# A Statistic for Measuring the Influence of Side Information in Investment

Charles Mathis
Electrical Engineering and Statistics Departments
Stanford University
Stanford, CA 94305-9505 USA
Email: cmathis@stanford.edu

Thomas M. Cover
Electrical Engineering and Statistics Departments
Stanford University
Stanford, CA 94305-9510 USA
Email: cover@stanford.edu

***Abstract*— Side information can increase the growth rate of wealth in stock market investment. The question is: when is the side information useful and when is it illusory? We propose a statistic to test this. We compare the wealth achieved by the best constant rebalanced portfolio in hindsight on the entire stock market sequence to the wealth resulting from the best constant rebalanced portfolio in hindsight on the subsequences of stocks identified by the states of the side information sequence. If one can't make money from side information given perfect hindsight one never can.**

**We prove, under the null hypothesis of independence of the side information sequence from the stock sequence, that the logarithm of the ratio of optimal wealth with and without side information is asymptotically distributed as half a chi-squared random variable with $(d-1)(m-1)$ degrees of freedom, where $d$ is the number of states of side information and $m$ is the number of stocks. This statistic has an asymptotic distribution that is independent of the particular stock sequence or the distribution of the side information sequence. Thus the use of side information can be accepted when the wealth ratio exceeds a certain fixed level given by the appropriate quantile of the chi-squared distribution.**

## I. Summary

Let $x_1, x_2, \ldots, x_n \in \mathbb{R}_+^m$ be a sequence of stock market vectors, where $x_{ij}$ denotes the price relative (the ratio of closing price to opening price that day) for stock $j$ on day $i$. Then $b^t x_i$ is the wealth relative on day $i$ resulting from using portfolio $b = (b_1, b_2, \ldots, b_m)$, where $b_j$ is the proportion of wealth invested in stock $j$. A portfolio $b$ must satisfy the constraint $\sum_{i=1}^m b_i = 1$. Let

$$S_n(b, x^n) = \prod_{i=1}^n b^t x_i$$

be the wealth resulting from using constant rebalanced portfolio $b$ on the stock sequence $x^n$. Let

$$S_n^*(x^n) = \max_b S_n(b, x^n)$$

be the maximum such wealth in hindsight achievable on the sequence $x^n = x_1 x_2 \cdots x_n$, where the maximum is over all portolios $b = (b_1, \ldots, b_m)$ with $\sum_{i=1}^m b_i = 1$. We note that this wealth can actually be achieved (to first order in the exponent) by a universal portfolio [1], [2], [3]. More precisely, the causal universal portfolio $\hat{b}_i(\cdot)$ specified in [2] achieves wealth $\hat{S}_n(x^n) = \prod_{i=1}^n b^t(x^{i-1}) x_i$, which, for $m = 2$ stocks, satisfies

$$\hat{S}_n(x^n)/S_n^*(x^n) \geq 1/2\sqrt{n+1}$$

for all $n$ and all market sequences. So $S_n^*$ is a useful measure of what can be reasonably achieved on the actual market sequence $x^n = (x_1, x_2, \ldots, x_n)$, $x_i \in \mathbb{R}_+^m$.

Now we consider a side information sequence $s^n = (s_1, s_2, \ldots, s_n)$, where $s_i \in \{1, 2, \ldots d\}$. Here $s_i$ denotes the state of side information on day $i$. This sequence can be generated in any fashion whatsoever, and may or may not be dependent on the stock sequence $x^n$. We shall test the null hypothesis that it is not.

While there are many techniques for showing dependence of one sequence on another, none of them relates to the financial value of this dependence, all involve some sort of quantization of $x^n$ or an assumption about the distribution of $x^n$, and none are appropriate for our needs.

We shall look directly at the performance of the best state-dependent portfolio $b(s)$ given hindsight after observing the entire sequence $x^n$. Thus consider

$$S_n^{**} = \max_{b(\cdot)} \prod_{i=1}^n b^t(s_i) x_i.$$

This can be re-expressed as the product of the performance of the best constant rebalanced portfolio on each subsequence $\{i : s_i = k\}$, $k = 1, 2, \ldots, d$, picked out by the state sequence $s^n$. Hence

$$S_n^{**} = S_{n,1}^* S_{n,2}^* \cdots S_{n,d}^*$$

where for each of the $d$ states,

$$S_{n,k}^* = \max_b \prod_{i : s_i = k} b^t x_i$$

is the maximum wealth achieved on the subsequence of times $i$ such that the corresponding state $s_i$ is equal to $k$.

A sequence $x^n$ of vectors will be said to be of full dimension if the convex hull of $x_1, x_2, \ldots x_n$ strictly contains a ray $\lambda \vec{1} \in \mathbb{R}^m$. A sequence $x^n$ will be said to be nondegenerate (with respect to $s^n$) if every subsequence induced by $s^n$ is of full dimension.

These conditions guarantee that we are really talking about $m$ stocks and that the $S_n^*$ maximizing portfolio is finite in each component.

We must clarify what is meant when we say that the state sequence $s^n$ is independent of $x^n$. Since we are stating an individual sequence result, one holding for each $x^n$, we can treat $x^n$ as a fixed sequence and thus $x^n$ is independent of any other random sequence $s^n$. So this would be trivial independence. What we need for the null hypothesis (that the state sequence is irrelevant to $x^n$) is that all permutations of $s^n$ are equally likely.

Define the statistic

$$T_n^* = \frac{S_n^{**}}{S_n^*}$$

to be the ratio of wealth induced by the best constant rebalanced portfolio with side information to the wealth without side information. Clearly $T_n^* \geq 1$, because the maximum over piecewise constant portfolios is greater than the maximum over constant portfolios.

We then have the following theorem:

**Theorem 1:** For any nondegenerate stock sequence $x^n$ and a state sequence $s^n$ drawn according to the uniform distribution over all permutations of a sequence with $n_k$ occurences of state $s_i = k, k = 1, 2, \ldots, d$, the distribution of $T_n^*$ is approximately $e^{(1/2)\mathcal{X}^2_{(d-1)(m-1)}}$, where $\mathcal{X}^2_{(d-1)(m-1)}$ is chi-square with $(d-1)(m-1)$ degrees of freedom. More precisely

$$\ln T_n^* \stackrel{\text{dist}}{\rightarrow} \frac{1}{2}\mathcal{X}^2_{(d-1)(m-1)}$$

as $n \to \infty$, $n_k \to \infty$, $k = 1, 2, \ldots, d$.

We note that the distribution of this wealth ratio does not depend on the stock sequence or the state sequence, just so long as the time ordering of the state sequence is random, i.e. all permutations of $s^n$ are equally likely.

The meaning and utility of this result is that a nonparametric statistical significance of the side information can be computed. The hypothesis of a uniform distribution on permutations of $s^n$ corresponds to the use of a permutation test statistic.

The derived chi-squared distribution of the theorem allows computation of the p-value of $S^{**}/S^*$, the probability that under random side information with the same marginal state frequencies, we could have done as well.

## Acknowledgment

This work was partially supported by NSF grant CCR-0311633.

## References

[1] T. Cover. Universal Portfolios. *Mathematical Finance*, 1(1): 1-29, January 1991.

[2] T. Cover and E. Ordentlich. Universal Portfolios with Side Information. *IEEE Transactions on Information Theory,* 42(2):348-363, March 1996.

[3] E. Ordentlich and T. Cover. The Cost of Achieving the Best Portfolio in Hindsight. *Mathematics of Operations Research*,23(4):960-982, November 1998.