Least Squares Predictions and Mean-Variance Analysis*

Enrique Sentana

(CEMFI, CEPR and LSE FMG)

Revised: October 2004

*This is a substantially revised version of Sentana (1999). I would like to thank Manuel Arellano, Gary Chamberlain, Ian Domowitz, Miguel Ferreira, Lars Hansen, Ravi Jagannathan, Javier Mencía, Jorge Padilla, Ignacio Peña, Francisco Peñaranda, Eric Renault, Rafael Repullo, Sushil Wadhwani, and audiences at the University of Alicante, CEMAF/ISCTE, CEMFI, CIRANO, Kellogg, LSE, Málaga, Minneapolis Fed, the III Spanish Meeting on Financial Economics (Bilbao, 1998), as well as the CFS Workshop on "New directions in financial risk management" (Frankfurt, 2003) for very useful comments and discussions. Two anonymous referees have also helped me to improve the paper. Of course, the usual caveat applies. Address for correspondence: Casado del Alisal 5, 28014 Madrid, Spain (tel: +34 91 429 0551, fax: +34 91 429 1056, e-mail: sentana@cemfi.es).

Abstract

We compare the Sharpe ratios of traders who combine one riskless and one risky asset following: buy and hold strategies (i); timing strategies with forecasts from simple (ii) or multiple (iii) regressions; and passive allocations of (i) and (ii) with mean-variance optimisers (iv). We show that (iv) implicitly uses the linear forecasting rule that maximises the Sharpe ratio of managed portfolios, but the remaining rankings are unclear. We also suggest GMM estimators to make (iv) operational, and evaluate their significance with spanning tests. Finally, we characterise the equivalence between (iii) and (iv), and propose moment tests to assess it.

Key words: Delegated Portofolio Management, Financial Forecasting, Portfolio Performance Evaluation, Sharpe ratios, Spanning Tests.

JEL: G11, C53

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 returns on the stock market 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. To do so, we study an economy with a safe asset and a risky one, and rank alternative predictions rules that can be used to dynamically determine the scale of managed portfolios in terms of the Sharpe ratios of the associated market timing strategies.

We choose the Sharpe ratio, which is defined as the ratio of the expected excess return of an investment to its standard deviation, because it is the most common measure used by financial market practitioners to rank fund managers and to evaluate the attractiveness of investment strategies in general. This is particularly true in the hedge fund industry, where portfolio composition is not usually reported, and hence performance evaluation must be based almost exclusively on observed track records. The ubiquity of the Sharpe ratio is obviously justified. Apart from its simplicity, and the fact that it is a rather natural risk-adjusted measure of performance, it has also the convenient property of being numerically invariant to the degree of leverage of the position. At the same time, though, the Sharpe ratio is not without its limitations, as the vast academic literature on performance evaluation has repeatedly made clear. Nevertheless, despite its problems, the Sharpe ratio remains a central concept for both researchers and practitioners.

Therefore, rather than stressing its shortcomings or suggesting alternative measures, we take an analogous view to Goetzmann al. (2002), and explain how to maximise the Sharpe ratio when investors may take positions in managed portfolios in the Hansen and Richard (1987) sense. Given that by construction such managed portfolios are linear functions of the predictor variables, we are able to provide a linear forecasting rule interpretation to the optimal (in the unconditional mean-variance sense) portfolio. Specifically, we obtain a closed-form analytical expression for the linear forecasting rule that maximises the unconditional Sharpe ratio of an actively traded portfolio, and discuss under which circumstances such an "optimal" forecast coincides (up to a proportionality factor) with the multiple least squares projection. We then confirm that this linear maximal Sharpe ratio forecasting rule is implicitly used by a passive portfolio manager who resorts to static mean-variance analysis to combine the returns of several individual investment funds, each of which follows dynamic portfolio allocations that use simple regressions to forecast excess returns.

This homeomorphism allows us to use mean-variance spanning tests to empirically assess whether or not a set of potential explanatory variable increases the Sharpe ratio of such market timing strategies. In addition, we also develop tests for the optimality of the least squares prediction rule in our context. In both cases, we use Hansen's (1982) generalised method of moments (GMM) framework to make our inferences robust to potential heteroskedasticity and serial correlation in the joint stochastic process that generates returns and predictors. As an illustration, we revisit the empirical application in Pesaran and Timmermann (1995), who used least squares regressions to generate predictions of excess returns on the Standard and Poor's 500 (SP500) index using only ex ante dated variables.

The rest of this paper is organised as follows. We introduce the theoretical set-up in section 1, derive the active and passive portfolio strategies mentioned above, and obtain general results in terms of Sharpe ratios. Then, in section 2, we make assumptions about the joint distribution of predictors and returns, and analyse in detail two special cases. Our proposed estimation and testing procedures are explained in section 3, while in section 4 we present the results of our empirical application. Finally, section 5 contains a discussion of our results in relation to several areas of current research interest in the finance and econometrics literatures, and suggests some extensions. Proofs of our propositions, together with some auxiliary results, are gathered in the appendix.

1 Investment Strategies and Sharpe Ratios

Let us consider a world with a safe asset and a risky one, whose gross returns (i.e. total pay-offs per unit invested) are R and R_0, respectively. Let $r = R - R_0$ denote the excess return on the risky asset, and suppose that there are k indicator variables, sometimes called instruments or signals, $\mathbf{x} = (x_1, \ldots, x_k)'$, which help predict r. Importantly, we assume that there are no transaction costs or other impediments to trade, and in particular, that short-sales are possible. In this way, we open the gate for managed portfolios, whose excess returns will be exactly proportional to r, with a factor of proportionality that depends on some or all of the signals (see Hansen and Richard (1987)). However, to keep the discussion as simple as possible, we explicitly exclude options and any other derivative assets whose pay-offs are non-linear functions of R (see Goetzmann et al. (2002) for the effects on including such non-linear pay-offs in an unconditional mean-variance set-up). Finally, we assume that the sizes of the investment funds are such that their behaviour does not alter the distribution of excess returns on the risky asset.

Let μ_r denote the expected excess return of an investment fund that effectively follows a simple buy and hold strategy. More properly, μ_r coincides with the expected pay-offs of an arbitrage (i.e. self-financing) position that buys one single unit of the risky asset by borrowing in the money market. If σ_r denotes the standard deviation of r, then its Sharpe ratio is defined as,

$$s(r) = \mu_r / \sigma_r.$$

Without loss of generality, we shall assume in what follows that $\mu_r \geq 0$. Otherwise, a fund manager could take negative positions in the risky asset and invest the proceedings in the safe asset, creating in this way an arbitrage portfolio whose pay-offs would be the opposite of the original pay-offs.

Let us now suppose that there is a fund manager, a say, with private information on $\mathbf{x}$, who pursues an active portfolio strategy. Specifically, we make the standard assumption in the literature that the fraction of the funds under her management invested in the risky asset is proportional to her forecast, which she generates by means of the linear least squares (LLS) regression of r on (some instantaneous transformation of) the indicators.[1]

Let $\boldsymbol{\beta}_x = \boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr}$ denote the slope coefficients in the (theoretical) multiple regression of excess returns on $\mathbf{x}$, and let $\beta_0 = \mu_r - \boldsymbol{\beta}_x'\boldsymbol{\mu}_x$ the associated intercept, where $\boldsymbol{\mu}_x = E(\mathbf{x})$, $\boldsymbol{\sigma}_{xr} = cov(\mathbf{x}, r)$ and $\boldsymbol{\Sigma}_{xx} = V(\mathbf{x})$, which we assume is positive definite. If we then define

$$\begin{aligned}
\mathbf{w} &= [1, (\mathbf{x} - \boldsymbol{\mu}_x)']', \\
\boldsymbol{\gamma}_{wr} &= E(\mathbf{w}r) = (\mu_r, \boldsymbol{\sigma}_{xr}')', \\
\boldsymbol{\Gamma}_{ww} &= E(\mathbf{w}\mathbf{w}') = \begin{pmatrix} 1 & 0 \\ 0 & \boldsymbol{\Sigma}_{xx} \end{pmatrix},
\end{aligned}$$

and

$$\boldsymbol{\beta}_w = (\mu_r, \boldsymbol{\beta}_x')',$$

we can write $\boldsymbol{\beta}_w = \boldsymbol{\Gamma}_{ww}^{-1}\boldsymbol{\gamma}_{wr}$. Similarly, let

$$f = \mu_r + \boldsymbol{\sigma}_{xr}'\boldsymbol{\Sigma}_{xx}^{-1}(\mathbf{x} - \boldsymbol{\mu}_x) = \boldsymbol{\beta}_w'\mathbf{w}$$

denote the predicted values from that regression,

$$u = r - \boldsymbol{\beta}_w'\mathbf{w}$$

the prediction errors,

$$\sigma_f^2 = \boldsymbol{\sigma}_{xr}'\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr}$$

the variance of the predicted values,

$$\sigma_u^2 = \sigma_r^2 - \boldsymbol{\sigma}_{xr}'\boldsymbol{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr}$$

the variance of the residuals, and finally

$$R^2 = \sigma_f^2/\sigma_r^2$$

the theoretical multiple correlation coefficient. With this notation, manager a's dynamic portfolio strategy produces excess returns, r_a say, which are proportional to

$$f \cdot r = \boldsymbol{\beta}'_w \mathbf{w} \cdot r = \boldsymbol{\beta}'_w \mathbf{z},$$

where $\mathbf{z} = \mathbf{w} \cdot r$ is such that $E(\mathbf{z}) = \boldsymbol{\mu}_z = \boldsymbol{\gamma}_{wr}$. But since

$$E(f \cdot r) = \boldsymbol{\beta}'_w \boldsymbol{\gamma}_{wr} = \boldsymbol{\gamma}'_{wr} \boldsymbol{\Gamma}^{-1}_{ww} \boldsymbol{\gamma}'_{wr} = \mu_r^2 + \boldsymbol{\sigma}'_{xr} \boldsymbol{\Sigma}^{-1}_{xx} \boldsymbol{\sigma}_{xr} \geq 0,$$

and

$$V(f \cdot r) = \boldsymbol{\beta}'_w \boldsymbol{\Sigma}_{zz} \boldsymbol{\beta}_w,$$

where $V(\mathbf{z}) = \boldsymbol{\Sigma}_{zz}$, the unconditional Sharpe ratio of such an active strategy will be given by the following expression:

$$s(r_a) = \frac{E(r_a)}{\sqrt{V(r_a)}} = \frac{\boldsymbol{\gamma}'_{wr} \boldsymbol{\Gamma}^{-1}_{ww} \boldsymbol{\gamma}_{wr}}{\sqrt{\boldsymbol{\beta}'_w \boldsymbol{\Sigma}_{zz} \boldsymbol{\beta}_w}} \geq 0.$$

In order to determine under which circumstances the LLS prediction rule that agent a uses maximises the Sharpe of the associated market timing strategy, let us first fully characterise the optimal linear prediction rule:

Proposition 1 $\boldsymbol{\gamma}'_{wr} \boldsymbol{\Sigma}^{-1}_{zz} \mathbf{w}$ *is (proportional to) the linear forecasting rule that maximises the ratio of excess mean return to standard deviation of an actively traded portfolio.*

In what follows, we shall refer to $\boldsymbol{\gamma}'_{wr} \boldsymbol{\Sigma}^{-1}_{zz} \mathbf{w}$ as the linear maximal unconditional Sharpe ratio (LMUSR) rule.

Let us call

$$r_p = \boldsymbol{\gamma}'_{wr} \boldsymbol{\Sigma}^{-1}_{zz} \mathbf{w} \cdot r = \boldsymbol{\gamma}'_{wr} \boldsymbol{\Sigma}^{-1}_{zz} \mathbf{z}$$

the excess returns of the associated market timing strategy. By definition, it must be the case that

$$s(r_p) = \frac{E(r_p)}{\sqrt{V(r_p)}} = \frac{\boldsymbol{\gamma}'_{wr} \boldsymbol{\Sigma}^{-1}_{zz} \boldsymbol{\gamma}_{wr}}{\sqrt{\boldsymbol{\gamma}'_{wr} \boldsymbol{\Sigma}^{-1}_{zz} \boldsymbol{\gamma}_{wr}}} = \sqrt{\boldsymbol{\gamma}'_{wr} \boldsymbol{\Sigma}^{-1}_{zz} \boldsymbol{\gamma}_{wr}} \geq s(r_a).$$

The following result specifies a necessary and sufficient condition for the LLS prediction rule to be optimal with respect to this alternative loss function:

Proposition 2 *$s(r_p) \geq s(r_a)$ with equality if and only if $\mathbf{\Gamma}_{ww}^{-1}\boldsymbol{\gamma}_{wr}$ is proportional to $\mathbf{\Sigma}_{zz}^{-1}\boldsymbol{\gamma}_{wr}$.*

In order to gain some intuition on why r_a is generally suboptimal from the point of view of unconditional Sharpe ratios, it is convenient to introduce another $k+1$ "active" fund managers, each of whom has information on a single component of $\mathbf{w}$ only. Obviously, the first manager, manager 0 say, simply buys and holds the risky asset, but the remaining k managers follow truly active market timing strategies with excess returns proportional to $(x_j - \mu_j) \cdot r = w_j \cdot r$ $(j = 1, \ldots, k)$. Hence, the vector $\mathbf{z} = \mathbf{w} \cdot r$ is (proportional to) the $k+1$ vector of excess returns on those funds.

Although the k truly active funds are redundant assets from the point of view of any agent who observes the signals because she can always unwind her positions, it is easy to see that an active strategy based on any linear prediction rule can be replicated by some passive strategy which combines $\mathbf{z}$ and the riskless asset. Specifically, $r_p = \boldsymbol{\varphi}_w^{+\prime}\mathbf{z}$, where $\boldsymbol{\varphi}_w^{+} = \mathbf{\Sigma}_{zz}^{-1}\boldsymbol{\gamma}_{wr} = \mathbf{\Sigma}_{zz}^{-1}\boldsymbol{\mu}_z$. Hence, r_p coincides with the excess returns of a passive fund manager, p say, who forms a portfolio of the $k+1$ individual funds and the safe asset with constant weightings optimally chosen according to the rules of unconditional mean-variance analysis. Since we know from the theory of mean-variance analysis with arbitrage portfolios that the mean-variance frontier will be spanned by the optimal portfolio alone, and that its Sharpe ratio will be the highest Sharpe ratio attainable, it is not surprising that the Sharpe ratio of r_p will be at least as high as the Sharpe ratios of r_a, r, and indeed any z_j.

In this respect, note that from the point of view of someone who does not observe the signals, manager a is observationally equivalent to a passive portfolio manager who is suboptimally allocating her wealth between the $k+1$ funds and the safe asset. In contrast, from point of view of someone who observes the signals, manager p is suboptimally forecasting excess returns because she is actively managing her portfolio on the basis of a linear prediction rule which is proportional to $\boldsymbol{\gamma}'_{wr}\boldsymbol{\Sigma}^{-1}_{zz}\mathbf{w}$ instead of the LLS predictions $\boldsymbol{\gamma}'_{wr}\boldsymbol{\Gamma}^{-1}_{ww}\mathbf{w}$. Nevertheless, given that the evaluation criterion of manager p is the unconditional Sharpe ratio of her fund performance, as opposed to the root mean square error of her predictions, or the correlation between those predictions and r, then she is outperforming everyone else, including manager a.

The relationship between $s(r_p)$, $s(r)$ and $s(z_j)$, where

$$s(z_j) = \frac{E(z_j)}{\sqrt{V(z_j)}} = \frac{|\sigma_{x_j r}|}{\sigma_{z_j}},$$

can also be made precise:

Proposition 3 *The Sharpe ratio of the optimal portfolio (in the unconditional mean-variance sense), $s(r_p)$, only depends on the vector of Sharpe ratios of the $k+1$ underlying funds, $s(\mathbf{z})$, and their correlation matrix, $\boldsymbol{\rho}_{zz}$, through the following quadratic form:*

$$s^2(r_p) = s(\mathbf{z})'\boldsymbol{\rho}^{-1}_{zz}s(\mathbf{z}).$$

The above expression, which for the case of $k = 1$ adopts the particularly simple form:

$$s^2(r_p) = \frac{1}{1-\rho^2_{z_1 r}}\left[s^2(r) + s^2(z_1) - 2\rho_{z_1 r}s(r)s(z_1)\right],$$

where $\rho_{z_1 r} = cor(z_1, r)$, 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,

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

The similarity is not merely coincidental. From the mathematics of the mean-variance frontier, we know that $E(z_j) = cov(z_j, r_p)E(r_p)/V(r_p)$, and therefore, that $s(z_j) = cor(z_j, r_p)s(r_p)$. In other words, the correlation coefficient between z_j and r_p is $s(z_j)/s(r_p)$, i.e. the ratio of their Sharpe ratios. Hence, the result in

Proposition 3 follows from (1) and the fact that the coefficient of determination in the multiple regression of r_p on $\mathbf{z}$ will be 1 because r_p is a linear combination of this vector.

In principle, one would expect that $s(r_a) \geq s(r)$ with equality if and only if $\boldsymbol{\sigma}_{xr} = 0$, since in the absence of parameter uncertainty, it would appear that superior information should always lead to superior performance. As we shall see in section 2.1 below, however, it turns out that this is not always the case. And although one would also expect $s(r_a) \geq s(z_j)$ for $j = 1, \ldots, k$, it is not possible to rank $s(r_a)$ and $s(z_j)$ either (see section 3.4 of Sentana (1999) for a counterexample). Therefore, 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, or no information at all.

2 Special cases

It is important to note that the results in the previous section are valid regardless of the joint distribution of signals and returns, as long as the required moments are bounded. In this respect, it is easy to see that the only higher moments involved are

$$\begin{aligned} &E[(r-\mu_r)(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{x}-\boldsymbol{\mu}_x)'], \\ &E[(r-\mu_r)^2(\mathbf{x}-\boldsymbol{\mu}_x)], \\ &E[(r-\mu_r)^2(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{x}-\boldsymbol{\mu}_x)']. \end{aligned} \tag{2}$$

For illustrative purposes, however, it is convenient to specify such moments. We shall do so in two different ways. In our first example, we shall assume that the distribution of excess returns given the signals has a linear conditional mean and a constant conditional variance. As we shall see, those two assumptions, together with assumptions about the third and fourth central moments of the marginal distribution of $\mathbf{x}$, fully specify (2) irrespective of the remaining characteristics of the conditional distribution of r given $\mathbf{x}$. In our second example, in contrast, we shall fully specify the joint distribution of r and $\mathbf{x}$.

2.1 Linear conditional means, constant conditional variances, and multivariate normal signals

Let us assume that

$$E(r|\mathbf{x}) = \mu_r + \boldsymbol{\beta}'_x(\mathbf{x} - \boldsymbol{\mu}_x) = \boldsymbol{\beta}'_w\mathbf{w},$$

and

$$V(r|\mathbf{x}) = \sigma_u^2,$$

so that the conditional mean coincides with the LLS prediction rule, and the conditional variance with the residual variance of the regression.

Further, let us also assume that the $\mathbf{x}$'s are jointly normally distributed.[2]

However, we shall deliberately make no assumptions about the remaining characteristics of the conditional distribution of r given $\mathbf{x}$, which therefore could be both skewed and leptokurtic.

Then, we can state the following result:

Proposition 4 *If* $E(r|\mathbf{x}) = \mu_r + \boldsymbol{\beta}'_x(\mathbf{x} - \boldsymbol{\mu}_x) = \boldsymbol{\beta}'_w\mathbf{w}$, $V(r|\mathbf{x}) = \sigma_u^2$ *and* $\mathbf{x} \sim N(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_{xx})$, *then*

1.

$$r_p = r \cdot \frac{\mu_r + [1 - s^2(r)][1 + s^2(r)]^{-1}\boldsymbol{\beta}'_x(\mathbf{x} - \boldsymbol{\mu}_x)}{\sigma_r^2\{1 + [1 - s^2(r)][1 + s^2(r)]^{-1}R^2\}}$$

2.

$$\begin{aligned} s(r_p) &= \sqrt{\frac{s^2(r) + [1 - s^2(r)][1 + s^2(r)]^{-1}R^2}{1 + [1 - s^2(r)][1 + s^2(r)]^{-1}R^2}} \\ &\geq \frac{s^2(r) + R^2}{\sqrt{s^2(r)(1 + 3R^2) + (1 + R^2)R^2}} = s(r_a) \end{aligned}$$

with equality if and only if either $R^2 = 0$ *or* $s(r) = 0$.

Several interesting results can be derived from this relationship:

a) The positions on $z_1, \ldots, z_k$ taken by managers a and p are proportional to the regression coefficients $\boldsymbol{\beta}_x$. As a result, if an indicator variable has no *additional* predictive power, so that the corresponding element of $\boldsymbol{\beta}_x$ is zero, the desired holdings of the relevant fund will be zero, even though the individual fund may be very profitable. However, the actual positions taken by the fund managers a and p could have different signs, depending on whether $s(r)$ exceeds 1 or not.

b) The correlation between the LLS prediction $\boldsymbol{\gamma}'_{wr}\boldsymbol{\Gamma}^{-1}_{ww}\mathbf{w}$ and the LMUSR forecast $\boldsymbol{\gamma}'_{wr}\boldsymbol{\Sigma}^{-1}_{zz}\mathbf{w}$ will be 1 if $s(r) < 1$, 0 if $s(r) = 1$, and -1 if $s(r) > 1$. However, the correlation between r_p and r_a will generally be less than 1, since the LLS and LMSUR rules are not usually proportional.

c) The Sharpe ratio of r_a is not necessarily higher than the Sharpe ratio of any z_j, including $s(r)$. In fact, fund manager a, who uses information on the entire vector $\mathbf{w}$, could be beaten in this metric by a simple buy and hold strategy (e.g. when $s^2(r) = 1$ and $0 < R^2 < 1$).

The equality between $s(r_p)$ and $s(r_a)$ when $R^2 = 0$ is rather obvious, because it corresponds to a situation in which the returns on the risky asset are stochastically independent from all the signals.

The special case of $s(r) = 0$, though, is far more interesting, as we can sharpen some of the previous results. In particular, we can show that the correlation between r_p and r_a will be exactly 1. As a result, the performance of both manager a and manager p, as measured by their common unconditional Sharpe ratio, is at least as good, and generally better, than the performance of any other fund manager.

2.2 Jointly log-normal returns and signals

Let us alternatively assume the $(k + 1)$-dimensional random vector $(\ln R, \ln x_1, \ldots, \ln x_k)$ is jointly normally distributed, with mean vector ν and covariance matrix Δ. Therefore, both the conditional mean and conditional variance of R given the signals will be non-linear functions of $\mathbf{x}$.[3]

In principle, it is straightforward to use the moment generating function of a multivariate normal vector to obtain all the moments of $r = R - R_0$ and $\mathbf{x}$ that appear in (2). In order to keep the algebra as simple as possible, though, we shall only consider the case of a single signal, a dynamic example of which would be a Gaussian AR(1) process for geometric (i.e. continuously compounded) returns.

In this context, we can prove the following result:

Proposition 5 *If* $\ln R$ *and* $\ln x_1$ *are jointly normally distributed, with means* ν_R *and* ν_1*, variances* δ_R^2 *and* δ_1^2*, and correlation coefficient* π_{R1}*, so that*

$$\begin{aligned}
\mu_R &= E(R) = \exp(\nu_R + .5\delta_R^2), \\
\sigma_R^2 &= V(R) = \mu_R^2\lambda_R^2, \\
\lambda_R &= \sigma_R/\mu_R = \sqrt{\exp(\delta_R^2) - 1} \\
s(r) &= \frac{1 - (R_0/\mu_R)}{\lambda_R}, \\
\mu_1 &= E(x_1) = \exp(\nu_1 + .5\delta_1^2), \\
\sigma_1^2 &= V(x_1) = \mu_1^2\lambda_1^2, \\
\lambda_1 &= \sigma_1/\mu_1 = \sqrt{\exp(\delta_1^2) - 1} \\
\rho_{r1} &= E\left[\left(\frac{R-\mu_R}{\sigma_R}\right)\left(\frac{x_1-\mu_1}{\sigma_1}\right)\right] = \frac{\exp(\pi_{R1}\delta_R\delta_1) - 1}{\lambda_1\lambda_R}, \\
\phi_{r1} &= E\left[\left(\frac{R-\mu_R}{\sigma_R}\right)^2\left(\frac{x_1-\mu_1}{\sigma_1}\right)\right] = \rho_{r1}[2\lambda_R + \rho_{r1}\lambda_1(1+\lambda_R^2)], \\
\phi_{1r} &= E\left[\left(\frac{R-\mu_R}{\sigma_R}\right)\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right] = \rho_{r1}[2\lambda_1 + \rho_{r1}\lambda_R(1+\lambda_1^2)], \\
\kappa_{r1} &= E\left[\left(\frac{R-\mu_R}{\sigma_R}\right)^2\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right] = 1 + 4\lambda_R\lambda_1\rho_{r1} \\
&\quad +2[1 + 2(\lambda_1^2 + \lambda_R^2) + 3\lambda_1^2\lambda_R^2]\rho_{r1}^2 + 4\lambda_R\lambda_1(1+\lambda_1^3)(1+\lambda_R^3)]\rho_{r1}^3 \\
&\quad +\lambda_1^2\lambda_R^2(1+\lambda_R^2)(1+\lambda_1^2)\rho_{r1}^4,
\end{aligned}$$

then

1.

$$
\begin{aligned}
r_p \;=\; & \frac{r}{\sigma_r\left[(1-\rho_{r1}^2)s^2(r)+2(\phi_{1r}-\rho_{r1}\phi_{r1})s(r)+(\kappa_{r1}-\phi_{r1}^2-\rho_{r1}^2)\right]} \\
& \times\left\{\left[s^3(r)+2\phi_{1r}s^2(r)+(\kappa_{r1}-2\rho_{r1}^2)s(r)-\rho_{r1}\phi_{r1}\right]\right. \\
& \left.+\left[-\rho_{r1}s^2(r)-\phi_{r1}s(r)+\rho_{r1}\right]\left(\frac{x_1-\mu_1}{\sigma_1}\right)\right\}.
\end{aligned}
$$

2.

$$
\begin{aligned}
s(r_p) &= \sqrt{\frac{s^4(r)+2\phi_{1r}s^3(r)+(\kappa_{r1}-3\rho_{r1}^2)s^2(r)-2\rho_{r1}\phi_{r1}s(r)+\rho_{r1}^2}{(1-\rho_{r1}^2)s^2(r)+2(\phi_{1r}-\rho_{r1}\phi_{r1})s(r)+(\kappa_{r1}-\phi_{r1}^2-\rho_{r1}^2)}} \\
&\geq \frac{s^2(r)+\rho_{r1}^2}{\sqrt{(1+3\rho_{r1}^2)s^2(r)+2\rho_{r1}(\phi_{1r}+\rho_{r1}\phi_{r1})s(r)+\rho_{r1}^2(\kappa_{r1}-\rho_{r1}^2)}} = s(r_a)
\end{aligned}
$$

with equality if and only if either $\pi_{R1}=0$ *or* $s(r)$ *solves the cubic equation*

$$
2\rho_{r1}s^3(r)+(2\rho_{r1}\phi_{1r}+\phi_{r1}^2)s^2(r)+\rho_{r1}(\kappa_{r1}-2\rho_{r1}^2-1)s(r)-\rho_{r1}^2\phi_{r1}^2=0.
$$

The equality between $s(r_p)$ and $s(r_a)$ when $\pi_{r1}=0$ is once again rather obvious, because it corresponds to a situation in which the returns on the risky asset are stochastically independent from the signal.

Unfortunately, this time the values of $s(r)$ that make the LLS and LMUSR rules equivalent are not so easy to interpret, although $s(r)=0$ is clearly excluded. Nevertheless, it is still worth looking at the limiting "risk neutral" situation of $\ln R_0=\nu_R+.5\delta_R^2$, in which the expected return on the risky asset equals the safe return. With this extra assumption, we will have that

$$
\begin{aligned}
r_a &\propto r\cdot\rho_{r1}\left(\frac{x_1-\mu_1}{\sigma_1}\right), \\
r_p &\propto r\cdot\left[-\rho_{r1}\phi_{r1}+\rho_{r1}\left(\frac{x_1-\mu_1}{\sigma_1}\right)\right],
\end{aligned}
$$

so that

$$
s(r_p)=\frac{|\rho_{r1}|}{\sqrt{\kappa_{r1}-\rho_{r1}^2-\phi_{r1}^2}}>\frac{|\rho_{r1}|}{\sqrt{\kappa_{r1}-\rho_{r1}^2}}=s(r_a)
$$

because $\phi_{r1}\neq 0$ when $\rho_{r1}\neq 0$. Intuitively, the co-skewness in the joint distribution of returns and signals implies that the excess returns on the risky asset r provide a hedge for the excess returns on the managed portfolio $z_1=r(x_1-\mu_1)$, which induces fund manager p to hold r in her unconditionally mean-variance efficient frontier portfolio even though its risk premia is 0.

3 Inference procedures

3.1 Parameter estimation

It is important to stress that the analytical expressions presented in the previous sections have been obtained under the assumption that the relevant moments of the joint distribution of r and $\mathbf{x}$ are known. However, it is clear that this is an unrealistic assumption. In practice, of course, those moments will have to be estimated, and therefore, the empirical analogues to the coefficients of the LLS and LMUSR rules will be subject to sampling variability.

As is well known, inferences about the theoretical least squares coefficients $\boldsymbol{\beta}_w = \boldsymbol{\Gamma}_{ww}^{-1}\boldsymbol{\gamma}_{wr}$ can be obtained by applying GMM to the following moment conditions:

$$E[\mathbf{w}_t(r_t - \mathbf{w}_t'\boldsymbol{\beta}_w)] = \mathbf{E}[\mathbf{h}_O(r_t, \mathbf{w}_t; \boldsymbol{\beta}_w)] = \mathbf{0}, \tag{3}$$

which correspond to the usual orthogonality conditions implicit in the normal equations, as the subscript O reminds us. Since these moment conditions exactly identify $\boldsymbol{\beta}_w$, its unrestricted GMM estimators trivially coincide with the OLS regression coefficients $\hat{\boldsymbol{\beta}}_w = \hat{\boldsymbol{\Gamma}}_{ww}^{-1}\hat{\boldsymbol{\gamma}}_{wr}$, where the elements of $\hat{\boldsymbol{\gamma}}_{wr}$ and $\hat{\boldsymbol{\Gamma}}_{ww}$ are the sample analogues of $\boldsymbol{\gamma}_{wr}$ and $\boldsymbol{\Gamma}_{ww}$, respectively.[4] Then, we can use a standard heteroskedasticity and autocorrelation consistent (HAC) estimator of the asymptotic covariance matrix of $\sqrt{T}\bar{\mathbf{h}}_{OT}(\boldsymbol{\beta}_w)$ to make robust inferences about $\boldsymbol{\beta}_w$, where $\bar{\mathbf{h}}_{OT}(\boldsymbol{\beta}_w)$ is the average value of $\mathbf{h}_O(r_t, \mathbf{w}_t; \boldsymbol{\beta}_w)$ in a sample of size T.

Exactly the same approach can be used to make inferences about the coefficients of the LMUSR rule described in Proposition 1. Specifically, given that $\boldsymbol{\varphi}_w^+ = \boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\mu}_z$, we can write

$$E\begin{bmatrix} \mathbf{z}_t - \boldsymbol{\mu}_z \\ (\mathbf{z}_t - \boldsymbol{\mu}_z)(\mathbf{z}_t - \boldsymbol{\mu}_z)'\boldsymbol{\varphi}_w^+ - \mathbf{z}_t \end{bmatrix} = E\begin{bmatrix} \mathbf{h}_M(\mathbf{z}_t; \boldsymbol{\mu}_z) \\ \mathbf{h}_C(\mathbf{z}_t; \boldsymbol{\varphi}_w^+, \boldsymbol{\mu}_z) \end{bmatrix} = E[\mathbf{h}_E(\mathbf{z}_t; \boldsymbol{\varphi}_w^+, \boldsymbol{\mu}_z)] = \mathbf{0},$$

where the first set of moment conditions, $E[\mathbf{h}_M(\mathbf{z}_t; \boldsymbol{\mu}_z)] = \mathbf{0}$, simply defines $\boldsymbol{\mu}_z$ as the mean (M) of $\mathbf{z}$, while the second set of moment conditions, $E[\mathbf{h}_C(\mathbf{z}_t; \boldsymbol{\varphi}_w^+, \boldsymbol{\mu}_z)] = \mathbf{0}$, defines the optimal weights $\boldsymbol{\varphi}_w^+$ in terms of centred (C) second moments.

In this context, parameter estimation is once more trivial because both $\boldsymbol{\varphi}_w^+$ and $\boldsymbol{\mu}_z$ are exactly identified. Specifically, we will end up with

$$\begin{aligned}\hat{\boldsymbol{\mu}}_z &= \hat{\boldsymbol{\gamma}}_{wr},\\ \hat{\boldsymbol{\varphi}}_w^+ &= \hat{\boldsymbol{\Sigma}}_{zz}^{-1}\hat{\boldsymbol{\gamma}}_{wr}.\end{aligned}$$

Similarly, it is also straightforward to make robust inferences about $\boldsymbol{\varphi}_w^+$ by using a HAC estimator of the covariance matrix of $\sqrt{T}\bar{\mathbf{h}}_{ET}(\boldsymbol{\varphi}_w^+, \boldsymbol{\mu}_z)$.

Interestingly, the moment conditions $E[\mathbf{h}_C(\mathbf{z}_t; \boldsymbol{\varphi}_w^+, \boldsymbol{\mu}_z)] = \mathbf{0}$ suggest an alternative interpretation of the optimal forecasting coefficients $\boldsymbol{\varphi}_w^+$. In particular, $\boldsymbol{\varphi}_w^{+\prime}\mathbf{z}$ corresponds in our context to the so-called *centred* mean representing portfolio introduced by Chamberlain and Rothschild (1983), z^{++} say, which is the unique portfolio of the arbitrage assets $\mathbf{z}$ such that $E(z) = cov(z, z^{++})$ for any z in the linear span of $\mathbf{z}$, $\langle \mathbf{z} \rangle$. In the next section, we shall exploit this interpretation to assess the contribution of a set of signals to the optimal prediction rule.

3.2 Testing the incremental value of some signals

From a practitioner's point of view, we may be interested in analysing the effects of adding new signals to an existing set. Specifically, let $\mathbf{x}_1$ denote our original k_1 signals, and $\mathbf{x}_2$ some k_2 additional ones, so that the expanded set of signals $\mathbf{x} = (\mathbf{x}_1', \mathbf{x}_2')'$ is of dimension $k = k_1 + k_2$, where $k_1 \geq 0$ and $k_2 \geq 1$. The question that we want to answer is whether the elements of $\mathbf{x}_2$ attract non-zero coefficients in the LMUSR rule. Two particularly important cases that arise in practice will be those in which $k_2 = 1$, when we want to assess the contribution of a single variable, and also $k_2 = k$, in which case we want to determine whether using the whole vector of signals $\mathbf{x}$ improves at all the risk-return trade-off of a simple buy and hold strategy.

Let $\mathbf{z} = (\mathbf{z}_1', \mathbf{z}_2')' = \mathbf{w} \cdot r$, where $\mathbf{w} = (\mathbf{w}_1', \mathbf{w}_2')'$, $\mathbf{w}_1 = [1, (\mathbf{x}_1 - \boldsymbol{\mu}_1)']'$ and $\mathbf{w}_2 = (\mathbf{x}_2 - \boldsymbol{\mu}_2)$, and partition $\boldsymbol{\varphi}_w^+$ conformably as $(\boldsymbol{\varphi}_{w_1}^{+\prime}, \boldsymbol{\varphi}_{w_2}^{+\prime})'$. In this context, the null hypothesis of interest is given by the following k_2 homogeneous parametric restrictions $H_0 : \boldsymbol{\varphi}_{w_2}^+ = \mathbf{0}$. Therefore, we can test the incremental value of $\mathbf{x}_2$ by using the classical trinity of GMM asymptotic tests: Wald, Lagrange Multiplier and Distance Metric tests (see e.g. Newey and McFadden (1994)). In addition, since $\boldsymbol{\varphi}_w^+$ is exactly identified under the alternative, the Distance Metric test coincides with the usual Overidentifying Restriction (or J) test in this case. As is well known, when the null hypothesis is true, all these tests will be asymptotically distributed as the same χ^2 random variable with k_2 degrees of freedom under standard regularity conditions under fairly weak assumptions on the distribution of $\mathbf{z}_t$, which depends on the joint distribution excess returns and signals.

Given the interpretation of $\boldsymbol{\varphi}_w^+$ as the weights of the centred mean representing portfolio of $\langle \mathbf{z} \rangle$, a test of $\boldsymbol{\varphi}_{w_2}^+ = \mathbf{0}$ is effectively a test of the equality of z^{++} and z_1^{++}, which is the centred mean representing portfolio corresponding to $\langle \mathbf{z}_1 \rangle$. Therefore, following Peñaranda and Sentana (2004), we can also interpret H_0 : $\boldsymbol{\varphi}_{w_2}^+ = \mathbf{0}$ as a mean-variance spanning restriction. The intuition is that since all the frontier arbitrage portfolios that can be generated from $\mathbf{z}$ are proportional to z^{++}, by testing that $z^{++} = z_1^{++}$, we will be testing the null hypothesis that the unconditional mean-variance frontier generated by $\mathbf{z}_1$ remains unaltered when we add $\mathbf{z}_2$ to the universe of available assets.

Such a homeomorphism allows us to apply to our problem some of the alternative mean-variance spanning tests that have been suggested in the literature.

For instance, let $z^+ = \boldsymbol{\phi}_w^{+\prime}\mathbf{z}$, where

$$\boldsymbol{\phi}_w^+ = E^{-1}(\mathbf{z}\mathbf{z}')E(\mathbf{z}) = [V(\mathbf{z})+E(\mathbf{z})E(\mathbf{z}')]^{-1}E(\mathbf{z}),$$

denote the unique *uncentred* mean representing portfolio of $\langle \mathbf{z} \rangle$, which is such that $E(z) = E(z, z^+)$ for any z in $\langle \mathbf{z} \rangle$ (see Chamberlain and Rothschild (1983)). In this context, we can alternatively test for spanning by testing the null hypothesis $H_0 : \boldsymbol{\phi}_{w_2}^+ = \mathbf{0}$ (see Peñaranda and Sentana (2004)). The main advantage of this procedure is that the uncentred (U) moment conditions on which it is based, namely

$$E[\mathbf{z}_t(\mathbf{z}_t'\boldsymbol{\phi}_w^+ - 1)] = E[\mathbf{h}_U(\mathbf{z}_t; \boldsymbol{\phi}_w^+)] = \mathbf{0}, \tag{4}$$

are linear in $\boldsymbol{\phi}_w^+$ and do not involve any nuisance parameters. Once more, estimation is trivial because $\boldsymbol{\phi}_w^+$ is exactly identified under the alternative.[5] As a result,

$$\hat{\boldsymbol{\phi}}_w^+ = (\hat{\boldsymbol{\Sigma}}_{zz} + \hat{\boldsymbol{\gamma}}_{wr}\hat{\boldsymbol{\gamma}}_{wr}')^{-1}\hat{\boldsymbol{\gamma}}_{wr}.$$

Likewise, it is also easy to make inferences about $\boldsymbol{\phi}_w^+$ which are robust to heteroskedasticity and autocorrelation in $\mathbf{h}_U(\mathbf{z}_t;\boldsymbol{\phi}_w^+)$. In addition, the linearity of both $\mathbf{h}_U(\mathbf{z}_t;\boldsymbol{\phi}_w^+)$ and H_0 in $\boldsymbol{\phi}_w^+$ implies that the trinity of asymptotic tests can be made numerically identical by using a common estimator of the asymptotic covariance matrix of $\sqrt{T}\bar{\mathbf{h}}_{UT}(\boldsymbol{\phi}_w^+)$ (see Newey and West (1987b)).

Similarly, we can also consider the Gibbons, Ross and Shanken (GRS) (1989) test, which is based on the multivariate regression of $\mathbf{z}_2$ on a constant and $\mathbf{z}_1$. Following MacKinlay and Richardson (1991), we can also cast such a multivariate regression in a GMM framework by using the orthogonality conditions

$$E\left[\begin{pmatrix} 1 \\ \mathbf{z}_{1t} \end{pmatrix}(\mathbf{z}_{2t} - \mathbf{a} - \mathbf{B}\mathbf{z}_{1t})\right] = E[\mathbf{h}_G(\mathbf{z}_t;\mathbf{a},\mathbf{b})] = \mathbf{0},$$

where $\mathbf{b} = vec(\mathbf{B})$, and the subscript G refers to the initial letter of GRS. In this context, the hypothesis of interest is simply $H_0 : \mathbf{a} = \mathbf{0}$.

The equivalence between the null hypotheses $H_0 : \boldsymbol{\varphi}_{w_2}^+ = \mathbf{0}$ and $H_0 : \boldsymbol{\phi}_{w_2}^+ = \mathbf{0}$ becomes immediately apparent if we use the Woodbury formula to show that

$$\boldsymbol{\phi}_w^+ = \frac{1}{1 + \boldsymbol{\gamma}_{wr}\boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\gamma}_{wr}}\boldsymbol{\varphi}_w^+,$$

so that z^+ and z^{++} are proportional. Similarly, it is also possible to prove that

$$z^{++} = z_1^{++} + \frac{1}{1 + \boldsymbol{\gamma}_{w_1r}'\boldsymbol{\Sigma}_{z_1z_1}^{-1}\boldsymbol{\gamma}_{w_1r}}\mathbf{a}'\boldsymbol{\Omega}^{-1}(\mathbf{z}_2 - \mathbf{a} - \mathbf{B}\mathbf{z}_1),$$

where $\boldsymbol{\Omega} = V(\mathbf{z}_2 - \mathbf{a} - \mathbf{B}\mathbf{z}_1)$, which confirms that the three null hypotheses are equivalent.

However, the fact that the parametric restrictions to test are equivalent does not necessarily imply that from a statistical point of view the corresponding GMM-based tests are equivalent too. The precise relationship between these three testing procedures is analysed in detail in Peñaranda and Sentana (2004). In this respect, they show that these three approaches to test for spanning are asymptotically equivalent under the null and sequences of local alternatives, and also, that they are consistent against fixed alternatives.[6]

3.3 Testing the optimality of the least squares prediction rule

In line with the motivation of our paper, we could also be interested in assessing whether the LLS prediction rule that agent a follows is proportional to the LMUSR rule. Once more, we can easily do so in a GMM framework by jointly considering the moment conditions

$$E\begin{bmatrix} \mathbf{w}_t(r_t - \mathbf{w}_t'\boldsymbol{\beta}_w) \\ \mathbf{z}_t(\mathbf{z}_t'\boldsymbol{\phi}_w^+ - 1) \end{bmatrix} = E\begin{bmatrix} \mathbf{h}_O(r_t, \mathbf{w}_t; \boldsymbol{\beta}_w) \\ \mathbf{h}_U(\mathbf{z}_t; \boldsymbol{\phi}_w^+) \end{bmatrix} = \mathbf{0},$$

or alternatively $E[\mathbf{h}_O(r_t, \mathbf{w}_t; \boldsymbol{\beta}_w)] = \mathbf{0}$ and $E[\mathbf{h}_E(\mathbf{z}_t; \boldsymbol{\varphi}_w^+, \boldsymbol{\mu}_z)] = \mathbf{0}$. In this context, the hypothesis of interest would be $H_0 : \boldsymbol{\phi}_w^+ = \lambda\boldsymbol{\beta}_w$, or equivalently, $H_0 : \boldsymbol{\varphi}_w^+ = \delta\boldsymbol{\beta}_w$, where λ and δ are additional scalar parameters to be estimated under H_0.

Again, we can apply the trinity of classical GMM tests to this problem, which will be asymptotically distributed as the same χ^2 random variable with k degrees of freedom under the null hypothesis. Moreover, since both $\boldsymbol{\beta}_w$ and $\boldsymbol{\phi}_w^+$ (or $\boldsymbol{\varphi}_w^+$) are exactly identified under the alternative, then the Distance Metric test still coincides with the usual Overidentifying Restrictions test.

4 Empirical illustration

In order to illustrate the estimation and testing procedures discussed in the previous section, we are going to revisit the empirical application in Pesaran and Timmermann (1995), who used least squares regressions to generate predictions of excess returns on Standard and Poor's 500 (SP500) portfolio using only ex ante dated variables. In particular, they estimated a monthly excess return regression over the period 1954(1) to 1992(12) in which the signals were:

1. x_{1t} : the first lag of the dividend yield defined as the ratio of dividends to share prices,
2. x_{2t} : the second lag of the rate of change of the twelve-month moving average of the producer prices index,
3. x_{3t} : the first lag of the change in the one-month T-bill rate, and
4. x_{4t} : the rate of change of the twelve-month moving average of the index of industrial production

(see Pesaran and Timmermann (1995) for data sources, transformations, and publication delays).

Their results are reproduced in the first column of Table 1, with the only difference that we have used the orthogonality conditions in (3) to obtain individual t-ratios and joint significance tests which are robust to heteroskedasticity and serial correlation. Nevertheless, our results confirm theirs, in the sense that both the Wald test and the overidentifying restriction test (J) clearly indicate that the null hypothesis of unpredictability of excess returns is rejected by the data.

The second column of the same table contains the coefficient estimates of the optimal linear forecasting rule in Proposition 1, which we have obtained by means of the GMM procedures described in Section 4. For the purposes of simplifying the comparisons between the two columns, we have chosen the arbitrary scale parameter by minimising the mean square error of the associated predictions. As for the individual t-ratios and joint significance tests, they have been obtained on the basis of the orthogonality conditions in equation (4). In this respect, note that the Wald and J versions of the *uncentred* representing portfolios spanning test of Peñaranda and Sentana (2004) indicate that the Sharpe ratio of a buy and hold strategy, which is .48216 in annual terms, is clearly dominated by a Sharpe ratio of 1.0178 for the optimal dynamic investment strategy. Exactly the same conclusion is reached with the *centred* representing portfolio and GRS versions of the spanning tests discussed in section 3, which we do not report for the sake of brevity.

The predictability of excess returns is confirmed by our robust version of the Henriksson-Merton-Pesaran-Timmermann statistic reported in Table 1, which we have computed as the HAC t-ratio of the slope coefficient in the regression of the sign of r on a constant and the sign of its predictions. Under the null hypothesis of lack of predictability, these directional market timing tests should be distributed as a standard normal random variable regardless of the potential heteroskedasticity and/or serial correlation in the joint data generating process for returns and signals, while they should have a positive mean under the alternative.

As expected, the correlation between actual and predicted excess returns is higher for the OLS rule than for the optimal rule. Similarly, the root mean square error is lower for the former than for the latter. In contrast, the Sharpe ratio of the optimal market timing rule is larger than the Sharpe ratio of the OLS-based rule, and the same applies to the correlation between the signs of the forecasts and the signs of the returns. However, the numerical differences are minor, reflecting the very high correlation (=.973) between the forecasts produced by the two empirical prediction rules. In fact, if we compute the test of optimality of the linear least squares projections described in Section 4.3, we find that we cannot reject the null hypothesis because the p-values of the HAC Wald and J tests are .468 and 1, respectively. Therefore, we can conclude that the least squares rule used by Pesaran and Timmermann (1995) to predict the SP500 excess returns was not statistically significantly different from the linear prediction rule that maximised the Sharpe ratio of the associated market timing strategy, which in turn reflects the fact that the joint empirical distribution of excess returns and signals is such that the moment condition in Proposition 2 cannot be rejected by the data.

5 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 managed 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 uses no information at all. We also show that the aforementioned passive portfolio allocation implicitly uses the linear forecasting rule that maximises the Sharpe ratio of actively traded portfolios. In addition, we discuss under what circumstances such an "optimal" forecast coincides (up to a factor of proportionality) with the least squares prediction.

In order to make such prediction rules operational, we have also developed a GMM estimation and testing strategy based on the moment conditions that define the optimal prediction rule. In this respect, we exploit the interpretation of our prediction rule in terms of mean-variance efficient portfolios, which allows us to use spanning tests to empirically assess whether or not a set of potential explanatory variable increases the Sharpe ratio of such market timing strategies. Finally, we propose a simple way of testing for the optimality of the least squares prediction rules.

An empirical illustration of our techniques with monthly excess returns on the SP500 portfolio confirms the predictability of this series on the basis of the regressors used by Pesaran and Timmermann (1995), and indicates that their linear least squares predictive regression was not statistically significantly different from the linear maximal unconditional Sharpe ratio rule.

Our theoretical results are not entirely surprising. First, we know from the asset pricing literature 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).

The fact that the passive fund manager is the best performer also 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 analysed 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.

Finally, 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 in the presence of return predictability, most conventional expected utility functions will lead to portfolio rules that either maximise the conditional Sharpe ratio, thus making the behaviour of the active fund manager superior to the behaviour of the passive fund manager, or are not even conditionally mean-variance efficient, which is a necessary but not sufficient condition for unconditional mean-variance efficiency, our results should also provide a note of warning regarding the indiscriminate use of such estimation methods.

For simplicity, we have considered a single risky asset, but in practice, portfolio decisions typically involve multiple assets. Perhaps the most natural approach to extend the analysis in this paper to a world with N risky assets whose excess returns $r_1, \ldots, r_N$ are predictable on the basis of a common set of signals, $x_1, \ldots, x_k$, would be to apply the rules of unconditional mean variance analysis to those N assets together with the Nk managed portfolios $z_{11} = r_1(x_1 - \mu_1), \ldots, z_{Nk} = r_N(x_k - \mu_k)$. In such a framework, we could compare the resulting unconditionally mean-variance efficient portfolio with the portfolio that would optimally combine the N managed portfolios whose positions are proportional to the least squares projections of each r_i on all the signals. Similarly, we could also deal with different predictors for different assets by simply eliminating some of the aforementioned managed portfolios from the investors' opportunity set.

It would be also be interesting to relax the linearity of the prediction rules that we have analysed, even though linearity is less restrictive that it may seem because instantaneous transformations involving one or several signals can be trivially accommodated as additional signals. Nevertheless, since the objective of such an exercise would be to obtain the maximal unconditional Sharpe ratio investment rule as a function of the signals $\mathbf{x}$, this is not a standard optimisation problem, but rather a problem in the calculus of variations. Hence, except in some simple examples, we should not expect to obtain closed form expressions. Instead, we should typically derive a first-order condition that must be satisfied by the optimal investment rule, analogous to the usual orthogonality condition that characterises conditional expectations as the non-linear prediction rule that minimises the unconditional mean square forecast error among all possible prediction rules (see e.g. Hansen and Sargent (1991)).

In addition, it would be worth extending the econometric analysis conducted in this paper, which like standard OLS prediction theory is in-sample in nature, to cover alternative sampling schemes, such as rolling and recursive in-samples, fixed or proportional out-of-sample sizes relative to the in-sample size, etc. (see e.g. McCracken (2000)).

Given the practical relevance of all these issues, they constitute obvious avenues for further research.

References

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

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

Britten-Jones, Mark. (1999). "The sampling error in estimates of mean-variance efficient portfolio weights." *Journal of Finance* 54, 655-671.

Chamberlain, Gary, and Michael Rothschild. (1983). "Arbitrage, factor structure, and mean-variance analysis on large asset markets." *Econometrica* 51, 1281-1304.

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

Geweke, John. (1981). "The approximate slopes of econometric tests." *Econometrica* 49, 1427-1442.

Gibbons, Michael R., Stephen A. Ross, and Jay Shanken. (1989). "A test of the efficiency of a given portfolio." *Econometrica* 57, 1121-1152.

Goetzmann, William N., Jonathan E. Ingersoll, Matthew I. Spiegel, and Ivo Welch. (2002). "Sharpening Sharpe ratios." Yale ICF Working Paper 02-08.

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

Hansen, Lars P. (1982). "Large sample properties of generalized method of moments estimators." *Econometrica* 50, 1029-1054.

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

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

MacKinlay, A. Craig, and Matthew P. Richardson. (1991). "Using GMM to test mean-variance efficiency." *Journal of Finance* 46, 511-527.

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

McCracken, Michael W. (2000). "Robust out-of-sample inference." *Journal of Econometrics* 99, 195-223.

Newey, Whitney K., and Daniel L. McFadden. (1994). "Large sample estimation and hypothesis testing." in Robert F. Engle and Daniel L. McFadden (eds.) *Handbook of Econometrics* vol. IV, 2111-2245, Elsevier.

Newey, Whitney K., and Kenneth D. West. (1987a). "A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix." *Econometrica* 55, 703-706.

Newey, Whitney K., and Kenneth D. West. (1987b). "Hypothesis testing with efficient method of moments estimation." *International Economic Review* 28, 777-787.

Peñaranda, Francisco, and Enrique Sentana. (2004). "Mean-variance tangency tests." mimeo, London School of Economics.

Pesaran, Hashem, and Allan Timmermann. (1995). "Predictability of stock returns: robustness and economic significance." *Journal of Finance* 50, 1201-28.

Sentana, Enrique (1999). "Least squares predictions and mean-variance analysis." CEPR Discussion Paper 2088.

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

Notes

[1]One formal way of rationalising such a behaviour is through conditional mean-variance analysis, under the maintained assumptions that the conditional expectations of returns are linear in 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 $\max_{w_r(I)} \{w_r(I)E(r|I) - .5\alpha w_r^2(I)V(r|I)\}$, where α is a positive risk aversion parameter, her optimal investment strategy will be $w_r^*(I) = \alpha^{-1}E(r|I)/V(r|I)$. Nevertheless, we would like to stress that it is not by any means necessary to make these assumptions for the validity of our results.

[2]In this way, we are implicitly guaranteeing 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.

[3]As explained in footnote 1, though, this is largely inconsequential for our resuts because we concentrate on linear prediction rules.

[4]In practice, $\mathbf{w}_t$ is unobserved, but we can either replace it with $(1, \mathbf{x}_t')'$, which only affects the definition of the "intercepts" of the different linear prediction rules, or else add the exactly identified moment conditions $E(\mathbf{x}_t - \boldsymbol{\mu}_x) = \mathbf{0}$ without altering our substantive results.

[5]As shown by Britten-Jones (1999), we can also obtain $\hat{\boldsymbol{\phi}}_w^+$ by simply regressing 1 on $\mathbf{z}$. The intuition is as follows. If we had a safe asset among our original set of assets, the mean-representing portfolio would be simply 1. Although this is not the case in our context, the mean-representing portfolio is the portfolio of $\mathbf{z}$ which is closest to 1 in the mean-square norm (see also Peñaranda and Sentana (2004)).

[6]In addition, these authors use Bahadur's concept of asymptotic relative efficiency to compare the power of the different spanning tests against fixed alternatives (see e.g. Geweke (1981)), and conclude that their ranking usually depends on the specific values of the parameters and the distributional assumptions.

Appendix

Auxiliary results

Cauchy-Schwartz Inequality

For any two vectors $\mathbf{a}$ and $\mathbf{b}$ of the same order, we have that

$$(\mathbf{a}'\mathbf{b})^2 \leq (\mathbf{a}'\mathbf{a})(\mathbf{b}'\mathbf{b}),$$

with equality if and only if $\mathbf{a}$ and $\mathbf{b}$ are linearly dependent.
(see e.g. Magnus and Neudecker (1988) for proofs and extensions).

Proof of Proposition 1

Formally, we can characterize (up to scale) the linear forecasting rule that maximises the Sharpe ratio, $\boldsymbol{\varphi}_w^{\oplus\prime}\mathbf{w}$, as

$$\boldsymbol{\varphi}_w^{\oplus\prime} = \arg\max_{\boldsymbol{\varphi}_w} \frac{\boldsymbol{\varphi}_w'\boldsymbol{\gamma}_{wr}\boldsymbol{\gamma}_{wr}'\boldsymbol{\varphi}_w}{\boldsymbol{\varphi}_w'\boldsymbol{\Sigma}_{zz}\boldsymbol{\varphi}_w}.$$

The solution to this well-known programme is an eigenvector associated with the maximum eigenvalue of the rank 1 matrix $\boldsymbol{\gamma}_{wr}\boldsymbol{\gamma}_{wr}'$ in the metric of $\boldsymbol{\Sigma}_{zz}$. That is,

$$\max_{\boldsymbol{\varphi}_w} \frac{\boldsymbol{\varphi}_w'\boldsymbol{\gamma}_{wr}\boldsymbol{\gamma}_{wr}'\boldsymbol{\varphi}_w}{\boldsymbol{\varphi}_w'\boldsymbol{\Sigma}_{zz}\boldsymbol{\varphi}_w} = \lambda_1(\boldsymbol{\Sigma}_{zz}^{-1/2}\boldsymbol{\gamma}_{wr}\boldsymbol{\gamma}_{wr}'\boldsymbol{\Sigma}_{zz}^{-1/2}) = \boldsymbol{\gamma}_{wr}'\boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\gamma}_{wr} = s^2(r_p)$$

where $\lambda_1(\mathbf{A})$ denotes the largest eigenvalue of the matrix $\mathbf{A}$. This confirms that $\boldsymbol{\varphi}_w^{\oplus} \propto \boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\gamma}_{wr}$, as required. □

Proof of Proposition 2

If we call $\mathbf{a} = \boldsymbol{\Sigma}_{zz}^{-1/2}\boldsymbol{\gamma}_{wr}$ and $\mathbf{b} = \boldsymbol{\Sigma}_{zz}^{1/2}\boldsymbol{\Gamma}_{ww}^{-1}\gamma_{wr}$, then by the Cauchy-Schwartz inequality,

$$\sigma_f^2 = (\boldsymbol{\gamma}_{wr}'\boldsymbol{\Gamma}_{ww}^{-1}\boldsymbol{\gamma}_{wr})^2 \leq (\boldsymbol{\gamma}_{wr}'\boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\gamma}_{wr})(\boldsymbol{\gamma}_{wr}'\boldsymbol{\Gamma}_{ww}^{-1}\boldsymbol{\Sigma}_{zz}\boldsymbol{\Gamma}_{ww}^{-1}\boldsymbol{\gamma}_{wr}) = (\boldsymbol{\gamma}_{wr}'\boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\gamma}_{wr})(\boldsymbol{\beta}_w'\boldsymbol{\Sigma}_{zz}\boldsymbol{\beta}_w),$$

so

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

and $s(r_p) \geq s(r_a)$ because they are both positive. Equality is achieved in the above inequality if and only if $\mathbf{\Sigma}_{zz}^{1/2}\mathbf{\Gamma}_{ww}^{-1}\boldsymbol{\gamma}_{wr} = \mathbf{\Sigma}_{zz}^{-1/2}\boldsymbol{\gamma}_{wr}\theta$, where θ is a non-zero scalar, or equivalently, if and only if $\mathbf{\Gamma}_{ww}^{-1}\boldsymbol{\gamma}_{wr} = \theta\mathbf{\Sigma}_{zz}^{-1}\boldsymbol{\gamma}_{wr}$, as stated. $\square$

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

$$\mathbf{\Gamma}_{ww}^{-1/2}\mathbf{\Sigma}_{zz}\mathbf{\Gamma}_{ww}^{-1/2}\frac{\mathbf{\Gamma}_{ww}^{-1/2}\boldsymbol{\gamma}_{wr}}{\sqrt{\boldsymbol{\gamma}_{wr}'\mathbf{\Gamma}_{ww}^{-1}\boldsymbol{\gamma}_{wr}}} = \theta\frac{\mathbf{\Gamma}_{ww}^{-1/2}\boldsymbol{\gamma}_{wr}}{\sqrt{\boldsymbol{\gamma}_{wr}'\mathbf{\Gamma}_{ww}^{-1}\boldsymbol{\gamma}_{wr}}},$$

so that $(\boldsymbol{\gamma}_{wr}'\mathbf{\Gamma}_{ww}^{-1}\boldsymbol{\gamma}_{wr})^{-1/2}\mathbf{\Gamma}_{ww}^{-1/2}\boldsymbol{\gamma}_{wr}$ can be regarded as a normalized eigenvector of the matrix $\mathbf{\Gamma}_{ww}^{-1/2}\mathbf{\Sigma}_{zz}\mathbf{\Gamma}_{ww}^{-1/2}$.

In this context, if we define $\mathbf{P}$ as an arbitrary symmetric positive semidefinite matrix of order $(k+1)$, then it is easy to see that the matrix

$$\theta\frac{\mathbf{\Gamma}_{ww}^{-1/2}\boldsymbol{\gamma}_{wr}\boldsymbol{\gamma}_{wr}'\mathbf{\Gamma}_{ww}^{-1/2}}{\boldsymbol{\gamma}_{wr}'\mathbf{\Gamma}_{ww}^{-1}\boldsymbol{\gamma}_{wr}} + \left(\mathbf{I}\text{-}\frac{\mathbf{\Gamma}_{ww}^{-1/2}\boldsymbol{\gamma}_{wr}\boldsymbol{\gamma}_{wr}'\mathbf{\Gamma}_{ww}^{-1/2}}{\boldsymbol{\gamma}_{wr}'\mathbf{\Gamma}_{ww}^{-1}\boldsymbol{\gamma}_{wr}}\right)\mathbf{P}\left(\mathbf{I}\text{-}\frac{\mathbf{\Gamma}_{ww}^{-1/2}\boldsymbol{\gamma}_{wr}\boldsymbol{\gamma}_{wr}'\mathbf{\Gamma}_{ww}^{-1/2}}{\boldsymbol{\gamma}_{wr}'\mathbf{\Gamma}_{ww}^{-1}\boldsymbol{\gamma}_{wr}}\right)$$

will have θ as one of its eigenvalues, and $(\boldsymbol{\gamma}_{wr}'\mathbf{\Gamma}_{ww}^{-1}\boldsymbol{\gamma}_{wr})^{-1/2}\mathbf{\Gamma}_{ww}^{-1/2}\boldsymbol{\gamma}_{wr}$ as the associated eigenvector. Hence, the above condition will be satisfied if and only if

$$\mathbf{\Sigma}_{zz} = \theta\frac{\boldsymbol{\gamma}_{wr}\boldsymbol{\gamma}_{wr}'}{\boldsymbol{\gamma}_{wr}'\mathbf{\Gamma}_{ww}^{-1}\boldsymbol{\gamma}_{wr}} + \left(\mathbf{\Gamma}_{ww}^{-1} - \frac{\boldsymbol{\gamma}_{wr}\boldsymbol{\gamma}_{wr}'}{(\boldsymbol{\gamma}_{wr}'\mathbf{\Gamma}_{ww}^{-1}\boldsymbol{\gamma}_{wr})}\right)\mathbf{Q}\left(\mathbf{\Gamma}_{ww} - \frac{\boldsymbol{\gamma}_{wr}\boldsymbol{\gamma}_{wr}'}{(\boldsymbol{\gamma}_{wr}'\mathbf{\Gamma}_{ww}^{-1}\boldsymbol{\gamma}_{wr})}\right),$$

where $\mathbf{Q} = \mathbf{\Gamma}_{ww}^{-1/2}\mathbf{P}\mathbf{\Gamma}_{ww}^{-1/2}$ is any arbitrary symmetric positive semidefinite matrix of dimension $k+1$.

To guarantee that $\mathbf{\Sigma}_{zz}$ has full rank, we can choose $\mathbf{P} = \mathbf{U\Theta U}'$, where $\mathbf{\Theta}$ is any diagonal positive definite matrix of order k, and $\mathbf{U}$ is any $(k+1)\times k$ orthogonal matrix such that $\mathbf{I} - \mathbf{\Gamma}_{ww}^{-1/2}\boldsymbol{\gamma}_{wr}(\boldsymbol{\gamma}_{wr}'\mathbf{\Gamma}_{ww}^{-1}\boldsymbol{\gamma}_{wr})^{-1}\boldsymbol{\gamma}_{wr}'\mathbf{\Gamma}_{ww}^{-1/2} = \mathbf{UU}'$.

Proof of Proposition 3

We have already seen that $r_p = \boldsymbol{\mu}_z'\mathbf{\Sigma}_{zz}^{-1}\mathbf{z}$, $E(r_p) = \boldsymbol{\mu}_z'\mathbf{\Sigma}_{zz}^{-1}\boldsymbol{\mu}_z$, and $V(\mathbf{r}_p) = \boldsymbol{\mu}_z'\mathbf{\Sigma}_{zz}^{-1}\boldsymbol{\mu}_z$. Therefore,

$$\begin{aligned} s^2(r_p) &= \boldsymbol{\mu}_z'\mathbf{\Sigma}_{zz}^{-1}\boldsymbol{\mu}_z = \boldsymbol{\mu}_z' dg^{-1/2}(\mathbf{\Sigma}_{zz}) dg^{1/2}(\mathbf{\Sigma}_{zz})\mathbf{\Sigma}_{zz}^{-1} dg^{1/2}(\mathbf{\Sigma}_{zz}) dg^{-1/2}(\mathbf{\Sigma}_{zz})\boldsymbol{\mu}_z \\ &= s'(\mathbf{z})\boldsymbol{\rho}_{zz}^{-1}s(\mathbf{z}), \end{aligned}$$

as required. $\square$

Proof of Proposition 4

Since in view of our assumptions

$$\begin{aligned} E[(r-\mu_r)|\mathbf{x}] &= \boldsymbol{\beta}_x'(\mathbf{x}-\boldsymbol{\mu}_x), \\ E[(r-\mu_r)^2|\mathbf{x}] &= \boldsymbol{\beta}_x'(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{x}-\boldsymbol{\mu}_x)'\boldsymbol{\beta}_x + \sigma_u^2, \end{aligned}$$

it follows from the law of iterated expectations that

$$E[(r-\mu_r)(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{x}-\boldsymbol{\mu}_x)'] = E[\boldsymbol{\beta}_x'(\mathbf{x}-\boldsymbol{\mu}_x)\cdot(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{x}-\boldsymbol{\mu}_x)'],$$

$$E[(r-\mu_r)^2(\mathbf{x}-\boldsymbol{\mu}_x)] = E\{[\boldsymbol{\beta}_x'(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{x}-\boldsymbol{\mu}_x)'\boldsymbol{\beta}_x]\cdot(\mathbf{x}-\boldsymbol{\mu}_x)\},$$

$$E[(r-\mu_r)^2(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{x}-\boldsymbol{\mu}_x)'] =$$

$$E\{[\boldsymbol{\beta}_x'(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{x}-\boldsymbol{\mu}_x)'\boldsymbol{\beta}_x]\cdot(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{x}-\boldsymbol{\mu}_x)'\}.$$

Then, we can use the following well known results on the multivariate normal distribution (see e.g. Arellano (1989))

$$E[(\mathbf{x}-\boldsymbol{\mu}_x)\otimes(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{x}-\boldsymbol{\mu}_x)'] = \mathbf{0},$$

$$E[(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{x}-\boldsymbol{\mu}_x)'\otimes(\mathbf{x}-\boldsymbol{\mu}_x)(\mathbf{x}-\boldsymbol{\mu}_x)'] =$$

$$(\boldsymbol{\Sigma}_{xx}\otimes\boldsymbol{\Sigma}_{xx})(\mathbf{K}+\mathbf{I}) + vec(\boldsymbol{\Sigma}_{xx})vec'(\boldsymbol{\Sigma}_{xx}),$$

where $\mathbf{K}$ is the commutation matrix (see e.g. Magnus and Neudecker (1988)), to show that

$$\boldsymbol{\Sigma}_{zz} = \begin{pmatrix} \sigma_r^2-\mu_r^2 & \mathbf{0}' \\ \mathbf{0} & (\sigma_r^2+\mu_r^2)\boldsymbol{\Sigma}_{xx} \end{pmatrix} + \begin{pmatrix} \mu_r \\ \boldsymbol{\sigma}_{xr} \end{pmatrix} \begin{pmatrix} \mu_r & \boldsymbol{\sigma}_{xr}' \end{pmatrix}.$$

If we use the Woodbury formula, then after some tedious algebra we end up with

$$\boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\gamma}_{rw} = \frac{1}{\sigma_r^2\{1+[1-s^2(r)][1+s^2(r)]^{-1}R^2\}}\left\{ \begin{array}{c} \mu_r \\ [1-s^2(r)][1+s^2(r)]^{-1}\boldsymbol{\beta}_x \end{array} \right\},$$

which proves part 1.

To prove part 2, we simply need to carefully compute

$$\begin{aligned} s(r_p) &= \sqrt{\boldsymbol{\gamma}_{wr}'\boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\gamma}_{wr}} = \sqrt{\frac{s^2(r)+[1-s^2(r)][1+s^2(r)]^{-1}R^2}{1+[1-s^2(r)][1+s^2(r)]^{-1}R^2}}, \\ s(r_a) &= \frac{\boldsymbol{\gamma}_{wr}'\boldsymbol{\Gamma}_{ww}^{-1}\boldsymbol{\gamma}_{wr}}{\sqrt{\boldsymbol{\beta}_w'\boldsymbol{\Sigma}_{zz}\boldsymbol{\beta}_w}} = \frac{s^2(r)+R^2}{\sqrt{s^2(r)(1+3R^2)+(1+R^2)R^2}}, \end{aligned}$$

whence it follows that

$$s^2(r_p)-s^2(r_a) = \frac{4s^6(r)R^2}{\{1+[1-s^2(r)][1+s^2(r)]^{-1}R^2\}[s^2(r)(1+3R^2)+(1+R^2)R^2]} \geq 0$$

as required. □

Proof of Proposition 5

The proof consists of two steps. First, we shall obtain all the required expressions for the case of a single signal in terms of $s(r)$, ρ_{r1}, ϕ_{r1}, ϕ_{1r} and κ_{r1} regardless of the joint distribution of returns and signals. Then, we shall obtain expressions for those quantities in the particular case of joint log-normally distributed returns and signal.

But before, we need to obtain expressions for $cov(r, z_1)$ and $V(z_1)$. In this respect, note that

$$\begin{aligned} cov(r, z_1) &= E\{(r-\mu_r)[r(x_1-\mu_1)-\sigma_{r1}]\} \\ &= \sigma_r^2\sigma_1 E\left[\left(\frac{r-\mu_r}{\sigma_r}\right)^2\left(\frac{x_1-\mu_1}{\sigma_1}\right)+\frac{\mu_r}{\sigma_r}\left(\frac{r-\mu_r}{\sigma_r}\right)^2\left(\frac{x_1-\mu_1}{\sigma_1}\right)\right] \\ &= \sigma_r^2\sigma_1[\phi_{r1}+s(r)\rho_{r1}] \end{aligned}$$

and

$$\begin{gathered} V(z_1) = E[r^2(x_1-\mu_1)^2]-\sigma_{r1}^2 \\ = E\{[(r-\mu_r)^2+2\mu_r(r-\mu_r)+\mu_r^2](x_1-\mu_1)^2\}-\sigma_{r1}^2 \\ = \sigma_r^2\sigma_1^2\left\{E\left[\left(\frac{r-\mu_r}{\sigma_r}\right)^2\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right]\right. \\ \left.+2s(r)E\left[\left(\frac{r-\mu_r}{\sigma_r}\right)\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right]+s^2(r)-\rho_{r1}^2\right\} \\ = \sigma_r^2\sigma_1[\kappa_{r1}+2s(r)\phi_{1r}+s^2(r)-\rho_{r1}^2]. \end{gathered}$$

Hence, we can write

$$\mathbf{\Sigma}_{zz} = \left\{\begin{array}{cc} \sigma_r^2 & \sigma_r^2\sigma_1[\phi_{r1}+s(r)\rho_{r1}] \\ \sigma_r^2\sigma_1[\phi_{r1}+s(r)\rho_{r1}] & \sigma_r^2\sigma_1^2[\kappa_{r1}+2s(r)\phi_{1r}+s^2(r)-\rho_{r1}^2] \end{array}\right\},$$

from where

$$\mathbf{\Sigma}_{zz}^{-1}\gamma_{wr} = \frac{1}{\sigma_r \left\{[\kappa_{r1} + 2s(r)\phi_{1r} + s^2(r) - \rho_{r1}^2] - [\phi_{r1} + s(r)\rho_{r1}]^2\right\}} \times \left\{ \begin{array}{c} s(r)[\kappa_{r1} + 2s(r)\phi_{1r} + s^2(r) - 2\rho_{r1}^2] - \rho_{r1}\phi_{r1} \\ \{\rho_{r1}[1 - s^2(r)] - s(r)\phi_{r1}\}\sigma_1^{-1} \end{array} \right\},$$

which directly gives us the weights of the LMUSR forecasting rule. In contrast, the least squares weights are simply

$$\mathbf{\Gamma}_{ww}^{-1}\gamma_{wr} = \sigma_r \left[\begin{array}{c} s(r) \\ \rho_{r1}\sigma_1^{-1} \end{array} \right]$$

Finally, we plug in these formulae in the expressions in section 1 to obtain the Sharpe ratios of r_a and r_p.

As for the required moments of the joint log-normal distribution, we only need to make repeated use of the expression

$$E(R^l x_1^j) = E[\exp(l \ln R + j \ln x_1)] = l\nu_R + j\nu_1 + .5(l^2\delta_R + j^2\delta_1 + 2lj\delta_{R1}).$$

In this respect, note that a necessary and sufficient condition for ρ_{r1} to be 0 is that $\pi_{R1} = 0$. In that case, though, the joint normality of $\ln R$ and $\ln x_1$ implies that R and x_1 are stochastically independent, so that $\phi_{r1} = \phi_{1r} = 0$ and $\kappa_{r1} = 1$.

In contrast, when $\rho_{r1} \neq 0$, the coefficient vectors $\mathbf{\Gamma}_{ww}^{-1}\gamma_{wr}$ will be proportional to $\mathbf{\Sigma}_{zz}^{-1}\gamma_{wr}$ when $s(r) > 0$ if and only if

$$\rho_{r1}s(r)[\kappa_{r1} + 2s(r)\phi_{1r} + 2s^2(r) - 2\rho_{r1}^2 - 1] - [\rho_{r1}^2 - s^2(r)]\phi_{r1} = 0,$$

which coincides with the cubic equation in the statement of the proposition.

Finally, note that when $\rho_{r1} \neq 0$ but $s(r) = 0$, then $\mathbf{\Sigma}_{zz}^{-1}\gamma_{wr}$ and $\mathbf{\Gamma}_{ww}^{-1}\gamma_{wr}$ cannot be proportional, since $\phi_{r1} \neq 0$ as long as $\pi_{R1} \neq 0$. $\square$

Table 1

Prediction rules for SP500 monthly excess returns

1954:1 to 1992:12 (468 observations)

Predictors	**OLS**	**Optimal rule**
constant	-.0240	-.0150
t-ratio	*-2.688*	*-1.148*
	(-2.489)	(-1.308)
Dividend yield	14.27	11.99
t-ratio	*4.752*	*2.800*
	(4.365)	(3.393)
Producer prices	-.2786	-.2868
t-ratio	*-4.096*	*-4.514*
	(-4.026)	(-5.365)
T-bill rates	$-.688\times10^{-2}$	$-.422\times10^{-2}$
t-ratio	*-2.068*	*-2.501*
	(-2.239)	(-2.079)
Industrial production	-.1586	-.1834
t-ratio	*-5.615*	*-4.442*
	(-4.328)	(-4.400)
R^2	.0870	.0824
RMSE	.04050	.04061
Sharpe ratio	.989	1.018
Directional correlation	.181	.216
Wald tests for	*53.22*	*33.85*
zero slopes	(44.55)	(46.94)
J tests for	*14.26*	*13.66*
zero slopes	(32.06)	(30.41)
PT tests	*3.639*	*4.303*
	(3.930)	(4.735)

Notes: The test statistics in italics have been computed by means of the usual Newey-West (1987a) expressions with 8 lags, while the values in brackets use 0 lags. The 1% critical value of a χ^2 with 4 degrees of freedom is 13.3. Directional correlation refers to the linear correlation coefficient between the signs of the forecasts and the signs of the returns. Finally, the J statistic is the usual Sargan-Hansen overidentifying restriction test, while the PT statistics correspond to the Henriksson-Merton-Pesaran-Timmermann market timing test.