_f + \beta_i (\mu_m - R_f) \text{ where } \alpha_i \neq 0 \]

- You estimate \( \mu_i \) using the sample mean and CAPM:

\[
\hat{\mu}_i^{(1)} = \frac{1}{T} \sum_{t=1}^{T} R_{it}
\]

\[
\hat{\mu}_i^{(2)} = R_f + \hat{\beta}_i (\hat{\mu}_m - R_f)
\]
Mean Standard Error (MSE)

- The quality of estimates is evaluated based on the Mean Squared Error (MSE)

\[ MSE^{(1)} = E \left( \hat{\mu}_i^{(1)} - \mu_i \right)^2 \]

\[ MSE^{(2)} = E \left( \hat{\mu}_i^{(2)} - \mu_i \right)^2 \]
MSE Bias and Noise of Estimates

• Notice that $MSE = bias^2 + Var(estimate)$

• Of course, the sample mean is unbiased thus

$$MSE^{(1)} = Var(\hat{\mu}_i^{(1)})$$

• However, the CAPM is rejected, thus

$$MSE^{(2)} = \alpha_i^2 + Var(\hat{\mu}_i^{(2)})$$
The Bias – Variance Tradeoff

• It might be the case that $\text{Var}(\hat{\mu}^{(2)})$ is significantly lower than $\text{Var}(\hat{\mu}^{(1)})$ - thus even when the CAPM is rejected, still zeroing out $\alpha_i$ could produce a smaller mean square error.
When is The Rejected CAPM Superior?

\[
Var \left( \hat{\mu}_i^{(1)} \right) = \frac{1}{T} \sigma^2 (R_i) = \frac{1}{T} \left[ \beta_i^2 \sigma^2 (R_m) + \sigma^2 (\varepsilon_i) \right]
\]

\[
Var \left( \hat{\mu}_i^{(2)} \right) = Var (\hat{\beta}_i \hat{\mu}_m)
\]
When is The Rejected CAPM Superior?

Using variance decomposition

\[ \text{Var}(\hat{\mu}_i^{(2)}) = \text{Var}(\hat{\beta}_i \hat{\mu}_m) = E[\text{Var}(\hat{\beta}_i \hat{\mu}_m | \hat{\mu}_m)] + \text{Var}(E(\hat{\beta}_i \hat{\mu}_m | \hat{\mu}_m)) \]

\[ = E \left[ (\hat{\mu}_m^e)^2 \frac{\sigma^2(\varepsilon_i)}{\hat{\sigma}_m^2} \cdot \frac{1}{T} \right] + \text{Var}(\beta_i \hat{\mu}_m) = \frac{1}{T} \sigma^2(\varepsilon_i)E \left[ \left( \frac{\hat{\mu}_m^e}{\hat{\sigma}_m^2} \right)^2 \right] + \beta_i^2 \frac{\hat{\sigma}_m^2}{T} \]

\[ = \frac{1}{T} \left[ SR_m \sigma^2(\varepsilon_i) + \beta_i^2 \hat{\sigma}_m^2 \right] \]

where

\[ SR_m = E \left[ \left( \frac{\hat{\mu}_m^e}{\hat{\sigma}_m^2} \right)^2 \right] \]
When is The Rejected CAPM Superior?

• Then

\[
\frac{\text{Var} \left( \hat{\mu}_i^{(2)} \right)}{\text{Var} \left( \hat{\mu}_i^{(1)} \right)} = \frac{SR_m \sigma^2(\varepsilon_i) + \beta_i^2 \sigma_m^2}{\sigma^2(R_i)} = SR_m (1 - R^2) + R^2
\]

where \( R^2 \) is the \( R \) squared in the market regression.

• Since \( SR_m \) is small -- the ratio of the variance estimates is smaller than 1.
Example

• Let

\[ \sigma^2(R_i) = 0.01 \]
\[ E \left( \frac{\hat{\mu}_m^e}{\hat{\sigma}_m} \right)^2 = 0.05 \]
\[ R^2 = 0.3 \]

• For what values of \( \alpha_i \neq 0 \) it is still preferred to use the CAPM?

• Find \( \alpha_i \) such that the MSE of the CAPM is smaller.
Example

\[
\frac{MSE^{(2)}}{MSE^{(1)}} = SR_m(1 - R^2) + R^2 + \frac{\alpha_i^2}{MSE^{(1)}} < 1
\]

\[
0.05 \times 0.7 + 0.3 + \frac{\alpha_i^2}{\frac{1}{60} \cdot 0.01} < 1
\]

\[
\alpha_i^2 < \frac{1}{60} \cdot 0.01 \times 0.665
\]

\[
|\alpha_i| < 0.01528 = 1.528\%
\]
We can test finance theories by the GMM of Hansen (1982).

Let us describe the basic concepts of GMM and propose some applications.

Let \( \Theta \) be an \( m \times 1 \) vector of parameters to be estimated from a sample of observations \( x_1, x_2, \ldots, x_T \).

Our conditional beta factor model does not have that unconditional multifactor interpretation since the firm-level \( Size_i \) and \( BM_j \) are not common across all test assets.
• One drawback in the maximum likelihood principle is that it requires specifying the joint density of the observations.

• The ML principle is indeed a parametric one.

• ML typically makes the IID, Normal, and homoskedastic assumptions.

• All these assumptions can be relaxed in the GMM framework.

• The GMM only requires specification of certain moment conditions (often referred as orthogonality conditions) rather than the full density.
• It is therefore considered a nonparametric approach.

• Do not get it wrong: The GMM is not necessarily perfect.

• First, it may not make efficient use of all the information in the sample.

• Second, nonparametric approaches typically have low power in out of sample tests possibly due to over-fitting.

• Also the GMM is asymptotic and can have poor, even measurable, final sample properties.
Let $f_t(\Theta)$ be an $r \times 1$ vector of moment conditions.

Note that $f_t$ is not necessarily linear in the data or the parameters, and it can be heteroskedastic and serially correlated.

If $r = m$, i.e., if there is the same number of parameters as there are moments, then the system is exactly identified.

In this case one could find the GMM estimate $\hat{\Theta}$, which satisfies

\[ E(f_t(\hat{\Theta})) = 0. \]  

(9.6.1/1)
• However, in testing economic theories, there should be more moment conditions than there are parameters.

• In this case, one cannot set all the moment conditions to be equal to zero, as I show below.

• Let us analyze both cases of exact identification \( r = m \) and over identification \( r > m \).

• To implement the GMM first compute the sample average of \( E[f_t(\Theta)] \) given by

\[
g_T(\Theta) = \frac{1}{T} \sum_{t=1}^{T} f_t(\Theta). \tag{9.6.1/2}
\]
• If \( r = m \), then the GMM estimator \( \hat{\Theta} \) solves

\[
g_T(\hat{\Theta}) = \frac{1}{T} \sum_{t=1}^{T} f_t (\hat{\Theta}) = 0. \tag{9.6.1/3}
\]

• Otherwise, the GMM estimator minimizes the quadratic form

\[
J_T(\Theta) = g_T(\Theta)'W_T g_T(\Theta), \tag{9.6.1/4}
\]

• where \( W_T \) is some \( r \times r \) weighing matrix to be discussed later.
Differentiating (9.6.1/4) with respect to \( \Theta \) yields

\[
D_T(\Theta)'W_Tg_T(\Theta),
\]  

where

\[
D_T(\Theta)'W_Tg_T(\Theta),
\]

The GMM estimator \( \hat{\Theta} \) solves

\[
D_T(\hat{\Theta})'W_Tg_T(\hat{\Theta}) = 0. \]

Observe from (9.6.1/7) that the left (and obviously the right) hand side is an \( m \times 1 \) vector. Therefore, as \( r > m \) only a linear combination of the moments, given by \( D_T(\hat{\Theta})'W_T \), is set to zero.
Hansen (1982, theorem 3.1) tells us the asymptotic distribution of the GMM estimate is

$$\sqrt{T}(\hat{\Theta} - \Theta) \sim N(0, V),$$

where

$$V = (D_0'WD_0)^{-1}D_0'WSWD_0(D_0'WD_0)^{-1},$$

$$S = \lim_{T \to \infty} V \text{ar} [\sqrt{T}g_T(\Theta)],$$

$$= \sum_{j=-\infty}^{\infty} E \left[ f_t(\Theta)f_{t-j}(\Theta)' \right],$$

$$D_0 = E \left[ \frac{\partial g_T(\Theta)}{\partial \Theta} \right] = \frac{1}{T} \sum_{t=1}^{T} E \left[ \frac{\partial f_t(\Theta)}{\partial \Theta} \right].$$

To implement the GMM one would like to replace $S$ with its sample estimate.
• If the moment conditions are serially uncorrelated then

\[ S_T = \frac{1}{T} \sum_{t=1}^{T} f_t(\Theta)f_t(\hat{\Theta})'. \]  \hspace{1cm} (9.6.1/13)

• We have not yet addressed the issue of how to choose the optimal weighting matrix.

• Hansen shows that optimally \( W = S^{-1} \).

• The optimal \( V \) matrix is therefore

\[ V^* = (D_0'S^{-1}D_0)^{-1}. \] \hspace{1cm} (9.6.1/14)
Moreover, if $W = S^{-1}$, i.e., if the weighting matrix is chosen optimally, then an overidentifying test statistic is given by

$$TJ_T(\Theta) \sim \chi_{r-m}^2.$$  \hspace{1cm} (9.6.1/15)

- This statistic is quite intuitive.
- In particular, note from (9.6.1/10) that $S_T = TVar[g_T(\Theta)]$. 

• Thus, the test statistic in (9.6.1/15) can be expressed as the minimized value of the model errors (in asset pricing context pricing errors) weighted by their covariance matrix

\[ g_T(\hat{\Theta})' \{\text{var}[g_T(\hat{\Theta})]\}^{-1} g_T(\hat{\Theta}) \sim \chi^2_{r-m}. \]  

(9.6.1/16)

• Below we display several applications of the GMM.

• The work of Hansen and Singleton (1982, 1983) is, to my knowledge, the first to apply the GMM in general and in the context of asset pricing in particular.
Application# 1: Estimating the mean of a time series

- You observe $x_1, x_2, \ldots, x_T$ and want to estimate the sample mean.
- In this case there is a single parameter $\Theta = \mu$ and a single moment condition
  \[ f_t(\Theta) = (x_t - \mu). \]  
  
  (9.6.2/1)
- The system is exactly identified.
• Notice that

\[ g_T(\Theta) = \frac{1}{T} \sum_{t=1}^{T} (x_t - \mu). \] (9.6.2/2)

• Setting

\[ g_T(\hat{\Theta}) = 0. \] (9.6.2/3)

• Then the GMM estimate for \( \mu \) is the sample mean

\[ \hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} x_t. \] (9.6.2/4)
Moreover, if $x_t$’s are uncorrelated then

$$S = E[f_t(\Theta)f_t(\Theta)'], \quad (9.6.2/5)$$

estimated using

$$S_T = \frac{1}{T} \sum_{t=1}^{T} f_t(\hat{\Theta})f_t(\hat{\Theta})' = \frac{1}{T} \sum_{t=1}^{T} (x_t - \hat{\mu})^2 \quad (9.6.2/6)$$

To compute the variance of the estimate we need to find $D_0$:

$$D_0 = E\left[\frac{\partial g_T(\theta')}{\partial \theta}\right] = \frac{1}{T} \sum_{t=1}^{T} E\left[\frac{\partial f_t(\hat{\Theta})}{\partial \theta'}\right] = -1. \quad (9.6.2/7)$$
The optimal $V$ matrix is then given by

$$V = (D_0' S^{-1} D_0)^{-1} = S.$$  \hspace{1cm} (9.6.2/8)

Use $S_T$ as a consistent estimator.

Asymptotically

$$\hat{\mu} \sim N\left(\mu, \frac{1}{T} V\right).$$
Application # 2: Estimating the market model coefficients when the residuals are heteroskedastic and serially uncorrelated

- In this application we will focus on a single security.
- The next application expands the analysis to accommodate multiple assets.
- Here is the market model for security $i$

$$r_{i,t} = \alpha_i + \beta_i r_{m,t} + \epsilon_{i,t}. \quad (9.6.3/1)$$
There are two parameters: $\Theta = [\alpha_i, \beta_i]'$.

There are also two moment conditions

\[
\begin{align*}
    f_t(\Theta) &= \left[ r_{i,t} - \alpha_i - \beta_i r_{m,t} \right] (6) \\
    &= (r_{i,t} - \alpha_i - \beta_i r_{m,t}) r_{m,t} (7).
\end{align*}
\] (9.6.3/2)

Let us rewrite the moment conditions compactly using the following form

\[
    f_t(\Theta) = x_t (r_t - x_t' \beta),
\] (9.6.3/3)
where

\[ r_t = r_{i,t}, \quad (9.6.3/4) \]

\[ x_t = [1, r_{m,t}]', \quad (9.6.3/5) \]

\[ \epsilon_t = \epsilon_{i,t}, \quad (9.6.3/6) \]

\[ \beta = [\alpha_i, \beta_i]'. \quad (9.6.3/7) \]

Let us now compute the sample moment

\[ g_T(\Theta) = \frac{1}{T} \sum_{t=1}^{T} x_t (r_t - x_t' \beta). \quad (9.6.3/8) \]
Since the system is exactly identified setting $g_T(\hat{\Theta}) = 0$ yields the GMM estimate

$$
\hat{\beta} = (\sum_{t=1}^{T} x_t x_t')^{-1} (\sum_{t=1}^{T} x_t r_t),
$$

(9.6.3/9)

$$
= (X'X)^{-1} X'R,
$$

(9.6.3/10)

where $X = [x_1, x_2, ..., x_T]'$ and $R = [r_1, r_2, ..., r_T]'$.

The GMM estimator for $\beta$ is the usual OLS estimator.
9. Estimating and Evaluating Asset Pricing Models > 9.6 GMM > 9.6.3 Application # 2: Estimating the market model coefficients when the residuals are heteroskedastic and serially uncorrelated

- To find $V$ first compute
  \[
  \frac{\partial f_t(\theta)}{\partial \theta'} = -x_t x_t', \quad (9.6.3/11)
  \]
  \[
  \frac{\partial g_T(\theta)}{\partial \theta'} = D_T(\Theta) = -\frac{1}{T} \sum_{t=1}^{T} x_t x_t' = -\frac{X'X}{T}. \quad (9.6.3/12)
  \]

- Moreover, if $\epsilon'$s are serially uncorrelated then
  \[
  S_T = \frac{1}{T} \sum_{t=1}^{T} f_t(\hat{\Theta}) f_t(\hat{\Theta})', \quad (9.6.3/13)
  \]
  \[
  = \frac{1}{T} \sum_{t=1}^{T} x_t \hat{\epsilon}_t \hat{\epsilon}_t' x_t', \quad (9.6.3/14)
  \]
  \[
  = \frac{1}{T} \sum_{t=1}^{T} x_t x_t' \hat{\epsilon}_t^2. \quad (9.6.3/15)
  \]
Using the optimal weighting matrix, it follows that

\[ \text{Var}(\hat{\beta}) = \frac{1}{T}(D_0' S^{-1} D_0)^{-1}, \]  

estimated by

\[ \widehat{\text{Var}}(\hat{\beta}) = \frac{1}{T}(D_T' S_T^{-1} D_T)^{-1}, \]

\[ = (X'X)^{-1}(\sum_{t=1}^{T} x_t \ x_t' \hat{\epsilon}_t^2)(X'X)^{-1}. \]

Application # 3: Testing the CAPM

• Here, we derive the CAPM test described on pages 208-210 in Campbell, Lo, and MacKinlay.

• The specification that we have is

\[ r_t^e = \alpha + \beta r_{mt}^e + \epsilon_t. \]  

(6.5.4/1)

• The CAPM says \( \alpha = 0 \).

• The \( 2N \times 1 \) parameter vector in the CAPM model is described by \( \Theta = [\alpha', \beta']' \).
• In the following I will give a recipe for implementing the GMM in estimating and testing the CAPM.

1. Start with identifying the $2N$ moment conditions:

$$f_t(\Theta) = x_t \otimes \epsilon_t = \begin{bmatrix} 1(26) \\ r^e_{m,t} \end{bmatrix} \otimes \epsilon_t = \begin{bmatrix} \epsilon_t(27) \\ r^e_{m,t} \epsilon_t \end{bmatrix}, \quad (9.6.4/2)$$

where $\epsilon_t = r^e_t - (x_t' \otimes I_N) \Theta$
2. Compute $D_0$.

$$\frac{\partial f_t(\Theta)}{\partial \Theta} = x_t \otimes -(x_t' \otimes I_N), \quad (9.6.4/3)$$

$$= -\begin{bmatrix} 1 & r^e_m(t) \\ r^e_m(t) & r^{e^2}_{m,t} \end{bmatrix} \otimes I_N. \quad (9.6.4/4)$$

Moreover,

$$D_0 = E \left[ \frac{\partial g_T(\Theta)}{\partial \Theta} \right] \quad (9.6.4/5)$$

$$= E \left[ \frac{\partial f_t(\Theta)}{\partial \Theta} \right] \quad (9.6.4/6)$$

$$= -\begin{bmatrix} 1 & \mu_m(34) \\ \mu_m & \mu^2_m + \sigma^2_m \end{bmatrix} \otimes I_N, \quad (9.6.4/7)$$

where $\mu_m = E(r_{m,t})$ and $\sigma^2_m = var(r_{m,t})$. 

3. In implementing the GMM, $D_0$ will be replaced by its sample estimate, which amounts to replacing the population moments $\mu_m$ and $\sigma^2_m$ by their sample analogs $\hat{\mu}_m$ and $\hat{\sigma}^2_m$.

4. That is,

$$D_T = -\begin{bmatrix} \frac{1}{\hat{\mu}_m} & \hat{\mu}_m(36) \\ \hat{\mu}_m & \hat{\mu}_m^2 + \hat{\sigma}^2_m \end{bmatrix} \otimes I_N,$$

(9.6.4/8)

5. There are as many moments conditions as there are parameters.

6. Still, you can test the CAPM since you only focus on the model restriction $\alpha = 0$. 
7. In particular, compute $g_T(\Theta)$ and find the GMM estimator

$$g_T(\Theta) = \frac{1}{T} \sum_{t=1}^{T} f_t(\Theta),$$

$$= \frac{1}{T} \sum_{t=1}^{T} (x_t \otimes \epsilon_t).$$

The GMM estimator $\hat{\Theta}$ satisfies $g_T(\hat{\Theta}) = 0$

$$\frac{1}{T} \sum_{t=1}^{T} (x_t \otimes \hat{\epsilon}_t) = 0$$

$$\frac{1}{T} \sum_{t=1}^{T} (x_t \otimes [r_t - (x_t' \otimes I_N)\hat{\Theta}]) = 0$$
The GMM estimator is thus given by

$$\hat{\Theta} = \left[ \begin{pmatrix} 1 & \hat{\mu}_m \\ \hat{\mu}_m^2 + \hat{\sigma}_m^2 & -\hat{\mu}_m \\ -\hat{\mu}_m & 1 \end{pmatrix} \otimes I_N \right] \left( \begin{pmatrix} \hat{\mu} \\ \hat{\sigma}_{rm} + \hat{\mu}\hat{\mu}_m \end{pmatrix} \right),$$

$$= \left[ \frac{1}{\hat{\sigma}_m^2} \begin{pmatrix} \hat{\mu}_m^2 + \hat{\sigma}_m^2 & -\hat{\mu}_m \\ -\hat{\mu}_m & 1 \end{pmatrix} \otimes I_N \right] \left( \begin{pmatrix} \hat{\mu} \\ \hat{\sigma}_{rm} + \hat{\mu}\hat{\mu}_m \end{pmatrix} \right),$$

$$= \begin{bmatrix} \hat{\mu} - \frac{\hat{\sigma}_{rm}}{\hat{\sigma}_m^2} \hat{\mu}_m \\ \frac{\hat{\sigma}_{rm}}{\hat{\sigma}_m^2} \hat{\mu}_m \end{bmatrix} = \begin{bmatrix} \hat{\alpha} \\ \hat{\beta} \end{bmatrix}. \quad (9.6.4/14)$$

These are the OLS estimators for the CAPM parameters.
8. Estimate \( S \) assuming that the moment conditions are serially uncorrelated

\[
S_T = \frac{1}{T} \sum_{t=1}^{T} f_t (\hat{\Theta}) f_t (\hat{\Theta})',
\]

\[
= \frac{1}{T} \sum_{t=1}^{T} (x_t x_t' \otimes \hat{\epsilon}_t \hat{\epsilon}_t')
\]  

(9.6.4/15)

9. Given \( S_T \) and \( D_T \) compute \( V_T \) the sample estimate of the optimal variance matrix

\[
V_T = (D_T' S_T^{-1} D_T)^{-1}
\]

(9.6.4/17)

10. The asymptotic distribution of \( \hat{\Theta} \) is given by

\[
\hat{\Theta} = \left( \hat{\alpha} \right) \sim N \left( \left( \alpha \right), \frac{1}{T} V \right),
\]

so you should substitute \( V_T \) for \( V \).
11. Now, we can derive the test statistic. In particular, let $\alpha = R\Theta$ where $R = [I_N, 0_N]$ and note that under the null hypothesis $H_0: \alpha = 0$ the asymptotic distribution of $R\hat{\Theta}$ is given by
\[
R\hat{\Theta} \sim N \left( 0, R \left( \frac{1}{T} V \right) R' \right).
\]
(9.6.4/19)

The statistic $J_7$ in CLM is derived using the Wald statistic
\[
J_7 = \bar{\Theta}' R' \left( R \left( \frac{1}{T} V_T \right) R' \right)^{-1} R\hat{\Theta},
\]
\[
= T\hat{\alpha}' \left( R (D_T' S_T^{-1} D_T)^{-1} R' \right) \hat{\alpha}.
\]
(6.5.4/21)

12. Under the null hypothesis $J_7 \sim \chi^2_N$. 
It should be noted that if the regression errors are both serially uncorrelated and homoscedastic then the matrix $S_T$ in (9.6.4/16) is

$$\frac{1}{T} (X'X) \otimes \Sigma = \begin{bmatrix} 1 & \hat{\mu}_m(55) \\ \hat{\mu}_m & \hat{\mu}_m^2 + \hat{\sigma}_m^2 \end{bmatrix} \otimes \Sigma \quad (9.6.4/22)$$

Thus $J_7$ in (9.6.4/21) becomes the traditional Wald statistic in (#REMOVED#).

So, the GMM test statistic is a generalized version of Wald correcting for heteroscedasticity.

You can also relax the non serial correlation assumption.
Application # 4: Asset pricing tests based on over-identification

- Harvey (1989) nicely implements the GMM to test conditional asset pricing models.

- The conditional CAPM, wlog, implies that

\[
\mathbb{E}(r_t|z_{t-1}) = \text{cov}(r_t, r_{mt}|z_{t-1})\lambda_t, \quad (9.6.5/1)
\]

\[
= \mathbb{E}[(r_t - \mathbb{E}[r_t|z_{t-1}])(r_{mt} - \mathbb{E}[r_{mt}|z_{t-1}])|z_{t-1}]\lambda_t,
\]

where \(z_t\) denotes a set of \(M\) instruments observed at time \(t\).
Let us assume that $\lambda_t$ is constant, that is $\lambda_t = \lambda$ for all $t$.

Let $x_t = [1, z_t']'$.

Moreover,

$$
\mathbb{E}[r_t|z_{t-1}] = \delta_r x_{t-1},
$$

$$
\mathbb{E}[(r_{mt}|z_{t-1}) = \delta_m x_{t-1}. \tag{9.6.5/3}
$$

Then, let us define several residuals

$$
u_{rt} = r_t - \delta_r x_{t-1}, \tag{9.6.5/4}
$$

$$
u_{mt} = r_{mt} - \delta_m x_{t-1}, \tag{9.6.5/5}
$$

$$
e_t = r_t - \mathbb{E}[(r_t - \mathbb{E}[r_t|z_{t-1}) (r_{mt} - \mathbb{E}[r_{mt}|z_{t-1}])|z_{t-1}] \lambda
= r_t - (r_t - \delta_r x_{t-1})(r_{mt} - \delta_m x_{t-1}) \lambda. \tag{9.6.5/6}
$$
Collecting the residuals into one vector yields

\[ f_t(\Theta) = [u_{rt}', u_{mt}, e_t']', \]  

where \( \Theta = [\text{vec}(\delta_r)', \delta_m', \lambda]' \).

That is, there are \( M(N + 1) + 1 \) parameters.

How many moment conditions do we have? More than you think!

Note that

\[ \mathbb{E}[f_t(\Theta)|z_{t-1}] = 0, \]  

which means that we have the following \( 2N + 1 \) moment conditions.
\[ \mathbb{E}[f_t(\Theta)] = 0, \quad (9.6.5/9) \]

as well as \( M(2N + 1) \) additional moment conditions involving the instruments

\[ \mathbb{E}[f_t(\Theta) \otimes z_{t-1}] = 0. \quad (9.6.5/10) \]

- Overall, there are \((2N + 1)(M + 1)\) moment conditions.
- You have more moment conditions than parameters.
- Hence, you can test the model using the \( \chi^2 \) over identifying test.
- Harvey considers several other generalizations.
Application # 5: Testing return predictability over long horizon

• Here is a nice GMM application in the context of return predictability over long return horizon.

• The model estimated is

\[ R_{t+k} = \alpha + \beta' z_t + \epsilon_{t+k} \]  \hspace{1cm} (9.6.6/1)

where

\[ R_{t+k} = \sum_{i=1}^{k} r_{t+i} \]  \hspace{1cm} (9.6.6/2)

with \( r_{t+i} \) being log return at time \( t + i \).
• Fama and French (1989) observe a dramatic increase in the sample $R^2$ as the return horizon grows from one month to four years.

• Kirby (1997) challenges the Fama-French findings using a GMM framework that accounts for serial correlation in the residuals.

• Let us formalize his test statistic.

• There are $M + 1$ parameters $\Theta = (\alpha, \beta')'$, where $\beta$ is a vector of dimension $M$. 
From Hansen (1982)

$$\sqrt{T}(\Theta - \hat{\Theta}) \sim N(0, V),$$

where

$$V = (D_0' S^{-1} D_0)^{-1}.$$

There are $M + 1$ moment conditions:

$$f_t(\Theta) = \begin{bmatrix} R_{t+k} - \alpha - \beta' z_t(3) \\ (R_{t+k} - \alpha - \beta' z_t) z_t(4) \end{bmatrix}.$$

Compute $D_0$:

$$D_0 = \mathbb{E} \left[ \frac{\partial f_t(\Theta)}{\partial \theta'} \right] = \begin{bmatrix} -1 & -\mu_z(6) \\ -\mu_z' & -(\Sigma_z + \mu_z \mu_z') \end{bmatrix}.$$

- It would be useful to take the inverse of $D_0$

$$D_0^{-1} = \begin{bmatrix}
-(1 + \mu_z'\Sigma_z^{-1}\mu_z) & \mu_z'\Sigma_z^{-1} \\
\Sigma_z^{-1}\mu_z & -\Sigma_z^{-1}
\end{bmatrix}. \quad (9.6.6/7)$$

- And the matrix $S$ is given by

$$S = \sum_{j=-\infty}^{\infty} \mathbb{E} \begin{bmatrix}
\epsilon_{t+k}\epsilon_{t+k-j} & \epsilon_{t+k}\epsilon_{t+k-j}Z_{t-j} \\
\epsilon_{t+k}\epsilon_{t+k-j}Z_t & \epsilon_{t+k}\epsilon_{t+k-j}Z_tZ_{t-j}'
\end{bmatrix}. \quad (9.6.6/8)$$

- In estimating predictive regressions we scrutinize the slope coefficients only.

- And we know that

$$\sqrt{T}(\beta - \hat{\beta}) \sim N(0, \tilde{V}), \quad (9.6.6/9)$$

where $\tilde{V}$ is the $M \times M$ lower-right sub-matrix of $V = D_0^{-1}S D_0^{-1}'$. 
It follows that

\[
\tilde{V} = \sum_{j=-\infty}^{\infty} \mathbb{E} \begin{bmatrix}
\mu_z' \Sigma_z^{-1} \\
-\Sigma_z^{-1}(13)
\end{bmatrix}' \begin{bmatrix}
\epsilon_{t+k} \epsilon_{t+k-j} \\
\epsilon_{t+k} \epsilon_{t+k-j} Z_t
\end{bmatrix} \begin{bmatrix}
\mu_z' \Sigma_z^{-1} \\
-\Sigma_z^{-1}(14)
\end{bmatrix}',
\]

\[
= \Sigma_z^{-1} \left[ \sum_{j=-\infty}^{\infty} \mathbb{E} \left( \epsilon_{t+k} \epsilon_{t+k-j} \right) (Z_t - \mu_z) (Z_{t-j} - \mu_z)' \right] \Sigma_z^{-1},
\]

\[
= \Sigma_z^{-1} \left[ \sum_{j=-\infty}^{\infty} \mathbb{E} \left( \delta_{t+k} \delta_{t+k-j}' \right) \right] \Sigma_z^{-1},
\]

(9.6.6/10)

where

\[
\delta_{t+k} = \epsilon_{t+k} Z_t,
\]

(9.6.6/11)

\[
\delta_{t+k-j} = \epsilon_{t+k-j} Z_{t-j}.
\]

(9.6.6/12)
• What if you are willing to assume that there is no autocorrelation in the residuals?

• Then,

\[ \tilde{V} = \Sigma_{z}^{-1} \left[ \mathbb{E} \left( \delta_{t+k} \delta_{t+k}' \right) \right] \Sigma_{z}^{-1}, \]  
\text{ (9.6.6/13)}

which can be estimated by

\[ \hat{V} = \hat{\Sigma}_{z}^{-1} \left[ \frac{1}{T} \sum_{t=1}^{T} \delta_{t+k} \delta_{t+k}' \right] \hat{\Sigma}_{z}^{-1}. \]  
\text{ (9.6.6/14)}

The estimator for \( \tilde{V} \) is identical to the heteroskedasticity-consistent covariance matrix estimator of White (1980).
• What if you are willing to assume that there is no autocorrelation in the residuals and there is no heteroskedasticity?

\[
\tilde{V} = \Sigma_z^{-1} \left[ \mathbb{E}(\epsilon_{t+k}\epsilon_{t+k-j}) \mathbb{E}(z_t - \mu_z) (z_{t-j} - \mu_z)' \right] \Sigma_z^{-1},
\]

(9.6.6/15)

• In that case

\[
\sqrt{T}(\hat{\beta} - \beta) \sim N(0, \sigma_{\epsilon}^2 \Sigma_z^{-1}).
\]

(9.6.6/16)

• Under the null hypothesis that \( \beta = 0 \)

\[
\sqrt{T}\hat{\beta} \sim N(0, \sigma_{\epsilon}^2 \Sigma_z^{-1}).
\]

(9.6.6/17)
• Note also that under the null $\sigma_\epsilon^2 = \sigma_r^2$, where $\sigma_r^2$ is the variance of the cumulative log return.

• Using properties of the $\chi^2$ distribution, it follows that

$$T \frac{\hat{\beta}' \hat{\Sigma} \hat{\beta}}{\hat{\sigma}_r^2} \sim \chi^2(M),$$

suggesting that

$$TR^2 \sim \chi^2(M).$$

• So we are able to derive a limiting distribution for the regression $R^2$. 

Kirby considers cases with heteroskedasticity and serial correlation.

Then the distribution of the regression slope coefficient and the $R^2$ are much more complex.

His conclusion: the $R^2$ in a predictive regression does not increase with the investment horizon.
Incorporating serial correlation

• In most applications in financial economics there is no a priori reason to believe that the regression residuals are serially uncorrelated.

• Consequently, a suitable scheme is required in order to obtain a consistent positive definite estimator of $S$.

• Notice that we cannot estimate the infinite sum in

$$S = \sum_{j=-\infty}^{\infty} E(u_t u_{t-j}).$$

(9.6.7/1)
Therefore, we must limit the number of terms.

More terms means more ability to pick up autocorrelation if there is any.

But this comes at the cost of losing efficiency in finite samples.

Newey and West (1987) propose a popular weighting scheme

\[
\hat{S} = \sum_{j=-k}^{k} \frac{|k-j|}{k} \left( \frac{1}{T} \sum_{t=j+1}^{T} (u_t - \bar{u})(u_{t-j} - \bar{u})' \right),
\]

(9.6.7/2)

where \( \bar{u} \) denotes the sample mean and \( k - 1 \) is the maximum lag length that receives a nonzero weight.
• The value $k$ should increase with the sample size but not too rapidly.

• This estimator guarantees positive definiteness by down-weighting higher order auto-covariances, and it is consistent because the down-weighting disappears asymptotically.

• The Newey and West (1987) weighting scheme is the most commonly used.

• Andrews (1991) develops a more complex, albeit useful, estimator.
Another way to estimate $S$ is by first running a VAR(1) for $u_t$

$$u_t = A u_{t-1} + v_t.$$  \hfill (9.6.7/3)

With this structure you can estimate $S$ via

$$\hat{S} = (I - A)^{-1} E(v_t v_t') (I - A)^{-1'}.$$  \hfill (9.6.7/4)
9. Estimating and Evaluating Asset Pricing Models > 9.6 The HJ Distance Measure
> 9.6.1 Motivation

The Hansen Jagannathan — HJ — (1997) Distance Measure

• The HJ measure is used for comparing and testing asset pricing models.

• Suppose you want to compare the performance of competing, not necessarily nested, asset pricing models.

• If there is only one asset, then you can compare the pricing error, i.e., the difference between the market price of an asset and the hypothetical price implied by a particular SDF.
However, when there are many assets, it is rather difficult to compare the pricing errors across the different candidate SDFs unless pricing errors of one SDF are always smaller across all assets.

One simple idea would be to examine the pricing error on the portfolio (there are infinitely many such portfolios) that is most mispriced by a given model.

Then, the superior model is the one with the smallest pricing error.
However, there is a practical problem in implementing this simple idea.

Suppose there are at least two assets which do not have the same pricing error for a given candidate SDF.

Let $R_{1t}$ and $R_{2t}$ denote the corresponding gross returns.

Suppose that (i) the date $t - 1$ market prices of these payoffs are both unity, and (ii) the model assigns prices of $1 + \psi_i$, i.e., the pricing errors are $\psi_1$ and $\psi_2$. 
Consider now forming a zero-investment portfolio by going long one dollar in security 1 and short one dollar in security 2.

The pricing error of this zero-cost position is $\psi_1 - \psi_2$.

That is, as long as the difference is not zero the pricing error of any portfolio of the two assets can be arbitrarily large by adding a scale multiple of this zero-investment portfolio.
The HJ idea

- HJ propose a way of normalization.
- They suggest examining the portfolio which has the maximum pricing errors among all portfolio payoffs that have the unit second moments.
- Let us demonstrate.
• Suppose that the SDF is modeled as

\[ \xi_t(\Theta) = \Theta_0 + \Theta_{vw}R_t^{vw} + \Theta_{prem}R_{t-1}^{prem} + \Theta_{labor}R_t^{labor} = \Theta'Y_t \]  

(9.6.2/1)

• where

\[ \Theta = [\Theta_0, \Theta_{vw}, \Theta_{prem}, \Theta_{labor}]', \]  

(9.6.2/2)

\[ Y_t = [1, R_t^{vw}, R_{t-1}^{prem}, R_t^{labor}]' \]  

(9.6.2/3)

• Moreover, let \( R_t = [R_{1t}, R_{2t}, ..., R_{Nt}]' \), and let

\[ f_t(\Theta) = R_t\xi_t(\Theta) - \iota_N = R_tY_t'\Theta - \iota_N. \]  

(9.6.2/4)

• Observe that \( E[f_t(\Theta)] \) is the vector of pricing errors.
• In unconditional models, the number of moment conditions is equal to \( N \), the number of test assets.

• HJ show that the maximum pricing error per unity norm of any portfolio of these \( N \) assets is given by

\[
\delta = \sqrt{E[f_t(\Theta)'][E(R_t R_t')^{-1}E[f_t(\Theta)]}. 
\]  

(9.6.2/5)

• This is the HJ distance measure (which is not the HJ bound).
Since the vector $\Theta$ is unknown, a natural way to estimate the system is to choose those values that minimize (6.6.2/5).

We can then assess the specification error of a given stochastic discount factor by examining the maximum pricing error $\delta$.

Next, compute some sample moments

$$D_T = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial f_t(\Theta)}{\partial \Theta} = \frac{1}{T} \sum_{t=1}^{T} R_t Y'_t = \frac{1}{T} R' Y, \quad (9.6.2/6)$$

$$g_T(\Theta) = \frac{1}{T} \sum_{t=1}^{T} f_t (\Theta) = D_T \Theta - \eta_N, \quad (9.6.2/7)$$
\[ G_T = \frac{1}{T} \sum_{t=1}^{T} R_t R'_t = \frac{1}{T} R'R, \]  
(9.6.2/8)

- where

\[ R = [R_1, R_2, ..., R_T]', \]  
(9.6.2/9)

\[ Y = [Y_1, Y_2, ..., Y_T]. \]  
(9.6.2/10)

- The sample analog of the HJ distance is thus

\[ \delta_T = \sqrt{\min_{\Theta} g_T(\Theta)' G_T^{-1} g_T(\Theta)}. \]  
(9.6.2/11)

- The first order condition of the minimization problem

\[ \min_{\Theta} g_T(\Theta) G_T^{-1} g_T(\Theta), \]  
(9.6.2/12)
is given by

\[ D_T' G_T^{-1} g_T(\Theta) = 0, \]  

(9.6.2/13)

which gives an analytic expression for the sample minimizer

\[ \hat{\Theta} = (D_T' G_T^{-1} D_T)^{-1} D_T' G_T^{-1} \eta_N, \]  

(9.6.2/14)

\[ = T(Y'R(R'R)^{-1}R'Y)^{-1}Y'R(R'R)^{-1}\eta_N. \]  

(9.6.2/15)

• It follows that

\[ g_T(\hat{\Theta}) = R'Y(Y'R(R'R)^{-1}R'Y)^{-1}Y'R(R'R)^{-1}\eta_N - \eta_N \]  

(9.6.2/16)
• From Hansen (1982) the asymptotic variance of \( \hat{\Theta} \) is given by

\[
\text{var}(\hat{\Theta}) = \frac{1}{T} (D_T' G_T^{-1} D_T)^{-1} D_T' G_T^{-1} S_T G_T^{-1} D_T (D_T' G_T^{-1} D_T)^{-1} \tag{9.6.2/17}
\]

• where, if the data is serially uncorrelated, the estimate of the variance matrix of pricing errors is given by

\[
S_T = \frac{1}{T} \sum_{t=1}^{T} f_t (\hat{\Theta}) f_t (\hat{\Theta})' \tag{9.6.2/18}
\]
• That is, the estimator $\hat{\Theta}$ is equivalent to a GMM estimator defined by Hansen (1982) with the moment condition $E[f(\Theta)] = 0$ and the weighting matrix $G^{-1}$.

• If the weighing matrix is optimal in the sense of Hansen (1982), then $T \delta_T^2$ is asymptotically a random variable of $\chi^2$ distribution with $N - m$ dof, where $m$ is the dimension of $\Theta$.

• Moreover, the optimal variance of $\hat{\Theta}$ becomes

$$\text{var}(\hat{\Theta}) = \frac{1}{T} (D_T' S_T^{-1} D_T)^{-1}. \quad (9.6.2/19)$$

However, $G$ is generally not optimal, and thus the distribution of $T \delta_T^2$ is not $\chi^2_{N-m}$.
• Instead, the limiting distribution of this statistic is given by

\[ u = \sum_{j=1}^{N-m} \lambda_j \nu_j, \]  

(9.6.2/20)

• where \( \nu_1, \nu_2, \ldots, \nu_{N-m} \) are independent \( \chi^2(1) \) random variables, and \( \lambda_1, \lambda_2, \ldots, \lambda_{N-m} \) are \( N - m \) nonzero eigenvalues of the matrix \( A \) given by

\[
A = S^{0.5} G'^{-0.5} (I_N - (G^{-0.5})D[D'G^{-1}D]^{-1} D'G'^{-0.5})(G^{-0.5})(S^{0.5})',
\]  

(9.6.2/21)

• and where \( S^{0.5} \) and \( G^{0.5} \) are the upper-triangle matrices from the Cholesky decomposition of \( S \) and \( G \).
As long as we have a consistent estimate $S_T$ of the matrix $S$, we can estimate the matrix $A$ by replacing $S$ and $G$ by $S_T$ and $G_T$, respectively.

One can generate a large number of draws from the nonstandard distribution (6.6.2/20) to determine the $p$-value of the HJ distance measure, or whether or not it is equal to zero.

You can follow the below-described algorithm to compute the empirical $p$-value:

1. Use results from (9.6.2/8) and (9.6.2/16) to compute
   $$T \delta_T^2 = T g_T(\Theta) G_T^1 g_T(\Theta).$$
2. Obtain the $N - m$ largest eigenvalues of $\hat{A}$, a consistent estimate of $A$.

3. Generate $N - m$ independent draws from $\chi^2(1)$. For example, using the Matlab command $g = \text{chi2rnd}(\nu, 1000, 1)$ generates 1000 independent draws from $\chi^2(\nu)$.

4. Based on these independent draws, compute the statistic $u_i$ in (9.6.2/20)

5. If $u_i > T \delta_i^2$ set $I_i = 1$. Otherwise set $I_i = 0$.

6. Repeat steps 3-5 100,000 times.

7. The empirical $p$-value is given by $\frac{1}{100,000} \sum_{i=1}^{100,000} I_i$. 

9. Estimating and Evaluating Asset Pricing Models > 9.6 The HJ Distance Measure
> 9.6.2 The HJ Idea
Let us now demonstrate the implementation of the HJ measure when the pricing kernel takes the form

\[
\xi_{t+1} = (\Theta_0'X_t) + (\Theta_1'X_t)f_{t+1}^1 + \cdots + (\Theta_K'X_t)f_{t+1}^K, \quad (9.7/1)
\]

where \( X_t' = [1, Z_t'] \) and \( Z_t \) is an \( M \times 1 \) vector of information variables and \( f_{t+1}^k (k = [1,2, \ldots, K]) \) denotes a proxy for marginal utility growth, or a macroeconomy factor.
As noted earlier the first order condition implies that
\[
E[R_{t+1}((\Theta_0'X_t) + (\Theta_1'X_t)f_{t+1}^1 + \cdots + (\Theta_K'X_t)f_{t+1}^K)|Z_t] = \iota_N .
\] (9.7/2)

We collect the vector of errors,
\[
f_{t+1} = R_{t+1}((\Theta_0'X_t) + (\Theta_1'X_t)f_{t+1}^1 + \cdots + (\Theta_K'X_t)f_{t+1}^K) - \iota_N ,
\]
\[= R_{t+1}Y_{t+1}'\Theta - \iota_N ,
\] (9.7/3)

where
\[
\Theta' = [\Theta_0', \Theta_1', \ldots, \Theta_K'],
\]
\[
Y_{t+1}' = [X_t', X_t'f_{t+1}^1, \ldots, X_t'f_{t+1}^K] .
\] (9.7/4)

- Overall there are \((K + 1)(M + 1)\) parameters to estimate.
- Equation (6.6.3/2)) implies

\[
E[f_{t+1}|Z_t] = 0,
\]

and therefore

\[
E[(f_{t+1} \otimes Z_t)|Z_t] = 0.
\]

- This forms a set of \(N \times (M + 1)\) moment conditions given by the compact notation

\[
g_T(\Theta) = \frac{1}{T} \sum_{t=0}^{T-1} [f_{t+1} \otimes X_t].
\]
To estimate and test the model we minimize the quadratic form

\[ \delta^2 = g_T(\Theta)'G_T^{-1}g_T(\Theta), \]  

(9.7/9)

with \( D \) and \( G \) being estimated by

\[ D_T = \frac{1}{T} \sum_{t=0}^{T-1} [(R_{t+1} \otimes X_t)Y_{t+1}'], \]  

(9.7/10)

\[ G_T = \frac{1}{T} \sum_{t=0}^{T-1} [(R_{t+1} \otimes X_t)(R_{t+1} \otimes X_t)']. \]  

(9.7/11)
Finally, the sample analog of (??0) is

\[ g_T(\hat{\Theta}) = D_T \hat{\Theta} - (\iota_N \otimes \bar{X}), \]  \hspace{1cm} (9.7/12)

where

\[ \hat{\Theta} = (D_T' G_T^{-1} D_T)^{-1} D_T' G_T^{-1} (\iota_N \otimes \bar{X}), \]  \hspace{1cm} (9.7/13)

\[ \bar{X} = \frac{1}{T} \sum_{t=0}^{T-1} X_t. \]  \hspace{1cm} (9.7/14)
HJ distance measure vs. the standard GMM

- Both the GMM and HJ distance measure are cross sectional tests of asset pricing models.

1. Since the distance measure is formed using a weighting matrix that is invariate across competing SDF candidates it can be used to compare the performance of nested and nonnested asset pricing models.

2. In the standard GMM the optimal weighting matrix $S^{-1}$ varies across competing specifications.

3. Therefore, the standard GMM cannot be used for comparing misspecification across competing models.
4. The HJ distance measure avoids the pitfall embedded in the standard GMM of favoring pricing models that produce volatile pricing errors.

5. In the HJ distance measure the weighting matrix is not a function of the parameters, which may result in a more stable estimation procedure.

6. On the other hand, the optimal GMM provides the most efficient estimate among estimates that use linear combinations of pricing errors as moments, in the sense that the estimated parameters have the smallest asymptotic covariance.
References


References


References


References


• Dispersion in analyst’s earnings forecasts and credit rating. forthcoming in Journal of Financial Economics.


References


References


References


References


References


References


References


• Cooper, Michael, Roberto Gutierrez, and Bill Marcum, 2001. On the predictability of stock returns in real time. Unpublished working paper, Purdue University.


References

References


References


References


References

References


References


References


References


References


References


References


• Mankiw, N. Gregory, and Matthew Shapiro, 1986. Do we reject too often? Small sample properties of tests of rational expectations models. Economic Letters 20, 139-145.

References


References


References


References


References

References


References


