Journal of Empirical Finance 23 (2013) 93–116

Contents lists available at ScienceDirect

Journal of Empirical Finance

journal homepage: www.elsevier.com/locate/jempfin

Testing for monotonicity in expected asset returns[☆]

Joseph P. Romano [a,1], Michael Wolf [b,*,2]

[a] Departments of Economics and Statistics, Stanford University, Stanford, CA 94305, USA
[b] Department of Economics, University of Zurich, CH-8032 Zurich, Switzerland

ARTICLE INFO

Article history:
Received 22 June 2012
Received in revised form 3 April 2013
Accepted 3 May 2013
Available online 12 May 2013

JEL classification:
C12
C58
G12
G14

Keywords:
Bootstrap
CAPM
Monotonicity tests
Non-monotonic relations

ABSTRACT

Many postulated relations in finance imply that expected asset returns strictly increase in an underlying characteristic. To examine the validity of such a claim, one needs to take the entire range of the characteristic into account, as is done in the recent proposal of Patton and Timmermann (2010). But their test is only a test for the *direction* of monotonicity, since it requires the relation to be monotonic from the outset: either weakly decreasing under the null or strictly increasing under the alternative. When the relation is non-monotonic or weakly increasing, the test can break down and falsely 'establish' a strictly increasing relation with high probability. We offer some alternative tests that do not share this problem. The behavior of the various tests is illustrated via Monte Carlo studies. We also present empirical applications to real data.

1. Introduction

Many postulated relations in finance imply that expected asset returns strictly increase in a certain characteristic. When the assets are equities, examples of such characteristics are CAPM beta, book-to-market, size, momentum, and reversal. When the assets are bonds, examples of such characteristics are maturity and rating quality. The search for new such characteristics is a never-ending quest, partly in hopes of creating novel trading strategies to 'beat the market'.

It is, therefore, of interest to test whether a particular characteristic indeed generates expected asset returns that are strictly increasing. Say there are a total of $N + 1$ categories, ordered according to the underlying characteristic. The postulated relation says that as one moves up from one category to the next, then the expected return should strictly increase. (The opposite case of expected asset returns being supposedly monotonically decreasing can be handled analogously by simply multiplying all returns by negative one or, alternatively, by reversing the order of the various return categories considered.)

For a long time, the standard in the field has been to simply test for a difference in expected returns between the highest and the lowest return category. Such a test is easily carried out, since the parameter of interest is univariate, being the difference

☆ We thank two anonymous referees, Olivier Ledoit, Markus Leippold, and various seminar participants for helpful comments.
* Corresponding author. Tel.: +41 44 634 5096.
E-mail addresses: romano@stanford.edu (J.P. Romano), michael.wolf@econ.uzh.ch (M. Wolf).
[1] Research supported by NSF Grant DMS-0707085.
[2] Research supported by the Swiss National Science Foundation (NCCR FINRISK, Module A3).

0927-5398/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jempfin.2013.05.001

between two expected values. Therefore, a conventional t-test can be applied, though one has to account for potential serial correlation of returns in computing the standard error that appears in the denominator of the test statistic.

A new test has been proposed recently by Patton and Timmermann (2010), abbreviated by PT henceforth. As they point out, simply testing for a difference in expected returns between the highest and the lowest category can be misleading. It could happen that the relation of expected returns is flat, or even decreasing, for intermediate return categories while a positive difference 'highest minus lowest' still exists. In this case, providing sufficient data are collected, the simple t-test is likely to falsely decide in favor of a strictly monotonic relation. Take the example of five return categories, ordered from lowest to highest, with respective expected returns of 1.0, 1.5, 1.1, 1.4 and 1.6. In this example, the overall relation is non-monotonic, even though the difference 'highest minus lowest' is positive.

Therefore, a more comprehensive approach is needed to establish a strictly monotonic relation over the entire range of return categories. When moving from one category up to the next, the difference in expected returns must be established as significantly positive every time. In other words, all N expected return differentials 'higher minus lower' must be established as significantly greater than zero. Such a test is more complex, since the underlying parameter is now an N-dimensional vector rather than a univariate number.

A natural test statistic for the more comprehensive approach is obtained as follows. Compute individual t-test statistics for each of the N expected return differentials, where in each case the alternative corresponds to the expected return differential being positive. Then take the minimum of the individual test statistics as the overall test statistic. If the resulting min-t statistic is 'sufficiently' large, one decides in favor of a strictly monotonic relation. The statistical question is how to obtain a proper critical value for this test. PT use a bootstrap method, resampling from a certain *worst-case* null distribution. Their proposed method is based on the implicit assumption that if the relation is not strictly increasing, it must be weakly decreasing. That is, if the expected return differentials are not all strictly positive, then they must be all weakly negative (meaning less than or equal to zero). As a result, the test of PT is only a test for the *direction* of monotonicity, not for monotonicity itself, since monotonicity is assumed from the outset.

While the assumption of monotonicity may hold for some applications, it certainly cannot be invoked always. When the true relation is non-monotonic, or even increasing but only weakly instead of strictly, the PT test can break down and no longer successfully control the rejection probability under the null. In fact, as shown in Remark B.1, its size can be arbitrarily close to one: it can falsely decide in favor of the alternative of a strictly monotonic increasing relation with a probability arbitrarily close to one. In this paper, we first discuss this problem of the PT test and then offer some alternative tests that do not share the problem and are, therefore, safer to use in general settings.

Having said this, all the new tests we present have the choice of alternative hypothesis in common with the PT test. The alternative postulates that a monotonic relation (in the sense of being strictly increasing) exists. In other proposals, this postulate becomes the null hypothesis instead; for example, see Wolak (1987, 1989) and also Fama (1984). Such an approach is unnatural, though, if the goal is to establish the existence of a monotonic relation. By not rejecting the null, one can never claim (quantifiable) statistical evidence in favor of the associated hypothesis. Hence, we will not include such tests in our paper. For a theoretical discussion of such tests and some examination of finite-sample performance, see PT.

The remainder of the paper is organized as follows. Section 2 describes the formal setup and the testing problem of interest. Section 3 presents the various approaches for designing tests for monotonicity. Section 4 details how the various tests are implemented in practice. Since the available data is assumed to be a multivariate time series, an appropriate bootstrap method is used to calculate critical values in a nonparametric fashion. Section 5 examines finite-sample performance via Monte Carlo studies. Section 6 contains empirical applications to real-life data. Finally, Section 7 concludes. Mathematical results are relegated to the Appendix.

2. Formal setup and testing problem

Our notation generally follows the notation of PT. One observes a strictly stationary time series of return vectors $r_t \equiv (r_{t,0}, r_{t,1}, \dots, r_{t,N})'$ of dimension $N+1$. The observation period runs from $t=1$ to $t=T$, so the sample size is given by T. Denote the expected return vector by $\mu \equiv (\mu_0, \mu_1, \dots, \mu_N)'$ and define the associated expected return differentials as

$$\Delta_i \equiv \mu_i - \mu_{i-1} \quad \textit{for} \;\; i = 1, \dots, N. \tag{2.1}$$

To avoid any possible confusion, note that the characteristic according to which the $N+1$ return categories are ordered is always assumed to be predetermined and not data-dependent. We also introduce the following notation for the observed return differentials:

$$\boldsymbol{d}_t \equiv \left(d_{t,1}, \dots, d_{t,N}\right)' \equiv \left(r_{t,1} - r_{t,0}, \dots, r_{t,N+1} - r_{t,N}\right)'. \tag{2.2}$$

Therefore, letting $\boldsymbol{\Delta} \equiv (\Delta_1, \dots \Delta_N)'$, one can also write $\boldsymbol{\Delta} = \mathbb{E}(\mathbf{d}_t)$.