Estimating and Evaluating Asset Pricing Models
Why Caring About Asset Pricing Models

- An essential question that arises is why both academics and practitioners invest huge resources attempting to develop and test of 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.
- Likewise, in the single factor CCAPM 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.
- 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

\[ w_{TP} = \frac{\nu^{-1}\mu^e}{\nu'\nu^{-1}\mu^e} \]

- Under the CAPM, the vector of expected returns and the covariance matrix are given by:

\[
\mu^e = \beta \mu^e_m \\
V = \beta\beta'\sigma^2_m + \Sigma
\]

where \( \Sigma \) is the covariance matrix of the residuals in the time-series asset pricing regression.

- The corresponding quantities under the FF model are

\[
\mu^e = \beta_{MKT}\mu^e_m + \beta_{SML}\mu_{SML} + \beta_{HML}\mu_{HML} \\
V = \beta \Sigma_F \beta' + \Sigma
\]

where \( \Sigma_F \) is the covariance matrix of the market, size, and book-t-market 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:

\[
P V = \sum_{t=1}^{T} \frac{C F_t}{(1+k)^t}
\]

- In practical applications, expected returns are typically assumed to be constant over time, an unrealistic assumption.

- Indeed, thus far we have examined models with constant beta and constant risk premiums

\[
\mu^e = \beta' \lambda
\]

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 could 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):

\[
r^e_t = \alpha + \beta_{MKT} \times r^e_{MKT,t} + \beta_{SMB} \times SMB_t \\
+ \beta_{HML} \times HML_t + \beta_{WML} \times WML_t + \varepsilon_t
\]

5 - Corporate Finance

- 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.
5 - Corporate Finance

- Examining the long run performance of SEO firms
- Analyzing abnormal performance of stocks going through splits and reverse splits.
- 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.
Methods of evaluation model pricing abilities

- The finance literature has used three main approaches to evaluate asset pricing models: calibration, cross sectional and time series asset pricing tests, and out-of-sample fit.

- With calibration (e.g., Long run risk, habit formation, prospect theory), values for the parameters of the underlying model are chosen, and the model is solved at these parameter values for the prices of financial assets.

- The model-generated series of prices and returns are examined to see if their moments match key moments of actual asset prices.

- In asset pricing tests, model parameters are optimally chosen to fit a panel of economic series and asset returns.

- Standard errors for the parameter estimates quantify their precision.

- Statistical hypothesis tests about the parameters are conducted and the residuals of the model are examined to assess the fit to the sample.

- Estimation typically challenges a model in more dimensions at once than calibration. For example, a calibrated parameter value may not be the value that maximizes the likelihood, indicating that more issues are going on in the data than captured by the calibration.
The practical utility of an asset pricing model ultimately depends on its ability to fit out-of-sample returns, as most practical applications are, in some sense, out of sample.

For example, firms want to estimate costs of capital for future projects, portfolio and risk managers want to know the expected compensation for future risks, and academic researchers will want to make risk adjustments to expected returns in future data.

Many of these applications rely on out-of-sample estimates for the required or ex ante expected return, where the model parameters are chosen based on available data.

This perspective leads naturally to the mean squared pricing error (MSE) criterion: a better model produces lower MSE.

Ex post out-of-sample performance of asset allocation decisions would establish a solid economic metric for model pricing abilities.

More on asset allocation is in the Bayesian section of the class notes.

This section provides the econometric paradigms of asset pricing tests.
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:
  \[ r_{1t}^e = \alpha_1 + \beta_1 r_{mt}^e + \varepsilon_{1t} \]
  
  : 
  
  \[ r_{Nt}^e = \alpha_N + \beta_N r_{mt}^e + \varepsilon_{Nt} \]

- 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 \( \alpha \times \) Recall, \( a \) is asset mispricing.

The time series regressions can be rewritten using a vector form as:

\[
r_t^e = \alpha_{N \times 1} + \beta_{N \times 1} \cdot r_{mt}^e + \varepsilon_t^{N \times 1}
\]

Let us assume that

\[
\varepsilon_t^{N \times 1} \sim iid N(0, \Sigma_{N \times N})
\]

for \( t = 1,2,3,\ldots,T \)

Let \( \Theta = (\alpha', \beta', vech(\varepsilon)')' \) be the set of all parameters.

Under normality, the likelihood function for \( \varepsilon_t \) is

\[
L(\varepsilon_t | \theta) = c |\Sigma|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (r_t^e - \alpha - \beta r_{mt}^e)' \Sigma (r_t^e - \alpha - \beta r_{mt}^e) \right]
\]

where \( c \) is the constant of integration (recall the integral of a probability distribution function is unity).
Moreover, the IID assumption suggests that

\[ L(\varepsilon_1, \varepsilon_2, ..., \varepsilon_N | \theta) = c^T |\Sigma|^{-\frac{T}{2}} \]

\[ \times \exp \left[ -\frac{1}{2} \sum_{t=1}^{T} (r_t^e - \alpha - \beta r_{mt}^e)' \Sigma^{-1} (r_t^e - \alpha - \beta r_{mt}^e) \right] \]

Taking the natural log from both sides yields

\[ \ln (L) \propto -\frac{T}{2} \ln (|\Sigma|) - \frac{1}{2} \sum_{t=1}^{T} (r_t^e - \alpha - \beta r_{mt}^e)' \Sigma^{-1} (r_t^e - \alpha - \beta r_{mt}^e) \]

Asymptotically, we have \( \theta - \hat{\theta} \sim N(0, \Sigma(\theta)) \)

where

\[ \Sigma(\theta) = \left[ -E \left[ \frac{\partial^2 \ln (L)}{\partial \theta \partial \theta'} \right] \right]^{-1} \]
Let us estimate the parameters

\[
\frac{\partial \ln (L)}{\partial \alpha} = \Sigma^{-1} \left[ \sum_{t=1}^{T} (r_t^e - \alpha - \beta r_{mt}^e) \right]
\]

\[
\frac{\partial \ln (L)}{\partial \beta} = \Sigma^{-1} \left[ \sum_{t=1}^{T} (r_t^e - \alpha - \beta r_{mt}^e) \times r_{mt}^e \right]
\]

\[
\frac{\partial \ln (L)}{\partial \Sigma} = -\frac{T}{2} \Sigma^{-1} + \frac{1}{2} \Sigma^{-1} \left[ \sum_{t=1}^{T} \varepsilon_t \varepsilon_t' \right] \Sigma^{-1}
\]

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,

\[ \hat{\Sigma} = \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_t \hat{\varepsilon}'_t \]

\[ \hat{\mu}^e = \frac{1}{T} \sum_{t=1}^{T} \hat{r}_t^e \]

\[ \mu_m^e = \frac{1}{T} \sum_{t=1}^{T} r_{mt}^e \]

Recall our objective is to find the variance-covariance matrix of \( \hat{\alpha} \).

Standard errors could be found using the information matrix:

\[ I(\theta) = -E \begin{bmatrix} \frac{\partial^2 \ln (L)}{\partial \alpha \partial \alpha'} & \frac{\partial^2 \ln (L)}{\partial \alpha \partial \beta'} & \frac{\partial^2 \ln (L)}{\partial \alpha \partial \Sigma'} \\ \frac{\partial \alpha \partial \beta'} & \frac{\partial \alpha \partial \beta'} & \frac{\partial \alpha \partial \Sigma'} \\ \frac{\partial^2 \ln (L)}{\partial \beta \partial \alpha'} & \frac{\partial^2 \ln (L)}{\partial \beta \partial \beta'} & \frac{\partial \beta \partial \Sigma'} \\ \frac{\partial^2 \ln (L)}{\partial \Sigma \partial \alpha'} & \frac{\partial^2 \ln (L)}{\partial \Sigma \partial \beta'} & \frac{\partial \Sigma \partial \Sigma'} \end{bmatrix} \]
The Distribution of the Parameters

- Try to establish yourself the information matrix.
- Notice that \( \alpha \) and \( \beta \) are independent of \( \Sigma \) - thus, you can ignore the second derivatives with respect to \( \Sigma \) in the information matrix if your objective is to find the distribution of \( \alpha \) and \( \beta \).
- If you aim to derive the distribution of \( \Sigma \) 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,

\[
\hat{\beta} \sim N \left( \beta, \frac{1}{T} \cdot \frac{1}{\hat{\sigma}_m^2} \Sigma \right)
\]

\[
T \hat{\Sigma} \sim W(T - 2, \Sigma)
\]

- 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

\[ X \sim N(\mu, \Sigma) \text{ then } (X - \mu)'\hat{\Sigma}^{-1}(X - \mu) \sim \chi^2(N) \]

- 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 \( \hat{\alpha}'\hat{\Sigma}_{\alpha}^{-1}\hat{\alpha} \sim \chi^2(N) \), which becomes:

\[ J_1 = T \left[ 1 + \left( \frac{\hat{\mu}_m^e}{\hat{\sigma}_m} \right)^2 \right]^{-1} \hat{\alpha}'\hat{\Sigma}_{\alpha}^{-1}\hat{\alpha} = T \frac{\hat{\alpha}'\hat{\Sigma}_{\alpha}^{-1}\hat{\alpha}}{1 + \bar{S}\hat{R}_m^2} \]

where \( \bar{S}\hat{R}_m \) is the Sharpe ratio of the market factor.
The algorithm for implementing the statistic is as follows:

- Run separate regressions for the test assets on the common factor:
  \[ r^e_{t1} = X_{Tx1} \theta_1 + \varepsilon_1 \]
  \[ \vdots \]
  \[ r^e_{TN} = X_{TxN} \theta_N + \varepsilon_N \]

- Where
  \[ X_{Tx2} = [1, r^e_{m1}] \]
  \[ \vdots \]
  \[ X_{Tx2} = [1, r^e_{mT}] \]
  \[ \theta_i = [\alpha_i, \beta_i]' \]

- Retain the estimated regression intercepts
  \[ \hat{\alpha} = [\hat{\alpha}_1, \hat{\alpha}_2, ..., \hat{\alpha}_N]' \]
  \[ \hat{\varepsilon} = [\hat{\varepsilon}_1, ..., \hat{\varepsilon}_N] \]

- 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:

  \[
  \text{un: } r_t^e = \alpha + \beta r_{mt}^e + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma) \\
  \text{res: } r_t^e = \beta^* r_{mt}^e + \varepsilon_t^*, \quad \varepsilon_t^* \sim N(0, \Sigma^*)
  \]

- Using MLE, we get:

  \[
  \hat{\beta}^* = \frac{\sum_{t=1}^{T} r_t^e r_{mt}^e}{\sum_{t=1}^{T} (r_{mt}^e)^2} \\
  \hat{\Sigma}^* = \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_t^* \hat{\varepsilon}_t'^* \\
  \hat{\beta}^* \sim N \left( \beta, \frac{1}{T} \left[ \frac{1}{\hat{\beta}_m^2 + \hat{\sigma}_m^2} \right] \Sigma \right) \\
  T \hat{\Sigma}^* \sim W(T - 1, \Sigma)
  \]
The LR Test

\[ LR = \ln (L^*) - \ln (L) = -\frac{T}{2} [\ln |\hat{\Sigma}^*| - \ln |\hat{\Sigma}|] \]

\[ J_2 = -2LR = T[\ln |\hat{\Sigma}^*| - \ln |\hat{\Sigma}|] \sim \chi^2(N) \]

- 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) – Finite Sample Test

Theorem: let

\[ X_{N \times 1} \sim N(0, \Sigma) \]

let

\[ A_{N \times N} \sim W(\tau, \Sigma) \]

where \( \tau \geq N \) and A and X are independent. Then:

\[ \frac{\tau - N + 1}{N} X' A^{-1} X \sim F_{N, \tau - N + 1} \]
In our context:

\[
X = \sqrt{T} \left[ 1 + \left( \frac{\hat{\mu}_m}{\hat{\sigma}_m} \right)^2 \right]^{\frac{1}{2}} \hat{\alpha} \overset{H_0}{\sim} N(0, \Sigma)
\]

\[
A = T \hat{\Sigma} \sim W(\tau, \Sigma)
\]

where

\[
\tau = T - 2
\]

Then:

\[
J_3 = \left( \frac{T - N - 1}{N} \right) \left[ 1 + \left( \frac{\hat{\mu}_m}{\hat{\sigma}_m} \right)^2 \right]^{-1} \hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha} \sim F(N, T - N - 1)
\]

**GMM**

The GMM test statistic (derivation comes up later in the notes) is given by

\[
J_4 = T \hat{\alpha}' (R(D_T'S_T'^{-1}D_T)^{-1}R')^{-1} \cdot \overset{H_0}{\hat{\alpha}} \sim \chi^2(N)
\]

where

\[
R_{N \times 2N} = \begin{bmatrix} I_N & 0 \\ N \times N & N \times N \end{bmatrix}
\]

\[
D_T = - \begin{bmatrix} 1, & \hat{\mu}_m^e \\ \hat{\mu}_m^e, & (\hat{\mu}_m^e)^2 + \hat{\sigma}_m^2 \end{bmatrix} \otimes I_N
\]
Assume no serial correlation but heteroscedasticity:

\[
S_T = \frac{1}{T} \sum_{t=1}^{T} (x_t x_t' \otimes \hat{\epsilon}_t \hat{\epsilon}_t')
\]

where

\[
x_t = [1, r_{mt}']
\]

Under homoscedasticity 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

\[
r_t^e = \alpha + \beta \cdot F_t + \epsilon_t
\]

\[
J_1 = T \left( 1 + \hat{\mu}_F' \hat{\Sigma}_F^{-1} \hat{\mu}_F \right)^{-1} \hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha} \sim \chi(N)
\]

\( J_2 \) 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:

\[
\hat{\mu}_F = \begin{bmatrix}
\hat{\mu}_e \\
\hat{\mu}_{SMB} \\
\hat{\mu}_{HML}
\end{bmatrix}
\]

\[
\hat{\Sigma}_F = \begin{bmatrix}
\hat{\sigma}_m^2, \hat{\sigma}_{mSMB}, \hat{\sigma}_{mHML} \\
\hat{\sigma}_{SMB,m}, \hat{\sigma}_{SMB}^2, \hat{\sigma}_{SMBHML} \\
\hat{\sigma}_{HML,m}, \hat{\sigma}_{HMLSMB}, \hat{\sigma}_{HML}^2
\end{bmatrix}
\]

### The Economics of Time Series Test Statistics

Let us summarize the first three test statistics:

\[
J_1 = T \frac{\hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha}}{1 + SR_m^2}
\]

\[
J_2 = T \cdot \ln \left( \frac{J_1}{T} + 1 \right)
\]

\[
J_3 = \frac{T - N - 1}{N} \frac{\hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha}}{1 + SR_m^2}
\]
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 heteroscedasticity and often even potential serial correlation.
- Notice that all test statistics depend on the quantity $\hat{\alpha} \hat{\Sigma}^{-1} \hat{\alpha}$
- 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

$$\hat{\lambda}_{(N+1) \times 1} = [\hat{\mu}_m', \hat{\mu}_e']'$$

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

$$
\hat{\Phi}_{(N+1)\times(N+1)} = \begin{bmatrix}
\hat{\sigma}_m^2, & \hat{\beta}'\hat{\sigma}_m^2 \\
\hat{\beta}\hat{\sigma}_m^2, & \hat{\Sigma}
\end{bmatrix}
$$

where $\hat{\sigma}_m^2$ is the estimated variance of the market factor, $\hat{\beta}$ is the N-vector of market loadings, and $\hat{\Sigma}$ is the covariance matrix of the N test assets.

- 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

$$
S\hat{R}_{tp}^2 = \hat{\lambda}'\hat{\Phi}^{-1}\hat{\lambda}
$$

- Notice also that the inverse of the covariance matrix is

$$
\hat{\Phi}^{-1} = \begin{bmatrix}
(\hat{\sigma}_m^2)^{-1} + \hat{\beta}'\hat{\Sigma}^{-1}\hat{\beta}, & -\hat{\beta}'\hat{\Sigma}^{-1} \\
-\hat{\Sigma}^{-1}\hat{\beta}, & \hat{\Sigma}^{-1}
\end{bmatrix}
$$
Understanding the Quantity $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$

- Thus, the squared Sharpe ratio of the tangency portfolio could be represented as

$$SR_{TP}^2 = \left(\frac{\hat{\mu}_m}{\hat{\sigma}_m}\right)^2 + \left[(\hat{\mu}^e - \hat{\beta}\hat{\mu}_m)'\hat{\Sigma}^{-1}(\hat{\mu}^e - \hat{\beta}\hat{\mu}_m)\right]$$

$$SR_{TP}^2 = SR_m^2 + \hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$$

or

$$\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha} = SR_{TP}^2 - SR_m^2$$

- 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.

- 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 $\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha}$

$$\hat{\alpha}'\hat{\Sigma}^{-1}\hat{\alpha} = \Phi_1^2 - \Phi_2^2$$
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

\[ r_{i,t} = \alpha_i + \beta_i f_t + \epsilon_{i,t}. \]

Then run the cross-section regression of average returns on the betas

\[ \bar{r}_i = \beta_i \lambda + \nu_i. \]

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 \( \bar{r} = [\bar{r}_1, \bar{r}_2, ..., \bar{r}_N]' \), and let \( \beta = [\beta_1, \beta_2, ..., \beta_N]' \).
Cross Sectional Regressions

- The OLS cross-sectional estimates are
  \[ \hat{\lambda} = (\beta' \beta)^{-1} \beta' \bar{r}, \]
  \[ \hat{v} = \bar{r} - \hat{\lambda} \beta. \]

- Furthermore, let \( \Sigma \) be the covariance matrix of asset returns, then it follows that
  \[ \sigma^2(\hat{\lambda}) = \frac{1}{T} (\beta' \beta)^{-1} \beta' \Sigma \beta (\beta' \beta)^{-1}, \]
  \[ \text{cov}(\hat{v}) = \frac{1}{T} (I - \beta (\beta' \beta)^{-1} \beta') \Sigma (I - \beta (\beta' \beta)^{-1} \beta'). \]

- We could test whether all pricing errors are zero with the statistic
  \[ \hat{v}' \text{cov}(\hat{v})^{-1} \hat{v} \sim \chi^2_{N-1}. \]

- We could also test whether a factor is priced
  \[ \frac{\hat{\lambda}}{\sigma(\hat{\lambda})} \sim t_{N-1} \]

- Notice that we assume \( \beta \) in is known.

- However, \( \beta \) is estimated in the time-series regression and therefore is unknown.

- So we have the EIV problem.
Cross Sectional Regressions

- 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

\[ \sigma^2_{eiv}(\hat{\lambda}) = \sigma^2(\hat{\lambda})Y + \frac{1}{T}\Omega_f \]

\[ \text{cov}_{eiv}(\hat{\nu}) = \text{cov}(\hat{\nu})Y, \]

where \( \Omega_f \) is the variance-covariance matrix of the factors and \( Y = 1 + \lambda'\Omega_f^{-1}\lambda \).

- 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

\[ r_{i,t} = \delta_{0,t} + \beta_i' \delta_{1,t} + \epsilon_{i,t}. \]

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

\[ \hat{\delta}_t = (X'X)^{-1}X'r_t, \]

\[ \hat{\epsilon}_{i,t} = r_{i,t} - X_i'\hat{\delta}_t. \]
Fama and MacBeth (FM) Procedure

FM suggest that we estimate $\delta$ and $\epsilon_i$ as the averages of the cross-sectional estimates

$$
\hat{\delta} = \frac{1}{T} \sum_{t=1}^{T} \delta_t,
$$

$$
\hat{\epsilon}_i = \frac{1}{T} \sum_{t=1}^{T} \hat{\epsilon}_{i,t}.
$$

They suggest that we use the cross-sectional regression estimates to generate the sampling error for these estimates

$$
\sigma^2(\hat{\delta}) = \frac{1}{T^2} \sum_{t=1}^{T} (\delta_t - \hat{\delta})^2,
$$

$$
\sigma^2(\hat{\epsilon}_i) = \frac{1}{T^2} \sum_{t=1}^{T} (\hat{\epsilon}_{i,t} - \hat{\epsilon}_i)^2.
$$

In particular, let $\hat{\epsilon} = [\hat{\epsilon}_1, \hat{\epsilon}_2, ..., \hat{\epsilon}_N]'$, then the variance-covariance matrix of the sample pricing errors is

$$
cov(\hat{\epsilon}) = \frac{1}{T^2} \sum_{t=1}^{T} (\hat{\epsilon}_t - \hat{\epsilon})(\hat{\epsilon}_t - \hat{\epsilon})'.
$$

Then we can test whether all pricing errors are zero using the WALD test statistic.
Let us now present the methodology in Avramov and Chordia (2006) for asset pricing tests based on individual stocks.

In particular, assume that returns are generated by a conditional version of a $K$-factor model

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

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

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

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:

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

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.

The 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, run the cross-sectional regression

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

where $Z_{mj-1}$ 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 ($R_{jt}^* = c_{0t} + \sum_{m=1}^{M} c_{mt} Z_{mj-1} + e_{jt}$).
To examine significance, we estimate the vector of characteristics rewards each month as

$$\hat{c}_t = (Z'_{t-1}Z_{t-1})^{-1}Z'_{t-1}R^*_t,$$

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 using the generic form

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

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.
Avramov and Chordia (AC) Procedure

- 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

$$
\beta_{jt-1} = \beta_{j1} + \beta_{j2}z_{t-1} + (\beta_{j3} + \beta_{j4}z_{t-1})Size_{jt-1} + (\beta_{j5} + \beta_{j6}z_{t-1})BM_{jt-1},
$$

where $Size_{jt-1}$ and $BM_{jt-1}$ are the market capitalization and the book-to-market ratio at time $t - 1$. 
Avramov and Chordia (AC) Procedure

- The first pass time series regression for the very last specification is
  \[ r_{jt} = \alpha_j + \beta_{j1} r_{mt} + \beta_{j2} z_{t-1} r_{mt} + \beta_{j3} \text{Size}_{jt-1} r_{mt} \]
  \[ + \beta_{j4} z_{t-1} \text{Size}_{jt-1} r_{mt} \]
  \[ + \beta_{j5} \text{BM}_{jt-1} r_{mt} + \beta_{j6} z_{t-1} \text{BM}_{jt-1} r_{mt} + u_{jt}, \]
  where \( r_{jt} = R_{jt} - R_{Ft} \) and \( r_{mt} \) is excess return on the value-weighted market index.

- Then, \( R_{jt}^\star \) in \( (R_{jt}^\star = c_{0t} + \sum_{m=1}^{M} c_m z_{mt} + e_{jt}) \), the dependent variable in the cross-section regression, is given by \( \alpha_j + u_{jt} \).

- The time series regression \( (r_{jt} = \alpha_j + \beta_{j1} r_{mt} + \beta_{j2} z_{t-1} r_{mt} + \beta_{j3} \text{Size}_{jt-1} r_{mt} \]
  \[ + \beta_{j4} z_{t-1} \text{Size}_{jt-1} r_{mt} \]
  \[ + \beta_{j5} \text{BM}_{jt-1} r_{mt} + \beta_{j6} z_{t-1} \text{BM}_{jt-1} r_{mt} + u_{jt}, \]
  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.
Other Procedures

- 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.
- 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.
Understanding GMM: Econometrics Setup and Applications
We can test theories in financial economics 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 $x_1, x_2, \ldots, x_T$.

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 homoscedastic 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 ideal.
- 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 deliver 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

\[ E(f_t(\hat{\Theta})) = 0. \]
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 (just a linear combination of the moment conditions as shown 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 $E[f_t(\Theta)]$ as

$$g_T(\Theta) = \frac{1}{T} \sum_{t=1}^{T} f_t (\Theta).$$

If $r = m$, then the GMM estimator $\hat{\Theta}$ solves

$$g_T (\hat{\Theta}) = \frac{1}{T} \sum_{t=1}^{T} f_t (\hat{\Theta}) = 0.$$

Otherwise, the GMM estimator minimizes the quadratic form

$$J_T(\Theta) = g_T(\Theta)' W_T g_T(\Theta).$$

Here, $W_T$ is some $r \times r$ weighing matrix to be discussed later.
Differentiating with respect to $\Theta$ yields

$$D_T(\Theta)'W_Tg_T(\Theta),$$

The GMM estimator $\hat{\Theta}$ solves

$$D_T(\hat{\Theta})'W_Tg_T(\hat{\Theta}) = 0.$$

Observe 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

$$V = (D_0'WD_0)^{-1}D_0'WSWD_0(D_0'WD_0)^{-1},$$

$$S = \lim_{T \to \infty} V ar[\sqrt{T}g_T(\Theta)],$$

$$= \sum_{j=-\infty}^{\infty} E[f_t(\Theta)f_{t-j}(\Theta)'],$$

$$D_0 = E\left[\frac{\partial g_T(\Theta)}{\partial \theta'}\right] = \frac{1}{T} \sum_{t=1}^{T} E\left[\frac{\partial f_t(\Theta)}{\partial \theta'}\right].$$

To implement the GMM one would like to replace $S$ with its sample estimate.
If the moment conditions are serially uncorrelated then
\[ S_T = \frac{1}{T} \sum_{t=1}^{T} f_t(\Theta)f_t(\Theta)'. \]

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
\[ V^* = (D_0'S^{-1}D_0)^{-1}. \]

Moreover, if \( W = S^{-1} \), i.e., if the weighting matrix is chosen optimally, then an over identifying test statistic is given by
\[ T J_T(\hat{\Theta}) \sim \chi^2_{r-m}. \]

This statistic is quite intuitive.

In particular, note that \( S_T = TVar[g_T(\Theta)] \).

Thus, the test statistic can be expressed as the minimized value of the model errors (in asset pricing context pricing errors) weighted by their covariance matrix
\[ g_T(\Theta)'\{var[g_T(\Theta)]\}^{-1}g_T(\Theta) \sim \chi^2_{r-m}. \]
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 $x_1, x_2, ..., x_T$ and want to estimate the sample mean.
- In this case there is a single parameter $\Theta = \mu$ and a single moment condition

$$f_t(\Theta) = (x_t - \mu).$$

- The system is exactly identified.

- Notice that

$$g_T(\Theta) = \frac{1}{T} \sum_{t=1}^{T} (x_t - \mu).$$

- Setting

$$g_T(\bar{\Theta}) = 0.$$

- Then the GMM estimate for $\mu$ is the sample mean

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} x_t.$$
Moreover, if $x_t$’s are uncorrelated then

$$S = E[f_t(\Theta)f_t(\Theta')]$$

estimated using

$$S_T = \frac{1}{T} \sum_{t=1}^{T} f_t(\hat{\Theta})f_t(\hat{\Theta}') = \frac{1}{T} \sum_{t=1}^{T} (x_t - \hat{\mu})^2$$

To compute the variance of the estimate we need to find $D_0$:

$$D_0 = E\left[\frac{\partial g_T(\theta')}{\partial \theta}\right] = \frac{1}{T} \sum_{t=1}^{T} E\left[\frac{\partial f_t(\theta)}{\partial \theta'}\right] = -1.$$

The optimal $V$ matrix is then given by

$$V = (D_0'S^{-1}D_0)^{-1} = S.$$

Use $S_T$ as a consistent estimator.

Asymptotically, we get

$$\hat{\mu} \sim N\left(\mu, \frac{1}{T}V\right).$$
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, while the follow up expands the analysis to accommodate multiple assets.
- Here is the market model for security $i$

$$r_{i,t} = \alpha_i + \beta_i r_{m,t} + \epsilon_{i,t}.$$ 

- There are two parameters: $\Theta = [\alpha_i, \beta_i]'$.
- There are also two moment conditions

$$f_t(\Theta) = \begin{bmatrix} r_{i,t} - \alpha_i - \beta_i r_{m,t} \\ (r_{i,t} - \alpha_i - \beta_i r_{m,t}) r_{m,t} \end{bmatrix}.$$ 

- Let us rewrite the moment conditions compactly using the following form

$$f_t(\Theta) = x_t (r_t - x_t' \beta),$$

where

$$r_t = r_{i,t},$$

$$x_t = [1, r_{m,t}]',$$

$$\epsilon_t = \epsilon_{i,t},$$ and $\beta = [\alpha_i, \beta_i]'$. 


Let us now compute the sample moment

\[ g_T(\Theta) = \frac{1}{T} \sum_{t=1}^{T} x_t (r_t - x_t' \beta). \]

Since the system is exactly identified setting \( g_T(\hat{\Theta}) = 0 \) yields the GMM estimate

\[ \hat{\beta} = \left( \sum_{t=1}^{T} x_t x_t' \right)^{-1} \left( \sum_{t=1}^{T} x_t r_t \right), \]

\[ = (X'X)^{-1}X'R, \]

where \( X = [x_1, x_2, \ldots, x_T]' \) and \( R = [r_1, r_2, \ldots, r_T]' \).

The GMM estimator for \( \beta \) is the usual OLS estimator.

To find \( V \) first compute

\( \frac{\partial f_t(\Theta)}{\partial \theta'} = -x_t x_t' \),

\[ \frac{\partial g_T(\Theta)}{\partial \theta'} = D_T(\Theta) = -\frac{1}{T} \sum_{t=1}^{T} x_t x_t' = -\frac{X'X}{T}. \]
Moreover, if $\epsilon$’s are serially uncorrelated then

$$S_T = \frac{1}{T} \sum_{t=1}^{T} f_t (\hat{\Theta}) f_t (\hat{\Theta})',$$

$$= \frac{1}{T} \sum_{t=1}^{T} x_t \hat{\epsilon}_t \hat{\epsilon}_t' x_t',$$

$$= \frac{1}{T} \sum_{t=1}^{T} x_t x_t' \hat{\epsilon}_t^2.$$

Using the optimal weighting matrix, it follows that

$$Var(\hat{\beta}) = \frac{1}{T} (D_0' S^{-1} D_0)^{-1},$$

estimated by

$$\hat{Var}(\hat{\beta}) = \frac{1}{T} (D_T' S_T^{-1} D_T)^{-1},$$

$$= (X' X)^{-1} (\sum_{t=1}^{T} x_t x_t' \hat{\epsilon}_t^2) (X' X)^{-1}.$$
Here, we derive the CAPM test described on pages 208-210 in Campbell, Lo, and MacKinlay.

The specification that we have is

\[ r_t^e = \alpha + \beta r_{mt}^e + \epsilon_t. \]

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:

\[
 f_t(\Theta) = x_t \otimes \epsilon_t = \begin{bmatrix}
 1 \\
 r_{m,t}^e \\
 \end{bmatrix} \otimes \epsilon_t = \begin{bmatrix}
 \epsilon_t \\
 r_{m,t}^e \epsilon_t \\
 \end{bmatrix},
\]

where \( \epsilon_t = r_t^e - (x_t' \otimes I_N)\Theta \)
2. Compute $D_0$.

$$\frac{\partial f_t(\theta)}{\partial \theta'} = x_t \otimes - (x_t' \otimes I_N),$$

$$= - \begin{bmatrix} 1 & r_{e_{m,t}} \\ r_{e_{m,t}} & r_{e_{m,t}^2} \end{bmatrix} \otimes I_N.$$

Moreover,

$$D_0 = E \left[ \frac{\partial g_T(\theta)}{\partial \theta'} \right]$$

$$= E \left[ \frac{\partial f_t(\theta)}{\partial \theta'} \right]$$

$$= - \begin{bmatrix} 1 & \mu_m \\ \mu_m & \mu_m^2 + \sigma_m^2 \end{bmatrix} \otimes I_N,$$

where $\mu_m = E(r_{e_{m,t}})$ and $\sigma_m^2 = var(r_{e_{m,t}})$.

3. In implementing the GMM, $D_0$ will be replaced by its sample estimate, which amounts to replacing the population moments $\mu_m$ and $\sigma_m^2$ by their sample analogs $\hat{\mu}_m$ and $\hat{\sigma}_m^2$.

4. That is,

$$D_T = - \begin{bmatrix} 1 & \hat{\mu}_m \\ \hat{\mu}_m & \hat{\mu}_m^2 + \hat{\sigma}_m^2 \end{bmatrix} \otimes I_N,$$
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 $g_T(\hat{\Theta}) = 0$

$$\frac{1}{T} \sum_{t=1}^{T} (x_t \otimes \hat{\epsilon}_t) = 0$$

$$\frac{1}{T} \sum_{t=1}^{T} (x_t \otimes [r_t - (x_t' \otimes I_N)\hat{\Theta}]) = 0$$
\[
\frac{1}{T} \sum_{t=1}^{T} (x_t \otimes r_t) = \frac{1}{T} \sum_{t=1}^{T} (x_t \otimes x'_t \otimes I_N) \hat{\Theta}
\]

The GMM estimator is thus given by

\[
\hat{\Theta} = \left[ \begin{array}{c} \frac{1}{\hat{\sigma}_m^2} \hat{\mu}_m \\ \hat{\mu}_m^2 + \hat{\sigma}_m^2 \end{array} \right]^{-1} \otimes I_N \begin{pmatrix} \hat{\mu} \\ \hat{\sigma}_{rm} + \hat{\mu} \hat{\mu}_m \end{pmatrix},
\]

\[
= \left[ \frac{1}{\hat{\sigma}_m^2} \begin{pmatrix} \hat{\mu}_m^2 + \hat{\sigma}_m^2 & -\hat{\mu}_m \\ -\hat{\mu}_m & 1 \end{pmatrix} \otimes I_N \right] \begin{pmatrix} \hat{\mu} \\ \hat{\sigma}_{rm} + \hat{\mu} \hat{\mu}_m \end{pmatrix},
\]

\[
= \begin{pmatrix} \hat{\mu} - \frac{\hat{\sigma}_{rm}}{\hat{\sigma}_m^2} \hat{\mu}_m \\ \frac{\hat{\sigma}_{rm}}{\hat{\sigma}_m^2} \hat{\mu}_m \end{pmatrix} = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix}.
\]

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. Given $S_T$ and $D_T$ compute $V_T$ the sample estimate of the optimal variance matrix

$$V_T = (D_T' S^{-1}_T D_T)^{-1}$$

10. The asymptotic distribution of $\hat{\Theta}$ is given by

$$\hat{\Theta} = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} \sim N \left( \begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \frac{1}{T} V \right),$$

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 $\mathcal{H}_0: \alpha = 0$ the asymptotic distribution of $R \hat{\Theta}$ is given by

$$R \hat{\Theta} \sim N \left( 0, R \left( \frac{1}{T} V \right) R' \right).$$

12. The statistic $J_7$ in CLM is derived using the Wald statistic

$$J_7 = \hat{\Theta}' R' \left( R \left( \frac{1}{T} V_T \right) R' \right)^{-1} R \hat{\Theta},$$

$$= T \hat{\alpha}' (R (D_T' S^{-1}_T D_T)^{-1} R')^{-1} \hat{\alpha}.$$
13. 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 is

$$\frac{1}{T}(X'X) \otimes \Sigma = \begin{bmatrix} 1 \\
\hat{\mu}_m \\
\hat{\mu}_m^2 + \hat{\sigma}_m^2 \end{bmatrix} \otimes \Sigma$$

- Thus $J_7$ becomes the $J1$ Wald statistic derived earlier

- 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

$$E(r_t|z_{t-1}) = \text{cov}(r_t, r_{mt}|z_{t-1})\lambda_t,$$

$$= E[(r_t - E[r_t|z_{t-1}])(r_{mt} - E[r_{mt}|z_{t-1}])|z_{t-1}]\lambda_t,$$

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,

\[
\mathbb{E}[r_t|z_{t-1}] = \delta_r x_{t-1},
\]

\[
\mathbb{E}[r_{mt}|z_{t-1}] = \delta_m x_{t-1}.
\]

Then, let us define several residuals

\[
\begin{align*}
u_{rt} &= r_t - \delta_r x_{t-1}, \\
u_{mt} &= r_{mt} - \delta_m x_{t-1},
\end{align*}
\]

\[
e_t = r_t - [(r_t - \mathbb{E}[r_t|z_{t-1}]) (r_{mt} - \mathbb{E}[r_{mt}|z_{t-1}])] z_{t-1} \lambda \\
&= r_t - (r_t - \delta_r x_{t-1}) (r_{mt} - \delta_m x_{t-1}) \lambda.
\]

Collecting the residuals into one vector yields

\[
f_t(\Theta) = [u_{rt}', u_{mt}', e_t']',
\]

where $\Theta = [vec(\delta_r)', \delta_m', \lambda]'$.

That is, there are $(M + 1)(N + 1) + 1$ 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

\[ \mathbb{E}[f_t(\Theta)] = 0, \]

as well as \(M(2N + 1)\) additional moment conditions involving the instruments

\[ \mathbb{E}[f_t(\Theta) \otimes z_{t-1}] = 0. \]

Overall, there are \((2N + 1)(M + 1) + N\) 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: Estimating Standard Errors in the presence of correlation among firms

- This application builds on Avramov, Chordia, and Goyal (2006b).

- Consider $N$ stocks with $T$ observations.

- The dependent variable is denoted by $y$ and the set of $K$ independent variables by $x$.

- For instance, $y$ could denote volatility and $x$ could include lags of volatility, day of week dummies, and trading-related variables.

- The regression equation is as follows:
  \[ y_{it} = x_{it}' \beta_i + \epsilon_{it} \]

- Let us introduce some notation. Let $\beta = (\beta_1', ..., \beta_N')'$, let $x_t = (x_{1t}', ..., x_{Nt}')'$, $X_t = (x_{i1}, ..., x_{iT})'$, and let $Y_t = (y_{i1}, ..., y_{iT})'$.

- Moment conditions are written as $E(f(x_t, \beta)) = 0$, where $f(x_t, \beta)$ is an $NK$ valued function given by

\[
 f(x_t, \beta) = \left( \begin{array}{c} x_{1t} \epsilon_{1t} \\ \vdots \\ x_{Nt} \epsilon_{Nt} \end{array} \right) = \left( \begin{array}{c} x_{1t} (y_{1t} - x_{1t}' \beta_1) \\ \vdots \\ x_{Nt} (y_{Nt} - x_{Nt}' \beta_N) \end{array} \right)
\]
Calculation of Standard Errors

- Since the number of moment conditions is exactly equal to the number of parameters, the system is exactly identified.
- Thus, we do not need the usual weighting matrix to carry out the optimization.
- The solution is, of course, given by the usual OLS
  \[ \hat{\beta}_i = (X_i'X_i)^{-1}X_i'Y_i \]
- The variance of the estimator is given from the GMM formula as
  \[ T \times \text{cov}(\hat{\beta}) = D^{-1}SD^{-1}' \]
  where the score matrix \( D \) and the spectral density matrix \( S \) are given by
  \[ D = E \left( \frac{\partial f(x_t, \beta)}{\partial \beta'} \right) \]
  \[ S = \sum_{s=-\infty}^{\infty} E[f(x_t, \beta)f(x_{t-s}, \beta)'] \]
Calculation of Standard Errors

- The special structure of $f$ and $\beta$ makes the computation of $D$ especially straightforward.
- In particular, we have

$$\frac{\partial f(x_t, \beta)}{\partial \beta'} = \begin{bmatrix} -x_{1t}x_{1t}' & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & -x_{Nt}x_{Nt}' \end{bmatrix}$$

- Thus,

$$\hat{D} = \hat{E} \left( \frac{\partial f(x_t, \beta)}{\partial \beta'} \right) = \frac{1}{T} \begin{bmatrix} -X_1'X_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & -X_N'X_N \end{bmatrix}$$

where $0$ is suitably defined $K \times K$ matrices.
- Thus, the matrix $D$ consists of $K^2$ blocks with the $(i, j)$th block equal to zero if $i \neq j$ and equal to $X_i'X_i$ when $i = j$.
- The diagonal nature of this matrix also makes the computation of the inverse of $D$ trivial.
Calculation of Standard Errors

- The matrix $S$ consists of $K^2$ blocks, where the $(i,j)$th block is given by

$$S_{ij} = \sum_{s=-\infty}^{\infty} E[x_{it}x_{jt-s}\epsilon_{it}\epsilon_{jt-s}]$$

and is estimated using the Newey–West estimation technique with $L$ lags as

$$\hat{S}_{ij} = \frac{1}{T} \left[ \sum_{t=1}^{T} (x_{it}x'_{jt}\epsilon_{it}\epsilon'_{jt}) + \sum_{s=1}^{L} \frac{L-s}{L} \sum_{t=s+1}^{T} (x_{it}x'_{jt-s}\epsilon_{it}\epsilon'_{jt-s} + x_{jt}x'_{it-s}\epsilon_{jt}\epsilon'_{it-s}) \right]$$

- Combining all the pieces together, we get

$$\text{cov}(\hat{\beta}_i, \hat{\beta}_j) = (X'_iX_i)^{-1}TS_{ij}(X'_jX_j)^{-1}$$

- Finally, can consider the average $\beta$ coefficient across all stocks.

- This is estimated as

$$\hat{\beta} = \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_i$$

with the variance given by

$$\text{cov} \left( \hat{\beta} \right) = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \text{cov}(\hat{\beta}_i, \hat{\beta}_j)$$
Application # 6: 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

\[ R_{t+k} = \alpha + \beta' z_t + \epsilon_{t+k} \]

where \( R_{t+k} = \sum_{i=1}^{k} r_{t+i} \), 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)
\[
\sqrt{T}(\Theta - \overline{\Theta}) \sim N(0, V),
\]
where
\[
V = (D_0' S^{-1} D_0)^{-1}.
\]
There are $M + 1$ moment conditions:
\[
f_t(\Theta) = \left[ \frac{R_{t+k} - \alpha - \beta' z_t}{(R_{t+k} - \alpha - \beta' z_t) z_t} \right].
\]
Compute $D_0$:
\[
D_0 = \mathbb{E} \left[ \frac{\partial f_t(\Theta)}{\partial \theta'} \right] = \left[ \begin{array}{cc}
-1 & -\mu_z \\
-\mu_z' & -(\Sigma_z + \mu_z \mu_z')
\end{array} \right].
\]
It would be useful to take the inverse of $D_0$
\[
D_0^{-1} = \left[ \begin{array}{cc}
-(1 + \mu_z' \Sigma_z^{-1} \mu_z) & \mu_z' \Sigma_z^{-1} \\
\Sigma_z^{-1} \mu_z & -\Sigma_z^{-1}
\end{array} \right].
\]
And the matrix $S$ is given by
\[
S = \sum_{j=-\infty}^{\infty} \mathbb{E} \left[ \begin{array}{ccc}
\epsilon_{t+k} \epsilon_{t+k-j} & \epsilon_{t+k} \epsilon_{t+k-j} z'_{t-j} \\
\epsilon_{t+k} \epsilon_{t+k-j} z_t & \epsilon_{t+k} \epsilon_{t+k-j} z_t z'_{t-j}
\end{array} \right].
\]
In estimating predictive regressions we scrutinize the slope coefficients only.

And we know that

\[
\sqrt{T}(\beta - \hat{\beta}) \sim N(0, \tilde{V}),
\]

where \( \tilde{V} \) is the \( M \times M \) lower-right sub-matrix of \( V = D_0^{-1}S D_0^{-1}' \).

It follows that

\[
\tilde{V} = \sum_{j=-\infty}^{\infty} \mathbb{E} \left[ \begin{bmatrix} \mu_z' \Sigma_z^{-1} \\ -\Sigma_z^{-1} \end{bmatrix}' \begin{bmatrix} \epsilon_{t+k}\epsilon_{t+k-j} & \epsilon_{t+k}\epsilon_{t+k-j} Z'_{t-j} \\ \epsilon_{t+k}\epsilon_{t+k-j} Z_{t} & \epsilon_{t+k}\epsilon_{t+k-j} Z_{t} Z'_{t-j} \end{bmatrix} \begin{bmatrix} \mu_z' \Sigma_z^{-1} \\ -\Sigma_z^{-1} \end{bmatrix} \right],
\]

\[
= \Sigma_z^{-1} \left( \sum_{j=-\infty}^{\infty} \mathbb{E} (\epsilon_{t+k}\epsilon_{t+k-j})(z_{t} - \mu_z)(z_{t-j} - \mu_z)' \right) \Sigma_z^{-1},
\]

\[
= \Sigma_z^{-1} \left[ \sum_{j=-\infty}^{\infty} \mathbb{E} (\delta_{t+k}\delta_{t+k-j}') \right] \Sigma_z^{-1},
\]

where

\[
\delta_{t+k} = \epsilon_{t+k} Z_{t},
\]

\[
\delta_{t+k-j} = \epsilon_{t+k-j} Z_{t-j}.
\]
What if you are willing to assume that there is no autocorrelation in the residuals?

Then,

\[
\tilde{V} = \Sigma_z^{-1} \left[ \mathbb{E}(\delta_{t+k} \delta_{t+k}') \right] \Sigma_z^{-1},
\]

which can be estimated by

\[
\tilde{V} = \hat{\Sigma}_z^{-1} \left[ \frac{1}{T} \sum_{t=1}^{T} \delta_{t+k} \delta_{t+k}' \right] \hat{\Sigma}_z^{-1}.
\]

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 heteroscedasticity?

\[
\tilde{V} = \Sigma_z^{-1} \left[ \mathbb{E}(\epsilon_{t+k} \epsilon_{t+k-j}) \mathbb{E}(z_t - \mu_z)(z_{t-j} - \mu_z)' \right] \Sigma_z^{-1},
\]

In that case

\[
\sqrt{T}(\hat{\beta} - \beta) \sim N(0, \sigma_{\epsilon}^2 \Sigma_z^{-1}).
\]

Under the null hypothesis that \( \beta = 0 \)

\[
\sqrt{T}\hat{\beta} \sim N(0, \sigma_{\epsilon}^2 \Sigma_z^{-1}).
\]
Note also that under the null $\sigma_e^2 = \sigma_r^2$, where $\sigma_r^2$ is the variance of the cumulative log return.

Using properties of the $\chi^2$ distribution, it follows that

$$T \frac{\hat{\beta}' \Sigma_z \hat{\beta}}{\hat{\sigma}_r^2} \sim \chi^2(M),$$

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 heteroscedasticity 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.

That is almost a no brainer.
Application # 7: Yet, another Testing return predictability over long horizon

- Boudoukh, Richardson, and Whitelaw (2006) is another interesting application of the GMM in the context of long-horizon return predictability
- Like Kirby (1997), they show that the sample evidence does not support predictability
- Further, they show that for persistent predictors the estimates of slope coefficients are almost perfectly correlated across horizons under the null hypothesis of no predictability
- They consider regression systems of the following type:
  
  $$R_{t,t+1} = \alpha_1 + \beta_1 Z_t + \epsilon_{t,t+1}$$
  
  $$\vdots$$
  
  $$R_{t,t+j} = \alpha_j + \beta_j Z_t + \epsilon_{t,t+j}$$
  
  $$\vdots$$
  
  $$R_{t,t+K} = \alpha_K + \beta_K Z_t + \epsilon_{t,t+K}$$

- Under the null hypothesis of no predictability
  
  $$\beta_1 = \cdots = \beta_j = \cdots = \beta_K = 0$$
And the same number of moment conditions corresponding to the regression system as

\[
f_t(\theta) = \begin{bmatrix}
(R_{t,t+1} - \alpha_1 - \beta_1 Z_t) \\
(R_{t,t+1} - \alpha_1 - \beta_1 Z_t)Z_t \\
\vdots \\
(R_{t,t+j} - j\alpha_1 - \beta_j Z_t)Z_t \\
\vdots \\
(R_{t,t+K} - K\alpha_1 - \beta_K Z_t)Z_t
\end{bmatrix}
\]

where \( \theta = (\alpha_1, \beta_1, \ldots, \beta_k)' \)

- There are \( k+1 \) parameters and moment conditions
- Under the null, the regression estimate \( \theta \) has an asymptotic normal distribution with mean \( (\alpha_1, 0)' \) and covariance matrix \( (D'_0 S_0^{-1} D_0)^{-1} \) where \( D_0 = E\left(\frac{\partial f_t}{\partial \theta}\right) \) and \( S_0 = \sum_{j=\infty}^{\infty} E(f_t f_{t-j}) \)
- \( D_0 \) is easily calculated as

\[
D_0 = -\begin{pmatrix}
1 & \mu_Z & 0 & \cdots & \cdots & \cdots \\
\mu_Z & \mu_Z^2 + \sigma_Z^2 & 0 & \cdots & \cdots & \cdots \\
\vdots & 0 & \mu_Z^2 + \sigma_Z^2 & 0 & \cdots & \cdots \\
j\mu_Z & \cdots & 0 & \mu_Z^2 + \sigma_Z^2 & 0 & \cdots \\
j\mu_Z & \cdots & \cdots & 0 & \mu_Z^2 + \sigma_Z^2
\end{pmatrix}
\]

Under the assumption of homoscedasticity and AR(1) for the predictor.
Long story short, for the typical 1- through 5-year horizons and for $\rho = 0.953$ the covariance matrix of slope estimates under the null of no predictability is given by

$$S_0 = \begin{bmatrix}
\sigma_R^2 & \sigma_R^2 \mu_Z & \cdots & j\sigma_R^2 \mu_Z & \cdots & k\sigma_R^2 \mu_Z \\
\sigma_R^2 \mu_Z & \sigma_R^2 \left(j\mu_Z^2 + \sigma_Z^2 \right) & \cdots & \sigma_R^2 \left[j\mu_Z^2 + \sigma_Z^2 \left[1 + \Sigma_{l=1}^{j-1} \rho_l \right] \right] & \cdots & \sigma_R^2 \left(k\mu_Z^2 + \sigma_Z^2 \left[1 + \Sigma_{l=1}^{j-1} \rho_l \right] \right) \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
j\sigma_R^2 \mu_Z & \sigma_R^2 \left(j\mu_Z^2 + \sigma_Z^2 \left[1 + \Sigma_{l=1}^{j-1} \rho_l \right] \right) & \cdots & \sigma_R^2 \left[j\mu_Z^2 + \sigma_Z^2 \left[1 + \Sigma_{l=1}^{j-1} \rho_l \right] \right] & \cdots & \sigma_R^2 \left(k\mu_Z^2 + \sigma_Z^2 \left[1 + \Sigma_{l=1}^{j-1} \rho_l \right] \right) \\
k\sigma_R^2 \mu_Z & \sigma_R^2 \left(k\mu_Z^2 + \sigma_Z^2 \left[1 + \Sigma_{l=1}^{j-1} \rho_l \right] \right) & \cdots & \sigma_R^2 \left[k\mu_Z^2 + \sigma_Z^2 \left[1 + \Sigma_{l=1}^{j-1} \rho_l \right] \right] & \cdots & \sigma_R^2 \left(k\mu_Z^2 + \sigma_Z^2 \left[1 + \Sigma_{l=1}^{j-1} \rho_l \right] \right)
\end{bmatrix}$$

$$T\text{cov}(\hat{\beta}_1, \ldots, \hat{\beta}_5) = \frac{\sigma_R^2}{\sigma_Z^2} \begin{pmatrix}
1 & 0.988 & 0.974 & 0.959 & 0.945 \\
1 & 0.993 & 0.982 & 0.970 \\
1 & 0.995 & 0.986 \\
1 & 0.996 \\
1
\end{pmatrix}$$
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

$$S = \sum_{j=-\infty}^{\infty} E(u_t u_{t-j}).$$

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.
The covariance matrix of the slope estimates

- Recall the regression estimates based on either the GMM or OLS are given by
  \[ \hat{\beta} = \beta + (X'X)^{-1}X'U \]

- Considering the base case of no serial correlation (SC) and no heteroscedasticity (HS), we have
  \[ \text{Var}(\hat{\beta}|X) = (X'X)^{-1}X' \text{Var}[U|X] X(X'X)^{-1} \]
  \[ \text{Var}[U|X] = \sigma^2 I_T \]

- And the variance is estimated by:
  \[ \hat{\text{Var}}(\hat{\beta}|X) = (X'X)^{-1}\hat{\sigma}^2 \]
  \[ \hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} \hat{u}_t^2 \]

- Now let us assume the presence of HS
  \[ \hat{\text{Var}}(\hat{\beta}|X) = (X'X)^{-1}\hat{\text{Var}}[X'U|X] (X'X)^{-1} \]
  \[ \hat{\text{Var}}[X'U|X] = \sum_{t=1}^{T} x_t x_t' \hat{u}_t^2 \]
  \[ \hat{\text{Var}}(\hat{\beta}|X) = (X'X)^{-1} \left( \sum_{t=1}^{T} x_t x_t' \hat{u}_t^2 \right) (X'X)^{-1} \]
Let us assume now that there is serial correlation only

In particular the error term obeys the $AR(1)$ process

$$y_t = x_t' \beta + u_t$$

$$u_t = \rho u_{t-1} + e_t$$

where

$$Var(e_t) = \sigma^2 \quad \forall \ t$$

Therefore

$$Var(u_t) = \frac{\sigma^2}{1-\rho^2} \quad \forall \ t$$

Then we get:

$$Var(\hat{\beta}) = [(X'X)^{-1}X'E(UU')X(X'X)^{-1}]$$

$$E(UU') = E\left[ \begin{bmatrix} u_1^2 & u_1u_2 & \cdots & u_1u_T \\ u_1u_2 & u_2^2 & \cdots & u_2u_T \\ \vdots & \vdots & \ddots & \vdots \\ u_1u_T & \cdots & \cdots & u_T^2 \end{bmatrix} \right]$$
\[
\begin{bmatrix}
\sigma^2 & \rho \sigma^2 & \ldots & \\
\frac{1 - \rho^2}{\rho \sigma^2} & \frac{\rho}{1 - \rho^2} & \frac{\rho \sigma^2}{1 - \rho^2} & \ldots \\
\vdots & \vdots & \ddots & \\
\rho^{T-1} \sigma^2 & \cdots & \cdots & \\
\frac{1 - \rho^2}{\rho \sigma^2} & \\
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\gamma(0) & \gamma(1) & \gamma(2) & \ldots & \gamma(T-1) \\
\gamma(1) & \gamma(0) & \gamma(1) & \ldots & \gamma(T-2) \\
\vdots & \vdots & \ddots & \\
\gamma(T-1) & \gamma(T-2) & \ldots & \gamma(0) \\
\end{bmatrix}
\]

\[
= \frac{\sigma^2}{1 - \rho^2} \begin{bmatrix}
1 & \rho & \rho^2 & \ldots & \rho^{T-1} \\
\rho & 1 & \rho & \ldots & \rho^{T-2} \\
\vdots & \vdots & \ddots & \\
\rho^{T-1} & \rho^{T-2} & \ldots & 1 \\
\end{bmatrix}
\]
\[
\begin{align*}
\sigma^2 & = \frac{\sigma^2}{1 - \rho^2} I_T + \frac{\sigma^2}{1 - \rho^2} \begin{bmatrix}
0 & \rho & \rho^2 & \cdots & \rho^{T-1} \\
\rho & 0 & \rho & \cdots & \rho^{T-2} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\rho^{T-1} & \rho^{T-2} & \cdots & \cdots & 0
\end{bmatrix} \\
\end{align*}
\]

- Let

\[
\bar{\rho} \equiv \begin{bmatrix}
0 & \rho & \rho^2 & \cdots & \rho^{T-1} \\
\rho & 0 & \rho & \cdots & \rho^{T-2} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\rho^{T-1} & \rho^{T-2} & \cdots & \cdots & 0
\end{bmatrix}
\]

- The co-variance matrix is eventually given by

\[
\text{Var} (\hat{\beta}) = (X'X)^{-1} \frac{\sigma^2}{1 - \rho^2} + (X'X)^{-1} X' \left( \frac{\sigma^2}{1 - \rho^2 \bar{\rho}} \right) X (X'X)^{-1}
\]
Notice that with zero autocorrelation we are back with the base-case covariance matrix.

Assume now that the error term obeys the $AR(2)$ process

$$u_t = \phi_1 u_{t-1} + \phi_2 u_{t-2} + e_t$$

$$\text{Var}(e_t) = \sigma^2 \quad \forall \ t$$

$$E(UU') = \begin{bmatrix}
\gamma(0) & \gamma(1) & \gamma(2) & \ldots & \gamma(T - 1) \\
\gamma(1) & \gamma(0) & \gamma(1) & \ldots & \gamma(T - 2) \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\gamma(T - 1) & \gamma(T - 2) & \ldots & \gamma(0)
\end{bmatrix}$$

- \(\gamma(0) = \frac{(1 - \phi_2)\sigma^2}{(1 + \phi_2)(1 - \phi_1 - \phi_2)(1 + \phi_1 - \phi_2)}\)

- \(\gamma(1) = \gamma(0) \frac{\phi_1}{1 - \phi_2}\)

- \(\gamma(2) = \gamma(0) \left(\frac{\phi_1^2}{1 - \phi_2} + \phi_2\right)\)
Newey and West (1987)

- If both HS and SC, Andrews (1991) develops a complex, albeit useful, estimator.
- We focus here on Newey and West (1987) who propose the following covariance matrix

\[
\hat{\text{Var}}[X'U|X] = \sum_{t=1}^{T} x_t x'_t \hat{u}_t^2 + \sum_{j=1}^{k} \frac{k-j}{k} \sum_{t=j+1}^{T} \hat{u}_t \hat{u}_{t-j} (x_t x'_{t-j} + x_{t-j} x'_t)
\]

- E.g.,
  - \( k=1 \)
    \[
    \hat{\text{Var}}[X'U|X] = \sum_{t=1}^{T} x_t x'_t \hat{u}_t^2
    \]
  - \( k=2 \)
    \[
    \hat{\text{Var}}[X'U|X] = \sum_{t=1}^{T} x_t x'_t \hat{u}_t^2 + \frac{1}{2} \sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1} (x_t x'_{t-1} + x_{t-1} x'_t)
    \]
  - \( k=3 \)
    \[
    \hat{\text{Var}}[X'U|X] = \sum_{t=1}^{T} x_t x'_t \hat{u}_t^2 + \frac{2}{3} \sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1} (x_t x'_{t-1} + x_{t-1} x'_t) + \frac{1}{3} \sum_{t=3}^{T} \hat{u}_t \hat{u}_{t-2} (x_t x'_{t-2} + x_{t-2} x'_t)
    \]
Bayesian Econometrics
Bayes Rule

- Let $x$ and $y$ be two random variables.
- Let $P(x)$ and $P(y)$ be the two *marginal* probability distribution functions of $x$ and $y$.
- Let $P(x|y)$ and $P(y|x)$ denote the corresponding *conditional* pdfs.
- Let $P(x, y)$ denote the *joint* pdf of $x$ and $y$.
- It is known from the law of total probability that the joint pdf can be decomposed as
  \[ P(x, y) = P(x)P(y|x) = P(y)P(x|y) \]
- Therefore
  \[ P(y|x) = \frac{P(y)P(x|y)}{P(x)} = cP(y)P(x|y) \]
  where $c$ is the constant of integration (see next page).
- The Bayes Rule is described by the following proportion
  \[ P(y|x) \propto P(y)P(x|y) \]
Bayes Rule

- Notice that the right hand side retains only factors related to $y$, thereby excluding $P(x)$

- $P(x)$, termed the marginal likelihood function, is

$$P(x) = \int P(y)P(x|y)dy$$

$$= \int P(x, y)dy$$

as the conditional distribution $P(y|x)$ integrates to unity.

- The marginal likelihood $P(x)$ is an essential ingredient in computing an important quantity - model posterior probability.

- Notice from the second equation above that the marginal likelihood obtains by integrating out $y$ from the joint density $P(x, y)$.

- Similarly, if the joint distribution is $P(x, y, z)$ and the pdf of interest is $P(x, y)$ one integrates $P(x, y, z)$ with respect to $z$. 
Bayes Rule

- The essence of Bayesian econometrics is the Bayes Rule.

- Ingredients of Bayesian econometrics are parameters underlying a given model, the sample data, the prior density of the parameters, the likelihood function describing the data, and the posterior distribution of the parameters.

- A predictive distribution could also be involved.

- In the Bayesian setup, parameters are stochastic while in the classical (non Bayesian) approach parameters are unknown constants.

- Decision making is based on the posterior distribution of the parameters or the predictive distribution of next period quantities as described below.

- On the basis of the Bayes rule, in what follows, $y$ stands for unknown stochastic parameters, $x$ for the data, $P(y|x)$ for the posterior distribution, $P(y)$ for the prior, and $P(x|y)$ for the likelihood.

- The Bayes rule describes the relation between the prior, the likelihood, and the posterior, or put differently it shows how prior beliefs are updated to produce posterior beliefs:

$$P(y|x) \propto P(y)P(x|y)$$

- Zellner (1971) is an excellent source of reference.
Bayes Econometrics in Financial Economics

- You observe the returns on the market index over $T$ months: $r_1, ..., r_T$
- Let $R: [r_1, ..., r_T]'$ denote the $T \times 1$ vector of all return realizations
- Assume that $r_t \sim N(\mu, \sigma_0^2)$ for $t = 1, ..., T$
  
  where
  
  $\mu$ is a stochastic random variable denoting the mean return
  $\sigma_0^2$ is the variance which, at this stage, is assumed to be a known constant
  and returns are IID (independently and identically distributed) through time.

- By Bayes rule
  
  $$P(\mu|R, \sigma_0^2) \propto P(\mu)P(R|\mu, \sigma_0^2)$$
  
  where
  
  $P(\mu|R, \sigma_0^2)$ is the posterior distribution of $\mu$
  $P(\mu)$ is the prior distribution of $\mu$
  and $P(R|\mu, \sigma_0^2)$ is the joint likelihood of all return realizations.
The likelihood function of a normally distributed return realization is given by

\[
P(r_t | \mu, \sigma_0^2) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp \left( -\frac{1}{2\sigma_0^2} (r_t - \mu)^2 \right)
\]

- Since returns are assumed to be IID, the joint likelihood of all realized returns is

\[
P(R | \mu, \sigma_0^2) = (2\pi \sigma_0^2)^{-\frac{T}{2}} \exp \left( -\frac{1}{2\sigma_0^2} \sum_{t=1}^{T} (r_t - \mu)^2 \right)
\]

- Notice:

\[
\sum (r_t - \mu)^2 = \sum [(r_t - \hat{\mu}) + (\hat{\mu} - \mu)]^2
\]

\[
= vs^2 + T(\mu - \hat{\mu})^2
\]

since the cross product is zero, and

\[
v = T - 1
\]

\[
s^2 = \frac{1}{T - 1} \sum (r_t - \hat{\mu})^2
\]

\[
\hat{\mu} = \frac{1}{T} \sum r_t
\]
The prior is specified by the researcher based on economic theory, past experience, past data, current similar data, etc. Often, the prior is diffuse or non-informative.

For the next illustration, it is assumed that $P(\mu) \propto c$, that is, the prior is diffuse, non-informative, in that it apparently conveys no information on the parameters of interest.

I emphasize “apparently” since innocent diffuse priors could exert substantial amount of information about quantities of interest which are non-linear functions of the parameters.

Informative priors with sound economic appeal are well perceived in financial economics.

For instance, Kandel and Stambaugh (1996), who study asset allocation when stock returns are predictable, entertain informative prior beliefs weighted against predictability. Pastor and Stambaugh (1999) introduce prior beliefs about expected stock returns which consider factor model restrictions. Avramov, Cederburg, and Kvasnakova (2017) study prior beliefs about predictive regression parameters which are disciplined by consumption based asset pricing models including habit formation, prospect theory, and long run risk.

Computing posterior probabilities (as opposed to posterior densities) of competing models (e.g., Avramov (2002)) necessitates the use of informative priors. Diffuse priors won’t fit.
The Posterior Distribution of Mean Return

- With diffuse prior and normal likelihood, the posterior is proportional to the likelihood function:

\[
P(\mu|R, \sigma_0^2) \propto \exp \left( -\frac{1}{2\sigma_0^2} [v s^2 + T (\mu - \hat{\mu})]^2 \right)
\]

\[
\propto \exp \left( -\frac{T}{2\sigma_0^2} (\mu - \hat{\mu})^2 \right)
\]

- The bottom relation follows since only factors related to \( \mu \) are retained

- The posterior distribution of the mean return is given by

\[
\mu|R, \sigma_0^2 \sim N(\hat{\mu}, \sigma_0^2 / T)
\]

- In classical econometrics:

\[
\hat{\mu}|R, \sigma_0^2 \sim N(\mu, \sigma_0^2 / T)
\]

- That is, in classical econometrics, the sample estimate of \( \mu \) is stochastic while \( \mu \) itself is an unknown non-stochastic parameter.
Informative Prior

- The prior on the mean return is often modeled as

\[
P(\mu) \propto (\sigma_a)^{-\frac{1}{2}} \exp \left( -\frac{1}{2\sigma_a^2} (\mu - \mu_a)^2 \right)
\]

where \( \mu_a \) and \( \sigma_a \) are prior parameters to be specified by the researcher.

- The posterior obtains by combining the prior and the likelihood:

\[
P(\mu|R, \sigma_0^2) \propto P(\mu)P(R|\mu, \sigma_0^2)
\]

\[
\propto \exp \left( -\frac{(\mu - \mu_a)^2}{2\sigma_a^2} - \frac{T(\mu - \bar{\mu})^2}{2\sigma_0^2} \right)
\]

\[
\propto \exp \left( -\frac{1}{2} \frac{(\mu - \bar{\mu})^2}{\sigma^2} \right)
\]

- The bottom relation obtains by completing the square on \( \mu \)

- Notice, in particular,

\[
\frac{\mu^2}{\sigma_a^2} + \frac{T}{\sigma_0^2} \mu^2 = \frac{\mu^2}{\hat{\sigma}^2}
\]

\[
\left( \frac{1}{\sigma_a^2} + \frac{T}{\sigma_0^2} \right) = \frac{1}{\hat{\sigma}^2}
\]
The Posterior Mean

- Hence, the posterior variance of the mean is

\[ \tilde{\sigma}^2 = \left[ \frac{1}{\sigma_a^2} + \frac{1}{\sigma_0^2/T} \right]^{-1} = (\text{prior precision} + \text{likelihood precision})^{-1} \]

- Similarly, the posterior mean of \( \mu \) is

\[ \tilde{\mu} = \tilde{\sigma}^2 \left[ \frac{\mu_a}{\sigma_a^2} + \frac{T\hat{\mu}}{\sigma_0^2} \right] = w_1 \mu_a + w_2 \hat{\mu} \]

where

\[ w_1 = \frac{1/\sigma_a^2}{1/\sigma_a^2 + \frac{1}{\sigma_0^2/T}} = \frac{\text{prior precision}}{\text{prior precision} + \text{likelihood precision}} \]

\[ w_2 = 1 - w_1 \]

- Intuitively, the posterior mean of \( \mu \) is the weighted average of the prior mean and the sample mean with weights depending on prior and likelihood precisions, respectively.
What if \( \sigma \) is unknown? – The case of Diffuse Prior

- Bayes: \( P(\mu, \sigma|R) \propto P(\mu, \sigma)P(R|\mu, \sigma) \)
- The non-informative prior is typically modeled as
  \[
  P(\mu, \sigma) \propto P(\mu)P(\sigma) \\
  P(\mu) \propto c \\
  P(\sigma) \propto \sigma^{-1}
  \]
- Thus, the joint posterior of \( \mu \) and \( \sigma \) is
  \[
  P(\mu, \sigma|R) \propto \sigma^{-(T+1)}\exp\left(-\frac{1}{2\sigma^2}[\nu s^2 + T(\mu - \hat{\mu})]^2\right)
  \]
- The conditional distribution of the mean follows straightforwardly
  \[
  P(\mu|\sigma, R) \sim N\left(\hat{\mu}, \frac{\sigma^2}{T}\right)
  \]
- More challenging is to uncover the marginal distributions, which are obtained as
  \[
  P(\mu|R) = \int P(\mu, \sigma|R)d\sigma \\
  P(\sigma|R) = \int P(\mu, \sigma|R)d\mu
  \]
Solving the Integrals: Posterior of $\mu$

- Let $\alpha = \nu s^2 + T(\mu - \hat{\mu})^2$
- Then,

$$P(\mu|R) \propto \int_{\sigma=0}^{\infty} \sigma^{-(T+1)} \exp \left( -\frac{\alpha}{2\sigma^2} \right) d\sigma$$

- We do a change of variable

$$x = \frac{\alpha}{2\sigma^2}$$

$$\frac{d\sigma}{dx} = -2^{-\frac{1}{2}} \alpha^{-\frac{1}{2}} x^{-\frac{1}{2}}$$

$$\sigma^{-T+1} = \left( \frac{\alpha}{2x} \right)^{-\frac{T+1}{2}}$$

- Then

$$P(\mu|R) \propto 2^{\frac{T-2}{2}} \alpha^{-\frac{T}{2}} \int_{x=0}^{\infty} x^{\frac{T}{2}-1} \exp(-x) dx$$

- Notice

$$\int_{x=0}^{\infty} x^{\frac{T}{2}-1} \exp(-x) dx = \Gamma \left( \frac{T}{2} \right)$$

- Therefore,

$$P(\mu|R) \propto 2^{\frac{T-2}{2}} \alpha^{-\frac{T}{2}} \Gamma \left( \frac{T}{2} \right) \alpha^{-\frac{T}{2}}$$

$$\alpha \left[ \nu s^2 + T(\mu - \hat{\mu})^2 \right]^{-\frac{v+1}{2}}$$

- We get $t = \frac{\mu - \hat{\mu}}{s/\sqrt{T}} \sim t(\nu)$, corresponding to the Student $t$ distribution with $\nu$ degrees of freedom.
The Marginal Posterior of $\sigma$

- The posterior on $\sigma$

\[
P(\sigma|R) \propto \int \sigma^{-(T+1)} \exp \left( -\frac{1}{2\sigma^2} [\nu s^2 + T(\mu - \hat{\mu})]^2 \right) d\mu
\]

\[
\propto \sigma^{-(T+1)} \exp \left( -\frac{\nu s^2}{2\sigma^2} \right) \int \exp \left( -\frac{T}{2\sigma^2} (\mu - \hat{\mu})^2 \right) d\mu
\]

- Let $z = \frac{\sqrt{T}(\mu - \hat{\mu})}{\sigma}$, then

\[
\frac{dz}{d\mu} = \frac{\sqrt{T}}{\sigma}
\]

\[
P(\sigma|R) \propto \sigma^{-T} \exp \left( -\frac{\nu s^2}{2\sigma^2} \right) \int \exp \left( -\frac{z^2}{2} \right) dz
\]

\[
\propto \sigma^{-T} \exp \left( -\frac{\nu s^2}{2\sigma^2} \right)
\]

\[
\propto \sigma^{-(v+1)} \exp \left( -\frac{\nu s^2}{2\sigma^2} \right)
\]

which corresponds to the inverted gamma distribution with $v$ degrees of freedom and parameter $s$

- The explicit form (with constant of integration) of the inverted gamma is given by

\[
P(\sigma|v, s) = \frac{2}{\Gamma \left( \frac{v}{2} \right)} \left( \frac{\nu s^2}{2} \right)^{v/2} \sigma^{-(v+1)} \exp \left( -\frac{\nu s^2}{2\sigma^2} \right)
\]
The Multiple Regression Model

- The regression model is given by

\[ y = X\beta + u \]

where

- \( y \) is a \( T \times 1 \) vector of the dependent variables
- \( X \) is a \( T \times M \) matrix with the first column being a \( T \times 1 \) vector of ones
- \( \beta \) is an \( M \times 1 \) vector containing the intercept and \( M-1 \) slope coefficients
- \( u \) is a \( T \times 1 \) vector of residuals.

- We assume that \( u_t \sim N(0, \sigma^2) \) \( \forall \ t = 1, \ldots, T \) and IID through time

- The likelihood function is

\[
P(y|X, \beta, \sigma) \propto \sigma^{-T} \exp \left( -\frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta) \right) \\
\propto \sigma^{-T} \exp \left\{ -\frac{1}{2\sigma^2} \left[ vs^2 + (\beta - \hat{\beta})'X'X(\beta - \hat{\beta}) \right] \right\}
\]
The Multiple Regression Model

where

\[ \nu = T - M \]
\[ \hat{\beta} = (X'X)^{-1}X'y \]
\[ s^2 = \frac{1}{\nu} (y - X\hat{\beta})'(y - X\hat{\beta}) \]

It follows since

\[ (y - X\beta)'(y - X\beta) = [y - X\hat{\beta} - X(\beta - \hat{\beta})]'[y - X\hat{\beta} - X(\beta - \hat{\beta})] \]
\[ = (y - X\hat{\beta})'(y - X\hat{\beta}) + (\beta - \hat{\beta})'X'X(\beta - \hat{\beta}) \]

while the cross product is zero
Assuming Diffuse Prior

- The prior is modeled as

\[ P(\beta, \sigma) \propto \frac{1}{\sigma} \]

- Then the joint posterior of \( \beta \) and \( \sigma \) is

\[ P(\beta, \sigma|y, X) \propto \sigma^{-(T+1)} \exp \left\{ -\frac{1}{2\sigma^2} \left[ \nu s^2 + (\beta - \hat{\beta}) \left( X'X (\beta - \hat{\beta}) \right) \right] \right\} \]

- The conditional posterior of \( \beta \) is

\[ P(\beta|\sigma, y, X) \propto \exp \left( -\frac{1}{2\sigma^2} (\beta - \hat{\beta}) \left( X'X (\beta - \hat{\beta}) \right) \right) \]

which obeys the multivariable normal distribution

\[ N(\hat{\beta}, (X'X)^{-1}\sigma^2) \]
Assuming Diffuse Prior

What about the marginal posterior for $\beta$?

$$P(\beta|y,X) = \int P(\beta, \sigma|y,X) d\sigma$$

$$\propto \left[ \nu s^2 + (\beta - \hat{\beta})' X'X (\beta - \hat{\beta}) \right]^{-T/2}$$

which pertains to the multivariate student t with mean $\hat{\beta}$ and $T-M$ degrees of freedom

What about the marginal posterior for $\sigma$?

$$P(\sigma|y,X) = \int P(\beta, \sigma|y,X) d\beta$$

$$\propto \sigma^{-(\nu+1)} \exp \left( -\frac{\nu s^2}{2\sigma^2} \right)$$

which stands for the inverted gamma with $T-M$ degrees of freedom and parameter $s$

You can simulate the distribution of $\beta$ in two steps without solving analytically the integral, drawing first $\sigma$ from its inverted gamma distribution and then drawing from the conditional of $\beta$ given $\sigma$ which is normal as shown earlier. This mechanism generates draws from the Student t distribution.
Bayesian Updating/Learning

- Suppose the initial sample consists of $T_1$ observations of $X_1$ and $y_1$.
- Suppose further that the posterior distribution of $(\beta, \sigma)$ based on those observations is given by:

$$P(\beta, \sigma | y_1, X_1) \propto \sigma^{-(T_1+1)} \exp \left[ -\frac{1}{2\sigma^2} (y_1 - X_1 \beta)'(y_1 - X_1 \beta) \right]$$

$$\propto \sigma^{-(T_1+1)} \exp \left\{ -\frac{1}{2\sigma^2} \nu_1 s_1^2 + (\beta - \hat{\beta}_1)'X_1'X_1(\beta - \hat{\beta}_1) \right\}$$

where

$$\nu_1 = T_1 - M$$

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1y_1$$

$$\nu_1 s_1^2 = (y_1 - X_1\hat{\beta}_1)'(y_1 - X_1\hat{\beta}_1)$$

- You now observe one additional sample $X_2$ and $y_2$ of length $T_2$ observations.
- The likelihood based on that sample is

$$P(y_2, X_2 | \beta, \sigma) \propto \sigma^{-T_2} \exp \left[ -\frac{1}{2\sigma^2} (y_2 - X_2 \beta)'(y_2 - X_2 \beta) \right]$$
Combining the posterior based on the first sample (which becomes the prior for the second sample) and the likelihood based on the second sample yields:

\[
P(\beta, \sigma | y_1, y_2, X_1, X_2) \propto \sigma^{-(T_1+T_2+1)} \exp\left\{ -\frac{1}{2\sigma^2} \left[ (y_1 - X_1\beta)'(y_1 - X_1\beta) + (y_2 - X_2\beta)'(y_2 - X_2\beta) \right] \right\}
\]

\[
\propto \sigma^{-(T_1+T_2+1)} \exp\left\{ -\frac{1}{2\sigma^2} \left[ v s^2 + (\beta - \tilde{\beta})'\mu(\beta - \tilde{\beta}) \right] \right\}
\]

where

\[
\mu = X_1'X_1 + X_2'X_2
\]
\[
\tilde{\beta} = \mu^{-1}(X_1'y_1 + X_2'y_2)
\]
\[
vs^2 = (y_1 - X_1\tilde{\beta})'(y_1 - X_1\tilde{\beta}) + (y_2 - X_2\tilde{\beta})'(y_2 - X_2\tilde{\beta})
\]
\[
v = T_1 + T_2 - M
\]

Then the posterior distributions for \( \beta \) and \( \sigma \) follow using steps outlined earlier.

With more observations realized you follow the same updating procedure.

Notice that the same posterior would have been obtained starting with diffuse priors and then observing the two samples consecutively \( Y=[y_1', y_2']' \) and \( X = [X_1', X_2']' \).
In finance and economics you often use predictive regressions of the form

$$y_{t+1} = a + b'z_t + u_{t+1}$$

$$u_{t+1} \sim N(0, \sigma^2) \forall t = 1, \ldots, T - 1 \text{ and IID}$$

where $y_{t+1}$ is an economic quantity of interest, be it stock or bond return, inflation, interest rate, exchange rate, and $z_t$ is a collection of $M - 1$ predictive variables, e.g., the term spread.

At this stage, the initial observation of the predictors, $z_0$, is assumed to be non stochastic.

Stambaugh (1999) considers stochastic $z_0$. Then, some complexities emerge as shown later.

The predictive regression can be written more compactly as

$$y_{t+1} = x_t'\beta + u_{t+1}$$

where

$$x_t = [1, z_t]'$$

$$\beta = [a, b]'$$

In a matrix form, comprising all time-series observations, the normal regression model obtains

$$y = X\beta + u$$
You are interested to uncover the predictive distribution of the unobserved $y_{T+1}$

Let $\Phi$ denote the observed data and let $\theta$ denote the set of parameters $\beta$ and $\sigma^2$

The predictive distribution is:

$$P(y_{T+1}|\Phi) = \int_\theta P(y_{T+1}|\Phi, \theta)P(\theta|\Phi) \, d\theta$$

where

$P(y_{T+1}|\Phi, \theta)$ is the conditional or classical predictive distribution

$P(\theta|\Phi)$ is the joint posterior of $\beta$ and $\sigma^2$

Notice that the predictive distribution integrates out $\beta$ and $\sigma$ from the joint distribution $P(y_{T+1}, \beta, \sigma|\Phi)$

since

$$P(y_{T+1}, \beta, \sigma|\Phi) = P(y_{T+1}|\Phi, \theta)P(\theta|\Phi)$$
Predictive Distribution

- The conditional distribution of the next period realization is
\[
P(y_{T+1}|\theta, \Phi) \propto \sigma^{-1}\exp\left[-\frac{1}{2\sigma^2}(y_{T+1} - x_T'\beta)^2\right]
\]

- Thus \( P(y_{T+1}, \beta, \sigma|\Phi) \) is proportional to \( \sigma^{-(T+2)}\exp\left[-\frac{1}{2\sigma^2}[(y - X\beta)'(y - X\beta) + (y_{T+1} - x_T'\beta)^2]\right] \)

- On integrating \( P(y_{T+1}, \beta, \sigma|\Phi) \) with respect to \( \sigma \) we obtain
\[
P(y_{T+1}, \beta|\Phi) \propto [(y - X\beta)'(y - X\beta) + (y_{T+1} - x_T'\beta)^2]^{-(T+1)/2}
\]

- Now we have to integrate with respect to the \( M \) elements of \( \beta \)

- On completing the square on \( \beta \) we get
\[
(y - X\beta)'(y - X\beta) + (y_{T+1} - x_T'\beta)^2
= y'y + y_{T+1}^2 + \beta'\Sigma\beta - 2\beta'(X'y + x_T'y_{T+1})
= y'y + y_{T+1}^2 - (y'X + y_{T+1}'x_T)\Sigma^{-1}(X'y + x_T'y_{T+1})
+ [\beta - \Sigma^{-1}(X'y + x_T'y_{T+1})\Sigma^{-1}(X'X + x_T'x_T)]
\]

where
\[
\Sigma = X'X + x_T'x_T
\]
Predictive Distribution

- Integrating with respect to $\beta$ yields

$$P(y_{T+1}|\Phi) \propto [y'y + y_{T+1}^2 - (y'X + y_{T+1}'x_T)\nu^{-1}(X'y + x_T'y_{T+1})]^{-\frac{\nu+1}{2}}$$

where

$$\nu = T - M$$

- With some further algebra it can be shown that the predictive distribution is

$$P(y_{T+1}|\Phi) \propto [\nu + (y_{T+1} - x_T'\hat{\beta})H(y_{T+1} - x_T'\hat{\beta})]^{-\frac{\nu+1}{2}}$$

where

$$H = \frac{1}{s^2} (1 - x_T'\nu^{-1}x_T)$$

$$\nu s^2 = (y - X\hat{\beta})'(y - X\hat{\beta})$$

$$\hat{\beta} = (X'X)^{-1}X'y$$
Predictive Moments

- The first and second predictive moments, based on the t-distribution, are

\[ \mu_y = E(y_{T+1} | \Phi) = x_T \hat{\beta} \]

\[ E(y_{T+1} - \mu_y)^2 = \frac{v}{v-2} H^{-1} \]

\[ = \frac{vs^2}{v-2} (1 - x_T'I^{-1}x_T)^{-1} \]

\[ = \frac{vs^2}{v-2} (1 + x_T'(X'X)x_T) \]

- With diffuse prior, the predictive mean coincides with the classical (non Bayesian) mean.
- The predictive variance is slightly higher due to estimation risk.
- Kandel and Stambaugh (JF 1996) provide more economic intuition about the predictive density.
- The estimation risk effect on the predictive variance is analytically derived by Avramov and Chordia (JFE 2006) in a multi-asset (asset pricing) context.
- Later, we will use the predictive distribution to recover asset allocation under estimation risk and even under model uncertainty considering informative priors.
Consider the multivariate form \( (N \text{ dependent variables}) \) of the predictive regression

\[
R = XB + U
\]

where \( R \) and \( U \) are both a \( T \times N \) matrix, \( X \) is a \( T \times M \) matrix, \( B \) is an \( M \times N \) matrix

\[
\text{vec}(U) \sim N(0, \Sigma \otimes I_T)
\]

and where \( \text{vec} \) denotes the vectorization operator and \( \otimes \) is the kronecker product.

The priors for \( B \) and \( \Sigma \) are assumed to be the normal inverted Wishart (conjugate priors)

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

\[
P(\Sigma) \propto |\Sigma|^{-\frac{\nu_0 + N + 1}{2}} \exp \left( -\frac{1}{2} \text{tr}[S_0 \Sigma^{-1}] \right)
\]
Multivariate regression

where

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

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

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

where \( \text{tr} \) stands for the trace operator. This can be rewritten in a more convenient form as

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

where

\[ \hat{S} = (R - X\hat{B})'(R - X\hat{B}) \]

\[ \hat{B} = (X'X)^{-1}X'R \]
Multivariate regression

- An equivalent representation for the likelihood function is given by

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

where

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

- Combining the likelihood with the prior and completing the square on \( b \) yield

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

\[
P(\Sigma|R, X) \propto |\Sigma|^{-\frac{\nu+N+1}{2}} \exp \left( -\frac{1}{2} \text{tr}[\hat{\Sigma}^{-1}] \right)
\]
Multivariate regression

where

$$\tilde{\Psi} = \Psi_0 + X'X$$
$$\tilde{b} = \text{vec}(\tilde{B})$$
$$\tilde{B} = \tilde{\Psi}^{-1}X'X\tilde{B} + \tilde{\Psi}^{-1}\Psi_0B_0$$
$$\tilde{S} = \tilde{S} + S_0 + \tilde{B}'X'X\tilde{B} + B_0'\Psi_0B_0 - \tilde{B}'\tilde{\Psi}\tilde{B}$$
$$\tilde{\nu} = \nu_0 + T$$

- So the posterior for $B$ is normal and for $\Sigma$ is inverted Wishart.
- Again, that is the conjugate prior idea - the prior and posterior have the same distributions but with different parameters.
- Not surprisingly, $\tilde{B}$ is a weighted average of $B_0$ and $\tilde{B}$:

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

where $W = I - \tilde{\Psi}^{-1}X'X$. Notice, the weights are represented by matrices.
What if the posterior does not obey a well-known expression?

- Thus far, the posterior densities can readily be identified.
- However, what if the posterior does not obey a well known expression?
- Markov Chain Monte Carlo (MCMC) methods can be employed to simulate from the posterior.
- The basic intuition behind MCMC is straightforward.
- Suppose the distribution is $P(x)$ which is unrecognized.
- The MCMC idea is to define a Markov chain over possible values of $x$ ($x_0, x_1, x_2, ...$) such that as $n \to \infty$, we can guarantee that $x_n \sim P(x)$, that is, that we have a draw from the posterior.
- As the number of draws (each draw pertains to a distinct chain) gets larger you can simulate the posterior density.
- The simulation gets more precise with increasing number of draws.
- There are various ways to set up such Markov chains
- Here, we cover two MCMC methods: the Gibbs Sampling and the Metropolis Hastings.
Gibbs Sampling

- This paper advocates a Bayesian method in which to test the APT of Ross (1976).
- Both APT and ICAPM motivate multiple factors – extending the CAPM.
- While APT motivates statistical based factors, as shown below, the ICAPM motivates economic factors related to the marginal utility of the investor – such as consumption growth.
- The basic APT model assumes that returns on $N$ risky portfolios are related to $K$ pervasive unknown factors ($K<N$).
- The relation is described by the $K$ factor model
  \[ r_t = \mu + \beta f_t + \epsilon_t \]
  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, rather, they are latent).
Specifically,

\[ E\{ f_t \} = 0 \]
\[ E\{ f_t f'_t \} = I_K \]
\[ E\{ \varepsilon_t | f_t \} = 0 \]
\[ E\{ \varepsilon_t \varepsilon'_t | f_t \} = \Sigma = \text{diag}(\sigma^2_1, \ldots, \sigma^2_N) \]
\[ \beta = [\beta_1, \ldots, \beta_K] \]

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

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

Notice that \( \lambda_0 \) is the component of expected return unrelated to factor exposures.

The original APT model is about an approximated relation.

An exact version is derived by Huberman (1982) among others.

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

- 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

$$\beta^* = [1_N, \beta]$$

- Recovering the sampling distribution of $Q^2$ is hopeless.

- Notice that one cannot even recover an analytic expression for the posterior density of model parameters $P(\theta|R)$. $P(\theta|R, F)$ is something known – but this is not the posterior.

- However, using Gibbs sampling, we can simulate the posterior distribution of $Q^2$ as well as simulate the posterior density of all parameters and latent factors.

- In what follows, we assume that observed returns and latent factors are jointly normally distributed:

$$\begin{bmatrix} f_t \\ r_t \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ \mu \end{bmatrix}, \begin{bmatrix} I_K & \beta' \\ \beta \beta' + \Sigma \end{bmatrix} \right)$$
Gibbs Sampling

- Here are some additional notations:
  - Data: \( R = [r_1', ..., r_T']' \)
  - Parameters: \( \Theta = [\mu', \text{vec}(\beta)', \text{vech}(\Sigma)']' \) where \( \text{vech} \) denotes the distinct elements of the matrix
  - Latent factors: \( f = [f_1', ..., f_T']' \)

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

- We draw from the joint posterior in a slightly different manner than that suggested in the Geweke-Zhou paper.

- First, we employ a multivariate regression setting. Moreover, the well-known identification (of the factors) problem is not accounted for to simplify the analysis.

- The prior on the diagonal covariance matrix is assumed to be non-informative

\[
P_0(\Theta) \propto |\Sigma|^{-\frac{1}{2}} = (\sigma_1 ... \sigma_N)^{-1}
\]
Gibbs Sampling

- Re-expressing the arbitrage pricing equation, we obtain:
  \[ r'_t = F'_t B' + \epsilon'_t \]

  where
  \[ F'_t = [1, f'_t] \]
  \[ B = [\mu, \beta] \]

- Rewriting the system in a matrix notation, we get
  \[ R = FB' + E \]

- 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:

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

Therefore, we can compute the full conditional posterior densities:

$$P(B|\Sigma, F, R)$$
$$P(\Sigma|B, F, R)$$
$$P(F|B, \Sigma, R)$$

The Gibbs sampling chain is formed as follows:

1. Specify starting values $\Sigma^{(0)}$ 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.
Gibbs Sampling

- 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:
  
  \[
  P(B^{(m)}|R) \to P(B|R) \\
  P(\Sigma^{(m)}|R) \to P(\Sigma|R) \\
  P(F^{(m)}|R) \to P(F|R) \\
  P(B^{(m)}, \Sigma^{(m)}, F^{(m)}|R) \to P(B, \Sigma, F|R)
  \]

- For $m$ (burn-in draws) large enough, the $G$ values
  
  \[(B^{(g)}, \Sigma^{(g)}, F^{(g)})_{g=m+1}^{m+G}\]

  are a sample from the joint posterior.

- What are the full conditional posterior densities?

- Note:

  \[
  P(B|\Sigma, F, R) \propto \exp \left\{ -\frac{1}{2} \text{tr}[R - FB']' [R - FB'] \Sigma^{-1} \right\}
  \]
Let $b = \text{vec}(B')$. Then

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

where

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

Therefore,

$$b|\Sigma, F, R \sim N(\hat{b}, \Sigma \otimes (F'F)^{-1})$$

Also note:

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

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

$$[R - FB']'[R - FB']$$
Gibbs Sampling

suggesting that:

\[
\frac{TS_i^2}{\sigma_i^2} \sim \chi^2(T)
\]

Finally,

\[f_t|\mu, \beta, \Sigma, r_t \sim N(M_t, H_t)\]

where

\[M_t = \beta'(\beta\beta' + \Sigma)^{-1}(r_t - \mu)\]
\[H_t = I_K - \beta'(\beta\beta' + \Sigma)^{-1}\beta\]

Here, we are basically done with the GS implementation.

Having all the essential draws form the joint posterior at hand, we can analyze the simulated distribution of the pricing errors and make a call about model’s pricing abilities.
Metropolis Hastings (MH)

- Indeed, Gibbs Sampling is intuitive, easy to implement, and convergence to the true posterior density is accomplished relatively fast and with mild regularity conditions.

- Often, however, integrable expressions for the full conditional densities are infeasible.

- Then it is essential to resort to more complex methods in which draws from the posterior distributions could be highly correlated and convergence could be rather slow. Still, such methods are useful.

- One example is the Metropolis Hastings (MH) algorithm, a MCMC procedure introduced by Metropolis et al (1953) and later generalized by Hastings (1970).

- The basic idea in MH is to make draws from a candidate distribution which seems to be related to the target (unknown) distribution.

- The candidate draw from the posterior is accepted with some probability – the Metropolis rule. Otherwise, it is rejected and the previous draw is retained. As in the Gibbs Sampling, the Markov Chain starts with some initial value set by the researcher.

- The Gibbs Sampling is a special case of MH in which all draws are accepted with probability one.
Metropolis Hastings (MH)

- I will display two applications of the MH method in the context of financial economics.

- The first application goes to the seminal work of Jacquier, Polson, and Rossi (1994) on estimating a stochastic volatility (SVOL) model. Coming up on the next page.

- The second application is based on Stambaugh (1999) who analyzes predictive regressions when the first observation of the predictive variable is stochastic.

- Mostly, analyses of predictive regressions are conducted based on the assumption that the first observation is fixed non-stochastic.

- While analytically tractable this assumption does not seem to hold true.

- Relaxing that assumption entertains several complexities and the need to use MH to draw from the joint posterior distribution of the predictive regression parameters.

- I will discuss that application in the section on asset allocation when stock returns are predictable.
Stochastic Volatility (SVOL)

- The SVOL model is given by

\[
\begin{align*}
y_t &= \sqrt{h_t} u_t \\
\ln(h_t) &= \alpha + \delta \ln(h_{t-1}) + \sigma_v v_t \\
\begin{pmatrix} u_t \\ v_t \end{pmatrix} &\sim N(0, I_2)
\end{align*}
\]

- Notice that volatility varies through time rather than being constant. While in ARCH, GARCH, EGARCH models there is no stochastic innovation, here volatility is stochastic.

- Now let

\[
\begin{align*}
h' &= [h_1 \ldots h_T] \\
\beta' &= [\alpha, \delta] \\
w' &= [\alpha, \delta, \sigma_v] \\
y' &= [y_1 \ldots y_T]
\end{align*}
\]

- The posterior of \( w \) given values of \( h \) is available from the standard regression model described earlier: \( \beta \) has the multivariable normal and \( \sigma_v \) has the inverted gamma distribution.

- However, drawing from \( h|w, y \) requires more efforts.
We cannot really draw $h$ at once, rather, we have to break down the joint posterior of the entire $h$ vector by considering the series of uni-variate conditional densities.

$$P(h_t | h_{t-1}, h_{t+1}, w, y_t) \text{ for } t = 1, \ldots, T$$

If it were possible to draw directly these uni-variate densities, the algorithm would reduce to a Gibbs sampler in which we would draw successively from $P(w|h, y)$ and then each of the $T$ univariate conditionals in turn to form one step in the Markov chain.

The uni-variate conditional densities, however, exhibit an unusual form:

$$P(h_t | h_{t-1}, h_{t+1}, w, y_t)$$

$$\propto P(y_t | h_t)P(h_t | h_{t-1})P(h_{t+1} | h_t)$$

$$\propto h_t^{-1/2} \exp \left( -\frac{1}{2} \frac{y_t^2}{h_t} \right) \frac{1}{h_t} \exp \left[ -\frac{(\ln(h_t) - \mu_t)^2}{2\sigma^2} \right]$$

where

$$\mu_t = [\alpha(1 - \delta) + \delta(\ln(h_{t+1}) + \ln(h_{t-1}))/2] / (1 + \delta^2)$$

$$\sigma^2 = \frac{\sigma_v^2}{1 + \delta^2}$$
Stochastic Volatility (SVOL)

- The result follows by combining two likelihood normal terms and completing the square on \( \ln(h_t) \).
- Notice, the density is not of a standard form. It is proportional to:

\[
\frac{\exp \left( -\frac{y_t^2}{2h_t} - \frac{1}{2\sigma^2} (\ln(h_t) - \mu)^2 \right)}{(h_t)^{3/2}}
\]

\[
= \frac{\exp \left\{ -\frac{y_t^2}{2h_t} - \frac{1}{2\sigma^2} [\ln(h_t) - (\mu - \sigma^2/2)]^2 \right\}}{h_t}
\]

- A good proposal here can be the lognormal density given by

\[
\frac{1}{\sqrt{2\pi \sigma h_t}} \exp \left( -\frac{1}{2\sigma^2} [\ln(h_t) - \mu + \sigma^2/2]^2 \right)
\]
Stochastic Volatility (SVOL)

- From here we compute the ratio of the target to the proposal as:

\[
\frac{\text{target}}{\text{proposal}} \propto \frac{\exp \left( -\frac{y_t^2}{2h_t} - \frac{1}{2\sigma^2} \left[ \ln(h_t) - (\mu - \sigma^2 / 2) \right]^2 \right)}{\exp \left( -\frac{1}{2\sigma^2} \left[ \ln(h_t) - (\mu - \sigma^2 / 2) \right]^2 \right) / h_t}
\]

\[
\propto \exp \left( -\frac{y_t^2}{2h_t} \right)
\]

- For the MH algorithm, the relevant ratio is

\[
\exp \left( \frac{y_t^2}{2h_t} - \frac{y_t^2}{2h_t^*} \right)
\]

where \( h_t^* \) is the new proposed draw and \( h_t \) is the current state (or previously accepted draw).
Bayesian Portfolio Analysis

- We next present topics on Bayesian portfolio analysis based upon a review paper of Avramov and Zhou (2010) that came up in the *Annual Review of Financial Economics*.

- We first study asset allocation when stock returns are assumed to be IID.

- We then incorporate potential return predictability based on macro economy variables.

- What are the benefits of using the Bayesian approach?

- There are at least three important benefits 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 meaningful prior beliefs about the distribution of future returns.
Bayesian Asset Allocation

- We start with the mean variance framework
- 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

$$ R_t \equiv r_t - r_{ft} 1_N $$

are excess returns on the $N$ risky assets, where $1_N$ is an $N \times 1$ vector of ones.
- 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/her portfolio weights $w$, so as to maximize the quadratic objective function

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

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
  \[
  w^* = \frac{1}{\gamma} V^{-1} \mu
  \]
  and the maximized expected utility is
  \[
  U(w^*) = \frac{1}{2\gamma} \mu' V^{-1} \mu = \frac{\theta^2}{2\gamma}
  \]
  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.
Bayesian Asset Allocation

- Given a sample size of $T$, the standard maximum likelihood estimators are

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^{T} R_t$$

$$\hat{V} = \frac{1}{T} \sum_{t=1}^{T} (R_t - \hat{\mu})(R_t - \hat{\mu})'$$

- 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,

$$\hat{w}^{ML} = \frac{1}{\gamma} \hat{V}^{-1} \hat{\mu}$$

- 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 $U(w^*)$. 

Bayesian Asset Allocation

- Denote by $\theta$ the 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,

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

and the uncertainty or estimation errors are ignored.

- To account for estimation risk, let us specify the posterior distribution of the parameters as

$$p(\mu, V \mid \Phi_T) = p(\mu \mid V, \Phi_T) \times p(V \mid \Phi_T)$$

with

$$p(\mu \mid V, \Phi_T) \propto |V|^{-1/2} \exp\left\{-\frac{1}{2} \text{tr} \left[ T(\mu - \hat{\mu})(\mu - \hat{\mu})'V^{-1} \right]\right\}$$

$$P(V) \propto |V|^{-\nu/2} \exp\left\{-\frac{1}{2} \text{tr} \left[ V^{-1}T\hat{\mu} \right]\right\}$$

where $\nu = T + N$. 
Bayesian Asset Allocation

- The predictive distribution is:

\[ p(R_{T+1}|\Phi_T) \propto \left| \hat{V} + (R_{T+1} - \hat{\mu})(R_{T+1} - \hat{\mu})'/(T + 1) \right|^{-T/2} \]

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 portfolio choice.

Bayesian Asset Allocation

- Later, Bawa, Brown, and Klein (1979) provide an excellent review of the early literature.
- Under the diffuse prior, it is known that the Bayesian optimal portfolio weights are
  \[ \hat{w}_{\text{Bayes}} = \frac{1}{\gamma} \left( \frac{T - N - 2}{T + 1} \right) \hat{V}^{-1} \hat{\mu} \]
- In contrast with the classical weights \( \hat{w}_{\text{ML}} \), the Bayesian portfolio is 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.
Bayesian Asset Allocation

- However, in the classical framework, $\hat{w}^{ML}$ is a biased estimator of the true weights since, under the normality assumption,

\[
E\hat{w}^{ML} = \frac{T - N - 2}{T}w^* \neq w^*
\]

- Let

\[
\tilde{V}^{-1} = \frac{T - N - 2}{T} \hat{V}^{-1}
\]

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

- The unbiased estimator of $w^*$ is

\[
\hat{w} = \frac{1}{\gamma} \frac{T - N - 2}{T} \hat{V}^{-1} \hat{\mu}
\]

which is a scalar adjustment of $\hat{w}^{ML}$. 
Bayesian Asset Allocation

- 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.

- In the following we show how factor models can be employed to form priors.
The Black-Litterman Model

- The Black-Litterman (BL) approach attempts to propose new estimates for expected returns.
- Indeed, the sample means are simply too noisy.
- Asset pricing models - even if misspecified - could potentially deliver a good guidance.
- To illustrate, you consider a $K$-factor model (factors are portfolio spreads) and run the time series regression

$$ r^e_t = \alpha + \beta_1 f_{1t} + \beta_2 f_{2t} + \cdots + \beta_K f_{Kt} + e_t $$

- Then the estimated excess mean return is given by

$$ \hat{\mu}^e = \hat{\beta}_1 \hat{\mu}_f + \hat{\beta}_2 \hat{\mu}_f + \cdots + \hat{\beta}_K \hat{\mu}_f $$

where $\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_K$ are the sample estimates of the factor loadings, and $\hat{\mu}_f, \hat{\mu}_f, \ldots, \hat{\mu}_f$ are the sample estimates of the factor mean returns.
The Black-Litterman Model

- 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

\[ \mu_{BL} = \left[ \left( \tau \Sigma \right)^{-1} + P' \Omega^{-1} P \right]^{-1} \left[ \left( \tau \Sigma \right)^{-1} \mu^{eq} + P' \Omega^{-1} \mu^{v} \right] \]

- 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.
The Black-Litterman Model

- The $\mu^e$, which is the equilibrium vector of expected return, is constructed as follows.
- Generate $\omega_{MKT}$, the $N \times 1$ 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 that price of risk.
- One could pick a range of values for $\gamma$ and examine performance for 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 under the IID assumption.
- It does change the Sharpe ratio, however, as the standard deviation grows with the horizon by the square root of the period, while the expected return grows linearly.
The Black-Litterman Model

- Having at hand both $\omega_{MKT}$ and $\gamma$, the equilibrium return vector is given by

$$\mu^{eq} = \gamma \Sigma \omega_{MKT}$$

- 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

$$w_{TP} = \frac{\Sigma^{-1} \mu^e}{\mu' \Sigma^{-1} \mu^e}$$

- Then using $\mu^{eq}$ as the vector of excess returns on the $N$ assets would deliver $\omega_{MKT}$ as the tangency portfolio.
The question being – would you get the same vector of equilibrium mean return if you directly use the CAPM?

Yes, if...

Under the CAPM the vector of excess returns is given by

\[
\beta = \frac{\text{cov} (r^e, r^e_m)}{\sigma^2_m} = \frac{\mu^e}{\sum w_{MKT}} = \frac{\gamma \sum w_{MKT}}{\sigma^2_m}
\]
The Black-Litterman Model

Since

\[ \mu^e_m = (\mu^e)'w_{MKT} \quad \text{and} \quad r^e_m = (r^e)'w_{MKT} \]

Then

\[ \mu^e = \frac{\mu^e_m}{\sigma^2_m} \sum w_{MKT} = \mu^{eq} \]

So indeed, if you use (i) the sample covariance matrix, rather than any other specification, as well as (ii)

\[ \gamma = \frac{\mu_m - Rf}{\sigma^2_m} \]

Then the BL equilibrium expected returns and expected returns based on the CAPM are identical.
The Black-Litterman Model

- 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 third entry which is one.
- Likewise, the second row is all zero except for the fifth entry which is one.
The Black-Litterman Model

- Let us consider now two "relative views".
- Here we could incorporate market anomalies into the BL paradigm.
- 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.
- 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.
The Black-Litterman Model

- 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.
- 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.
The Black-Litterman Model

- 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.
- The same applies to value versus growth stocks.
The Black-Litterman Model

- Rule: the sum of the row corresponding to absolute views is one, while the sum of the row corresponding to relative views is zero.

- $\mu^v$ 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).

- $\Omega$ 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.
The Black-Litterman Model

- 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.
- 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.
Consider a sample of size $T$, e.g., $T=60$ monthly observations.

Let us estimate the mean and covariance 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

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

where

$$
\Delta = \left[ (\tau \Sigma)^{-1} + P' \Omega^{-1} P \right]^{-1}
$$
Pástor (2000) and Pástor and Stambaugh (1999) 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

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

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

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

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

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

$$
\alpha | \Sigma \sim N \left( 0, \sigma_\alpha^2 \left( \frac{1}{s_\Sigma^2} \Sigma \right) \right)
$$

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.
Baks, Metrick, and Wachter (JF 2001) (henceforth BMW) and Pastor and Stambaugh (JFE 2002) have explored the role of prior information about fund performance in making investment decisions.

BMW consider a mean variance optimizing investor who is largely 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 to the case where prior uncertainty is not only about managerial skills but also about model pricing abilities.

In particular, starting from Jensen (1965), mutual fund 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.

However, the intercept in such time series regressions could reflect 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 nonbenchmark 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. This SRI cost depends crucially on the investor’s views about asset pricing models and stock-picking skill by fund managers.
Mutual Funds: Prior dependence across funds

- BMW and Pastor and Stambaugh assume that the prior on alpha is independent across funds.
- Jones and Shanken (JFE 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 display 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 towards the aggregate estimate, mitigating extreme views.
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

$$r_t = a + b'z_{t-1} + u_t$$
$$z_t = \theta + \rho z_{t-1} + v_t$$

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, as shown earlier. Here, the evolution is crucial for understanding expected return and risk over long horizons.

The regression residuals are assumed to obey the normal distribution and are IID.
Bayesian Asset Allocation

- In particular, let $\eta_t = [u_t, v_t']'$ then $\eta_t \sim N(0, \Sigma)$
  where
  \[
  \Sigma = \begin{bmatrix}
  \sigma_u^2 & \sigma_{uv} \\
  \sigma_{vu} & \Sigma_v
  \end{bmatrix}
  \]
- The distribution of $r_{T+1}$, e.g., 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
  \[
  \omega^* = \arg \max_{\omega} \int_{r_{T+1}} \left[ (1 - \omega) \exp (r_f) + \omega \exp (r_f + r_{T+1}) \right]^{1-\gamma} P(r_{T+1}|\Phi_T) \, dr_{T+1}
  \]
  subject to $\omega$ being nonnegative.
Bayesian Asset Allocation

- It is infeasible to have analytic solution for the optimal portfolio.
- Then, considering a power utility investor with parameter of relative risk aversion denoted by $\gamma$ the optimization formulation is

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

subject to $\omega$ being nonnegative.
- Then, 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

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

subject to $\omega$ being nonnegative.
Let us revisit the predictive regression with one predictor only
\[ r_t = a + bz_{t-1} + u_t \]
\[ z_t = \theta + pz_{t-1} + v_t \]
\[ \begin{bmatrix} u_t \\ v_t \end{bmatrix} \sim \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{pmatrix} \sigma_u^2 & \sigma_{uv} \\ \sigma_{uv} & \sigma_v^2 \end{pmatrix} \]

The system of two equations can be re-expressed in the normal multivariate form
\[ R = X\beta + u \]
where \( R \) is a \( T \times 2 \) matrix with the first (second) column consisting of excess stock return (current values of the predictors) and \( X \) is a \( T \times 2 \) matrix with the first (second) column consisting of a \( T \times 1 \) vector of ones (lagged values of the predictors).

Previously we analyzed such multivariable regressions, assuming the initial observation is non stochastic.

In particular, the posterior is given by (recall from the section on multivariate regression)
\[ |\Sigma|^{\frac{T+N+1}{2}} \exp \left( -\frac{1}{2} (\beta - \bar{\beta}) [\Sigma^{-1} \otimes \bar{\Psi}] (\beta - \bar{\beta}) + tr[\bar{\Sigma}^{-1}] \right) \]
Stochastic Initial Observation

- In the case of stochastic initial observation we multiply this posterior by $P(z_0|b, \Sigma)$ which is

$$
\left( \frac{1 - p^2}{2\pi \sigma_v^2} \right)^{\frac{1}{2}} \exp \left\{ -\frac{1 - p^2}{2\sigma_v^2} \left( z_0 - \frac{\theta}{1 - p} \right)^2 \right\}
$$

- Notice that this is the normal distribution with unconditional mean and variance of the predictive variables.

- Integrating the posterior analytically to obtain the marginal posterior density of the parameters does not appear to be feasible.

- Instead, the posterior density can be obtained using the MH algorithm.

- The candidate distribution is normal/inverted Wishart.

- See Stambaugh (1999) for a detailed discussion.
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.


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.
Bayesian Asset Allocation

- 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

$$R_{T+K}|\Theta, \Phi_T \sim N(\lambda, Y)$$

where

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

$$Y = K\sigma_u^2 + \sum_{k=1}^{K} \delta(k)\sigma_v\delta(k)' + \sum_{k=1}^{K} \sigma_{uv} \delta(k)' + \sum_{k=1}^{K} \delta(k)\sigma_{vu}$$

$$\delta(k) = b'[\rho^{k-1} - I_M](\rho - I_M)^{-1}$$

- Drawing from the Bayesian predictive distribution is done in two steps.
- First, draw the model parameters $\Theta$ from their posterior distribution with either fixed or stochastic first observation.
- Second, conditional on model parameters, draw $R_{T+K}$ from the normal distribution.
- The optimal portfolio is accomplished by numerically maximizing

$$\frac{1}{G} \sum_{g=1}^{G} \left\{ (1 - \omega) \exp (r_f) + \omega \exp (r_f + R_{T+K}^{(g)}) \right\}^{1-\gamma} \frac{1}{1 - \gamma}$$
The mean and variance in an IID world increase linearly with the investment horizon.

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 return predictability and estimation risk, Barberis (2000) shows that investors allocate considerably more heavily to equity the longer their horizon.

This is not clear ex ante - there is a tradeoff between mean reversion and estimation risk. The mean reversion effect appears to be stronger.
Risk for the Long Run – Mean Reversion

- 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 investment horizon, as long as estimation risk is overlooked.
- However, let us assume that stock returns can be predictable by the dividend yield:

$$r_t = \alpha + \beta \text{div}_{t-1} + \epsilon_t$$
$$\text{div}_t = \phi + \delta \text{div}_{t-1} + \eta_t$$

where $\text{var}(\epsilon_t) = \sigma_1^2$, $\text{var}(\eta_t) = \sigma_2^2$, and $\text{cov}(\epsilon_t, \eta_t) = \sigma_{12}$ and the residuals are uncorrelated in all leads and lags.
Risk for the Long Run: Mean Reversion

- 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 the empirical evidence.

- This is the mean-reversion property.

- While mean reversion makes stocks appear less risky with the investment horizon, there are other effects (estimation risk, uncertainty about current and future mean return) which make stocks appear riskier.

- In the next slide, we decompose the predictive variance of long horizon return to all these effects.
Risk for the Long Run: The Predictive Variance of Long Horizon Cumulative Return

In the traditional predictive regression setup, the predictive variance consists of four parts:

\[
\text{Var}(r_{T,T+k} \mid D_T) = \\
a. \quad E(k\sigma_r^2 \mid D_T) \\
b. \quad +E\left[\sum_{i=1}^{k-1} 2b_r(I - B_x)^{-1}(I - B_x^i) \Sigma_{xr} \mid D_T\right] \\
c. \quad +E\left[\sum_{i=1}^{k-1} \left(b_r(I - B_x)^{-1}(I - B_x^i)\right)\Sigma_x \left(b_r(I - B_x)^{-1}(I - B_x^i)\right)' \mid D_T\right] \\
d. \quad +\text{Var}\left[k a_r + b_r(I - B_x)^{-1}\left((kI - (I - B_x)^{-1}(I - B_x^k)) a_x + (I - B_x^k)x_T\right) \mid D_T\right]
\]

- **a. iid uncertainty**
- **b. mean reversion**
- **c. future expected return uncertainty**
- **d. estimation risk**
The Predictive Variance of Long Horizon Cumulative Return

- The first is the IID component
- The second is the mean-reversion component
- The third reflects the uncertainty about future mean return
- The fourth is the estimation risk component
- In the predictive system of Pastor and Stambaugh (2012) there is a fifth component describing uncertainty about current mean return.
- Accounting for model uncertainty induces one more component of the predictive variance
Pastor and Stambaugh (2012) implement a predictive system to show that stocks may be more risky over long horizons from an investment perspective.

They exhibit one additional source of uncertainty about current mean return, which increases with the horizon.

Avramov, Cederbug, and Kvasnakova (2017) suggest model based priors on the return dynamics and show that per-period variance can be either higher or lower with the horizon.

In particular, prospect theory and habit formation (Long Run Risk) investors perceive less (more) risky equities due to strong (weak) mean reversion.

Bottom line: Economic theory could give important guidance about investments for the long run as the sample is not informative enough.

The next slide gives more intuition about mean reversion.
Financial economists have 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, 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 is attractive, as it explicitly incorporates model uncertainty.

The idea is to compute posterior probability for each candidate return forecasting model – then predicted return is the weighted average of return forecasting models with weights being the posterior model probabilities.

Avramov (2002) derives analytic expressions for model posterior probability. Often numerical methods are proposed.

Assuming equal prior model probability, the posterior probability is a normalized version of the marginal likelihood, which in turn is a mix of two components standing for model complexity and goodness-of-fit.
In the context of asset allocation, 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

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

where $P(M_j|\Phi_T)$ is the posterior probability that model $M_j$ 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.
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.
Bayesian Asset Allocation

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

\[
\begin{align*}
r_t &= \alpha(z_{t-1}) + \beta f_t + u_{rt}, \\
\alpha(z_{t-1}) &= \alpha_0 + \alpha_1 z_{t-1} \\
f_t &= \lambda(z_{t-1}) + u_{ft} \\
\lambda(z_{t-1}) &= \lambda_0 + \lambda_1 z_{t-1}
\end{align*}
\]

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

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

for all $t$, where $\mathbb{E}$ stands for the expected value operator.

- The model imposes restrictions on parameters and goodness of fit in the multivariate predictive regression

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

where $\mu_0$ is an $N$-vector and $\mu_1$ is an $N \times M$ matrix of slope coefficients.
Bayesian Asset Allocation

\[ r_t = (\mu_0 - \beta \lambda_0) + (\mu_1 - \beta \lambda_1) z_{t-1} + \beta f_t + u_{rt} \]

\[ \begin{align*}
\mu_0 &= \alpha_0 + \beta \lambda_0 \\
\mu_1 &= \alpha_1 + \beta \lambda_1
\end{align*} \]

- That is, under pricing model restrictions where \( \alpha_0 = \alpha_1 = 0 \) it follows that:
  \[ \begin{align*}
  \mu_0 &= \beta \lambda_0 \\
  \mu_1 &= \beta \lambda_1
  \end{align*} \]

- 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 \( \alpha_1 \neq 0 \).
Bayesian Asset Allocation

- Avramov and Chordia (2006) 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).
Bayesian Asset Allocation

- Simplistically speaking, moments used for asset allocation are
  \[
  \mu = \omega \mu_d + (1 - \omega) \mu_m
  \]
  \[
  \Sigma = \omega \Sigma_d + (1 - \omega) \Sigma_m
  \]
  where \(\omega\) is the shrinkage factor

- 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 \(0 < \omega < 1\).

- 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).
Bayesian Asset Allocation

- 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.
Could one exploit predictability to design outperforming trading strategies?

- Avramov and Chordia (2006), Avramov and Wermers (2006), and Avramov, Kosowski, Naik, and Teo (2009) are good references here.


- 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.
Bayesian Asset Allocation - The Evolution of Stock Returns

- The underlying statistical models for excess stock returns, the market premium, and macro variables are

\[ r_t = \alpha(Z_{t-1}) + \beta(Z_{t-1}) mkt_t + \nu_t \]
\[ \alpha(Z_{t-1}) = \alpha_0 + \alpha_1 Z_{t-1} \]
\[ \beta(Z_{t-1}) = \beta_0 + \beta_1 Z_{t-1} \]
\[ mkt_t = a + b' z_{t-1} + \eta_t \]
\[ z_t = c + d z_{t-1} + e_t \]

- 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 ($b \neq 0$).
Bayesian Asset Allocation

- 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.

- In the mutual fund and hedge fund context alpha reflects skill (but can also entails mispricing).

- Indeed, alpha reflects skill only if the benchmarks used to measure performance are able to price all passive payoffs.
Bayesian Asset Allocation - The Proposed Strategy

Avramov and Chordia (2006) form optimal portfolios from the universe of AMEX-NYSE stocks over the period 1972 through 2003 with monthly rebalancing on the basis of various models for stock returns.

For instance, when predictability in alpha, beta, and the equity premium is permissible, the mean and variance used to form optimal portfolios are

\[
\mu_{t-1} = \hat{\alpha}_0 + \hat{\alpha}_1 z_{t-1} + \hat{\beta}(z_{t-1})[\hat{\alpha} + \hat{\beta} z_{t-1}]
\]

\[
\Sigma_{t-1} = \hat{\beta}(z_{t-1})\hat{\beta}(z_{t-1})'\sigma^2_{mkt} + \Phi
\]

\[
+ \delta_1 \hat{\beta}(z_{t-1})\hat{\beta}(z_{t-1})'\sigma^2_{mkt} + \delta_2 \Psi.
\]

The trading strategy is obtained by maximizing

\[
w_t = \arg \max_{w_t} \left\{ w_t' \mu_{t-1} - \frac{1}{2(1/\gamma_t - r_{ft})} w_t' \left[ \Sigma_{t-1} + \mu_{t-1} \mu_{t-1}' \right] w_t \right\}
\]

where \( \gamma_t \) is the risk aversion level.

We do not permit short selling of stocks but we do allow buying on margin.
Bayesian Asset Allocation - 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.
- Altogether, we form 377 optimal portfolios on a monthly basis for each model under consideration.
- We record the realized excess return on any strategy
  \[ r_{p,t+1} = \omega_t' r_{t+1}. \]
- We evaluate the ex-post out-of-sample performance of the trading strategies based on the realized returns.
- 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.
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.

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.
Bayesian Asset Allocation - 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 2015 about $6 trillion is currently invested in U.S. equity mutual funds, making them a fundamental part of the portfolio of a domestic investor.
  - Active fund management underperforms, on average, passive benchmarks.
  - Strategies that attempt to identify subsets of funds using information variables such as past returns or new money inflows (“hot hands” or “smart money” strategies) underperform 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 generate substantial positive performance.
Understanding the SDF approach and the Hansen-Jagannathan Distance Measure
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:
  \[ E[\xi_{t+1}R_{t+1}|J_t] = 1, \]
- \( E[\cdot |J_t] \) is the expectation operator conditioned on \( J_t \), the full set of information available to investors at time \( t \).
- The fundamental pricing equation 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
  \[ \xi_{t+1} = a_t + b' f_{t+1}. \]
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

$$\xi_{t+1} = \frac{U'(W_{t+1})}{U'(W_t)}.$$

The Taylor’s series expansion of the pricing kernel around $U'(W_t)$ is

$$\xi_{t+1} = 1 + \frac{W_t U''(W_t)}{U'(W_t)} r_{w,t+1} + o(W_t),$$

$$= a + b r_{w,t+1},$$

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
\[
\xi_{t+1} = 1 + \frac{W_t u''(W_t)}{u'(W_t)} r_{w,t+1} + \frac{W_t^2 u'''(W_t)}{2u'(W_t)} r_{w,t+1}^2 + o(W_t),
\]
\[
= a + b r_{w,t+1} + c r_{w,t+1}^2.
\]
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.

Are pricing kernel parameters fixed or time-varying?

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

\[ U(c_t, x_t) = \frac{(c_t - H_t)^{1-\gamma}}{1-\gamma}, \]

where \( H_t \) is the consumption habit.

Then the pricing kernel parameters are time-varying even when the risk aversion parameter is constant.

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:

\[ \xi_{t+1} = a(z_t) + b(z_t)f_{t+1}, \]
\[ a_t = a_0 + a_1 z_t, \]
\[ b_t = b_0 + b_1 z_t. \]

Then a conditional single-factor model becomes an unconditional multifactor model

\[ \xi_{t+1} = a_0 + a_1 z_t + b_0 f_{t+1} + b_1 f_{t+1} z_t. \]

The set of factors is \([z_t, f_{t+1}, f_{t+1} z_t]'\).

The multi-factor representation is

\[ \xi_{t+1} = a_0 + a_1' z_t + b_0' f_{t+1} + b_1' [f_{t+1} \otimes z_t]. \]
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

\[ \xi_t(\Theta) = \Theta_0 + \Theta_{vw} R_{vw}^t + \Theta_{prem} R_{t-1}^{prem} + \Theta_{labor} R_{labor}^t = \Theta' Y_t \]

where

\[ \Theta = [\Theta_0, \Theta_{vw}, \Theta_{prem}, \Theta_{labor}]', \]

\[ Y_t = [1, R_{vw}^t, R_{t-1}^{prem}, R_{labor}^t]' \]

Moreover, let \( R_t = [R_{1t}, R_{2t}, ..., R_{Nt}]' \), and let

\[ f_t(\Theta) = R_t \xi_t(\Theta) - \iota_N = R_t Y_t' \Theta - \iota_N. \]

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

\[ \delta = \sqrt{E[f_t(\Theta)'][E(R_t R_t')^{-1}E[f_t(\Theta)]]}. \]

This is the HJ distance measure - 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 the function.

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

$$D_T = \frac{1}{T} \sum_{t=1}^{T} \frac{\partial f_t(\Theta)}{\partial \Theta} = \frac{1}{T} \sum_{t=1}^{T} R_t Y'_t = \frac{1}{T} R'Y,$$

$$g_T(\Theta) = \frac{1}{T} \sum_{t=1}^{T} f_t (\Theta) = D_T \Theta - t_N,$$

$$G_T = \frac{1}{T} \sum_{t=1}^{T} R_t R_t' = \frac{1}{T} R'R,$$

where

$$R = [R_1, R_2, \ldots, R_T]'$$

$$Y = [Y_1, Y_2, \ldots, Y_T].$$

The sample analog of the HJ distance is thus

$$\delta_T = \sqrt{\min_{\Theta} g_T(\Theta)'G_T^{-1}g_T(\Theta)}.$$

The first order condition of the minimization problem

$$\min_{\Theta} g_T(\Theta)G_T^{-1}g_T(\Theta),$$
is given by

\[ D_T'G_T^{-1}g_T(\Theta) = 0, \]

which gives an analytic expression for the sample minimizer

\[ \hat{\Theta} = (D_T'G_T^{-1}D_T)^{-1}D_T'G_T^{-1}t_N, \]

\[ = T(Y'R(R'R)^{-1}R'Y)^{-1}Y'R(R'R)^{-1}t_N. \]

- It follows that

\[ g_T(\hat{\Theta}) = R'Y(Y'R(R'R)^{-1}R'Y)^{-1}Y'R(R'R)^{-1}t_N - t_N \]

- From Hansen (1982) the asymptotic variance of \( \hat{\Theta} \) is given by

\[ var(\hat{\Theta}) = \frac{1}{T}(D_T'G_T^{-1}D_T)^{-1}D_T'G_T^{-1}S_TG_T^{-1}D_T(D_T'G_T^{-1}D_T)^{-1} \]

- If the data is serially uncorrelated, the estimate of the variance matrix of pricing errors is given by

\[ S_T = \frac{1}{T} \sum_{t=1}^{T} f_t(\hat{\Theta})f_t(\hat{\Theta})' \]
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

$$\text{var}(\hat{\Theta}) = \frac{1}{T} (D_T'S_T^{-1}D_T)^{-1}. $$

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

$$u = \sum_{j=1}^{N-m} \lambda_j \nu_j,$$

where $\nu_1, \nu_2, ... \nu_{N-m}$ are independent $\chi^2(1)$ random variables, and $\lambda_1, \lambda_2, ... \lambda_{N-m}$ are $N - m$ nonzero eigenvalues of the matrix $A$ given by

$$A = S^{0.5}G'^{-0.5}(I_N - (G^{-0.5})D[D'G^{-1}D]^{-1}D'G'^{-0.5})(G^{-0.5})(S^{0.5})',$$

and where $S^{0.5}$ and $G^{0.5}$ are the upper-triangle matrices from the Cholesky decomposition of $S$ and $G$. 

Professor Doron Avramov: Topics in Asset Pricing
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 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. Compute $T\delta^2_T = Tg_T(\hat{\Theta})G_T^1g_T(\hat{\Theta})$.

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 $u_i$.

5. If $u_i > T\delta^2_T$ 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 $\frac{1}{100,000}\sum_{i=1}^{100,000} I_i$. 
Considering conditional pricing models

Let us now demonstrate the implementation of the HJ measure when the pricing kernel takes the form

\[ \xi_{t+1} = (\Theta_0'X_t) + (\Theta_1'X_t)f_{t+1}^1 + \cdots + (\Theta_K'X_t)f_{t+1}^K, \]

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

\[ E[R_{t+1}((\Theta_0'X_t) + (\Theta_1'X_t)f_{t+1}^1 + \cdots + (\Theta_K'X_t)f_{t+1}^K)|Z_t] = \iota_N. \]

We collect the vector of errors,

\[ f_{t+1} = R_{t+1}((\Theta_0'X_t) + (\Theta_1'X_t)f_{t+1}^1 + \cdots + (\Theta_K'X_t)f_{t+1}^K) - \iota_N, \]

\[ = R_{t+1}Y_{t+1}'\Theta - \iota_N, \]

where

\[ \Theta' = [\Theta_0', \Theta_1', \ldots, \Theta_K'], \]

\[ Y_{t+1}' = [X_t', X_t'f_{t+1}^1, \ldots, X_t'f_{t+1}^K]. \]
Overall there are \((K + 1)(M + 1)\) parameters to estimate.

Observe that

\[
E[f_{t+1}|Z_t] = 0,
\]

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

\[
\delta^2 = g_T(\Theta)'G_T^{-1}g_T(\Theta),
\]

with \(D\) and \(G\) being estimated by

\[
D_T = \frac{1}{T} \sum_{t=0}^{T-1} [(R_{t+1} \otimes X_t)Y_{t+1}'],
\]

\[
G_T = \frac{1}{T} \sum_{t=0}^{T-1} [(R_{t+1} \otimes X_t)(R_{t+1} \otimes X_t)'].
\]
Finally, we get

$$g_T(\Theta) = D_T \hat{\Theta} - (\iota_N \otimes \bar{X}),$$

where

$$\hat{\Theta} = (D_T' G_T^{-1} D_T)^{-1} D_T' G_T^{-1} (\iota_N \otimes \bar{X}),$$

$$\bar{X} = \frac{1}{T} \sum_{t=0}^{T-1} x_t.$$

**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 non-nested 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.

7. While the idea looks neat be careful using the HJ distance measure.

8. For one, the more factors you throw in (as meaningless as they can get) the smaller the distance to the extent that artificial factors are not rejected by the test.

9. Moreover, if one model displays smaller distance it is considered better. That is imprecise as the gap of distance measures is a random quantity.

10. Indeed, the HJ statistic applies to large samples while small sample performance is troublesome.

11. In my opinion, Bayesian methods are more plausible to test and compare models. See in particular, Avramov and Chao (2006).

12. Bayesian econometrics is coming up next.
Spectral Analysis in Asset Pricing
For a comprehensive coverage of the econometrics of spectral analysis it is suggested to consult the textbook of John H. Cochrane http://faculty.chicagobooth.edu/john.cochrane/research/papers/time_series_book.pdf.

Surprisingly, there are only a few papers in asset pricing implementing spectral analysis.

One of the first studies goes back to Daniel and Marshall (1998) – henceforth DM.

There is a nice follow up work by Yu (2012).

Here is the motivation of both studies.

The sample correlation between the market excess return and consumption growth (on a quarterly frequency) is only 0.15.

Such low correlation makes it difficult for consumption based models to match the data.

Low correlation is often attributed to short term frictions including transaction and adjustment costs.

Such factors could be meaningful in high frequencies (short horizons) but they should not interrupt the model over low frequencies (long horizons).

DM and Yu perform coherence analysis of the consumption growth and excess market return.

Essentially, what the coherence analysis does is to split each of the two series into a set of Fourier (periodic) components at different frequencies, and then to determine the correlation of a set of Fourier components for the two series around each frequency.
Such correlation is not a single number but rather varies at each frequency.

Since coherence is always positive, the sign of the correlation at different frequencies cannot be determined from the coherence spectrum.

To identify the sign of the correlation, the co-spectrum can be examined.

The co-spectrum at frequency $\omega$ can be interpreted as the portion of the covariance between consumption growth and asset returns that is attributable to cycles with frequency $\omega$.

Since the covariance can be positive or negative, the co-spectrum can also be positive or negative.

Spectral analysis also yields the phase relation between the two series, which is a measure of how far the series must be shifted to maximize the correlation of the sets of Fourier components.

The slope of the phase spectrum at any frequency $\omega$ is the group delay at frequency $\omega$ and precisely measures the number of leads or lags between consumption growth and asset returns.

When this slope is positive, consumption leads the market return and vice versa.

Therefore, the coherence, co-spectrum, and phase spectrum provide a convenient tool for analyzing the lead-lag relation and the correlations at different frequencies between time series.

DM show that while there is a complete lack of correlation between asset returns and consumption growth at high frequencies, the coherence/correlation between the two series at lower frequencies is above 60%.
Yu shows how the presence of a persistent habit process leads to an attenuated correlation between consumption and returns at low frequencies.

Specifically, he shows that as long as the external habit model produces a countercyclical risk premium or a pro-cyclical price-dividend ratio, the model implies that the covariation between consumption and returns is greater in high-frequency components, whereas in the data, the opposite occurs.

Parker and Julliard (2006) build on the long run correlation to revive the CCAPM in the cross section of average return.

Rather than measuring risk by the contemporaneous covariance of an asset’s return and consumption growth, they measure risk by the covariance of an asset’s return and consumption growth cumulated over many quarters.

They find that while contemporaneous consumption risk explains little of the variation in average returns across the 25 Fama-French portfolios, their measure of ultimate consumption risk at a horizon of three years explains a large fraction of this variation.

While they don’t use spectral analysis their paper is motivated by the findings of DM Otrok, Ravikumarb, and Whiteman (2002) is another paper that implements spectral analysis to habit-formation preferences.
They show that habit agents are much more averse to high-frequency fluctuations than to low-frequency fluctuations, and further, the relatively high equity premium in the habit model is determined by a relatively insignificant amount of high-frequency volatility in U.S. consumption.


As noted by Dew-Becker and Giglio (2016) since the price of the asset reflects a combination of preferences and dynamics, it is impossible to evaluate the relative importance of the two.

Instead, Dew-Becker and Giglio quantify preferences over the dynamics of shocks by deriving frequency-specific risk prices that capture the price of risk of consumption fluctuations at each frequency.

The frequency-specific risk prices are derived analytically for leading models.

The decomposition helps measure the importance of economic fluctuations at different frequencies.

They precisely quantify the meaning of long-run in the context of Epstein-Zin preferences and measure the exact relevance of business-cycle fluctuations.

Last, they estimate frequency-specific risk prices and show that cycles longer than the business cycle long-run risks are significantly priced in the equity market.
Recall, following the notation of Bansal and Yaron (2004) the consumption and dividend dynamics are

\[ x_{t+1} = \rho x_t + \varphi_e \sigma e_{t+1} \]
\[ g_{c,t+1} = \mu + x_t + \sigma \eta_{t+1} \]
\[ g_{d,t+1} = \mu_d + \phi x_t + \varphi_d \sigma u_{t+1} \]

Then we can show that when Epstein-Zin (1989) recursive preferences excess return on the dividend claim can be written as:

\[ r_{t+1}^e \approx \bar{r} + \frac{\phi - 1}{\psi} k_{1m} \varphi_e \sigma e_{t+1} + \varphi_d \sigma u_{t+1} \]

where \( \psi \) is the IES, \( k_{im} \) is the constant in the CS log linearization and \( \rho = \text{corr}(\eta_{t+1}, u_{t+1}) \)
The spectral representations of \( \{g_{c,t}\}, \{r_{t}^{ex}\}, \{x_{t}\}, \{e_{t}\}, \{u_{t}\}, \text{and} \ \{\eta_{t}\} \) are

\[
dZ_{g_{c}} = e^{-i\lambda} dZ_{x} + \sigma dZ_{\eta}
\]

\[
dZ_{x} = \rho e^{-i\lambda} dZ_{x} + \varphi_{e} \sigma dZ_{e}
\]

\[
dZ_{r} = \frac{\phi - 1}{1 - k_{1m} \rho} k_{1m} \varphi_{e} \sigma dZ_{e} + \varphi_{d} \sigma dZ_{u}
\]

Define \( A_{1m} = \frac{\phi - 1}{1 - k_{1m} \rho} \) rearrange, and solve for \( dZ_{g_{c}} \) and \( dZ_{r} \) to obtain:

\[
dZ_{g_{c}} = \frac{\varphi_{e} \sigma e^{-iw}}{1 - \rho e^{-iw}} dZ_{e} + \sigma dZ_{\eta}
\]

\[
dZ_{r} = k_{1m} A_{1m} \varphi_{e} \sigma dZ_{e} + \varphi_{d} \sigma dZ_{u}
\]
Thus, the multivariate spectrum is given by

\[
\begin{align*}
    f_{rr} &= (k_1 A_1 \varphi \sigma)^2 + (\varphi_d \sigma)^2 \\
    f_{gg} &= \left| e^{-i\lambda} \frac{\varphi \sigma}{1 - \rho e^{-i\omega}} \right|^2 + \sigma^2 \\
    f_{gr} &= \left( \frac{e^{-i\omega} - \rho}{1 + \rho^2 - 2\rho \cos(\omega)} \right) k_1 A_1 \varphi_e^2 \sigma^2 + \varphi_d \sigma^2 \rho \eta_u \\
        &= \left( \frac{e^{-i\omega} - \rho}{1 + \rho^2 - 2\rho \cos(\omega)} \right) k_1 A_1 \varphi_e^2 \sigma^2 + \varphi_d \sigma^2 \rho \eta_u
\end{align*}
\]

Solving for the cospectrum \( C_{sp}(\omega) \), the real part of the cross spectrum \( f_{12}(\omega) \), yields:

\[
C_{sp}(\omega) = \left( \frac{\cos(\omega) - \rho}{1 + \rho^2 - 2\rho \cos(\omega)} \right) k_1 A_1 \varphi_e^2 \sigma^2 + \varphi_d \sigma^2 \rho \eta_u
\]
Taking the derivative and rearranging the equation yield

\[ C'_{sp}(w) = \frac{\sin(w)(\rho^2 - 1)}{(1 + \rho^2 - 2\rho \cos(w))^2} \cdot \frac{\phi - \frac{1}{\psi}}{1 - k_1m\rho}k_1m\varphi_e^2\sigma^2 \]

From the expression of \( f_{gr} \) - under the assumption of \( \rho_{nu} = 0 \), we can solve for the phase spectrum \( \phi_{12}(w) \):

\[ \tan(\phi_{12}(w)) = \frac{-\sin(w) - \rho}{\cos(w) - \rho} \]

Thus, taking the derivative, it follows that

\[ \phi'_{12}(w) \propto \frac{\delta(-\sin(w) - \rho)}{\cos(w) - \rho} \]
\[ \propto -\cos(w)^2 + \rho \cos(w) - \sin(w)^2 - \rho \sin(w) \]
\[ = \rho (\cos(w) - \sin(w)) - 1 < 0 \]

where ‘\( \propto \)’ denotes that both sides of ‘\( \propto \)’ have the same sign. Thus, the phase spectrum is always decreasing.
The cospectrum, phase, and Habit formation (based on Jianfeng, Yu 2012)

- The growth of log consumption is modeled as
  \[ g_{c,t} = \mu_c + \varepsilon_{c,t} \]

- The approximated excess return on dividend claim is
  \[ r_{t+1}^e \approx \alpha - \beta_s \sum_{j=1}^{\infty} \phi_s^{j-1} g_{c,t+1-j} + \beta_c \varepsilon_{c,t+1} + \beta_\delta \varepsilon_{\delta,t+1} \]

- Now let us understand the cospectrum and phase of the joint consumption return process.

- We first replace the consumption and return dynamics using inverse FT

  \[ dZ_{g_c}(w) = dZ_{\varepsilon_c}(w) \]

  \[ dZ_r(w) = -\beta_s \sum_{j=1}^{\infty} \phi_s^{j-1} \exp(-iw)dZ_{g_c}(w) + \beta_c dZ_{\varepsilon_c}(w) + \beta_\delta dZ_{\varepsilon_\delta}(w) \]
Notice that

\[\sum_{j=1}^{\infty} \phi_s^{j-1} \exp(-ijw) = \frac{\exp(-iw)}{1 - \phi_s \exp(-iw)}\]

So:

\[dZ_r(w) = \left(\frac{-\beta_s \exp(-iw)}{1 - \phi_s \exp(-iw)} + \beta_c\right) dZ_{\epsilon_c}(w) + \beta_\delta dZ_{\epsilon_\delta}(w)\]

Then, the multivariate spectrum is given by

\[2\pi f_{11}(w) = \sigma_c^2\]

\[2\pi f_{22}(w) = \left|\frac{-\beta_s \exp(-iw)}{1 - \phi_s \exp(-iw)} + \beta_c\right|^2 \sigma_c^2 + \beta_\delta^2 \sigma_\delta^2 + 2Re\left(\frac{-\beta_s \exp(-iw)}{1 - \phi_s \exp(-iw)} + \beta_c\right)\beta_\delta \sigma_c \delta\]
\[ 2\pi f_{12}(w) = \left( \frac{-\beta_S \exp(-iw)}{1 - \phi_s \exp(-iw)} + \beta_c \right)' \sigma_c^2 + \beta_\delta \sigma_{c\delta} \]
\[ = \left( \frac{\beta_S (\phi_s - \exp(iw))}{1 + \phi_s^2 - 2\phi_s \cos(w)} + \beta_c \right) \sigma_c^2 + \beta_\delta \sigma_{c\delta} \]

- For the cospectrum \( C_{sp}(w) \) the real part of the cross spectrum \( f_{12}(w) \)
\[ 2\pi C_{sp}(w) = \left( \frac{\beta_S (\phi_s - \cos(w))}{1 + \phi_s^2 - 2\phi_s \cos(w)} + \beta_c \right) \sigma_c^2 + \beta_\delta \sigma_{c\delta} \]

- Therefore, the derivative of the cospectrum is
\[ C_{sp}'(w) = \beta_s \sigma_c^2 \frac{\sin(w)(1 + \phi_s^2 - 2\phi_s \cos(w)) + 2\phi_s \sin(w)(\cos(w) - \phi_s)}{(1 + \phi_s^2 - 2\phi_s \cos(w))^2} \]
\[ = \frac{\beta_s \sigma_c^2 \sin(w)}{(1 + \phi_s^2 - 2\phi_s \cos(w))^2} (1 - \phi_s^2)^2 \geq 0 \]

and the portion of covariance contributed by components at frequency \( w \) is increasing in the frequency \( w \).
By definition, the coherence and the phase are, respectively

\[ h(w) = \frac{|f_{12}|}{\sqrt{f_{11}f_{22}}} \]

\[ \tan(\phi_{12}(w)) = \frac{\beta_S \sin(w)}{1 + \phi_S^2 - 2\phi_S \cos(w)} \sigma_c^2 \]

\[ = \frac{\beta_S \sin(w)}{1 + \phi_S^2 - 2\phi_S \cos(w)} + \beta_S \sigma_c \]

At the frequency \( w = 0 \), the cospectrum is

\[ C_{sp}(0) = \left( -\beta_S \frac{1 - \phi_S}{1 + \phi_S^2 - 2\phi_S} + \beta_c \right) \sigma_c^2 + \beta_S \sigma_c \]

\[ = \left( 1 - a_1 k_1 - \frac{a_1 (1 - k_1)}{S (1 - \phi_S)} \right) \sigma_c^2 + \beta_S \sigma_c \]
Therefore, the low-frequency correlation between consumption growth and asset returns is negative if and only if

\[
(1 - a_1 k_1 - \frac{a_1 (1 - k_1)}{S(1 - \phi_s)}) + \beta_\delta \frac{\sigma_{c\delta}}{\sigma_c^2} < 0
\]

By differentiating the following equation

\[
tag(\phi_{12}(w)) = \frac{\beta_s \sin(w)}{1 + \phi_s^2 - 2\phi_s \cos(w)} \sigma_\epsilon^2
\]

\[
= \left( \frac{\beta_s(\phi_s - \cos(w))}{1 + \phi_s^2 - 2\phi_s \cos(w)} + \beta_c \right) \sigma_c^2 + \beta_\delta \sigma_{c\delta}
\]

the sign of the slope of the phase spectrum can be examined.
To see this, start with

\[ \phi'_{12}(w) \]

\[ \propto -\left[ \beta_s (\phi_s - \cos(w)) \sigma_C^2 + (\beta_c \sigma_c^2 + \beta_\delta \sigma_c \delta)(1 + \phi_s^2 - 2\phi_s \cos(w)) \right] \cdot \beta_s \cos(w) + \beta_s \sin(w) \left[ \beta_s \sin(w) \sigma_C^2 + 2\phi_s (\beta_c \sigma_c^2 + \beta_\delta \sigma_c \delta) \sin(w) \right] \]

where \( \propto \) denotes that both sides of \( \propto \) have the same sign.

Rearrange and simplify to obtain

\[ \phi'_{12}(w) \]

\[ \propto - \frac{a_1(k_1 \phi_s)^{-1}}{S} + 2\phi_s \left( 1 + a_1 k_1 \frac{1 - S}{S} \right) + 2\phi_s \beta_\delta \frac{\sigma_c \delta}{\sigma_C^2} \]

\[ - \left\{ 1 + a_1 k_1 \frac{1 - S}{S} + \beta_\delta \frac{\sigma_c \delta}{\sigma_C^2} + \phi_s^2 + \frac{-a_1 k_1 S \phi_s^2}{S} + \frac{a_1 \phi_S}{S} + \beta_\delta \phi_s^2 \frac{\sigma_c \delta}{\sigma_C^2} \right\} \cos(w) \]

\[ \geq - \left[ 1 - a_1 k_1 - a_1 \frac{1 - k_1}{S(1 - \phi_s)} + \beta_\delta \frac{\sigma_c \delta}{\sigma_C^2} \right] (1 - \phi_s)^2 \]
The inequality above requires the assumption

\[(1 - a_1 k_1)(1 + \phi_s^2) + \frac{a_1 k_1}{S} + \frac{a_1 \phi_s}{S} + \beta \delta (1 + \phi_s^2) \frac{\sigma_c \delta}{\sigma_c^2} > 0\]

which is true if the correlation between the innovations of return and consumption is positive.

Thus, the phase spectrum is increasing as long as

\[\left[ 1 - a_1 k_1 - \frac{a_1 (1 - k_1)}{S (1 - \phi_s)} + \beta \delta \frac{\sigma_c \delta}{\sigma_c^2} \right] + \beta \delta \frac{\sigma_c \delta}{\sigma_c^2} < 0\]

and the correlation between the innovations of return and consumption is positive.
Appendix
Quality Investing using Return on Invested Capital

- The Little Book That Beats the Market - by Joel Greenblatt: The idea here is to combine only two financial ratios – earnings yield (EBIT / enterprise value) and return on capital (EBIT/net fixed assets plus working capital). Greenblatt suggests the “magic formula”: purchasing 30 cheap stocks with a high earnings yield and a high return on capital. The receipt is below.

1. Decide on minimum market capitalization (usually greater than $50 million).
2. Exclude utility and financial stocks.
3. Exclude foreign companies (American Depositary Receipts).
4. Compute company's earnings yield = EBIT / enterprise value.
5. Compute company's return on capital = EBIT / (net fixed assets + working capital).
6. Rank all companies above the threshold market capitalization by highest earnings yield and highest return on capital.
7. Invest in 20–30 highest ranked companies.
8. Re-balance portfolio once per year, selling losers one week before the year-mark and winners one week after the year mark.
Quality investing based on the G Score

**G Score:** The G-Score is due to Mohanram (2005). It combines traditional fundamentals, such as earnings and cash flows, with measures tailored for growth firms, such as earnings stability, growth stability and intensity of R&D, capital expenditure and advertising. A long–short strategy based on GSCORE earns significant excess returns, though most of the returns come from the short side. Thus, to form an attractive trading strategy one could take long positions based on the F-Score or the F-score combined with the book-to-market ratio and short positon based on the G-score.

The formation of the G-score based on 8 binary variables as follow:

- **G1** is equal 1 if a firm’s ROA is greater than the contemporaneous median ROA for all low BM firms in the same industry and 0 otherwise. ROA, defined as the ratio of net income before extraordinary items scaled by average total assets.
- **G2**=1 if a firm’s cash flow ROA exceeds the contemporaneous median for all low BM firms in the same industry and 0 otherwise. Cash flows ROA is similar to the above-defined ROA except that operating cash flows replace net income.
- **G3**=1 if a firm’s cash flow from operations exceeds net income and 0 otherwise.
Quality investing based on the G-Score

- $G_4 = 1$ if a firm’s earnings variability is less than the contemporaneous median for all low BM firms in the same industry and 0 otherwise.
- $G_5 = 1$ if a firm’s sales growth variability is less than the contemporaneous median for all low BM firms in the same industry and 0 otherwise.
- $G_6, G_7$ and $G_8$ are defined to equal 1 if a firm’s R&D, capital expenditure and advertising intensity respectively, are greater than the contemporaneous medians of the corresponding variables for all low BM firms in the same industry and 0 otherwise. The intensity of R&D, capital expenditure and advertising are measured by deflating these variables by beginning assets.

The signals relating to profitability and cash flows ($G_1$-$G_3$) as well as those related to conservatism ($G_6$-$G_8$) are created using the annualized financials. The two signals earnings variability and sales growth variability ($G_4$-$G_5$) are generated from quarterly financials of the past 4 years, with the constraint that at least six quarters of information be available. While quarterly information might induce variability owing to seasonality, the industry adjustment should mitigate this.
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. 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. The $\Delta XFIN$ measure can be decomposed as

$$\Delta XFIN = \Delta CEquity + \Delta PEquity + \Delta Debt$$

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
Downside Risk

- Downside risk is the financial risk associated with losses.
- There are various downside risk measures which quantify the risk of losses, the expected loss given the realization of a loss, or even the worst case scenario characterizing a particular investment.
- All downside risk measures exclusively focus on the left tail of the return distribution, whereas volatility measures are both about the upside and downside outcomes.
- Typical downside risk measures include the Value at Risk (VaR), expected shortfall, semi-variance, maximum drawdown, downside beta, and shortfall probability. To establish the trading strategy one can focus on VaR, which is a very well-used measure in risk management.
- Downside risk measures are often positively and often negatively associated with average returns.
Tail risk as a common factor and equity premium predictor

- Kelly and Jiang (2014) propose the tail exponent as both a common factor and a predictor of the equity premium.

- The tail exponent is an aggregate variable constructed based on daily returns.

- In particular, consider daily returns of all stocks within a particular month and identify the 5th percentile of the cross sectional distribution or the return threshold.

- Then only for those daily returns which fall below the return threshold take the simple average of the natural log of return divided by the return threshold

\[
\lambda_t = \frac{1}{k_t} \sum_{k=1}^{k_t} \ln \frac{r_{kt}}{u_t}
\]

where \(k_t\) is the number of exceedances, \(u_t\) is the return threshold, \(r_{kt}\) is the daily return that falls below the threshold.

- Perhaps an open question would be how tail risk is associated with other anomalous patterns in the cross section of average returns.
Proof of result on the maximum Sharpe ratio

If the SDF is conditionally log-normally distributed, hence we can apply the general formula:

\[
SR_t = \sqrt{\exp(Var_t(\log M_{t+1}) - 1)} \approx \sigma_t(\log M_{t+1})
\]

Proof: use the log-normal formula

\[
E(\exp(X)) = \exp\left( E(X) + \frac{1}{2} Var(X) \right)
\]

and compute

\[
E_t(M_{t+1}) = \exp\left( E_t(\log M_{t+1}) + \frac{1}{2} Var_t(\log M_{t+1}) \right)
\]

\[
Var_t(M_{t+1}) = E_t(M_{t+1}^2) - [E_t(M_{t+1})]^2
\]

\[
= \exp\left( 2E_t(\log M_{t+1}) + 2Var_t(\log M_{t+1}) \right)
\]

\[-\exp\left( 2E_t(\log M_{t+1}) + Var_t(\log M_{t+1}) \right)
\]

The result follows.
References
Abel, Andrew B. 1990. *Asset Prices Under Habit Formation and Catching Up with the Joneses*.


- Avramov, Doron, Si Cheng, and Allaudeen Hameed. 2016. Mutual funds and mispriced stocks.


Campbell, John Y. 1992. *Intertemporal Asset Pricing without Consumption Data*.


Feng, Guanhao, Stefano Giglio, and Dacheng Xiu, 2017, Taming the factor zoo, Working Paper.


Geczy, Christopher, and Mikhail Samonov. 2013. 212 years of price momentum (the World’s longest backtest: 1801–2012). *Available at SSRN*.


Zellner, Arnold. 1971. *An Introduction to Bayesian Inference in Econometrics*.


