127

Reihe Ökonomie

Economics Series

The Day-of-the-Week Effect Revisited: An Alternative Testing Approach

Raimund Alt, Ines Fortin, Simon Weinberger

127

Reihe Ökonomie

Economics Series

The Day-of-the-Week Effect Revisited: An Alternative Testing Approach

Raimund Alt, Ines Fortin, Simon Weinberger

November 2002

Institut für Höhere Studien (IHS), Wien
Institute for Advanced Studies, Vienna

Contact:

Raimund Alt
Department of Economics and Finance
Institute for Advanced Studies
Stumpergasse 56, A-1060 Vienna, Austria
☎: +43/1/599 91-159
email: alt@ihs.ac.at

Ines Fortin
Department of Economics and Finance
Institute for Advanced Studies
Stumpergasse 56, A-1060 Vienna, Austria
☎: +43/1/599 91-165
email: fortin@ihs.ac.at

Simon Weinberger
email: simonweinberger@hotmail.com

Abstract

This paper questions traditional approaches for testing the day-of-the-week effect on stock returns. We propose an alternative approach based on the closure test principle introduced by Marcus, Peritz and Gabriel (1976), which has become very popular in Biometrics and Medical Statistics. We test all pairwise comparisons of daily expected stock returns, while the probability of committing any type I error is always kept smaller than or equal to some prespecified level a for each combination of true null hypotheses. We confirm day-of-the-week effects for the S&P 500, the FTSE 30 and the DAX 30 found in earlier studies, but find no evidence for the 1990's.

Keywords

Day-of-the-week effect, multiple hypotheses testing, multiple comparisons, closed test procedures, multiple level a test

JEL Classifications

C12, C20, G14

Contents

1 Introduction

In recent decades the testing for market anomalies in stock returns has become an active field of research in empirical finance. Some anomalies have attracted much attention not only in academic journals but also in the financial press. Among the more well-known anomalies are the size effect, the January effect and the day-of-the-week effect (Monday effect).

Since Cross (1973) observed negative returns on Mondays for the US stock market, numerous studies have been devoted to the examination of day-of-the-week effects, in particular irregularities affiliated with Monday returns (see e.g. French (1980)). Negative Monday returns were found to be robust over time and different markets early after (see Jaffe and Westerfield (1985) and Keim and Stambaugh (1984)). Also more recent studies, examining data until the early 1990's document the existence of a day-of-the-week effect in major markets (see e.g. Dubois and Louvet (1996)). Meanwhile various methodological issues related among others to time series properties of stock returns have been addressed (see e.g Connolly (1991) and Chang, Pinegar and Ravichandran (1993)).

Though testing for market anomalies often involves the use of several tests, it is a matter of fact that empirical studies often do not appropriately account for the multiplicity effect. This may lead to an inflated occurrence of multiple type I errors producing spurious significance with regard to the test results. To overcome this problem we propose an alternative approach for testing the day-of-the-week effect, which differs from traditional approaches in two respects. First, we test all pairwise comparisons of daily expected stock returns. Second, the null hypotheses of interest are tested in such a way, that the multiple level α is controlled, i.e. the probability of committing any type I error is always smaller than or equal to the given level α. The latter property, which holds for each combination of true null hypotheses, is guaranteed by the application of the so-called closure test principle introduced by Marcus, Peritz and Gabriel (1976).

Closure tests have become very popular in Biometrics and Medical Statistics, which is due to the fact that they are often more powerful than classic multiple test procedures (see Pigeot (2000) for a recent survey). Aside from a few exceptions like Neusser (1991) or Madlener and Alt (1996), closed test procedures, as they are also called, seem to be quite unknown in the economics and financial literature. In particular, Savin (1984) did not mention them in his survey, although these procedures were already used in the 1970's.

We are not the first to address the problem of controlling type I error probabilities in empirical studies on market anomalies. Recently, Greenstone and Oyer (2000) have suggested the use of the Bonferroni procedure, a simple but rather conservative method. Another approach was suggested by Sullivan, Timmermann and White (2001), who use a computationally intensive bootstrapping procedure to account for the multiplicity effect in studies on calendar anomalies. In some sense our proposal can be seen as a middle-of-the-road approach. On the one hand, closed test procedures are based on a relatively simple construction involving level α tests for all intersection hypotheses. On the other hand, the resulting multi-step or stepwise procedures are often more powerful than classic single-step procedures, like for example, the Bonferroni procedure.

The paper is organized as follows. Section 2 discusses traditional approaches for testing the day-of-the-week effect. Section 3 explains the closure test principle and its implementation for the problem at hand. Section 4 provides a description of the used data set. Section 5 presents the results of the closed test procedure applied to daily returns of the S&P 500, the FTSE 30 and the DAX 30 for the period 1971-2001 as well as three subperiods. Section 6 concludes the paper.

2 Traditional Testing Approaches

The testing for day-of-the-week effects on stock returns usually involves regression models with dummy variables. Meanwhile there are numerous extensions, but for convenience let us restrict ourselves to the following basic model set-up. Let R_{is} denote the return on day i during the week s. Then R_{is} can be written as

$$R_{is} = E(R_{is}) + \varepsilon_{is}, \tag{1}$$

where $E(R_{is})$ is the expected value of R_{is} and ε_{is} is the corresponding error term. Usually it is assumed that the expected return on day i is constant over all weeks of the observation period. In this case (1) becomes

$$R_{is} = \gamma_i + \varepsilon_{is}, \tag{2}$$

where $\gamma_i \equiv E\left(R_{is}\right)$.

The above equations can be formulated jointly as a regression equation with dummy variables:

$$R_t = \gamma_1 d_{1t} + \gamma_2 d_{2t} + \gamma_3 d_{3t} + \gamma_4 d_{4t} + \gamma_5 d_{5t} + \varepsilon_t. \tag{3}$$

R_t is now the return on any day of the week, the dummy variables d_{it}, $1 \leq i \leq 5$, indicate the day of the week, on which the return R_t was observed, and ε_t is the error term, $1 \leq t \leq T$, T being the number of all observations. In particular, $d_{1t} = 1$ indicates that the return was observed on a Monday, $d_{2t} = 1$ indicates a Tuesday return and so on. Whenever $d_{it} = 1$ for some i, then $d_{jt} = 0$ for all $j \neq i$.

The usual approach is then to test the global null hypothesis,

$$H_0\colon \quad \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5,$$

that all expected daily returns are equal. If H_0 is not rejected at a given level α, then one concludes that there is no evidence that the expected daily returns are different. Otherwise, in case that H_0 is rejected, it is common practice, to look at the values of the t-statistics for the single regression coefficients, assuming $\gamma_i = 0$. In other words, the additional null hypotheses

$$H_i\colon \; \gamma_i = 0 \qquad 1 \leq i \leq 5$$

are tested, though this is often done in a more implicit way.

An alternative formulation of (2) is

$$R_{is} = \begin{cases} \delta_1 + \varepsilon_{1s} & \text{if } i = 1 \\ \delta_1 + \delta_i + \varepsilon_{is} & \text{if } 2 \leq i \leq 5 \end{cases}$$

where $\delta_1 \equiv \gamma_1$ has the same meaning as in (2), while the differences $\delta_i \equiv \gamma_i - \gamma_1$ measure the expected excess return of some day i of the week over a specific other day (in this case Monday). The regression equation then looks as follows:

$$R_t = \delta_1 + \delta_2 d_{2t} + \delta_3 d_{3t} + \delta_4 d_{4t} + \delta_5 d_{5t} + \varepsilon_t.$$

The global null hypothesis that all expected daily returns are equal can be formulated as:

$$H_0\colon \ \ \delta_2 = \delta_3 = \delta_4 = \delta_5 = 0.$$

If H_0 is not rejected at level α, then one concludes that there is no evidence that the expected excess returns are different from zero. Otherwise, if H_0 is rejected, the null hypotheses

$$H_i\colon \ \delta_i = 0 \qquad 2 \leq i \leq 5$$

are tested as well. The rejection of, say H_i, would indicate that there is some evidence that the expected returns between day i and Monday, respectively, are different.

We think that both testing approaches suffer from some serious drawbacks. With regard to the first approach, the testing of the null hypotheses H_i: $\gamma_i = 0$, $1 \leq i \leq 5$, is questionable, to say the least. If there is some evidence that the expected daily returns are not equal, what is the logic behind the next step to test the null hypotheses that each expected daily return is equal to zero? One of the problems of the second approach is that it is obviously biased towards Monday, in the sense that it is only the expected Monday return which is compared to the expected return of each other day. In particular, if H_2 is tested and not rejected, there is still a chance for Monday to be different, namely when H_3 is tested, and so on. But there is another problem inherent in both approaches. Each of the described testing strategies requires the testing of several null hypotheses. If this is done in the traditional way by using tests, each having a significance level of 5%, then this may result in spurious significance in the employed tests. The reason for this is that the type I error probability for each testing approach may be much larger than 5%.

To overcome these problems we suggest an alternative approach, which differs from the traditional approaches in two ways. First, if the global null hypothesis H_0 is rejected, we test a new set of null hypotheses, namely the multiple comparisons H_{ij}: $\gamma_i = \gamma_j$, for all $1 \leq i < j \leq 5$, resulting in an unbiased, symmetric test design. Second, the null hypotheses are tested in such a way, that the probability of committing any type I error is always kept smaller than or equal to some given level α, for each combination of true null hypotheses. The testing procedure is based on the so-called closure test principle, which we describe in the following section.

3 Closed Test Procedures - An Alternative Approach

Closed test procedures, which were introduced by Marcus, Peritz and Gabriel (1976), belong to the class of multiple level α tests. A multiple level α test is a test procedure with the property that the probability of committing any type I error is always smaller than or equal to the given level α, for each combination of true null hypotheses. A simple example for a multiple level α test is the well-known Bonferroni procedure, where each of the null hypotheses $H_{01}, ..., H_{0n}$ is tested at the level α/n. The multiple level α property of the Bonferroni procedure can be shown by applying the Bonferroni inequality to the type I error probability

$$\begin{aligned}
& P\left(\bigcup_{i\in I}\{H_{0i} \text{ is rejected}\}\right) \\
\leq & \sum_{i\in I} P\left(H_{0i} \text{ is rejected}\right) \\
= & |I| \cdot \frac{\alpha}{n} \\
\leq & \alpha
\end{aligned}$$

where I is the index set of the true null hypotheses and $|I|$ is the number of elements in I. A major disadvantage of the Bonferroni procedure is the fact that the type I error probability can sometimes become very small compared to the given level α, for example, if the number $|I|$ of true null hypotheses is small. A recent application of the Bonferroni procedure in finance is due to Greenstone and Oyer (2000), who tested various calendar based anomalies. Meanwhile, many classic single-step procedures have been replaced by closed test procedures, since the latter are usually considered to be more powerful. There are some empirical studies, e.g. Chow and Denning (1993) or Huber (1997), where the use of the Sidak inequality results in slightly larger significance levels for the single tests compared with those of the Bonferroni procedure. For practical purposes, however, the additional gain in power is negligible.

The application of closed test procedures is based on two assumptions. First, the set $\mathcal{H} = \{H_{01}, ..., H_{0n}\}$ of interesting null hypotheses has to be closed under intersection, i.e. it is required that $H_{0i} \cap H_{0j} \in \mathcal{H}$ for any two indices $i \neq j$. Second, there should be available a multiple test procedure $\Phi = (\phi_1, ..., \phi_n)$, where each test ϕ_i is a level α test for the null hypothesis H_{0i},

$1 \le i \le n$. Given these two assumptions one can apply the so-called 'closure test principle', resulting in a test procedure which keeps the multiple level α, i.e. the probability of committing any type I error is always less than or equal to the given level α, for each combination of true null hypotheses.

The closure principle works as follows. One defines another test procedure $\Psi = (\psi_1, ..., \psi_n)$, which combines the results of the tests $\phi_1, ..., \phi_n$ in such a way that the test procedure Ψ is a multiple level α test. Each test $\psi_i, i = 1 \le i \le n$, is defined by the following rule:

$$\text{Reject } H_{0i}\text{, if each subhypothesis } H_{0j} \subset H_{0i}$$
$$\text{is rejected by its level } \alpha \text{ test } \phi_j,$$

where a subhypothesis of H_{0i} is any element in $\mathcal{H}$ which is contained in H_{0i}, including H_{0i} itself. Marcus, Peritz and Gabriel (1976) have shown that the test procedure $\Psi = (\psi_1, ..., \psi_n)$ keeps the multiple level α.

With regard to regression equation (3) we want to test all pairwise comparisons

$$H_{ij}\colon\ \gamma_i = \gamma_j \qquad 1 \le i < j \le 5,$$

provided that the global null hypothesis

$$H_0\colon\ \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5$$

has been rejected.

In order to apply the closure test principle we have to consider the two assumptions described above. Obviously the set of null hypotheses

$$\mathcal{H} = \{H_0, H_{12}, H_{13}, H_{14}, H_{15}, H_{23}, H_{24}, H_{25}, H_{34}, H_{35}, H_{45}\}$$

we are interested in, is not closed under intersection, e.g. the intersection $H_{12} \cap H_{13} = \{\gamma_1 = \gamma_2 = \gamma_3\}$ is not an element of $\mathcal{H}$. Therefore, in order to get a closed set $\mathcal{H}^*$ of null hypotheses, we have to add as auxiliary hypotheses all possible intersections of the sets in $\mathcal{H}$. Table 1 lists all pairwise comparisons $H_{ij} : \gamma_i = \gamma_j, 1 \leq i < j \leq 5$, together with all possible intersections.

Since we are interested in testing all pairwise comparisons H_{ij}: $\gamma_i = \gamma_j$, $1 \leq i < j \leq 5$, the closure principle requires that all subhypotheses of each comparison have to be tested by their corresponding level α tests. It will turn out, however, that this is not always necessary.

If, for example, H_0 is not rejected at level α, the procedure stops and none of the null hypotheses will be rejected. This is so, because H_0 is a subhypothesis of each hypothesis in $\mathcal{H}^*$ and, therefore, given the closure test rule described above, no null hypothesis can be rejected by the tests $\psi_1, ..., \psi_n$. On the other hand, if H_0 is rejected at level α, one has to consider the subhypotheses of each comparison H_{ij}: $\gamma_i = \gamma_j$. Whenever a subhypothesis of H_{ij} cannot be rejected at level α, this immediately implies that H_{ij} cannot be rejected by the closure test.

H_{ij}	H_{ijk}	H_{ijkl}	$H_{12345} = H_0$
$\gamma_1 = \gamma_2$	$\gamma_1 = \gamma_2 = \gamma_3$	$\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4$	$\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5$
$\gamma_1 = \gamma_3$	$\gamma_1 = \gamma_2 = \gamma_4$	$\gamma_1 = \gamma_2 = \gamma_3 = \gamma_5$	
$\gamma_1 = \gamma_4$	$\gamma_1 = \gamma_2 = \gamma_5$	$\gamma_1 = \gamma_2 = \gamma_4 = \gamma_5$	
$\gamma_1 = \gamma_5$	$\gamma_1 = \gamma_3 = \gamma_4$	$\gamma_1 = \gamma_3 = \gamma_4 = \gamma_5$	
$\gamma_2 = \gamma_3$	$\gamma_1 = \gamma_3 = \gamma_5$	$\gamma_2 = \gamma_3 = \gamma_4 = \gamma_5$	
$\gamma_2 = \gamma_4$	$\gamma_1 = \gamma_4 = \gamma_5$		
$\gamma_2 = \gamma_5$	$\gamma_2 = \gamma_3 = \gamma_4$		
$\gamma_3 = \gamma_4$	$\gamma_2 = \gamma_3 = \gamma_5$		
$\gamma_3 = \gamma_5$	$\gamma_2 = \gamma_4 = \gamma_5$		
$\gamma_4 = \gamma_5$	$\gamma_3 = \gamma_4 = \gamma_5$		

$H_{ij,kl}$	$H_{ijk,lm}$
$\gamma_1 = \gamma_2 \wedge \gamma_3 = \gamma_4$	$\gamma_1 = \gamma_2 = \gamma_3 \wedge \gamma_4 = \gamma_5$
$\gamma_1 = \gamma_2 \wedge \gamma_3 = \gamma_5$	$\gamma_1 = \gamma_2 = \gamma_4 \wedge \gamma_3 = \gamma_5$
$\gamma_1 = \gamma_2 \wedge \gamma_4 = \gamma_5$	$\gamma_1 = \gamma_2 = \gamma_5 \wedge \gamma_3 = \gamma_4$
$\gamma_1 = \gamma_3 \wedge \gamma_2 = \gamma_4$	$\gamma_1 = \gamma_3 = \gamma_4 \wedge \gamma_2 = \gamma_5$
$\gamma_1 = \gamma_3 \wedge \gamma_2 = \gamma_5$	$\gamma_1 = \gamma_3 = \gamma_5 \wedge \gamma_2 = \gamma_4$
$\gamma_1 = \gamma_3 \wedge \gamma_4 = \gamma_5$	$\gamma_1 = \gamma_4 = \gamma_5 \wedge \gamma_2 = \gamma_3$
$\gamma_1 = \gamma_4 \wedge \gamma_2 = \gamma_3$	$\gamma_2 = \gamma_3 = \gamma_4 \wedge \gamma_1 = \gamma_5$
$\gamma_1 = \gamma_4 \wedge \gamma_2 = \gamma_5$	$\gamma_2 = \gamma_3 = \gamma_5 \wedge \gamma_1 = \gamma_4$
$\gamma_1 = \gamma_4 \wedge \gamma_3 = \gamma_5$	$\gamma_2 = \gamma_4 = \gamma_5 \wedge \gamma_1 = \gamma_3$
$\gamma_1 = \gamma_5 \wedge \gamma_2 = \gamma_3$	$\gamma_3 = \gamma_4 = \gamma_5 \wedge \gamma_1 = \gamma_2$
$\gamma_1 = \gamma_5 \wedge \gamma_2 = \gamma_4$	
$\gamma_1 = \gamma_5 \wedge \gamma_3 = \gamma_4$	
$\gamma_2 = \gamma_3 \wedge \gamma_4 = \gamma_5$	
$\gamma_2 = \gamma_4 \wedge \gamma_3 = \gamma_5$	
$\gamma_2 = \gamma_5 \wedge \gamma_3 = \gamma_4$	

Table 1: Pairwise Comparisons and All Intersection Hypotheses.

4 Data

In order to empirically investigate the day-of-the-week effect we consider three major stock markets, the US (S&P 500), the UK (FTSE 30) and Germany (DAX 30). We extract daily closing levels of the selected indices from Thomson Financial/Datastream from January 1, 1971 to December 31, 2001 and compute daily returns on the indices as logarithmic returns

$$R_t = \ln(\frac{P_t}{P_{t-1}}),$$

where P_t are daily closing prices. To each return observation we assign the corresponding day of the week and generate a set of dummy variables indicating that day. As Monday return e.g. we consider the change from Friday close to Monday close. We exclude days, when markets are closed due to national holidays. So if Wednesday, for example, is a national holiday, the Thursday return is computed from the Tuesday to the Thursday close, resulting in a total of four observations for that particular week. In contrast to the S&P 500 and the FTSE 30, the DAX 30 is a performance index, i.e. it explicitly accounts for dividend payments. As we are not aware of any empirical evidence that dividend payments across all shares contained in the index occur primarily on any particular day of the week, the impact of single dividends on index returns should be of minor importance for our investigation.

To get a more detailed picture of our results we divide the total sample period into three subsamples, subperiod 1 (January 1, 1971 to December 31, 1980), subperiod 2 (January 1, 1981 to December 31, 1990), and subperiod 3 (January 1, 1991 to December 31, 2001). All tests we employ will thus be carried out first for the total period ranging from 1971 to 2001, and second for each of the three subperiods. Figures 1 to 3 plot the average rate of return per day of the week for the S&P 500, the FTSE 30 and the DAX 30. The first graph (black coloured) shows the average rates of return of the total period, the remaining graphs (grey coloured) show the corresponding numbers for the three subperiods.

Considering the total period, we observe a negative average Monday return in each of the markets, whereas all other average returns are positive, except for the FTSE 30 Thursday returns. If we take a closer look at the subsample returns, though, the empirical pattern found in the total period does not carry over to all subperiods. The finding of a supposedly negative Monday return seems to be strongly driven by the first two subperiods, where we see negative Monday returns for the S&P 500, FTSE 30 and DAX 30. In the last subperiod, on the other hand, Monday returns are positive across all markets examined. In general, there seem to exist differences between average day-of-the-week returns, but not in a consistent way over all markets and subperiods. Just looking at the graphs, there is a clear indication that differences in the average rate of return as to the day of the week, if any, decrease over time. A thorough statistical analysis will have to shed light on the question, whether the S&P 500, the FTSE 30 and the DAX 30 exhibit any significant day-of-the-week effects. As opposed to most other comparable investigations, we do not restrict ourselves to a potential Monday effect, but consider all days of the week likewise by applying the closure test principle as described in Section 3.

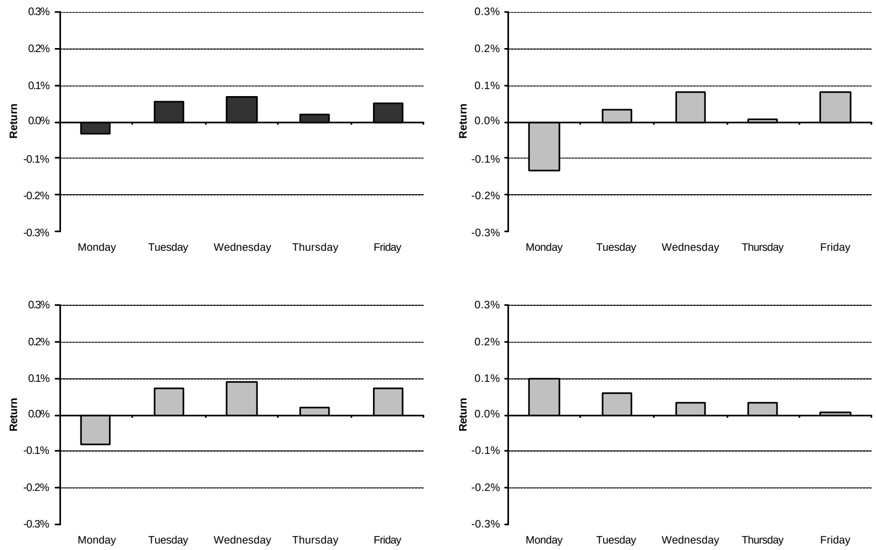

Figure 1: S&P 500 average day-of-the-week returns for the total period ranging from 1971 to 2001 (upper left graph), and the three subperiods 1971-1980 (upper right), 1981-1990 (lower left), 1991-2001 (lower right).

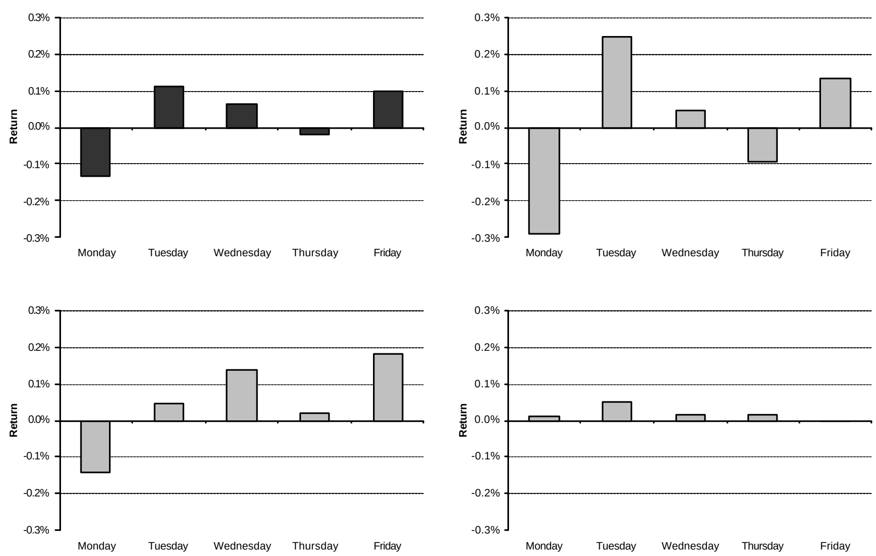

Figure 2: FTSE 30 average day-of-the-week returns for the total period ranging from 1971 to 2001 (upper left graph), and the three subperiods 1971-1980 (upper right), 1981-1990 (lower left), 1991-2001 (lower right).

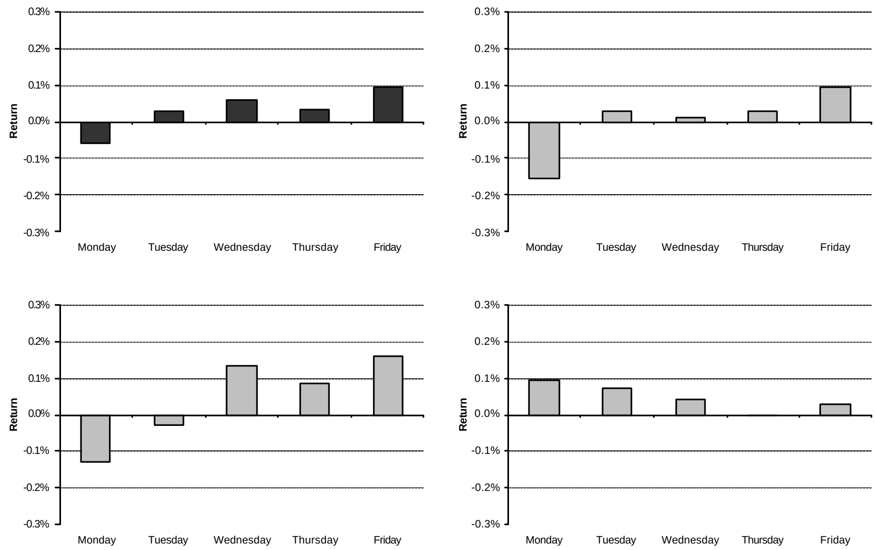

Figure 3: DAX 30 average day-of-the-week returns for the total period ranging from 1971 to 2001 (upper left graph), and the three subperiods 1971-1980 (upper right), 1981-1990 (lower left), 1991-2001 (lower right).

5 Results

To apply the test procedure described in Section 3, we first estimate the regression equation

$$R_t = \gamma_1 d_{1t} + \gamma_2 d_{2t} + \gamma_3 d_{3t} + \gamma_4 d_{4t} + \gamma_5 d_{5t} + \varepsilon_t,$$

as given by equation (3), by standard OLS. For all markets we estimate the given model for the total period 1971-2001 as well as for the three subperiods. We then apply the closure test principle by performing tests on the null hypotheses we are interested in, i.e. all hypotheses contained in

$$\mathcal{H} = \{H_0, H_{12}, H_{13}, H_{14}, H_{15}, H_{23}, H_{24}, H_{25}, H_{34}, H_{35}, H_{45}\},$$

and on the corresponding subhypotheses. The complete set $\mathcal{H}^*$ of hypotheses we need to test is listed in Section 3, Table 1.

The first step in the closed test procedure is to test the global null hypothesis $H_0 : \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5$. If the null hypothesis of equal day-of-the-week returns (no day-of-the-week effect) is not rejected at the 5% significance level, we do not need to check any other hypotheses, since the closure principle tells us that in this case none of the null hypotheses stating pairwise equality can be rejected, either. On the other hand, if the global null hypothesis is rejected at the 5% significance level, we test all ten primary null hypotheses H_{ij}: $\gamma_i = \gamma_j$, $1 \leq i < j \leq 5$ and their corresponding sets of subhypotheses.

Table 2 presents the p-values corresponding to the global F-test, i.e. the test of H_0, for all the different periods and markets considered. The null hypothesis of equal returns can be rejected for the entire period in all markets, but not for all subperiods. In the last subperiod H_0 cannot be rejected across all three markets. As the global null hypothesis is contained in the subhypotheses set of any pairwise comparison, we cannot reject any pairwise equalities for this period, either. This indicates that there is no evidence of a day-of-the-week effect in US, UK and German stock returns in subperiod 1991-2001. In the remaining subperiods, the F-test suggests a day-of-the-week effect for the 1970's across all markets, whereas in the 1980's only the FTSE 30 and the DAX 30 seem to exhibit a day-of-the-week anomaly.

For the periods where the global F-test does suggest a day-of-the-week effect, we need to test the complete set of subhypotheses of all primary hypotheses in order to locate the source for the anomaly. Table 3 exemplarily states all subhypotheses contained in the primary hypothesis H_{12} that expected Mon-

	S&P 500	FTSE 30	DAX 30
Total Period: 1971-2001	0.0339*	0.0000*	0.0046*
Subperiod 1: 1971-1980	0.0007*	0.0000*	0.0001*
Subperiod 2: 1981-1990	0.0900	0.0001*	0.0005*
Subperiod 3: 1991-2001	0.6193	0.9028	0.7811

Table 2: Global F-Test Results.
Numbers represent the p-values corresponding to the global F-Test for all markets and subperiods. * indicates significance at the 5 % level.

day returns equal expected Tuesday returns and reports the corresponding p-values for the total period of S&P 500 returns.

In closed test procedures, a given primary hypothesis is rejected if each subhypothesis is rejected. We thus report an adjusted p-value for H_{12}, which is defined as the maximum of all p-values corresponding to the subhypotheses contained in the given primary hypothesis. The value in the first row of Table 3 shows the adjusted p-value, while the other values represent traditional p-values. In the given example the adjusted p-value of 0.046, which is the maximum of the p-values of all subhypotheses of H_{12}, is roughly four times the traditional p-value of 0.012. The null hypothesis, however, can still be rejected at the 5% significance level. In this particular example, therefore, the closed test procedure and conventional testing yield the same test result, where by conventional testing we mean separate testing of the comparisons H_{ij} by using tests each having a significance level of 5%.

In a similar manner we determine adjusted p-values for the complete set of pairwise comparisons. The results for the US, the UK, and Germany are summarized in Tables 4 to 6. For periods where already the global test suggests no day-of-the-week effects we still report adjusted p-values for information purposes. Taking a closer look at the S&P 500 results, we see that the overall day-of-the-week effect observed in the total period can be attributed to significant differences between Monday and Tuesday, and Monday and Wednesday returns. For the first subperiod we additionally observe significant differences between Monday and Friday returns. In the second and third subperiod, however, the data do not reveal any week-day anomaly in the US market.

The FTSE 30 seems to exhibit the classical Monday effect in the total period, since Monday returns are statistically distinguishable from all other days of the week. Furthermore, the equality of Thursday returns with Tuesday and Friday returns can be rejected. For the 1970's and 1980's we cannot reject the equality of Monday and Thursday returns as well as the equality of Thursday and Friday returns. While Tuesday and Thursday returns differ significantly in the first subperiod their equality cannot be rejected at very high significance levels for subperiod 2. The global null hypothesis has already ruled out any day-of-the-week effect for the 1990's.

The analysis of the DAX 30 for the total period shows that Monday returns differ significantly from Wednesday and Friday returns. In the 1970's Monday returns are statistically distinguishable from all other days of the week. In the 1980's the equality of Monday and Tuesday returns cannot be rejected any longer. Subperiod 3 has already been found not to exhibit any day-of-the-week effect.

Since the main focus of our work is the test design we choose to restrict our presentation here to the results obtained by standard OLS estimation. In additional computations we account for heteroscedasticity by using the White correction when estimating the regression equation and allow for non-normality by using a more general chi-square test instead of the standard

F-Test.[4] The non-standard results further weaken the evidence of a day-of-the-week effect in the US stock market. H_0 cannot be rejected anymore for the total period, which means that there is no evidence for a day-of-the-week effect relating to the period 1971-2001. Results concerning pairwise comparisons are quite similar to the ones reported. In addition to the differences found already we see significant differences between Thursday and Friday returns for the 1970's and 1980's in the FTSE 30. Furthermore, the equality of Tuesday and Friday returns in the 1980's can be rejected for the DAX 30.

An interesting question that remains is whether the closure test principle and the conventional test methodology yield qualitatively different results. Tables 7 to 9 present p-values corresponding both to the closure test principle (upper value) and to the conventional approach (lower value in brackets), whenever a rejection of the primary hypothesis (based on the conventional approach) would have been the wrong conclusion.

In case of the S&P 500, the conventional approach would have yielded a different conclusion once for the total period and the first subperiod and three times for subperiod two. If we look at the primary hypothesis that Monday returns equal Friday returns in the total period, the adjusted p-value of 0.1147 does not reject the null hypothesis, whereas the conventional p-value of 0.0172 does reject.

While conventional testing would have found a Monday effect over the first two subperiods in the UK market, Monday returns are indistinguishable from Thursday returns when considering adjusted p-values. The same holds true for the comparison of Thursday and Friday returns.

Monday returns are no longer significantly different from Tuesday and Thurs-

[4] Results are available on request.

day returns when adopting the closure test principle for the DAX 30 data. Including the subperiods there are four cases where conventional testing would have lead to a different conclusion. For example, looking at the second subperiod we would have rejected the equality of Tuesday returns to Wednesday and Friday returns, although not all subhypotheses can be rejected.

		p-value
Primary hypothesis	H_{12}: $\gamma_1 = \gamma_2$	0.046* (adjusted)
Subhypotheses	H_{12}: $\gamma_1 = \gamma_2$	0.012*
	H_{123}: $\gamma_1 = \gamma_2 = \gamma_3$	0.008*
	H_{124}: $\gamma_1 = \gamma_2 = \gamma_4$	0.040*
	H_{125}: $\gamma_1 = \gamma_2 = \gamma_5$	0.019*
	H_{1234}: $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4$	0.021*
	H_{1235}: $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_5$	0.018*
	H_{1245}: $\gamma_1 = \gamma_2 = \gamma_4 = \gamma_5$	0.046*
	$H_{12,34}$: $\gamma_1 = \gamma_2 \wedge \gamma_3 = \gamma_4$	0.017*
	$H_{12,35}$: $\gamma_1 = \gamma_2 \wedge \gamma_3 = \gamma_5$	0.037*
	$H_{12,45}$: $\gamma_1 = \gamma_2 \wedge \gamma_4 = \gamma_5$	0.029*
	$H_{123,45}$: $\gamma_1 = \gamma_2 = \gamma_3 \wedge \gamma_4 = \gamma_5$	0.016*
	$H_{124,35}$: $\gamma_1 = \gamma_2 = \gamma_4 \wedge \gamma_3 = \gamma_5$	0.018*
	$H_{125,34}$: $\gamma_1 = \gamma_2 = \gamma_5 \wedge \gamma_3 = \gamma_4$	0.021*
	$H_{345,12}$: $\gamma_3 = \gamma_4 = \gamma_5 \wedge \gamma_1 = \gamma_2$	0.043*
	$H_{12345}{=}H_0$: $\gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 = \gamma_5$	0.034*

Table 3: Example of a Set of Subhypotheses.
The table lists p-values for the primary null hypothesis H_{12} and the set of all subhypotheses for the S&P 500, 1971-2001. The first value is the adjusted p-value, while the other values are traditional p-values.
* indicates significance at the 5 % level.

		Tuesday	Wednesday	Thursday	Friday
Total period	Monday	0.0459*	0.0427*	0.4545	0.1147
1971-2001	Tuesday		0.8931	0.5986	0.8970
	Wednesday			0.5986	0.8931
	Thursday				0.6601
Subperiod 1	Monday	0.0122*	0.0012*	0.0660	0.0016*
1971-1980	Tuesday		0.6157	0.8967	0.6157
	Wednesday			0.4705	0.9396
	Thursday				0.4705
Subperiod 2	Monday	0.1062	0.1037	0.5493	0.1738
1981-1990	Tuesday		0.9502	0.7365	0.9885
	Wednesday			0.7365	0.9502
	Thursday				0.7365
Subperiod 3	Monday	0.8794	0.6869	0.6652	0.6193
1991-2001	Tuesday		0.8686	0.8686	0.8468
	Wednesday			0.9500	0.8794
	Thursday				0.8794

Table 4: Test Results for the S&P 500.
The table presents adjusted p-values for hypotheses
$H_{ij} : \gamma_i = \gamma_j,\ 1 \leq i < j \leq 5$ for the total period and the three subperiods.
* indicates significance at the 5 % level.

		Tuesday	Wednesday	Thursday	Friday
Total period	Monday	0.0000*	0.0001*	0.0416*	0.0000*
1971-2001	Tuesday		0.5157	0.0109*	0.7674
	Wednesday			0.1489	0.5157
	Thursday				0.0197*
Subperiod 1	Monday	0.0000*	0.0040*	0.0899	0.0001*
1971-1980	Tuesday		0.1061	0.0033*	0.2402
	Wednesday			0.1674	0.3544
	Thursday				0.0540
Subperiod 2	Monday	0.0270*	0.0011*	0.0606	0.0001*
1981-1990	Tuesday		0.1924	0.7780	0.1275
	Wednesday			0.1924	0.7780
	Thursday				0.0630
Subperiod 3	Monday	0.9028	0.9901	0.9994	0.9794
1991-2001	Tuesday		0.9028	0.9084	0.9028
	Wednesday			0.9994	0.9794
	Thursday				0.9794

Table 5: Test Results for the FTSE 30.
The table presents adjusted p-values for hypotheses
$H_{ij} : \gamma_i = \gamma_j,\ 1 \leq i < j \leq 5$ for the total period and the three subperiods.
* indicates significance at the 5 % level.

		Tuesday	Wednesday	Thursday	Friday
Total period	Monday	0.0928	0.0235*	0.0530	0.0046*
1971-2001	Tuesday		0.7148	0.8427	0.3779
	Wednesday			0.7148	0.7041
	Thursday				0.3779
Subperiod 1	Monday	0.0019*	0.0083*	0.0020*	0.0001*
1971-1980	Tuesday		0.9025	0.9958	0.4348
	Wednesday			0.9025	0.3956
	Thursday				0.4352
Subperiod 2	Monday	0.4035	0.0050*	0.0223*	0.0016*
1981-1990	Tuesday		0.1021	0.3285	0.0666
	Wednesday			0.5768	0.6841
	Thursday				0.5768
Subperiod 3	Monday	0.9635	0.9002	0.8442	0.8442
1991-2001	Tuesday		0.9002	0.8442	0.8442
	Wednesday			0.9490	0.9635
	Thursday				0.9490

Table 6: Test Results for the DAX 30.
The table presents adjusted p-values for hypotheses
$H_{ij} : \gamma_i = \gamma_j$, $1 \leq i < j \leq 5$ for the total period and the three subperiods.
* indicates significance at the 5 % level.

		Tuesday	Wednesday	Thursday	Friday
Total period 1971-2001	Monday	$\surd$	$\surd$	$\surd$	0.1147 (0.0172*)
	Tuesday		$\surd$	$\surd$	$\surd$
	Wednesday			$\surd$	$\surd$
	Thursday				$\surd$
Subperiod 1 1971-1980	Monday	$\surd$	$\surd$	0.0660 (0.0126*)	$\surd$
	Tuesday		$\surd$	$\surd$	$\surd$
	Wednesday			$\surd$	$\surd$
	Thursday				$\surd$
Subperiod 2 1981-1990	Monday	0.1062 (0.0269*)	0.1037 (0.0132*)	$\surd$	0.1738 (0.0288*)
	Tuesday		$\surd$	$\surd$	$\surd$
	Wednesday			$\surd$	$\surd$
	Thursday				$\surd$

Table 7: Closure Test Principle and Conventional Testing, S&P 500.
The table shows a comparison of test results on the basis of the closure test principle (adjusted p-value) and conventional testing (traditional p-value in brackets).
* indicates significance at the 5 % level. $\surd$ indicates same result for both testing approaches.

		Tuesday	Wednesday	Thursday	Friday
Total Period 1971-2001	Monday	√	√	√	√
	Tuesday		√	√	√
	Wednesday			√	√
	Thursday				√
Subperiod 1 1971-1980	Monday	√	√	0.0899 (0.0466*)	√
	Tuesday		0.1061 (0.0346*)	√	√
	Wednesday			√	√
	Thursday				0.0540 (0.0167*)
Subperiod 2 1981-1990	Monday	√	√	0.0606 (0.0233*)	√
	Tuesday		√	√	0.1275 (0.0476*)
	Wednesday			√	√
	Thursday				0.0630 (0.0203*)

Table 8: Closure Test Principle and Conventional Testing, FTSE 30.
The table shows a comparison of test results on the basis of the closure test principle (adjusted p-value) and conventional testing (traditional p-value in brackets).
* indicates significance at the 5 % level. √ indicates same result for both testing approaches.

		Tuesday	Wednesday	Thursday	Friday
Total Period 1971-2001	Monday	0.0928 (0.0347*)	√	0.0530 (0.0219*)	√
	Tuesday		√	√	√
	Wednesday			√	√
	Thursday				√
Subperiod 1 1971-1980	Monday	√	√	√	√
	Tuesday		√	√	√
	Wednesday			√	√
	Thursday				√
Subperiod 2 1981-1990	Monday	√	√	√	√
	Tuesday		0.1021 (0.0372*)	√	0.0666 (0.0128*)
	Wednesday			√	√
	Thursday				√

Table 9: Closure Test Principle and Conventional Testing, DAX 30.
The table shows a comparison of test results on the basis of the closure test principle (adjusted p-value) and conventional testing (traditional p-value in brackets).
* indicates significance at the 5 % level. √ indicates same result for both testing approaches.

6 Conclusion

A drawback of many empirical studies of market anomalies is that the multiplicity effect is not accounted for appropriately or even ignored completely. This may result in an inflated occurrence of type I errors producing spurious significance with regard to the test results. To overcome this problem, we propose a special type of multiple test procedure based on the closure test principle introduced by Marcus, Peritz and Gabriel (1976). This procedure, which has strongly influenced the field of multiple hypotheses testing in recent decades, seems to be quite unknown in the financial literature. The underlying principle guarantees the construction of test procedures with the property that the probability of rejecting at least one of the true null hypotheses is always smaller than or equal to a given level α, independent of how many and which null hypotheses are actually true. For further results on closed test procedures see, for example, Hochberg and Tamhane (1987), Hsu (1996) or Pigeot (2000).

In our paper we apply the closure test principle to test for day-of -the-week effects in the U.S., the British and the German stock market during the period 1971 to 2001, including three subperiods. Our results confirm the existence of a day-of-the-week effect for the total and the first subperiod in all markets, and for the second subperiod in the British and German market only, though not as pronounced as documented in other studies. We find no evidence for a day-of-the-week effect during the 1990's. In contrast to other studies we examine all pairwise comparisons of expected week-day returns and do not restrict ourselves to anomalies affiliated with Monday returns. We also compare test decisions resulting from the closed test procedure and conventional testing. In some cases conventional testing rejects pairwise equalities, while according to the closure test principle the pairwise equality cannot be rejected. One has to keep in mind that conventional testing does not guarantee that type I error probabilities are less or equal than α.

A final point worth mentioning is the fact that the closure test principle is not only a generally applicable method for constructing multiple level α tests. It can also be used to characterize these procedures. Under the assumption that no primary hypothesis is a proper subset of any other, one can show that for each multiple level α test there exists a closed test procedure, such that the test decisions of both procedures coincide with respect to the primary null hypotheses $H_{01}, ..., H_{0n}$. This implies that under the assumption stated above, a test procedure keeps the multiple level α if and only if there exists an 'equivalent' closed test procedure. In particular this result holds for the case of testing the multiple comparisons H_{ij}: $\gamma_i = \gamma_j$, $1 \leq i < j \leq 5$, which we used to test for day-of-the-week effects.

References

[1] Chang, E. C.; J. M. Pinegar; and R. Ravichandran. "International Evidence on the Robustness of the Day-of-the-Week Effect", *Journal of Financial and Quantitative Analysis*, 28 (1993), 497-513.

[2] Chow, K.V, and K. Denning. "A Simple Multiple Variance Ratio Test", *Journal of Econometrics*, 58 (1993), 385-401.

[3] Cross, F. "The Behaviour of Stock Prices on Friday and Monday", *Financial Analysts Journal*, 29 (1973), 67-69.

[4] Connolly, R.A. "A Posterior Odds Analysis of the Weekend Effect", *Journal of Econometrics*, 49 (1991), 51-104.

[5] Dubois, M., and P. Louvet. "The Day-of-the-Week Effect: The International Evidence", *Journal of Banking and Finance*, 20 (1996), 1463-1484.

[6] French, K. R. "Stock Returns and the Weekend Effect", *Journal of Financial Economics*, 8 (1980), 55-69.

[7] Greenstone, M., and P. Oyer. "Are There Sectoral Anomalies Too? The Pitfalls of Unreported Multiple Hypothesis Testing and a Simple Solution", *Review of Quantitative Finance and Accounting*, 15 (2000), 37-55.

[8] Hochberg, Y., and A.C. Tamhane. *Multiple Comparison Procedures*, New York: John Wiley (1987).

[9] Hsu, J.C. *Multiple Comparisons*, London: Chapman & Hall (1996).

[10] Huber, P. "Stock Market Returns in Thin Markets: Evidence from the Vienna Stock Exchange", *Applied Financial Economics*, 7 (1997), 493-498.

[11] Jaffe, J. F., and R. Westerfield. "The Week-End Effect in Common Stock Returns: The International Evidence", *Journal of Finance*, 40 (1985), 433-454.

[12] Keim, D. B., and R.F. Stambaugh. "A Further Investigation of the Weekend Effect in Stock Returns", *Journal of Finance*, 39 (1984), 819-835.

[13] Madlener, R., and R. Alt. "Residential Energy Demand Analysis: An Empirical Application of the Closure Test Principle", *Empirical Economics*, 21 (1996), 203-220.

[14] Marcus, R.; E. Peritz; and K.R. Gabriel. "On Closed Testing Procedures with Special Reference to Ordered Analysis of Variance", *Biometrika*, 63 (1976), 655-660.

[15] Neusser, K. "Testing the Long-Run Implications of the Neoclassical Growth Model", *Journal of Monetary Economics*, 27 (1991), 3-37.

[16] Pigeot, I. "Basic Concepts of Multiple Tests - A Survey", *Statistical Papers*, 41 (2000), 3-36.

[17] Savin, N.E. "Multiple Hypothesis Testing." In: *Handbook of Econometrics* (1984), Vol. II, Z. Griliches and M.D. Intriligator, eds. Amsterdam: North-Holland.

[18] Sullivan, R.; A. Timmermann; and H. White. "Dangers of Data Mining: The Case of Calendar Effects in Stock Returns", *Journal of Econometrics*, 105 (2001), 249-286.

Authors: Raimund Alt, Ines Fortin, Simon Weinberger

Title: The Day-of-the-Week Effect Revisited: An Alternative Testing Approach

Reihe Ökonomie / Economics Series 127

ISSN: 1605-7996

Stumpergasse 56, A-1060 Vienna • ☎ +43 1 59991-0 • Fax +43 1 59991-555 • http://www.ihs.ac.at

ISSN: 1605-7996