

# The Empirical Application of Finance

Carlo A. Favero,<sup>1</sup>

<sup>1</sup>Bocconi University & CEPR

Chapter 3

# Model Specification: the Constant Expected Return Model

Our objective is the specification of a statistical model for asset prices and returns. To this end consider the (naive) log random walk (LRW) hypothesis:

$$\ln P_t = \mu + \ln P_{t-1} + \epsilon_t$$

where the 'innovations'  $\epsilon_t$  are assumed to be uncorrelated across time ( $cov(\epsilon_t; \epsilon_{t'}) = 0 \quad \forall t \neq t'$ ), with constant expected value 0 and constant variance  $\sigma^2$ . Sometimes, a further hypothesis is added and the  $\epsilon_t$  are assumed to be jointly normally distributed.

If we take prices as inclusive of dividends we can write the following model for total log-returns

$$r_{t,t+1} = \mu + \sigma \epsilon_t$$
$$\epsilon_t \sim i.i.d.(0, 1)$$

# CER Model Properties

This simple specification has some appealing properties for the  $n$  period returns  $r_{t,t+n}$ :

If we assume the LRW and consider a sequence of  $n$  log returns  $r_t^*$  at times  $t, t-1, t-2, \dots, t-n+1$  (just for the sake of simplicity in notation we suppose each time interval  $\Delta$  to be of length 1 and drop the generic  $\Delta$ ) we have that:

$$E(r_{t,t+n}) = E\left(\sum_{i=1}^n r_{t+i,t+i-1}\right) = \sum_{i=1}^n E(r_{t+i,t+i-1}) = n\mu$$

$$Var(r_{t,t+n}) = Var\left(\sum_{i=1}^n r_{t+i,t+i-1}\right) = \sum_{i=1}^n Var(r_{t+i,t+i-1}) = n\sigma^2$$

## Model Estimation

Model specification has led us to the following description for the vector of one-period returns on assets used to build a portfolio:

$$\begin{aligned}\mathbf{r}_{t,t+1} &= \boldsymbol{\mu} + \mathbf{H}\epsilon_{t+1} \\ \Sigma &= \mathbf{H}\mathbf{H}' \\ \epsilon_{t+k} &\sim \mathcal{D}(\mathbf{0}, \mathbf{I})\end{aligned}$$

where  $\mathbf{r}_{t,t+k}$  is the vector of returns between time  $t$  and time  $t+k$  in which we are interested,  $\boldsymbol{\mu}$  is the vector of mean returns and the matrix  $\mathbf{H}$  determines the time invariant variance-covariance matrix of returns.

## Model Estimation

Model estimation allows to find values for  $\mu$ ,  $\Sigma$ . In the case of CER this step is easily solved by n OLS regressions of the n returns on a constant.

$$\hat{\mu}^i = \frac{1}{T} \sum_{t=1}^T r_{t,t+1}^i$$

$$\hat{\sigma}_{ii} = \frac{1}{T} \sum_{t=1}^T \left( r_{t,t+1}^i - \hat{\mu}^i \right)^2$$

$$\hat{\sigma}_{ij} = \frac{1}{T-1} \sum_{t=1}^T \left( r_{t,t+1}^i - \hat{\mu}^i \right) \left( r_{t,t+1}^j - \hat{\mu}^j \right)$$

## Model Estimation: the OLS method

The CER is a special case of a linear model , consider the following general representation of a linear model :

$$\mathbf{y} = \mathbf{X}\beta + \epsilon,$$

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ y_N \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdot & \cdot & x_{1k} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ x_{N1} & x_{N2} & \cdot & \cdot & x_{Nk} \end{pmatrix},$$
$$\beta = \begin{pmatrix} \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_k \end{pmatrix}, \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_N \end{pmatrix}.$$

# Model Estimation: the OLS method I

- The simplest way to derive estimates of the parameters of interest is the ordinary least squares (OLS) method.
- Such a method chooses values for the unknown parameters to minimize the magnitude of the non-observable components.
- The best fit is obtained by minimizing the sum of squared vertical deviations of the data points from the fitted line.

## Model Estimation: the OLS method II

Define the following quantity:

$$\mathbf{e}(\beta) = \mathbf{y} - \mathbf{X}\beta,$$

where  $\mathbf{e}(\beta)$  is a  $(n \times 1)$  vector. If we treat  $\mathbf{X}\beta$ , as a (conditional) prediction for  $\mathbf{y}$ , then we can consider  $\mathbf{e}(\beta)$  as a forecasting error. The sum of the squared errors is then

$$\mathbf{S}(\beta) = \mathbf{e}(\beta)' \mathbf{e}(\beta).$$



## Model Estimation: the OLS method III

The OLS method produces an estimator of  $\beta$ ,  $\hat{\beta}$ , defined as follows:

$$\mathbf{S}(\hat{\beta}) = \min_{\beta} \mathbf{e}(\beta)' \mathbf{e}(\beta).$$

Given  $\hat{\beta}$ , we can define an associated vector of residual  $\hat{\epsilon}$  as  $\hat{\epsilon} = \mathbf{y} - \mathbf{X}\hat{\beta}$ . The OLS estimator is derived by considering the necessary and sufficient conditions for  $\hat{\beta}$  to be a unique minimum for  $\mathbf{S}$ :

1.  $\mathbf{X}'\hat{\epsilon} = 0$ ;
2.  $\text{rank}(\mathbf{X}) = k$ .

Condition 1 imposes orthogonality between the  $\mathbf{X}$  variables and the OLS residuals, it ensures that residuals have zero mean when a constant is included among the regressors. Condition 2 requires that the columns of the  $\mathbf{X}$  matrix are linearly independent.

## Model Estimation: the OLS method IV

From 1 we derive an expression for the OLS estimates:

$$\mathbf{X}'\hat{\epsilon} = \mathbf{X}'(\mathbf{y} - \mathbf{X}\hat{\beta}) = \mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{X}\hat{\beta} = 0,$$

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}.$$

$$\hat{\sigma}^2 = \frac{\hat{\epsilon}'\hat{\epsilon}}{T - k}$$

## OLS in the CER

In the CER for asset with return  $r_t$  we have:

$$\mathbf{y} = \mathbf{X}\beta + \epsilon,$$

$$\mathbf{y} = \begin{pmatrix} r_1 \\ \cdot \\ \cdot \\ \cdot \\ r_T \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{pmatrix}, \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_T \end{pmatrix}.$$

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = (T)^{-1} \left( \sum_{t=1}^T r_t \right) = \mu$$

## Model Simulation

Once parameters in the CER have been estimated the model can be simulated to derive the distribution of asset returns in the future, this is done by generating pseudo data from the model.

- split the sample in two part, a training sample and a test sample.
- Use the training sample to estimate model parameters
- use the model to simulate artificial observation for the test sample.

# Monte-Carlo Simulation

- To use Monte-Carlo Simulation to generate pseudo data from the CER model, some estimates of  $\mu$   $\sigma$  are necessary.
- Given this estimates an assumption must be made on the distribution of  $\epsilon_t$ . Then an artificial sample for  $\epsilon_t$  of the length matching that of the available can be computer simulated.
- The simulated residuals are then mapped into simulated returns via  $\mu$ ,  $\sigma$ .
- This exercise can be replicated N times (and therefore a Monte-Carlo simulation generates a matrix of computer simulated returns whose dimension are defined by the sample size T and by the number of replications N).
- The distribution of model predicted returns can be then constructed at any point in time (including the future)

# Bootstrap Simulation

- One of the possible limitation of the Monte-Carlo approach is the choice a distribution from which the residuals are to be drawn.
- Bootstrap methods overcome this problem by sampling residuals from their empirical distribution.
- All the steps in a bootstrap simulation are the same with the Monte-Carlo simulation except that different observations for residuals are constructed by taking the deviation of returns from their sample mean putting them in an urn and resampling from the urn with replacement.

# Asset Allocation with the CER

- The code `prog_port.R` after the usual preliminaries ( setting working directory, upload relevant packages) use the inbuilt database BERNDINVEST in the package Econfin to perform optimal asset allocation using the CER model for US stocks.
- First, Data transformation is applied via a loop to construct, from monthly returns monthly prices, i.e. the value over-time of a buy and hold portfolio in each stock, and monthly log-prices.
- Second, descriptive graphical analysis is implemented using the facilities in the package ggplot.
- Third, the relevant parameters in the CER are estimated and optimal asset allocation is found by computing weights for the tangency portfolio.
- Lastly, the utilization of the package fPortfolio in R is described. The programme illustrates how to get the data in the appropriate format, set constraints for the portfolio optimization, compute efficient frontiers and optimal portfolio weights and provide graphic illustration of the results.

# Model Simulation with the CER: backtesting and VaR

- To illustrate model simulation we consider the application to backtesting and Value at Risk.
- The code **prog\_sim.R** starts by building the tangency and the minimum variance portfolio on the stocks considered in the previous section and by implementing within sample evaluation of the properties of the optimized portfolios.
- This type of exercise suffers from the well-known problem of "look-ahead bias" as data non available in real time have been used to construct weights.
- True backtesting when the available data are divided into a "training Sample" and a test sample is then implemented using the package fPortfolio. A procedure is used according to which a rolling sample is used to build the allocation that reflect the information available in real time at the end of sample, allocation are then evaluated out-of-sample and then re-optimized. The possibility of smoothing optimal weights in the rolling procedure is also considered. Backtesting is then conducted by assessing ex-post the performance of each allocation.
- Finally, in the last part of the code, the CER model is applied to the Tangency portfolio to simulate, via bootstrap and Monte-Carlo procedures, the distribution of the returns and to produce one- month ahead Value-at-Risk.