Relatórios Coppead é uma publicação do Instituto COPPEAD de Administração da Universidade Federal do Rio de Janeiro (UFRJ)

Comissão de Pesquisa
Angela Rocha
Rebecca Arkader
Ricardo Leal

Gerência de Publicações
Regina Helena Meira de Castro

Editoração Eletrônica
Regina Helena Meira de Castro

Referenciação e Ficha Catalográfica
Ana Rita Mendonça de Moura

Leal, Ricardo Pereira Câmara.
Using robust portfolios techniques in emerging markets / Ricardo Pereira Câmara Leal; Beatriz Vaz de Melo Mendes. – Rio de Janeiro: COPPEAD/UFRJ, 2003.

13 p. ; 27 cm. – (Relatórios COPPEAD; 357)

ISBN 85-7508-048-2
ISSN 1518-3335

1. Finanças. I. Mendes, Beatriz Vaz de Melo. II. Título. Série.

CDD - 332

Pedidos para Biblioteca
Caixa Postal 68514 – Ilha do Fundão
21941-970 – Rio de Janeiro – RJ
Telefone: 21-2598-9837
Telefax: 21-2598-9835
e-mail: biblioteca@coppead.ufrj.br
Home-page: http://www.coppead.ufrj.br

USING ROBUST PORTFOLIOS TECHNIQUES IN EMERGING MARKETS

Ricardo Pereira Câmara Leal[1]
Beatriz Vaz de Melo Mendes

Financial data are heavy tailed containing some proportion of extreme observations. We propose to use a robust covariance estimator to define the center and orientation of the data. We provide an illustration of the usefulness of the proposed procedure to efficiently allocate among emerging stock markets. We show that the resulting robust portfolios may yield higher cumulative returns and have more stable weights. We strongly recommend that a robust covariance matrix is used to solve emerging stock markets allocation problems. We believe that our technique has a key advantage. Because all we change is the covariance matrix, we can use any commercially available optimizer to obtain robust portfolio weights.

Key Words: Robust Estimation; Multivariate Financial Data; Outliers; Mean-Variance Optimal Portfolios.

1. INTRODUCTION

Several models in finance rely on simplified assumptions. For example, the Mean-Variance (MV) model of Markowitz (1959) assumes the multivariate normal distribution for a collection of independent and identically distributed (iid) assets. Based on this assumption, the resulting procedure simply requires estimates of the center and covariance matrix of the data as inputs to obtain the efficient frontier weights.

In this context, the classical sample mean and sample covariance estimators are the maximum likelihood estimators and possess desirable statistical properties. However, their

[1] Acknowledgements: Ricardo Leal thanks Datastream, the financial support of CNPq, the Brazilian Scientific and Technological Development Council, and research support by Daniel C. Thomas.

asymptotic breakdown point is equal to zero (Maronna, 1976), which means that they are badly affected by extreme observations and may become meaningless.

Extreme observations are even more common in emerging markets. They may or may not be considered outliers (this is a polemic discussion topic), but they certainly seem to be related to a data generating process different from the one generating the vast majority of the observations. Even though these atypical observations constitute a small proportion of the data set, they are often associated with some type of crisis and are of great interest in finance. However, statistical analysis of low probability tail events should be made through specific models, such as extreme value models (Hartman, Straetmans and De Vries (2001) and Embrechts, Klüppelberg, and Mikosch (1997)), or regime-switching models (Ang and Bekaert, 2002).

The effects of atypical points on the ellipsoid associated to an estimate of the covariance structure (Johnson and Wichern, 1990) are at least two: (1) they may inflate its volume; (2) they may tilt its orientation. The first effect is related to inflated scale estimates. The second is the worst one, and may show up as switching the correlations' signs. The concept of breakdown point is related to the amount of extreme values which can "break down" the estimator. It is a measure which tells us what is the maximum fraction of atypical values in the sample with which the estimator still gives reliable information. For example, the breakdown point of the sample mean is zero, the smallest possible value, reflecting its high sensitivity to extreme values.

In this paper we use a variation of the well known high breakdown point Minimum Covariance Determinant (MCD) estimator to obtain robust efficient frontiers and construct robust portfolios in emerging markets. The remaining of this paper is organized as follows. In Section 2 we propose a robust estimation procedure for the inputs of the MV model. To illustrate, we use emerging markets data in Section 3. We compare the performances of the robust and classical MV optimal portfolios, and show that the robust portfolios may yield higher cumulative returns and seem to possess more stable weight structures. In Section 4 we summarize the results.

2. INPUTS FOR THE MV-MODEL

To obtain a good representation for the p-dimensional data we propose to estimate the covariance matrix using

$(1-\varepsilon)\mathbf{\Sigma}_1 + \varepsilon\mathbf{\Sigma}_2$ (1)

where ε is some contaminating proportion.

The $p \times p$ covariance matrix $\mathbf{\Sigma_1}$ represents the (predominant) dependence structure of the usual business days, or, in other words, the covariance structure of the data cloud without the outliers. $\mathbf{\Sigma_2}$ is the covariance matrix of an extended data cloud containing also most of the atypical points.

In equation 1, the ellipsoids associated to $\mathbf{\Sigma_1}$ and $\mathbf{\Sigma_2}$, for fixed $\mathbf{x}$ have the same orientation but different volumes. These characteristics are derived from the choice of the same eigenvectors for $\mathbf{\Sigma_1}$ and $\mathbf{\Sigma_2}$. In practice, because ε is small, the contaminating distribution in equation 1 typically produces spurious extreme observations seeming to follow an orientation structure different of that observed during usual days, or $\mathbf{\Sigma_1}$. These are the observations occurring during stress periods when we may observe different (greater) correlations. When using the classical sample covariance matrix S, these few points can tilt the orientation of the axes of its associated ellipsoid.

To avoid this problem, we propose to estimate the correct orientation of the data using the high breakdown point Minimum Covariance Determinant (MCD) estimate (Rousseeuw, 1985). The volumes of $\mathbf{\Sigma_1}$ and $\mathbf{\Sigma_2}$ will be estimated based, respectively, on the volumes of the robust MCD and classical S covariance estimates. We estimate the proportion ε empirically.

The MCD is a covariance affine equivariant estimator which attains the maximum possible breakdown point (approximately 0.5). For a given integer h, the MCD location estimator $\hat{\mu}_h(\mathbf{X})$ is defined as *the mean of the h points of* $\mathbf{X} = (x_1, x_2, ..., x_n)$ *for which the determinant of the sample covariance is minimal*.

Let us denote by h^+ the number of points used to obtain the MCD. Thus, $n - h^+$ extreme points were not used to compute the covariance estimate and this information is used to empirically compute the proportion ε. We can interpret the MCD estimator as able to measure the “outlyingness” of any data point relatively to the center of the collection.

In summary, we assume that a set of assets returns $X \in \Re^p$ possess a distribution that is a mixture of two elliptical distributions with same center, and covariance matrices intended to represent the usual days and most atypical days. This fraction ε of observations results from days when larger volatility is observed with no change on the strength and sign of correlations. Contamination here does not mean errors and it is just a mechanism to model the data. We denote the robust estimators of the location and covariance matrix of the data by $\left(\widehat{\mu}_h, \widehat{\Sigma}_{\mathbf{emp}}\right)$, where ε is estimated empirically and, unlike the MV-model inputs, no assumption is made about the data's underlying distribution.

In the next section we provide a practical illustration of the financial applications of the proposed model in emerging stock markets country allocations. In all MV optimizations carried on we use positive weights and ex post μ.

3. PERFORMANCE OF ROBUST PORTFOLIOS: HIGHER ACCUMULATED YIELDS?

The MV Model is probably the most used in practical asset allocation applications. Estimation of the efficient frontier is almost always done via the sample mean $\bar{x}$ and sample covariance matrix ***s***. However, different statistical estimates define different efficient frontiers. One of the most important limitations of MV optimization in practice is the lack of optimality presented by these classical estimates.

We now use the estimates $\left(\widehat{\mu}_h, \widehat{\Sigma}_{\mathbf{emp}}\right)$ as inputs for an asset allocation exercise with the MV model to construct robust portfolios that should reflect the behavior of both the usual and the higher volatility days. We stress that they do not reflect extreme crises. We note that portfolios constructed based on high breakdown point estimates are meant to be used for long term objectives, since they capture the dynamics of the majority of the business days.

The seven markets in our emerging market portfolio exercise are: Argentina, Brazil, Korea, China, Taiwan, South Africa, and Mexico. We have 2151 daily returns from July 3, 1995 through September 29, 2003. The indices have been obtained from Datastream and are all computed by S&P from the former IFC (International Financial Corporation) Global indices. The indexes are market capitalization weighed. The market capitalization of the constituents of the S&P/IFCG indices exceeds 75% of all domestic shares listed on the local exchange. Index computation details may be obtained at Standard and Poor's website. We used both local currency and dollar returns.

3.1 First Empirical Exercise

We perform out-of-the-sample analysis of several aspects of the optimal robust and classical portfolios and investigate, in particular, which one could yield higher cumulative returns. To this end, we split the data in two parts. The first part of the data, the estimating period (1870 daily observations), is used to compute the robust and classical inputs for the MV optimization procedure. The second part, the testing period (281 daily observations), is used in the comparisons. We are interested in the cumulative returns at the end of the testing period.

Thus, we analyze the trajectory of the portfolios' cumulative returns in the testing period. Three portfolios in the efficient frontier were used in the comparisons: (a) the portfolio possessing a fixed target daily percentage return v, say, $v=0.041\%$; (b) the minimum risk and (c) the maximum return portfolios. Note that even though the robust and the classical portfolios have the same target expected daily return value of 0.041%, they belong to completely different regions in their respective efficient frontiers. The robust one lies in a low risk region while the classical lies in a high risk region. The reason is that the two efficient frontiers lie on different regions of $\Re^2$.

The portfolios' performances are assessed by computing their allocations at baseline $t = 1870$ (given in Table 1), which is the end of the estimating period, through $t = 2151$, the end of the testing period. The weights were kept fixed during the testing period. The three robust portfolios have lower risk than their classical counterparts. The asset allocation for the robust portfolios is also better distributed among markets and is quite different from the classical portfolio's weights. Our first empirical exercise indicates that our robust contamination technique yields portfolio weights that dominate the classical portfolio weights in emerging markets country allocations, at least for the period examined.

Table 1

Portfolios compositions at the baseline $t=1870$, based on the robust and classical inputs. Daily US dollar returns from S&P/IFCG indexes reported.

	% Daily Return	% Risk (St. Dev.)	WEIGHTS						
			ARG	BRAZ	KOR	CHI	TAI	AFR	MEX
(a) Fixed Target (0.00041) Return Portfolios									
Robust	0.041	0.775	0.085	0.051	0.016	0.280	0.134	0.305	0.129
Classical	0.041	1.539	0.000	0.000	0.000	0.897	0.000	0.000	0.103
(b) Minimum Risk Portfolios									
Robust	0.033	0.770	0.130	0.029	0.030	0.252	0.147	0.299	0.112
Classical	0.000	1.003	0.088	0.012	0.015	0.296	0.201	0.238	0.151
(c) Maximum Return Portfolios									
Robust	0.090	1.419	0.000	0.000	0.000	1.000	0.000	0.000	0.000
Classical	0.043	1.699	0.000	0.000	0.000	1.000	0.000	0.000	0.000

Figure 1 shows the cumulative returns of the fixed 0.041% return portfolios. It displays returns accumulated over the one year testing period. The portfolio constructed using $\left(\hat{\mu}_h, \hat{\Sigma}_{\mathbf{emp}}\right)$ (the black line) dominates the classical one (the gray line). We note that even though the mean returns of the robust and the classical portfolios are the same, it seems that the robust method, being more truthful to the data, is more successful when composing the portfolios. We repeated the analysis using local currency returns. The results are qualitatively the same and we do not report them, but make them available upon request, as usual.

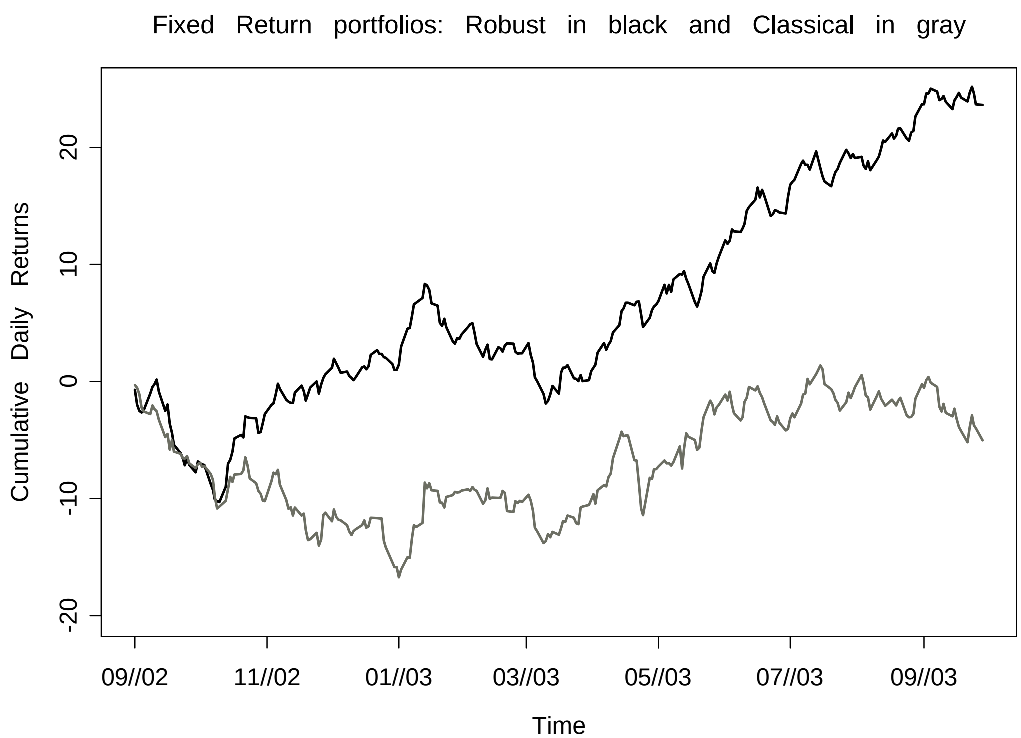

Figure 1: Cumulative (%) daily returns of portfolios with target daily return equal to 0.041%. The black line corresponds to the robust portfolio. The gray line to the classical portfolio.

3.2 Second Empirical Exercise

In our first exercise we obtained the portfolio weights at the base day $t=1870$ and studied the portfolio behavior in the following 281 days. However, it is possible that the portfolio should be rebalanced more often or that this time horizon is too long. In our second exercise we rebalance the portfolio by computing its covariance matrix and weights every 10 days. The baseline times now are t = 1870, 1880, 1890, ..., 2050. We have 19 baseline or estimation times. Thus we have 19 baseline portfolios and for each one we compute 19 trajectories and 19 accumulated returns over the following baseline plus 100 days period. There are two objectives to this second exercise: (1) to assess the stability of the covariance estimates, as this stability carries over to the weights; (2) to assess and compare accumulated gains over a shorter time horizon of about 4 months (the first exercise assumed approximately a 1 year horizon). We update the portfolio weights more frequently because economic changes would be captured by the estimates. At each baseline point we are enlarging the data set, and adding more information. Then we examine the distributions of the returns and risks of the robust and classical portfolios at two points of the time: at the baselines (t=1870, 1880, ..., 2050), and also at the end of each of the 100-day periods (t=1971, ..., 2151).

The out-of-the-sample performance of the portfolios depends upon their intrinsic characteristics, as well as on whether or not the testing and the estimating periods are compatible. In other words, for the comparisons to be meaningful, the inputs computed with and without the observations in the testing period should be close.

We compute six portfolios at each baseline: the minimum risk (Mi), the maximum return (Ma), and a "central" portfolio (Me), using the classical and the robust covariance matrix. The central portfolio return is the average between the returns of the portfolios of minimum risk and maximum return over its respective efficient frontier. By choosing the "central" portfolio, we aim to characterize a portfolio designed for investors with about the same degree of risk aversion, half way between the minimum and maximum risks for any given efficient frontier. The cumulative return over a 100-day period is computed for each portfolio. Then, the following 10 observations (t = 1871 to t = 1880) are added to the estimating sample. All computations are repeated, robust and classical portfolios of the three types (Mi, Ma, Me) are obtained at the baseline t=1880, and estimates for the final value of the 100-day accumulated returns of all portfolios are saved. We repeat this process until we have 2051 observations in the sample, thus obtaining 19 representations of the returns of the (6) portfolios at the baselines and at the end of each 100-day period.

Figure 2 shows the distribution of the portfolio returns and risk at the baselines. These are their past returns at the baselines. The notations RMi, RMe, and RMa (CMi, CMe, CMa) stand for the robust (classical) portfolios of the three types. The plot at left shows the returns. We observe that the robust portfolios are more promising, with a distribution located at higher values and possessing smaller variability. For example, for the minimum variance portfolios, the smaller observed robust value was greater than the highest observed for the classical portfolios. We also carried out a formal paired t-test to verify the equality of the means of returns. For the three portfolio types the p-value was zero against the alternative hypothesis of the robust mean return being greater than the classical. The risks associated to the 19 portfolios are box-plotted at the right hand side of Figure 2. We also note a smaller variability of the (also smaller) robust quantities.

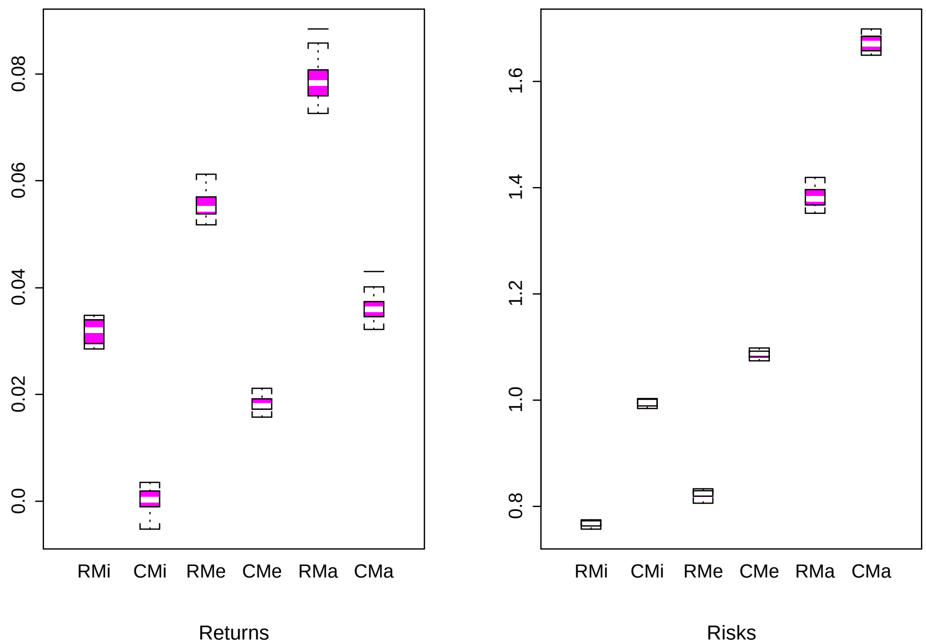

Figure 2: Distribution of the returns (left) and risks (right) of the 19 baseline portfolios (t = 1870, 1880,...) for each portfolio type (minimum risk, central, and maximum return) under the robust (R) and classical (C) approaches.

All of this was the past. Do the robust portfolios deliver in the testing period? We investigate the distribution of the US dollar returns by examining their 100-day accumulated values associated to the 19 baseline portfolios. Table 2 summarizes our results. We observe that the accumulated returns distributions of the robust portfolios are located to the right of the classical ones for all portfolio types. For example, the central robust portfolio median is 6.37%, whereas the classical central portfolio median is 4.77%. There is a 160 basis point return difference in 100 days, which is highly significant from a financial stand point. The maximum return portfolios are not given but their results are qualitatively similar, albeit less relevant.

Table 2

Quantiles of the daily US dollar return distribution of the 100-days cumulative returns of the minimum variance and central portfolios according to the robust and classical estimation procedures.

	Probabilities				
	0.05	0.25	0.50	0.75	0.95
Minimum Risk Portfolios (Mi)					
Robust	3.5458	6.0519	10.025	15.3432	20.7501
Classical	0.7757	5.2593	8.5151	14.3825	20.7104
Central Portfolios (Me)					
Robust	-0.3210	4.3651	6.3793	11.9707	18.304
Classical	-3.7725	1.0911	4.7786	11.8908	17.3109

We could look to the results by observing the trajectory after the 100 days for each baseline portfolio. We observe the 19 differences between the robust and the classical (100 days) accumulated returns for the minimum variance portfolio with the weights obtained at baselines 1870, 1880, ..., 2050. Thus, for each of the 19 time periods, are the performances of the minimum variance robust portfolios greater? The answer is yes. The boxplots in figure 3 show the differences at the end of each of the 19 trajectories for the minimum variance and the central portfolios. The maximum return portfolios behave the same compositions with weight 100% on China and are omitted. The formal t-tests reject that the mean difference is equal to zero with zero p-values for both portfolio types.

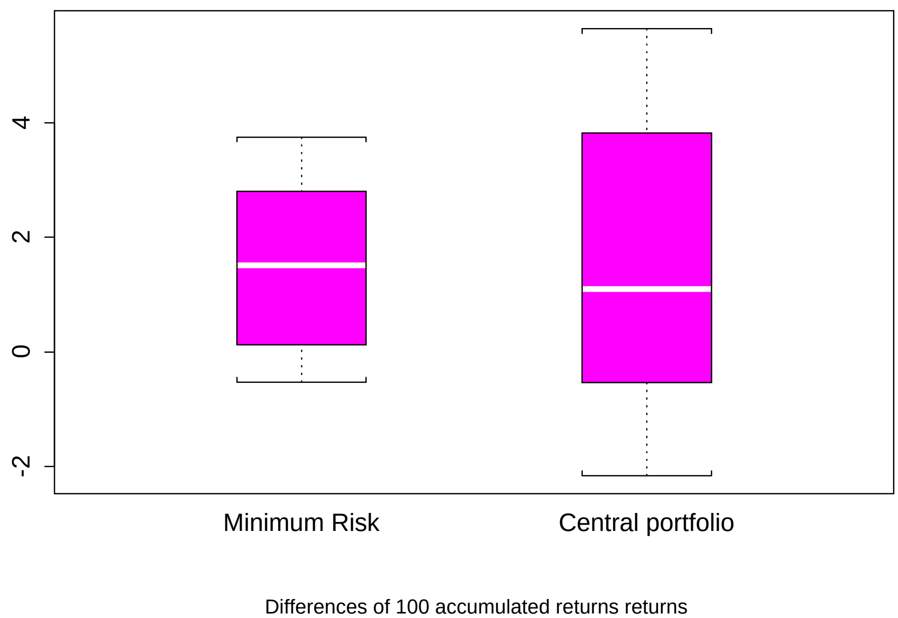

Figure 3: Differences (robust minus classical) between the accumulated returns at the end of each of the 19 100-day trajectories of the minimum variance (risk) and the central portfolios.

3.3 Weights Stability

Now we investigate the stability of the robust and classical portfolio weights. This is important because the portfolio holdings remained fixed during some time. We estimate the weights and rebalance the portfolios daily for 200 days. We form a data set of 200 weights attached to each emerging market index in our study and examine its concentration, or stability. Successive days were incorporated into the analysis one at a time. The sample started with 1951 observations and increased until it reached 2151 observations. At each of the 200 baselines we computed the robust and classical MV inputs. The idea is to observe, for a given portfolio, how do the weights change as long as new data points are incorporated in the analysis. The minimum risk, the maximum return, and the "central" portfolios are used.

Figure 4 shows the boxplot of the weights associated to the 7 components of the minimum risk portfolio under the robust (left) and classical (right) approaches for the central portfolio. The robust weights presented less variability for all 7 components. However, variables 2 (Brazil), 3 (Korea), and 6 (South Africa) were never used by the classical procedure and should not be used in comparisons. The robust portfolios seem to have more stable weights, thus reducing portfolio rebalancing costs. The weights are very stable for the minimum risk portfolios, under both robust and classical procedures.

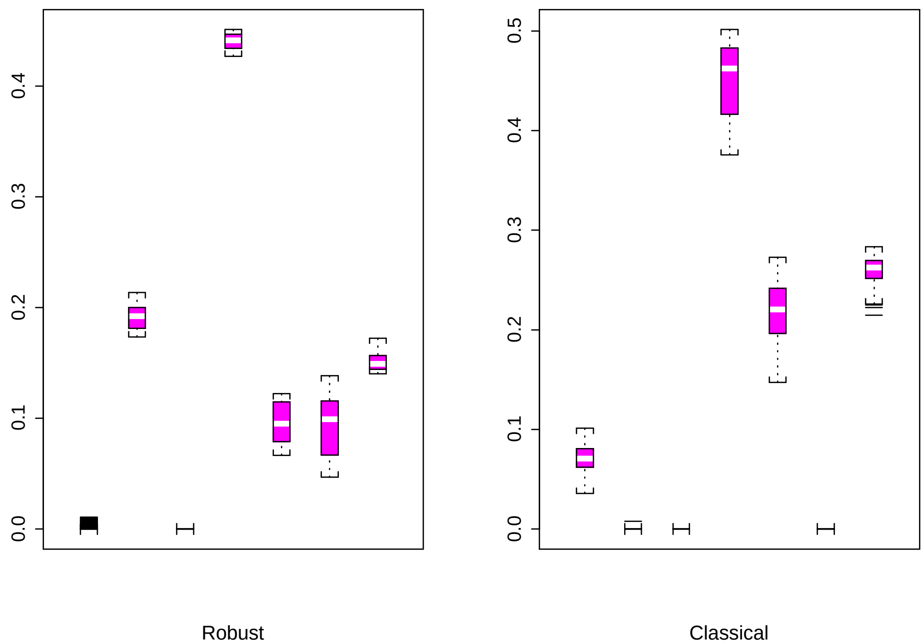

Figure 4: Boxplots of robust (left) and classical (right) weigths associated to the seven emerging market indexes for the central portfolio. Each box plot in each panel represents an index, from left to right: Argentina, Brazil, Korea, China, Taiwan, South Africa, and Mexico.

4. CONCLUSIONS

In this paper we proposed the use of robust inputs for the MV-model. The main motivations for this work were the fact that for long horizon investments with no frequent updating of portfolio weights, a robust estimate should capture the correlations observed in the vast majority of the business days; and the fact that extreme observations may show up locally or globally, and whenever they occur this may result in spurious correlations if zero breakdown point classical estimates are used.

Several aspects of the out-of-sample performance of the robust and classical portfolios were investigated. We found that robust portfolios typically yield higher accumulated returns. Also, for any given type of portfolio in the efficient frontier, the robust portfolios showed a more concentrated distribution with higher expected returns. We also concluded that the baseline choice has a stronger effect on classical portfolios than on the robust ones. In other words, due to their definition and statistical properties, the robust estimates were able to reduce the instability of the optimization process. Finally, we found that this stability property carried over to the weights associated to the robust portfolios. We strongly recommend that a robust covariance matrix is used to solve emerging stock markets allocation problems. We believe that our technique has a key advantage. Because all we change is the covariance matrix, we can use any commercially available optimizer to obtain robust portfolio weights.

5 REFERENCES

]

ANG, A.; BEKAERT, G. International asset allocation with regime shifts. **Review of Financial Studies**, v.15, n. 4, p. 1137-1187, 2002 .

EMBRECHTS, P.; KLÜPPELBERG, C.; MIKOSCH, T. **Modelling extremal events for insurance and finance.** Berlin: Springer-Verlag, 1997. 645p. (Applications of Mathematics; 33)

HARTMAN, P.; STRAETMANS, S.; DE VRIES, C. G. **Asset market linkages in crisis periods**. July 2001. (European Central Bank, n. 71).

JOHNSON, R.; WICHERN, D. **Applied multivariate statistical analysis**. Englewood: Prentice Hall, 1990. 642 p.

MARONNA, R. A. Robust m-estimators of multivariate location and scatter. **The Annals of Statistics**, v. 4, p. 51-56, 1976.

MARKOWITZ, H. M. **Portfolio selection**: efficient diversification of investments. New York: J. Willey, 1959.

ROUSSEEUW, P.J. Multivariate estimation with high breakdown point. In**:** GROSSMANN, W.; PFLUG, G.; VINCZE, I.; WERTZ, W. (Eds.) **Mathematical Statistics and Applications**, Reidel: Dordrecht, v. B, 283-297, 1985.