Portfolio Management: New Methods

Francisco Peñaranda (University of Alicante and FMG/LSE)
<penaranda@merlin.fae.ua.es>

- Traditional Portfolio Management
- Black-Litterman Approach to Estimation Error
- Portfolios of Non-Normal Returns
- Benchmark-Relative Management
- Indexing and Statistical Arbitrage: New Methods

Portfolio Management: New Methods

Francisco Peñaranda (University of Alicante and FMG/LSE)
<penaranda@merlin.fae.ua.es>

1 Traditional Portfolio Management

- Foundations of Mean-Variance Approach

- Analysis of Implied Risk and Return

- Objections to Mean-Variance Paradigm

References:

• Brandt, .M W. (2004): Portfolio Choice Problems, in Y. Ait-Sahalia and L.P. Hansen (eds.), *Handbook of Financial Econometrics*, forthcoming.

Campbell, J.Y., and L.M. Viceira (2002): *"Strategic Asset Allocation: Portfolio Choice for Long-Term Investors"*, Oxford University Press.

Litterman, B. (2003): *"Modern Investment Management: An Equilibrium Approach"*, Wiley.

• Scherer, B. (2004): *"Portfolio Construction and Risk Budgeting"*, Risk Books.

1.1 Foundations of Mean-Variance Approach

- Notation:

A gross return is denoted by R and its rate of return by r,

$$R = 1 + r.$$

The risk-free rate of return is denoted by c and an excess return by e,

$$e = r - c.$$

We assume the investor has access to N risky assets and a safe asset. The corresponding vector of excess returns is

$$\mathbf{e} = \mathbf{r} - c\mathbf{1},$$

where $\mathbf{1}$ is a vector of ones. The corresponding means ($N \times 1$ vector) and variances ($N \times N$ matrices) are

$$E(\mathbf{r}) = \boldsymbol{\nu}, \qquad E(\mathbf{e}) = \boldsymbol{\mu},$$
$$Var(\mathbf{r}) = Var(\mathbf{e}) = \boldsymbol{\Sigma}.$$

We also define

$$H = \boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu},$$

and we will use less frequently some standard notation of traditional mean-variance analysis without safe asset,

$$B = \boldsymbol{\nu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\nu}, \qquad A = \boldsymbol{\nu}'\boldsymbol{\Sigma}^{-1}\mathbf{1}, \qquad C = \mathbf{1}'\boldsymbol{\Sigma}^{-1}\mathbf{1},$$
$$D = BC - A^2.$$

Portfolios will be characterized by a vector of weights $\mathbf{w}$, with associated return

$$r_p = (1 - \mathbf{w}'\mathbf{1})\, c + \mathbf{w}'\mathbf{r} = c + \mathbf{w}'\mathbf{e},$$

where we normalize portfolio cost to one. The corresponding notation for its moments is

$$\begin{gathered} E(r_p) = (1 - \mathbf{w}'\mathbf{1})\, c + \mathbf{w}'\boldsymbol{\nu} = \nu_p, \\ E(e_p) = \mathbf{w}'\boldsymbol{\mu} = \mu_p, \\ Var(r_p) = Var(e_p) = \sigma_p^2 = \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}. \end{gathered}$$

Its Sharpe ratio is defined as

$$SR_p = \frac{\mu_p}{\sigma_p}.$$

Unless otherwise noted, we assume our relevant investor maximizes expected utility of final wealth with an initial wealth normalized to 1 as the cost of R_p.

$$\max_{\mathbf{w}} E\left[u(R_p)\right].$$

The budget constraint is implicit in the definition of $\mathbf{w}$. We could add other constraints such as no short sales.

- Mean-variance efficiency rests on one of the following assumptions:

▶Investors exhibit quadratic utility

$$u(R_p) = R_p - \frac{b}{2} R_p^2 \Rightarrow E\left[u(R_p)\right] = U\left(\nu_p, \sigma_p^2\right),$$

where $b > 0$.

But this utility shows decreasing marginal utility from some point and increasing absolute risk aversion.

▶Returns are multivariate normal.

Usual justification, but it can be more general⇒Elliptical distributions. See Chamberlain (1983).

It is skewness, more than kurtosis, what is a problem for mean-variance analysis.

- Mean-variance optimization in an asset-only world:

Optimal portfolio is given by any of the following equivalent problems

$$\min_{\mathbf{w}} Var\left(e_p\right) \qquad s.t. \qquad E\left(e_p\right) = \mu_p \Rightarrow \mathbf{w}_{mv}\left(\mu_p\right) = \left(\frac{\mu_p}{H}\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu},$$
$$\max_{\mathbf{w}} E\left(e_p\right) \qquad s.t. \qquad Var\left(e_p\right) = \sigma_p^2 \Rightarrow \mathbf{w}_{mv}\left(\sigma_p^2\right) = \sqrt{\frac{\sigma_p^2}{H}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu},$$
$$\max_{\mathbf{w}} E\left(e_p\right) - \frac{1}{2\lambda} Var\left(e_p\right) \Rightarrow \mathbf{w}_{mv}\left(\lambda\right) = \lambda\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu},$$

where we can interpret $\lambda > 0$ as risk tolerance.

Note that we can infer λ from chosen portfolio

$$\lambda = \frac{\sigma_p^2}{\mu_p}$$

and the efficient frontier is such that

$$\sigma_p^2 = \frac{1}{H}\mu_p^2.$$

Efficient frontier: Linearity and separation

- The efficient frontier in space $\left(\mu_p, \sigma_p\right)$ is linear.
- Every efficient portfolio is a portfolio of c and a particular portfolio of risky assets (market portfolio).

Brandt (2004): Figure 1

This linearity justifies the use of the Sharpe ratio as a measure of portfolio efficiency⇒The Sharpe ratio of every optimal portfolio is

$$SR = \sqrt{H}.$$

- Mean-variance frontier if there is not a safe asset:

We will only use these results in the chapter on tracking error. Now

$$r_p = \mathbf{w}'\mathbf{r}.$$

Traditional frontier without safe asset is the solution to

$$\begin{aligned} &\min_{\mathbf{w}} Var\left(r_p\right) \qquad s.t. \qquad E\left(r_p\right) = \nu_p, \qquad \mathbf{1}'\mathbf{w} = 1 \\ \Rightarrow\ &\mathbf{w}_{mv}\left(\nu_p\right) = \frac{1}{D}\left(B - A\nu_p\right)\mathbf{\Sigma}^{-1}\mathbf{l} + \frac{1}{D}\left(C\nu_p - A\right)\mathbf{\Sigma}^{-1}\boldsymbol{\nu}. \end{aligned}$$

The frontier in the $\left(\mu_p, \sigma_p^2\right)$ space is

$$\sigma^2\left(\nu_p\right) = \frac{1}{D}\left(C\nu_p^2 - 2A\nu_p + B\right).$$

It defines an hyperbola in $\left(\mu_p, \sigma_p\right)$ space.

Any two portfolios on the frontier span the whole frontier.

- Asset-Liability Management (ALM):

The mean-variance framework can be easily extended to cover background risks. For instance liabilities for ALM, which is the relevant context for pension funds.

▶ Management of difference between assets and liabilities (Surplus):
We can define surplus returns as change in surplus relative to assets

$$\frac{\Delta Surplus}{Assets} = R_p - fR_l, \qquad f = \frac{Liabilities}{Assets},$$

and interpret liabilities as a benchmark or background risk

$$r_b = fr_l,$$

taking f as given.

▶ Surplus-efficient frontier:
We can easily apply the mean-variance analysis as

$$\min_{\mathbf{w}} Var\left(r_p - r_b\right) \qquad s.t. \qquad E\left(r_p\right) = \nu_p,$$

which will be studied in the chapter on benchmark-relative management.

The main novelty is a new component in the optimal portfolio to hedge liability risk

$$\mathbf{w}_{br}\left(\nu_p\right) = \mathbf{w}_{mv}\left(\nu_p\right) + \mathbf{w}_h,$$

such that there is a parallel movement in the $\left(\nu_p, \sigma_p^2\right)$ space.

Scherer (2004): Figure 1.1

▶ Cases when surplus and asset-only efficient frontiers coincide:

· Liabilities are cash or covariance between assets and liabilities is zero.

· The covariance of assets with liabilities has a special structure.

· There exists a liability mimicking asset and it lies on the efficient frontier.

1.2 Analysis of Implied Risk and Return

- Basic risk decomposition:

Risk budget⇒Sources of risk.

Marginal (or absolute) contribution to risk is defined as

$$\mathbf{MCTR} = \frac{\mathrm{d}\sigma_p}{\mathrm{d}\mathbf{w}} = \frac{\boldsymbol{\Sigma}\mathbf{w}}{\sigma_p} = \boldsymbol{\beta}_p \sigma_p,$$

$$\boldsymbol{\beta}_p = \frac{Cov\left(\mathbf{r}, r_p\right)}{Var\left(r_p\right)}.$$

Given that

$$\sigma_p = \mathbf{w}' \frac{\mathrm{d}\sigma_p}{\mathrm{d}\mathbf{w}},$$

we can define percentage (or relative) contribution to risk as

$$\mathbf{PCTR} = \frac{diag\left(\mathbf{w}\right)}{\sigma_p} \frac{\mathrm{d}\sigma_p}{\mathrm{d}\mathbf{w}} = \mathbf{w} \odot \boldsymbol{\beta}_p,$$

$$\mathbf{PCTR}'\mathbf{1} = 1,$$

where $diag\left(\mathbf{w}\right)$ is a diagonal matrix with $\mathbf{w}$ on its diagonal and $\odot$ denotes Hadamard product.

Scherer (2004): Table 1.3-4

Risk budgeting vs. portfolio optimization:

- Equivalence of VAR and classic risk measures under normality.
- If returns are not normal then we should not use the mean-variance framework and maybe VAR is a good alternative. However, risk budgeting is still the same as portfolio optimization.

- Implied view analysis: Reverse optimization.

If $\mathbf{w}$ is optimal then we can map positions to implicit views

$$\boldsymbol{\mu} = \left(\frac{1}{\lambda}\right)\boldsymbol{\Sigma}\mathbf{w} = \left(\frac{\mu_p}{\sigma_p^2}\right)\boldsymbol{\Sigma}\mathbf{w} = \boldsymbol{\beta}_p \mu_p.$$

In addition, if $\mathbf{w}$ is the market portfolio then we can give an equilibrium interpretation to those views.

Scherer (2004): Table 1.4; Figure 1.4

This is one of the key elements of Black-Litterman approach to portfolio management that will be studied later.

1.3 Objections to Mean-Variance Paradigm

Some short-comings of mean-variance framework:

- Relative magnitude of estimation errors:

Inputs $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ in portfolio optimization are uncertain.

Variance can be estimated more accurately than mean if returns follow diffusions as Merton (1980) shows.

Black-Scholes model implies that log-returns for a period Δt are normally distributed as

$$d = \Delta \ln S \sim N\left(\nu \Delta t, \sigma^2 \Delta t\right),$$

where S is the price of the asset under study, and $\left(\nu, \sigma^2\right)$ are the drift and diffusion parameters.

Let us assume that we have N observed log-returns $(d_1, ..., d_N)$ during a period H, so that $H = \Delta t N$ (think of time measured in years and daily log-returns). The corresponding sample mean and variance are

$$\bar{d} = \frac{1}{N}\sum_{i=1}^{N} d_i \sim N\left(\nu \Delta t, \frac{\sigma^2 \Delta t}{N}\right),$$

$$S_d^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(d_i - \bar{d}\right) \sim \left(\frac{\sigma^2 \Delta t}{N-1}\right)\chi^2(N-1).$$

Therefore, we will estimate $\left(\nu, \sigma^2\right)$ by

$$\hat{\nu} = \frac{\bar{d}}{\Delta t}, \qquad \hat{\sigma}^2 = \frac{S_d^2}{\Delta t}.$$

The variance of those estimators is

$$Var(\hat{\nu}) = \frac{Var\left(\bar{d}\right)}{(\Delta t)^2} = \frac{\sigma^2}{H}, \qquad Var\left(\hat{\sigma}^2\right) = \frac{Var\left(S_d^2\right)}{(\Delta t)^2} = \frac{2\sigma^4}{N-1}.$$

That is, the variance of $\hat{\nu}$ depends on H, while the variance of $\hat{\sigma}^2$ depends on N, so that the letter can be made arbitrarily accurate.

Scherer (2004): Table A1-2

We will dedicate one chapter to study how to deal with estimation error.

- How well does the mean-variance framework approximate reality?

We can express a general expected utility through its Taylor expansion around the mean of R_p

$$E\left[u\left(R_p\right)\right] = E\left[\sum_{j=0}^{\infty} \frac{u^{(j)}\left(1+v_p\right)}{j!}\left(r_p - v_p\right)^j\right] = \sum_{j=0}^{\infty} \frac{u^{(j)}\left(1+v_p\right)}{j!} E\left[\left(r_p - v_p\right)^j\right],$$

where

$$u^{(j)}\left(1+v_p\right) = \frac{\mathrm{d}^j u\left(1+v_p\right)}{\mathrm{d}r_p^j}$$

Mean-variance makes the approximation (the component related to $u^{(1)}(\cdot)$ cancels out)

$$E\left[u\left(R_p\right)\right] \simeq u\left(1+v_p\right) + \frac{u^{(2)}\left(1+v_p\right)}{2}\sigma_p^2.$$

Accuracy of approximation depends on utility function and distribution of returns.

Let us study briefly non-normality in return data. Recall comment on elliptical distributions (skewness vs. kurtosis).

▶ Single-period returns: Visualizing and testing non-normality
Returns are not normal.

Scherer (2004): Figure 2.1

Not only because of skewness and asymmetry in univariate distributions, also because of asymmetric correlation in joint distributions.

Scherer (2004): Table 1.5

▶ Statistical significance of deviations: Statistics for testing asymmetry and fat tails

$$skewness = \frac{1}{T}\sum_{t=1}^{T}\frac{(r_t-\bar{r})^3}{\hat{\sigma}^3}, \qquad kurtosis = \frac{1}{T}\sum_{t=1}^{T}\frac{(r_t-\bar{r})^4}{\hat{\sigma}^4}-3,$$
$$JB = T\left(\frac{skewness^2}{6}+\frac{kurtosis^2}{24}\right)\sim\chi^2(2).$$

Scherer (2004): Table 2.1

These tests rely on i.i.d. observations, which is not the case of return data⇒But there are versions for time series data. See Bai and Ng (2001).

▶ Normality and multi-period returns:
Central limit theorem gives convergence to normality, but there might be problems in its application to returns. Specifically, dependence and unbounded variance.

Scherer (2004): Figure 2.3-4

▶ Modelling non-normality with a mixture of normal distributions:
Combination of 2 regimes of high and low volatility (and different means) to get skewness and kurtosis

$$f_{mix}(r) = pf_{high}(r)+(1-p)f_{low}(r),$$
$$f_{high}(r) = \frac{1}{\sqrt{2\pi}\sigma_{high}}\exp\left(-\frac{1}{2}\left(\frac{r-v_{high}}{\sigma_{high}}\right)^2\right),$$
$$f_{low}(r) = \frac{1}{\sqrt{2\pi}\sigma_{low}}\exp\left(-\frac{1}{2}\left(\frac{r-v_{low}}{\sigma_{low}}\right)^2\right).$$

Scherer (2004): Figure 1.5, 2.5

We can also introduce persistence in regimes⇒Markov-switching models. See Ang and Bekaert (2002).

We will dedicate one chapter to study how to deal with non-normality.

- How well does the one-period solution approximate multi-period optimality?

Campbell and Viceira (2002): Figure 1.1; Table 1.1

►Assuming no rebalancing in a long-horizon context, let us see two fallacies of long-term choice

►►Does mean-variance frontier change as investment horizon lengthens?
Assuming uncorrelated and normally distributed returns⇒Frontier does not change with horizon T

$$w_{mv}\left(\lambda\right) \simeq \lambda\frac{\mu T}{\sigma^2 T} = \lambda\frac{\mu}{\sigma^2}.$$

Caveat: Sharpe ratio rises with horizon H

$$SR \simeq \sqrt{T}\frac{\mu}{\sigma},$$

but risk does not diversify across time.

Campbell and Viceira (2002): Figure 4.2

►►Should all long-term investors maximize geometric return ?
Growth-optimal portfolio outperforms any other portfolio in the long run with probability reaching one.
But this is only optimal for particular preferences: Log-utility

$$E\left[u\left(R_p\right)\right] = E\left[\ln\left(R_p\right)\right].$$

Sometimes this portfolio gives a great loss and this really hurts an investor with higher risk aversion.
Intuition in a mean-variance context: A portfolio geometric return can be approximated as

$$g_p \simeq \mu_p - \frac{1}{2}\sigma_p^2,$$

so its maximization is only optimal for a risk tolerance of 1, which represents a fairly aggressive investor.

▶Allowing rebalancing during the investment horizon,

▶▶Does repeatedly investing in myopic (one-period efficient) portfolios result in multi-period efficient portfolios?

The optimal portfolio has an interesting decomposition in continuous time. If there is only one state variable that drives the conditional distribution of returns (such as short-term rate or dividend yield) then

$$\mathbf{w}^* = a\mathbf{m} + b\mathbf{h}$$

- The myopic component $a\mathbf{m}$ is the one-period solution⇒ $\mathbf{m}$ is mean-variance portfolio and a is related to relative risk tolerance.
- Investor tries to hedge against changing investment opportunities through hedging component $b\mathbf{h}$ ⇒ $\mathbf{h}$ is portfolio with maximum correlation with state variable and b is a measure of aversion to changes in state variable.

Recall that liabilities introduced a hedging component in the optimal ALM portfolio⇒Now we have an intertemporal hedging demand instead of a liability hedging demand.

Campbell and Viceira (2002): Figure 3.3, 4.1

Equivalence of intertemporal and myopic strategies only holds under one of the following contexts:

▶▶▶Investment opportunities are not stochastic: $\mathbf{h} = \mathbf{0}$ ⇒Nothing to hedge.

▶▶▶Preferences are represented by log-utility: $b = 0$ ⇒No effect of state variables on marginal utility of wealth.

We will not study multi-period choice in this course. It would require an introduction to continuous time mathematics and models of return dynamics. See Campbell and Viceira (2002).

We will only study some dynamic strategies in the last chapter.

2 Black-Litterman Approach to Estimation Error

- Estimation Error in Mean-Variance Inputs

- Bayesian Portfolio Choice

- Combination of Equilibrium and Views

References:

• Black, F. and R. Litterman (1992): Global Portfolio Optimization, *Financial Analysts Journal*, September-October, 28-43.

• Brandt, .M W. (2004): Portfolio Choice Problems, in Y. Ait-Sahalia and L.P. Hansen (eds.), *Handbook of Financial Econometrics*, forthcoming.

Litterman, B. (2003): *"Modern Investment Management: An Equilibrium Approach"*, Wiley.

Scherer, B. (2004): *"Portfolio Construction and Risk Budgeting"*, Risk Books.

2.1 Estimation Error in Mean-Variance Inputs

We use estimated parameters from a time series with T observations⇒Estimation error in mean-variance inputs.

First we will study 3 simple approaches to handle estimation error.

- Portfolio Resampling:

This is a heuristic method (no theoretical justification) to deal with estimation error.

We can visualize the effect of estimation error on optimal portfolios by means of portfolio resampling, a Monte Carlo procedure⇒Translation of input $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ uncertainty into output $\mathbf{w}_{mv}(\lambda)$ uncertainty.

If our sample estimates of mean-variance parameters are $\left(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}\right)$ we can simulate artificial samples from a normal distribution with that mean and variance.

Let us assume we simulate S samples of size T, then we can compute the corresponding estimates for each sample

$$\left(\hat{\boldsymbol{\mu}}_1, \hat{\boldsymbol{\Sigma}}_1\right), \left(\hat{\boldsymbol{\mu}}_2, \hat{\boldsymbol{\Sigma}}_2\right), ..., \left(\hat{\boldsymbol{\mu}}_S, \hat{\boldsymbol{\Sigma}}_S\right),$$

which define S different mean-variance frontiers

$$\mathbf{w}_s(\lambda) = \lambda \hat{\boldsymbol{\Sigma}}_s^{-1} \hat{\boldsymbol{\mu}}_s, \qquad s = 1, ..., S.$$

Those frontier portfolios are not efficient when evaluated under $\left(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}\right)$.

Scherer (2004): Figure 3.1-2

Resampled efficiency: We can average simulated mean-variance efficient portfolios⇒Resampled frontier.

Resampled portfolio m, associated to some return level μ_m or risk tolerance λ_m, is then

$$\mathbf{w}_m^{resampled} = \frac{1}{S} \sum_{i=1}^{S} \mathbf{w}_{mi}.$$

Resample portfolios show higher diversification and less sharp changes than traditional mean-variance portfolios.

Scherer (2004): Table 3.7-8

There are many important pitfalls of portfolio resampling. For instance, resamplings are derived from sample estimates. We will focus on different approaches.

- Link to linear regression:

There is a more formal approach to measure uncertainty about optimal portfolio weights. Britten-Jones (1999) uses the simple linear regression

$$\mathbf{1} = \mathbf{E}\mathbf{w} + \mathbf{u},$$
$$\hat{\mathbf{w}} = \left(\mathbf{E}'\mathbf{E}\right)^{-1}\mathbf{E}'\mathbf{1} = \left(\frac{1}{1 + \hat{\boldsymbol{\mu}}'\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}}}\right)\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}},$$

where $\mathbf{E}$ is a $T \times N$ matrix with our data on excess returns and $\mathbf{1}$ is a vector of ones.

The regression framework gives the corresponding standard errors of $\hat{\mathbf{w}}$ to measure uncertainty and construct tests.

We only have to rescale the regression coefficients to get the efficient portfolio $\mathbf{w}_{mv}(\lambda)$.estimator

$$\hat{\mathbf{w}}_{mv}(\lambda) = \lambda\hat{\boldsymbol{\Sigma}}^{-1}\hat{\boldsymbol{\mu}} \propto \hat{\mathbf{w}}.$$

This approach is based on classic inference⇒We will focus on Bayesian inference in next sections.

- Constrained Optimization

An approach to avoid unreasonable mean-variance portfolios: Impose constraints to enforce diversification,

$$\max_{\mathbf{w}} E\left(e_p\right) - \frac{1}{2\lambda} Var\left(e_p\right) \qquad s.t. \qquad \mathbf{w}^{lower} \leq \mathbf{A}\mathbf{w} \leq \mathbf{w}^{upper},$$

for some matrix $\mathbf{A}$.

Effect on portfolio choice: Constraints change the distribution of implied views⇒Higher conviction on forecasts, but constraints where introduced because of estimation error.

Scherer (2004): Figure 3.14

2.2 Bayesian Portfolio Choice

This is a formal way of dealing with estimation error: Strong foundation in decision theory.

- Theoretical foundations:

Frequentist vs. Bayesian approach to statistical inference: Objective vs. subjective interpretation of probability (frequency in repeated sampling vs. degree of belief)⇒Estimators vs. posteriors.

Standard i.i.d. normal context with known σ^2

$$\bar{e} \sim N\left(\mu, \frac{\sigma^2}{T}\right) \qquad vs. \qquad \mu \mid \mathbf{e} \sim N\left(\bar{e}, \frac{\sigma^2}{T}\right)$$

Confidence interval vs. probability interval in this context⇒Same answer to different questions.

Let us assume that we have a time series of a particular asset's excess returns $(e_1, e_2, ..., e_T)$, which we represent by a $T \times 1$ vector $\mathbf{e}$, and we want to learn about $\left(\mu, \sigma^2\right)$, which we represent by parameter vector $\boldsymbol{\theta}$.

To fix ideas, we assume again a simple normal model with i.i.d. observations

$$\mathbf{e} \mid \boldsymbol{\theta} \sim N\left(\mu\mathbf{1}, \sigma^2\mathbf{I}\right),$$

where $\mathbf{1}$ is a vector of ones and $\mathbf{I}$ is an identity matrix.

The Bayesian approach needs two inputs and produces two outputs:

· Inputs: Prior beliefs $p\left(\boldsymbol{\theta}\right)$ and data likelihood $p\left(\mathbf{e} \mid \boldsymbol{\theta}\right)$.

· Outputs: Posterior beliefs $p\left(\boldsymbol{\theta} \mid \mathbf{e}\right)$ and predictive distribution $p\left(e_{T+1} \mid \mathbf{e}\right)$.

Let us study both outputs.

▶ Posterior beliefs:

Following Bayes' rule, posterior beliefs $p\left(\boldsymbol{\theta} \mid \mathbf{e}\right)$ are a combination of prior beliefs $p\left(\boldsymbol{\theta}\right)$ and sample information given by data likelihood $p\left(\mathbf{e} \mid \boldsymbol{\theta}\right)$

$$p\left(\boldsymbol{\theta} \mid \mathbf{e}\right) \propto p\left(\boldsymbol{\theta}\right) p\left(\mathbf{e} \mid \boldsymbol{\theta}\right),$$

where $\propto$ means proportional to. The notation $p\left(y \mid x\right)$ means the distribution of y conditional on x.

▶ Predictive distribution:

In the context of asset allocation, we are interested in the distribution of future excess returns given our data. We will focus on next period e_{T+1} and distribution of interest $p\left(e_{T+1} \mid \mathbf{e}\right)$. This object is called predictive distribution and can be computed as follows

$$p\left(e_{T+1} \mid \mathbf{e}\right) = \int_{\mathbb{R}^2} p\left(e_{T+1} \mid \boldsymbol{\theta}, \mathbf{e}\right) p\left(\boldsymbol{\theta} \mid \mathbf{e}\right) \mathrm{d}\boldsymbol{\theta}.$$

To fix ideas, the previous normal model states that

$$e_{T+1} \mid \boldsymbol{\theta}, \mathbf{e} \sim N\left(\mu, \sigma^2\right).$$

If there is not an analytical expression for the predictive distribution, as it is often the case, we can still draw from it via Monte Carlo methods$\Rightarrow$Simulate $\boldsymbol{\theta}$ from the posterior and then draw e_{T+1} from $p\left(e_{T+1} \mid \boldsymbol{\theta}, \mathbf{e}\right)$ evaluated at that $\boldsymbol{\theta}$.

Note that we have only dealt with parameter uncertainty, but the Bayesian approach can also deal with model uncertainty by means of Bayesian model averaging. See Avramov (2002) and Tu and Zhou (2004).

Bayesian inference is the approach advocated in decision theory but analytical expressions are available only for few simple cases such as two examples that will study below.

We need complex tools (Markov Chain Monte Carlo methods) in general and they must be applied on a case by case basis. See Barberis (2000) for an application to multi-period choice and non-i.i.d. returns.

- Noninformative priors:

Let us assume

$$\mathbf{e} \mid \boldsymbol{\theta} \sim N\left(\mu\mathbf{1},\sigma^2\mathbf{I}\right),$$

and an uninformative prior

$$p\left(\boldsymbol{\theta}\right) = \sigma^{-2}.$$

In this context, the posterior is

$$\sigma^2 \mid \mathbf{e} \sim IG\left(\frac{T-1}{2},\frac{1}{2}\mathbf{u}'\mathbf{u}\right), \qquad \mathbf{u} = \mathbf{e}-\bar{e}\mathbf{1},$$

$$\mu \mid \mathbf{e},\sigma^2 \sim N\left(\bar{e},\frac{\sigma^2}{T}\right),$$

where $IG\left(\cdot\right)$ means an inverse gamma distribution.

We can also integrate out σ^2 in the last posterior

$$\mu \mid \mathbf{e} \sim t\left(T-1;\bar{e},\frac{S^2}{T}\right), \qquad S^2 = \frac{1}{T-1}\mathbf{u}'\mathbf{u},$$

where $t\left(\cdot\right)$ means a Student-t distribution (its first parameter is degrees of freedom).

We can simulate the corresponding predictive distribution of some future e_{T+1} as it was commented above. First, sample $\left(\mu,\sigma^2\right)$ from posterior. Second, sample e_{T+1} from the corresponding $N\left(\mu,\sigma^2\right)$.

But we have an explicit distribution in this simple case

$$e_{T+1} \mid \mathbf{e} \sim t\left(T-1;\bar{e},S^2+\frac{S^2}{T}\right).$$

Scherer (2004): Figure 4.1

- Adding prior information:

Let us assume we know σ^2 and therefore we only need to learn about μ. In addition, we have the following prior

$$p(\mu) \sim N(\rho_0, \varphi_0),$$

while $p(\mathbf{e} \mid \boldsymbol{\theta})$ is the same as before before.

Our posterior is then

$$p(\mu \mid \mathbf{e}) \sim N(\rho_T, \varphi_T),$$

$$\varphi_T = \left(\frac{1}{\varphi_0} + \frac{T}{\sigma^2}\right)^{-1}, \qquad \rho_T = \varphi_T \left(\frac{\rho_0}{\varphi_0} + \frac{T\bar{e}}{\sigma^2}\right).$$

Properties:
▶Importance of prior is inversely related to time series length.
▶Volatile assets are more amenable to the application of prior information.
▶Highly uncertain prior results in little importance of prior.

Scherer (2004): Figure 4.2-4

We can simulate the corresponding predictive distribution of some future e_{T+1} as in previous case. First, sample μ from posterior. Second, sample e_{T+1} from the corresponding $N\left(\mu, \sigma^2\right)$.

But we also have an explicit distribution in this simple case

$$e_{T+1} \mid \mathbf{e} \sim N\left(\rho_T, \sigma^2 + \varphi_T\right).$$

2.3 Combination of Equilibrium and Views

Traditional mean-variance portfolios are badly behaved when using sample estimates $\left(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}}\right)$ as inputs ⇒We will study an approach developed in Black and Litterman (1992) that trusts the sample variance as a good estimator but not the sample mean.

Recall the comments on sample estimators of mean and variance of returns.

This approach represents a simple and convenient approximation to Bayesian portfolio choice: It combines an equilibrium model with subjective views on risk premia ⇒The use of an equilibrium model links the chosen portfolio to the market portfolio, which is well behaved.

Black and Litterman (1992): Table 1, 3.

- Inputs:

▶ Sample information is represented by an equilibrium model

$$\mathbf{e} \sim N\left(\boldsymbol{\pi}, \tau\hat{\boldsymbol{\Sigma}}\right),$$

where

· $\boldsymbol{\pi}$ are implied risk premia from market portfolio under the CAPM,

$$\boldsymbol{\pi} = \left(\frac{1}{\lambda}\right)\hat{\boldsymbol{\Sigma}}\mathbf{w}_{market}.$$

· $\hat{\boldsymbol{\Sigma}}$ is the estimated variance matrix of excess returns

· τ reflects the degree of confidence on equilibrium returns (negative link).

Note that we do not use a sample estimate of risk premia⇒Bayesian?

▶ On the other hand, prior information is expressed as some subjective views on risk premia. Defining $\mathbf{Q}$ as a $K \times N$ matrix ($K \leq N$) that reflects investor's views,

$$\mathbf{Q}\boldsymbol{\mu} = \mathbf{q} + \boldsymbol{\eta}, \qquad \boldsymbol{\eta} \sim N\left(\mathbf{0}, \boldsymbol{\Omega}\right),$$

where $\boldsymbol{\Omega}$ reflects the degree of confidence on those views.

Note the difference to previous Bayesian example of known σ^2. There we used:
▶ A prior $p(\mu)$, where we can introduce the CAPM as our prior belief on risk premia.
▶ A likelihood $p(\mathbf{e} \mid \boldsymbol{\theta})$, which introduces sample data on risk premia.
See Pástor (2000) for this approach.

- Outputs:

▶ Both pieces of information together yield the following posterior risk premia

$$\boldsymbol{\mu}_{post} = \left[\mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{Q} + \left(\tau\hat{\boldsymbol{\Sigma}}\right)^{-1}\right]^{-1} \left[\mathbf{Q}'\boldsymbol{\Omega}^{-1}\mathbf{q} + \left(\tau\hat{\boldsymbol{\Sigma}}\right)^{-1}\boldsymbol{\pi}\right],$$

which is similar to previous expressions in a general Bayesian framework.

We might interpret $\boldsymbol{\mu}_{post}$ in a GLS framework. We can join the two sources of information in a system like two samples

$$\begin{pmatrix} \mathbf{q} \\ \boldsymbol{\pi} \end{pmatrix} = \begin{pmatrix} \mathbf{Q} \\ \mathbf{I} \end{pmatrix} \boldsymbol{\mu} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix},$$

so that we can define

$$\mathbf{X} = \begin{pmatrix} \mathbf{Q} \\ \mathbf{I} \end{pmatrix}, \qquad \mathbf{y} = \begin{pmatrix} \mathbf{q} \\ \boldsymbol{\pi} \end{pmatrix}, \qquad \mathbf{V} = \begin{pmatrix} \boldsymbol{\Omega} & \mathbf{0} \\ \mathbf{0} & \tau\hat{\boldsymbol{\Sigma}} \end{pmatrix},$$

and finally get the GLS estimator as

$$\boldsymbol{\mu}_{post} = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\right).$$

▶ The portfolio advocated by this approach is

$$\mathbf{w}_{mv}(\lambda) = \lambda\hat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\mu}_{post}.$$

Black and Litterman (1992): Table 6, 7, 13, 14.

This approach anchors portfolio choice to market portfolio: Low confidence in subjective views with respect to equilibrium (or simply neutral views)⇒Investor holds the market portfolio.

Scherer (2004): Figure 4.5

Portfolio Management: New Methods

Francisco Peñaranda (University of Alicante and FMG/LSE)
<penaranda@merlin.fae.ua.es>

3 Portfolios of Non-Normal Returns

- From Variance to Lower Partial Moments

- Introducing Skewness and Kurtosis

- Performance Measures: Alternatives to Sharpe Ratio

References:

• Jondeau, E., and M. Rockinger (2004): The Allocation of Assets under Higher Moments, mimeo.
Keating, C., and F. Shadwick (2002): A Universal Performance Measure, *Journal of Performance Measurement* 6, ?-?.
• Scherer, B. (2004): *"Portfolio Construction and Risk Budgeting"*, Risk Books.
Stutzer, M. (2000): A Portfolio Performance Index, *Financial Analysts Journal* 56, 52-61.

3.1 From Variance to Lower Partial Moments

Lower partial moments (LPM) are a convenient way to work under non-normality (asymmetry) since the optimization framework is similar to the traditional one⇒Only change concept of risk to downside risk: Falling below a target.

- Definition of LPM:

General decomposition of returns

$$r = h + \max(r - h, 0) - \max(h - r, 0).$$

LPMs are a measure of downside risk. Given a density $f(r)$,

$$LPM(h,k) = E\left[\max\left((h-r)^k, 0\right)\right]$$
$$= \int_{-\infty}^{h} (h-r)^k f(r)\,\mathrm{d}r = \Pr(r \le h)\, E\left[(h-r)^k \mid r \le h\right].$$

where

▶ h is a target value: Risk-free rate, benchmark,...

▶ k is linked to risk aversion: 0, 1, 2,....⇒ $k = 1$ separates risk-seeking from risk-averse behaviour for $r \le h$.

- $LPM(h,0)$ is simply $\Pr(r \le h)$ like Roy's safety-first approach.
- $LPM(h,1)$ takes into account $E[r \mid r \le h]$ too.
- We will focus on $LPM(h,2)$, which is called (target) semivariance $SV(h)$.

Scherer (2004): Table 2.3
Caveat: Scherer (2004) uses $LPM(h,k) = \int_{-\infty}^{h} (r-h)^k f(r)\,\mathrm{d}r$.

Theoretical appealing: The use of LPM as a measure of risk can be linked to stochastic dominance (studying every h), which orders portfolios for general utility functions⇒That is not the case with variance.

See Bawa and Lindenberg (1977).

We can relate this set-up to expected utility: The corresponding preferences are

$$u\left(R_p\right) = \begin{cases} r_p, & r_p > h \\ r_p - \gamma\left(h - r_p\right)^2, & r_p \leq h \end{cases} \Rightarrow E\left[u\left(R_p\right)\right] = \nu_p - \gamma SV_p\left(h\right),$$

for some $\gamma > 0$.

- Portfolio choice when $h = c$:

Then we can express the optimal portfolio as the solution to

$$\min_{\mathbf{w}} SV_p\left(c\right) \qquad s.t. \qquad E\left(e_p\right) = \mu_p \Rightarrow \mathbf{w}_{msv}\left(\mu_p\right).$$

Note that the criterion function is convex and the feasible set is convex too⇒Standard optimization, we know efficient frontier $SV_p\left(c\right)$ is convex in μ_p, but without explicit solution.

Following Bawa and Lindenberg (1977), we find a similar outcome to mean-variance approach (if $\sqrt{SV_p\left(c\right)}$ is convex): Linearity and separation

▶ The efficient frontier in space $\left(\mu_p, \sqrt{SV_p\left(c\right)}\right)$ is linear.

▶ Every efficient portfolio is a portfolio of c and a particular portfolio of risky assets (market portfolio).

Bawa and Lindenberg (1977): Figure 1, 2

(General h ⇒Efficient frontier $SV_p\left(h\right)$ is convex in μ_p too and similar separation. See Harlow and Rao (1989))

- Comparison of lower partial moments-based and variance-based portfolio choice:

Scherer (2004): Figure 2.8

Differences increase with skewness and decrease with return requirements and threshold.

- Correct estimation of lower partial moments:

The corresponding ex-post expression of $SV(h)$ is

$$sv(h) = \frac{1}{T}\sum_{t=1}^{T}\left(h - r_t\right)^2 I\left(r_t \leq h\right),$$

where $I(\cdot)$ is the indicator function (1 if event is true, 0 otherwise).

Scherer (2004): Table 2.4

Estimation error can be a serious problem with LPMs⇒Using empirical distribution function as in previous expressions might not be accurate enough.

Scherer (2004): Figure 2.6

One way that might improve estimation is using a fitted distribution $\hat{f}(r)$ instead

$$\mathrm{sv}(h) = \int_{-\infty}^{h}\left(h - r\right)^2 \hat{f}(r)\,\mathrm{d}r.$$

3.2 Introducing Skewness and Kurtosis

We can use a Taylor expansion of expected utility to easily handle asset allocation under nonnormal returns. This methodology is developed in Jondeau and Rockinger (2004).

- Portfolio allocation for general utility:

Investor's problem is

$$\max_{\mathbf{w}} E\left[u\left(R_p\right)\right],$$

where the solution is usually characterized by a set of FOCs.

Its solution might be difficult to compute since we must evaluate the integral that defines $E\left[u\left(R_p\right)\right]$. We will study now an approximation that makes that evaluation simple.

- Approximation by fourth-order Taylor series expansion:

We already saw that we can translate expected utility to moments by a Taylor expansion (assuming convergence)

$$E\left[u\left(R_p\right)\right] = E\left[\sum_{j=0}^{\infty} \frac{u^{(j)}\left(1+v_p\right)}{j!}\left(r_p - v_p\right)^j\right] = \sum_{j=0}^{\infty} \frac{u^{(j)}\left(1+v_p\right)}{j!} E\left[\left(r_p - v_p\right)^j\right],$$

and that mean-variance analysis is an approximation that only takes into first two components.

Now we will also take into account another two components

$$\begin{gathered} E\left[u\left(R_p\right)\right] \simeq u\left(1+v_p\right) + \frac{u^{(2)}\left(1+v_p\right)}{2}\sigma_p^2 + \frac{u^{(3)}\left(1+v_p\right)}{3!}s_p + \frac{u^{(4)}\left(1+v_p\right)}{4!}\kappa_p, \\ s_p = E\left[\left(r_p - v_p\right)^3\right], \qquad \kappa_p = E\left[\left(r_p - v_p\right)^4\right], \end{gathered}$$

which are moments that represent skewness and kurtosis.

▶ Application to a constant absolute risk aversion (CARA) utility function:
That function is

$$u(R_p) = -\exp(-\lambda R_p),$$

which allows convergence of Taylor expansion.

The approximation becomes

$$E[u(R_p)] \simeq -\exp(-\lambda(1+v_p))\left[1+\frac{\lambda^2}{2}\sigma_p^2-\frac{\lambda^3}{3!}s_p+\frac{\lambda^4}{4!}\kappa_p\right].$$

We can also compute analytical FOCs from that approximation.

▶ Application to a constant relative risk aversion (CRRA) utility function:
That function is

$$u(R_p) = \begin{cases} \frac{R_p^{1-\gamma}}{1-\gamma}, & \gamma \geq 0 \\ \ln(R_p), & \gamma = 1 \end{cases},$$

which allows convergence of Taylor expansion at $R_p \in [0, 2(1+v_p)]$.

The approximation becomes

$$E[u(R_p)] \simeq \frac{(1+v_p)^{1-\gamma}}{1-\gamma} - \frac{\gamma}{2}(1+v_p)^{-(\gamma+1)}\sigma_p^2 + \frac{\gamma(\gamma+1)}{3!}(1+v_p)^{-(\gamma+2)}s_p - \frac{\gamma(\gamma+1)(\gamma+2)}{4!}(1+v_p)^{-(\gamma+3)}\kappa_p$$

for $\gamma \neq 1$ and

$$E[u(R_p)] \simeq \ln(1+v_p) - \frac{1}{2(1+v_p)^2}\sigma_p^2 + \frac{1}{3(1+v_p)^3}s_p - \frac{1}{4(1+v_p)^4}\kappa_p$$

for $\gamma = 1$.

We can also compute analytical FOCs from that approximation.

- Solution to approximated asset-allocation problem:

Given the previous set-up, we only need the computation of relevant moments to construct the optimal portfolio. Defining

$$\mathbf{u} = \mathbf{r} - \boldsymbol{\nu},$$

we can express the relevant moments as

$$\begin{gathered}
\mathbf{M}_1 = E\left[\mathbf{u}\right] = \mathbf{0}, \\
\mathbf{M}_2 = E\left[\mathbf{u}\mathbf{u}'\right] = \boldsymbol{\Sigma}, \\
\mathbf{M}_3 = E\left[\mathbf{u}' \otimes \mathbf{u}\mathbf{u}'\right], \\
\mathbf{M}_4 = E\left[vec'\left(\mathbf{u}\mathbf{u}'\right) \otimes \mathbf{u}\mathbf{u}'\right],
\end{gathered}$$

where $\otimes$ is Kronecker product and $vec\left(\cdot\right)$ vectorizes a matrix. $\mathbf{M}_3$ is an $N \times N^2$ matrix of co-skewness and $\mathbf{M}_4$ is an $N \times N^3$ matrix of co-kurtosis.

The previous notation help us to write portfolio moments as

$$\begin{gathered}
v_p = c + \mathbf{w}'\boldsymbol{\mu}, \\
\sigma_p^2 = \mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}, \\
s_p = \mathbf{w}'\mathbf{M}_3\left(\mathbf{w} \otimes \mathbf{w}\right), \\
\kappa_p = \mathbf{w}'\mathbf{M}_4\left(\mathbf{w} \otimes \mathbf{w} \otimes \mathbf{w}\right),
\end{gathered}$$

and the FOCs are easily computed from derivatives of the previous expressions.

Jondeau and Rockinger (2002): Table 2-3; Figure 1-3

3.3 Performance Measures: Alternatives to Sharpe Ratio

Recall that Sharpe ratio of a portfolio is

$$SR_p = \frac{\mu_p}{\sigma_p}.$$

It is justified as a performance measure under normal returns since then the efficient frontier in the $\left(\mu_p, \sigma_p\right)$ space is a straight line giving a maximum Sharpe ratio

$$SR = \sqrt{H}.$$

There are other standard performance measures under normality (Treynor index and Jensen's alpha), but Sharpe ratio will be our benchmark here.

Alternative performance measures when returns are not normal (asymmetric).

They can also be used as a criterion for portfolio choice⇒Choose portfolio that maximizes one of those measures.

- Sortino ratio:

The Sortino ratio is very similar to the Sharpe ratio, but it is based on semivariance for some target h as a measure of risk

$$S_p\left(h\right) = \frac{v_p - h}{\sqrt{SV_p\left(h\right)}}.$$

There is a modified version with better theoretical foundations, fixing $h = c$

$$S\left(c\right) = \frac{\mu_p}{\sqrt{SV_p\left(c\right)}}.$$

In that context, we saw that the $\left(\mu_p, \sqrt{SV_p\left(c\right)}\right)$ frontier satisfies the properties of two fund spanning and linearity.

See Pedersen and Satchell (2002).

- Omega function:

The Omega function is defined in Keating and Shadwick (2002). It uses lower and upper partial moments to measure the risk-return trade-off.

For some target h, it is defined as

$$\Omega_p(h) = \frac{\int_h^{+\infty} (1 - F(r_p)) \, \mathrm{d}r_p}{\int_{-\infty}^{h} F(r_p) \, \mathrm{d}r_p} = \frac{E[\max(r_p - h, 0)]}{E[\max(h - r_p, 0)]} = \frac{UPM_p(h,1)}{LPM_p(h,1)},$$

where $F(\cdot)$ is the portfolio return's cumulative distribution function.

Keating and Shadwick (2002): Diagram 1.1-3; Diagram 2.5-8

Some properties of this function are

$$\frac{\mathrm{d}\Omega_p(h)}{\mathrm{d}h} < 0,$$
$$\Omega_p(v_p) = 1.$$

Keating and Shadwick (2002): Page 18,20-22,31; Diagram 4.1-3

- Stutzer index:

Stutzer (2000) develops a performance measure that is based on the hypothesis that a fund manager is averse to receiving a nonpositive time-averaged excess return above some benchmark. We will assume the benchmark is the risk-free asset and then its objective is to minimize the probability of underperformance

$$\Pr\left(\bar{e}_p \leq 0\right).$$

Note that this is like a time series version of Roy's safety-first approach.

If portfolio returns are such that $\mu_p > 0$ and we have a long i.i.d. (this can be generalized) time series then a law of large numbers shows

$$\Pr\left(\bar{e}_p \leq 0\right) \simeq 0,$$

which is not very useful to order portfolios.

Large deviation theory is more informative since it gives the rate of decay of that probability, which is related to the moment generating function of e_p

$$\begin{gathered} \Pr\left(\bar{e}_p \leq 0\right) \simeq \exp\left(-I_p T\right), \\ I_p = \max_{\theta} - \ln E\left[\exp\left(\theta e_p\right)\right]. \end{gathered}$$

The rate of decay $I_p \geq 0$ is the proposed performance index: $I_p > 0$ if $\mu_p > 0$, and $I_p = 0$ otherwise.

Note that I_p is independent of portfolio scale and we need a well defined moment generating function for the application of this theory.

It gives the same ranking as the Sharpe ratio under normality because in that case

$$I_p = \frac{1}{2} SR_p^2.$$

Obviously, we do not know the true moment generating function, but we can compute an ex-post performance index

$$\hat{I}_p = \max_{\theta} - \ln\left(\frac{1}{T}\sum_{t=1}^{T}\exp\left(\theta e_{pt}\right)\right)$$

Stutzer (2000): Table 1-2

See Stutzer (2003) for a new application of this framework.

Portfolio Management: New Methods

Francisco Peñaranda (University of Alicante and FMG/LSE)
<penaranda@merlin.fae.ua.es>

4 Benchmark-Relative Management

- General Framework: Background Risks

- Tracking Error Optimization

- Tracking Error vs. Mean-Variance Efficiency

References:

Grinold, R.C., and R.N. Kahn (1999): *"Active Portfolio Management"*, McGraw-Hill.
• Roll. R. (1992): A Mean-Variance Analysis of Tracking Error, *Journal of Portfolio Management* 18, 13-22
Scherer, B. (2004): *"Portfolio Construction and Risk Budgeting"*, Risk Books.
• Solnik, B. (1978): Inflation and Optimal Portfolio Choices, *Journal of Financial and Quantitative Analysis* 13, 903-925.

4.1 General Framework: Background Risks

- Motivation:

Background risks, with rate of return r_b, cover several situations:

▶ Liabilities in an ALM, which was introduced in a previous chapter.

▶ Fixed assets such as human capital.

▶ Inflation when managing real returns.

▶ Benchmark in active management.

Solnik (1978) studies the case of inflation and real returns but it will be the basis for our analysis of background risks in general.

Instead of managing the risk-return of r_p, we manage the risk-return of $r_p - r_b$ ⇒That difference can be surplus, real return or active return depending on the context.

We introduce new notation

$$\begin{aligned} \omega_p^2 &= Var\left(r_p - r_b\right), \qquad E\left(r_p - r_b\right) = \delta_p, \\ \boldsymbol{\gamma} &= Cov\left(\mathbf{e}, r_b\right), \qquad F = \boldsymbol{\gamma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}. \end{aligned}$$

- Background risks in a mean-variance framework:

The optimal portfolio is now given by

$$\min_{\mathbf{w}} Var\left(r_p - r_b\right) \qquad s.t. \qquad E\left(r_p - r_b\right) = \delta_p.$$

We can express the portfolio problem as

$$\min_{\mathbf{w}} Var\left(e_p\right) - 2Cov\left(e_p, r_b\right) \qquad s.t. \qquad E\left(e_p\right) = \mu_p$$
$$\Rightarrow \mathbf{w}_{br}\left(\mu_p\right) = \left(\frac{\mu_p - F}{H}\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}.$$

There is a useful representation of optimal portfolios

$$\begin{aligned} \mathbf{w}_{br}\left(\mu_p\right) &= \left(\frac{\mu_p}{H}\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \left[\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma} - \left(\frac{F}{H}\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}\right] \\ &= \mathbf{w}_{mv}\left(\mu_p\right) + \mathbf{w}_h. \end{aligned}$$

They are equal to the traditional optimal portfolio for μ_p plus a constant term, a hedging demand due to background risk.

In terms of rates of return,

$$r_{br} = r_{mv} + r_h.$$

Note that r_h is independent of chosen μ_p (its cost is not zero) and

$$E\left(r_h\right) = 0, \qquad Cov\left(r_{mv}, r_h\right) = 0.$$

There is two-fund spanning of mean-variance frontier, but none of the two funds is the safe asset.

- Background risk efficient frontier:

The frontier is a hyperbola in the (δ_p, ω_p) space.
But it is more interesting to compare background risk and traditional frontiers in the $\left(\mu_p, \sigma_p\right)$ space.

Using the previous representation, we have the decomposition

$$Var\left(r_{br}\right) = Var\left(r_{mv}\right) + Var\left(r_h\right),$$

so the background risk frontier represents a parallel parabola in the $\left(\mu_p, \sigma_p^2\right)$ space, while it represents a hyperbola with asymptotes equal to traditional line in the $\left(\mu_p, \sigma_p\right)$ space.

Solnik (1978): Figure 4

Cases when background risk and asset-only efficient frontiers coincide:

► Background risks are cash or covariance $\boldsymbol{\gamma}$ between assets and them is $\mathbf{0}$.

►The covariance of assets with background risks is proportional to $\boldsymbol{\mu}$.

►There exists a background risk mimicking asset, $r_b = c + \mathbf{w}_b'\mathbf{e}$, and it lies on the efficient frontier.
This last case of risk as a portfolio will be the focus of next sections.

- Similar results if there is not a safe asset:

Now

$$r_p = \mathbf{w}'\mathbf{r}.$$

We need some additional notation

$$B_b = \boldsymbol{\nu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}, \qquad C_b = \mathbf{1}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma}.$$

The optimal frontier is given by

$$\min_{\mathbf{w}} Var\left(r_p\right) - 2Cov\left(r_p, r_b\right) \qquad s.t. \qquad E\left(r_p\right) = \nu_p, \qquad \mathbf{w}'\mathbf{1} = 1$$
$$\Rightarrow \mathbf{w}_{br}\left(\nu_p\right) = \mathbf{w}_{mv}\left(\nu_p\right) + \mathbf{w}_h,$$
$$\mathbf{w}_h = \boldsymbol{\Sigma}^{-1}\boldsymbol{\gamma} - \frac{1}{D}\left(BC_b - AB_b\right)\boldsymbol{\Sigma}^{-1}\mathbf{l} - \frac{1}{D}\left(CB_b - AC_b\right)\boldsymbol{\Sigma}^{-1}\boldsymbol{\nu},$$

where $\mathbf{w}_{mv}\left(\nu_p\right)$ is the traditional optimal portfolio without a safe asset.

In terms of rates of return,

$$r_{br} = r_{mv} + r_h,$$

and r_h is independent of chosen v_p, its cost is zero, and

$$E\left(r_h\right) = 0, \qquad Cov\left(r_{mv}, r_h\right) = 0.$$

There is also two-fund spanning of mean-variance frontier.

Solnik (1978): Figure 1, 2

4.2 Tracking Error Optimization

Benchmark-relative investment management: Tracking error (TE) is a measure of relative investment risk.

Given our general framework of background risks, we focus now on the special case where that risk is given by a portfolio of risky assets (benchmark)

$$r_b = \mathbf{w}_b' \mathbf{r}.$$

In addition, we will not use the safe asset in our portfolio.

- Notation:

Let us define active returns as portfolio returns minus benchmark returns,

$$a = r_p - r_b = (\mathbf{w}_p - \mathbf{w}_b)' \mathbf{r} = \mathbf{w}_a' \mathbf{r}.$$

We define TE as the volatility of active returns:
►The ex-ante TE, an unknown parameter, is

$$\sigma_a = \sqrt{\mathbf{w}_a' \mathbf{\Sigma} \mathbf{w}_a}.$$

►The ex-post TE, our estimator, is

$$\hat{\sigma}_a = S_a = \sqrt{\frac{1}{T-1} \sum_{t=1}^{T} (a_t - \bar{a})^2},$$

where $\bar{a}$ is the sample average of active returns.

Time aggregation of tracking error: If we construct monthly portfolios and we assume for simplicity that annual returns are additive, the square of annual TE and the square of annualized TE are

$$Var\left(\sum_{i=1}^{12} a_i\right) \qquad vs. \qquad 12\sigma_a^2,$$

and they are not equal if there is autocorrelation in active returns. This also applies to annualized Sharpe ratio.

- Tracking error optimization:

We will use the framework in Roll (1992) from now on. This problem can be solved within the traditional framework, imposing a self-financing constraint

$$\min_{\mathbf{w}_a} Var\left(r_a\right) \qquad s.t. \qquad E\left(r_a\right)=\nu_a, \qquad \mathbf{1}'\mathbf{w}_a{=}0$$
$$\Rightarrow \mathbf{w}_a\left(\nu_a\right)=\nu_a\left(\frac{C}{D}\right)\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\nu}-\left(\frac{A}{C}\right)\boldsymbol{\Sigma}^{-1}\mathbf{1}\right).$$

Note that this is simply a traditional optimal portfolio (without safe asset) when we impose zero instead of unit cost

$$\mathbf{w}_a\left(\nu_a\right)=\mathbf{w}_{zc}\left(\nu_a\right).$$

In this context, the optimal active portfolio is independent of the particular benchmark.

Changes in ν_a only scale the optimal weights⇒The slope of TE-efficient frontier remains the same in the (ν_a, σ_a) space.

Scherer (2004): Figure 7.1

The total optimal portfolio $\mathbf{w}_p$ is therefore $(\nu_p = \nu_a + \nu_b)$

$$\mathbf{w}_{te}\left(\nu_p\right)=\mathbf{w}_{zc}\left(\nu_p\right)+\mathbf{w}_b,$$

that is, the traditional zero cost optimal portfolio plus a constant term, the benchmark.

In terms of rates of return,

$$r_{te}=r_{zc}+r_b.$$

4.3 Tracking Error vs. Mean-Variance Efficiency

- How do TE efficient portfolios compare with mean-variance efficient portfolios in mean-variance space?

TE optimization might be detrimental to overall portfolio efficiency.

Recall that traditional frontier in the $\left(\mu_p, \sigma_p^2\right)$ space is

$$\sigma^2\left(\nu_p\right) = \frac{1}{D}\left(C\nu_p^2 - 2A\nu_p + B\right).$$

Given the previous representation of optimal TE portfolios,

$$\begin{gathered} Var\left(r_{te}\right) = Var\left(r_{zc}\right) + Var\left(r_b\right) + 2Cov\left(r_{zc}, r_b\right) \\ = \frac{1}{D}\left(C\nu_p^2 - 2A\nu_p + B\right) + \left[\sigma_b^2 - \frac{1}{D}\left(C\nu_b^2 - 2A\nu_b + B\right)\right] \\ = \sigma^2\left(\nu_p\right) - \left[\sigma_b^2 - \sigma^2\left(\nu_b\right)\right]. \end{gathered}$$

▶ TE frontier represents a parallel parabola in the $\left(\mu_p, \sigma_p^2\right)$ space.
TE frontier passes through the benchmark and is dominated at all return levels by mean-variance frontier. The distance between both frontiers is constant.

▶ If r_b is efficient then both frontiers coincide.
TE efficient portfolios $\mathbf{w}_{te}$ are not mean-variance efficient $\mathbf{w}_{mv}$ unless the benchmark $\mathbf{w}_b$ is itself mean-variance efficient.
Importance of setting appropriate benchmark: No room to improve on an inefficient benchmark later.

Scherer (2004): Figure 7.2
Roll (1992): Figure 1

- There is a portfolio that, if added to a TE efficient portfolio, results in a mean-variance efficient portfolio:

Let us use the general results under background risks without safe asset. Now

$$\boldsymbol{\gamma} = \boldsymbol{\Sigma} \mathbf{w}_b,$$

and therefore

$$B_b = \nu_b, \qquad C_b = 1.$$

The optimal TE frontier is simply

$$\begin{gathered} \mathbf{w}_{br}\left(\nu_p\right) = \mathbf{w}_{mv}\left(\nu_p\right) + \mathbf{w}_h, \\ \mathbf{w}_h = \mathbf{w}_b - \mathbf{w}_{mv}\left(\nu_b\right). \end{gathered}$$

Note the link with previous TE results

$$\mathbf{w}_{te}\left(\nu_p\right) = \mathbf{w}_{zc}\left(\nu_p\right) + \mathbf{w}_b,$$

so that

$$\mathbf{w}_{zc}\left(\nu_p\right) = \mathbf{w}_{mv}\left(\nu_p\right) - \mathbf{w}_{mv}\left(\nu_b\right).$$

- Beta constraints (sensitivity with respect to benchmark) are sometimes included into TE optimization:

The optimal TE portfolios are such that

$$\beta_{te} = \frac{Cov\left(r_{te}, r_b\right)}{Var\left(r_b\right)} = 1 + \nu_a \left(\frac{C}{D}\right)\left(\nu_b - \frac{A}{C}\right).$$

Therefore, it will be frequent a situation like (A/C is mean of global minimum variance portfolio)

$$\nu_a > 0, \qquad \nu_b > \frac{A}{C} \Rightarrow \beta_{te} > 1$$

but portfolios that dominate r_b have $\beta_p < 1$.

Adding constraint on beta

$$\min_{\mathbf{w}_a} Var\left(r_a\right) \qquad s.t. \qquad E\left(r_a\right) = \nu_a, \qquad \mathbf{1}'\mathbf{w}_a{=}0, \qquad Cov\left(r_p, r_b\right) = \beta_p \sigma_b^2$$

$$\Rightarrow \mathbf{w}_a\left(\nu_a, \beta_p\right) = \lambda_1 \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}{+}\lambda_2 \boldsymbol{\Sigma}^{-1}\mathbf{1}{+}\lambda_3 \mathbf{w}_b.$$

Optimal $\mathbf{w}_a$ will depend on benchmark now.

Constraint on beta might give a better portfolio in the $\left(\mu_p, \sigma_p^2\right)$ space (not in the TE space).

Roll (1992): Figure 2-4

Portfolio Management: New Methods

Francisco Peñaranda (University of Alicante and FMG/LSE)
<penaranda@merlin.fae.ua.es>

5 Indexing and Statistical Arbitrage: New Methods

- Econometric Foundations: Cointegration

- Application to Pairs Trading

- Application to Index Tracking

References:

• Alexander, C. (2001): *"Market Models: A Guide to Financial Data Analysis"*, Wiley.

• Alexander, C., and A. Dimitriu (2005): Indexing and Statistical Arbitrage, *Journal of Portfolio Managament* 31, 50-63.

Gatev, E.G., W.N. Goetzmann, and K.G. Rowenhorst (1999): Pairs Trading: Performance of a Relative Value Arbitrage Rule, NBER WP 7032.

Vidyamurthy, G. (2004): *"Pairs Trading: Quantitative Methods and Analysis"*, Wiley.

5.1 Econometric Foundations: Cointegration

Nuevo contexto en la gestión de carteras:

- Hasta ahora, hemos supuesto que los activos sobre los que teníamos acceso estaban bien valorados ⇒Análisis estático de la relación riesgo vs. rentabilidad.
- Pasaremos a explotar situaciones en las que ciertos activos estén mal valorados ⇒Estrategias dinámicas que apuestan por la desaparición de esas situaciones (arbitraje estadístico).

Por ejemplo, analizaremos estrategias que apuestan por reversión a la media en el diferencial de precios entre dos activos (negociación de pares).

También veremos una aplicación de estas técnicas a la replicación de índices.

Novedades:

- Este análisis requiere un uso intensivo de econometría de series temporales.
- Pasamos de estudiar rentabilidades a estudiar precios⇒Correlación vs. cointegración.

El uso de modelos dinámicos de precios permite explotar relaciones de largo plazo entre activos que perdemos con el uso de rentabilidades, pero introduce complicaciones en términos de inferencia estadística.

Esta sección desarrolla la econometría de series temporales necesaria para dichas estrategias. Tratará dos temas:

▶ Procesos integrados:

Algunas series financieras, como los precios de activos, muestran tendencias y por lo tanto no son estacionarias. La relevancia de la estacionariedad radica en que su falta puede invalidar el uso de inferencia asintótica estándar.

Hay dos formas básicas de introducir tendencias en una serie temporal. El más obvio, y menos relevante para datos financieros, es una tendencia determinística. Nosotros nos centraremos en una segunda alternativa de tendencias estocásticas, basadas en procesos autoregresivos (AR).

▶ Cointegración:

Analizaremos el caso de un vector autoregresivo (VAR) con raíz unitaria. Por ejemplo, se da cuando el VAR se compone de precios de activos.

Dentro de esta clase de procesos VAR, nos interesan aquellos que cumplen una propiedad llamada cointegración. Se trata de una propiedad útil para desarrollar las estrategias de negociación comentadas arriba.

La notación que usaremos es independiente de los temas anteriores. Por ejemplo, e no denotará un exceso de rentabilidad.

- Integrated Processes I: Unit Roots and Random Walks

Un modelo AR de orden 1, AR(1), se construye a partir de ruido blanco u_t como

$$y_t = \phi_0 + \phi_1 y_{t-1} + u_t,$$

donde se impone

$$|\phi_1| < 1$$

para que el proceso pueda ser estacionario.

Vamos a eliminar dicho supuesto y estudiar el comportamiento bajo $\phi_1 = 1$. Por ello, estos procesos se denominan de *Raíz Unitaria*. También supondremos inicialmente $\phi_0 = 0$ para simplificar la exposición. El proceso es entonces

$$y_t = y_{t-1} + u_t,$$

lo que da lugar a un *Paseo Aleatorio*. Nótese que su varianza incondicional no está bien definida.

Podemos construir una serie estacionaria r_t con sólo tomar primeras diferencias

$$r_t = \Delta y_t = y_t - y_{t-1} = u_t,$$

y vemos que esta serie es simplemente ruido blanco en este contexto. En este caso, se dice que y_t es un *Proceso Integrado de Orden 1*

$$y_t \sim I(1)$$

y como r_t es estacionaria, se dice que es integrada de orden 0

$$r_t \sim I(0).$$

Este contexto es muy relevante para datos financieros. Se puede tomar y como el logaritmo del precio de un activo y r será entonces la rentabilidad en tasa continua (excluyendo dividendos).

Alexander (2001): Figura 11.1,3
Vidyamurthy (2004): Figura 2.6

Una extensión relevante de este modelo es el *Paseo Aleatorio con Deriva.* No imponemos $\phi_0 = 0$ y a ese parámetro se le denomina *Deriva.* El proceso es entonces

$$y_t = \phi_0 + y_{t-1} + u_t.$$

Nótese la similitud con el modelo log-normal de Black-Scholes si y_t es el log-precio de una acción.

La deriva introduce una tendencia temporal sobre la que se van acumulando las innovaciones. Por eso se denomina también a estos proceso de *Tendencia Estocástica*, como alternativa a una tendencia determinística.

Hamilton (1994): Figura 15.3
Tsay (2002): Figura 2.8

En el contraste que veremos más adelante, expresamos el proceso como

$$\Delta y_t = \phi_0 + \phi_1^* y_{t-1} + u_t, \qquad \phi_1^* = \phi_1 - 1,$$

y por lo tanto un proceso $I(1)$ es tal que $\phi_1^* = 0$.

Hasta ahora nos hemos centrado en el caso AR(1), pero podemos ser más generales. Podemos reexpresar un AR(p) como

$$\Delta y_t = \phi_0 + \phi_1^* y_{t-1} + \phi_2^* \Delta y_{t-1} + ... + \phi_p^* \Delta y_{t-(p-1)} + u_t,$$

para los correspondientes $\left(\phi_1^*, \phi_2^*, ..., \phi_p^*\right)$. En concreto,

$$\phi_1^* = \phi_1 + \phi_2 + ... + \phi_p - 1,$$

y de nuevo un proceso $I(1)$ equivale a $\phi_1^* = 0$.

- Integrated Processes II: Unit Root Tests

Es difícil diferenciar un proceso con raíz unitaria de un proceso estacionario cercano a raíz unitaria (por ejemplo, tipos de interés). Se necesitan muchas observaciones para ello.

Aunque existen otros contrastes, el contraste más conocido de raíz unitaria es el *Contraste de Dickey-Fuller* (DF). Se utiliza la parametrización vista con ϕ_1^* y las hipótesis a enfrentar son

$$H_0 : \phi_1^* = 0,$$
$$H_1 : \phi_1^* < 0,$$

mediante el estadístico t de MCO

$$DF = \frac{\hat{\phi}_1^*}{SE\left(\hat{\phi}_1^*\right)}.$$

El modelo anterior es estacionario bajo la alternativa. El modelo más general que se llega a estimar es

$$\Delta y_t = \phi_0 + \delta t + \phi_1^* y_{t-1} + u_t$$

para poder enfrentar una tendencia estocástica bajo la nula contra una tendencia determinística bajo la alternativa.

Veremos una tabla con los valores críticos del contraste, pero supone que no hay correlación en el residuo. Si la hay, no la podemos utilizar. En cambio, el que haya heteroscedasticidad no impide usar la tabla.
Hasta ahora nos hemos centrado en el caso AR(1), pero podemos ser más generales y usar un AR(p) en la forma expresada anteriormente. El contraste que nos interesa es la misma nula de antes sobre ϕ_1^* y se aplica la misma tabla, puesto que no depende de p. Cuando p>1, se dice que hacemos un *Contraste de Dickey-Fuller Aumentado* (ADF).

Tabla de valores críticos asintóticos para cada nivel de significatividad del ADF

Modelo estimado	Modelo verdadero	1%	5%	10%
$\phi_0 = 0, \delta = 0$	$\phi_0 = 0, \delta = 0$	-2.58	-1.95	-1.62
$\phi_0 \neq 0, \delta = 0$	$\phi_0 = 0, \delta = 0$	-3.43	-2.86	-2.57
$\phi_0 \neq 0, \delta = 0$	$\phi_0 \neq 0, \delta = 0$	$N(0,1)$		
$\phi_0 \neq 0, \delta \neq 0$	ϕ_0 gral., $\delta = 0$	-3.96	-3.41	-3.12

Fuente: Hamilton (1994) - Tablas 17.1 y B.6.

Usaremos los valores del caso 4 para una serie con tendencia clara (precios de acciones), y los del caso 2 en caso contrario (tipos de interés).

- Cointegration I: Definition and Properties

Ahora queremos estudiar conjuntamente más de una serie temporal. Para simplificar la exposición, supondremos que nos interesan 2 series temporales que organizaremos en un vector

$$\mathbf{y}_t' = (y_{1t}, y_{2t}).$$

Suponemos que

$$y_{1t} \sim I(1), \qquad y_{2t} \sim I(1)$$

es decir, las dos series son integradas del mismo orden y ese orden es 1.

Decimos que (y_{1t}, y_{2t}) están *Cointegradas* si además encontramos una combinación lineal estacionaria de ambas series (distinta de la trivial $\mathbf{0}$)

$$e_t = \mathbf{b}'\mathbf{y}_t = y_{1t} - by_{2t} \sim I(0),$$

donde hemos normalizado los pesos a $\mathbf{b} = (1, -b)'$ para que la combinación sea única. Al vector $\mathbf{b}$ se le denomina *Vector de Cointegración.*

Nótese que sólo puede haber una o ninguna relación de cointegración. Si hubiera dos relaciones estacionarias, las dos variables no podrían ser $I(1)$.

Cointegración implica que (y_{1t}, y_{2t}) no presentan reversión a la media, pero y_{1t} no puede alejarse mucho de forma permanente de by_{2t}. Un ejemplo clásico en finanzas son los precios a contado y futuro de un activo (otro, en economía, es el consumo y la renta). En las aplicaciones que veremos, (y_{1t}, y_{2t}) serán (log) precios e interpretaremos e_t como un diferencial de precios.

Hamilton (1994): Figura 19.1
Vidyamurthy (2004): Figura 5.1-2

Centraremos nuestra exposición en un modelo VAR de orden 1, VAR(1), de dimensión dos. Se construye a partir de un vector de ruido blanco como

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \phi_{01} \\ \phi_{02} \end{pmatrix} + \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{pmatrix} \begin{pmatrix} y_{1(t-1)} \\ y_{2(t-1)} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix},$$

$$\mathbf{y}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1 \mathbf{y}_{t-1} + \mathbf{u}_t.$$

Para que el proceso pueda ser estacionario, se supone que los autovalores de $\mathbf{\Phi}_1$ son menores que 1 en valor absoluto. Es decir, los valores λ tales que

$$|\mathbf{\Phi}_1 - \lambda \mathbf{I}_2| = 0$$

cumplen la restricción

$$|\lambda| < 1.$$

Vamos a analizar el caso de que no se cumpla esta restricción. Suponemos $\boldsymbol{\phi}_0 = \mathbf{0}$ para simplificar y tomamos primeras diferencias (como en las expresiones usadas para los contrastes de raíz unitaria)

$$\begin{pmatrix} \Delta y_{1t} \\ \Delta y_{2t} \end{pmatrix} = \begin{pmatrix} \phi_{11} - 1 & \phi_{12} \\ \phi_{21} & \phi_{22} - 1 \end{pmatrix} \begin{pmatrix} y_{1(t-1)} \\ y_{2(t-1)} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix},$$

$$\Delta \mathbf{y}_t = \mathbf{\Phi}_1^* \mathbf{y}_{t-1} + \mathbf{u}_t, \qquad \mathbf{\Phi}_1^* = \mathbf{\Phi}_1 - \mathbf{I}_2.$$

El que las variables estén cointegradas impone una restricción en el VAR. El rango de $\mathbf{\Phi}_1^*$ debe ser 1.

En general, la matriz $\mathbf{\Phi}_1^*$ puede tener rangos 0, 1, y 2. Nuestra hipótesis mantenida es que las variables son $I\,(1)$. Dado que $\Delta \mathbf{y}_t$ y $\mathbf{u}_t$ son estacionarios , pero $\mathbf{y}_{t-1}$ no lo es, el caso de rango 2 es imposible en este contexto. Luego sólo podemos encontrar rango de $\mathbf{\Phi}_1^*$ 0 o 1.

Si es 0 entonces no hay relación de cointegración, $\mathbf{\Phi}_1^* = \mathbf{0}$. Si el rango es 1, la matriz $\mathbf{\Phi}_1^*$ debe tener una estructura sólo degenerada en parte como

$$\mathbf{\Phi}_1^* = \mathbf{c}\mathbf{b}',$$

para ciertos $\mathbf{c}$ y $\mathbf{b}$, donde $\mathbf{b}$ es el vector de cointegración, y por lo tanto

$$\Delta \mathbf{y}_t = \mathbf{c} e_{t-1} + \mathbf{u}_t.$$

A esta forma de expresar el proceso de $\mathbf{y}_t$ se le denomina *Vector de Corrección del Error* (VEC). Un VEC es un VAR donde se ha impuesto la restricción de cointegración.

Se interpreta $e_{t-1} = 0$ como una relación de largo plazo o de equilibrio a la que el sistema tiende. Si $e_{t-1} \neq 0$ se produce un efecto sobre $\Delta \mathbf{y}_t$ que corrige esa desviación. El vector $\mathbf{c}$ mide la velocidad de esa convergencia.

Se pueden extender los conceptos anteriores a un contexto más general. Hay cointegración si un vector de dimensión N es tal que todas sus componentes son $I(1)$ y podemos encontrar una o varias combinaciones lineales de dicho vector (diferentes a la nula) que son $I(0)$.

Partimos de un VAR(p) con constante

$$\mathbf{y}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1 \mathbf{y}_{t-1} + ... + \boldsymbol{\Phi}_p \mathbf{y}_{t-p} + \mathbf{u}_t,$$

donde las componentes de $\mathbf{y}_t$ son todas $I(1)$. Para los correspondientes $\left(\boldsymbol{\Phi}_1^*, \boldsymbol{\Phi}_2^*, ..., \boldsymbol{\Phi}_p^*\right)$, podemos expresarlo como

$$\Delta\mathbf{y}_t = \boldsymbol{\phi}_0 + \boldsymbol{\Phi}_1^* \mathbf{y}_{t-1} + \boldsymbol{\Phi}_2^* \Delta\mathbf{y}_{t-1} + ... + \boldsymbol{\Phi}_p^* \Delta\mathbf{y}_{t-(p-1)} + \mathbf{u}_t.$$

La restricción derivada de cointegración ($\boldsymbol{\Phi}_1^*$ de rango degenerado, pero no igual a $\mathbf{0}$) es

$$\boldsymbol{\Phi}_1^* = \boldsymbol{\Phi}_1 + ... + \boldsymbol{\Phi}_p - \mathbf{I}_2 = \mathbf{C}\mathbf{B}',$$

y si la imponemos en el VAR tenemos el VEC

$$\Delta\mathbf{y}_t = \boldsymbol{\phi}_0 + \mathbf{C}\mathbf{e}_{t-1} + \boldsymbol{\Phi}_2^* \Delta\mathbf{y}_{t-1} + ... + \boldsymbol{\Phi}_p^* \Delta\mathbf{y}_{t-(p-1)} + \mathbf{u}_t.$$

En un contexto general, $\mathbf{B}$ es una matriz de vectores de cointegración, y $\mathbf{e}_{t-1}$ es un vector de relaciones de equilibrio.

Cuando se realiza inferencia con variables integradas de orden 1 $\mathbf{y}_t$ (log-precios), se tiende a trabajar con series en diferencias $\Delta\mathbf{y}$ (rentabilidades). Nótese que un VAR aplicado exclusivamente a las diferencias $\Delta\mathbf{y}_t$ es lo correcto cuando las variables no están cointegradas, pero estaría mal especificado cuando sí lo están puesto que olvidamos $\mathbf{e}_{t-1}$.

Las relaciones a largo plazo o de equilibrio entre las variables se pierden al tomar diferencias. Nótese la diferencia entre estudiar correlación (rentabilidades, corto plazo) y cointegración (precios, largo plazo).

Alexander (2001): Figura 12.1

- Cointegration II: Estimation and Tests

Pasamos a estudiar dos cuestiones claves que van de la mano: Estimar el vector de cointegración y contrastar cointegración. Como ocurre con los contrastes de raíz unitaria, se necesitan series temporales largas para responder a estas cuestiones.

Antes de aplicar los métodos que veremos, debemos haber aplicado contrastes ADF a las series (y_{1t}, y_{2t}) y no haber rechazado que sean $I(1)$.

Hay dos enfoques para estudiar cointegración y trabajar con un VEC:

- Enfoque uniecuacional o de Engle-Granger (EG): Será la metodología que usemos puesto que es sencilla y se basa en lo ya estudiado para contrastar raíces unitarias.
- Enfoque multiecuacional o de Johansen: Es una metodología más completa y compleja que se puede consultar en las referencias.

Pasamos a estudiar el enfoque de Engle-Granger, que se divide en dos etapas.

▶ Primera etapa de EG:

Se estima la relación de cointegración de y_{1t} sobre y_{2t} mediante MCO

$$y_{1t} = a + by_{2t} + v_t,$$

donde se suele introducir a (si es conveniente una constante en el VAR).

Se debe comprobar que el R^2 sea cercano a 1 con muestras grandes. Sabemos que debe converger a 1.

Esta regresión entre variables $I(1)$ sólo se justifica en un contexto de cointegración (regresiones espurias en otro caso). Es decir, cuando $v_t \sim I(0)$ y por tanto b define el vector de cointegración. Pero incluso en este caso las propiedades de $\hat{b}$ no son las usuales.

No se puede aplicar la inferencia asintótica estándar:

- Por un lado, $\hat{b}$ es superconsistente y el R^2 tiende a 1 conforme aumenta el tamaño muestral.
- Por otro lado, la distribución asintótica de $\hat{b}$ no es normal en general, con lo que es difícil contrastar significatividad.
- Tampoco se aplica en general la distribución usual del contraste t (ni la del contraste F cuando tenemos más de dos series)
- Además, las propiedades en muestras pequeñas pueden ser malas. Por ejemplo, el sesgo puede ser alto.

Posteriormente, usamos el estimador de MCO para estimar v_t

$$\hat{v}_t = y_{1t} - \hat{a} - \hat{b}y_{2t}$$

y analizamos si el residuo es estacionario vía un contraste tipo ADF. En concreto, contrastamos la nula de no cointegración mediante

$$H_0 : \varphi_1 = 0,$$
$$H_1 : \varphi_1 < 0,$$

en la ecuación

$$\Delta\hat{v}_t = \varphi_1\hat{v}_{t-1} + \varphi_2\Delta\hat{v}_{t-1} + ... + \varphi_p\Delta\hat{v}_{t-(p-1)} + \xi_t.$$

No es necesario introducir constante aquí ya que se incluyó en la regresión de cointegración. Lo llamaremos contraste AEG.

Sólo podemos usar los valores críticos vistos para ADF si observamos el verdadero v_t (si sabemos b), pero no si sólo conocemos $\hat{v}_t$. Además, esos valores dependen del número de variables en la ecuación de cointegración y otros parámetros.

Tabla: Valores críticos ($T = 500$) para cada nivel de significatividad de AEG

1) 2 variables

EC estimada	VAR verdadero	1%	5%	10%
$a = 0$	$\boldsymbol{\phi}_0 = \mathbf{0}$	-3.39	-2.76	-2.45
$a \neq 0$	$\boldsymbol{\phi}_0 = \mathbf{0}$	-3.96	-3.37	-3.07
$a \neq 0$	$\boldsymbol{\phi}_0 \neq \mathbf{0}$	-3.98	-3.42	-3.13

2) 6 variables

EC estimada	VAR verdadero	1%	5%	10%
$a = 0$	$\boldsymbol{\phi}_0 = \mathbf{0}$	-4.99	-4.40	-4.14
$a \neq 0$	$\boldsymbol{\phi}_0 = \mathbf{0}$	-5.28	-4.71	-4.43
$a \neq 0$	$\boldsymbol{\phi}_0 \neq \mathbf{0}$	-5.36	-4.74	-4.46

Fuente: Hamilton (1994) - Tabla B.9.

Usaremos los valores del caso 3 si observamos una tendencia clara y los del caso 2 en caso contrario.

Alexander (2001): Figura 12.2-3

▶ Segunda etapa de EG:
Se estima por MCO el VEC

$$\Delta \mathbf{y}_t = \boldsymbol{\phi}_0 + \mathbf{c}\hat{e}_{t-1} + \mathbf{\Phi}_2^* \Delta \mathbf{y}_{t-1} + ... + \mathbf{\Phi}_p^* \Delta \mathbf{y}_{t-(p-1)} + \mathbf{u}_t,$$

usando el vector de cointegración estimado

$$\hat{e}_t = y_{1t} - \hat{b} y_{2t}.$$

Hay un interés especial en la estimación de $\mathbf{c}$, pues mide la velocidad de ajuste a equilibrio.

Todas las variables involucradas son estacionarias y por tanto podemos aplicar inferencia asintótica estándar. Además, la distribución asintótica de los estimadores no depende del uso de $\hat{e}_{t-1}$ en vez de e_{t-1} por la superconsistencia de $\hat{b}$.

Se deben tener en cuenta las limitaciones del enfoque EG:

• Si hay más de dos variables en el análisis entonces puede haber más de un vector de cointegración. Éste es el problema más grave de EG. Pero las aplicaciones que veremos más adelante no sufren este problema.

• Cuando sólo analizamos dos variables, hay una pérdida de eficiencia en la estimación de b respecto al enfoque de Johansen en general.

Además, hay dos normalizaciones a estudiar, aplicando MCO a y_{1t} sobre y_{2t} y a y_{2t} sobre y_{1t}. Las dos estimaciones pueden dar resultados diferentes al contrastar cointegración, pero no será el caso con muestras grandes.

5.2 Application to Pairs Trading

Es la estrategia neutral al mercado más sencilla que podemos estudiar, puesto que se basa en negociar 2 activos. Se analizan los precios relativos de ambos activos para ver si cierto diferencial es estable en el tiempo. Es decir, si muestra reversión a su media.

Si se da esta propiedad, cuando observemos que el diferencial se aleja de su media abriremos una posición corta en el activo sobrevalorado y una larga en el activo infravalorado. La cerraremos cuando revierta a su media, que es lo esperado dado lo ocurrido en el pasado, y así ganaremos la variación en el diferencial.

Se trata de una apuesta sobre la reversión a la media del diferencial. Obviamente, corremos el riesgo de que esto tarde mucho en ocurrir o que ni siquiera ocurra.

Véase Gatev et al. (1999) para una evaluación de estas estrategias.

Podemos traducir de forma sencilla esta situación al contexto de cointegración estudiado:

- Los precios de los activos (o índices que incluyen la reinversión de dividendos) son (y_{1t}, y_{2t}) y sabemos que son $I\left(1\right)$.

- La posición larga-corta a tomar viene dada por la relación de cointegración $e_t = y_{1t} - by_{2t}$, llamada diferencial en este contexto.

- Strategy design: 3 stages

▶ Primera etapa: Identificación de posibles pares

Podemos estudiar todos los pares posibles dentro de cierto conjunto de activos (hacer todos los contrastes posibles de cointegración), pero sería demasiado costoso.

Es recomendable hacer una selección de posibles pares a través de un análisis más fundamental como por ejemplo vía modelos factoriales. Podemos centrarnos en pares de activos con betas similares.

▶ Segunda etapa: Análisis de cointegración

Aplicaremos el enfoque EG visto.

Nótese que podemos imponer directamente cierto valor de b (por ejemplo 1) que nos parezca razonable y estudiar la estacionariedad del diferencial correspondiente.

▶ Tercera etapa: Regla de negociación

Aquí es clave la dinámica del diferencial (dada por el VEC) y los costes de transacción. Podemos simular distintas estrategias, comparar su comportamiento y seleccionar la óptima.

- Se debe determinar el nivel del diferencial para el que abriremos la posición. Por ejemplo, podríamos abrir la posición cuando el valor absoluto de la desviación del diferencial respecto a su media es mayor que 2 veces su desviación típica. Se debe compensar el beneficio por operación con el número de operaciones. También es relevante la horquilla de precios.
- Otra decisión es cuándo cerrar la posición. Se puede cerrar cuando el diferencial revierta a la media o cuando tome el valor contrario al de apertura de la posición. La segunda opción disminuye los costes de transacción (menos operaciones), pero también mantiene la posición abierta durante más tiempo.
- También se debe estudiar el tamaño de la posición.

5.3 Application to Index Tracking

Pasamos a ver la replicación de índices mediante cointegración según Alexander and Dimitriu (2005) en vez de la práctica habitual de minimizar tracking error. La primera alternativa explota la información de largo plazo en los precios, mientras que la segunda se basa en la información de corto plazo de las rentabilidades (correlación).

Un índice bursátil es una cartera de acciones, por lo que una cartera suficientemente grande debería estar cointegrada con el índice (suponiendo que los pesos no cambian demasiado con el tiempo).

Alexander and Dimitriu (2005): Exhibit 1

- Strategy design: 3 stages

▶ Primera etapa: Selección de los activos

Se debe seleccionar un número reducido de activos que sean representativos de la evolución del índice, puesto que buscamos una cartera cointegrada con él.

▶ Segunda etapa: Análisis de cointegración

Aplicaremos el enfoque EG visto.

Usamos MCO sobre log-precios,

$$y_t = a + b_1 x_{1t} + ... + b_K x_{Kt} + e_t,$$

donde y_t es el log-índice y $(x_{1t}, ..., x_{Kt})$ son K log-precios de acciones. Los coeficientes $(b_1, ..., b_K)$ se normalizan para que sumen a uno y por tanto sean interpretables como pesos de cartera.

▶ Tercera etapa: Regla de negociación

Hay que calibrar dos parámetros adicionales: El periodo de estimación (se requieren series largas, varios años de datos diarios) y el de reajuste de la cartera.

La calibración se debe realizar vía evaluación dentro y fuera de la muestra usada para la estimación:

- Podemos estudiar medidas estadísticas como el contraste AEG y el error estándar de la regresión dentro de la muestra.
- Podemos estudiar medidas económicas como el tracking error y el ratio de información fuera de la muestra.

Posteriormente, habría que hacer backtesting para comprobar la robustez de la estrategia a lo largo del tiempo.

Alexander and Dimitriu (2005): Exhibit 2-4

Additional References

Ang, A., and G. Bekaert (2002): International Asset Allocation with Regime Shifts, *Review of Financial Studies* 15, 1137-1187.

Avramov, D. (2002): Stock Return Predictability and Model Uncertainty, *Journal of Financial Economics* 64, 423-458.

Bai, J., and S. Ng (2001): Test for Skewness, Kurtosis, and Normality for Time Series Data, mimeo.

Barberis, N. (2000): Investing in the Long Run when Returns are Predictable, *Journal of Finance* 55, 225-264.

Bawa, V., and E. Lindenberg (1977): Capital Market Equilibrium in a Mean-Lower Partial Moment Framework, *Journal of Financial Economics* 5, 189-200.

Britten-Jones, M. (1999): The Sampling Error in Estimates of Mean-Variance Efficient Portfolio Weights, *Journal of Finance* 54, 655-671.

Chamberlain, G. (1983): A Characterization of the Distributions that Imply Mean-Variance Utility Functions, *Journal of Economic Theory* 29, 185-201.

Hamilton, J.D. (1994): *"Time Series Analysis"*, Princeton.

Harlow, W., and R. Rao (1989): Asset Pricing in a Generalised Mean-Lower Partial Moment Framework: Theory and Evidence, *Journal of Financial and Quantitative Analysis* 24, 285-311.

Ingersoll, J.E. (1987): *"Theory of Financial Decision Making"*, Rowman & Littlefield.

Lee, W. (2001): *"Theory and Methodology of Tactical Asset Allocation"*, John Wiley & Sons Inc.

Merton, R.C. (1980): On Estimating the Expected Return on the Market: An Exploratory Investigation, *Journal of Financial Economics* 8, 323-361.

Pástor, L. (2000): Portfolio Selection and Asset Pricing Models, *Journal of Finance* 55, 179-223.

Pedersen, C.S., and S.E. Satchell (2002): On the Foundation of Performance Measure under Asymmetric Returns, *Quantitative Finance* 2, 217-223.

Sharpe, W.F., and L.G. Tint (1990): Liabilities - A New Approach, *Journal of Portfolio Management*, Winter 1990, 5-10.

Stutzer, M. (2003): Portfolio Choice with Endogenous Utility: A Large Deviations Approach, *Journal of Econometrics* 116, 365-386.

Tsay, R.S. (2002): *"Analysis of Financial Time Series"*, Wiley.

Tu, J., and G. Zhou (2004): Data-Generating Process Uncertainty: What Difference Does it Make in Portfolio Decisions?, *Journal of Financial Economics* 72, 385-421.