Least Squares Predictions and Mean-Variance Analysis

Enrique Sentana
CEMFI

Working Paper No. 9711
September 1997

I would like to thank Manuel Arellano, Rafael Repullo and Sushil Wadhwani for very useful discussions. Of course, the usual caveat applies. (e-mail: sentana@cemfi.es).

CEMFI, Casado del Alisal 5, 28014 Madrid, Spain.
Tel: 341 4290551, fax: 341 4291056, http://www.cemfi.es.

Abstract

In an economy with one riskless and one risky asset, we compare the Sharpe ratios of investment funds that follow: i) timing strategies which forecast the market 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 ranking of i) and ii) is generally unclear. We also discuss under what circumstances the performance of ii) and iii) coincides.

1 Introduction

From a formal point of view, mean-variance analysis and linear projections are very closely related, as both are the result of the minimization of a mean square norm over a closed linear subspace of the set of all random variables with finite second moments. 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 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 note is to determine to what extent this intuition is correct. We do so in the context of a model with a safe asset and a risky one, in which closed-form analytical solutions are available. Under the maintained assumption that fund managers are conditional mean-variance optimizers, we consider alternative linear 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 the market; ii) an active strategy which uses multiple regression instead; and iii) a passive portfolio allocation which combines the funds in i) with constant weightings. Furthermore, we obtain an expression for the linear forecasting rule that maximizes the Sharpe ratio of an actively traded portfolio, and discuss under which circumstances such "optimal" forecasts coincide with least squares predictions.

The rest of this note is organized as follows. We introduce the theoretical set-up in section 2, and derive the active and passive portfolio strategies mentioned above. Then, in section 3, we discuss in detail a special case of our model in which returns and predictor variables are jointly normally distributed. General results in terms of Sharpe ratios are obtained in section 4. Finally, section 5 contains a discussion of our results in relation to several areas of current research interest in the finance and econometrics literatures.

2 Basic Set-up

Let's consider a world with a safe asset and a risky one (i.e. the "market"). Let r_t be the excess return on the risky asset, and suppose that there are k indicator variables, $\mathbf{x}_t' = (x_{1t}, \ldots, x_{kt})$, known in period $t-1$, which help predict r_t.

If uninformed investors allocate their wealth between the two assets according to standard (i.e. unconditional) mean-variance analysis, they will invest a constant fraction $E(r_t)/[\alpha V(r_t)]$ of their wealth in the risky asset, where α is their common risk aversion parameter. To make the comparisons simpler, we assume that $E(r_t) = 0$, so that in the absence of information (our benchmark case), individual investors only hold cash.

Let's now suppose that there are k fund managers, each endowed with information on a single indicator variable, x_{jt}, $j = 1, \ldots, k$, who pursue active portfolio strategies according to a variant of conditional mean-variance analysis, in which conditional expectations are replaced by linear projections, and conditional variances by mean square forecast errors. More precisely, we assume that the objective function of manager j at time $t-1$ is

$$\max_{w(x_{jt})} \left\{ w(x_{jt}) E^*(r_t \mid x_{jt}) - \frac{\alpha}{2} w^2(x_{jt}) E\left[r_t - E^*(r_t \mid x_{jt})\right]^2 \right\}$$

where $E^*(y \mid \mathbf{z})$ denotes the least squares projection of y on the linear span generated by a constant and $\mathbf{z}$, and $E^*(y) = E(y)$ (see e.g. Hansen and Sargent (1991)). 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 $\tilde{x}_{jt} = x_{jt} - \nu_j$ as the demeaned value of the j^{th} predictor variable, $\delta_j = \sigma_{jr}/\sigma_{rr}$ as the coefficient in the (theoretical) simple regression of r_t on x_{jt}, $\varepsilon_{jt} = r_{jt} - \delta_j x_{jt}$ as the associated prediction error, $\sigma_{\varepsilon_j\varepsilon_j} = \sigma_{rr} - \sigma_{jr}^2/\sigma_{jj}$ as its variance, and $\rho_{jr} = \sigma_{jr}/\sqrt{\sigma_{rr}\sigma_{x_jx_j}}$ as the theoretical correlation coefficient between r_{jt} and x_{jt}. Then, excess returns on each fund will be

$$r_{jt} = \frac{1}{\alpha} \cdot \frac{\delta_j \tilde{x}_{jt}}{\sigma_{\varepsilon_j \varepsilon_j}} \cdot r_t \tag{1}$$

so that

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

with equality if and only if the j^{th} indicator variable has no predictive power at all. Notice that r_{jt} 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_{jt}. However, such a timing strategy is also riskier, since obviously $V(r_{jt}) \geq 0$. For that reason, and in line with standard practice, we shall use the unconditional risk-return trade-off (or Sharpe ratio) of manager $j's$ portfolio to evaluate its performance taking into account its risk. In particular, since

$$V(r_{jt}) = \frac{1}{\alpha^2} \frac{\delta_j^2 \lambda_{jj}}{\sigma_{\varepsilon_j \varepsilon_j}^2} \tag{3}$$

where $\lambda_{jj} = V(\tilde{x}_{jt} r_t)$, manager $j's$ Sharpe ratio is

$$s_j = \frac{E(r_{jt})}{\sqrt{V(r_{jt})}} = \frac{|\sigma_{jr}|}{\sqrt{\lambda_{jj}}} \tag{4}$$

Suppose that there is another fund manager, a say, who also follows an active investment strategy based on the same mean-variance analysis rule as the first k managers, but this time knowing the whole of $\mathbf{x}_t$. Let $\boldsymbol{\beta} = \boldsymbol{\Sigma}_{xx}^{-1} \boldsymbol{\sigma}_{xr}$ be the coefficients of the (theoretical) multiple regression of returns on the indicators, $\hat{r}_t = E(r_t) + \boldsymbol{\sigma}_{xr}' \boldsymbol{\Sigma}_{xx}^{-1} (\mathbf{x}_t - \boldsymbol{\nu}) = \boldsymbol{\beta}' \tilde{\mathbf{x}}_t$ the fitted values from that regression, $u_t = r_t -$

$\boldsymbol{\beta}'\tilde{\mathbf{x}}_t$ 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 an excess return of

$$r_{at} = \frac{1}{\alpha}\frac{\boldsymbol{\beta}'\tilde{\mathbf{x}}_t}{\sigma_{uu}} \cdot r_t \tag{5}$$

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

$$E(r_{at}) = \frac{1}{\alpha}\frac{\sigma_{\hat{r}\hat{r}}}{\sigma_{uu}} = \frac{1}{\alpha} \cdot \frac{R^2}{1 - R^2} \tag{6}$$

so that $E(r_{jt}) \leq E(r_{at})$ for $j = 1, \ldots, k$, as $R^2 \geq \rho_{jr}^2$. Also, since

$$V(r_{at}) = \frac{1}{\alpha^2}\frac{\boldsymbol{\beta}'\boldsymbol{\Lambda}\boldsymbol{\beta}}{\sigma_{uu}^2} \tag{7}$$

where $\boldsymbol{\Lambda} = V(\tilde{\mathbf{x}}_t r_t)$, manager a's Sharpe ratio will be

$$s_a = \frac{\sigma_{\hat{r}\hat{r}}}{\sqrt{\boldsymbol{\beta}'\boldsymbol{\Lambda}\boldsymbol{\beta}}} \tag{8}$$

Finally, suppose that there is yet another manager, p say, who does not observe $\mathbf{x}_t$ at all, but constructs 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 $\boldsymbol{\mu} = E(\mathbf{r}_t)$ and $\mathbf{-} \ = V(\mathbf{r}_t)$, where $\mathbf{r}_t$ is the vector of excess returns on each fund, i.e. $\mathbf{r}'_t = (r_{1t}, \ldots, r_{kt})$. 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 vector notation, we can write

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

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

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

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

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

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

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

From here, it is straightforward to see that

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

$$V(r_{pt}) = \frac{1}{\alpha^2}\boldsymbol{\sigma}'_{xr}\mathbf{\Lambda}^{-1}\boldsymbol{\sigma}_{xr} \tag{14}$$

and

$$s_p = \frac{E(r_{pt})}{\sqrt{V(r_{pt})}} = \sqrt{\boldsymbol{\sigma}'_{xr}\mathbf{\Lambda}^{-1}\boldsymbol{\sigma}_{xr}} \tag{15}$$

3 A multivariate normal example

Let's initially assume that r_t and $\mathbf{x}_t$ are jointly normally distributed.[1] Using well known results on fourth moments of the multivariate normal distribution (see e.g. Fang, Kotz and Ng (1990)), we have that $\boldsymbol{\Lambda} = \sigma_{rr}\boldsymbol{\Sigma}_{xx} + \boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}$. Then, on the basis of the Woodbury formula, we get:

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

After some algebraic manipulations, we finally obtain that the returns on the passive strategy will be

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

Therefore, r_{pt} is exactly proportional to r_{at} (cf. (5)), with a time-invariant 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_{pt} and r_{at} is trivially one. Hence, although the mean and variance of r_{at} 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.

[1] Note that in this case, conditional expectations and linear projections on the one hand, and conditional variances and mean square errors on the other, coincide.

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

d) Since we know from the theory of mean-variance analysis with a safe asset that the Sharpe ratio of the optimal portfolio will be higher than the Sharpe ratio of any other portfolio, including the original assets, the Sharpe ratio of r_{at}, will be at least as high as the Sharpe ratio of any r_{jt}. Therefore, fund manager a, who uses information on the entire vector $\mathbf{x}_t$, will do at least as well as any manager who only uses information on a particular x_{jt}, or indeed a subset of them.

More explicitly, since in this case

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

the unconditional Sharpe ratio of r_{pt} and r_{at} is

$$s_a = s_p = \sqrt{\frac{\sigma_{\hat{r}\hat{r}}}{(\sigma_{rr} + \sigma_{\hat{r}\hat{r}})}} = \sqrt{\frac{R^2}{1+R^2}}$$

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

$$s_j = \frac{|\sigma_{jr}|}{\sqrt{\sigma_{rr}\sigma_{jj} + \sigma_{jr}^2}} = \frac{|\rho_{jr}|}{\sqrt{1+\rho_{jr}^2}}$$

As a consequence, the Sharpe ratio of an individual fund will also be higher the more correlated x_{jt} is with r_t (in absolute value), but it could never exceed the Sharpe ratio of r_{at}.

4 A general inequality

The above results, though, depend crucially on the normality assumption. Since

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

and

$$E(r_t^2 \tilde{\mathbf{x}}_t \tilde{\mathbf{x}}_t') = E\left[E(r_t^2 \mid \tilde{\mathbf{x}}_t)\tilde{\mathbf{x}}_t \tilde{\mathbf{x}}_t'\right]$$

we could alternatively make assumptions about $E(r_t^2 \mid \tilde{\mathbf{x}}_t)$, or $V(r_t \mid \tilde{\mathbf{x}}_t)$ and $E(r_t \mid \tilde{\mathbf{x}}_t)$. In particular, if we make the somewhat contrived assumption that $E(r_t^2 \mid \tilde{\mathbf{x}}_t) = E(r_t^2) = \sigma_{rr}$, we obtain the remarkable result that $r_{pt} = r_{at}$. In general, though, we would not expect r_{pt} and r_{at} to be proportional. Nevertheless, we can still compare their Sharpe ratios.

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

$$s_p^2 \geq s_a^2 \tag{16}$$

and $s_p \geq s_a$ given that they are both positive. Hence, 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 this respect, it is worth mentioning that equality is achieved in (16) not only under Gaussianity, but also for all multivariate elliptically symmetric distributions with bounded fourth moments. The reason is that for this family of distributions

$$\mathbf{\Lambda} = (\kappa + 1)\sigma_{rr}\mathbf{\Sigma}_{xx} + (2\kappa + 1)\boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}$$

where κ is the coefficient of multivariate (excess) kurtosis (see Fang, Kotz and Ng (1990)), so that $\mathbf{\Lambda}^{-1/2}\boldsymbol{\sigma}_{xr}$ and $\mathbf{\Lambda}^{1/2}\boldsymbol{\beta}(= \mathbf{\Lambda}^{1/2}\mathbf{\Sigma}_{xx}^{-1}\boldsymbol{\sigma}_{xr})$ are proportional. Examples include the multivariate normal in section 3 ($\kappa = 0$), the multivariate t with $\nu > 4$ degrees of freedom ($\kappa = 2/(\nu - 4)$), as well as uniform distributions on the unit sphere.

As discussed in the previous section, we also know that $s_p \geq s_j$ for all j. However, we cannot rank in general s_a and s_j, so that manager a, who uses information on the entire vector $\mathbf{x}_t$, may do better or worse than a manager who only uses information on a particular x_{jt}, despite the fact that expected excess returns for a are always higher. In principle, we would expect $s_p \geq s_a \geq s_j$ for all j. However, it is possible to construct numerical counterexamples in which $s_a < s_j < s_p$ for some j.

Finally, notice that manager p's behaviour is observationally equivalent to that of an active portfolio manager who used $(\sigma_{uu})\boldsymbol{\sigma}'_{xr}\mathbf{\Lambda}^{-1}\tilde{\mathbf{x}}_t$ instead of $\boldsymbol{\sigma}'_{xr}\mathbf{\Sigma}_{xx}^{-1}\tilde{\mathbf{x}}_t$ as her linear prediction rule. In fact, it turns out that $\boldsymbol{\sigma}'_{xr}\boldsymbol{\Lambda}^{-1}\tilde{\mathbf{x}}_t$ is (proportional to) the linear forecasting rule, $\boldsymbol{\gamma}^{*\prime}\tilde{\mathbf{x}}_t$ say, that maximizes the ratio of excess mean return to standard deviation of an actively traded portfolio. The proof of this result, which generalizes (16), is based on the fact that $\boldsymbol{\sigma}'_{xr}\boldsymbol{\Lambda}^{-1}$ is (proportional to) 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, $\max_{\gamma} \frac{\boldsymbol{\gamma}'\boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}\boldsymbol{\gamma}}{\boldsymbol{\gamma}'\boldsymbol{\Lambda}\boldsymbol{\gamma}} = \lambda_1(\boldsymbol{\Lambda}^{-1/2}\boldsymbol{\sigma}_{xr}\boldsymbol{\sigma}'_{xr}\boldsymbol{\Lambda}^{-1/2}) = \boldsymbol{\sigma}'_{xr}\boldsymbol{\Lambda}^{-1}\boldsymbol{\sigma}_{xr} = s_p^2$, where $\lambda_1(\mathbf{A})$ denotes the largest eigenvalue of the matrix $\mathbf{A}$.

5 Summary and Discussion

In an economy with one riskless and one risky asset, we show that a dynamic portfolio strategy which combines multiple regression with a mean-variance optimizer, 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. In fact, it is possible to construct counterexamples in which the manager who uses all the available information will perform strictly worse than a manager who only uses information on a particular variable. We also show that such a passive portfolio allocation implicitly uses the linear forecasting rule that maximizes the Sharpe ratio of actively traded portfolios. Nevertheless, we prove that if excess returns and predictor variables are jointly elliptically distributed, least squares regression-based forecasts are optimal.

Our results, though, 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 maximize unconditional Sharpe ratios. At the same time, since the behaviour of fund manager a is, in economic terms, to be preferred to the behaviour of fund manager p, our results also provide a note of warning regarding the use of such estimation methods.

References

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

Fang, K.-T.; Kotz, S. and Ng, K.-W. (1990): *Symmetric Multivariate and Related Distributions*, Chapman and Hall, London.

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.

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