

The aggregate “portfolio”: Econometrics of economic rates of return with a Portuguese illustration

By Ana Paula MARTINS †

Abstract. This research forwards estimation procedures – applications of weighted and generalized least squares techniques - designed to infer expected values, variances and covariances of rates of return in the presence of variously correlated sample observations but uncorrelated “sample waves” (strata of) and availability of already aggregate data – under which inference must rely on averages (means) of averages of averages... The same principles are extended to the method of order statistics, appropriate for univariate inference of a truncated distribution parameters. Simple tests of portfolio – market - efficiency based on correlations (or special rank correlations) between actual and estimated optimal shares are also proposed. Illustrative estimates for Portuguese economic sectors are provided – relying on yearly, semi-aggregate information for firms with 20 or more employees, covering the period 1996-2002: on the one hand, sector means, variances and covariances of economic returns to unitary (tangible and intangible asset) applications are presented and reduced by principal components. On the other, optimal (“unrestricted”) portfolios for nested subsets are reported, having been generated by a stepwise elimination procedure. Industries’ betas are approximated and market efficiency tested. Finally, parameter MOS estimates under (univariate) truncated normal assumptions are obtained.

Keywords. Industry economic rates of return; Firm size; Optimal portfolio; Mean – Variance; CAPM; Market efficiency; Weighted least squares; “Weighted” SUR; Weighted method of order statistics; Weighted principal components; Dummy variables. Index numbers; Aggregation.

JEL. G11; G12; G30. C39; C43; C51; C61. C24. L16; L25..

1. Introduction

Finance is full of examples where weighting by some relevant budget share is required to generate appropriate test statistics. With respect to means, as is well documented in statistical textbooks, weighting unitary dimensions – prices, rates of return - is of common usage, usually generating aggregate indices. Less explored is the appropriate refinement to weight and extract variances and co-variances – in particular, from published series on aggregate or already averaged information -, either to be assessed *per se* or used as input in other empirical research. It is the purpose of this note to digress over the subject, illustrating with some basic financial econometric applications.

† Universidade Católica Portuguesa, Department of Economics. This research started while the author was Invited Professor at Faculdade de Economia da Universidade Nova de Lisboa, Portugal.

☎. (351) 217 21 42 48 ✉. apm@ucp.pt

Journal of Economics Bibliography

Firstly, we note that appropriately weighted least squares of a regression of profits on the capital stock can produce an adequate estimate of the average rate of return. Then, for a multiple and interconnected assets pooled sample, the use of an equation system of such single regressions provides a convenient device to proceed to multivariate (assets) inference. Such weighting is also valid under univariate treatment: if a particular distribution function is assumed for the unit rates of return, one can use the same principles to estimate, using the method of order statistics (Martins, 2005b), the parameters of a potentially truncated distribution, with a lower truncation point having an obvious meaning in finance applications.

Inquiry of factors composing observed returns can also be subject to weighting. Principal component techniques are also modified in order to account for different representativeness of each variable's observations – the method involves intermediate regressions and decomposes an uncentered cross-moment matrix.

Finally, inference of assets' betas – measuring each asset's contribution to the overall variance of a portfolio - obviously face the same concern, as well as market efficiency (see [Sharpe, 1964](#); [Lintner, 1965](#); [Black, 1972](#)) testing.

The several issues are exemplified with Portuguese data on trading profits and tangible plus intangible assets for the period 1996 to 2002, and interaction with average firm size and aggregate importance of the industry briefly inspected. With an overall average economic return of around 11%, empirical traits of the pursued inquiries apparently generated different clustering – or ranking - of the sectors – and different from that implied by the industries' relative importance in the total existing stock. Simulated optimal portfolios do not seem to generate the later; yet, the actual "market" efficiency does not seem to be overall rejected.

In the empirical applications, the weighting methods are used by force of available data, based on aggregate yearly information by industry; yet, they have general applicability whenever observations pertain to elements of different sizes themselves. Naturally, international comparisons, as well as with micro data estimates – where the time length of the investments also arises as another weighting dimension -, suggest themselves for further developments.

The analysis proceeds as follows: Section 2, justifies econometric estimation procedures based on regressions without constant term and with multiplicative dummy variables on a theoretical level – special cases of SUR (and GLS) estimation; conformable estimates are forwarded in section 3, where data is introduced and statistically synthesized. Section 4 presents the implied efficient portfolios, contrasting the sectors' optimal with the actual asset shares in the economy. Aggregate betas towards the effective portfolio are provided along with some tests in section 5. Section 6 forwards parameter estimates for the unit rates of return distribution function under a truncated normal assumption. Concluding appraisals are summarized in this introduction.

2. (Weighted) Least squares rate of return estimates

The appropriate concept of the average return in an industry depends on its purposes. As an overall – descriptive - performance indicator, it should take the form of an aggregate index. As is well-known, that involves a weighted average by the amounts invested. Denote by R_{it} aggregate profits generated in year t by sector i and K_{it} the value of the assets allocated to that sector in the same year; the average return on the monetary unit invested in the sector in year t is $r_{it} = \frac{R_{it}}{K_{it}}$. Assume that the annual return of a “monetary” unit investment in sector i is a stable aggregate over the years, a random variable with fixed expected value, r_i , and variance, $\sigma_i^2 (= \sigma_{ii})$; the contemporaneous covariance structure (σ_{ij} 's) between – the joint pdf of - the unit returns in any two (and of all...) sectors is also stable. Across time, all such returns are statistically independent². Being our objective to infer the first and second centered moments (and cross-moments) of the relevant distributions, the appropriate yearly estimator may differ from the previous aggregate ratio; an estimator derived from multiple-year samples may not have a direct correspondence with it. On the other hand, available information may condition our choice.

We may take different views of the statistical properties of the K_{it} observations sampled in year $t - t = 1, 2, \dots, T$ - for sector $i - i = 1, 2, \dots, n$ -, each of which indeed yields a unit return $r_{it}, 1=1,2,\dots, K_{it}$ unobserved by the

researcher that only has access to $R_{it} = \sum_{t=1}^{K_{it}} r_{it}$ - and thus, for each year and

sector, to the yearly mean $r_{it} = \frac{R_{it}}{K_{it}}$ over the K_{it} units.

i) Let the observations be independent within each sector i .

Then, a natural estimator of the expected value r_i is the weighted (by $\frac{K_{it}}{\sum_{t=1}^T K_{it}}$)

average of the T yearly means, the r_{it} 's, i.e.:

$$\bar{r}_i = \sum_{t=1}^T \frac{R_{it}}{K_{it}} \frac{K_{it}}{\sum_{t=1}^T K_{it}} = \frac{\sum_{t=1}^T R_{it}}{\sum_{t=1}^T K_{it}}, i=1, 2, \dots, n \quad (1)$$

² This may, of course, be an unlikely assumption, even if we are not in the presence of a panel. Nevertheless, given the small number of time observations in the sample, it will be maintained throughout this research.

Journal of Economics Bibliography

As \bar{r}_i is identical to the standard mean over the $\sum_{t=1}^T K_{it}$ observations,

$$Var(\bar{r}_i) = \frac{\sigma_{ii}}{\sum_{t=1}^T K_{it}}. \text{ Theoretically, } Var(r_{it}) = \frac{\sigma_{ii}}{K_{it}} \text{ - which is not constant over}$$

t – and therefore $Var(\bar{r}_i) = \sum_{t=1}^T \frac{K_{it}}{\sum_{t=1}^T K_{it}} \frac{Var(r_{it})}{T}$: being \bar{r}_i the average over T of

the r_{it} 's, an estimator for $Var(\bar{r}_i)$, $Var(\bar{r}_i)$, would relate to the estimators of $Var(r_{it})$ according to:

$$Var(\bar{r}_i) = \frac{Var(r_{it})}{T} \quad (2)$$

Of course, an estimator of $Var(r_{it})$ is $(\frac{R_{it}}{K_{it}} - \bar{r}_i)^2$:

$$\overline{Var(r_{it})} = \sum_{t=1}^T \frac{(\frac{R_{it}}{K_{it}} - \bar{r}_i)^2}{\sum_{t=1}^T K_{it}} = \sum_{t=1}^T \frac{(\frac{R_{it}}{K_{it}})^2}{\sum_{t=1}^T K_{it}} - \bar{r}_i^2, \quad i = 1, 2, \dots, n \quad (3)$$

In \bar{r}_i , all observations of sector i share the same status, no matter from which year they come from. Then, the variance of the monetary unit return in sector i, if we assumed each to be an independent draw, would be approximated by:

$$Var(r_{it}) = \sum_{t=1}^T K_{it} Var(\bar{r}_i) = (\sum_{t=1}^T \frac{R_{it}^2}{K_{it}} - \sum_{t=1}^T K_{it} \bar{r}_i^2) / T, \quad i = 1, 2, \dots, n \quad (4)$$

An alternative estimation approach – and another indirect justification of the proposed estimators - would rely on the estimator of the parameter of the regression on the sum of the sector's yearly returns:

$$R_{it} = r_i K_{it} + \otimes_{it}, \quad t = 1, 2, \dots, T; \quad i = 1, 2, \dots, n \quad (5)$$

\otimes_{it} is a null mean error term. It is reasonable to assume it heteroscedastic, with variance proportional to the number of invested units – once they are assumed independent and summed to form each equation i,t, i.e.:

$$Var(\otimes_{it}) = \otimes_i^2 K_{it}, \quad t = 1, 2, \dots, T; \quad i = 1, 2, \dots, n \quad (6)$$

Journal of Economics Bibliography

If the error terms are uncorrelated, in time as across sectors, single equation estimates by weighted least squares, using $\frac{1}{K_{it}}$ as weights³, or applying ordinary least squares to the transformed system:

$$\frac{R_{it}}{\sqrt{K_{it}}} = r_i \sqrt{K_{it}} + \otimes_{it}^* \quad , \quad t = 1, 2, \dots, T; \quad i = 1, 2, \dots, n \quad (7)$$

will provide as estimate of r_i :

$$\hat{r}_i = \frac{\sum_{t=1}^T R_{it}}{\sum_{t=1}^T K_{it}} = \bar{r}_i \quad , \quad i = 1, 2, \dots, n \quad (8)$$

The estimated variance will be:

$$Var(\hat{r}_i) = \sigma_i^2 \frac{1}{\sum_{t=1}^T K_{it}} \quad , \quad i = 1, 2, \dots, n \quad (9)$$

where:

$$\sigma_i^2 = \frac{\sum_{t=1}^T e_{it}^{*2}}{T-1} = \frac{1}{T-1} \left(\sum_{t=1}^T \frac{R_{it}^2}{K_{it}} - \sum_{t=1}^T K_{it} \bar{r}_i^2 \right) = \frac{T}{T-1} Var(r_{it}) \quad , \quad i = 1, 2, \dots, n \quad (10)$$

with e_{it}^* denoting the estimated residuals from (7).

Hence:

$$Var(\hat{r}_i) = \frac{T}{T-1} Var(\bar{r}_i) \quad , \quad i = 1, 2, \dots, n \quad (11)$$

Admit we also want to estimate the covariances of unit returns across sectors. Under independence, $Cov(r_{it}, r_{jt}) = Cov(r_{itl}, r_{jtl}) = \otimes_{ij}$; estimators of both covariances should coincide.

If we admit independence of sector observations, $Cov\left(\frac{R_{it}}{K_{it}}, \frac{R_{jt}}{K_{jt}}\right) = \otimes_{ij}$:

$Cov(r_{it}, r_{jt})$ would reproduce $Cov(r_{itl}, r_{jtl})$ and we can propose a weighted (by $K_{it} K_{jt}$) estimator:

³ Standard weighted least squares as commonly presented in econometric textbooks– see Greene (2003), for example.

Journal of Economics Bibliography

$$Cov(r_{it}, r_{jt}) = \frac{\sum_{t=1}^T \frac{(R_{it} - \bar{r}_i)(R_{jt} - \bar{r}_j)K_{it}K_{jt}}{\sum_{t=1}^T K_{it}K_{jt}}}{\sum_{t=1}^T K_{it}K_{jt}} = \frac{\sum_{t=1}^T (R_{it} - \bar{r}_i K_{it})(R_{jt} - \bar{r}_j K_{jt})}{\sum_{t=1}^T K_{it}K_{jt}} \quad (12)$$

and – as we can show that $Cov(\bar{r}_i, \bar{r}_j) = \frac{\sum_{t=1}^T K_{it}K_{jt}}{\sum_{t=1}^T K_{it} \sum_{t=1}^T K_{jt}}$ \otimes_{ij} - it is reasonable

to infer:

$$Cov(\bar{r}_i, \bar{r}_j) = \frac{\sum_{t=1}^T K_{it}K_{jt}}{\sum_{t=1}^T K_{it} \sum_{t=1}^T K_{jt}} Cov(r_{it}, r_{jt}) = \frac{\sum_{t=1}^T \frac{(R_{it} - \bar{r}_i)(R_{jt} - \bar{r}_j)K_{it}K_{jt}}{\sum_{t=1}^T K_{it} \sum_{t=1}^T K_{jt}}}{\sum_{t=1}^T K_{it} \sum_{t=1}^T K_{jt}} = \frac{\sum_{t=1}^T (R_{it} - \bar{r}_i K_{it})(R_{jt} - \bar{r}_j K_{jt})}{\sum_{t=1}^T K_{it} \sum_{t=1}^T K_{jt}} \quad (13)$$

Consistently, under independence, the errors of the distribution of the yearly sum of returns of (5) would be correlated in such a way that:

$$Cov(\otimes_{it}, \otimes_{jt}) = \otimes_{ij} K_{it} K_{jt} \quad , \quad t = 1, 2, \dots, T; \quad i \otimes j, \quad i, j = 1, 2, \dots, n \quad (14)$$

Then, referring to (7), $Var(\otimes_{it}^*) = \otimes_{ii} = \otimes_i^2$. Yet:

$$Cov(\otimes_{it}^*, \otimes_{jt}^*) = \otimes_{ij} \sqrt{K_{it}K_{jt}} \quad , \quad t = 1, 2, \dots, T; \quad i \otimes j, \quad i, j = 1, 2, \dots, n \quad (15)$$

Let $\otimes^* = (\otimes_{1t}^*, \otimes_{2t}^*, \dots, \otimes_{Tt}^*)'$ represent the error term(s) of system (7) in vector form with \otimes_{it}^* denoting the error of the t-th period observations, i.e., $\otimes_{it}^* = (\otimes_{1t}^* \otimes_{2t}^* \dots, \otimes_{nt}^*)'$ - ordering observations by sector and forming yearly blocks. Then:

$$Cov(\otimes \otimes) = V = \quad (16)$$

Journal of Economics Bibliography

$$\begin{matrix}
 \sigma_{11} & \sigma_{12}\sqrt{K_{11}K_{21}} & \dots & \sigma_{1n}\sqrt{K_{11}K_{n1}} & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\
 \sigma_{21}\sqrt{K_{21}K_{11}} & \sigma_{22} & \dots & \sigma_{2n}\sqrt{K_{21}K_{n1}} & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\
 & & \dots & & & & & & & & & & \\
 \sigma_{n1}\sqrt{K_{n1}K_{11}} & \sigma_{n2}\sqrt{K_{n1}K_{21}} & \dots & \sigma_{nn} & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\
 0 & 0 & & 0 & \sigma_{11} & \sigma_{12}\sqrt{K_{12}K_{22}} & \dots & \sigma_{1n}\sqrt{K_{12}K_{n2}} & \dots & 0 & 0 & \dots & 0 \\
 0 & 0 & & 0 & \sigma_{21}\sqrt{K_{22}K_{12}} & \sigma_{22} & \dots & \sigma_{2n}\sqrt{K_{22}K_{n2}} & \dots & 0 & 0 & \dots & 0 \\
 & & & & \dots & & & & & & & & \\
 0 & 0 & & 0 & \sigma_{n1}\sqrt{K_{n2}K_{12}} & \sigma_{n2}\sqrt{K_{n2}K_{22}} & \dots & \sigma_{nn} & \dots & 0 & 0 & \dots & 0 \\
 & & & & \dots & & & & & & & & \\
 0 & 0 & & 0 & 0 & 0 & \dots & 0 & \dots & \sigma_{11} & \sigma_{12}\sqrt{K_{1T}K_{2T}} & \dots & \sigma_{1n}\sqrt{K_{1T}K_{nT}} \\
 0 & 0 & & 0 & 0 & 0 & \dots & 0 & \dots & \sigma_{21}\sqrt{K_{2T}K_{1T}} & \sigma_{22} & \dots & \sigma_{2n}\sqrt{K_{2T}K_{nT}} \\
 & & & & \dots & & & & & & & & \\
 0 & 0 & & 0 & 0 & 0 & \dots & 0 & \dots & \sigma_{n1}\sqrt{K_{nT}K_{1T}} & \sigma_{n2}\sqrt{K_{nT}K_{2T}} & \dots & \sigma_{nn}
 \end{matrix}$$

Take the single equation OLS estimates – by sector – and obtain:

$$\sigma_{ii} = \frac{\sum_{t=1}^T e_{it}^{*2}}{T-1}, \quad i = 1, 2, \dots, n \tag{17}$$

For σ_{ij} we could propose:

$$\sigma_{ij} = \frac{\sum_{t=1}^T \frac{e_{it}^* e_{jt}^*}{\sqrt{K_{it}K_{jt}}}}{T-1} = Cov(r_{it}, r_{jt}), \quad i \neq j, \quad i, j = 1, 2, \dots, n \tag{18}$$

Alternatively – consistently with (12) – we can infer it from the OLS estimator of σ_{ij} of the equation: $e_{it}^* e_{jt}^* = \sigma_{ij} \sqrt{K_{it}K_{jt}} + error_t$, $t = 1, 2, \dots, T$. Then we get:

$$\sigma_{ij} = \frac{\sum_{t=1}^T e_{it}^* e_{jt}^* \sqrt{K_{it}K_{jt}}}{\sum_{t=1}^T K_{it}K_{jt}} \tag{19}$$

On the one hand, the appropriate covariance matrix of the OLS estimator $\hat{r} = \bar{r} = (\bar{r}_1, \bar{r}_2, \dots, \bar{r}_n)'$ may be obtained from:

$$Cov(\bar{r}) = (X'X)^{-1} X' V X (X'X)^{-1} \tag{20}$$

where, in V , σ_{ij} is replaced by σ_{ij} . Then, one can show that for $i \neq j$, $i, j = 1, 2, \dots, n$:

$$Cov(\bar{r}_i, \bar{r}_j) = \frac{\sum_{t=1}^T K_{it} K_{jt}}{\sum_{t=1}^T K_{it} \sum_{t=1}^T K_{jt}} \frac{\sum_{t=1}^T (\frac{R_{it}}{K_{it}} - \bar{r}_i)(\frac{R_{jt}}{K_{jt}} - \bar{r}_j)}{T-1}$$

or

$$\frac{\sum_{t=1}^T (R_{it} - \bar{r}_i K_{it})(R_{jt} - \bar{r}_j K_{jt})}{\sum_{t=1}^T K_{it} \sum_{t=1}^T K_{jt}} \quad (21)$$

according to which estimation approach we followed and we expect to approximate $Cov(r_{it}, r_{jt})$ from:

$$Cov(r_{it}, r_{jt}) = \frac{\sum_{t=1}^T K_{it} \sum_{t=1}^T K_{jt}}{\sum_{t=1}^T K_{it} K_{jt}} Cov(\bar{r}_i, \bar{r}_j) \quad (22)$$

On the other, SUR estimation of equation system (7) is naturally suggested⁴: a second step would involve applying GLS to it with that same replacement. Denoting by Y the vector containing $\frac{R_{it}}{\sqrt{K_{it}}}$ stacked by period, and by X that of the corresponding $\sqrt{K_{it}}$, containing n multiplicative sector dummy variables: DI_{it} multiplied by $\sqrt{K_{it}}$, where $DI_{it} = 1$ for $i = I$, 0 for $i \neq I$, $I = 1, 2, \dots, n$ ⁵:

$$\hat{r}_{GLS} = (X' V^{-1} X)^{-1} X' V^{-1} Y \quad (23)$$

and

$$Cov(\hat{r}_{GLS}) = (X' V^{-1} X)^{-1} \quad (24)$$

Through Cholesky decomposition, one can obtain P such that $V = P P'$ and estimate by OLS, as is well-known, of the transformed model:

$$P^{-1} Y = P^{-1} X + v \quad (25)$$

⁴ For identification, we would need an adequately sized T – possibly not smaller than $n + 1$ (or $n + 2$ - see Gibbons, Ross & Shanken (1989), footnote 3 -, once we are estimating n parameters and $n(n+1)/2$ different elements of V with only nT observations. Given the special structure, a smaller sample could perform the task.

⁵ Standard generalized least squares – see Dhrymes (1978), for example.

Journal of Economics Bibliography

is legitimate, originating much the same results. In terms of relative order of magnitude, we would expect that matrix (24) would approximate:

1) in the diagonal,

$$\frac{T}{T-1} \text{Var}(\bar{r}_i),$$

and $\frac{T}{T-1} \frac{1}{\sum_{t=1}^T K_{it}} \text{Var}(r_{it})$ (26)

2) off the diagonal

$$\text{Cov}(\bar{r}_i, \bar{r}_j),$$

and $\frac{\sum_{t=1}^T K_{it} K_{jt}}{\sum_{t=1}^T K_{it} \sum_{t=1}^T K_{jt}} \text{Cov}(r_{it}, r_{jt})$ (27)

Of course, new – more efficient – estimates \hat{r}_i will also be produced – as well as σ_{ij} 's.

ii) Independence between observations of the same class or category is not the logic behind the portfolio variance estimation: rather, perfect (and positive) correlation between observations of the same type/asset is assumed – variances equating covariances for applications on the same title – rather than independence. That has important implications: consider a vector of observations X and define vector $L = (1, 1, \dots, 1)'$, so that $\bar{X} = L'X / n$. Then $\text{Var}(\bar{X}) = L' \text{Cov}(X) L / n^2$. If the observations in X are independent and share a common variance σ^2 , $\text{Cov}(X) = \sigma^2 I_n$, and, in fact, $\text{Var}(\bar{X}) = \sigma^2 / n$. But if they are perfectly (positively) correlated, $\text{Cov}(X) = \sigma^2 L L'$ and $\text{Var}(\bar{X}) = \sigma^2$. The same would not be true for weighted means.

. Admit that, yearly, we are in the presence of a “portfolio” and that within each sector and year the correlation is perfect between the unit application returns (which is hardly the case for aggregate data like ours). Then the σ_{it} of (5) will more adequately be such that:

$$\text{Var}(\sigma_{it}) = \sigma_i^2 K_{it}^2 = \sigma_{ii} K_{it}^2, \quad t = 1, 2, \dots, T; \quad i = 1, 2, \dots, n \quad (28)$$

Journal of Economics Bibliography

and (14) will still hold. If the error terms are uncorrelated, single equation estimates by weighted least squares, using $\frac{1}{K_{it}^2}$ as weights, or applying ordinary least squares to the transformed system:

$$\frac{R_{it}}{K_{it}} = r_i + \otimes_{it}^* \quad , \quad t = 1, 2, \dots, T; \quad i = 1, 2, \dots, n \quad (29)$$

will provide a first step to obtain $\sigma_{ij} = \frac{\sum_{t=1}^T e_{it}^* e_{jt}^*}{T-1}$ or $\frac{\sum_{t=1}^T e_{it}^* e_{jt}^*}{T}$, for all $i, j = 1, 2, \dots, n$, elements of the n by n covariance matrix \otimes of any t -th year's residuals. Notice that now, the first step estimator of r_i is the unweighted yearly mean of the weighted yearly means, i.e.: .

$$\hat{r}_i = \frac{\sum_{t=1}^T \frac{R_{it}}{K_{it}}}{T} \quad , \quad i = 1, 2, \dots, n \quad (30)$$

One can show that

$$Var(r_i) = \frac{\sigma_{ii}}{T} \quad , \quad i = 1, 2, \dots, n; \quad \text{and} \quad Cov(r_i, r_j) = \frac{\sigma_{ij}}{T} \quad , \quad i, j = 1, 2, \dots, n \quad (31)$$

As now

$$Cov(\otimes^*) = V = I_T \otimes \otimes \quad (32)$$

(\otimes denotes the Kronecker product of matrices) we could apply GLS to the system formed by the n equations of type (29) to obtain estimates of variances, covariances and of the estimators of r_i . However, under the new format, the right hand-side variables of the system are the same for all equation groups (a single constant term, always equal to 1) – rendering single-equation-OLS and GLS estimates identical. That is, $Cov(\hat{r}_{GLS}) = Cov(\hat{r}_{OLS})$ - obtainable as in (20) but with the new V - and we would approximate:

1) in the diagonal,

$$Var(r_i), \quad \text{and} \quad \frac{1}{T} Var(r_{it}) \quad (33)$$

2) off the diagonal

$$Cov(r_i, r_j), \quad \text{and} \quad \frac{1}{T} Cov(r_{it}, r_{jt}) \quad (34)$$

Journal of Economics Bibliography

One can show that under the assumption of independence between the

observations of each sector,
$$\text{Var}\left(\frac{\sum_{t=1}^T \frac{R_{it}}{K_{it}}}{T}\right) = \frac{1}{T^2} \sum_{t=1}^T \frac{1}{K_{it}} \textcircled{ii} > \text{Var}\left(\frac{\sum_{t=1}^T R_{it}}{\sum_{t=1}^T K_{it}}\right) =$$

$$\frac{1}{\sum_{t=1}^T K_{it}} \textcircled{ii} \text{ (the arithmetic mean is larger than the harmonic mean) - hence,}$$

the aggregate estimator is more efficient ⁶. Under perfect correlation of

within sector observations,
$$\text{Var}\left(\frac{\sum_{t=1}^T R_{it}}{\sum_{t=1}^T K_{it}}\right) = \frac{\sum_{t=1}^T K_{it}^2}{\left(\sum_{t=1}^T K_{it}\right)^2} \textcircled{ii} > \text{Var}\left(\frac{\sum_{t=1}^T \frac{R_{it}}{K_{it}}}{T}\right) =$$

$\frac{\sigma_{ii}}{T}$ (because a variance – of K_{it} – is positive) – then, the simple mean of yearly rates is more efficient than the aggregate ratio.

iii) Finally, admit that within each year t and for sector i , we have n_{it} – say, the existing firms in the industry – sets each with $\frac{K_{it}}{n_{it}}$ perfectly correlated observations among themselves, yet, independent across sets of the same sector. One can show that then $\text{Var}\left(\frac{R_{it}}{K_{it}}\right) = \frac{1}{n_{it}} \textcircled{ii}$ and $\text{Cov}\left(\frac{R_{it}}{K_{it}}, \frac{R_{jt}}{K_{jt}}\right) = \textcircled{ij}$.

$$\frac{R_{it}}{K_{it}}, \frac{R_{jt}}{K_{jt}} = \textcircled{ij}.$$

One can advance another weighted estimator for r_i based on n_{it} and available $\frac{R_{it}}{K_{it}}$:

$$\hat{r}_i = \frac{\sum_{t=1}^T \frac{R_{it}}{K_{it}} n_{it}}{\sum_{t=1}^T n_{it}}, \quad i = 1, 2, \dots, n \quad (35)$$

Its variance is now given by:

$$\text{Var}(\hat{r}_i) = \text{Var}\left(\frac{\sum_{t=1}^T \frac{R_{it}}{K_{it}} n_{it}}{\sum_{t=1}^T n_{it}}\right) = \frac{1}{\sum_{t=1}^T n_{it}} \textcircled{ii} \quad (36)$$

⁶ This is not surprising, once the LS regressions are weighted to provide conditions for increased efficiency of the estimators and we proved equivalence between the weighted means and the estimators of the first-step under independence, of the simple mean under perfect correlation...

Journal of Economics Bibliography

The average variance of the yearly means $Var(r_{it})$ would be estimated by:

$$\overline{Var(r_{it})} = \sum_{t=1}^T \left(\frac{R_{it}}{K_{it}} - r_i \right)^2 \frac{n_{it}}{\sum_{t=1}^T n_{it}} = \sum_{t=1}^T \left(\frac{R_{it}}{K_{it}} \right)^2 \frac{n_{it}}{\sum_{t=1}^T n_{it}} - \hat{r}_i^2 = T Var(\hat{r}_i), \quad i = 1, 2, \dots, n \quad (37)$$

Hence, an estimator for \otimes_{ii} is:

$$Var(r_{it}) = \sum_{t=1}^T n_{it} \left(\sum_{t=1}^T \left(\frac{R_{it}}{K_{it}} \right)^2 \frac{n_{it}}{\sum_{t=1}^T n_{it}} - \hat{r}_i^2 \right) / T, \quad i = 1, 2, \dots, n \quad (38)$$

We would propose:

$$Cov(\hat{r}_i, \hat{r}_j) = \sum_{t=1}^T \left(\frac{R_{it}}{K_{it}} - \bar{r}_i \right) \left(\frac{R_{jt}}{K_{jt}} - \bar{r}_j \right) \frac{n_{it} n_{jt}}{\sum_{t=1}^T n_{it} \sum_{t=1}^T n_{jt}}, \quad i \neq j, \quad i, j = 1, 2, \dots, n \quad (39)$$

and – as under the assumptions, $Cov(\hat{r}_i, \hat{r}_j) = \frac{\sum_{t=1}^T n_{it} n_{jt}}{\sum_{t=1}^T n_{it} \sum_{t=1}^T n_{jt}} \otimes_{ij}$ – for $i \neq j$, $i, j = 1, 2, \dots, n$:

$$Cov(r_{it}, r_{jt}) = \frac{\sum_{t=1}^T n_{it} \sum_{t=1}^T n_{jt}}{\sum_{t=1}^T n_{it} n_{jt}} Cov(\hat{r}_i, \hat{r}_j) = \frac{\sum_{t=1}^T \left(\frac{R_{it}}{K_{it}} - \hat{r}_i \right) \left(\frac{R_{jt}}{K_{jt}} - \hat{r}_j \right) n_{it} n_{jt}}{\sum_{t=1}^T n_{it} n_{jt}} \quad (40)$$

Under the assumptions, the error of (5) would be such that:

$$Var(\otimes_{it}) = \otimes_i^2 \frac{K_{it}^2}{n_{it}} = \otimes_{ii} \frac{K_{it}^2}{n_{it}}, \quad t = 1, 2, \dots, T; \quad i = 1, 2, \dots, n \quad (41)$$

The transformed system:

$$\frac{R_{it}}{K_{it}} \sqrt{n_{it}} = r_i \sqrt{n_{it}} + \otimes_{it}^* \quad t = 1, 2, \dots, T; \quad i = 1, 2, \dots, n \quad (42)$$

with \otimes_{it}^* homoscedastic, i.e., $Var(\otimes_{it}^*) = \otimes_{ii}$ but with $Cov(\otimes_{it}^*, \otimes_{jt}^*) = \sqrt{n_{it} n_{jt}} \otimes_{ij}$ for $i \neq j$, $i, j = 1, 2, \dots, n$. The single-equation estimators coincide with (35). The application of the GLS procedures of i) to the new format is straightforward. We would expect that the inferred $Cov(\hat{r}_{GLS})$ would approximate:

1) in the diagonal,

$$\frac{T}{T-1} \text{Var}(\hat{r}_i), \text{ and } \frac{T}{T-1} \frac{1}{\sum_{t=1}^T n_{it}} \text{Var}(r_{it}) \quad (43)$$

2) off the diagonal

$$\text{Cov}(\hat{r}_i, \hat{r}_j), \text{ and } \frac{\sum_{t=1}^T n_{it} n_{jt}}{\sum_{t=1}^T n_{it} \sum_{t=1}^T n_{jt}} \text{Cov}(r_{it}, r_{jt}) \quad (44)$$

Obviously, for $n_{it} = K_{it}$ expressions resume to those derived for case i). For $n_{it} = 1$, to case ii).

Under the new assumptions, $\text{Var}\left(\frac{\sum_{t=1}^T R_{it}}{\sum_{t=1}^T K_{it}}\right) = \frac{\sum_{t=1}^T \frac{K_{it}}{n_{it}} K_{it}}{\left(\sum_{t=1}^T K_{it}\right)^2} \otimes_{ii}$ and

$$\text{Var}\left(\frac{\sum_{t=1}^T \frac{R_{it}}{K_{it}}}{T}\right) = \frac{1}{T^2} \sum_{t=1}^T \frac{1}{n_{it}} \otimes_{ii},$$

the latter larger than (36) – the current weighted average is more efficient.

If each unit within each set was in fact independent and not perfectly

correlated as assumed, $\text{Var}\left(\frac{\sum_{t=1}^T \frac{R_{it}}{K_{it}} n_{it}}{\sum_{t=1}^T n_{it}}\right) = \frac{\sum_{t=1}^T \frac{n_{it}}{K_{it}} n_{it}}{\left(\sum_{t=1}^T n_{it}\right)^2} \otimes_{ii}$ and expected to be

larger than $\text{Var}\left(\frac{\sum_{t=1}^T R_{it}}{\sum_{t=1}^T K_{it}}\right) = \frac{1}{\sum_{t=1}^T K_{it}} \otimes_{ii}$. (Of course we are always assuming

that K_{it} and n_{it} – and $\frac{K_{it}}{n_{it}}$ – are deterministic, otherwise, we would stumble

into the fact that the variance of the product of even two statistically independent variables is related to their second moments in a different way that the expected value of their product...). If all the K_{it} observations are

perfectly correlated, $\text{Var}\left(\frac{\sum_{t=1}^T \frac{R_{it}}{K_{it}} n_{it}}{\sum_{t=1}^T n_{it}}\right) = \frac{\sum_{t=1}^T n_{it}^2}{\left(\sum_{t=1}^T n_{it}\right)^2} \otimes_{ii} > \text{Var}\left(\frac{\sum_{t=1}^T \frac{R_{it}}{K_{it}}}{T}\right) = \frac{\sigma_{ii}}{T}$:

the simple yearly mean is a more efficient estimator.

Some crude heteroscedasticity tests were performed on residuals of equation (29) in the hope to validate one of the three cases. Unfortunately results – summarized in the Appendix – are very vague.

Journal of Economics Bibliography

Possibly, we would like to have information at firm level – rather than by sector - and apply model ii) to the corresponding data set; arrangement iii) would somehow capture that pattern. Also, iii) would be preferable to i) due to arbitrariness of the relevant “unit” in i) – the one in which the K_{it} 's are measured: the estimated variance of the “unit” return is proportional to the sum of K_{it} 's and depends on its measurement unit – the units within 1 unit would be those we would be assuming to be able to replicate to, say, form a portfolio, which would not make much sense. Therefore, in the empirical research below, we only report results for cases ii) and iii). Still, i) has the intuitive advantage of producing for the OLS estimate of the expected value of a sector's unit return the general weighted mean of all the observed returns. Moreover, it provides the adequate procedures to produce variances and covariances of individuals' attributes from averaged (or aggregate) data.

Nevertheless, in the formation of a portfolio from assets of either type, one should only use the estimated σ_{ij} 's rather than $Cov(\hat{r})$ if one admits that only one firm can be picked to represent the i-th title – which would be rather unusual.

A final comment can be made with respect to environment iii): it maybe not be reasonable to assume that some units within a sector are mutually independent and yet consistently correlated to other sectors'... One could justify that theoretically only if there are links between sectors' – say, through intermediate product exchange – but not so much between units of the same sector. On the other extreme – and favoring now i) -, one could argue that at a given point in time a firm's assets are the result of different waves of investments, which indeed give a different return. Yet, we are left with the unknown measure of the relevant unit...

3. Sector economic rates of return: Source data and estimation results

Aggregate information on (gross, end-of-the-year) “Tangible Assets” and “Intangible Assets” of firms' balance sheets by industry and the corresponding yearly Trading Profits (“Excedente Bruto de Exploração”), along with the number of sampled units, is published by the official Portuguese statistical institution, I.N.E; we collected it from the statistical periodical: “Sistema de Contas Integradas das Empresas”, using information from the issues 1996-1997, 1997-1998, 1998-1999, (from all of which only the oldest year was used in the research), 1999-2000 and 2001-2002 ⁷. We constructed a pooled data set with information by industry (totaling $n = 23$

⁷ Each (annual) volume has information for two consecutive years and the number of covered enterprises of each sector and category – for representativeness purposes, we presume, which is convenient under our estimation logic - change even in the same issue. The periodical is quite recent, starting in 1994-1995 and data for firms with 20 to 99 people are only available from the 1996-1997 issue onwards; aggregates are reported in euros since the 1998-1999 issue.

Journal of Economics Bibliography

sectors, excluding Agriculture and most Financial Services and Institutions, with different accounting rules) and year (1996 to 2002 – T = 7) pertaining to firms with 20 or more employees⁸. Measurement units of aggregate data were all converted to 10⁶ PTE – at the rate 200.482 PTE per Euro. The coding used for sectors is reported below:

0. Total

3. Mining Industries

40. Manufacturing Industries: 4 to 17

4. Food, Beverages and Tobacco

5. Textiles and Clothing

6. Leather and Leather Articles

7. Woodwork and Cork Manufacturing

8. Paper, Graphical Arts and Publishing

910. Chemical Industries from Oil and Coke

11. Rubber and Plastic Articles

12. Other non-Metallic Minerals

13. Heavy Metallurgy and Metallic Products

14. Machinery and Equipment

15. Electric and Optical Equipment

16. Transportation Material

17. Non-Specified Manufacturing Industries

18. Electricity, Water and Gas

19. Construction and Public Infrastructure

20. Commerce

21. Restoration and Lodging

22. Transportation, Storage and Communications

24. Real Estate and Service to Firms

26. Education

27. Health and Social Service

28. Other Collective and Personal Services

Proceeding to the estimates of section 2, we inferred the sector rates of return, estimates of variances and covariances of both the rates as their estimators. Results are reported in Tables 1, 1.A and 1.B below. The first column exhibits the inferred sector average rate of return. In the last seven rows of the Tables, we report the simple correlation coefficients between each series and the estimates of the rates of return (\hat{r}_i), the variance of the rates ($Var(r_i)$), of the estimates ($Var(\hat{r}_i)$ - given the proportionality between the two variances under hypothesis ii), only one of them is reported), the sector's (yearly average) assets' share on the total (\bar{s}_{Ki}), the sector's profits

⁸ Published information offers disaggregation for each sector and year by two firm size classes – 20 to 99 people and 100 or more. Refinements of the procedures to benefit from it could be developed and are left for future research.

Journal of Economics Bibliography

share (\bar{s}_{Ri}), firms' (sampled units) share (\bar{s}_{ni}), the average firm's asset size ($\frac{\bar{K}_i}{n_i}$), the average firm's profit size ($\frac{\bar{R}_i}{n_i}$), and average – from a regression of the logarithm of the sector series on a time trend – annual growth rates of K_i , R_i , n_i , $\frac{K_i}{n_i}$, $\frac{R_i}{n_i}$ and r_i . (For 23 observations, correlations are significant at the 5% level for $|r|$ larger than 0.4142985; at the 10% level, for $|r|$ larger than 0.3522931.)

For hypothesis ii) of section 2 - results are depicted in Table 1 -, the elements of the covariance matrix of the residuals was derived according to (29). The average rate of return in the economy is around 11%. By industry – one can inspect the (descending) order/rank of the reported rates and variances by industry in the first columns -, the highest returns are found in sectors 20, 19, 27, 15, 11, 17 and 12; lowest, in 910, 3, 5, 22, 18, 21 and 28. The highest volatility occurs in sectors 6, 28, 27, 16, 26, 24 and 910; sectors 17, 14, 22, 21, 13, 18 and 4 exhibit low variances. The correlation between the average sector rates of return and their variance was -0.042122, with a p-value of .985. Hence, non-significant.

Tables 1.A and 1.B present the corresponding weighted estimates under assumption iii) of section 2. The correlation between the variances of these two tables is 0.53811. The correlation between the mean returns and the estimated variances of the unit application⁹ – Table 1.A - is 0.304706 and almost significant (which could be almost expected: higher unit variances would require higher returns – if we disregard the covariances effect); that with the variance of the mean returns – Table 1.B - is -0.043765.

The ordering of the variances of the mean returns – Table 1.B – coincides with the ones inferred in Table 1. The same is true for the estimated mean returns – even if under weighting, estimates are slightly larger¹⁰.

With respect to the correlations with the other variables, we note a consistent positive and significant relation between the firms share – and hence the number of firms in the sector – and the average returns, suggesting that more profitable sectors would attract entry, less profitable would push exit. And a negative relation between the sector's capital and profit share (hence total assets and profits importance) and the variance of the estimator of the rate of return – not present for the variance of the rate of return itself for case iii).

⁹ We report approximations inferred after (43)-(44), indeed close to the direct (first-step GLS) calculations of σ_{ij} .

¹⁰ Average yearly economic rates of return by sampled firm – an average of which would be similar to what we capture in our weighted estimates of type iii - were reported in the statistical publications of 1998-1999 and 1999-2000 only. In general, such average rates were larger than the average rates calculated from the ratios of the total sector's aggregate profits by the aggregate assets.

Journal of Economics Bibliography

A first attempt to isolate some pattern in the covariance structure involved the extraction of principal components¹¹ of the covariance matrix itself. The factor loadings – simple correlations between the components (linear combinations of the variables – here, the columns of the covariance matrix – designed to be linearly independent and explain decreasing importance of the total variance) and the variables are depicted in Table 2 for case ii), and 2.A. for the covariance of the estimates of case iii) (the factor loadings for the decomposition of $Cov(r_{it}, r_{jt})$ revealed themselves quite sparse – resulting from the very small correlations between the original elements – and all the eigenvalues are close to 1. So we only report the results of the decomposition of $Cov(\hat{r}_i, \hat{r}_j)$). We shade darkest the highest loading in each row, mildest the second highest – the most important components contributing for the variable's explanation. Correlations with the other series are in the last rows.

¹¹ Computed directly by TSP 4.4 PRIN routine, which standardizes the input variables – see Hall & Cummins (1997) and (1998).

Journal of Economics Bibliography

Table 2. *Principal Components: (Unweighted) Covariance Matrix, ii)*

Sector	PC1	PC2	PC3	PC4	PC5
Eigenv.	15.846333	2.7318894	2.1066607	1.6309184	0.64726145
% Cum. Exp Var.	0.68897100	0.80774880	0.89934274	0.97025224	0.99839404
Factor Loadings:					
3	0.92105	0.025139	0.37846	0.050718	0.070096
4	-0.93595	0.089876	-0.10038	0.27766	0.16563
5	0.95929	-0.097942	-0.069040	0.25239	-0.038378
6	0.98778	0.023115	0.090049	0.017406	0.12085
7	0.95583	0.10389	0.084197	0.17070	-0.18479
8	-0.76450	-0.11303	0.57164	0.24056	-0.13403
910	0.99507	-0.015713	-0.031166	0.076666	-0.050166
11	0.95086	-0.089440	-0.27484	0.088203	0.067467
12	0.63165	0.36936	-0.37403	0.54338	-0.14509
13	-0.76423	0.29174	0.34946	0.44815	0.047639
14	-0.86456	-0.18877	-0.15791	0.43293	0.060115
15	0.65875	-0.42719	-0.33871	0.44753	-0.25814
16	0.33356	-0.23152	0.61982	0.62099	0.25294
17	0.38922	0.67639	-0.52567	0.14468	0.29989
18	0.97822	-0.053728	-0.13658	0.040812	-0.14043
19	0.42147	-0.89219	-0.038287	-0.093359	-0.11587
20	0.80066	-0.39023	-0.28761	0.049184	0.34794
21	-0.85677	-0.35684	-0.29232	0.22132	-0.052460
22	0.98657	0.13776	0.048966	-0.057903	0.042433
24	0.62691	0.66432	0.28011	0.013829	-0.29496
26	0.97838	-0.089132	0.17732	0.038940	-0.040249
27	0.92156	0.25315	0.24191	-0.098975	0.13016
28	-0.85085	0.34588	-0.32986	0.13991	-0.16687
Cor. with					
\hat{r}_i	0.21261	-0.24588	0.03256	-0.09414	0.33360
$Var(\hat{r}_i)$	0.51124 *	0.29503	0.20053	0.26420	0.06375
\bar{s}_{ki}	-0.02685	-0.07695	-0.07604	-0.32090	-0.04399
\bar{s}_{Ri}	-0.02333	-0.20708	-0.12739	-0.32789	0.09768
\bar{s}_{ni}	-0.01079	-0.30895	-0.19869	-0.26630	0.17046
\bar{K}_i	-0.01487	-0.05911	-0.05066	-0.17245	-0.10448
\bar{n}_i					
\bar{R}_i	-0.00162	-0.08271	-0.03582	-0.13959	-0.10616
\bar{n}_i					
Kgr	-0.03499	0.51462 *	-0.07256	-0.38653 **	-0.10871
Rgr	-0.55093 *	0.43269 *	-0.17721	-0.22557	-0.05490
ngr	-0.04977	0.40939 **	0.10646	-0.51997 *	-0.08025
(K/n)gr	0.00938	0.33692	-0.28926	0.05981	-0.08035
(R/n)gr	-0.66029 *	0.31820	-0.27866	0.00000	-0.02506
rgr	-0.81022 *	0.23728	-0.21040	-0.02683	0.00546

In both cases, four components have eigenvalues larger than 1 – a possible criteria to choose the number of relevant components -, explaining more than 95% of the total variance. The first (most important) component seems to be in line with the sectors' variance of the rate of return estimate and with a negative trend of the rate of return. The covariances of sectors 3, 5, 6, 7, 910, 11, 12, 15, 18, 20, 22, 26 and 27 are strongly and positively related to it; those of sectors 4, 8, 13, 14, 21, and 28, negatively.

The second component moves oppositely to the number of firms in the sector. Sectors 17 and 24's covariances are positively related to it, sector 19's negatively.

Journal of Economics Bibliography

The fourth – associated to sector 16's covariances - moves oppositely to the sectors' capital and profits importance and to capital and number of firms' growth in the economy. The fifth appears slightly positively correlated with the estimated rates of return.

Table 2.A. Principal Components: Covariance Matrix, ii , $Cov(\hat{r}_i, \hat{r}_j)$

Sector	PC1	PC2	PC3	PC4	PC5
Eigenv.	15.157883	3.0995653	2.0962431	1.7937795	0.72412471
% Cum. Exp Var.	0.65903840	0.79380211	0.88494311	0.96293352	0.99441720
Factor Loadings:					
3	0.90841	0.025576	0.40609	0.013973	0.063052
4	-0.93483	0.11315	-0.079591	0.25107	0.20471
5	0.95752	-0.052869	-0.039204	0.27384	-0.023915
6	0.97844	0.036525	0.11463	-0.00056511	0.14536
7	0.94354	0.17324	0.11423	0.16816	-0.17271
8	-0.75831	-0.10939	0.58797	0.22355	-0.11579
910	0.99315	0.030872	-0.0039299	0.085126	-0.055300
11	0.95187	-0.066665	-0.26251	0.11469	0.072013
12	0.59887	0.43473	-0.32240	0.57306	-0.061627
13	-0.76193	0.37564	0.35615	0.37296	0.074364
14	-0.85614	-0.13392	-0.11170	0.46947	0.10643
15	0.67488	-0.31874	-0.26857	0.54900	-0.24188
16	0.29435	-0.15064	0.66235	0.58390	0.32848
17	0.31368	0.71697	-0.52345	0.067356	0.32334
18	0.96870	-0.060413	-0.15872	0.097423	-0.14998
19	0.46611	-0.87394	-0.032730	0.0046347	-0.11931
20	0.80294	-0.40108	-0.27865	0.044643	0.33556
21	-0.79548	-0.40912	-0.29779	0.32368	-0.052354
22	0.98750	0.12341	0.012724	-0.067480	0.048352
24	0.41222	0.80259	0.27793	0.075926	-0.31914
26	0.96777	-0.089952	0.18925	0.069453	-0.072180
27	0.86549	0.31599	0.19950	-0.25513	0.17159
28	-0.80643	0.43040	-0.33003	0.16005	-0.15014
Cor. with					
\hat{r}_i	0.17636	-0.27321	-0.00654	-0.11439	0.32669
$Var(\hat{r}_i)$	0.43971 *	-0.02306	-0.02696	-0.03727	0.21942
$Var(\hat{r}_i)$	0.50914 *	0.27675	0.24045	0.26254	0.13798
\bar{s}_{ki}	-0.02829	-0.12680	-0.10365	-0.27066	-0.07989
\bar{s}_{Ri}	-0.01501	-0.27150	-0.16121	-0.25849	0.06411
\bar{s}_{ni}	0.00179	-0.34988	-0.22450	-0.20319	0.14651
\bar{K}_i	-0.01205	-0.08085	-0.06149	-0.13885	-0.12880
\bar{n}_i					
\bar{R}_i	0.00112	-0.10075	-0.04389	-0.10613	-0.12588
\bar{n}_i					
Kgr	-0.09995	0.45957 *	-0.14511	-0.40008 **	-0.13801
Rgr	-0.57967 *	0.37416	-0.23727	-0.21191	-0.07181
ngr	-0.12801	0.32873	0.01443	-0.53006 *	-0.12244
(K/n)gr	0.00597	0.35518 **	-0.28623	0.04992	-0.07172
(R/n)gr	-0.65380 *	0.28884	-0.30378	0.02250	-0.02332
rgr	-0.80077 *	0.19325	-0.24242	0.00507	0.00371

Principal components on the average returns themselves (using only the 7 yearly observations) yielded the results in Tables 3 to 3.B, where the loadings of the first five components are depicted. We complemented the

Journal of Economics Bibliography

analysis computing the explanation ¹² of each PC (principal component) in the subsequent five columns – and dashed the highest contributors accounting at least 90% of each PC's variance (Usually, shaded cells are also dashed cells, even if not conversely.) We completed the analysis inspecting the correlations with yearly outside series – a Trend, total capital (K_t), profits (R_t), number of firms in the sample (n_t), average firm size in terms of assets ($\frac{K_t}{n_t}$) and of profits ($\frac{R_t}{n_t}$), reported in the last seven rows of the Tables - , and with the yearly capital shares of each specific sector and of the average firm asset size - reported in the last 10 columns of the Tables.

(For 7 observations, correlations are significant at the 5% level for $|r|$ larger than 0.7545450; at the 10% level, for $|r|$ larger than 0.6694306.)

Table 3 presents the results for standard principal components (with only 7 observations to generate the correlation matrix, a total of 6 components are extracted...). With high loadings in the:

- first component (all positive) are sectors 3, 5, 6, 7, 910, 11, 18, 22, 26 and 27. The component is negatively and very strongly related to the trend in the series (that also drives the outside series), negatively (but weakly) to the average rate of return.

- second component (all positive), 4, 12, 13, 14 and 28 - including metallurgy in general. It appears to be negatively (even if weakly) affected by the number of firms in the industry, positively (weakly) by the rate of return level.

- third component, 15, 19, 20, 21 and 24 (the only one positive);

- fourth component 8, 16 and 17 (the only one positive).

Table 3.A is derived from the application of principal components to the residuals of equation (42) – those presumed homoscedastic -, a procedure that subscribes to Weighted Principal Components ¹³ applied to those same average returns – decomposing the cross-moment matrix of the (hence, transformed) residuals of regression (42) divided by the estimated standard error. Results are in Table 3.A and implied the same clustering of the sectors and general results as the standard algorithm.

Finally, using the correlation matrix inferred from the covariance matrix of the weighted mean returns in the internal decomposition generated the results of Table 3.B, closer to those of Table 3: the first and second components remain mostly unaltered relative to Table 3, but not the others: with high loadings in the:

- first component are sectors 3, 4 (the only negative), 5, 6, 7, 910, 11, 18, 22, 26 and 27; negatively related to the trend.

¹² Computations were programmed with TSP 4.4, relying on matrix and database facilities – see Hall & Cummins (1997) and (1998). See Martins (2004) for further explanation and examples.

¹³ Computations were programmed with TSP 4.4, relying on matrix and database facilities – see Hall & Cummins (1997) and (1998). See Martins (2004) for a more detailed explanation of the method and some examples.

4. The unrestricted optimal portfolios

With mean returns, vector μ containing \hat{r}_i 's, and corresponding covariance matrix, $\otimes = [Cov(r_i, r_j)]$, we could proceed to the inference of the unrestricted optimal portfolio W , that is satisfying:

$$\underset{W}{Min} \quad W' \otimes W \quad \text{s.t.} \quad \mu' W \geq \alpha, \quad W_i \geq 0, \quad i = 1, 2, \dots, n \quad (45)$$

(of course, $W_i < \frac{\sum_{t=1}^T K_{it}}{T}$, or on average, it will be infeasible.) The general solution for the included assets/titles will be of the type ¹⁴:

$$W^* = \otimes^{-1} \mu (\mu' \otimes^{-1} \mu)^{-1} \alpha \quad (46)$$

provided that, being g the budget to be allocated and L a column vector of 1's ¹⁵ $\frac{\alpha}{g} < (\mu' \otimes^{-1} \mu) / (L' \otimes^{-1} \mu)$ – in other words, with unrestrictive funds.

We will concentrate on solution (46) and report the optimal shares:

$$\frac{W_\alpha^*}{L' W_\alpha^*} = \frac{\Omega^{-1} \mu}{L' \Omega^{-1} \mu} \quad (47)$$

If the problem has a unique solution, there will be a typical “optimal portfolio”, the shares of which are mean (α) invariant, that every investor buys a portion of and – under free market conditions - we would expect that,

for any i , W_i^* to be close to observed K_{it} ¹⁶ (or $\frac{W_i^*}{\sum_{i=1}^n W_i^*}$ to $\frac{K_{it}}{\sum_{i=1}^n K_{it}}$). Also:

$$\frac{\alpha}{\sigma^*} = (\mu' \otimes^{-1} \mu)^{1/2} \quad (48)$$

where σ^* denotes the square root of the optimized minimand; then $\frac{\alpha}{\sigma^*}$ (or its inverse) is independent of α . The unit return of the optimal portfolio, r^* , will also be α invariant:

$$r^* = \frac{\mu' W_\alpha^*}{L' W_\alpha^*} = \frac{\alpha}{L' W_\alpha^*} = \frac{\mu' \Omega^{-1} \mu}{L' \Omega^{-1} \mu} \quad (49)$$

¹⁴ See Tobin (1958), or Martins (2005).

¹⁵ See Martins (2005).

¹⁶ Again, one can argue that such assertion is only compatible with assumption ii) of section 2.

Journal of Economics Bibliography

Implicitly, investors will combine the optimal bundle, yielding this unit rate, with borrowing or lending at a zero rate... Of course, its standard deviation will obey (48).

The solution of (45) is that of a classical quadratic programme¹⁷. We relied on EXCEL's SOLVER for optimization to generate the efficient portfolio. Corner solutions – i.e., $W_1^* = 0$ – were found for several sectors and cases¹⁸. We therefore proceeded to a stepwise elimination of the highest optimal share sector and recalculated the optimal portfolio shares; we report in the Tables 4 and 4.A below (now, second moments were not multiplied by 1000000) - 4.A also contains the shares for the overall optimal portfolio when using the covariances of the returns themselves under case iii, once this resulted in no corner solutions - the results of such procedure applied to the covariance of average returns, for unweighted and weighted according hypothesis iii) respectively – where the iteration/order in which the sector was eliminated is registered, as well as the implicit number of “iterations” in which the sector was either included in the optimal portfolio or had already been eliminated, with the associated rankings. In two subsequent rows in the Tables, the inverse and the optimal standard deviation of the unitary mean yield is registered - α/σ^* (that coincides with Sharpe's performance index of the optimal portfolio when the risk-free rate is zero¹⁹) and σ^*/α ²⁰ - the slopes of the implicit opportunity locus or market line- below the effective shares we report the effective ratio (for case ii, the ratio of the arithmetic mean over the standard deviation of the yearly aggregate rates; for case iii, obtained dividing by the square root of $T=7$ the t-ratio of the coefficient of the regression, weighted by number of firms, of the yearly aggregate rates of return on a constant term). The next two rows report r^* and its standard deviation (the latter inferred by the product of r^* by the reported σ^*/α). The last seven rows register the usual correlations with the outside series.

The most important sectors in the sample period in terms of tangible and intangible assets – (actual) average industry capital shares are reported in the first column of the Tables, its descending order in the second - were 22, 18, 20, 4, 5, 910 and 24. The least important, 7, 3, 11, 6, 17, 26 and 27. (Correlation between the rankings of the actual sector shares weighted and unweighted estimates is very high: 0.99901.)

¹⁷ See Intriligator (1971), Taha (1982).

¹⁸ Notice that for case ii the estimated covariance matrix of all 23 sectors must be singular once only 7 observations are being used for its calculation – we would require at least 24 observations for its non-singularity. That implies that under such case we would not expect more than 6 ($T - 1$, for the corresponding estimated covariance matrix to be nonsingular - and hence invertible) sectors to be relevant to achieve an optimal portfolio...

¹⁹ Rolling tests based on the comparison of adjacent Sharpe ratios – in the spirit of Gibbons, Ross, & Shanken (1989)'s statistic (see also Glen & Jorion (1993), Campbell *et al* (1997), p.196) - may not be appropriate – both because n is smaller than $T - 1$, as even included sectors are sometimes smaller $T - 1$... That is, because we invariably register corner solutions and effectively included sectors do not intersect.

²⁰ For aggregate profits, the ratios are 7.26089 and 0.137724.

Journal of Economics Bibliography

For unweighted estimates, the correlations of the effective capital shares with: sector returns are -0.26331 ; estimated variances, -0.35924 . For weighted estimates, the correlations of the effective capital shares with: sector average returns are -0.26250 ; estimated variances, -0.039259 ; variances of the mean return: -0.35543 . That is, most signs are negative, but only significant (at 10%) for variances. Industry capital shares are close to profits shares, but not very much to the sector's number of firms; they are also positively related to firms' size but negatively related to size growth.

The first sectors to be found most representative in the optimal portfolio of unweighted estimates – Table 4 - were 4, 19, 21, 20, 17, 14 and 11; the last, 16, 5, 7, 910, 24, 28 and 6. Therefore, it appeared to be no correspondence between these and the actual relative importance of the sectors in the aggregate economy's assets (of course, in an open economy that would not be expected in any case). In fact, the Spearman's rank correlation test with the effective capital share order - on the rows labeled "Rank Corr" of the tables, along with the two-tailed critical value – inferred from Newbold (1995), Table 9, p. 843 - generated a positive but non-significant relation between the two series. The importance of a sector in the portfolio (negatively represented by the rank "Order") would appear – as expected – to be negatively related to the variances and positively to the average sector rate; no outside variables influence is significant.

For weighted estimates, we report two types of results in Table 4.A: those for $Cov(r_{it}, r_{jt})$, which generated a portfolio *per se*, and the stepwise results on $Cov(\hat{r}_i, \hat{r}_j)$ – the correlations between the rankings of sector importance implied by the two procedures is positive, but low: 0.38735.

The highest representation in the optimal portfolio from $Cov(r_{it}, r_{jt})$ of weighted estimates were found for 18, 4, 11, 3, 14, 27 and 15; the smallest for 19, 26, 28, 24, 20, 5 and 6. Spearman's rank correlation with the effective shares is negative and non-significant; yet, the (standard) correlation between the effective and optimal shares themselves – reported in the Table - is positive and significant: 0.54748. Moreover, firm size appears to be strongly and positively related to representation in the optimal bundle – with the correlations with the sectors ranking suggesting: a strong and negative influence of the variance of the returns, and of the number of firms in the industry in the determination of the sectors' portfolio representation.

The first sectors to be found most representative in the optimal portfolio from $Cov(\hat{r}_i, \hat{r}_j)$ of weighted estimates - were 4, 19, 21, 20, 18, 14 and 8; the last, 24, 7, 28, 910, 16, 26 and 6. Spearman's rank correlation test with the actual capital shares is again positive and almost significant at the 10% significance level.

The correlation between the sector portfolio rankings of the unweighted with those of the weighted estimates for $Cov(\hat{r}_i, \hat{r}_j)$ is, in any case, – and as expected - very high: 0.92292. We recover most traits of the pattern of association with outside variables: it is the variance – sector risk –

A.P. Martins, JEB, 9(3), 2022, p.99-136.

5. The effective market “Betas”

A complementary question, entailing a quite different perspective, would unravel how far away is each sector from the economy’s “market line” – assuming the current distribution of asset values across sectors to be optimal (of course, this can only be assumed in the long-run) – and if far enough to reject the possibility of aggregate “optimality”. To some extent, the positive correlations found between the optimal portfolio’s shares and actual market shares and on the corresponding sectors’ rankings previously advanced can be seen as a sign that on the aggregate the argument could hold – and the inspection of the significance of such correlations as tests on the market financial efficiency. Another perspective is derived from theoretical implications of optimization – implying regression analysis, sometimes pursued in empirical work:

With a riskless asset of zero yield, with u_i denoting the aggregate return of sector i – u the corresponding column vector for the n assets -, and $R^M = u'W^*$ that of the optimal portfolio W^* , we expect that ²¹:

$$E[u_i] = E[R^M] \frac{Cov(u_i, u'W^*)}{Cov(u'L, u'W^*)}$$

If we are assuming the current portfolio to be the optimal one, $W^* = L$ and $Cov(u'L, u'W^*) = Var(u'W^*)$. This suggests:

$$\bar{R}_i = \bar{R} \frac{Cov(R_i, R)}{Var(R)} \quad (50)$$

where $R_t = \sum_{i=1}^n R_{it}$, approximating total trading profits in year t . This suggests calculating $\frac{\bar{R}_i}{\bar{R}}$ and $\frac{Cov(R_i, R)}{Var(R)}$ by sector, then, regressing, over i (the whole n sectors), one on the other – excluding the intercept – and testing the equality to 1 of the unique parameter.

Of course, the expression also applies to unit returns ²², r_{it} 's – interchangeability being appropriate under assumption ii) of section 2. Then:

$$\bar{r}_i = \bar{r} \frac{Cov(r_i, r)}{Var(r)} \quad (51)$$

²¹ See Martins (2005). Also, Campbell, Lo & MacKinlay (1997).

²² See Brealey & Myers (2003).

Journal of Economics Bibliography

with $\frac{Cov(r_i, r)}{Var(r)} = \beta_i$ corresponding to the well-known beta of asset i .

We considered two approaches to the estimation of $\frac{Cov(R_i, R/n)}{Var(R/n)}$ - for

which we used for $R = \sum_{i=1}^n R_{it}$ and $n = 23$ ²³ -, and of $\frac{Cov(r_i, r)}{Var(r)}$:

a) directly, calculating the numerator and denominator of such “betas” (generating estimator B1 below).

b) from the (n) coefficient(s) of the regression(s) $r_{it} = a_i + b_i r_t + v_{it}$. As the estimate of b_i would approach the former, a_i was fixed to 0 (estimator B2).

Optimality testing then relies on the fact – arising from (50) and (51) - that the ratio between mean returns - $\frac{\bar{r}_i}{r}$, $n \frac{\bar{R}_i}{R}$ - should equal such estimates²⁴.

The several indicators are depicted in Table 5 below – where we also register in two middle columns the relative firm size of the sector in terms of both assets and profits (positively related to sector’s capital and profit shares).

A first general test on the significance of a trend in a regression on the yearly rates by sector – in the last columns of the table - resulted in the non rejection of the null for most (but not all) cases – negative and significant at 5% only for sectors 5, 6, 7, 9, 10, 11, 12, 18, 22 and 26. The same is true for its squared deviation from the general mean (only five sectors show a significant trend at 5%). Hence, stationarity of mean unit returns, and possibly of their variances – is suggested. The shares (s_{it} ’s) on total assets did exhibit a significant trend for the majority of the 23 sectors, though (correlation between the trend coefficients of rates and share regressions were non-significant: -0.038410). Interestingly, the size of the trend coefficient on the rates of return regressions is significantly negatively related to the sector’s rate of return variance. Of the capital share regressions, negatively and strongly related to average firm size.

The correlation of B1 and B2 is high for aggregate profits (0.81556), but low and even negative for rates (-0.32617). Correlation between the two B1’s is again positive, 0.39925, and between the two B2’s 0.38581 (both significant at 10%) – yet, as noted, we would not expect the latter coincidence. The correlations of the relative rates and betas with outside variables only exhibited a consistent (positive) sign for number of firms in the sector.

²³ That provides some standardization; even so, $\frac{Cov(r_i, r)}{Var(r)}$ and $\frac{Cov(R_i, R/n)}{Var(R/n)}$ are, of course, expected to differ.

²⁴ One can argue that such test would be more reliable for direct betas, once indirect betas already use (50) and (51). Nevertheless, it would stand in this case as an overall performance test.

6. (Weighted) Method of order statistics under truncated normal assumptions

. If one advances a particular cdf for the distribution of the unit rates of return, weighted principles can again be applied to produce adequate inference. Of general importance, truncated distribution assumptions, specially lower truncation, have meaning in finance, where a lower bound – even if it does not have to...-may correspond to cut-off, acceptance, points of investment profitability. Moreover, range of the random variable – implicit in double truncation - is also important as a dispersion (risk) measure. We consider below the three cases of a truncated normal – that is, of a truncated standard normal of the standardized residuals of adequate single-parameter regression models. We use the method of order statistics (MOS)²⁵, providing the Minimum Distance estimators, corresponding standard errors, and the identifying restrictions statistic $e' W^{-1} e$ where W denotes the appropriate covariance matrix.

Table 6. provides the inference for form (29), i.e. compatible with case ii) of section 2:

$$\frac{Rank_i(r_{it})}{T+1} \approx E\left\{\frac{\Phi\left(\frac{r_{it}-r_i}{\sigma_i}\right)-\alpha_i}{\beta_i-\alpha_i}\right\}, \quad t=1,2,\dots,T \quad , \quad \alpha_i < \Phi\left(\frac{r_{it}-r_i}{\sigma_i}\right) < \beta_i \quad \text{for all } i \quad (52)$$

For each sector i , the lower and upper truncation probabilities were translated into truncation points on the rates of return range through:

$$a_i = \hat{r}_i + \hat{\sigma}_i \Phi^{-1}(\hat{\alpha}_i) < r_{it} < \hat{r}_i + \hat{\sigma}_i \Phi^{-1}(\hat{\beta}_i) = b_i, \quad \text{for all } i \quad (53)$$

Estimates are obtained from nonlinear least squares applied to (51). The covariance matrix of the estimates for sector i from $C\hat{O}V(\hat{\theta}_{MDMOS}) = [G(\hat{\theta})' G(\hat{\theta})]^{-1} G(\hat{\theta})' W G(\hat{\theta}) [G(\hat{\theta})' G(\hat{\theta})]^{-1}$, where $W = [w_{it}] = \left[\frac{Min[Rank_i(r_{it}), Rank_i(r_{it})]}{T+1} \left\{ 1 - \frac{Max[Rank_i(r_{it}), Rank_i(r_{it})]}{T+1} \right\} \right]$ and $G(\hat{\theta})$ the

matrix of derivatives of $\frac{\Phi\left(\frac{r_{it}-r_i}{\sigma_i}\right)-\alpha_i}{\beta_i-\alpha_i}$ with respect to the four

parameters. In $e' W^{-1} e$, the identifying restrictions test, e refers to the vector estimated differences (residuals...) between the left and right hand-side of

²⁵ Martins (2005b). The range restriction was not embedded in estimation.

²⁶ Ranks are measured ascendingly here: Rank of r_{it} goes, for each sector i , from 1 (smallest r_{it}) to T (largest).

Journal of Economics Bibliography

(52), expected to have under the correct distribution $\otimes_{(T-k)}^2$ where k denotes the number of estimated parameters.

In Table 6.A stages an analogous setting to form (42) – case iii): denoting by e_{it}^* the (first-step) OLS residual for the t -th observation of the i -th sector of such equation, MOS fits ²⁷:

$$\frac{Rank_i(e_{it}^*)}{T+1} \approx E\left\{ \frac{\Phi\left[\frac{\sqrt{n_{it}}(r_{it}-r_i)}{\sigma_i}\right] - \alpha_i}{\beta_i - \alpha_i} \right\}, \quad t = 1, 2, \dots, T, \quad \alpha_i < \Phi\left[\frac{\sqrt{n_{it}}(r_{it}-r_i)}{\sigma_i}\right] < \beta_i \text{ for all } i \quad (54)$$

The truncation points were evaluated at ²⁸:

$$a_i = \hat{r}_i + \frac{\hat{\sigma}_i}{\sqrt{Max_t n_{it}}} \Phi^{-1}(\hat{\alpha}_i) < r_{it} < \hat{r}_i + \frac{\hat{\sigma}_i}{\sqrt{Min_t n_{it}}} \Phi^{-1}(\hat{\beta}_i) = b_i, \quad \text{for all } i \quad (55)$$

For each Table, we produce the untruncated distribution estimates (for fixed $\alpha = 0$, $\beta = 1$); the lower single - (for fixed $\beta = 1$) truncation case, and the doubly-truncated one.

The estimates, including the range ones, exhibit general consonance regardless of the heteroscedasticity assumption (ii or iii) – of course, they differ for each according to the truncation context. Also, identifying restrictions would not seem to reject normality (at – upper tail - 5%, $\otimes_{(5)}^2 = 11.07$; $\otimes_{(4)}^2 = 9.49$; $\otimes_{(3)}^2 = 7.81$; therefore, the order restrictions would not be rejected in most cases at that significance level – moreover, the tests are only asymptotically valid...)

Not always truncation seems required – evident from the number of cases with empty cells inferring truncation points, corresponding to estimates of α smaller than 0 or β larger than 1.

(Notice that, say for case ii), standard errors of the plain mean estimates would be the reported sd divided by $\sqrt{7}$ – which are, for our sample, in general, smaller than the reported standard errors of \hat{r}_i from the untruncated MDMOS estimates... We are not implying, therefore, that the latter should be used instead – even if GMOS would eventually perform better. For truncated distributions, they are an alternative available to infer the true distribution parameters.)

²⁷ For case i, the same procedure could – and was – applied, with n_{it} replaced by K_{it} .

²⁸ Notice that, effectively, the distribution and distribution range changes with $\sqrt{n_{it}}$. We evaluate the extremes at such points...

Journal of Economics Bibliography

they range (even if the range has a lesser meaning under iii ³¹) between 2.8 to 29.1%, in sectors (from longer to shorter range) 6, 11, 26, 7, 20, 5, 21, 13. Some sectors of short rate of return range are also sectors of low (untruncated variance estimates) volatility.

³¹ See footnote 30.

Journal of Economics Bibliography

References

- Black, F. (1972). Capital market equilibrium with restricted borrowing. *Journal of Business*, 45, 444-455. doi. [10.1086/295472](https://doi.org/10.1086/295472)
- Brealey, R.A., & Myers, S.C. (2003). *Principles of Corporate Finance*. McGraw-Hill Higher Education. 7th Edition.
- Campbell, J.Y., Lo, A.W., & MacKinlay, A.C. (1997). *The Econometrics of Financial Markets*. Princeton University Press.
- Cuthbertson, K. (1996). *Quantitative Financial Economics*. John Wiley and Sons.
- Dhrymes, P.J. (1978). *Introductory Econometrics*. Springer-Verlag.
- Gibbons, M.R., Ross, S.A., & Shanken, J. (1989). A test of the efficiency of a given portfolio. *Econometrica*, 57(5), 1121-1152. doi. [10.2307/1913625](https://doi.org/10.2307/1913625)
- Glen, J., & Jorion, P. (1993). Currency hedging for international portfolios. *Journal of Finance*, 48(5), 1865-1886. doi. [10.1111/j.1540-6261.1993.tb05131.x](https://doi.org/10.1111/j.1540-6261.1993.tb05131.x)
- Greene, W.H. (2003). *Econometric Analysis*. Prentice-Hall. 5th Edition.
- Hall, B.H., & Cummins, C. (1998). *TSP 4.4 Reference Manual*. TSP International.
- Hall, B.H., & Cummins, C. (1997). *TSP 4.4 User's Guide*. TSP International.
- Instituto Nacional de Estatística. *Sistema de Contas Integradas das Empresas*. Lisboa, I.N.E.
- Intriligator, M.D. (1971). *Mathematical Optimization and Economic Theory*. Prentice-Hall.
- Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, 47(1), 13-37. doi. [10.2307/1924119](https://doi.org/10.2307/1924119)
- Martins, A.P. (2004). *Segmented Life-cycle Labor Markets – Portuguese Evidence*. Mimeo, presented at UCP, Lisbon.
- Martins, A.P. (2005). Portfolio selection – A technical note. Mimeo, presented at UCP, Lisbon, and at the *EEFS 2005 Conference*, Coimbra.
- Martins, A.P. (2010). On ordered principles: Order regression, inter-quantile inference and truncated distributions. *International Journal of Statistics and Economics*, 5(Special), A10, 64-103.
- Newbold, P. (1995). *Statistics for Business and Economics*. Prentice-Hall. 4th Edition.
- Sharpe, W. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3), 425-442. doi. [10.1111/j.1540-6261.1964.tb02865.x](https://doi.org/10.1111/j.1540-6261.1964.tb02865.x)
- Taha, H.A. (1982). *Operations Research: An Introduction*. Macmillan. 3rd Edition.
- Tobin, J. (1958). Liquidity preference as behavior towards risk. *Review of Economic Studies*, 25(2), 68-85. doi. [10.2307/2296205](https://doi.org/10.2307/2296205)



Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by-nc/4.0>).

