Review Session 1
Part 1
The Econometrics of Asset Management

CLEFIN Course: 20192

Financial Econometrics 2

The Econometrics of Asset Management

Important **takeway** from this part:

- Econometric tools outlined in this course cannot be used regardless of the environment where you are using them (theoretical framework and data characteristics).
- e.g. is your main aim to predict returns or to manage portfolio risk? Are you working with low or high frequency data?
- Different answers to these questions imply different assumptions and hence different econometrics!

The Econometrics of Asset Management

- Classical Constant Expected Returns (CER)
- Dynamic Dividend Growth Model
- The DDGM as a bivariate cointegrated model
- Asset Pricing with predictable returns
- Quantitative risk management
- Exam exercise

CER

- Efficient market hypothesis
- Example of return estimation
- Static Allocation Problem with CER
- Improvements to standard estimation
  1. Resamples optimal mean-variance portfolio
  2. Black and Litterman’s approach

Efficient market hypothesis

(Fama, 1970)

- The CER model assumes that an asset's return over time is normally distributed with constant mean and constant covariance.
- The model allows for the returns on different assets to be contemporaneously correlated, but independent over time both across and within the same asset

$$
\begin{aligned}
r_{i,t} &= \mu_i + \sigma_i \epsilon_{it} \\
\epsilon_{it} &\sim NID(0,1) \\
cov(\epsilon_{it}, \epsilon_{js}) &= \begin{cases} \sigma_{ij} & t = s \\ 0 & t \neq s \end{cases}
\end{aligned}
$$

Implications of CER model

- CAPM is a good measure of risk and thus a good explanation of why some stocks earn higher average returns than others

$$\boldsymbol{\mu} - r^f \mathbf{e} = \boldsymbol{\beta} \odot [(\mu_M - r^f)\mathbf{e}]$$

- Excess returns are close to unpredictable; any predictability is a statistical artifact or cannot be exlpoited after transaction costs are taken into account

$$E[\mathbf{r}_{t+1} - r^f \mathbf{e} | \mathcal{I}_t] = E[\mathbf{r}_{t+1} - r^f \mathbf{e}] = \boldsymbol{\mu} - r^f \mathbf{e}$$

$$\Sigma_t \equiv Var[\mathbf{r}_{t+1} - r^f \mathbf{e} | \mathcal{I}_t] = Var[\mathbf{r}_{t+1} - r^f \mathbf{e}] = \Sigma$$

- Asset prices behave as a (log) random walk with drift

Regression Model Representation

- The simplest case of the CER also assumes that all residuals are both contemporaneously and serially uncorrelated, with diagonal covariance matrix. Then, because the diagonal structure of the covariance matrix, classical OLS equation by equation can be applied. Consider for example the observations on the i^th^ return.

$$\mathbf{y}_i = \mathbf{e}_T \delta_i + \mathbf{u}_i,$$

$$\mathbf{y}_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix}, \quad \mathbf{X}_i = \mathbf{e}_T = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$$

The OLS estimates of the relevant parameters are then simply:

$$\hat{\delta}_i = \frac{1}{T}\sum_{t=1}^{T} r_{it} = \bar{r}_i \quad \hat{\sigma}_{11} = \hat{\sigma}_1^2 = \frac{1}{T}\sum_{t=1}^{T} (r_{it} - \bar{r}_i)^2$$

Given this we can compute the *tangency portfolio:* $\hat{\mathbf{w}}^T = \dfrac{\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu} - r^f \mathbf{e}\right)}{\mathbf{e}'\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\mu} - r^f \mathbf{e}\right)}$

What happens in practice?

- The traditional, simple approach to portfolio allocation can lead to dramatic swings in optimal portfolio weights for small changes in investment views and conditions, as given by the estimates/forecasts of μ and Σ. There is a simple reason for this common finding: too much sampling error in the estimation of the vector of expected returns and, due to this, an asset allocation which is idiosyncratic to the specific estimation sample.
- Luckily, there are solutions to this problem:
  1. **Resampled optimal mean-variance portfolio:** use methods that keep the simplest possible estimates of μ and Σ but fully recognize that the resulting estimates are simply realizations of sample estimators that may imply considerable parameter uncertainty;
  2. **Black and Littermann's approach:** allowing for more complex econometric models of returns that are capable of exploiting predictability.

Focus on Resampled Optimal-Mean Variance Portfolio

- Implement **bootstrap methods** to derive the optimal portfolio allocation.
- Consider the estimation of a simple multivariate model, in which the only regressor is a constant for the returns r^i_t on N assets, $i = 1, 2, ..., N$

$$\begin{aligned} r_{1t} &= \hat{\mu}_1 + \hat{u}_{1t} \\ r_{2t} &= \hat{\mu}_2 + \hat{u}_{2t} \\ &\dots \\ r_{Nt} &= \hat{\mu}_N + \hat{u}_{Nt} \end{aligned}$$

$$\begin{bmatrix} \hat{u}_{1t} & \hat{u}_{2t} & \dots & \hat{u}_{Nt} \end{bmatrix}' \sim \mathcal{N}\left(\mathbf{0}, \hat{\Sigma}\right)$$

Steps of the bootstrap procedure

1. Collect of the residuals from estimation in the following TxN matrix

$$\hat{U} \equiv \begin{bmatrix} \hat{u}_{11} & \hat{u}_{21} & \ldots & \hat{u}_{N1} \\ \hat{u}_{12} & \hat{u}_{22} & \ldots & \hat{u}_{N2} \\ \vdots & \ldots & \ddots & \vdots \\ \hat{u}_{1T} & \hat{u}_{2T} & \ldots & \hat{u}_{NT} \end{bmatrix}$$

2. Draw a new sample of size T of residuals by extracting randomly T rows from $\hat{U}$
   Given these new, re-sampled residuals collected in a vector $\hat{u}_t^1$ (t = 1, 2, ..., T) and the estimates $\boldsymbol{\mu}$, we proceed to generate a new artificial sample of returns using

$$\mathbf{r}_t^1 = \boldsymbol{\mu} + \hat{u}_t^1$$

   where the subscript "1" alludes to the fact that this represents the first iteration of the algorithm. At this point, a new OLS estimation of the model is performed on this artificial data, obtaining as an outcome a pair of new, bootstrapped estimates, $\hat{\mu}^1$ and $\hat{\Sigma}^1$ and, using the classical formul $\hat{w}^1$
3. Iterate the algorithm B times, where B is in general a large number (let's say 5,000 or 10,000 times), using the fact that at the b^th^ iteration one simply draws a new sample of size T of residuals by extracting randomly T rows from $\hat{U}$ **,generate a new artificial sample of returns.**

This total number B of replications of this procedure will generate B optimal portfolio allocations $\{\hat{w}^b\}_{b=1}^B$

Resampled Optimal-Mean Variance Portfolio

- Now that we have B optimal portfolio allocations $\{\hat{w}^b\}_{b=1}^B$, the desired vector of re-sampled, optimized portfolio weights may be represented by the average, across the B bootstraps:

$$\breve{w}^{boot} = \frac{1}{B}\sum_{b=1}^{B}\hat{w}^b$$

- This method and the resulting average portfolio allocation across bootstraps acknowledges the effects of estimation uncertainty and is **generally more stable across different sample** whenever the instability in the portfolio allocation is generated by estimation uncertainty rather than by a true structural break (or other forms of statistical instability such as regimes) in the distribution of the vector of risky asset returns.

Empirical challenges to the CER model: the DDG model

- Practictioners implementing portfolio allocation based on the CER model experienced rather soon a number of problems that stressed limitations of this model but it was the work of Robert Shiller and co-authors that led te profession to go beyond the CER model.

Simple log price-dividend ratio framework

- Starting from the definition of logarithmic, continuously compounded returns, defined as:

$$r^s_{t+1} = \log\left(1 + H^s_{t+1}\right) = \log\left(\frac{P_{t+1} + D_{t+1}}{P_t}\right)$$

- We can arrive at this expression of the log dividend yield (see lecture notes and slides to check computation):

$$(p_t - d_t) = \frac{\kappa}{1-\rho} + \sum_{j=1}^{m} \rho^{j-1}\left(\Delta d_{t+j}\right) - \sum_{j=1}^{m} \rho^{j-1}\left(r^s_{t+j}\right) + \rho^m\left(p_{t+m+1} - d_{t+m+1}\right)$$

which shows that p_t - d_t measures the value of a very long-term investment strategy (buy and hold) which is equal to the stream of future dividend growth discounted at the appropriate rate, which reflects the risk free rate plus risk premium required to hold risky assets

$$(p_t - d_t) = \frac{\kappa}{1-\rho} + \sum_{j=1}^{m} \rho^{j-1} \left(\Delta d_{t+j}\right) - \sum_{j=1}^{m} \rho^{j-1} \left(r_{t+j}^{s}\right) + \rho^{m} \left(p_{t+m+1} - d_{t+m+1}\right)$$

- Predictability of returns is ruled out under the null that expected returns are constant.
- In this case the price-dividend ratio should be completely determined by the process generating dividends and determining their expectations.
- The empirical evidence is strongly against this prediction (see the Shiller(1981) and Campbell-Shiller(1987)).
- Stock prices are too volatile to be determined only by expected dividends.

The empirical evidence is against the CER prediction that price-dividend ratio should be completely determined by the process generating dividends

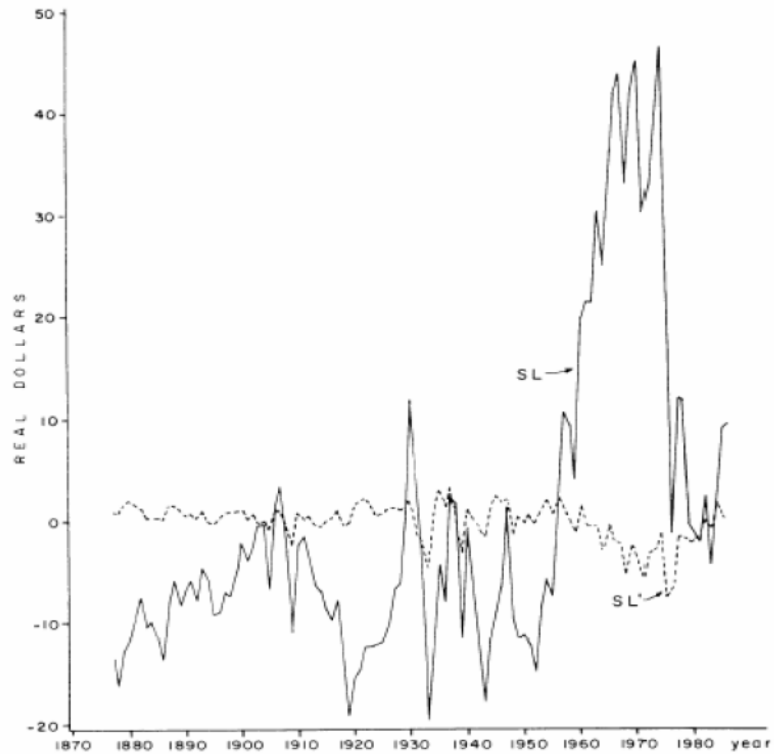


FIG. 2.—Stock market: deviations from means of actual spread (SL_t = Price$_t$ − θ Dividend$_{t-1}$) and theoretical spread SL_t', θ = 12.195.

- The following figure, taken from Campbell-Shiller (1987) illustrates the point by reporting the observed price-dividend ratio and a counterfactual price-dividend ratio which is obtained by assuming constant future expected returns and by using a Vector Autoregressive Model to predict future dividend-growth:

Predictability at different horizons

$$r^s_{t+1} = \kappa + \rho E_t\left(p_{t+1} - d_{t+1}\right) + E_t \Delta d_{t+1} - \left(p_t - d_t\right) + \rho u^{pd}_{t+1} + u^{\Delta d}_{t+1}$$

$$\sum_{j=1}^{m} \rho^{j-1} r^s_{t+j} = \frac{\kappa}{1-\rho} + \sum_{j=1}^{m} \rho^{j-1} E_t\left(\Delta d_{t+j}\right) - \left(p_t - d_t\right) + \underbrace{\boxed{\rho^m E_t\left(p_{t+m} - d_{t+m}\right)}}_{\downarrow\; 0 \text{ as } m \to \infty} + \boxed{\rho^m u^{pd}_{t+m} + \sum_{j=1}^{m} \rho^{j-1} u^{\Delta d}_{t+j}} \longrightarrow \text{Const.}$$

- These two expressions illustrate that when the price dividends ratio is a noisy process, such noise dominates the variance of one-period returns and the statistical relation between the price dividend ratio and one period returns is weak.
- However as the horizon over which returns are defined gets longer, noise tends to be dampened and the predictability of returns given the price dividend ratio increases. Note that expected returns are constant the price dividend ratio should be only determined by future dividend growth.

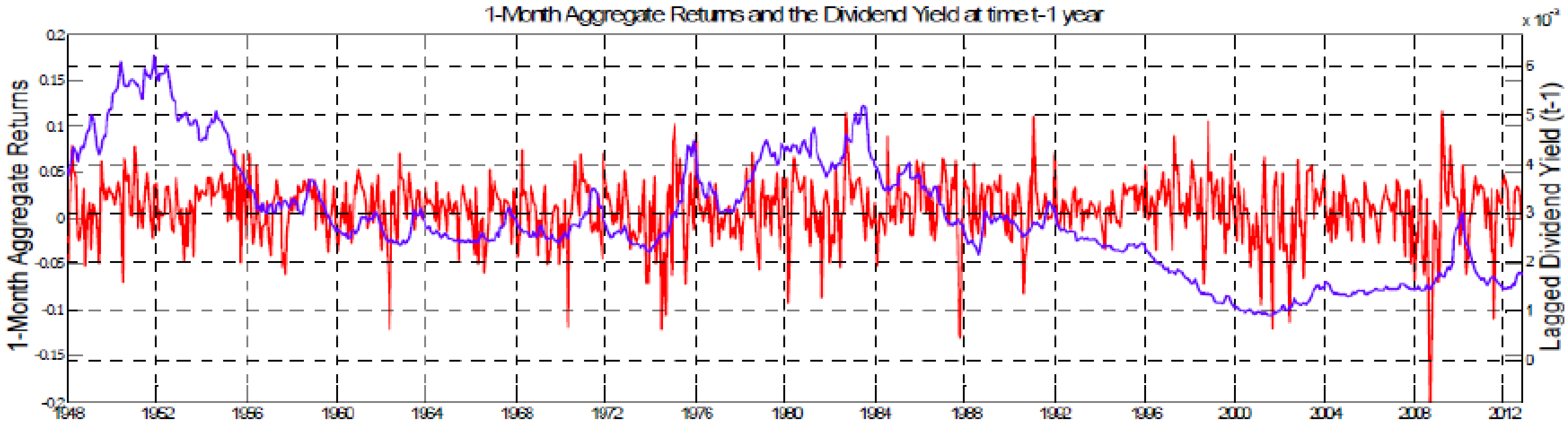
Empirical evidence

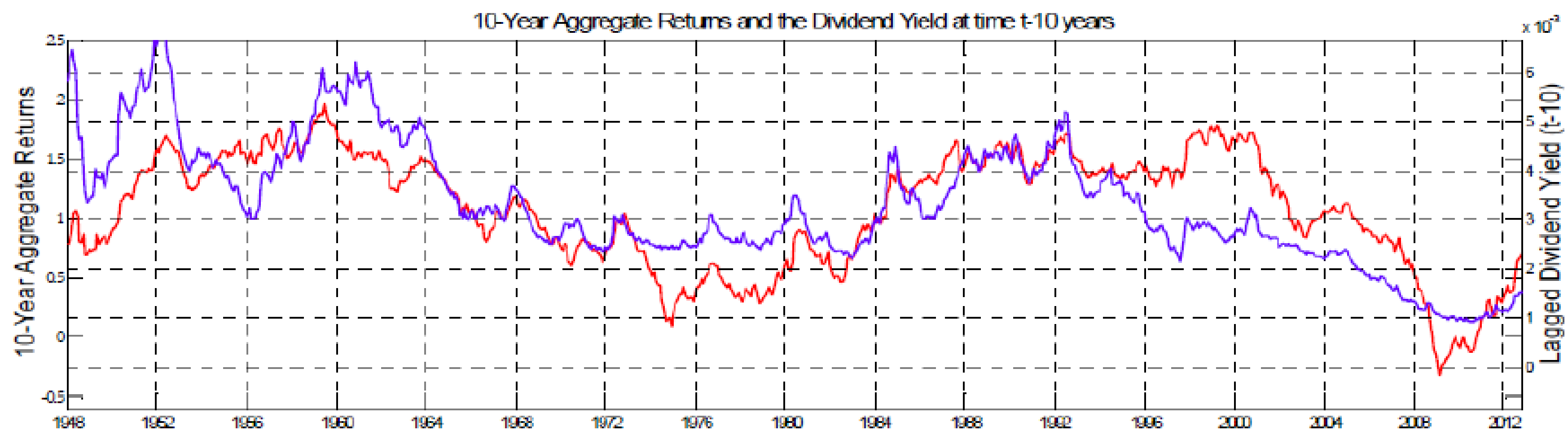
- Here we report the slopes, the adjusted R2, as well as the adjusted t-stats as in Valkanov (2003), of the following predictive regression

$$r_{t:t+k} = \alpha_k + \beta_k \log\left(D_t / P_t\right) + \sigma\varepsilon_{t+k} \qquad \varepsilon_{t+k} \sim N(0,1)$$

- where rt:t+k the aggregate US stock market returns from t to t + k, Dtthe aggregate dividend, Pt the index, *#t+k an idiosyncratic error*component and *s its corresponding risk.*

Horizon k	$\hat{\beta}$	$t/\sqrt{T}$	R^2
1	0.726	0.092	0.007
4	3.369	0.187	0.032
8	7.105	0.269	0.066
16	15.96	0.412	0.144
24	23.59	0.523	0.214
60	54.69	0.976	0.487


1-Month Aggregate Returns and the Dividend Yield at time t-1 year
1-Month Aggregate Returns
Lagged Dividend Yield (t-1)
x 10^-3


10-Year Aggregate Returns and the Dividend Yield at time t-10 years
10-Year Aggregate Returns
Lagged Dividend Yield (t-10)
x 10^-3

DDGM as a bivariate cointegrated model

- The DDGM that we have just seen states that the log of the dividend price (p_t-d_t) is stationary. However, we know from evidence showed in class that taken indipendently the log of prices and the log of dividends series are non-stationary!
- And we know that when two or more series are individually integrated (for example they present a unit-root, like in this case) but some linear combination of them has a lower order of integration, then the series are said to be **cointegrated**.

Error Correction Model

- Let's outline this case and see how we will end up to the same theoretical conclusions of the DDGM

$$lp_t = a_0 + a_1 lp_{t-1} + a_2 ld_{t-1} + \epsilon_{1t}$$
$$ld_t = b_0 + b_1 ld_{t-1} + \epsilon_{2t}.$$

- Take differences and you end up with the **Error Correction Model**

$$\Delta lp_t = a_0 + \alpha \left(lp_{t-1} - \beta_1 ld_{t-1} \right) + \epsilon_{1t}$$
$$\Delta ld_t = b_0 + \epsilon_{2t}$$
$$\alpha \equiv (a_1 - 1) \qquad \beta_1 \equiv -\frac{a_2}{a_1 - 1}.$$

$$\Delta lp_t = a_0 + \alpha\,(lp_{t-1} - \beta_1 ld_{t-1}) + \epsilon_{1t}$$
$$\Delta ld_t = b_0 + \epsilon_{2t}$$
$$\alpha \equiv (a_1 - 1) \qquad \beta_1 \equiv -\frac{a_2}{a_1 - 1}.$$

- If we interpret βld_{t-1} as the long-run equilibrium level for the log-stock price, $lp^*_{t-1} = \beta ld_{t-1}$, then you understand the meaning of the correction part
- If $a_1 < 1$, then $\alpha \equiv a_1 - 1 < 0$, and when $lp^*_{t-1} < \beta ld_{t-1}$ implies that $\Delta lp_t > 0$, i.e., when prices are below their long-run equilibrium defined by dividends, then prices will increase
- When $lp_{t-1} > \beta ld_{t-1}$ implies that $\Delta p_t < 0$, i.e., when prices are above their long-run equilibrium defined by dividends, then prices will decrease

- The parameter α in the ECM specification determines the **speed of adjustment** in the presence of disequilibrium. The system defined by the ECM based on a cointegrating relationship is **self-equilibrating**

$$\Delta lp_t = a_0 + \alpha\left(lp_{t-1} - \beta_1 ld_{t-1}\right) + \epsilon_{1t}$$
$$\Delta ld_t = b_0 + \epsilon_{2t}$$
$$\alpha \equiv (a_1 - 1) \qquad \beta_1 \equiv -\frac{a_2}{a_1 - 1}.$$

- Finally, note that the prediction of the dividend growth model covered in our past lecture is that: $\beta_1 = \frac{a_2}{1-a_1} = 1$
- Because *Δlp* represents most of the return variation, cointegration implies that we can use the log of the dividend price ratio as a predictor for stock market returns. This allows us to re-interpret in terms of cointegration between prices and dividends the results of the predictive regressions of stock market returns on the dividend price ratio that we have obtained and commented earlier.

Asset Pricing with Predictable Returns

- Different approaches have been used in finance to model time-varying expected returns. Consider a situation in which in each period, *k* state of nature can occur and each state has a probability *p(k). In the absence of arbitrage* opportunities the price of an asset *i* at time t can be written as follows:

$$P_{i,t} = \sum_{s=1}^{k} \pi_{t+1}(s) m_{t+1}(s) X_{i,t+1}(s)$$

- m_{t+1} (s) is the discounting weight attributed to future pay-offs, which (as the probability *p) is independent from the asset i.*

- $X_{i,t+1}$ (s) are the payoffs of the assets and therefore returns on assets are defined as: $1 + R_{i,t+1} = \frac{X_{i,t+1}}{P_{i,t}}$

- After some passage (and exploiting the definition of covariance, see lecture notes) we can write:

$$E_t\left(R_{i,t+1} - R_{s,t+1}\right) = -\left(1 + R_{s,t+1}\right) cov\left(m_{t+1} R_{i,t+1}\right)$$

- Assets whose returns are low when the stochastic discount factor is high (i.e. when agents values payoffs more) require an higher risk premium, i.e. an higher excess return on the risk-free rate.
- Consider now the case where the period t is made by two points in time very close to each other (a short holding horizon), in this case m_{t+1} can be safely approximated by a constant (very close to one) and excess returns are not predictable. As the point in time that define the period becomes further and further separated, then time variation in m cannot be discounted anymore and future excess returns becomes predictable if their covariance with m is predictable.

Quantitative Risk Management

- Having solved the portfolio weights (preferentially exploting the info we now have have achieved on predictability of relevant future returns), there is a different role that econometrics can play at high frequencies.
- When k is small the following framework is normally referred to:

$$
\begin{aligned}
R_{t,t+k} &= \sigma_{k,t} u_{t+k} \\
\sigma^2_{k,t} &= f(\mathcal{I}_t) \qquad u_{t+k} \sim IID\ \mathcal{D}(0,1).
\end{aligned}
$$

The following features of the model at high frequency are noteworthy:

1. The distribution of returns is centered around a mean of zero, and the zero mean model dominates any alternative model based on predictors.
2. The variance is time-varying and predictable, given the information set, It, available at time t.
3. The distribution of returns at high frequency is not normal, i.e., D(0, 1) may often differ from N(0, 1)

The VaR is the percentage loss obtained with a probability at most of α percent:

$$\Pr\left(R^p < -VaR_\alpha\right) = \alpha.$$

If the distribution of returns is normal, then α-percent VaR_α is obtained as follows (assume $\alpha \in (0,1)$):

$$\begin{aligned}\Pr\left(R^p < -VaR_\alpha\right) &= \alpha \iff \Pr\left(\frac{R^p - \mu_p}{\sigma_p} < -\frac{VaR_\alpha + \mu_p}{\sigma_p}\right) = \alpha \\ &\iff \Phi\left(-\frac{VaR_\alpha + \mu_p}{\sigma_p}\right) = \alpha,\end{aligned}$$

where $\Phi(\cdot)$ is the cumulative density of a standard normal. At this point, defining $\Phi^{-1}(\cdot)$ as the inverse CDF function of a standard normal, we have that

$$-\frac{VaR_\alpha + \mu_p}{\sigma_p} = \Phi^{-1}(\alpha) \iff VaR_\alpha = -\mu_p - \sigma_p\Phi^{-1}(\alpha).$$

Exam excercise

$$
\begin{aligned}
r_{t,t+4}^{UK} &= \underset{(0.16)}{1.04} + \underset{(0.05)}{0.29}\, dp_t^{UK} + \varepsilon_{2,t+4}; \\
NObs. &= 141, \qquad R^2 = 0.203
\end{aligned}
$$

```
F-statistic vs.  constant model:  35.3 (p-value = 2.13e-08)
```

10a (1 point) If the mean of dp_t^{UK} over the estimation sample is -3.23, what is the average real return $r_{t,t+4}^{UK}$ over the sample?

10b (1 point) Are real stock market returns predictable at a 1-year horizon?

10c (2 points) Can the model specification adopted above be considered an Error Correction Model? If so, please indicate the potentially cointegrated variables and the cointegrating vector.

10d (1 point) Do you think that the residuals from the regression might feature a Moving Average structure? If so, explain why and give indications as to the resulting structure.