Estimating a Hedge Fund Return Model Based on a Small Number of Samples *

Dmitriy Levchenkov † Thomas F. Coleman ‡ Yuying Li §

Do not distribute: October 9, 2008

Abstract

Optimal portfolio selection decisions hinge on the availability of return models which capture important statistical properties of the assets under consideration. In comparison to traditional assets, estimating a return model for hedge funds proves to be much more complex and challenging. Extended factor models, with option returns as additional factors, have been proposed to model index and individual hedge fund returns, see, e.g., Agarwal and Naik (2004), Fung and Hsieh (1997, 2001). Given that typically only a small number of return samples are available, it is very difficult to identify dominant risk exposures in estimation of return models, particularly for individual hedge funds. Assuming a small sample set, we consider a hypothetic market timing investment strategy from a universe of investment assets and use it for evaluating predictive quality of models estimated using different methods. Ordinary least squares, ridge regression, and support vector regression (SVR) methods are compared. We illustrate that, for predicting individual hedge fund returns, the more sophisticated ridge regression and SVR regression methods, which limit generalization error in addition to minimizing empirical error, perform significantly better than simple ordinary least squares methods with a heuristic factor selection. In addition, we compare and contrast the predictive quality of the estimated extended asset-class index factor model with the extended individual asset-based factor models. We illustrate that, using more sophisticated estimation methods, the extended individual-asset based model can perform better than the extended asset-class index based model.

1 Introduction

In recent years, the hedge fund industry has been experiencing a rapid growth. One reason for this growth is the belief that hedge funds may be able to deliver "pure alpha", see e.g., [7]. In addition, hedge funds can be used to diversify risk. Consequently, there has been tremendous interest in including hedge funds in portfolio investment decisions.

Sound hedge fund portfolio selection decisions require analytic tools to qualify and quantify fund return characteristics, similar to traditional assets. To this end, typically a return model, which characterizes the fund return probabilistically, is required. Based on the estimated return model, fund allocation decisions can then be made to achieve desired expected return and risk consideration objectives.

For mutual funds and traditional assets, asset-class index models are typically estimated to explain and predict fund returns. For example, Sharpe describes in [14] a twelve-asset class model for mutual fund returns; these asset classes include bills, bonds, mortgage-related securities, and various stock indices. Using data from 1985–1989, Sharpe has discovered that a substantial portion of the variance in the monthly mutual fund returns is explained by this asset class model. The asset-class index model can then be used for fund style analysis, performance evaluation, and fund selection decisions.

*The authors would like to thank anonymous referees whose comments have improved the presentation of the paper. This work was supported by the National Sciences and Engineering Research Council of Canada and by TD Securities Inc. The views expressed herein are solely from the authors.

†School of Operation Research, Cornell University, Ithaca, NY, 14850, email: dvl7@cornell.edu

‡Faculty of Mathematics, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1, email: tfcoleman@uwaterloo.ca

§David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1, email: yuying@uwaterloo.ca

Similarly, it is desirable to establish a link between hedge fund returns and asset returns and use that information for creating portfolios which include hedge funds. However, while mutual funds typically use long-only and buy-and-hold strategies on standard asset classes in order to achieve a relative return target, hedge funds are typically based on absolute return targets. In addition, hedge fund managers usually employ dynamic trading strategies, possibly trading derivative instruments directly. In contrast to traditional asset investment, a hedge fund manager has almost unlimited freedom in types of investment instruments and types of trades. Strategies implemented by managers are proprietary and currently there is no legal requirement for disclosure of the trading strategies and trading positions. Moreover, hedge fund returns typically have low correlation with the market, see e.g., [5]. In addition, hedge fund performance depends on the skill of the manager, which is hard to model.

There seems to be a consensus that mean-variance portfolio optimization is not appropriate for selecting hedge funds, either in a fund of funds portfolio or in augmenting a traditional investment portfolio. In addition it is not clear what the most appropriate risk measure is when constructing an optimal portfolio, which contains hedge funds. Possible alternative risk measures that have appeared in the current literature include modified VaR, CVaR, and probability of shortfall.

Current market turmoil clearly indicates the importance of accurate evaluation of risk in complex financial instruments. No matter what risk measure one decides to use in portfolio analysis and portfolio selection, one requires a return model which authentically describes important statistical properties of uncertain fund returns. Unfortunately, simple models and simple estimation methods are not sufficient due to the complexity in hedge fund returns. One of the main objectives of this paper is to identify robust methods for suitable return models for hedge funds to accurately evaluate the return and risk in hedge funds.

The unique risk management issues in hedge funds are beginning to attract interest from academic research. In [11] and his subsequent papers on risk management for hedge funds, Lo argues that there is need for a new set of dynamic risk analytics specifically targeted for hedge fund investing. Fung and Hsieh [5] show that, while the behavior of mutual funds can be predicted relatively well by an asset-class index model similar to the Sharpe's model in [14], the same factor model poorly explains index and individual hedge fund return data.

It has been observed that many hedge fund returns resemble option returns, resulting from nonlinear risk exposures, see, e.g., [5], [6], [1]. It seems intuitive that including option returns as variables increases the explanatory power for a hedge fund return model. Indeed, Haugh and Lo [9] show that the optimal continuous asset allocation based utility maximization can be well approximated with a simple buy-and-hold option strategy. Merton [12] shows that market timing can be implemented using options. Browne [2] shows that dynamic shortfall probability minimization trading strategy is equivalent to the replicating strategy for a European digital option. In addition hedge fund managers sometimes trade derivatives directly or implement dynamic trading strategies that resemble option trading strategies. It seems that using option returns as explanatory variables, in addition to asset returns, may lead to a model with better explanatory power. Subsequently we refer to factor models with traditional asset returns, augmented with option returns, as *extended* factor models.

Recently empirical studies of the extended factor models for index and individual hedge funds have been conducted to explain and predict hedge fund returns, see e.g., [5], [6], [1]. For example, in [5], a set of asset-based style (ABS) factors for hedge funds and fund indices is proposed. Both trend following trading and convergence trading are linked to the returns of lookback straddles. Although the lookback option prices are not directly observable from the market, they can be synthesized by observable liquid standard options. In a similar approach, risk factors are constructed in [1] by calculating returns on portfolios that dynamically trade at-the-money and out-of-the money puts and calls on the S&P 500 index.

Most of the current literature, however, has focused on modeling returns of hedge fund indices; estimating individual fund return is much more challenging. In [5], Fung and Hsieh discover that their estimated extended asset-class index model gives poor out-of-sample-R^2 for individual hedge fund returns. The empirical results in [1], based on individual hedge funds in HFR and TASS data, show that out-of-sample-R^2 for hedge fund indices are generally much higher than that for individual hedge funds.

The difficulty of risk exposure estimation for hedge funds has been well recognized in hedge fund literature. The challenges in identifying the dominant risk factors for hedge funds are due partially to the fact that there are typically only a small number of fund return samples and numerous possible market and trading strategies that hedge funds can follow. These difficulties are certainly exacerbated for individual hedge funds. Researchers have attempted to address this problem by using a stepwise ordinary least squares regression procedure either explicitly or implicitly to identify dominant risk factors. For example, this estimation approach is used in [1], [5], and [6].

The classical ordinary least squares regression minimizes the empirical error. When many explanatory variables exist and observations are noisy, the ordinary least squares method is known to be problematic. A stepwise least squares regression procedure is a heuristic procedure which attempts to limit the number of explanatory variables. However, usefulness of an estimated model is measured against its out-of-sample performance. Ridge regression methods and more recent support vector regression (SVR) methods, see, e.g., [15], are more robust methods, which attempt to limit the out-of-sample generalization error, in addition to minimizing empirical error. In particular, a SVR regression method has a theory of uniform convergence and has been shown to give good out-of-sample performance on a wide variety of application problems.

One question, which is of both the theoretical and practical importance, is whether these more sophisticated methods can lead to better return models for individual hedge funds. However, quantifying the improvements on historical data is likely to be problematic because available samples are typically very small, e.g., only 3 years of monthly observations. Even evaluating out-of-sample performance using cross-validation techniques may not be reliable enough.

In this paper, we consider the following approach to evaluate models and estimation methods for obtaining a return model of a hedge fund. We consider a hedge fund example by modeling a generic dynamic trading strategy, i.e., market timing. For our purpose, it is not essential to reproduce all details of the fund operation. Rather it is more important to capture key statistical properties of hedge fund returns identified by previous research. Our approach then allows us to generate data for in-sample fitting and out-of-sample evaluation, on which conventional OLS-based methods, as well as more sophisticated ridge regression and SVR methods can be tested.

Specifically, in our approach, we:

- **Specify Evolution of Asset Returns.** We assume, consistent with the recognized results for traditional assets, that the investment universe consists of assets whose returns follow asset-class index models.

- **Model Hedge Fund Returns.** We assume that a hedge fund manager employs a hypothetic market timing strategy using options on assets in the universe.

- **Specify Return Models to Be Estimated.** We consider extended asset-class index models with asset/index returns and options on assets/indices as factors.

- **Generate In-sample and Out-of Sample Data.** We assume that the return sample size is small; 36 monthly return samples are used in our computations for fitting.

- **Evaluate the Regression Methods.** We apply the regression methods to the in-sample data and evaluate their performance on the out-of-sample data.

This seemingly ideal setup nonetheless reflects some of the major causes of difficulties in hedge fund return model estimation, i.e., small sample sizes and complex strategies in a universe of diverse assets. We ignore, however, potential non-stationary properties of asset returns in this study.

Note that our framework can be easily modified or extended; the above specification serves only as an illustration of the approach. If more specific information about hedge fund behavior and asset universe is available, the described methodology for evaluating performance of possible models and regression techniques can similarly be applied.

One issue worthy of discussion is the usage of returns of individual assets as well as returns of the derivatives on these assets rather than index returns. For individual funds, it seems that a more complex factor model with individual asset returns as well as returns of options on these assets has potential to better explain and forecast returns, compared to an extended asset-class index model with returns of indices and options on indices. By considering returns of individual assets and options, the complexity of the regression function hypothesis space is increased and naturally one is concerned with the additional difficulty in identifying dominant risk exposure. The difficulty of gathering a much larger data set aside, it is not clear which of the *estimated* models is superior in performance. In this paper we investigate whether and when more sophisticated regression methods lead to better performance with the more complex factor models.

Presentation of the paper is organized as follows. Firstly, we describe, in Section 2, a simple one-period model to illustrate that using option returns can potentially improve the predictive power of a return model for a dynamic trading strategy; theoretical analysis of forecasting quality in this simple model is provided in Appendix A. This analysis underscores challenges and provides a benchmark for the subsequent computational results. We proceed to describe regression formulations for ordinary least squares, ridge regression, and support vector regression in Section 3. The assumed asset universe is described in Section 4 and the assumed hedge fund trading strategy is presented in Section 5. Description of experiments carried out to compare different regression techniques is given in Section 6. The results of these experiments are discussed in Section 7. We conclude in Section 8.

2 Predicting a Market Timing Strategy Based on Option Returns

In order to illustrate that using option returns can potentially improve the predictive power of a return model for a dynamic trading strategy, we first consider a simple one period model with the universe consisting of the following two assets:

- Stock with prices S_0 and S_1 at time moments 0 and 1.
- Risk-free asset (bond) with the initial price $B_0 = 1$ and the fixed interest rate r.

In the next sections and Appendix A, we consider more realistic multi-asset and multi-period models.

Denote the rate of return of the stock as $r_1 = S_1/S_0 - 1$. Assume that $E[r_1] = \mu$ and $Var[r_1] = \sigma^2$.

Suppose that, at time moment 0, we start with capital V_0. We can use this capital to buy ξ_0 units of the stock and η_0 units of the bond, forming a portfolio (ξ_0, η_0) with the initial portfolio value $V_0 = S_0\xi_0 + \eta_0$.

At the time moment 1, the value of the portfolio will be $V_1 = S_1\xi_0 + (1+r)\eta_0$ and its rate of return will be

$$\rho = V_1/V_0 - 1 = \frac{(S_1 - S_0)\xi_0 + r\eta_0}{S_0\xi_0 + \eta_0}.$$

Note that the rate of return r is guaranteed if no investment in stock is undertaken, i.e., $\xi_0 = 0$. We will concentrate on excess returns of portfolios. The rate of return of the portfolio exceeding r is given by

$$\rho - r = \frac{S_0\xi_0}{V_0}(r_1 - r).$$

The expected value and variance of the excess return are

$$E[\rho - r] = \frac{S_0\xi_0}{V_0}(\mu - r), \quad Var[\rho - r] = \left(\frac{S_0\xi_0}{V_0}\sigma\right)^2.$$

If $\mu > r$, then investing in the stock (i.e., $\xi_0 > 0$) will yield a positive gain on average. However, high value of ξ_0 also increases $Var[\rho - r]$. This leads to a larger probability of losing a lot of money. In general, the investment decision depends on the risk preferences of the manager.

We now consider a simple market timing strategy. Similar to the work [13] by Merton, we assume that a fund manager is able to predict the sign of the excess asset return $r_1 - r$. In addition, we suppose that the manager attempts to predict the true value of the return r_1 and the error of his prediction is a random noise ϵ. The manager may be able to do that by examining a company's past and current performance. However, the exact nature of the prediction is not relevant here.

More precisely, we assume that the manager's prediction $\tilde{r}_1$ is centered around the true value of r_1 with an additive noise ϵ, i.e.,

$$\tilde{r}_1 = r_1 + \epsilon, \quad E[\epsilon] = 0, \quad Var[\epsilon] = \sigma_\epsilon^2.$$

We refer to the ratio $\frac{1}{\sigma_\epsilon}$ as the skill of the manager since it determines the quality of the prediction.

Using this prediction the manager invests in the stock with a holding position $\tilde{\xi}_0 = \tilde{\xi}_0(\tilde{r}_1)$ and receives return $\tilde{\rho}$:

$$\tilde{\rho} - r = \frac{S_0 \tilde{\xi}_0(\tilde{r}_1)}{V_0}(r_1 - r).$$

We make the following assumptions on the form of $\tilde{\xi}_0(\tilde{r}_1)$. Assume that the fund manager makes investment decisions based on his own prediction. If we assume $\tilde{\xi}_0(\tilde{r}_1) = a\frac{V_0}{S_0}\,\mathrm{sgn}(\tilde{r}_1 - r)$, then the framework is similar to the market timing strategy discussed in [13]. However the magnitude of $\tilde{r}_1$ is not considered in this strategy. The simplest variant which takes the magnitude of $\tilde{r}_1$ into account is $\tilde{\xi}_0(\tilde{r}_1) = a\frac{V_0}{S_0}(\tilde{r}_1 - r),\ a > 0$. This is consistent with the ideas considered in [10], which suggest a curvature in dependence of fund returns on the asset returns. Indeed, under this definition of $\tilde{\xi}_0(\tilde{r}_1)$, the rate of the excess return of the portfolio is

$$\tilde{\rho} - r = a(r_1 - r)(r_1 - r + \epsilon),$$

which is a quadratic function in r_1 for this simple dynamic strategy.

It can be easily shown that the expected excess return is $E[\tilde{\rho} - r] = a(\sigma^2 + (\mu - r)^2)$ and the variance of the excess return is

$$Var[\tilde{\rho} - r] = a^2\left(Var[(r_1 - r)^2] + \sigma_\epsilon^2((\mu - r)^2 + \sigma^2)\right).$$

We can regard the value of a as an indicator of the manager's appetite for risk. A larger value of a produces a larger expected portfolio return; but variance also increases. If two managers have the same a but different skills, the expected returns are equal. However the variance is bigger for the manager with lower asset return predicting abilities, i.e., higher σ_ϵ^2.

We are interested in how well stock return r_1 can describe the portfolio return $\tilde{\rho}$. Consider more generally the problem of using a random variable X to explain a random variable Y. Following the Ordinary Least Squares approach we would like to find function $f(x)$ that minimizes $E[(Y - f(X))^2]$. It can be readily shown that the minimizer is $f(x) = E[Y|X = x]$.

Substituting $f(X)$ by $E[Y|X]$ in the objective function of the minimization problem and using the properties of variance, we have $E[(Y - E[Y|X])^2] = E[Var[Y|X]]$.

If we do not know values of X then all we can say about Y is that it has mean $E[Y]$ and variance $Var[Y]$. However, if it is known that $X = x$, the distribution of Y is best described by $E[Y|X = x]$ and the residual (unexplained) variance is $Var[Y - E[Y|X = x]|X = x] = Var[Y|X = x]$. Thus, we can define the goodness-of-fit variable

$$R^2(X) = \frac{Var[Y] - Var[Y|X]}{Var[Y]} = 1 - \frac{Var[Y|X]}{Var[Y]},$$

which essentially describes the proportion of explained variance of Y for a given value of X.

We can average this quantity over X and consider its expectation

$$\mathbf{E}R^2 = 1 - \frac{E[Var[Y|X]]}{Var[Y]}.$$

In light of this approach, given variables X and Y we will calculate:

- $E[Y|X]$ - the distribution of Y for given X;
- $Var[Y|X]$ - unexplained variance for given X;
- $E[Var[Y|X]]$ - expected unexplained variance, and, finally,
- $\mathbf{E}R^2$ - expected proportion of explained variance.

Typically one wishes to maximize $\mathbf{E}R^2$. For our simple one period model, it is easy to calculate the conditional expectation of $\tilde{\rho} - r$:

$$E[\tilde{\rho} - r|r_1] = a(r_1 - r)(r_1 - r + E[\epsilon]) = a(r_1 - r)^2.$$

We observe that the dependence of $E[\tilde{\rho} - r|r_1]$ on r_1 is non-linear; it is obvious that linear regression on r_1 will not yield good results. However, this non-linear dependence can be easily approximated by a piecewise

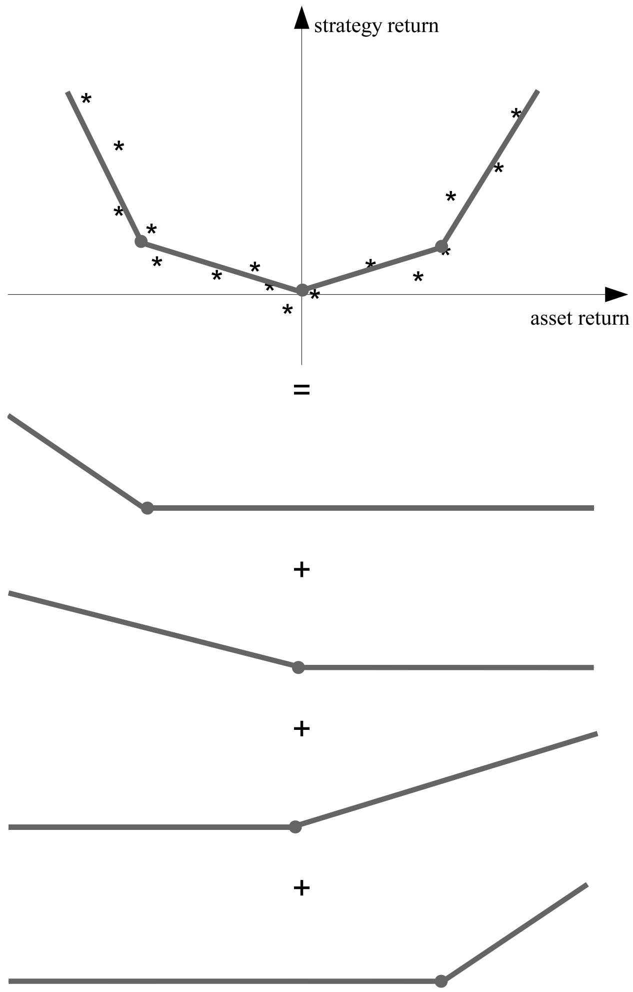


Figure 1: An illustration of how non-linear dependence can be captured by using option returns.

linear function, for example, by a linear combination of option returns, as illustrated in Figure 1. This consideration essentially suggests use of option returns for forecasting this simple market timing strategy returns in the case of an one-period model.

In Appendix A, we present the exact formula for $\mathbf{E}R^2$. In particular, the analytic formula shows that the average goodness-of-fit is proportional to the manager's skill $\frac{1}{\sigma_\epsilon}$. A generalization and analysis of this market timing strategy in a 2-period model is also presented in Appendix A.

3 Regression Techniques for Forecasting Fund Returns

In §2, we have illustrated why using option returns in a regression model can better explain the return of a complex market timing strategy. Now we consider different regression methods for estimating a return model based on a small number of observations.

There are potentially many reasons for poor predictive performance of estimated return models for individual hedge funds. In comparison to traditional asset-class index models, an extended factor model used for hedge funds consists of option returns as additional factors; these factors are necessary when dynamic trading or direct option trading is involved. On the other hand, as complexity of the assumed model increases, one runs into the potential problem of over-fitting. In addition, hedge fund return distributions can be asymmetric and non-normal. Moreover, additional significant explanatory variables may still be missing in a specified factor model and the distribution of hedge fund returns may not even be stationary.

We describe three types of estimation methods and discuss their particular relevance with respect to aforementioned difficulties. We ignore here, however, the potential non-stationarity issue.

Denote $\mathbf{H}$ as the hedge fund return and $\mathbf{X} \in \mathbb{R}^m$ as its explanatory random variables, e.g., asset-class index returns.

Assume that a finite set of return observations $\{(X_i, H_i)\}_{i=1}^n$ is generated from an unknown probability distribution for $(\mathbf{X}, \mathbf{H})$.

Let $F(H|X)$ denote the conditional distribution of hedge fund returns. The expected relationship between return $\mathbf{H}$ and explanatory variables $\mathbf{X}$ is given by the regression, i.e., the conditional expectation

$$f^*(X) = \int H dF(H|X).$$

Typically, estimation of $f^*(X)$ is used to explain or predict the random $\mathbf{H}$. For fund return models, estimation of the regression function is restricted to the linear functional space $\{f(X) : f(X) = \beta^T X + b, \beta \in \mathbb{R}^m\}$.

Let $F(H, X)$ denote the joint distribution of $(\mathbf{H}, \mathbf{X})$. The problem of determining the linear regression $f^*(X)$ is then reduced to minimize a quadratic loss function

$$\text{error}(f) = \int (H - f(X))^2 \, dF(H, X).$$

The minimum of the above quadratic risk functional can be shown to be the closest to the regression $f^*(X)$ in the L_2 metric. Indeed, error(f) measures how well the fund return $\mathbf{H}$ can be described by the linear model $\beta^T \mathbf{X} + b$.

Alternatively, the error functional can be measured by a general loss function $\Lambda(\cdot)$

$$\text{error}(f) = \int \Lambda\left(H - f(X)\right) dF(H, X).$$

Since the probability density is unknown one cannot directly evaluate error(f). One can only compute the empirical risk. The empirical risk corresponding to the observations $\{(X_i, H_i)\}_{i=1}^n$ is

$$\text{error}_{emp}(f) = \frac{1}{n} \sum_{i=1}^{n} \Lambda(H_i - f(X_i)).$$

The ordinary least squares approach minimizes the empirical risk error$_{emp}$ for the quadratic loss function, i.e.,

$$\min_{\beta, b} \sum_{i=1}^{m} (H_i - f(X_i))^2$$

Under the assumption that H is the result of measuring a regression function with normal additive noise, the above empirical risk minimization principal provides the best unbiased estimator of the regression function. In practice, this assumption does not hold.

Huber (1964) developed a theory that finds the best strategy for choosing the loss function using only general information about the noise model; the best minimax strategy regression approximation is obtained with the loss function

$$\Lambda(H, f(X)) = |H - f(X)|$$

if the only information about the density is that it is symmetric and smooth.

Unfortunately, methods based only on empirical minimization do not control out-of-sample generalization error. With the knowledge of a small set of samples, minimizing empirical error can lead to large error(f), which is the measure of the true error. Indeed, empirical minimization often leads to over-fitting and sensitivity to noisy data. For example, when the number of factors n in X exceeds the number of observations m, it is possible to find β that *provides a perfect fit* to the empirical observations. Unfortunately, this does not in general lead to a good out-of-sample performance.

One could try to improve OLS by introducing a *factor selection* process. For example, the forward stepwise feature selection (FSFS) method first chooses one variable out of all available that gives the best approximation to the data. Then the FSFS method fixes this variable and adds a second variable such that these two variables define the best approximation to the data. This process can be repeated by adding additional variables.

More sophisticated regression methods are based on the idea of balancing the empirical and generalization error via use of regularization terms. Ridge regression and support vector regression methods belong to this class of methods, see e.g., [4].

A quadratic (L_2) ridge regression solves

$$\min_{\beta,b} \quad \left(u \cdot \sum_{i=1}^{m} (H_i - \beta^T X_i - b)^2 + ||\beta||_2^2 \right), \tag{1}$$

where $u > 0$ is a regularization parameter. Alternatively, one may consider L_1 ridge regression:

$$\min_{\beta,b} \quad \left(u \cdot \sum_{i=1}^{m} |H_i - \beta^T X_i - b| + ||\beta||_1 \right). \tag{2}$$

If u is large, the empirical error can be made arbitrarily small, and this can lead to over-fitting. As u decreases, the empirical error gets balanced by norms of coefficients; an appropriately chosen u can produce better out-of-sample performance. In general, the most appropriate value for u may need to be determined using a cross-validation technique.

The most recent support vector regression uses the ϵ-insensitive loss function

$$\Lambda(H, \beta^T X + b) = ||H - \beta^T X - b||_\epsilon,$$

where $||v||_\epsilon = \sum_i |v_i|_\epsilon$ and $|z|_\epsilon = \begin{cases} 0 & \text{if } |z| \le \epsilon, \\ |z| - \epsilon & \text{otherwise.} \end{cases}$. In a support vector regression using the 2-norm (**SVR2**), one solves the following problem

$$\min_{\beta,b} \quad \left(u \cdot ||H - \beta^T X - b||_\epsilon + ||\beta||_2^2 \right). \tag{3}$$

Using the 1-norm, a linear programming Support Vector Regression variation **SVR1** solves:

$$\min_{\beta,b} \quad \left(u \cdot ||H - \beta^T X - b||_\epsilon + ||\beta||_1 \right). \tag{4}$$

The parameters ϵ and u are generally determined by cross-validation.

Notice that SVR methods differ from corresponding ridge regressions only in the presence of the ϵ-tolerant residual term. From a practical point of view, for SVR methods, one has to find a value not only for u, but also for ϵ. What benefits do SVR methods have that may justify this additional complexity? We illustrate the difference between these two types of methods by performing a simple example.

It is easy to notice that L_1 ridge regression is the SVR1 with ϵ set to 0. We concentrate on comparisons between these two methods. Let us assume that $H = \beta X + \epsilon$ where the dimension of X is $m = 50$. In addition, X are normal with mean 0 and variance 1, and pairwise correlation between any two factors is 0.8. Assume that $\beta = (1, 0, ..., 0)^T$, i.e., H depends only on the first factor, and the standard deviation of the noise ϵ is $\sigma = 0.8$.

Suppose that $n = 36$ observations are given for H and R. We now compute estimates of β by L_1 Ridge Regression with parameter $u \in [0.01, 1]$, and SVR1 with parameters $u \in [0.01, 1]$, $\epsilon \in [0.1, 2]$. We then compute the sample mean of out-of-sample R^2 based on 50 out-of-sample scenarios; the sample mean for each of the methods is graphed in Figure 2.

We observe that for L_1 Ridge Regression the out-of-sample performance measure peaks at u approximately equal to 0.15, and is low for high and low values of u. Similar behavior can be observed for SVR1 method with small $\epsilon \approx 0.1$, while for $\epsilon \approx 0.7$ the performance measure peaks at approximately 0.3 and does not rapidly decline when u becomes larger. However, the most important observation is that the heights of

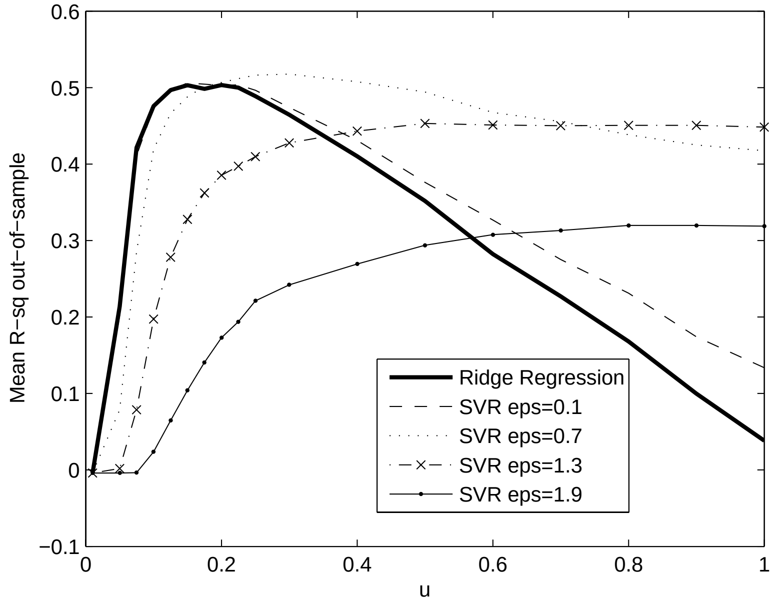


Figure 2: The simple experiment: the mean R^2 out-of-sample for L_1 Ridge Regression and SVR1.

the peaks are nearly the same, meaning that *using the optimal parameter choices* both methods will give roughly the same out-of-sample performance (at least, for this example).

In practice, values for u have to be determined from the data sample or using other considerations. It is likely that the chosen values will differ from the optimal ones. Notice that using a wrong value for u in the Ridge Regression might severely impact the out-of-sample performance, while for SVR any value above certain threshold will result in satisfactory performance. Moreover, reliable heuristic formulae for threshold values of u for a given value of ϵ are available (see, for example, [3]). This brings us to the following observation: SVR seems to be more robust than ridge regression in the sense that it is easier to choose parameters that result in good out-of-sample performance.

If we continue to increase ϵ, the residual term will play less and less role, reducing the overall fit of the model. This explains the behavior of out-of-sample performance for SVR1 with ϵ values 1.3 and 1.9. Thus choosing ϵ becomes the most important task for SVR methods. But the choice of ϵ can be achieved by cross-validation.

Vapnik illustrates in [16] that SV regression is advantageous relative to OLS and the forward stepwise feature selection (FSFS) method. One of our goals is to quantify the difference for the hedge fund return model estimation problem. We also would like to investigate if SV regressions are advantageous over ridge regressions in estimating a complex fund return model based on a few data samples.

4 A Factor Model for Asset Returns

Before modeling hedge fund returns we need to describe an asset universe. In this section we consider a standard factor model for asset returns and introduce related notation. We allow several market sectors with a non-trivial correlation structure. Without loss of generality we assume the risk-free rate to be 0%, as it has little effect on the performance of regression methods. However, this assumption can be easily relaxed.

We assume that a collection of A assets is divided into G classes $\{\mathcal{G}_g\}_{g=1,\ldots,G}$, which defines a partition of the set $\{1,\ldots,A\}$. These classes may correspond to different sectors of the market such as bonds, equities, etc.

For each asset class $g \in \{1,\ldots,G\}$, daily expected return and volatility are specified: $\tilde{r}_g$ and $\tilde{\sigma}_g$. In addition, we assume that the covariance matrix for the G factors is $C \in \mathbb{R}^{G\times G}$. More specifically, risk factors $f_t \in \mathbb{R}^G$ satisfy the following:

$$f_t = \Lambda \xi_t^{\mathbf{g}},$$

where Λ is Cholesky decomposition of C: $C = \Lambda\Lambda^T$ and $\xi_t^{\mathbf{g}}$ are independent and distributed as $\mathcal{N}(0_G, I_{G\times G})$.

We consider daily returns, which are denoted as r_t^a for asset a and day $t \in \{1, \dots, T\}$. For asset a in asset class g, daily returns are

$$r_t^a = \bar{r}_a + B_a f_t + \epsilon_{t,a},$$

where B_a is the exposure vector for asset a, $\epsilon_{t,a}$ are independent normals with mean 0 and variance $\breve{\sigma}_a$. Specification of B_a's is discussed in Appendix B.

Suppose we observe returns of A assets for N months. Denote monthly return for asset $a \in \{1, \dots, A\}$ in month $i \in \{1, \dots, N\}$ as R_i^a, where $T = 30N$ and T is the number of daily return observations.

Monthly and daily returns are assumed to be linked by the following formula:

$$1 + R_i^a = \prod_{t=1+30(i-1)}^{30i} (1 + r_t^a),$$

For each asset class $g \in \{1, \dots, G\}$, we define index returns as an average of returns of assets in the class:

$$\Phi_i^g = \frac{1}{|\mathcal{G}_g|} \sum_{a \in \mathcal{G}_g} R_i^a.$$

These index returns may be used as explanatory variables in the regression.

5 A Market Timing Fund

In order to evaluate performance of regression methods for the problem of forecasting hedge fund returns, we need to be able to generate data with statistical properties similar to ones for hedge fund returns. We attempt to achieve that by considering a hedge fund manager who employs a market-timing strategy in the described asset universe. Monthly returns of this fund are stationary and an extended factor model explains a significant portion (but not in its entirety) of the fund return, even when a set of option returns is included as explanatory variables. Note that the model presented is intentionally simple for illustrative purposes.

We assume that the fund manager invests in a fixed subset of assets which are selected from the investment universe described in §4. For each asset a the manager makes an investment decision based on his prediction of the next period asset return R_t^a.

We assume that the manager's prediction for month t has the following form:

$$\tilde{R}_t^a = R_t^a + \frac{\sigma^a}{\sqrt{\varsigma}} \upsilon_t.$$

This is an unbiased prediction of the true return with error $\frac{\sigma^a}{\sqrt{\varsigma}}\upsilon_t$, where σ^a is the standard deviation of asset returns R_t^a, and υ_t is a standard normal random variable independent[1] of the asset return R_t^a. The variance of the error term is $\frac{(\sigma^a)^2}{\varsigma}$; ς plays the role of manager's skill since it controls the precision of the prediction.

The skill parameter ς reflects the manager's capability to use additional information in order to predict the future asset monthly returns. To achieve financial gain from such a prediction the manager can invest in a call option if the asset return is predicted to be positive, and in a put option if negative.

It is reasonable to assume that the amount invested in either call or put option depends on the magnitude of the predicted asset return. We assume that the proportion invested in either put or call grows linearly with the absolute value of the predicted asset return until reaching 1 at value $\frac{\sigma^a}{\sqrt{\varsigma}}$. This way the manager will not face a large loss if the signs of the predicted and true values are different.

[1] To be precise, $\xi_t^{\mathbf{g}}$ and error terms $\epsilon_{t,a}$ and υ_t are independent as a collection of random variables.

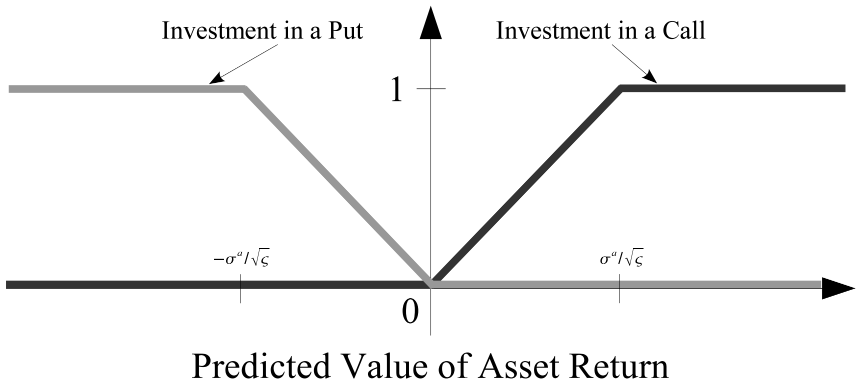


Figure 3: Hedge fund manager investment decision for a given predicted value of asset return.

The described market timing strategy is depicted in Figure 3. Mathematically, assuming w^a is a weight vector influencing amount to be invested in derivatives of asset a, then at time $t-1$ the proportion of the capital to be invested in at-the-money call option on that asset will be

$$^{c}w_{t-1}^{a} = w^a \begin{cases} 1, & \tilde{R}_t^a > \frac{\sigma^a}{\sqrt{\varsigma}} \\ \tilde{R}_t^a \frac{\sqrt{\varsigma}}{\sigma^a}, & 0 \le \tilde{R}_t^a \le \frac{\sigma^a}{\sqrt{\varsigma}} \\ 0, & \tilde{R}_t^a < 0 \end{cases} .$$

Similarly, for the put option the proportion invested is

$$^{p}w_{t-1}^{a} = w^a \begin{cases} 1, & \tilde{R}_t^a < -\frac{\sigma^a}{\sqrt{\varsigma}} \\ -\tilde{R}_t^a \frac{\sqrt{\varsigma}}{\sigma^a}, & -\frac{\sigma^a}{\sqrt{\varsigma}} \le \tilde{R}_t^a \le 0 \\ 0, & \tilde{R}_t^a > 0 \end{cases} .$$

The rest of the money is invested in a risk-free asset: $^{rf}w_{t-1}^{a} = w^a - {}^{c}w_{t-1}^{a} - {}^{p}w_{t-1}^{a}$.

We allow the fund manager to trade multiple derivative assets. For each fund, the number of underlying assets are fixed and chosen randomly a priori. The estimation problems become harder when more assets are being used since it is more difficult for regression methods to identify locations of the investment strategy. On the other hand, more factors become relevant for the problem. For simplicity we assume that the investment weights w^a do not change with time.

Formally, each hedge fund is assigned:

- a set $\mathcal{A}$ of underlying assets

$$\mathcal{A} = \{a \quad | \quad a \in \{1, \ldots, A\} \text{ is traded by the hedge fund } \}.$$

  The hedge fund is assumed to trade options on each of these assets.

- a static weight vector $w \in \{w : w \ge 0, \quad w^a = 0 \text{ if } a \notin \mathcal{A}, \quad \sum_a w^a \le 1\}^2$ for options on corresponding assets (either calls or puts depending on predictions);

- a skill parameter $\varsigma > 0$.

For month t and asset a, P_t^a denotes one-month return of an at-the-money European put option on asset $a \in \mathcal{A}$ with one month maturity. Similarly C_t^a denotes monthly return of an at-the-money call on asset $a \in \mathcal{A}$ with one month time to maturity.

[2] The sum of w^a's can be less than 1 to allow investment in the risk-free asset.

Then the overall return of the corresponding hedge fund is:

$$H_t = \sum_{a \in \mathcal{A}} \left({}^c w_{t-1}^a C_t^a + {}^p w_{t-1}^a P_t^a \right).$$

6 Experimental Setup

We now describe parameters, settings, and evaluation criteria used in the experiments.

Asset Universe

We assume that there are $G = 8$ asset classes in the universe. For factor model calibration, we use historical data for 8 traditional asset classes considered in [5]: MSCI U.S. equities, MSCI non-U.S. equities, IFC emerging market equities, JP Morgan U.S. government bonds, JP Morgan non-U.S. government bonds, 1-month Eurodollar deposit, gold, and Federal Reserves Trade weighted dollar index. The vector of expected daily returns $\{\tilde{r}_g\}_{g=1,\ldots,G}$, volatilities $\{\tilde{\sigma}_g\}_{g=1,\ldots,G}$, and the covariance matrix C were estimated using this data, see also Appendix C.

In each asset class, there are 10 assets with each asset return modeled according to the asset-class factor model described in §4. This gives rise to a total of $A = 80$ assets in the universe.

For simplicity, asset daily expected returns are set to that of the corresponding asset class, i.e., $\bar{r}_a = \tilde{r}_g$ for all $a \in \mathcal{G}_g$, $g = 1, \ldots G$. Similarly, the standard deviation of noise $\epsilon_{t,a}$ is set to $\breve{\sigma}_a = \tilde{\sigma}_g/4$ for all $a \in \mathcal{G}_g$, $g = 1, \ldots G$.

We generate a total of 108 monthly asset returns, out of which 36 are used to generate data for estimation, and 72 are used for out-of-sample performance evaluation.

Hedge Funds

We consider a total of 60 randomly specified market timing hedge funds as described in §5. Each hedge fund implements a monthly trading strategy on a set of randomly selected assets $\mathcal{A}$ in the asset universe; we compare the performance as the number of assets traded by each fund varies from 1 to 8. The investment proportions w^a are fixed at 5% for assets $a \in \mathcal{A}$. The inverse of the skill for each hedge fund, $\frac{1}{\varsigma}$, is selected uniformly in the interval $[0.3, 3]$.

Factor Models

For our hypothetic market, we consider asset-class index models **model {i}** and extended asset-class index models **model {i+oi}** similar to [1] and [5]:

- **model {i}**: the explanatory variables $\mathbf{X}$ consist of monthly index returns $\Phi_t^{\mathbf{g}}$. Here the regression function has the form $\beta^T \Phi_t^{\mathbf{g}}$.
- **model {i+oi}**: the explanatory variables $\mathbf{X}$ consist of monthly index returns $\Phi_t^{\mathbf{g}}$ *as well as* European option returns on indices $g = 1, \ldots, G$ with one month time to expiry and a fixed set of strike prices. The regression function in this case is assumed to be of the form $(\beta^{\mathbf{g}})^T \Phi_t^{\mathbf{g}} + (\beta^{\mathbf{p}})^T \tilde{O}_t^{\mathbf{g}}$, where $\tilde{O}_t^{\mathbf{g}}$ denotes the returns of options on indices in month t.

To stress-test estimation methods and investigate the potential benefits of more complex models, we also apply estimation methods to asset-based models below:

- **model {a}**: the explanatory variables $\mathbf{X}$ consist of monthly returns $R_t^{\mathbf{a}}$ of all assets in the universe. In this case, the hedge fund return is predicted using the regression function $\beta^T R_t^{\mathbf{a}}$.
- **model {a+oa}**: the explanatory variables $\mathbf{X}$ consist of monthly asset returns $R_t^{\mathbf{a}}$ *as well as* European option returns on assets, $a = 1, \ldots, A$, with one month to expiry and a fixed set of strike prices. The regression function in this case is assumed to be of the form $(\beta^{\mathbf{a}})^T R_t^{\mathbf{a}} + (\beta^{\mathbf{p}})^T O_t^{\mathbf{a}}$, where $O_t^{\mathbf{a}}$ denotes the returns of corresponding options.

Table 1: Number of Explanatory Variables in Different Categories.

	model {a}	model {i}	model {a+oa}	model {i+oi}
Assets/Indices	80	8	80	8
Options	0	0	80×8	8×8
Total	80	8	640	64

Number of explanatory variables in different categories are presented in this table. The following settings are used: the total number of assets is 80 and the total number of asset classes is 8. A total of 8 strike prices are used to define option-based explanatory variables.

The difference between **model {a+oa}** and **model {i+oi}** is that **model {a+oa}** contains option returns on *individual* assets, rather than indices. Note that, for each model considered, hedge fund return H_t can not be fully described by the explanatory variables specified.

Model {a} and **model {i}** correspond to asset class models used for predicting returns of mutual funds. However, it was found that these factors alone are often unable to describe behavior of hedge fund returns. We will verify that this remains to be the case here.

Since hedge funds considered in our setup use option-based strategies it is reasonable to anticipate better results for extended factor models **model {a+oa}** or **model {i+oi}**. Using indices is more practical because of data size and data quality. Indeed, as we can see in Table 1, **model {a+oa}** has more than 600 explanatory variables while **model {i+oi}** has 64. However, using indices and their derivatives means sacrificing some information about traded assets. One of our goals is to assess if this sacrifice is significant or not.

We consider option returns for different strike prices to construct a set of option-based variables for each asset or index. In our setting, we experimented with various choices of strikes and found that 8 evenly-spaced strike prices in the range ± 3 standard deviations from the mean provide an adequate set of variables for our problems.

It is also important to scale option-based factors. Scaling seems rather important for SVR methods. While for asset returns traditional scaling can be used (for example, dividing them by their empirical standard deviation), for option-based variables the situation is quite different.

To obtain an option return one divides the option payoff by the initial option price, which could be calculated by the standard Black-Scholes formula. However, this can be computationally unstable when the option price is small, e.g., the option is deeply out-of-the money. To avoid this problem, for a given strike price we consider either a put or a call option to make sure it is in-the-money. Theoretically, it does not matter whether to use put or call due to put-call parity for European options, but we found that this technique helps in practice.

Regressions and Cross Validation

In ridge regressions and support vector regressions, in addition to minimizing the empirical error, one attempts to limit out-of-sample errors by simplifying the model via minimizing additional term $||\beta||$. The performance of this type of method depends on parameters such as u and ϵ which balance minimizing empirical error versus minimizing out-of-sample error. For example, variables that do not appear to be significant can be suppressed by the term $||\beta||$. For ridge regressions and support vector regressions, a k-fold cross validation method can be used to determine the parameter u and any other parameters. In a k-fold cross validation, the available data is divided evenly into k distinct subsets. For each set of parameter values, a regression estimation is performed k times during which $(k-1)$ of the subsets are used as data for regression estimation and the remaining set is used for the performance evaluation. The set of parameter values which leads to the best performance measure is then selected and the entire data set is then used for regression under the selected set of parameter values. Note that this type of cross validation can become computationally expensive.

A heuristic technique can also be applied to an ordinary least squares method for model parsimony consideration to improve the performance of the estimated model. To incorporate this for an OLS method, we first implement the Bayesian information criterion (BIC), which penalizes models with additional parameters. This identifies important variables which contribute to the explanatory power for a fund return. Since there can be many correlated variables (e.g., option returns with different strikes) in models considered in this work, we use, in addition to BIC selection, a k-fold cross validation technique to determine the number of explanatory variables for the final regression estimation.

Specifically, we compare the following regression estimation methods:

- **CV-FS-OLS**: OLS regression with the combination of the BIC criterion and k-fold cross validation for factor selection. Once the subset Z of factors X is chosen, the estimate is obtained by solving the following minimization problem:

$$\min_{\tilde{\beta}} ||y - Z\tilde{\beta}||_2^2$$

  Though the CV-FS-OLS method is easier to implement, it has several potential pitfalls. One of them is that the factor selection, as well as the minimization problem, is sensitive to noise in the data. This may lead to omitting relevant variables and including irrelevant ones, consequently reducing out-of-sample performance.

- **CV-L_2-Ridge, CV-L_1-Ridge, CV-SVR$_2$, CV-SVR$_1$**: L_2-, L_1-ridge regressions, and L_2-, L_1-SVRs (described in (1)-(4) in §3) with k-fold cross-validation.

  Note that **CV-L_1-Ridge** is a particular case of **CV-SVR$_1$** with $\epsilon = 0$; thus, the performance of **CV-SVR$_1$** is no worse than that of **CV-L_1 Ridge**, provided that the value for ϵ is chosen appropriately.

Evaluation

Our objective is to compare different models and different regression methods in their capabilities to predict fund returns based on a small number of samples. To achieve this, for each model considered, we carry out performance evaluation for each of the estimation methods on a set of market timing funds as described in §5.

Specifically, we evaluate regression methods **CV-FS-OLS, CV-L_2-Ridge, CV-L_1-Ridge, CV-SVR$_1$, CV-SVR$_2$** for the following models:

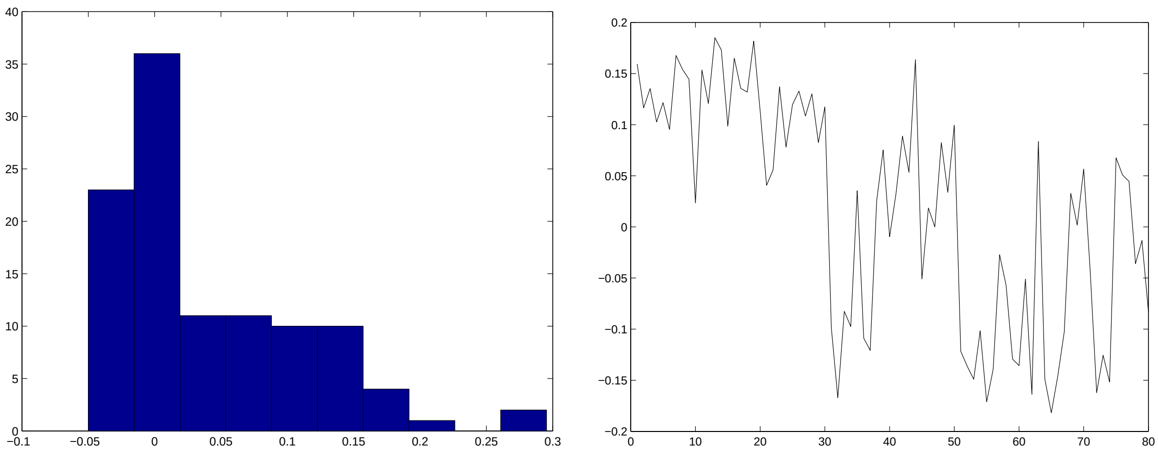
- Experiment **E1: model {a}**.
- Experiment **E2: model {i}**.
- Experiment **E3: model {a+oa}**.
- Experiment **E4: model {i+oi}**.

Experiments **E1** and **E2** are designed to investigate capability of the traditional factor models, which use asset or index returns alone, in predicting our market timing fund returns. Experiment **E3** and **E4** include additional option returns as explanatory variables. Comparison between Experiment **E3** and Experiment **E4** allows us to determine whether additional advantages can be achieved by including options on assets or indices for predicting our hedge fund returns.

As in previous studies, we are mainly concerned with out-of-sample performance of the models. We choose to use out-of-sample R^2 to measure the predictive capability:

$$R^2_{\text{out}} = 1 - \frac{\text{Mean Squared Prediction Error}}{\text{Hedge Fund Return Variance}}.$$

Because returns for different hedge funds have different variances, it is more appropriate to report here the out-of-sample R^2 rather than the mean squared error. The out-of-sample performance measure R^2_{out} is computed using 72 monthly return observations, as mentioned in Section 4.



a) b)

a) Histogram of monthly returns for a sample hedge fund (there is a single underlying asset,#34, for this fund). Distribution does not seem to be normal.
b) Correlations between monthly returns of a sample hedge fund and each asset return. Correlations are not high in absolute values.

Figure 4: Properties of a sample hedge fund return

One could also be interested in whether estimated coefficients uncover investment locations of hedge funds. This question is very challenging since only a very small sample return set is given. This challenge can be addressed in further research.

7 Computational Results

Before presenting results for experiments **E1-E4**, we first examine properties of the generated hedge fund returns. While the asset returns specified in §4 have normal distribution, this is not the case for the hedge fund returns. A sample histogram of the return of a market timing fund as described in §4 is presented in Figure 7 a). It is easily observed that the distribution is skewed and has a longer right tail than one would expect from a normal distribution. This characteristic is consistent with the available data on hedge fund returns [8].

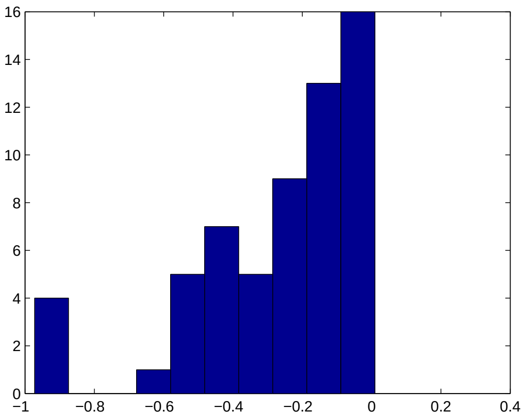
We can also plot correlations between the hedge fund returns and the asset returns, see Figure 7 b). We can see that the correlations are not high (in absolute values). In addition, this example illustrates that the correlation cannot be used in general to identify which asset the fund invests in; this particular hedge fund trades only options on a single asset.

Experiment 1 (E1) - regressing on asset returns only.

We study the R^2-out-of-sample for each of the regression techniques. As expected, using only asset returns as explanatory variables does not provide a good quality of forecast, regardless of the regression method. In Figure 5, we report performance in the case that there is a single underlying asset for hedge funds trading.

The distribution of R^2-out-of-sample for the OLS (CV-FS-OLS) method for all 60 hedge funds is displayed on the left subplot. The mean out-of-sample performance for other methods is summarized on the right panel of Figure 5. We observe that the OLS tends to overfit, which results in substantially negative values for R^2_{out}. On the contrary, regularization methods seem to correctly identify the lack of dependence between hedge fund returns and regressors, and produce only slightly negative values for R^2_{out}.

The performance remains roughly the same when hedge funds trade multiple assets, for all methods.



Distribution of R^2_{out} for CV-FS-OLS regression

Mean of R^2_{out}	
CV-FS-OLS	-0.27
CV-L_2-Ridge	-0.11
CV-L_1-Ridge	-0.09
CV-SVR$_2$	-0.07
CV-SVR$_1$	-0.06

Figure 5: Experiment 1 - regressing on asset returns when there is a single underlying asset for each hedge fund. Reported out-of-sample R^2 shows that asset returns alone are not able to explain hedge fund returns.

Experiment 2 (E2) - regressing on index returns only.

Similar observations can be made if index returns are used in the regressions. However, though the indices provide even less information than assets themselves, the reduced number of factors makes it easier not to choose irrelevant ones and, thus, reduce the number of outliers. The results are presented in Figure 6.

Experiment 3 (E3) - regressing on asset returns and option returns.

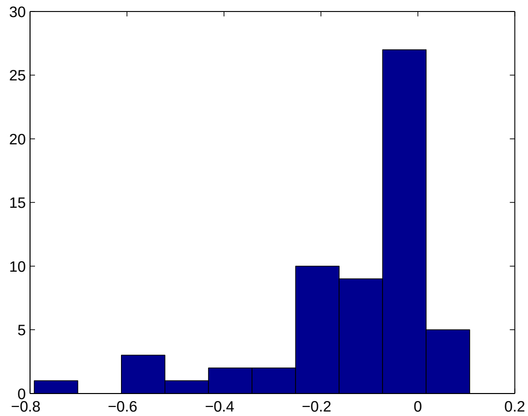
Out of the four computational experiments considered in this paper, this is perhaps the most important one since option returns for individual assets are expected to carry relevant information about fund returns in our setup. We report the mean of R^2-out-of-sample for OLS with factor selection, cross-validated ridge and support vector regressions in Figure 7. Note that comparisons between different regression methods are made when the number of assets invested equals to 1, 2, 4 and 8 respectively.

Firstly, performance of CV-FS-OLS is good only if the number of traded assets is small, i.e., 1 and 2. In all other cases it is clearly dominated by the other methods. This means that a practitioner using OLS with factor selection may conclude that the model can not be fitted well while it might not be the case.

Secondly, CV-L_2-Ridge works well, especially when the number of traded assets is large. This suggests that we can expect the L_2-Ridge regression to perform well when the parameter u is chosen correctly to prevent overfitting, since its residual term and the performance measure R^2_{out} are both based on the sum of squared residuals. It is important to note, however, that this residual term is not robust and in some cases, due to inability to find an appropriate value for u or bad scaling of the factors, the performance may be degraded. However, it does not seem to be a problem for our case.

CV-SVR$_1$ posts consistently good values of R^2_{out}, while CV-SVR$_2$ does not work as well in the case when the number of traded assets is small. This can be explained by the CV-SVR$_2$ regularization term based on the 2-norm of the coefficients, which favors larger number of non-zero coefficients. However, this is counterproductive in the case when the number of traded assets is small. On the contrary, the 1-norm regularization term in L_1-ridge regression typically produces only a few non-zero coefficients in the solution, which helps to achieve superior performance in case when only a few assets are traded by hedge funds.

We can summarize our recommendations as follows. While it is not advisable to use CV-FS-OLS within this framework, ridge and support vector regressions seem to achieve reasonable performance. In many cases, L_2-ridge regression produces good results. However, if robustness is very important, using support vector regressions may be preferred. In this case, if it is believed that the fund trades many assets or derivatives,



Distribution of R^2_{out} for CV-FS-OLS regression

Mean of R^2_{out}	
CV-FS-OLS	-0.13
CV-L_2-Ridge	-0.05
CV-L_1-Ridge	-0.04
CV-SVR$_2$	-0.06
CV-SVR$_1$	-0.04

Figure 6: Experiment 2 - regressing on index returns when there is a single underlying asset for hedge funds. Reported out-of-sample R^2 shows that index returns alone are not able to explain hedge fund returns.

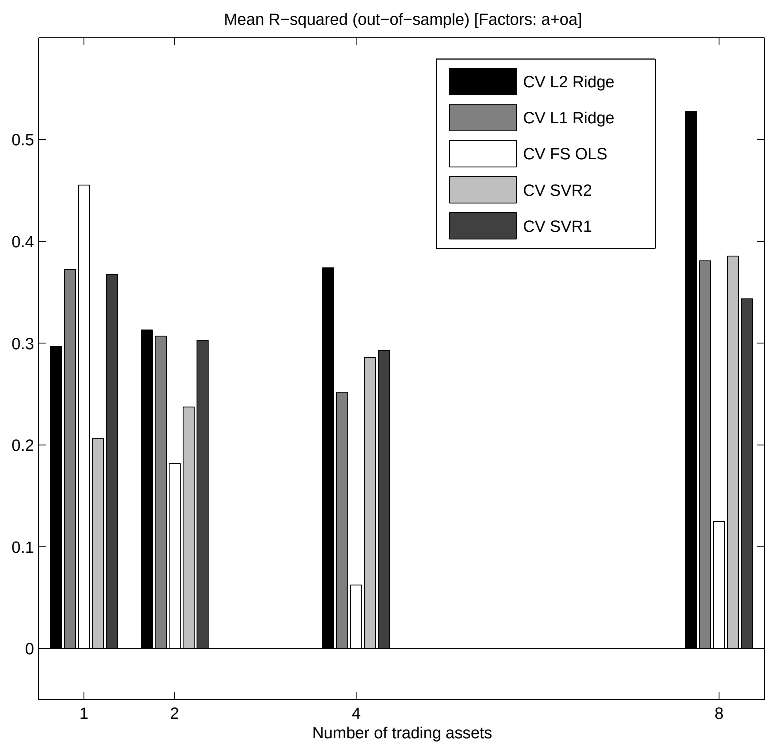


Figure 7: Experiment 3 - regressing on asset returns and option returns on assets. Out-of-sample performance (Mean of R^2_{out}) of cross-validated regression methods for various number of assets traded by hedge funds.

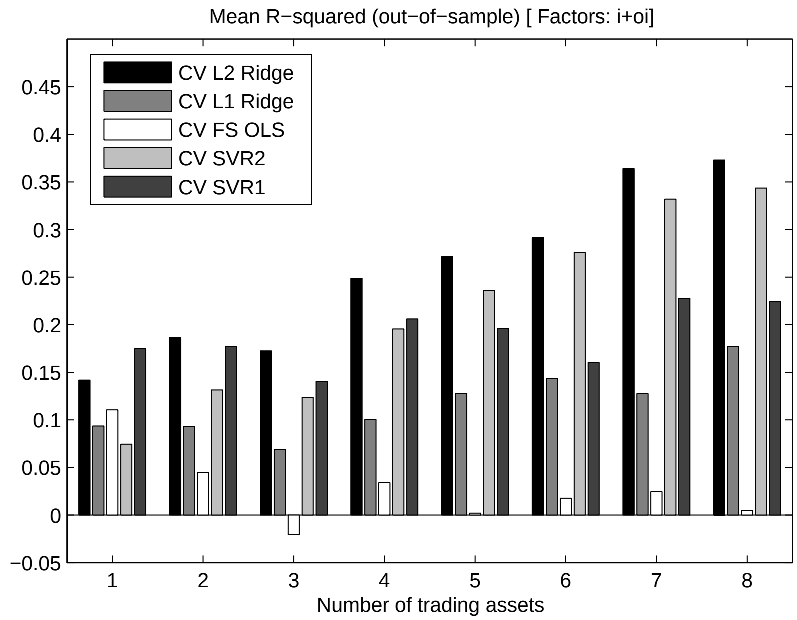


Figure 8: Experiment 4 - regressing on index returns and option returns on indices. Out-of-sample performance (Mean of R^2_{out}) of cross-validated regression methods for various number of assets traded by hedge funds.

2-norm SVR can be used, and 1-norm SVR otherwise.

Experiment 4 (E4) - regressing on index returns and index option returns.

This experiment allows us to answer the following question: what would happen if we use indices and index derivatives rather than individual assets and asset derivatives themselves? On the one hand, indices contain less information than assets; on the other hand, the number of factors is reduced, which may improve ability of the methods to select relevant ones.

Ignoring practical difficulty of data collecting, there is clear evidence from our experiments that the first argument overweighs the second in most of the cases, see Figure 8. The performance drops by about 25% for ridge and SV regressions, while CV-FS-OLS posts mediocre results when the number of traded assets is high. This illustrates that a model including individual asset returns and returns of the derivatives has potential to outperform the index based model.

For ridge and SV regressions, the trends that we observed in Experiment 3 remain in this case. CV-L_2-Ridge posts good results overall, performance of CV-L_1-Ridge is consistent, but lags behind other methods. CV-SVR$_2$ lags behind CV-SVR$_1$ when the number of traded assets is small, but dominates it for larger numbers of traded assets.

8 Conclusions

Due to the tremendous growth of the hedge fund industry, there is urgent need to consider the problem of optimally constructing and analyzing portfolios of hedge funds, as in a fund of funds, or the problem of optimally augmenting an existing portfolio of classical assets with hedge funds. It is not clear what risk measures are most appropriate when analyzing portfolios consisting of hedge funds. In addition, no matter what risk measure one decides to use in portfolio analysis and portfolio selection, one requires availability of a return model which authentically describes important statistical properties of uncertain fund returns. Identifying suitable return models and robust methods for estimating such return models for hedge funds becomes essential. Unfortunately, simple models and simple estimation methods are no longer sufficient.

In this paper we have considered four different models for predicting a hedge fund return, model of returns of all the assets (**model {a}**), model of index returns (**model {i}**), model of all the assets and options on the assets (**model {a+oa}**), and model of returns of indices and returns of options on the indices (**model {i+oi}**).

We investigate five types of regression methods with cross-validation, the classical OLS type method (CV-FS-OLS), the L_2 ridge regression (CV-L_2-Ridge), the L_1 ridge regression (CV-L_1-Ridge), and, finally, 1- and 2-norm versions of the Support Vector Regression (CV-SVR$_1$ and CV-SVR$_2$).

We compare the ability of different models and regression estimation methods in predicting return of a hypothetic fund with an option based market timing dynamic strategy using only a small number of samples. Although the trading strategy considered is artificial, it embodies some of the complex characteristics of hedge fund returns.

From our computational investigation, we observe that, while Ordinary Least Squares (OLS) with factor selection is a highly popular regression technique, it performs poorly in the complex environment of hedge fund returns. It is prone to over fitting if many explanatory variables are included in the regression; factor selection may sacrifice information, which is not desirable, especially when only a small number of data samples are available. Thus, using methods that have a flexible way of dealing with a large number of explanatory variables but at the same time can prevent over fitting is crucial. Support vector regressions (SVRs) are examples of such methods. These methods balance the trade-off between in-sample fitting and out-of-sample performance. In this paper we have considered two SVR methods, as well as their simpler variants - ridge regressions. It turns out that all SVR-like regressions are good alternatives and in most cases dominate the OLS with factor selection. SVR methods analyze all explanatory variables in one step simultaneously; choosing control parameters for the methods does not seem to be a problem. In addition, it is very important that these methods are able to work with individual asset data (rather than the data on indices), potentially increasing the forecasting accuracy. This observation opens a door for examining performance and fine-tuning of more sophisticated SVR methods for estimating models for hedge fund returns.

Directions for further research include analyzing a larger universe, e.g., thousands of assets, with industries as factors, considering more complicated models for hedge fund returns (perhaps, specific to a hedge fund type), and explaining investment locations of hedge funds based on estimated coefficients.

A Bounds on Quality of Forecasting for the Simple Market Timing Strategy

The purpose of this section is to show how bounds on the quality of the forecast can be derived for the simple one-period market timing strategy described in Section 2. First, we examine a one-period framework. Second, we consider extensions of the model, such as a two-period model and a model with unknown trading times.

The precision of the forecast can be evaluated by calculating the conditional variance of $\tilde{\rho} - r$:

$$Var[\tilde{\rho} - r|r_1] = E[(\tilde{\rho} - r)^2|r_1] - (E[\tilde{\rho} - r|r_1])^2 = a^2\sigma_\epsilon^2(r_1 - r)^2$$

and then averaging it over r_1:

$$E[Var[\tilde{\rho} - r|r_1]] = a^2\sigma_\epsilon^2(\sigma^2 + (\mu - r)^2)$$

The average proportion of explained variance (goodness-of-fit) will then be

$$\mathbf{E}R^2_{(1)} = 1 - \frac{E[Var[\tilde{\rho} - r|r_1]]}{Var[\tilde{\rho} - r]} = \frac{1}{1 + \frac{\sigma_\epsilon^2(\sigma^2 + (\mu - r)^2)}{Var[(r_1 - r)^2]}}$$

We observe that the average goodness-of-fit is proportional to the manager's skill ($\frac{1}{\sigma_\epsilon}$). Thus we can expect *good forecast precision for a skillful manager.* The expression also includes parameters of distribution of r_1. For example, assuming $r_1 \sim \mathcal{N}(\mu, \sigma^2)$, the expression becomes

$$\mathbf{E}R^2_{(1)} = \frac{1}{1 + \frac{\sigma_\epsilon^2(\sigma^2 + (\mu - r)^2)}{2\sigma^2(\sigma^2 + 2(\mu - r)^2)}} \geq \frac{1}{1 + \frac{\sigma_\epsilon^2}{2\sigma^2}}.$$

Thus, for the normal case we can obtain a convenient lower bound that depends on the ratio of σ_ϵ^2 and σ^2.

The one-period model showed what we can expect in terms of the forecast quality if the trading is done at the same frequency as the reporting of fund returns. However, it is not a practical assumption, since there could be several trades during the period of reporting, e.g., a month. Thus, it is necessary to consider multi-period models. Here we briefly discuss results for a two-period model and its variants.

In a two-period model trading occurs at the beginning of each period. We are interested in explaining the overall (over two periods) hedge fund return. Assuming equal period lengths, we can consider two variants:

A. both asset returns for periods 1 and 2 are known and could be used to explain the overall hedge fund return.

B. only the overall asset return is known (e.g., the trading is done twice a month, but only monthly returns are available for the asset) and can be used to explain the overall hedge fund return.

We show that for case A the bound on forecasting quality coincides with the one for the one-period model. However, for case B, assuming that asset returns are approximately normal, the quality of forecast of the market timing strategy is bounded by

$$\mathbf{E}R^2_{(2B)} \geq \frac{1/2}{1 + \frac{\sigma_\epsilon^2}{2\sigma^2}},$$

i.e., the quality is halved compared to case A. This is a penalty for omitting one of the trading times since the frequency of trading is double of that of reporting. For comparative purposes, we summarize the result in Table 2.

We also consider a situation in which the second trade occurs at some random time between time moments 0 and 2, while asset returns for periods $[0, 1]$ and $[1, 2]$ are known, e.g, the trading is done twice a month, but the exact moments are not known, and bi-weekly asset returns are used to explain the monthly hedge

$\sigma_\epsilon^2/\sigma^2$	$\mathbf{E}R^2_{(2A)} \geq$	$\mathbf{E}R^2_{(2B)} \geq$
0	1	1/2
1	2/3	1/3
2	1/2	1/4
3	2/5	1/5

Table 2: Lower bounds on expected proportions of explained variance for the two-period model (cases A and B).

fund return. If only monthly asset returns were used, the forecast would coincide with case B. However, use of bi-weekly asset returns leads to slightly better forecasting quality with the lower bound being

$$\frac{3/4}{1+\frac{\sigma_\epsilon^2}{2\sigma^2}}.$$

The model can be extended to three or more periods, but that would not add much illustrative value. The pattern is clear: if returns of a hedge fund manager strategy are consistent with the market timing concept, they can be forecasted by using piecewise linear functions of asset returns (option returns being an example of such functions); returns for managers with greater market timing ability (i.e., skill) are easier to forecast, but if we leave out some trading times out of consideration, the forecast's quality may degrade significantly. We also showed that in this situation it may be possible to improve the forecast by examining bi-weekly asset returns even if hedge fund returns are reported monthly.

The above analysis was conducted for a one-asset model. The results will still be valid for a multi-asset model if the investment in uncorrelated assets is done independently and the set of traded assets remains the same. However, it is obvious that in the real-world situation that might not be the case, especially in the long run.

B Choice of Matrix B_a

Recall that for asset a in asset class g daily returns are

$$r_{t,a}^{\mathbf{a}} = \bar{r}_a^{\mathbf{a}} + B_a f_t + \epsilon_{t,a},$$

The easiest way to choose B_a is to set it to $\mathbf{e}_g$ if $a \in \mathcal{G}_g$, where $\mathbf{e}_g$ is a vector that has 1 in gth position and all other components are 0. This way only the factor that corresponds to the asset class controls the asset return. However, this results in unrealistically high correlations between assets and corresponding indices. The correlation can be reduced by increasing the variance of the noise term $\epsilon_{t,a}$, but the increase must be rather big. We try to choose B_a appropriately to lower the correlation.

First we form matrix $\tilde{B}$. We assign $\tilde{B}_{a,g}$ random values (distributed as $\mathcal{N}(0, 0.15^2)$, independent of other rvs) if $a \notin \mathcal{G}_g$, and 0 if $a \in \mathcal{G}_g$. Finally, we set $B_{a,g} = 1$ for each pair $(a,g) : a \in \mathcal{G}_g$, and if $a \notin \mathcal{G}_g$, we set

$$B_{a,g} = \tilde{B}_{a,g} - \frac{1}{|\mathcal{G}_g|} \sum_{\alpha \in \mathcal{G}_g} \tilde{B}_{\alpha,g}.$$

It is easy to check that the matrix B constructed in such a way satisfies

$$\frac{1}{|\mathcal{G}_g|} \sum_{a \in \mathcal{G}_g} B_a = \mathbf{e}_g, \quad g = 1, \ldots, G,$$

which means that, though the asset returns depend on all factors, the index returns for the asset class will depend only on the corresponding factor since all others will cancel out.

C Data Calibration

We have 36 monthly returns for 8 factors (which are, in fact, returns of some indices). We calculate sample means and variances and then scale them to get daily means in excessive of the risk free rate (estimated using 3-month treasury bills data) and variances. This gives us the following table:

Table 3: Mean and Variance of Excessive Daily Index Returns.

Factors	1	2	3	4	5	6	7	8
Means	.0508%	.0223%	.0228%	.0217%	.0221%	-.0025%	-.0056%	.0018%
Std. Deviations	.70%	.88%	1.17%	.22%	.48%	1.16%	.68%	.35%

We also get a correlation matrix for the factors:

Table 4: Correlation Matrix of Daily Index Returns.

Factors	1	2	3	4	5	6	7	8
1	1.0000	0.5347	0.4360	0.3397	0.0861	-0.1454	-0.1718	0.0084
2	0.5347	1.0000	0.4424	0.1488	0.4490	-0.1274	0.0661	-0.4025
3	0.4360	0.4424	1.0000	-0.1204	-0.1032	0.0305	0.0370	0.0940
4	0.3397	0.1488	-0.1204	1.0000	0.3887	-0.4207	-0.0087	-0.1698
5	0.0861	0.4490	-0.1032	0.3887	1.0000	-0.2833	0.1158	-0.9051
6	-0.1454	-0.1274	0.0305	-0.4207	-0.2833	1.0000	-0.0507	0.1957
7	-0.1718	0.0661	0.0370	-0.0087	0.1158	-0.0507	1.0000	-0.1986
8	0.0084	-0.4025	0.0940	-0.1698	-0.9051	0.1957	-0.1986	1.0000

Scaling it by corresponding variances gives us the covariance matrix C.

References

[1] V. Agarwal and N. Naik. Risk and portfolio decisions involving hedge fund. *Review of Financial Studies*, 17, 2004.

[2] Sid Browne. Reaching goals by a deadline: Digital options and continuous time active portfolio management. *Advances in Applied Probability*, 31, 1999.

[3] Vladimir Cherkassky and Yunqian Ma. Selection of meta-parameters for support vector regression. In *Lecture Notes in Computer Science*, volume 2415, pages 687–693. ISBN 0302-9743, 2002.

[4] Nello Cristianini and John Shawe-Taylor. *Support Vector Machines and other kernel-based learning methods*. Cambridge, 2000.

[5] W. Fung and D.A. Hsieh. Empirical characteristics of dynamic trading strategies: The case of hedge funds. *The Review of Financial Studies*, 10, 1997.

[6] W. Fung and D.A. Hsieh. The risk in hedge fund strategies: Theory and evidence from trend followers. *The Review of Financial Studies*, 14, 2001.

[7] W. Fung and D.A. Hsieh. Extracting portable alphas from euity long-short hedge funds. *Journal of Investment Management*, 2, 2004.

[8] Andrew Grauberg. *Hedge Fund of Funds*. ISBN 9963-8936-0-0, 2004.

[9] M. B. Haugh and A. W. Lo. Asset allocation and derivatives. *Quantitative Finance*, 1, 2001.

[10] K. Mazuy J. Treynor. Can mutual funds outguess the market? *Harvard Business Review*, 44:131–136, 1966.

[11] Andrew W. Lo. Risk management for hedge funds: Introduction and overview. *Working Paper Series*, 2001.

[12] R. Merton. On market timing and investment performance. i. an equilibrium theory of value for market forcasts. *The Journal of Business*, 54:363–406, 1981.

[13] Robert Merton. On market timing and investment performance i: An equilibrium theory of value for market forecasts. *Journal of Business*, 54:363–407, 1981.

[14] W. Sharpe. Asset allocation: Management style and performance measurement. *Journal of Portfolio Management*, 18, 2001.

[15] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, 1995.

[16] V. N. Vapnik. *Statistical Learning Theory*. John Wiley & Sons, A Wiley Inter-Science Publication, 1998.