



Stock returns and investors' mood: Good day sunshine or spurious correlation?☆



Jae H. Kim

Department of Economics and Finance, La Trobe University, VIC 3086, Australia

ARTICLE INFO

Article history:

Received 27 February 2017

Accepted 23 May 2017

Available online 24 May 2017

JEL classification:

G12

G14

Keywords:

Anomaly

Data mining

Market efficiency

Sunspot numbers

Weather effect

ABSTRACT

This paper critically evaluates the significant weather effect on stock return reported in two seminal studies of investors' mood on stock market. It is found that their research design of maximizing statistical power by pooling as many data points as possible is statistically flawed, with a consequence that the test is severely biased against the null hypothesis of no effect. Coupled with small effect size estimates and test statistics inflated by massive sample sizes, this strongly suggests spurious statistical significance as an outcome of Type I error. The alternatives to the p -value criterion for statistical significance soundly support the null hypothesis of no weather effect. As an application, the effect of daily sunspot numbers on stock return is examined. Under the same research design as that of a seminal study, the number of sunspots is found to be highly statistically significant although its economic impact on stock return is negligible.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

The question as to whether investors' mood affects the stock market (i.e. their emotional states or feelings unrelated to market fundamentals or rational pricing of financial assets) has been an issue of considerable interest in economics and finance (see, for a survey, Lucey & Dowling, 2005). The seminal studies in this literature are Saunders (1993) and Hirshleifer and Shumway (2003) where statistically significant weather effects on stock return are reported. Subsequent studies overall support the existence of the weather effect (cloudiness, sunshine, temperature, or wind) on stock return or other trading activities: see Cao and Wei (2005), Dowling and Lucey (2005, 2008), Goetzmann and Zhu (2005), Chang, Chen, Chou, and Lin (2012), Chang, Nieh, Yang, and Yang (2006), Keef and Roush (2002, 2005, 2007), Yoon and Kang (2009), Kang, Jiang, Lee, and Yoon (2010), Lee and Wang (2011), Lu and Chou (2012), and Novy-Marx (2014). The literature has proliferated over the years in the publication of studies examining the effects of investors' moods derived from disparate sources such as: daylight saving (Kamstra, Kramer, & Levi, 2000), seasonal depression (Kamstra, Kramer, & Levi, 2003), sports events (Chang et al., 2012; Edmans,

García, & Norli, 2007; Kaplanski & Levy, 2010), lunar phases (Yuan, Zheng, & Zhu, 2006; Keef & Khaled, 2011), pollution (Lepori, 2015), and Ramadan (Bialkowski, Etebari, & Wisniewski, 2012). Most of these studies report statistically significant effects of investors' mood on the stock market, and their findings are presented as direct evidence for the anomalies against market efficiency.

On the other hand, there are studies that raise suspicions that a statistically significant weather effect may be the result of data mining or spurious correlation. In replicating Saunders' (1993) results using a German data set, Krämer and Runde (1997) report that statistical significance of the weather effect depends largely on how the null hypothesis is phrased. Trombley (1997) provides evidence that Saunders' (1993) results depend on the type of the return used and sample period employed. Loughran and Schultz (2004), in the context of localized trading of NASDAQ stocks, examine the weather effect in the city where the company is based and find that the weather effect is too slight to establish a profitable weather-based trading strategy. They (p.363) state that "we would not dismiss the possibility that the relationship between cloud cover in New York and stock returns is spurious". Jacobsen and Marquering (2008) argue that the documented weather effects might be the consequence of "data-driven inference based on spurious correlation". They provide evidence that the seasonal anomaly in stock return is unlikely to be caused by investors' mood changes due to weather variations, stating that it is premature to conclude that weather has an effect on stock return through mood changes of investors. In re-evaluating the effect of seasonal depression, Kelly and

☆ I would like to thank Peter Luc Bossaerts, Geoff Cumming, Patrick Kelly, Liam Lenten, Abul Shamsuddin, and Xiangkang Yin for their constructive comments on an earlier version of the paper. An earlier version of the paper was presented at the 2016 Australian Finance and Banking Conference.

E-mail address: j.kim@latrobe.edu.au.

Meschke (2010) document that the observed effect is mechanically driven by an overlapping dummy-variable specification.

The purpose of this study is to critically evaluate the statistical significance reported in two seminal studies of the weather effect on stock return, i.e. Saunders (1993) and Hirshleifer and Shumway (2003), in order to shed light on the possibility of a spurious statistical significance between investors' mood and stock return. First, paying attention to Hirshleifer and Shumway (2003), I evaluate whether their research design that “maximizes the power of the test by pooling the all available data jointly” is statistically sensible. This is important since many subsequent studies in this area adopt large or massive sample sizes in the same spirit. However, there is a danger that the use of a massive sample size produces spurious statistical significance (see, for example, McCloskey & Ziliak, 1996, p.101; Lockett, McWilliams, & Van Fleet, 2014, p.865; and Kim & Ji, 2015). Second, statistical significance reported in these seminal studies is re-evaluated using the Bayesian method (Zellner & Siow, 1980) and the adaptive level of significance (Perez & Pericchi, 2014). These are the alternatives to the p -value criterion for statistical significance that are exclusively adopted in the prior studies. Note that the American Statistical Association (Wasserstein & Lazar, 2016) recently issued a statement expressing grave concerns that improper use of the p -value criterion is distorting the scientific process and invalidating many scientific conclusions. Third, as an application, the effect of sunspot numbers on stock return is examined, under the same research design as that of Hirshleifer and Shumway (2003). This is to demonstrate that a variable with little economic relevance on stock return can be shown to be statistically significant through a simple data mining process.

The main finding of the paper is that statistically significant weather effects reported in the past studies is highly likely to be spurious and an outcome of Type I error. In particular, the research design employed by Hirshleifer and Shumway (2003) is problematic, with the probability of Type I error disproportionately higher than that of Type II. It is severely biased against the null hypothesis of no effect in its implied specification of loss function and prior probabilities. The alternatives to the p -value criterion show an overwhelming support for the null hypothesis of no weather effect. It is also demonstrated that a balanced research design in the context of the data employed by Hirshleifer and Shumway (2003) requires a sample of size < 2000 . The results from the empirical application further confirm these findings. While suspicions concerning spurious statistical significance of the weather effect on stock return have been raised previously, this study is the first to assess the underlying statistical issues in the research design of the seminal studies and re-evaluates statistical significance of their results. In the next section, the research design of Hirshleifer and Shumway (2003) is examined. Section 3 presents further analyses based on the alternative criteria for statistical significance; and discussion on the effect size estimates reported in the past seminal studies. Section 4 presents the empirical application, and Section 5 concludes the paper.

2. Issues related with the research design

In this section, the research design of Hirshleifer and Shumway (2003) is discussed with reference to the possibility of spurious statistical significance. The weather effect is typically tested in the regression model of the form:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \dots + \beta_K X_{Kt} + u_t, \quad (1)$$

where Y is the stock return (in percentage), X_1 is a weather variable (e.g. cloud cover), and other X 's represent the possible control variables. Hirshleifer and Shumway (2003) also consider a logit model where Y is an indicator variable. Under $H_0: \beta_1 = 0$, the weather has no effect on the stock return. The t -test statistic can be written as $t = \frac{b_1}{s/\sqrt{n}}$, where b_1 is the least-squares estimator for β_1 , n is the sample size,

and s/\sqrt{n} denotes the standard error of b_1 . Throughout the paper, following convention, the regression parameters are denoted as β_i 's; and the probability of Type I and II errors as α and β , respectively (the power $= 1 - \beta$).

2.1. Background

One common feature of the studies of the weather effect is the use of large or massive sample sizes. My survey of 20 papers in this literature finds that the average sample size used is around 6000 with the maximum being 92,808. In addition, the past studies conduct their statistical tests almost exclusively at the conventional level of significance such as 0.05. A number of authors warn that spurious statistical significance may occur in this scenario (Neal, 1987, p. 524; Connolly, 1989, p. 139; McCloskey & Ziliak, 1996, p.101; Lockett et al., 2014, p.865; and Kim & Ji, 2015). However, many researchers seem to believe that a large or massive sample size is necessarily a desirable feature of a research design, delivering strong power to their statistical tests. Hirshleifer and Shumway (2003, p.1014) justify their use of a massive panel data set, asserting that “the panel increases our power to detect an effect. ... Given high variability of returns, it is useful to maximize power by using a large number of markets”. However, as we shall see, this can cause statistical inference severely biased towards Type I error. The extreme power leads to an acute imbalance between α and β if a conventional level of significance is maintained. For example, suppose an extreme power ($1 - \beta$) of 0.99999 is achieved by pooling a massive panel data set. If the researcher conducts a test at the 5% level ($\alpha = 0.05$), the Type I error is 5000 times more likely to occur than the Type II error. As a result, if an error occurs, it is highly likely to be that of Type I, rejecting the true null hypothesis of no effect. This is particularly so when the effect size (e.g. the magnitude of the regression coefficient β_1 in Eq. (1)) is small.

Note that the null hypothesis is often violated by an economically trivial deviation (see Hodges & Lehmann, 1954; De Long & Lang, 1992). It is unrealistic that the null hypothesis of no effect holds exactly in practice. In reality, a null hypothesis is violated by a negligible margin even when the true effect is economically unimportant. That is, $\beta_1 = 0 + \Delta$, where Δ represents a deviation from the null hypothesis. As De Long and Lang (1992, p. 1269) find, all economic hypotheses are false with $\Delta \neq 0$. For example, in the context of market efficiency, Grossman and Stiglitz (1980) show that a perfectly efficient market ($\Delta = 0$) is impossible because if prices fully reflect all available information, traders would not have any incentive to acquire costly information. Rather, it is widely accepted that market efficiency is relative (Campbell, Lo, & MacKinlay, 1997), and that the degree of market efficiency depends on the prevailing market conditions (Lo, 2004).¹ The key question in empirical research is whether the value of Δ is large enough to be economically meaningful (see McCloskey & Ziliak, 1996). An important point is, even when the value of Δ is economically unimportant, a test statistic (in absolute value) approaches infinity as the sample size increases. In this case, if a fixed level of significance is maintained, the probability of rejecting the null hypothesis approaches one as the sample size increases (see, for details, Kim & Ji, 2015, Section 5.1).

Moreover, as we shall see in Section 3.2, the effect sizes reported in published seminal studies on the weather effect are indeed fairly small, which strongly suggests that the values of Δ are economically negligible. Hence, one may validly suspect that the statistically significant relationship between investors' mood and stock return reported in many studies are spurious, on the basis that their significance testing is conducted at a fixed conventional level (such as 0.05) under a large or massive sample size. A sensible strategy in this case is to adjust the level of significance as a decreasing function of sample size so that a reasonable balance between α and β is maintained (see, for example, Arrow,

¹ McCloskey and Ziliak (1996, p.98) also provide a similar example in the context of the purchasing power parity.

Download English Version:

<https://daneshyari.com/en/article/5084420>

Download Persian Version:

<https://daneshyari.com/article/5084420>

[Daneshyari.com](https://daneshyari.com)