No. 2088

LEAST SQUARES PREDICTIONS AND MEAN-VARIANCE ANALYSIS

Enrique Sentana

FINANCIAL ECONOMICS

ISSN 0265-8003

LEAST SQUARES PREDICTIONS AND MEAN-VARIANCE ANALYSIS

Enrique Sentana

Discussion Paper No.2088
February 1999

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: http://www.cepr.org

CEPR Discussion Paper No. 2088

February 1999

ABSTRACT

Least Squares Predictions and Mean-Variance Analysis*

We compare the Sharpe ratios of investment funds which combine one riskless and one risky asset following: (i) timing strategies which forecast excess returns using simple regressions; (ii) a strategy which uses multiple regression instead; and (iii) a passive allocation which combines the funds in (i) with constant weightings. We show that (iii) dominates (i) and (ii), as it implicitly uses the linear forecasting rule that maximizes the Sharpe ratio of actively traded portfolios, but the relative ranking of (i) and (ii) is generally unclear. We also discuss under what circumstances the performances of (ii) and (iii) coincide.

JEL Classification: G11
Keywords: sharpe ratios, portfolio allocation, market timing strategies, forecasting

Enrique Sentana
CEMFI
Casado del Alisal 5
28014 Madrid
SPAIN
Tel: (34 91) 429 0551
Fax: (34 91) 429 1056
Email: sentana@cemfi.es

*The author would like to thank Manuel Arellano, Gary Chamberlain, Ian Domowitz, Lars Hansen, Jorge Padilla, Ignacio Peña, Rafael Repullo, Sushil Wadhwani and audiences at CEMFI, LSE and the III Spanish Meeting on Financial Economics for very useful comments and discussions. Of course, the usual caveat applies.

Submitted 15 January 1999

NON-TECHNICAL SUMMARY

From a formal point of view, mean-variance analysis and least squares predictions are very closely related, as both are the result of the minimization of a mean square norm over a subset set of random variables with finite variance. From a practical point of view, they are also closely connected, since many financial market practitioners combine the predictions from their regression equations with a mean-variance optimizer in order to make dynamic portfolio allocation decisions. In fact, given a set of variables, or signals, which help predict stock market returns or other financial assets, one would think *a priori* that this is a rather natural way to time the market. The purpose of this paper is to determine to what extent this intuition is correct.

We analyse the portfolio allocation between a safe asset and a risky one. To make the comparisons simpler, we assume that the unconditional expected return on the risky asset equals the safe return. Since this implies second order stochastic dominance in the absence of information (our benchmark case), risk-averse uninformed investors will only hold cash. Importantly, we assume that there are no transaction costs or other impediments to trade and, in particular, that short-sales are allowed. We also assume that the sizes of the investment funds are such that their behaviour does not alter the distribution of returns.

We then compare the performance of investment funds that follow: (i) dynamic portfolio allocations which use simple regressions to forecast excess returns; (ii) an active strategy which uses multiple regression instead; and (iii) a passive portfolio allocation which combines the funds in (i) with constant weightings implied by standard mean-variance analysis. Importantly, we make the standard assumption that the fraction of the wealth invested in the risky asset by the managers of the funds in (i) and (ii) is linear in their forecasts. One formal way of rationalizing such behaviour is through conditional mean-variance analysis, under the maintained assumption that the conditional expectations of returns are linear in (some instantaneous transformation of) the signals and the corresponding conditional variances constant.

Given that the original signals, or their transformations, may well be private information, we compute the unconditional risk-return trade-off, or Sharpe ratio, of the different portfolios to evaluate their performance taking into account their risk. This measure, which is an industry standard, is defined as the ratio of average excess return to standard deviation of a portfolio.

We show that the dynamic strategy which combines multiple regression with a mean-variance optimizer, cannot beat in terms of unconditional Sharpe ratios, the passive portfolio strategy which combines individual funds that trade on the basis of a single information variable each. We cannot, however, rank in general (i) and (ii), so that the manager who uses information on the entire set of signals may do better or worse than a manager who only uses information on a particular signal, despite the fact that expected excess returns for the former are always higher. In fact, we present a counterexample in which the manager who uses all the available information will perform in this metric strictly worse than a manager who only uses information on a particular variable.

We also show that the aforementioned passive portfolio allocation implicitly uses the linear forecasting rule that maximizes the Sharpe ratio of actively traded portfolios and discuss under which circumstances such an 'optimal' forecast coincides (up to a proportionality factor) with the conditional expectation.

Finally, we show that the Sharpe ratio of the optimal portfolio (in the unconditional mean-variance sense) depends only on the vector of Sharpe ratios of the underlying funds and their correlation matrix, through a simple quadratic form, in exactly the same way as the R2 of a multiple regression depends on the correlations implied by the simple regressions.

Our results are not totally surprising. First, from the asset pricing literature, we know that conditional mean-variance efficiency does not necessarily imply unconditional mean-variance efficiency. Second, we also know from the portfolio evaluation literature, that one-parameter performance measures such as Sharpe ratios, designed to compare passive portfolio strategies, may often yield misleading results if fund managers pursue market timing strategies.

On the other hand, there has been increasing attention recently in the time series econometrics literature on the estimation of models based on alternative prediction loss functions. In this respect, our results can be understood as saying that the quadratic loss function implicit in least squares regressions will not generally lead to estimators which maximize unconditional Sharpe ratios. At the same time, since once the signals are observed, the behaviour of the fund manager who uses multiple regression is superior to the behaviour of the fund manager who follows the passive strategy, in terms of mean-variance preferences, our results also provide a note of warning regarding the use of such estimation methods. In any case, it is worth noting that our results could be used to develop an asymptotic distribution theory that would allow us to assess in practice whether knowledge of a particular signal significantly improves Sharpe ratios.

Finally, note that the fact that the passive fund manager is the best performer raises the question of why any other fund would make efforts to find and extract the signals when they can free ride on the others. It could justify, for instance, that in order to make sure that there is an incentive to find and do research on the information, fund managers charge management fees.

1 Introduction

From a formal point of view, mean-variance analysis and least squares predictions are very closely related, as both are the result of the minimisation of a mean square norm over a closed linear subspace of the set of all random variables with finite second moments (see e.g. Hansen and Sargent (1991)). From a practical point of view, they are also closely connected, since many financial market practitioners combine the predictions from their regression equations with a mean-variance optimiser in order to make dynamic portfolio allocation decisions. In fact, given a set of variables which help predict stock market returns or other financial assets, one would think a priori that this is a rather natural way to time the market.

The purpose of this paper is to determine to what extent this intuition is correct. We analyze the portfolio allocation between a safe asset and a risky one, and derive closed-form analytical solutions. We consider alternative prediction rules, and rank them in terms of the Sharpe ratios of the associated market timing strategies. In particular, we compare the performance of investment funds that follow: i) dynamic portfolio allocations which use simple regressions to forecast excess returns; ii) an active strategy which uses multiple regression instead; and iii) a passive portfolio allocation which combines the funds in i) with constant weightings implied by standard mean-variance analysis. Furthermore, we obtain an expression for the linear forecasting rule that maximises the Sharpe ratio of an actively traded portfolio, and discuss under which circumstances such an "optimal" forecast coincides (up to a proportionality factor) with the conditional expectation.

The rest of this paper is organised as follows. We introduce the theoretical set-up in section 2, derive the active and passive portfolio strategies mentioned above, and obtain general results in terms of Sharpe ratios. Then, in section 3, we make assumptions about the joint distribution of the signals, and analyze in detail several special cases. Finally, section 4 contains a discussion of our results in relation to several areas of current research interest in the finance and econometrics literatures. Proofs of our main propositions, together with some auxiliary results, are gathered in the appendix.

2 Investment Strategies and Sharpe Ratios

Let's consider a world with a safe asset and a risky one. Let r be the excess return on the risky asset, and suppose that there are k indicator variables, or signals, $\mathbf{x} = (x_1, \ldots, x_k)'$, which help predict r.

To make the comparisons simpler, we assume that the (unconditional) expected return on the risky asset equals the safe return. Since this implies second order stochastic dominance in the absence of information (our benchmark case), risk averse uninformed investors will only hold cash.

Let's now suppose that there are k fund managers, each endowed with private information on a single indicator variable, x_j, $j = 1, \ldots, k$, who pursue active portfolio strategies. Specifically, we make the standard assumption in the literature that the fraction of their wealth invested in the risky asset is linear in their information. One formal way of rationalizing such a behaviour is through conditional mean-variance analysis, under the maintained assumption that the conditional expectations of returns are linear in (some instantaneous transformation of) the signals, and the corresponding conditional variances constant. More precisely, if we assume that the optimisation problem of a manager endowed with information I can be expressed as

$$\max_{w_r(I)} \left\{ w_r(I)E(r \mid I) - \frac{\alpha}{2} w_r^2(I)V(r \mid I) \right\} \tag{1}$$

where α is a common positive risk aversion parameter, her optimal investment strategy will be

$$w_r^*(I) = \frac{1}{\alpha}\frac{E(r|I)}{V(r|I)} \tag{2}$$

In the case of manager j, in particular, $w_r^*(x_j)$ will be proportional to the demeaned value of the j^{th} predictor variable, $\tilde{x}_j = x_j - \nu_j$, as stated.

Importantly, we assume that there are no transaction costs or other impediments to trade, and in particular, that short-sales are allowed. We also assume that the sizes of the investment funds are such that their behaviour does not alter the distribution of returns.

To keep the notation simple, define $\delta_j = \sigma_{jr}/\sigma_{jj}$ as the coefficient in the (theoretical) simple regression of r on x_j, $\varepsilon_j = r_j - \delta_j \tilde{x}_j$ as the associated prediction error, $\sigma_{\varepsilon_j\varepsilon_j} = \sigma_{rr} - \sigma_{jr}^2/\sigma_{jj}$ as its variance, and $\rho_{jr} = \sigma_{jr}/\sqrt{\sigma_{rr}\sigma_{jj}}$ as the theoretical correlation coefficient between r and x_j. Then, the excess returns on each fund will be

$$r_j = \frac{1}{\alpha} \cdot \frac{\delta_j \tilde{x}_j}{\sigma_{\varepsilon_j\varepsilon_j}} \cdot r \tag{3}$$

so that

$$E(r_j) = \frac{1}{\alpha} \cdot \frac{\delta_j \sigma_{jr}}{\sigma_{\varepsilon_j\varepsilon_j}} = \frac{1}{\alpha} \cdot \frac{\rho_{jr}^2}{1 - \rho_{jr}^2} \geq 0 \tag{4}$$

with equality if and only if the j^{th} indicator variable has no predictive power at all. Therefore, r_j is not only more profitable on average than the benchmark strategy of holding cash, but also its profitability increases with the predictive power of x_j. However, such a timing strategy is also riskier, since obviously

$$V(r_j) = \frac{1}{\alpha^2}\frac{\delta_j^2 \lambda_{jj}}{\sigma_{\varepsilon_j\varepsilon_j}^2} \geq 0 \tag{5}$$

where $\lambda_{jj} = V(\tilde{x}_j r)$.

Let's now consider another fund manager, a say, who also follows an active strategy based on the same investment rule as the first k managers, but this time knowing the whole of $\mathbf{x}$.[1] Let $\boldsymbol{\beta} = \boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr}$ be the coefficients of the (theoretical) multiple regression of returns on the indicators, $\hat{r} = E(r) + \boldsymbol{\sigma}_{xr}'\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\nu}) = \boldsymbol{\beta}'\tilde{\mathbf{x}}$ the fitted values from that regression, $u = r - \boldsymbol{\beta}'\tilde{\mathbf{x}}$ the prediction errors, $\sigma_{\hat{r}\hat{r}} = \boldsymbol{\sigma}_{xr}'\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr}$ the variance of the predicted values, $\sigma_{uu} = \sigma_{rr} - \boldsymbol{\sigma}_{xr}'\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr}$ the variance of the residuals, and finally $R^2 = (\sigma_{\hat{r}\hat{r}}/\sigma_{rr})$ the theoretical multiple correlation coefficient. Such a dynamic portfolio strategy produces excess returns of

$$r_a = \frac{1}{\alpha}\frac{\boldsymbol{\beta}'\tilde{\mathbf{x}}}{\sigma_{uu}} \cdot r \tag{6}$$

Then, since $E(\tilde{\mathbf{x}}r) = \boldsymbol{\sigma}_{xr}$

$$E(r_a) = \frac{1}{\alpha}\frac{\sigma_{\hat{r}\hat{r}}}{\sigma_{uu}} = \frac{1}{\alpha} \cdot \frac{R^2}{1 - R^2} \geq 0 \tag{7}$$

so that $E(r_j) \leq E(r_a)$ for $j = 1, \ldots, k$, as $R^2 \geq \rho_{jr}^2$. But again, such a strategy is risky, since

$$V(r_a) = \frac{1}{\alpha^2}\frac{\boldsymbol{\beta}'\boldsymbol{\Lambda}\boldsymbol{\beta}}{\sigma_{uu}^2} \geq 0 \tag{8}$$

where $\boldsymbol{\Lambda} = V(\tilde{\mathbf{x}}r)$.

Finally, let's introduce yet another manager, p say, who does not observe $\mathbf{x}$ at all, but free-rides on the other managers by constructing an "umbrella" fund of the k individual funds and the safe asset with constant weightings, according to the rules of unconditional mean-variance analysis. Let's call $\boldsymbol{\mu} = E(\mathbf{r})$ and - $= V(\mathbf{r})$, where $\mathbf{r}$ is the vector of excess returns on each fund, i.e. $\mathbf{r} = (r_1, \ldots, r_k)'$. Let $\boldsymbol{\Phi}$ be a $k \times k$ diagonal matrix with typical element $\phi_{jj} = \delta_j/\sigma_{\varepsilon_j\varepsilon_j}$, so that in matrix notation, we can write

$$\mathbf{r} = \frac{1}{\alpha}\boldsymbol{\Phi}\tilde{\mathbf{x}}r \tag{9}$$

[1]Within a conditional mean-variance framework, this assumption implicitly imposes restrictions on the joint distribution of the signals, because we would be simultaneously assuming that the distribution of returns conditional on the whole of $\mathbf{x}$ and each of its elements has a linear mean and a constant variance. In this respect, it is important to mention that all the examples discussed in section 3 satisfy such a requirement.

$$\boldsymbol{\mu} = \frac{1}{\alpha}\boldsymbol{\Phi}\boldsymbol{\sigma}_{xr} \tag{10}$$

$$- = \frac{1}{\alpha^2}\boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi} \tag{11}$$

As is well known, the optimal proportions of manager p's resources invested in each fund will be given by the vector

$$\mathbf{w}_p^* = \frac{1}{\alpha} -^{-1}\boldsymbol{\mu} = \boldsymbol{\Phi}^{-1}\boldsymbol{\Lambda}^{-1}\boldsymbol{\sigma}_{xr}$$

Hence, the excess returns from her static portfolio allocation will be

$$r_p = \mathbf{w}_p^{*\prime}\mathbf{r} = \frac{1}{\alpha}\boldsymbol{\sigma}_{xr}'\boldsymbol{\Lambda}^{-1}\tilde{\mathbf{x}}r \tag{12}$$

From here, it is straightforward to see that

$$E(r_p) = \frac{1}{\alpha}\boldsymbol{\sigma}_{xr}'\boldsymbol{\Lambda}^{-1}\boldsymbol{\sigma}_{xr} \geq 0 \tag{13}$$

and

$$V(r_p) = \frac{1}{\alpha^2}\boldsymbol{\sigma}_{xr}'\boldsymbol{\Lambda}^{-1}\boldsymbol{\sigma}_{xr} \geq 0 \tag{14}$$

Given that the original signals, or their transformations, may well be private information, in line with standard practice we shall compute the unconditional risk-return trade-off, or Sharpe ratio, of the different portfolios to evaluate their performance taking into account their risk.

Let $s(r_j)$, $s(r_a)$ and $s(r_p)$ denote the Sharpe ratios of managers j, a and p respectively. In view of (4)-(5), (7)-(8), and (13)-(14) we have that

$$s(r_j) = \frac{E(r_j)}{\sqrt{V(r_j)}} = \frac{|\sigma_{jr}|}{\sqrt{\lambda_{jj}}} \tag{15}$$

$$s(r_a) = \frac{E(r_a)}{\sqrt{V(r_a)}} = \frac{\sigma_{\hat{r}\hat{r}}}{\sqrt{\boldsymbol{\beta}'\boldsymbol{\Lambda}\boldsymbol{\beta}}} \tag{16}$$

and

$$s(r_p) = \frac{E(r_p)}{\sqrt{V(r_p)}} = \sqrt{\boldsymbol{\sigma}_{xr}'\boldsymbol{\Lambda}^{-1}\boldsymbol{\sigma}_{xr}} \tag{17}$$

In order to compare $s(r_j)$, $s(r_a)$ and $s(r_p)$, we could make further assumptions on the joint distribution of the signals. We shall explore this avenue in the next section, but before, we can state the following general result:

Proposition 1 *$s(r_p) \geq s(r_a)$ with equality if and only if $\mathbf{\Lambda}\mathbf{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr}$ is proportional to $\boldsymbol{\sigma}_{xr}$.*

Therefore, in terms of unconditional risk-return trade-offs, manager p, who pursues a passive portfolio strategy, does always at least as well as, and often better than, manager a, who pursues an active portfolio strategy.

In order to gain some intuition on this result, it is convenient to understand what the behaviour of manager a looks like from manager p's perspective. In this respect, it is important to realize that manager a is indifferent between a strategy based on the underlying risky asset alone, or one based on some of the k funds, as she can always unwind their positions. In fact, it is trivial to see that one can replicate her original active strategy with the passive strategy $\sigma_{uu}^{-1}\boldsymbol{\sigma}_{xr}'\mathbf{\Sigma}_{xx}^{-1}\mathbf{\Phi}^{-1}\mathbf{r}$. Therefore, from the point of view of p, manager a is observationally equivalent to a passive portfolio manager who is suboptimally allocating her wealth between the k funds and the safe asset. Since we know from the theory of mean-variance analysis with a riskless asset that the Sharpe ratio of the optimal portfolio will be higher than the Sharpe ratio of any other portfolio, including the original assets, it is not surprising that the Sharpe ratio of r_p, will be at least as high as the Sharpe ratio of r_a, and indeed any r_j.

In contrast, from a's vantage point, manager p is conducting a suboptimal active investment strategy, in which the fraction of her wealth invested in the risky asset is $\alpha^{-1}\boldsymbol{\sigma}_{xr}'\mathbf{\Lambda}^{-1}\tilde{\mathbf{x}}$ as opposed to $\alpha^{-1}\sigma_{uu}^{-1}\boldsymbol{\sigma}_{xr}'\mathbf{\Sigma}_{xx}^{-1}\tilde{\mathbf{x}}$. Therefore, her behaviour is observationally equivalent to that of an active portfolio manager who used $(\sigma_{uu})\boldsymbol{\sigma}_{xr}'\mathbf{\Lambda}^{-1}\tilde{\mathbf{x}}$ instead of $\boldsymbol{\sigma}_{xr}'\mathbf{\Sigma}_{xx}^{-1}\tilde{\mathbf{x}}$ as her linear prediction rule. Given that the return on any active strategy based on a linear forecast can always be replicated by some passive strategy which combines $r_1, \ldots, r_k$ and the riskless asset, in view of our previous discussion, it is not surprising that we can characterize the optimality of manager p's forecasting rule as follows:

Proposition 2 $\boldsymbol{\sigma}'_{xr}\boldsymbol{\Lambda}^{-1}\tilde{\mathbf{x}}$ *is (proportional to) the linear forecasting rule,* $\boldsymbol{\gamma}^{*\prime}\tilde{\mathbf{x}}$ *say, that maximises the ratio of excess mean return to standard deviation of an actively traded portfolio.*

Note that Proposition 2 is stronger than Proposition 1, as it says that p is not only better than a in terms of Sharpe ratios, but also better than any other trading strategy which is linear in the signals.

However, we cannot rank in general $s(r_a)$ and $s(r_j)$, so that manager a, who uses information on the entire vector $\mathbf{x}$, may do better or worse than a manager who only uses information on a particular x_j, despite the fact that expected excess returns for a are always higher. In principle, we would expect $s(r_p) \geq s(r_a) \geq s(r_j)$ for all j. Nevertheless, it is possible to construct examples in which $s(r_a) < s(r_j) < s(r_p)$ for some j (see section 3.4 below).

Our final proposition makes the relationship between $s(r_p)$ and $s(r_j)$ precise:

Proposition 3 *The Sharpe ratio of the optimal portfolio (in the unconditional mean-variance sense),* $s(r_p)$*, depends only on the vector of Sharpe ratios of the* k *underlying funds,* $s(\mathbf{r})$*, and their correlation matrix,* $\boldsymbol{\Pi}$*, through the following quadratic form:*

$$s^2(r_p) = s(\mathbf{r})'\boldsymbol{\Pi}^{-1}s(\mathbf{r})$$

The above expression, which for the case of $k = 2$ adopts the particularly simple form:

$$s^2(r_p) = \frac{1}{1-\pi_{12}^2}\left[s^2(r_1) + s^2(r_2) - 2\pi_{12}s(r_1)s(r_2)\right]$$

where $\pi_{12} = cor(r_1, r_2)$, turns out to be remarkably similar to the formula that relates the R^2 of the multiple regression of r on (a constant and) $\mathbf{x}$ with the correlations of the simple regressions. Specifically,

$$R^2 = \boldsymbol{\rho}'_{xr}\boldsymbol{\rho}_{xx}^{-1}\boldsymbol{\rho}_{xr} \tag{18}$$

The similarity is not merely coincidental. From the mathematics of the mean-variance frontier, we know that $E(r_j) = cov(r_j, r_p)E(r_p)/V(r_p)$, and therefore, that $s(r_j) = cor(r_p, r_j)s(r_p)$. In other words, the correlation coefficient between r_p and r_j is the ratio of Sharpe ratios $s(r_j)/s(r_p)$. Then, the result in Proposition 3 follows from (18) and the fact that the coefficient of determination in the multiple regression of r_p on all k $r_j's$ will be 1.

3 Examples

3.1 Two binary signals

Let's analyze in detail the case of two binary signals, whose joint distribution is given by:

$x_2 \diagdown x_1$	1	-1	
1	$\frac{1}{4}(1+\rho_{12})$	$\frac{1}{4}(1-\rho_{12})$	$\frac{1}{2}$
-1	$\frac{1}{4}(1-\rho_{12})$	$\frac{1}{4}(1+\rho_{12})$	$\frac{1}{2}$
	$\frac{1}{2}$	$\frac{1}{2}$	

(19)

so that $E(x_1) = E(x_2) = 0$, $V(x_1) = V(x_2) = 1$ and $cov(x_1, x_2) = \rho_{12}$.

It is straightforward to prove that for the four possible combinations of the signals we will have

x_1	x_2	$E(r\vert x_1)$	$E(r\vert x_2)$	$E(r\vert x_1, x_2)$
1	1	σ_{1r}	σ_{2r}	$(\sigma_{1r}+\sigma_{2r})/(1+\rho_{12})$
1	-1	σ_{1r}	$-\sigma_{2r}$	$(\sigma_{1r}-\sigma_{2r})/(1-\rho_{12})$
-1	1	$-\sigma_{1r}$	σ_{2r}	$-(\sigma_{1r}-\sigma_{2r})/(1-\rho_{12})$
-1	-1	$-\sigma_{1r}$	$-\sigma_{2r}$	$-(\sigma_{1r}+\sigma_{2r})/(1+\rho_{12})$

with variances $\sigma_{\varepsilon_1\varepsilon_1}$, $\sigma_{\varepsilon_2\varepsilon_2}$ and σ_{uu} respectively. For convenience, but without loss of generality, we can assume that σ_{1r} and σ_{2r} are both non-negative, with $\sigma_{1r} \geq \sigma_{2r}$. Otherwise, we simply redefine the signals appropriately so that positive values indicate "good news" about returns, and negative values "bad news". In this way, managers 1 and 2 will always take long positions when their respective signals are positive, and short positions when they are negative.

In view of Corollary 1 in the appendix, it is easy to see that

$$s^2(r_j) = \frac{\rho_{jr}^2}{1-\rho_{jr}^2}$$

so that not only the expected returns, but also the unconditional Sharpe ratios of managers 1 and 2 portfolios are monotonic in the predictive ability of their corresponding signals. Manager p will then combine r_1, r_2 and the safe asset in order to maximise the Sharpe ratio of her portfolio.

An interesting situation arises when $\rho_{1r}^2 = \rho_{2r}^2$. In that case, we can use Corollary 1 to prove that $\mathbf{\Lambda\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr}$ is proportional to $\boldsymbol{\sigma}_{xr}$, and furthermore, that

$$s^2(r_p) = s^2(r_a) = \frac{R^2}{1-2R^2\rho_{12}/(1+\rho_{12})}$$

Since r_1 and r_2 are not perfectly correlated as long as $\rho_{12}^2 < 1$, it is possible to form portfolios with these two funds that maintain the mean but reduce the variance, which can then be combined with the riskless asset to achieve the desired level of risk. This is precisely what managers a and p will do. In particular, manager a, and effectively manager p, will take long positions when both signals are positive, short positions when they are both negative, but no position when the signals disagree.

If $\rho_{1r}^2 \neq \rho_{2r}^2$, though, managers 1 and a will take long (short) positions in the risky asset whenever $x_1 = 1$ (-1), while manager 2 will take long (short) positions when $x_2 = 1$ (-1). Therefore, the sign of these three managers positions will be the same when the signals coincide, but they will differ when the information in the signals is conflicting. In this case, it is fairly easy to find numerical examples in which not only $s(r_p) \geq s(r_a)$, but also $s(r_p) > s(r_1) > s(r_a) > s(r_2)$ (see section 3.4 below).

3.2 Multivariate normal signals

Let's now assume that the $\mathbf{x}$'s are jointly normally distributed. Since we can prove on the basis of Corollary 2 in the appendix that $\mathbf{\Lambda\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr} \propto \boldsymbol{\sigma}_{xr}$ under normality, then we know from Proposition 1 that $s(r_a) = s(r_p)$, so that manager a's behaviour is optimal in this set up. In fact, we can prove that with Gaussian signals, the excess returns on the passive strategy will be

$$r_p = \frac{1}{\alpha}\frac{\boldsymbol{\beta}'\tilde{\mathbf{x}}}{(\sigma_{rr} + \sigma_{\hat{r}\hat{r}})} \cdot r$$

Therefore, r_p is exactly proportional to r_a (cf. (6)), with a factor of proportionality equal to $\sigma_{uu}/(\sigma_{rr} + \sigma_{\hat{r}\hat{r}}) = (1 - R^2)/(1 + R^2) \leq 1$. Several interesting results can be derived from this relationship:

a) The correlation between r_p and r_a is trivially one. Hence, although the mean and variance of r_a are higher because manager a follows an apparently riskier strategy based on her superior information, the two Sharpe ratios coincide.

b) If an indicator variable has no *additional* predictive power, so that the corresponding element of $\boldsymbol{\beta}$ is zero, the desired holdings of the relevant fund will be zero, even though the individual fund may be very profitable.

c) Manager p's behaviour is observationally equivalent to that of a portfolio manager who, in order to time the market, uses the "shrinkage" rule $\frac{1-R^2}{1+R^2} \cdot \boldsymbol{\beta}'\tilde{\mathbf{x}}$ to produce her linear predictions.

d) The Sharpe ratio of r_a, will be at least as high as the Sharpe ratio of any r_j. Therefore, fund manager a, who uses information on the entire vector $\mathbf{x}$, will do at least as well as any manager who only uses information on a particular x_j, or indeed a subset of them.

More explicitly, since in this case

$$E(r_p) = \frac{1}{\alpha}\frac{\sigma_{\hat{r}\hat{r}}}{(\sigma_{rr} + \sigma_{\hat{r}\hat{r}})} = \frac{1}{\alpha} \cdot \frac{R^2}{1 + R^2}$$

the unconditional Sharpe ratio of r_p and r_a is

$$s(r_a) = s(r_p) = \sqrt{\frac{\sigma_{\hat{r}\hat{r}}}{(\sigma_{rr} + \sigma_{\hat{r}\hat{r}})}} = \sqrt{\frac{R^2}{1+R^2}}$$

Hence, not only the expected return but also the return to risk ratio of the actively managed fund improves with the predictability of returns. Similarly, the Sharpe ratio for each fund will be

$$s(r_j) = \frac{|\sigma_{jr}|}{\sqrt{\sigma_{rr}\sigma_{jj} + \sigma_{jr}^2}} = \frac{|\rho_{jr}|}{\sqrt{1+\rho_{jr}^2}}$$

As a consequence, the Sharpe ratio of an individual fund will also be higher the more correlated x_j is with r (in absolute value), but it could never exceed the Sharpe ratio of r_a.

3.3 Independent signals

Let's assume that each signal j ($j = 1, \ldots, k$) has an arbitrary distribution with bounded fourth moment $(\kappa_j + 3)\sigma_{jj}^2$, where the $\kappa_j's$ are the coefficients of excess kurtosis, but that they are stochastically independent.

In view of Corollary 3 in the appendix, it is straightforward to prove that

$$s^2(r_j) = \frac{\rho_{jr}^2}{1 + (\kappa_j + 1)\rho_{jr}^2}$$

so that once more, the unconditional Sharpe ratios of the individual funds will be monotonic functions of the predictive power of the signals on which they trade. Similarly, we can prove that

$$s^2(r_a) = \frac{R^2}{1 + R^2 + \sum_{j=1}^{k}(\kappa_j \rho_{jr}^4 / R^2)}$$

and

$$s^2(r_p) = \frac{\sum_{j=1}^{k} \rho_{jr}^2/(1+\kappa_j\rho_{jr}^2)}{1 + \sum_{j=1}^{k} \rho_{jr}^2/(1+\kappa_j\rho_{jr}^2)}$$

Then, given Proposition 1, we will have that $s(r_p) > s(r_a)$ for independent signals, except in the unlikely event in which $\kappa_j \rho_{jr}^2$ is the same for all j. Notice, though, that this restriction is trivially satisfied when $\mathbf{x}$ is Gaussian, and also in the example of section 3.1 when $\rho_{1r}^2 = \rho_{2r}^2$, as $\kappa_1 = \kappa_2 = -2$.

3.4 A ra- e with independent binary signals

Let's finally illustrate the different issues involved by means of the following very simple game of chance. Each ticket holder in a ra- e is entitled to two random prizes. The first prize is either worthless, or something whose cash value is 2, 4 or 6 dollars, while the second is an item worth 0, 1, 3 or 4 dollars. The two prizes are independently chosen, and all combinations are equally likely. Participants can either buy as many tickets (or fractions of a ticket) as they like and collect their prizes (a long position) or sell them and pay the prizes (a short position). Also, there is unlimited borrowing and lending at a zero rate.

The cost of a ticket must be 5 dollars to ensure that it is a fair game, although no uninformed risk averse individual will participate. Suppose, though, that agent 1 possesses a valuable signal which tells her whether the second prize is "high" (i.e. 3 or 4) or "low" (i.e. 0 or 1), while agent 2 knows whether it is "high on average" (i.e. 1 or 4) or "low on average" (i.e. 0 or 3). Note that the two binary signals are independent, but the first one is more useful for predicting payoffs. In particular, the expected total payoff from the point of view of agent 1 will be \$6.50 when she receives the "high" signal, and \$3.50 when she receives the "low" one, with a standard deviation of $\$\sqrt{21}/2$ in both cases. On the other hand, the expected payoff from the point of view of agent 2 will be \$5.50 when she receives the "high on average" signal, and \$4.50 when she receives the "low on average" one, with a constant standard deviation of $\$\sqrt{29}/2$. As a result, these agents will buy each a positive amount of ra- e tickets when their signals are "good", and will sell the same amount when they are "bad".

Finally, agent a, who knows both signals, effectively knows the actual value of the second prize. Therefore, the expected value of a ticket for her will be the figure she already knows, plus the expected value of the first prize (\$3), with a constant standard deviation of $\$\sqrt{5}$. Therefore, she will buy tickets in the ra- e whenever the second prize is worth 3 or 4 dollars, and will sell them otherwise. Specifically, she will buy (sell) twice as many tickets when the value of the second prize is 4 (0) than when it is 3 (1). In this respect, it is important to emphasize that since each ticket costs \$5, she may still incur in losses whatever the value of the second prize.

In terms of excess returns per dollar invested, the conditional mean-variance frontiers as viewed by the different informed agents are depicted in Figures 1a to 1d. The slopes of these frontiers, which correspond (in absolute value) to the Sharpe ratio of a 100% investment in the risky asset conditional on the relevant information, are $2/\sqrt{5}, 1/\sqrt{5}$, $\sqrt{3/7}$ and $1/\sqrt{29}$ respectively, and the associated number of tickets bought would be proportional (again in absolute value) to $2, 1, 10/7$ and $10/29$.

If we now feed in these numbers through the appropriate expressions (assuming for the sake of concreteness that $\alpha = 1$), we obtain that agent a would make an average 50% return on her investments, while agents 1 and 2 would make 42.86% and 3.45% respectively. On the other hand, agent p, who would borrow 56% of her wealth to be able to put 98% and 58% under management by agents 1 and 2, would make a 44% return on average. The resulting situation is depicted in Figure 2 from an unconditional perspective. Note that although the actions of all agents are mean-variance efficient given their information, only agent p is efficient in the unconditional mean-variance sense. Furthermore, if we take into account the riskiness of the different strategies by means of their unconditional Sharpe ratios, it turns out that:

$$s(r_p) = .6633 > s(r_1) = .6547 > s(r_a) = .6509 > s(r_2) = .1857$$

so that the performance of agent a, who knows the value of the second prize, in fact looks worse in this metric than the performance of agent 1, who only knows whether the second prize is "high" or "low".

4 Summary and Discussion

In the context of a portfolio allocation between one riskless and one risky asset, we show that a dynamic strategy which combines multiple regression with a mean-variance optimiser, cannot beat in terms of unconditional Sharpe ratios, a passive portfolio strategy which combines individual funds that trade on the basis of a single information variable each. Furthermore, we present a counterexample in which the manager who uses all the available information will perform in this metric strictly worse than a manager who only uses information on a particular variable. We also show that the aforementioned passive portfolio allocation implicitly uses the linear forecasting rule that maximises the Sharpe ratio of actively traded portfolios. Nevertheless, we discuss under what circumstances such an "optimal" forecast coincides (up to a factor of proportionality) with the conditional expectation.

Our results are not totally surprising. First, from the asset pricing literature, we know that conditional mean-variance efficiency does not necessarily imply unconditional mean-variance efficiency (see e.g. Hansen and Richard (1987)). Second, we also know from the portfolio evaluation literature, that one-parameter performance measures such as Sharpe ratios, designed to compare passive portfolio strategies, may often yield misleading results if fund managers pursue market timing strategies (see Chen and Knez (1996), and the references therein).

On the other hand, there has been increasing attention recently in the time series econometrics literature on the estimation of models based on alternative prediction loss functions (see e.g. Weiss (1996)). In this respect, our results can be understood as saying that the quadratic loss function implicit in least squares regressions will not generally lead to estimators which maximise unconditional Sharpe ratios. At the same time, since once the signals are observed, the behaviour of fund manager a is, in terms of mean-variance preferences, superior to the behaviour of fund manager p, our results also provide a note of warning regarding the use of such estimation methods. In any case, it is worth noting that we could use our Proposition 2 to develop an asymptotic distribution theory for the sample analogues of the coefficients $\boldsymbol{\gamma}^*$ associated with the "optimal" prediction rule, which would allow us to assess in practice whether knowledge of a particular signal significantly improves Sharpe ratios.

Finally, note that the fact that fund manager p is the best performer raises the question of why any other fund would make efforts to find and extract the signals when they can free-ride on the others. This issue was originally addressed by Grossman and Stiglitz (1980), and subsequently analyzed in several other papers (see e.g. Admati and Pfleiderer (1990) and the references therein). It could justify, for instance, that in order to make sure that there is an incentive to find and do research on the information, fund managers charge management fees.

References

Admati, A.R. and Pfleiderer, P. (1990): "Direct and indirect sale of information", *Econometrica*, 58, 901-928.

Arellano, M. (1989): "On the efficient estimation of simultaneous equations with covariance restrictions", *Journal of Econometrics* 42, 247-265.

Chen, Z. and Knez, P.J. (1996): "Portfolio performance measurement: Theory and applications", *Review of Financial Studies* 9, 511-555.

Grossman, S.J. and Stiglitz, J. (1980): "On the impossibility of informationally efficient markets", *American Economic Review* 70, 393-408.

Hansen, L.P. and Richard, S.F. (1987): "The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models", *Econometrica* 55, 587-613.

Hansen, L.P. and Sargent, T.J. (1991): "Lecture notes on least squares prediction theory", in L.P. Hansen and T.J. Sargent, eds. *Rational Expectations Econometrics*, Westview, Boulder, Colorado.

Magnus, J.R. and Neudecker, H. (1988): *Matrix differential calculus with applications in Statistics and Econometrics*, Wiley, Chichester.

Weiss, A.A. (1996): "Estimating time series models using the relevant cost function", *Journal of Applied Econometrics* 11, 539-560.

Appendix

Auxiliary results

Proof of Proposition 1

By the Cauchy-Schwartz inequality,

$$\sigma_{\hat{r}\hat{r}} = (\boldsymbol{\sigma}'_{xr}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr})^2 \leq (\boldsymbol{\sigma}'_{xr}\boldsymbol{\Lambda}^{-1}\boldsymbol{\sigma}_{xr})(\boldsymbol{\sigma}'_{xr}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\Lambda}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr}) = (\boldsymbol{\sigma}'_{xr}\boldsymbol{\Lambda}^{-1}\boldsymbol{\sigma}_{xr})(\boldsymbol{\beta}'\boldsymbol{\Lambda}\boldsymbol{\beta})$$

so

$$s^2(r_p) \geq s^2(r_a)$$

and $s(r_p) \geq s(r_a)$ given that they are both positive. Equality is achieved in the above inequality if and only if $\boldsymbol{\Lambda}^{1/2}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr} = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{\sigma}_{xr}\theta$, where θ is a non-zero scalar, or equivalently, if and only if $\boldsymbol{\Lambda}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr} = \theta\boldsymbol{\sigma}_{xr}$, as stated. □

It is in fact possible to fully characterize the matrices $\boldsymbol{\Lambda}$ for which $s(r_p) = s(r_a)$. To do so, it is convenient to re-write the necessary and sufficient condition as

$$\boldsymbol{\Sigma}_{xx}^{-1/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}_{xx}^{-1/2}\frac{\boldsymbol{\Sigma}_{xx}^{-1/2}\boldsymbol{\sigma}_{xr}}{\sqrt{\boldsymbol{\sigma}'_{xr}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr}}} = \theta\frac{\boldsymbol{\Sigma}_{xx}^{-1/2}\boldsymbol{\sigma}_{xr}}{\sqrt{\boldsymbol{\sigma}'_{xr}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr}}}$$

so that $(\boldsymbol{\sigma}'_{xr}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr})^{-1/2}\boldsymbol{\Sigma}_{xx}^{-1/2}\boldsymbol{\sigma}_{xr}$ can be regarded as a normalized eigenvector of the matrix $\boldsymbol{\Sigma}_{xx}^{-1/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}_{xx}^{-1/2}$. Since the spectral decomposition of this matrix will be given by

$$\theta\frac{\boldsymbol{\Sigma}_{xx}^{-1/2}\boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}\boldsymbol{\Sigma}_{xx}^{-1/2}}{\boldsymbol{\sigma}'_{xr}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr}} + \left(\mathbf{I} - \frac{\boldsymbol{\Sigma}_{xx}^{-1/2}\boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}\boldsymbol{\Sigma}_{xx}^{-1/2}}{\boldsymbol{\sigma}'_{xr}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr}}\right)\mathbf{U}\boldsymbol{\Theta}\mathbf{U}'\left(\mathbf{I} - \frac{\boldsymbol{\Sigma}_{xx}^{-1/2}\boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}\boldsymbol{\Sigma}_{xx}^{-1/2}}{\boldsymbol{\sigma}'_{xr}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr}}\right)$$

where $\boldsymbol{\Theta}$ is a diagonal positive semi-definite matrix of order $N-1$, and $\mathbf{U}$ is any $N \times (N-1)$ orthogonal matrix such that $\mathbf{I} - \boldsymbol{\Sigma}_{xx}^{-1/2}\boldsymbol{\sigma}_{xr}(\boldsymbol{\sigma}'_{xr}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr})^{-1}\boldsymbol{\sigma}'_{xr}\boldsymbol{\Sigma}_{xx}^{-1/2} = \mathbf{U}\mathbf{U}'$, we finally have that all admissible $\boldsymbol{\Lambda}$ could be written as

$$\boldsymbol{\Lambda} = \frac{\theta\boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}}{(\boldsymbol{\sigma}'_{xr}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr})} + \left(\boldsymbol{\Sigma}_{xx} - \frac{\boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}}{(\boldsymbol{\sigma}'_{xr}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr})}\right)\mathbf{Q}\left(\boldsymbol{\Sigma}_{xx} - \frac{\boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}}{(\boldsymbol{\sigma}'_{xr}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr})}\right)$$

where $\mathbf{Q}$ is an arbitrary symmetric positive semidefinite matrix of dimension N and rank $N-1$. In this respect, note that if we choose $\mathbf{Q}$ proportional to the Moore-Penrose inverse of $\boldsymbol{\Sigma}_{xx} - (\boldsymbol{\sigma}'_{xr}\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr})^{-1}\boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}$, then $\boldsymbol{\Lambda}$ will be a linear combination of $\boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}$ and $\boldsymbol{\Sigma}_{xx}$, as in Corollary 2.

Proof of Proposition 2

Formally, we can characterize the linear forecasting rule that maximises the Sharpe ratio, $\boldsymbol{\gamma}^{*\prime}\tilde{\mathbf{x}}$, as

$$\boldsymbol{\gamma}^* = \arg\max_{\gamma} \frac{\boldsymbol{\gamma}'\boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}\boldsymbol{\gamma}}{\boldsymbol{\gamma}'\boldsymbol{\Lambda}\boldsymbol{\gamma}} = \boldsymbol{\Lambda}^{-1/2}\arg\max_{\tilde{\gamma}} \frac{\tilde{\boldsymbol{\gamma}}'\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}\boldsymbol{\Lambda}^{-1/2}\tilde{\boldsymbol{\gamma}}}{\tilde{\boldsymbol{\gamma}}'\tilde{\boldsymbol{\gamma}}}$$

where $\tilde{\boldsymbol{\gamma}} = \boldsymbol{\Lambda}^{1/2}\boldsymbol{\gamma}$. The solution to this well-known programme is simply the eigenvector associated with the maximum eigenvalue of the rank 1 matrix $\boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}$ in the metric of $\boldsymbol{\Lambda}$. That is,

$$\max_{\tilde{\gamma}} \frac{\tilde{\boldsymbol{\gamma}}'\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}\boldsymbol{\Lambda}^{-1/2}\tilde{\boldsymbol{\gamma}}}{\tilde{\boldsymbol{\gamma}}'\tilde{\boldsymbol{\gamma}}} = \lambda_1(\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}\boldsymbol{\Lambda}^{-1/2}) = \boldsymbol{\sigma}'_{xr}\boldsymbol{\Lambda}^{-1}\boldsymbol{\sigma}_{xr} = s^2(r_p)$$

where $\lambda_1(\mathbf{A})$ denotes the largest eigenvalue of the matrix $\mathbf{A}$. Therefore, since $\tilde{\boldsymbol{\gamma}}^* = \boldsymbol{\Lambda}^{-1/2}\boldsymbol{\sigma}_{xr}/\sqrt{\boldsymbol{\sigma}'_{xr}\boldsymbol{\Lambda}^{-1}\boldsymbol{\sigma}_{xr}}$, then $\boldsymbol{\gamma}^* = \boldsymbol{\Lambda}^{-1}\boldsymbol{\sigma}_{xr}/\sqrt{\boldsymbol{\sigma}'_{xr}\boldsymbol{\Lambda}^{-1}\boldsymbol{\sigma}_{xr}}$ as required. □

Proof of Proposition 3

We have already seen that $r_p = \mathbf{w}_p^{*\prime}\mathbf{r}$, $E(r_p) = \frac{1}{\alpha}\boldsymbol{\mu}'\text{-}^{-1}\boldsymbol{\mu}$, and $V(\mathbf{r}_p) = \frac{1}{\alpha^2}\boldsymbol{\mu}'\text{-}^{-1}\boldsymbol{\mu}$, with $\mathbf{w}_p^* = \frac{1}{\alpha}\text{-}^{-1}\boldsymbol{\mu}$, $\boldsymbol{\mu} = E(\mathbf{r})$ and $\text{-} = V(\mathbf{r})$. Therefore, $s^2(r_p) = \boldsymbol{\mu}'\text{-}^{-1}\boldsymbol{\mu} = \boldsymbol{\mu}'dg^{-1/2}(\text{-})dg^{1/2}(\text{-})\text{-}^{-1}dg^{1/2}(\text{-})dg^{-1/2}(\text{-})\boldsymbol{\mu} = s'(\mathbf{r})\boldsymbol{\Pi}^{-1}s(\mathbf{r})$ as required.
□

Lemma 1

Let $\boldsymbol{\Upsilon}_4$ be the $k(k+1)/2 \times k(k+1)/2$ matrix which contains the fourth order cumulants of the signals. Then

$$\boldsymbol{\Lambda} = \sigma_{rr}\boldsymbol{\Sigma}_{xx} + \boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr} + \boldsymbol{\Psi}$$

where

$$vech(\boldsymbol{\Psi}) = \boldsymbol{\Upsilon}_4 \mathbf{D}'\mathbf{D} vech(\boldsymbol{\Sigma}_{xx}^{-1} \boldsymbol{\sigma}_{xr} \boldsymbol{\sigma}'_{xr} \boldsymbol{\Sigma}_{xx}^{-1})$$

and $\mathbf{D}$ is the duplication matrix.

Proof. Since

$$\boldsymbol{\Lambda} = V(\tilde{\mathbf{x}}r) = E(r^2 \tilde{\mathbf{x}}\tilde{\mathbf{x}}') - E(\tilde{\mathbf{x}}r)E(\tilde{\mathbf{x}}'r)$$

and

$$E(r^2 \tilde{\mathbf{x}}\tilde{\mathbf{x}}') = E\left[E(r^2|\tilde{\mathbf{x}})\tilde{\mathbf{x}}\tilde{\mathbf{x}}'\right] = E\left\{\left[E^2(r|\tilde{\mathbf{x}}) + V(r|\tilde{\mathbf{x}})\right]\tilde{\mathbf{x}}\tilde{\mathbf{x}}'\right\}$$

by the law of iterated expectations, we can write

$$\boldsymbol{\Lambda} = E[(\boldsymbol{\beta}'\tilde{\mathbf{x}})^2 \tilde{\mathbf{x}}\tilde{\mathbf{x}}'] + \sigma_{uu}\boldsymbol{\Sigma}_{xx} - \boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr} = E[(\boldsymbol{\sigma}'_{xr}\boldsymbol{\Sigma}_{xx}^{-1}\tilde{\mathbf{x}}\tilde{\mathbf{x}}'\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}'_{xr})\tilde{\mathbf{x}}\tilde{\mathbf{x}}'] + \sigma_{uu}\boldsymbol{\Sigma}_{xx} - \boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}$$

Vectorising the first term of this expression we get

$$\begin{aligned} vec\{E[(\boldsymbol{\sigma}'_{xr}\boldsymbol{\Sigma}_{xx}^{-1}\tilde{\mathbf{x}}\tilde{\mathbf{x}}'\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}'_{xr})\tilde{\mathbf{x}}\tilde{\mathbf{x}}']\} &= E[\tilde{\mathbf{x}}\tilde{\mathbf{x}}' \otimes \tilde{\mathbf{x}}\tilde{\mathbf{x}}'] vec(\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}\boldsymbol{\Sigma}_{xx}^{-1}) \\ &= E[(\tilde{\mathbf{x}} \otimes \tilde{\mathbf{x}})(\tilde{\mathbf{x}}' \otimes \tilde{\mathbf{x}}')] vec(\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}\boldsymbol{\Sigma}_{xx}^{-1}) \\ &= E[vec(\tilde{\mathbf{x}}\tilde{\mathbf{x}}')vec'(\tilde{\mathbf{x}}\tilde{\mathbf{x}}')] vec(\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}\boldsymbol{\Sigma}_{xx}^{-1}) \end{aligned}$$

Given that for any symmetric matrix $\mathbf{A}$, $vec(\mathbf{A}) = \mathbf{D}vech(\mathbf{A})$, $vech(\mathbf{A}) = \mathbf{D}^+ vec(\mathbf{A})$, and $\mathbf{D}^+\mathbf{D} = \mathbf{I}$, where $\mathbf{D}^+ = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$ is the Moore-Penrose inverse of $\mathbf{D}$, we have

$$vech\{E[(\boldsymbol{\sigma}'_{xr}\boldsymbol{\Sigma}_{xx}^{-1}\tilde{\mathbf{x}}\tilde{\mathbf{x}}'\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}'_{xr})\tilde{\mathbf{x}}\tilde{\mathbf{x}}']\} = \boldsymbol{\Delta}_4 \mathbf{D}'\mathbf{D} vech(\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}\boldsymbol{\Sigma}_{xx}^{-1})$$

where $\boldsymbol{\Delta}_4 = E[vech(\tilde{\mathbf{x}}\tilde{\mathbf{x}}')vech'(\tilde{\mathbf{x}}\tilde{\mathbf{x}}')]$. But since

$$\boldsymbol{\Delta}_4 = vech(\boldsymbol{\Sigma}_{xx})vech'(\boldsymbol{\Sigma}_{xx}) + 2\mathbf{D}^+(\boldsymbol{\Sigma}_{xx} \otimes \boldsymbol{\Sigma}_{xx})\mathbf{D}^{+\prime} + \boldsymbol{\Upsilon}_4$$

(see e.g. Arellano (1989)), and

$$\begin{aligned} & \mathbf{D}\left[vech(\boldsymbol{\Sigma}_{xx})vech'(\boldsymbol{\Sigma}_{xx}) + 2\mathbf{D}^+(\boldsymbol{\Sigma}_{xx} \otimes \boldsymbol{\Sigma}_{xx})\mathbf{D}^{+\prime}\right]\mathbf{D}'\mathbf{D} vech(\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}\boldsymbol{\Sigma}_{xx}^{-1}) \\ = & \left[vec(\boldsymbol{\Sigma}_{xx})vec'(\boldsymbol{\Sigma}_{xx}) + 2\mathbf{D}\mathbf{D}^+(\boldsymbol{\Sigma}_{xx} \otimes \boldsymbol{\Sigma}_{xx})\mathbf{D}^{+\prime}\mathbf{D}'\right] vec(\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}\boldsymbol{\Sigma}_{xx}^{-1}) \end{aligned}$$

$$= \quad [vec(\mathbf{\Sigma}_{xx})vec'(\mathbf{\Sigma}_{xx}) + 2(\mathbf{\Sigma}_{xx} \nwarrow\!\!\searrow \mathbf{\Sigma}_{xx})]\, vec(\mathbf{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}\mathbf{\Sigma}_{xx}^{-1})$$
$$= \quad vec(\mathbf{\Sigma}_{xx})tr(\mathbf{\Sigma}_{xx}\mathbf{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}\mathbf{\Sigma}_{xx}^{-1}) + 2vec(\mathbf{\Sigma}_{xx}\mathbf{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}\mathbf{\Sigma}_{xx}^{-1}\mathbf{\Sigma}_{xx})$$
$$= \quad (\boldsymbol{\sigma}'_{xr}\mathbf{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr})vec(\mathbf{\Sigma}_{xx}) + 2vec(\boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr})$$

because $\mathbf{D}\mathbf{D}^{+}(\mathbf{\Sigma}_{xx}\nwarrow\!\!\searrow\mathbf{\Sigma}_{xx}) = (\mathbf{\Sigma}_{xx}\nwarrow\!\!\searrow\mathbf{\Sigma}_{xx})\mathbf{D}^{+\prime}\mathbf{D}'$, $\mathbf{D}^{+\prime}\mathbf{D}'\mathbf{D}^{+\prime}\mathbf{D}' = \mathbf{D}^{+\prime}\mathbf{D}' = \frac{1}{2}(\mathbf{I}+\mathbf{K})$, where $\mathbf{K}$ is the commutation matrix (see e.g. Magnus and Neudecker (1988)), and $\frac{1}{2}(\mathbf{I}+\mathbf{K})vec(\mathbf{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}\mathbf{\Sigma}_{xx}^{-1}) = vec(\mathbf{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}\mathbf{\Sigma}_{xx}^{-1})$ by symmetry, then the result follows. □

Corollary 1

If x_1 and x_2 are two binary signals whose joint distribution is specified by (19), then

$$\mathbf{\Psi} = -2\sigma_{rr}\begin{pmatrix} \rho_{1r}^2 & \rho_{12}R^2 \\ \rho_{12}R^2 & \rho_{2r}^2 \end{pmatrix}$$

Proof. Given the joint distribution of the signals, it is easy to prove that

$$\mathbf{\Delta}_4 = \begin{pmatrix} 1 & \rho_{12} & 1 \\ \rho_{12} & 1 & \rho_{12} \\ 1 & \rho_{12} & 1 \end{pmatrix}$$

and

$$\mathbf{\Upsilon}_4 = -2\begin{pmatrix} 1 & \rho_{12} & \rho_{12}^2 \\ \rho_{12} & \rho_{12}^2 & \rho_{12} \\ \rho_{12}^2 & \rho_{12} & 1 \end{pmatrix}$$

Then the expression for $\mathbf{\Psi}$ follows on the basis of Lemma 1, as

$$\mathbf{\Upsilon}_4\mathbf{D}'\mathbf{D}vech(\mathbf{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}\mathbf{\Sigma}_{xx}^{-1}) = -\,2\begin{pmatrix} \sigma_{1r}^2 \\ \rho_{12}(\sigma_{1r}^2 + \sigma_{2r}^2 - 2\rho_{12}\sigma_{1r}\sigma_{2r})/(1-\rho_{12}^2) \\ \sigma_{2r}^2 \end{pmatrix}$$

and $\sigma_{\hat{r}\hat{r}} = (\sigma_{1r}^2 + \sigma_{2r}^2 - 2\rho_{12}\sigma_{1r}\sigma_{2r})/(1-\rho_{12}^2)$.

Corollary 2

If $\mathbf{x}$ is multivariate normal, then

$$\boldsymbol{\Lambda} = \sigma_{rr}\boldsymbol{\Sigma}_{xx} + \boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}$$

and

$$\boldsymbol{\Lambda}^{-1} = \frac{1}{\sigma_{rr}}\boldsymbol{\Sigma}_{xx}^{-1} - \frac{1}{\sigma_{rr}^2(1+R^2)}\boldsymbol{\beta}\boldsymbol{\beta}'$$

Proof. The first part is a trivial consequence of Lemma 1, since $\boldsymbol{\Upsilon}_4 = \mathbf{0}$ under normality. The expression for $\boldsymbol{\Lambda}^{-1}$ follows directly from the Woodbury formula.□

When the *joint* distribution of returns and signals is normal, the expression for $\boldsymbol{\Lambda}$ in Corollary 2 can be immediately derived on the basis of well known results on fourth moments of the multivariate normal distribution (see e.g. Arellano (1989)). However, note that the assumptions of Corollary 2 are weaker, since they only require a *marginal* Gaussian distribution for $\mathbf{x}$, and a *conditional* distribution for r given $\mathbf{x}$ with a linear mean and a constant variance.

Corollary 3

If the signals are independently distributed, then

$$\boldsymbol{\Psi} = diag\left[\kappa_1\sigma_{1r}^2, \ldots, \kappa_k\sigma_{1k}^2\right]$$

where the $\kappa_j's$ are the coefficients of excess kurtosis.

Proof. To keep the algebra simple, we shall only consider the case of $k = 2$. An analogous proof for arbitrary k is straightforward.

Since in this case:

$$\boldsymbol{\Delta}_4 = \begin{pmatrix} (\kappa_1+3)\sigma_{11}^2 & 0 & \sigma_{11}\sigma_{22} \\ 0 & \sigma_{11}\sigma_{22} & 0 \\ \sigma_{11}\sigma_{22} & 0 & (\kappa_2+3)\sigma_{22}^2 \end{pmatrix}$$

and

$$\boldsymbol{\Upsilon}_4 = \boldsymbol{\Delta}_4 - vech(\boldsymbol{\Sigma}_{xx})vech'(\boldsymbol{\Sigma}_{xx}) - 2\mathbf{D}^{+}(\boldsymbol{\Sigma}_{xx} \nwarrow \boldsymbol{\Sigma}_{xx})\mathbf{D}^{+\prime} = \begin{pmatrix} \kappa_1\sigma_{11}^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \kappa_2\sigma_{22}^2 \end{pmatrix}$$

then, the result follows on the basis of Lemma 1, as

$$\boldsymbol{\Upsilon}_4\mathbf{D}'\mathbf{D}vech(\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}_{xr}'\boldsymbol{\Sigma}_{xx}^{-1}) = \begin{pmatrix} \kappa_1\sigma_{1r}^2 \\ 0 \\ \kappa_2\sigma_{2r}^2 \end{pmatrix}$$

□

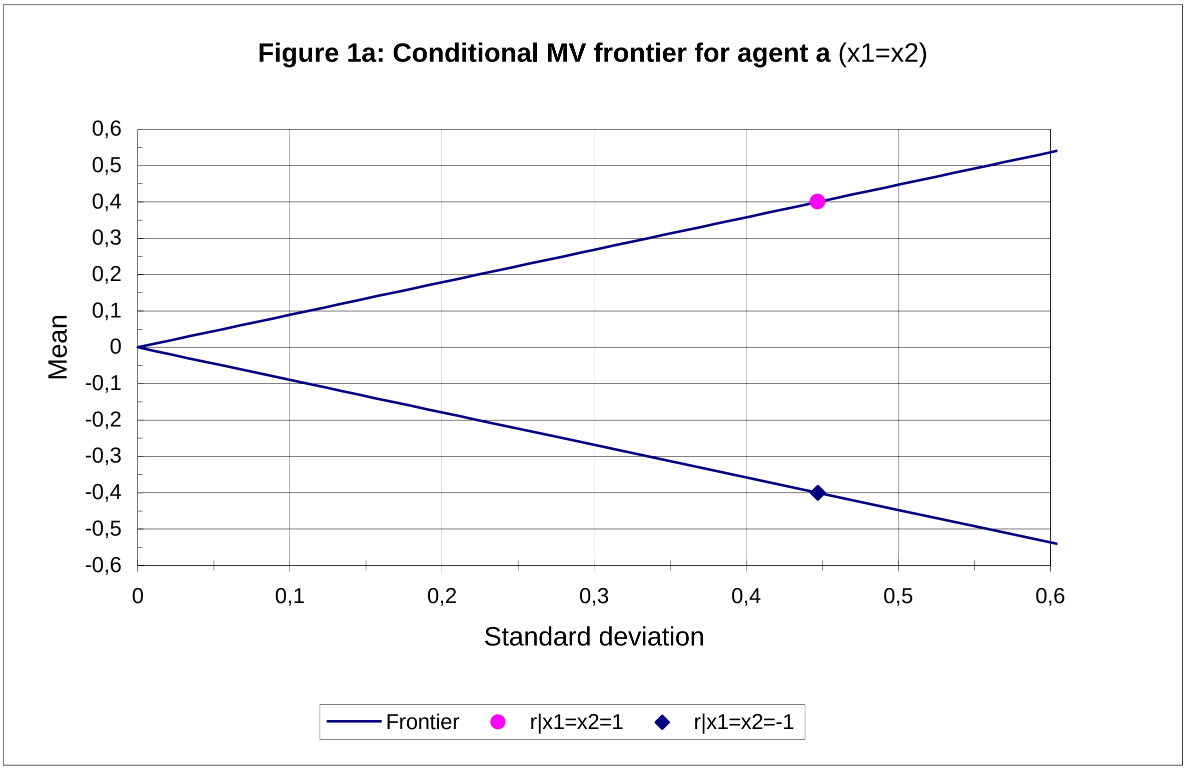
Figure 1a: Conditional MV frontier for agent a (x1=x2)
Mean
0,6
0,5
0,4
0,3
0,2
0,1
0
-0,1
-0,2
-0,3
-0,4
-0,5
-0,6
0
0,1
0,2
0,3
0,4
0,5
0,6
Standard deviation
Frontier
r|x1=x2=1
r|x1=x2=-1

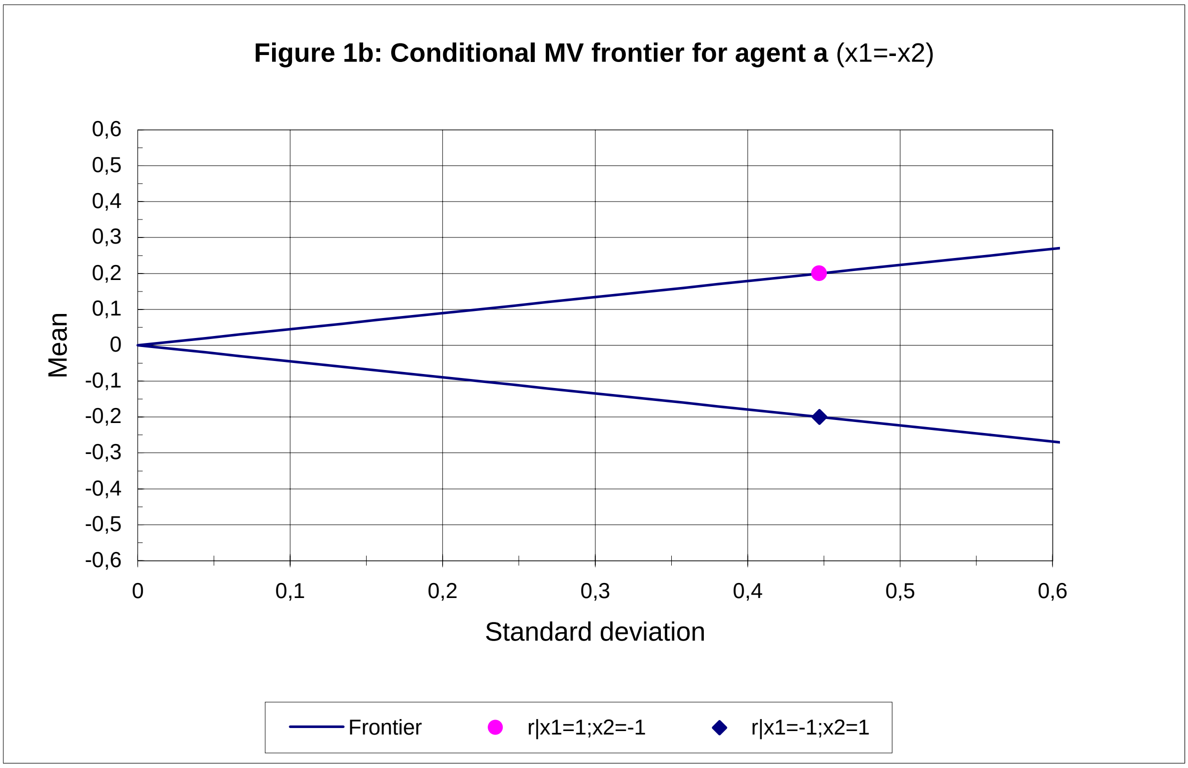
Figure 1b: Conditional MV frontier for agent a (x1=-x2)
Mean
0,6
0,5
0,4
0,3
0,2
0,1
0
-0,1
-0,2
-0,3
-0,4
-0,5
-0,6
0
0,1
0,2
0,3
0,4
0,5
0,6
Standard deviation
Frontier
r|x1=1;x2=-1
r|x1=-1;x2=1

Figure 1c: Conditional MV frontier for agent 1

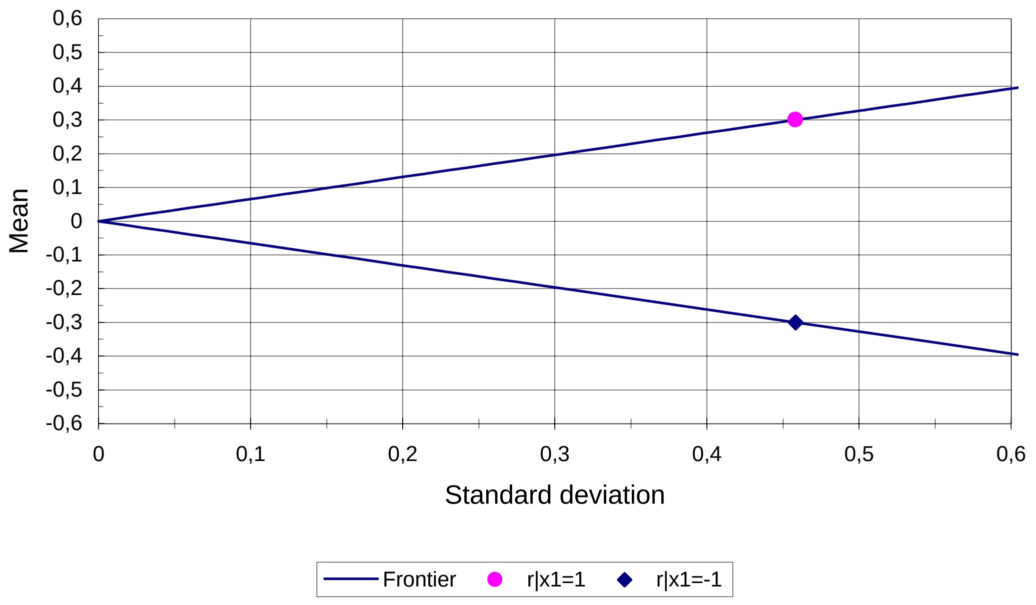

Figure 1d: Conditional MV frontier for agent 2

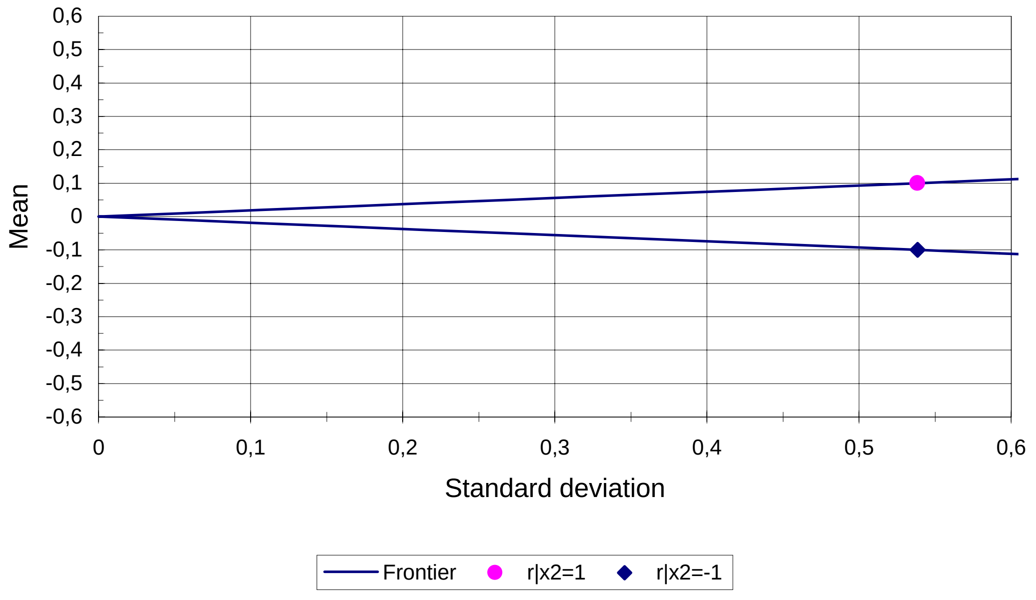

Figure 2: Unconditional MV frontier for agent p