JEL Classification: C58, G1, G12, G14
Keywords: U.S. equity exchange-traded funds, overnight and daytime returns, day-of-the-week effect, mean-variance, sharpe ratio, stochastic dominance

Revisiting Seasonality in Overnight and Daytime Returns in the U.S. Equity Markets: Mean-Variance, Sharpe Ratio and Stochastic Dominance Approaches*

João Dionísio MONTEIRO - Department of Management and Economics, NECE-UBI, University of Beira Interior, Covilhã, Portugal (jdm@ubi.pt), *corresponding author*

Ernesto Raúl FERREIRA - Department of Management and Economics, NECE-UBI, University of Beira Interior, Covilhã, Portugal (ferreira@ubi.pt)

Abstract

This paper examines the existence of the day-of-the-week effect in overnight and daytime period returns in a group of broad-index exchange-traded funds (ETFs) that track the major U.S. stock indexes (S&P 500 and NASDAQ 100 indices) over the period from 1996 to 2018. Previous empirical studies suggest that the positive overnight minus daytime mean return spread could be of economic significance. However, empirical evidence is not entirely consistent across studies. To examine this effect, we use various inference procedures: the mean-variance (MV), Sharpe ratio (SR), and stochastic dominance (SD) approaches. The MV and SR results suggest a decrease or even the disappearance of the positive overnight minus daytime mean return spread. The SD results show that overnight periods do not dominate and are not stochastically dominated by daytime period returns, in the sense of first-order SD. These SD findings suggest that no arbitrage opportunities exist in U.S. equity markets and investors could not increase their wealth and expected utilities by switching from any daytime to overnight periods, or vice versa, over weekdays. Overall, the results suggest that information impounding mechanisms have become more efficient in U.S. markets.

1. Introduction

As trading can be perceived as a continuous-time process throughout the entire day, theoretical and empirical studies have focused their attention on decomposing daily (close-to-close) returns into overnight (close-to-open) and daytime (open-to-close) periods and examining the implications of periodic market closure for trading and returns (Hong and Wang, 2000; Barclay and Hendershott, 2003; Branch and Ma, 2006; Cliff et al. 2008; Kelly and Clark, 2011; Berkman et al., 2012; Lachance, 2015; Lou et al., 2018). The evidence from these and other studies suggests statistically higher returns during the overnight period. In addition, empirical studies document that the second- and higher-order moments of the return-generating process are different over daytime and overnight periods (Cliff et al. 2008; Tompkins and Wiener, 2008) and that risk-adjusted overnight returns are significantly higher than risk-adjusted daytime

*The authors acknowledge financial support of the NECE-UBI - Research Unit in Business Science and Economics, funded by the Portuguese Foundation for the Development of Science and Technology, Project no. UID/GES/04630/2019. The authors are also grateful to the two anonymous reviewers for their invaluable

returns (Kelly and Clark, 2011). Furthermore, empirical studies suggest that these differences may have economic significance through the implementation of profitable trading strategies (Kelly and Clark, 2011; Lachance, 2015).

Thus, to the best of our knowledge, a comprehensive analysis using a longer period of data and robust approaches regarding day-of-the-week effects during daytime and overnight periods in the U.S. equity market remains to be done. Indeed, most previous studies on calendar effects, mainly using close-to-close daily returns, have employed the mean-variance (MV) criterion, capital asset pricing model (CAPM) statistics, or regression-based methods. These approaches use parametric statistics that rely on the normality assumption and depend primarily on the first two moments of return distributions. However, it has been well established that individual and portfolio stock distributions exhibit significant skewness and kurtosis (Schwert, 1990). Using these approaches to test for calendar effects in daytime and overnight returns could miss important information contained in higher moments.

Given that the aforementioned evidence points to the need for further research, this study attempts to verify this effect and examine its robustness using the framework of calendar effects. We ground our study in the rationale of market efficiency, and aim to add to the field of market efficiency an analysis of calendar effects during daytime and overnight period returns in the U.S. equity market of exchange-traded funds (ETFs). Although much less scrutinized with respect to daytime and overnight periods, and specifically within the ETF market, the U.S. equity markets have been extensively addressed in calendar effect studies (Pettengill, 2003).

Several empirical studies have employed a nonparametric stochastic dominance (SD) approach to examining market efficiency (Fong et al., 2005; Lean et al., 2007; Bai et al., 2015). Contrary to traditional parametric approaches, the SD approach in comparing portfolios is equivalent to the choices of assets by utility maximization. This approach guarantees the minimum assumptions of the investor's utility function and examines the entire distribution of returns. The SD analysis is superior to parametric tests because it is not distribution-type dependent, as the SD approach does not require any assumption about the nature of the distribution. When examining the entire distribution, the SD approach incorporates information about higher moments in the analysis, whereas traditional parametric tests, depending on the mean and variance, omit this information.

Thus, we examine the day-of-the-week effect in daytime and overnight returns in a set of ETFs that track the major U.S. stock indexes during the period 1996-2018. With the purpose of comparing current results with those of previous studies and to examine whether there is persistence, a decrease, disappearance or even the absence of the effect, we divide the entire period into two subperiods: 1996-2006 and 2007-2018. Three statistical procedures are used. First, over the two subperiods and for comparative purposes, we examine the day-of-the-week effect on overnight and daytime returns using the MV criterion. Second, we employ the Sharpe ratio (SR) statistic, using asymptotic distributions that are valid under very general conditions (i.e., stationary and ergodic returns). In the MV and SR approaches, we apply multiple testing procedures to control for the false discovery rate (FDR) in the significant dominances obtained. Third, we employ the SD test proposed by Davidson and Duclos (2000; hereinafter, the DD test) to examine the effect during daytime and overnight periods.

The remainder of the paper is organized as follows: Section 2 reviews theoretical causes, predictions, and empirical evidence for the behaviour of returns during daytime and overnight periods. We also review some empirical evidence on calendar effects during daytime and overnight periods. Section 3 presents the data; the MV, the SR, and the SD approaches; and the corresponding statistical tests. We present and discuss the findings in Section 4 and summarize them along with some concluding remarks in Section 5.

2. Literature Review

Across financial markets, information flows continuously around the clock, but price variations and trading are not continuous due to the periodic market closure. Changes in daily transaction regimes, when markets open and close, can have important implications for the return-generating process over daytime and overnight periods. Empirical studies have reported that the mean return, the trading volume, the volatility, and the bid-ask spreads in general have a U-shaped pattern during the intraday period across developed stock markets, with these variables being high at the open and close of the market and relatively flat during the middle of the intraday period (Wood et al., 1985; McInish and Wood, 1992; Foster and Viswanathan, 1993; Abhyankar et al., 1997; Hong and Wang, 2000; Chow et al., 2004). There is less consensus, however, about the behaviour of the mean return during daytime and overnight periods.

Theoretical papers have sought to model the implications of periodic market closure for price equilibrium (Longstaff, 1995; Hong and Wang, 2000). Hong and Wang's (2000) model predicts lower returns during overnight periods than during daytime periods, a prediction consistent with the observed higher volatility and information flow rates during daytime periods. Conversely, Longstaff's (1995) model predicts higher returns during overnight periods than during daytime periods to compensate liquidity providers for bearing additional risk (i.e., higher returns over overnight periods arise from a liquidity-related nonmarketability effect).

Wood et al. (1985) examine return patterns around the open and the close of the market. They document that the return and volatility are unusually high at the open and close of the daytime period. French and Roll (1986) document that stock returns are more volatile during the daytime period than during the overnight period, attributing the higher volatility during intraday hours to the differences in information flow rates between the two periods. Harris (1989) documents a large mean price increase at the market's open and before market closure, and this effect is persistent across stocks and days. George and Hwang (2001) examine the rate of information flow and find that the daytime information rate is about seven times higher than the overnight rate. Barclay and Hendershott (2003) find that there is less information asymmetry in the post-close period than in the preopen period of the market. Their findings suggest that there will be a higher fraction of liquidity-motivated trades in the post-close period and a higher fraction of informed trades in the preopen period.

With respect to return patterns over daytime and overnight periods, empirical evidence is not consistent across empirical studies. French (1980) first identified the weekend effect using U.S. daily stock returns from 1953 to 1977. French finds a weekend effect such that Monday's mean return is significantly negative, while the other day-of-the-week returns are significantly positive. Rogalski (1984) examines the

U.S. stock market from 1974 to 1984 to determine whether the weekend effect is a closed market effect by decomposing daily close-to-close returns into an overnight and daytime return. Rogalski finds that the negative weekend return is composed of a negative Monday overnight return (Friday close to Monday open) and a Monday daytime return (Monday open to close) identical to the daytime returns of other weekdays.

Cliff et al. (2008), using data sets of different asset classes for the 1993–2006 period, performed an extensive study in U.S. equity markets on overnight and daytime returns. They document that the U.S. equity premium during this decade is entirely due to overnight returns: returns during the night are strongly positive, and returns during the day are close to zero and sometimes negative. They also show that this daytime and overnight effect exists for individual equities, equity indexes, ETFs, and futures contracts on equity indexes.

Tompkins and Wiener (2008) examine returns for five global index futures markets over daytime and overnight periods. They find significant differences between daytime and overnight period returns. For the U.S. market, the mean return is higher for the daytime period than for the overnight period, with the overnight period showing significantly lower variance. For the four non-U.S. stock markets, the overnight period return is significantly higher than that of the daytime period. They attribute this positive overnight minus daytime return spread to differences in regulatory risk management requirements between U.S. and non-U.S. equity derivative market makers.

Kelly and Clark (2011) compare the intraday and overnight returns on a set of U.S. equity ETFs. Using the SR statistic, they find the overnight SR to significantly exceed the intraday SR, implying that the premium one receives by taking on risk is higher during the overnight period than during the daytime period. Berkman et al. (2012) examine intraday patterns in retail order flow and price formation in a sample of the largest U.S. stocks. Based on the theory of attention-based overpricing at the opening of the market, they report the existence of the overnight effect confined to a large U.S. stock group and attribute the significantly positive (negative) overnight (daytime) return to the trading activity of retail investors. Qiu and Cai (2013) examine the anomaly of superior overnight returns on international stock markets. Using stock index data for 32 countries, they find that the anomaly exists in 20 countries, including both developed and emerging markets, and that the superior overnight returns are not justified by the risk-return trade-off, as overnight returns are less volatile than daytime period returns. Using all listed U.S. stocks in the period 1995-2014, but not including ETFs, Lachance (2015) finds evidence that overnight returns are subject to highly persistent and positive biases in a large group of stocks. Lou et al. (2018) find that abnormal returns related to momentum portfolios are present overnight but not during the day while abnormal returns related to size and value portfolios occur only during the day.

Several arguments have been forwarded to explain this overnight effect, namely, the timing of earning announcements, asset liquidity, and investor trading heterogeneity. The timing-of-earnings-announcement hypothesis suggests that managers tend to disclose good news during the overnight period, particularly before the opening of the markets. However, empirical evidence is not consistent with this hypothesis (Patell and Wolfson, 1982; Damadoram, 1989; Bagnoli et al., 2005; Doyle and Magilke, 2009). The asset liquidity hypothesis (Amihud, 2002) suggests that the higher (lower) risk or transaction costs of low-liquidity (high-liquidity) stocks predict a greater (lower) overnight minus daytime return spread. However, evidence does not generally support this hypothesis (Cliff et al., 2008). Investor trading heterogeneity during daytime and overnight periods has also been suggested as a contributor to the effect. Barclay and Hendershott (2003) report that there is a higher fraction of liquidity-motivated trades in the post-close period and a higher fraction of informed trades in the preopen period, because trading in the preopen period is dominated by more informed investors. Kelly and Clark (2011) attribute the overnight effect to the behaviour of active day (semi-professional) traders. Berkman et al. (2012) and Lou et al. (2018) suggest that the trading activity of individual investors plays an important role in explaining higher overnight returns due to their herding behaviour in the high-attention stock group, which pushes opening prices up.

Overall, the results of the empirical studies we have examined are not entirely consistent. Some point to the existence of higher overnight returns across all the individual assets of the sample, others report that the effect is confined to a group of assets, and still others report an inverse effect. In addition, the evidence we have reviewed suggests that findings regarding the daytime and overnight period returns are not entirely consistent across markets, or even within the same market, and that these results might be sample-period-dependent and asset-specific. These empirical findings motivate us to carry out the present study to further investigate overnight and daytime effects in a set of ETFs that track major indices of the U.S. equity market, examining whether, in the day-of-the-week-effect context, these effects are actually manifest or if they diminished and disappeared.

3. Data and Methodology

3.1 Data

The data we employ in this study are actual opening and closing daily prices from a group of two ETFs that track major U.S. equity market indices. The two ETFs we use are the Standard and Poor's Depository Receipts (SPDRs or 'spiders' – ticker SPY, representing the S&P 500 index) and the Invesco PowerShares QQQ Trust (representing the NASDAQ 100 index, ticker QQQ). The data series were collected from Datastream. ETFs allow investors to trade a basket of stocks in a single transaction. ETFs shares, each representing a basket of stocks, trade in the secondary market just like ordinary shares. The creation and destruction features of ETFs ensure that prices on the exchange closely reflect the fair value of the underlying portfolio's components.

Analysis of ETFs' returns offers advantages over analysis of the indices' returns for two reasons. First, the share price of an ETF is the price for the entire portfolio, with no problem of asynchronous transactions on certain stocks in the index. Second, ETFs that track major stock market indices are highly liquid, with very low transaction costs (bid-ask spreads) involved in the trading of these instruments. In addition, two specific and useful features of ETFs are that the transaction is an in-kind trade (i.e., securities are traded for securities) and they are generally more tax efficient.

Kelly and Clark (2011) previously used the same ETF sample in their analysis of overnight and daytime SR over the sample period 1996-2006, but did not analyse the day-of-week effect. This set of ETFs began trading on the exchanges during different years. The SPY started trading in 1993, but its liquidity was poor during the first half of 1990s. To determine the starting point of the analysis, we follow the liquidity criterion used by Kelly and Clark. Thus, SPY time-series data are used from 01/02/1996 (mm/dd/yyyy) and the QQQ time-series data are used from 03/11/1999 to 12/31/2018.

For each ETF, we compute returns during the two daily subperiods: the overnight (close-to-open prices) and the daytime (open-to-close prices) returns. As in most of the analysis of daily and intradaily financial data, we work with continuously compounded returns, and we compute the overnight and daytime returns, respectively, as follows:

$$r_t^n = ln[P_t^o / P_{t-1}^c].100\%, \quad (1.1)$$

$$r_t^d = ln[P_t^c / P_t^o].100\% \quad (1.2)$$

where P_t^o is the ETF price level at the open of day t , P_t^c is the ETF price level at the close of day t, and P_{t-1}^c is the ETF price level at the close of day $t-1$. The average returns are geometric averages, and therefore, the sign indicates whether the ETF gained or lost value during these intraday ranges over the sample period. For each ETF return time series, we computed overnight and daytime returns for each day of the week. We computed Monday overnight returns as Friday close to Monday open, Monday daytime returns as Monday open to close, Tuesday overnight returns as Monday close to Tuesday open, Tuesday daytime returns as Tuesday open to close, and so on. We excluded any week with fewer than five trading days[1] from the study in order to fulfil the requirement of equal sample size across days of the week. Returns over the extended close following the September 11, 2001 tragedy were not taken into account. We formed the portfolio of each weekday subperiod (overnight and daytime) by grouping the returns of the same weekday subperiod over the entire sample period. For each ETF, after forming the portfolio of each weekday subperiod, we performed pairwise comparisons between all these portfolios, using inference procedures using three competing approaches to examine the hypothesis of overnight and daytime effects on U.S. equity ETFs: the MV, SR, and SD approaches.

3.2 The Mean-Variance (MV) Approach

The MV approach involves conducting a descriptive analysis and parametric tests of overnight and daytime returns by day of the week. For each ETF and among all daily subperiod returns of the week, we perform tests of equality of means and variances using parametric tests. Parametric testing is suitable because for large samples, sample means will be normally distributed even if the underlying variables are not normally distributed. In addition, parametric tests have more statistical power than their nonparametric counterparts. Next, we describe how we apply the MV criterion.

[1] We excluded weeks that included the following holidays: New Year Day, Martin Luther Jr. King Day (the third Monday in January), President's Day (the third Monday in February), Good Friday (Easter), Memorial Day (the last Monday in May), Independence Day (July 4), Labor Day (the first Monday in September), Thanksgiving Day (the fourth Thursday in November), and Christmas Day.

For any two risky portfolios with returns Y_i and Z_j with means μ_i and μ_j and standard deviations σ_i and σ_j, respectively, for risk-averse investors it is stated that Z_j dominates Y_i by the MV criterion, denoted as $Z_j \succ Y_i$, if $\mu_j \geq \mu_i$ and $\sigma_j \leq \sigma_i$, and the inequality holds in at least one of the two (Markowitz, 1952). In addition, for risk-averse investors, SD is equivalent to MV efficiency when the variables are normally distributed (Markowitz, 1952; Tobin, 1958). Wong (2007) also proved that if both Y_i and Z_j belong to the same location-scale family of distributions or the same linear combinations of location-scale families, $Z_j \succ Y_i$ implies that $E[u(Z_j)] \geq E[u(Y_i)]$, where $u' \geq 0, u'' \leq 0$ for any risk-averse investor.

3.3 The Sharpe Ratio (SR) Approach

The SR approach involves computing the SR for each ETF for overnight and daytime returns for each day of the week, testing statistical significance, and making pairwise comparison inferences among all daily subperiod returns of the week. The MV model and the SR metric are frequently used methods to evaluate the performance of investments. The consistency of these criteria with expected utility theory depends on the existence of normal return distributions and investors having preferences according to quadratic utility functions (Feldstein, 1969; Hakansson, 1972) or on both portfolio returns belonging to the same location-scale family or the same linear combinations of location-scale families (Wong, 2007). However, quadratic utility functions have undesirable features – namely, they exhibit increasing absolute and relative risk-aversion functions that are not consistent with empirical evidence on investor behaviours (Graves, 1979; Guiso and Paiella, 2008; Chiappori and Paiella, 2011). Nevertheless, given the broad use of the MV and SR in the evaluation and ranking of investments, we also use SR in this study.

This SR statistic is a risk-adjusted performance measure that assesses the average excess return (beyond a risk-free rate) of a portfolio relative to its volatility, as measured by its standard deviation. In the SR calculation, we use the following sample counterpart,[2] dividing the average risk premium by the volatility of the risk premium (Sharpe, 1966, 1994):

$$\widehat{SR} = \frac{\hat{\mu}_e}{\hat{\sigma}_e} \qquad (2)$$

where $\hat{\mu}_e$ is the sample mean of the excess returns of a portfolio beyond some risk-free rate (r_f), $\hat{\mu}_e = \sum_{t=1}^{T} r_{et}/T$, $r_{et} = (r_t - r_{ft})$, and $\hat{\sigma}_e = \left(\sum_{t=1}^{T}(r_{et} - \hat{\mu}_e)^2/(T-1)\right)^{1/2}$ is the sample standard deviation of the excess returns. The debate over the importance of this statistic for the evaluation of investment performance has been extensive. This

[2] In the use of this statistic, other versions are used in which it is assumed that although r_f is not literally constant during the period used to calculate SR, it has a very small variance in relation to risky investment, with their mean being treated as its constant value (Lo, 2002; Christie, 2005). Given that the risk-free rate proxy used in the present study varies over time, we chose to use the version of equation (2) instead of using the average risk-free rate over the period. This version does not influence calculations and inference because its covariance with overnight and daytime returns is null.

metric tends to be useful[3] although much of the research uses asymptotic distributions of $\widehat{SR}$ based on unsuitable assumptions (i.e., using independent and identically distributed [i.i.d.] normality of returns [Jobson and Korkie, 1981]). Opdyke (2007) derives the asymptotic distribution of $\widehat{SR}$ that generalizes the i.i.d. requirement of Mertens (2002) and simplifies the more complex formula of the asymptotic distributions of Christie (2005), making it more mathematically tractable and far easier to calculate and implement when conducting hypothesis tests. Opdyke (2007) demonstrates that his estimators, under real-world conditions of returns, show a reasonable level of control and good power for sample sizes up to 300 time periods.

We perform two statistical tests. First, for each ETF and for each day-of-the-week and daily subperiod return, we test the significance of the SR. This test enables us to examine whether respective daytime and overnight returns earned a significant, positive or negative, risk premium. Second, for each ETF and among days of the week and daily subperiods, we test whether SRs between any two portfolios are significantly different. This test enables us to examine whether a daytime or overnight period return earned a significantly higher risk-adjusted return than another daytime or overnight period. For this purpose, we use Opdyke's (2007) asymptotic distributions, which we present next.

3.3.1 Asymptotic Distributions for One- and Two-Sample Estimators for SR

To test the hypothesis of whether the SR of one portfolio is statistically significant – that is, $H_0: SR = 0$ versus $H_a: SR \neq 0$ – we use the following asymptotic distribution of $\widehat{SR}$, which is valid under very general conditions (i.e., stationary and ergodic returns):

$$\sqrt{T}(\widehat{SR} - SR) \sim N\left(0; 1 + \frac{SR^2}{4}\left[\frac{\mu_4}{\sigma^4} - 1\right] - SR\frac{\mu_3}{\sigma^3}\right) \tag{3}$$

where the estimated standard error is computed with respective sample counterparts.

To test the hypothesis of whether the SR of portfolio Y is significantly different from the SR of portfolio Z – that is, $H_0: SR_Y = SR_Z$ versus $H_a: SR_Y \neq SR_Z$ – we use the following asymptotic distribution of the difference between two SRs:[4] $\widehat{SR}_{diff} = \left(\widehat{SR}_y - \widehat{SR}_z\right) - (SR_y - SR_z)$,

$$\sqrt{T}\left(\widehat{SR}_{diff}\right) \sim N\left(0; Var_{diff}\right).^{5} \tag{4}$$

[3] Eling and Schuhmacher (2007) present strong evidence, even under highly non-normal return conditions, in support of the SR approach compared with other more complex risk-adjusted performance metrics.

[4] Opdyke (2007) only rigorously proves the validity of the variance formula for the two-sample test in the i.i.d. general case. Because this distribution was derived using the delta method, the same as that used to derive the distribution for the one-sample test that proved to be identical to Christie's (2005) more generally valid generalized method of moments, Opdyke conjectures that it is also valid under the more general conditions of stationarity and ergodicity.

[5] The difference variance formula of the asymptotic distribution of the difference between two SRs is presented in appendix.

For each ETF, day-of-the-week and daily subperiod respective returns are converted to risk premiums by subtracting a risk-free interest rate proxy obtained from the Federal Reserve Economic Data available at the St. Louis Federal Reserve website. The interest rate used is the 3-month US Treasury Bill daily rate. The number of days of interest subtracted from the daily subperiod returns is determined by the difference between the trading day and the settlement date, as payment for purchases and proceeds from sales are due on settlement date. We assume that transactions made on Monday and Tuesday have three calendar days of interest subtracted and that transactions made on Wednesday, Thursday, and Friday have five calendar days of interest subtracted.[6] The equivalent daily risk-free rate that is subtracted is calculated as $r^f_{daily,t} = Ln\left(1 + r^f_{annum,t}\right)/365$. Only the overnight returns have the risk-free rate subtracted. The daytime returns, which have both transactions in the same day, have the same settlement date. Because two offsetting trades with the same settlement date do not require funding, the realized daytime return is equal to the realized daytime risk premium.

3.4 The Stochastic Dominance (SD) Approach

3.4.1 Definitions and properties of SD

Consider that there are two investments with random returns Y and Z, with the corresponding cumulative distribution functions (CDFs) to be $F_Y(y)$ and $G_Z(z)$, respectively, with common support of $[a, b]$, where $a < b$. The respective expected returns on investments Y and Z are

$$\mu_Y \equiv E(Y) = \int_a^b y dF(y) \text{ and } \mu_Z \equiv E(Z) = \int_a^b z dG(z). \quad (5)$$

For each investment, the following expressions are designated as the corresponding higher-order cumulative functions:

$$H_j(x) = \int_a^x H_{j-1}(y) dy, \quad j = 1,2,3, \quad (6)$$

where $H_0(x)$ is the probability density function and $H_1(x)$ the CDF with $H = F$ or G. The definition of H_j in equation (6) is used to develop SD theory for risk-averse investors, and H_j is designated as j^{th}-order cumulative probability, where H_j is obtained by integrating H_{j-1} in ascending order from the leftmost point of the downside risk. Typically, risk-averse investors prefer assets that have a smaller probability of loss. To decide between two investments Y and Z, investors will compare their corresponding j^{th} orders on the integrals F_j and G_j and choose Y (Z) if F_j (G_j) is smaller because it has a lower probability of loss. The stochastic dominance definitions for risk-averse investors are as follows:

[6] See https://www.sec.gov/news/press-release/2017-68-0. This is due to T+3 settlement regime during the sample period.

Definition 1. Given two investments, Y and Z, with F_1 and G_1 as their respective CDFs, Y dominates Z in the sense of FSD, SSD, TSD, denoted by $Y \succcurlyeq_1 Z$, $Y \succcurlyeq_2 Z$, $Y \succcurlyeq_3 Z$, if and only if $F_1(x) \leq G_1(x)$, $F_2(x) \leq G_2(x)$, $F_3(x) \leq G_3(x)$, and $\mu_Y \geq \mu_G$, for each $x \in [a, b]$, where FSD, SSD, and TSD represent first-, second-, and third-order SD, respectively. In addition, if there is a strict inequality for any $x \in [a, b]$, Y dominates Z in the sense of strict FSD, SSD, and TSD, denoted as $Y \succ_1 Z$, $Y \succ_2 Z$, $Y \succ_3 Z$, respectively.

Several pioneering papers (Hadar and Russel, 1969; Whitmore, 1970) have shown that SD theory is related to utility maximization theory. To establish the relationship between these two theories, the following is the definition of the utility function sets for risk-averse investors (Bai et al., 2015):

Definition 2. For $j = 1,2,3$, U_j, U_j^s are the utility function sets u such that

$$U_j\left(U_j^s\right) = \{u: (-1)^i u^{(i)} \leq (<) 0, i = 1,2, \ldots, j\},$$

where $u^{(i)}$ is the i^{th} derivative of u. In choosing between Y and Z investments, according to a consistent set of preferences, investors will satisfy the consistency properties of the expected utility theory. Thus, investment Y is (strictly) preferred to Z if $\Delta Eu \equiv Eu(y) - Eu(z) \geq (>)\ 0$, where $Eu(y) = \int_a^b u(y) dF(y)$ and $Eu(z) = \int_a^b u(z) dG(z)$.

The SD approach is considered the most useful tool for ordering risky investments. Li and Wong (1999) show that ordering investments using this approach is equivalent to maximizing the expected utility of risk-averse investor preferences, as presented in the following proposition:

Proposition 1. Let Y and Z be risky investments with CDFs F_1 and G_1, respectively. Let u be a utility function. For $j = 1, 2, 3$, $F \succcurlyeq_j (\succ_j) G$ if and only if $Eu(y) \geq (>) Eu(z)$ for any u in $U_j\left(U_j^s\right)$.

The existence of SD implies that the risk-averse investor's expected utility is always greater when he holds the dominant asset than when he holds the dominated asset, and thus the investor chooses the dominant asset.

3.4.2 Estimation and inference: the DD test

Several studies have developed tests to determine statistical significance in the SD approach (Anderson, 1996; Davidson and Duclos-DD, 2000; Barret and Donald, 2003; Linton et al., 2005). Because the DD test is considered to have a high power and yet be less conservative in size (Tse and Zhang, 2004, Lean et al., 2008), we use it here. Davidson and Duclos (2000) derive the asymptotic sampling distribution of several estimators to test any SD order for any two random variables with CDFs F_1 and G_1 based on the conditional moments of the distributions, in which the size is well controlled by a Studentized Maximum Modulus (SMM) statistic. The DD test is based on a set of grid points on the distributions.

Let $\{y_i, z_i\}_{i=1}^n$ be return pairs drawn from the populations of the daily subperiod (overnight and daytime) returns of the ETFs with CDFs $F_1(x)$ and $G_1(x)$, respectively. Given a preselected set of grid points $\{x_1, x_2, \ldots, x_k\}$, the DD statistic of the j^{th} order, $T_j(x)$, $j = 1,2,3$, is

$$T_j(x) = \frac{\widehat{F_j}(x) - \widehat{G_j}(x)}{\sqrt{\widehat{V_j}(x)}}, \tag{7}$$

where
$$\widehat{H}_j(x) = \frac{1}{n(j-1)!} \sum_{i=1}^{n} (x - h_i)_+^{j-1},$$
$$\widehat{V}_j(x) = \widehat{V}_{F_j}(x) + \widehat{V}_{G_j}(x) - 2\widehat{V}_{FG_j}(x),$$
$$\widehat{V}_{FG_j}(x) = \frac{1}{n}\left[\frac{1}{n((j-1)!)^2} \sum_{i=1}^{n} (x - y_i)_+^{j-1} (x - z_i)_+^{j-1} - \widehat{F}_j(x)\,\widehat{G}_j(x)\right],$$
$$\widehat{V}_{H_j}(x) = \frac{1}{n}\left[\frac{1}{n((j-1)!)^2} \sum_{i=1}^{n} (x - h_i)_+^{2(j-1)} - \widehat{H}_j(x)^2\right], H = F, G;\ h = y, z,$$

where $(x - h_i)_+ \equiv Max[(x - h_i); 0]$ and the tests are performed in a predefined set of values of x. The following hypotheses are considered:

$$\begin{aligned} &H_0\colon F_j(x_i) = G_j(x_i), \text{ for all } x_i,\ i = 1,2, \ldots, k. \\ &H_A\colon F_j(x_i) \neq G_j(x_i), \text{ for some } x_i, \text{ but } F \not\succcurlyeq_j G, F \not\preccurlyeq_j G. \\ &H_{A1}\colon F_j(x_i) \leq G_j(x_i), \text{ for all } x_i \text{ and } F_j(x_i) < G_j(x_i), \text{ for some } x_i. \\ &H_{A2}\colon F_j(x_i) \geq G_j(x_i), \text{ for all } x_i \text{ and } F_j(x_i) > G_j(x_i), \text{ for some } x_i. \end{aligned} \tag{8}$$

In these hypotheses, H_A is defined as being exclusive of H_{A1} and H_{A2}, which means that if the test accepts H_{A1} or H_{A2}, it will not be classified as H_A. H_A is accepted if $F_j(x) > G_j(x)$ for some x_i and $F_j(x) < G_j(x)$ for some x_i. Under H_0, Davidson and Duclos (2000) show that $T_j(x_i)$ is asymptotically distributed as a SMM distribution (Richmond, 1982) to account for joint test size. To implement the DD test, the test statistic (7) is calculated on each grid point, and H_0 is rejected if the highest test statistic is significant. The SMM distribution with k and infinite degrees of freedom at the $\alpha\%$ significance level, denoted by $M_{\infty,\alpha}^k$, is used to control for the probability of rejecting H_0. The decision rules based on the percentile $(1 - \alpha)$ of $M_{\infty,\alpha}^k$, tabulated in accordance with Stoline and Ury (1979), are as follows:

$$\begin{aligned} &\text{If } |T_j(x_i)| < M_{\infty,\alpha}^k \text{ for } i = 1,2, \ldots, k, \text{ accept } H_0. \\ &\text{If } T_j(x_i) < M_{\infty,\alpha}^k \text{ for all } i \text{ and } -T_j(x_i) > M_{\infty,\alpha}^k \text{ for some } i, \text{ accept } H_{A1}. \\ &\text{If } -T_j(x_i) < M_{\infty,\alpha}^k \text{ for all } i \text{ and } T_j(x_i) > M_{\infty,\alpha}^k \text{ for some } i, \text{ accept } H_{A2}. \\ &\text{If } T_j(x_i) > M_{\infty,\alpha}^k \text{ for some } i \text{ and } -T_j(x_i) > M_{\infty,\alpha}^k \text{ for some } i, \text{ accept } H_A. \end{aligned} \tag{9}$$

If H_0 or H_A is accepted, there is no SD of a given overnight or daytime return over another. However, if H_{A1} or H_{A2} is accepted, for $j = 1$, for an ETF, a given overnight or daytime return stochastically dominates another overnight or daytime return at the first order. If H_{A1} or H_{A2} is accepted for $j = 2$ or $j = 3$, a given overnight or daytime return stochastically dominates another overnight or daytime return at the second or third order. There is no definite theoretical solution for the optimum number of grid points to maximize the power of the SD test. Tse and Zhang (2004) and Lean et al. (2008) suggest using 10-15 grid points in the empirical distributions for the SD tests. More information about the distribution would be revealed using more grid points, but the assumption of independence of grids statistics required by SMM distribution would be violated using an excessive number of grid points (Richmond, 1982).

We follow Fong et al. (2005) and Lean et al. (2007) and make 10 major partitions with 10 minor partitions within any two consecutive major partitions in each comparison (i.e., a total of 100 grid points), and statistical inference is based on the SMM distribution for $k = 10$ and infinite degrees of freedom. The critical value of SMM for $n = \infty$ and $k = 10$ at the 5% level is 3.254.

4. Empirical Results

Table 1 shows descriptive statistics for the daytime and overnight returns for the examined U.S. equity ETFs, decomposed by days of the week during the entire sample period. Specifically, the table presents the mean return, standard deviation, skewness, kurtosis, the Jarque-Bera test for the normality hypothesis, the Welch two-sample F-test for the equality of means, and the two-sample Brown-Forsythe test for the equality of standard deviations. Across ETFs and for days of the week, there is a tendency for the higher mean return to occur during overnight than during the daytime period. During the daytime period and for each day of the week, except on Friday for QQQ, average returns are not significantly different from zero. The aforementioned day is significantly negative.

For the overnight period on Monday and Tuesday in both ETFs and on Thursday for QQQ, average returns are significantly different from zero. In these five cases, mean returns are significantly positive. The overall overnight average returns on both ETFs are also significantly positive. Upon first consideration, there does not seem to be a marked day-of-the-week effect in the overnight and daytime returns across the ETF group. Across ETFs, there are three common patterns: the Monday, Tuesday and overall overnight mean returns are significantly positive.

Volatility of returns (as measured by standard deviation) for each day of the week and across ETFs is higher during the daytime period. This result is consistent with evidence obtained in previous studies (French and Roll, 1986; Lockwood and Linn, 1990; Cliff et al., 2008) and is in line with the hypothesis that the volume of information flow is higher during the daytime than during the overnight period (George and Hwang, 2001). The volatility is lower in the SPY than in the QQQ, reflecting the higher volatility of the individual stocks that make up the NASDAQ-100 index.

The distributional properties of the return series for both ETFs, days of the week, and daytime and overnight periods do not appear to be normal. For both ETFs and almost every day of the week and daytime and overnight period returns, the return skewness is significant but there does not seem to be a pattern regarding the signal of this parameter. The only exception seems to be Tuesday overnight periods where returns in both ETFs are significantly positively skewed. This result indicates a higher probability than under the normal distribution of obtaining extreme positive returns during the overnight period on Tuesday. The kurtosis across ETFs and days of the week and by daytime and overnight periods is significant, indicating leptokurtic distributions, with the number of extreme returns being greater than under the normal distribution. Finally, the Jarque-Bera statistics, shown in the sixth column of Table 1, reject the null hypothesis of normality of returns in ETFs, all weekdays, and daytime and overnight periods.

Table 1 Descriptive Statistics for Overnight and Daytime Returns by Day of the Week and Etfs: 1996-2018 Period

Exchange Traded Fund	*Mean (%)*	*Std. dev. (%)*	*Skewness*	*Kurtosis*	*J.B.-test* [a]	*W-test* [b]	*B.F.-test* [c]
SPY							
Daytime Monday	-0.026	1.048	-0.812	13.713	5131***	2.138**	70***
Tuesday	0.014	1.063	0.373	12.401	3999***	1.053	122***
Wednesday	0.009	1.021	0.107	12.032	3999***	0.465	151***
Thursday	-0.009	1.012	-0.802	12.774	4717***	0.903	135***
Friday	-0.039	0.942	0.051	5.788	364***	1.395	62***
All	-0.010	1.017	-0.212	11.788	18009***	2.668***	525***
Overnight Monday	0.055***	0.678	0.153	16.154	7567***	-	-
Tuesday	0.053***	0.591	0.261	9.622	1983***	-	-
Wednesday	0.024	0.569	-0.616	8.017	1308***	-	-
Thursday	0.021	0.588	-0.270	7.292	899***	-	-
Friday	0.010	0.730	-1.886	26.947	27549***	-	-
All	0.032***	0.633	-0.647	17.278	47818***	-	-
QQQ							
Daytime Monday	-0.049	1.484	0.483	13.987	4577***	1.969**	98***
Tuesday	-0.045	1.581	-0.693	7.696	924***	1.913*	134***
Wednesday	0.017	1.679	0.879	20.053	12380***	0.311	118***
Thursday	0.026	1.520	-0.091	7.410	807***	0.684	135***
Friday	-0.108**	1.403	0.019	8.808	1365***	2.639***	60***
All	-0.031	1.538	0.170	12.641	18636***	3.277***	540***
Overnight Monday	0.063**	0.878	-0.311	18.091	8583***	-	-
Tuesday	0.065**	0.762	0.605	8.591	1261***	-	-
Wednesday	0.035	0.863	0.182	11.096	2767***	-	-
Thursday	0.064**	0.844	0.648	9.045	1585***	-	-
Friday	0.036	0.983	-0.441	14.003	4930***	-	-
All	0.052***	0.869	0.044	13.064	20281***	-	-

Notes: Overnight and daytime return time series are from 01/02/1996 to 12/31/2018 for SPY and from 03/11/1999 to 12/31/2018 for QQQ. [a] Jarque-Bera test for normality hypothesis. [b] The Welch F-test for equality of means is the square of the statistic $t = (\bar{x}_1 - \bar{x}_2)/\big((s_1^2/n_1) + (s_2^2/n_2)\big)^{0.5}$ (unequal variances), where n_1 and n_2 are the sample sizes, $\bar{x}_1$ and $\bar{x}_2$ are the sample means, and s_1^2 and s_2^2 are the samples variances, that follow the *t*-distribution with υ degrees of freedom, where $\upsilon = ((s_1^2/n_1 + s_2^2/n_2)^2)/\big(((s_1^2/n_1)^2/(n_1-1)) + ((s_2^2/n_2)^2/(n_2-1))\big)$. [c] The test for equality of variances is the Brown-Forsythe test that is robust against many types of non-normality. *, **, *** denote values that are statistically significant at the 10, 5 and 1% levels, respectively.

Table 2 Mean-Variance Dominance Results of Overnight and Daytime Returns by Days of the Week for SPY

ETF – daily period	*Day_T*	*Day_W*	*Day_Th*	*Day_F*	*Night_M*	*Night_T*	*Night_W*	*Night_Th*	*Night_F*
					1996– 2006				
Day(time) M(onday)	ND	ND	ND	ND	≺	≺	≺	≺	≺
T(uesday)		ND	ND	ND	≺	≺	≺	≺	≺
W(ednesday)			ND	ND	≺	≺	≺	≺	≺
Th(ursday)				ND	≺	≺	≺	≺	≺
F(riday)					≺	≺	≺	≺	≺
(Over)night M(onday)						ND	ND	ND	ND
T(uesday)							ND	ND	≻
W(ednesday)								ND	≻
Th(ursday)									≻
					2007– 2018				
Day(time) M(onday)	ND	ND	ND	ND	≺	≺	≺	≺	≺
T(uesday)		ND	ND	ND	≺	≺	≺	≺	≺
W(ednesday)			ND	ND	≺	≺	≺	≺	≺
Th(ursday)				≺	≺	≺	≺	≺	≺
F(riday)					≺	≺	≺	≺	≺
(Over)night M(onday)						ND	ND	ND	ND
T(uesday)							ND	ND	ND
W(ednesday)								ND	ND
Th(ursday)									ND

Notes: The table shows the results of the MV tests for day of the week by daytime and overnight period returns for SPY. The sample period is from 01/02/1996 to 12/31/2018. The results in this table are read based on row versus column. $Y \succ [\prec] X$ means that Y (row) dominates [is dominated by] X (column) by the MV criteria if $\mu_Y > (\geq)[< (\leq)]\ \mu_X$ and $\sigma_Y \leq (<)\ [\geq (>)]\ \sigma_X$ at the 5% significance level. ND denotes that Y do not dominates and is not dominated by X by the MV criterion.

4.1 Parametric Tests of the Mean and Volatility Return Differences

The last two columns of Table 1 present the test results for equality of means and standard deviations. For each ETF and day of the week, the statistical values are presented for the tests of the difference between the daytime and the overnight period returns. For both ETFs and each day of the week, the null hypothesis of the same variances is rejected at the 1% level, with the overnight volatility being significantly lower than the daytime volatility. These results are consistent with the hypothesis that the volume of information flow that occurs during daytime is significantly higher than that observed during overnight (Stoll and Whaley, 1990).

Regarding average returns, for SPY and throughout days of the week, the results indicate that, in general, differences in average returns between daytime and overnight periods are not, with some exceptions, significant. The Monday and overall overnight average returns are significantly higher than the corresponding daytime average returns. However, for SPY and the same subperiod used in previous studies, i.e. 1996-2006,[7] our results are consistent with those obtained by Cliff et al. (2008). For QQQ and across days of the week, the results indicate some significant differences. The overnight average return on overall, Monday and Friday are significantly higher than the corresponding daytime average returns. For QQQ and for the same subperiod, i.e. 1999-2006, our results on overall overnight and daytime average returns and volatility are also consistent with those obtained by Kelly and Clark (2011).[8] However, for both ETFs and the subperiod 2007-2018, the results indicate that there are no significant differences in average returns between daytime and overnight periods. Thus, upon initial consideration, it appears that the overnight and daytime effect found previously has significantly diminished or even disappeared. The effect that persists and is pervasive is the volatility of overnight being significantly lower than the volatility of the corresponding daytime periods.

4.2 Mean-Variance (MV) Results

In this section, we present and examine the results of the MV criterion. For comparative purposes, the results of the MV, SR, and SD tests are based on the risk premium series obtained as described in the methodology. Throughout the paper, to examine whether the daytime and overnight effect previously found actually manifests, whether it has diminished or even disappeared, we base our analysis on two subperiods: 1996-2006 and 2007-2018.

The MV criterion results for SPY and QQQ appear in Table 2 and Table 3, respectively, and the dominance relationships use a 5% significance level. Pairwise comparisons between daytime and overnight period returns are delimited by dotted lines. For the SPY and QQQ, and for the subperiod 1996-2006, mean equality test results for the 45 pairwise comparisons[9] among the daytime and overnight returns on weekdays show that only 3 and 7 differences are significant at the 5% levels, respectively. For both ETFs and over the subperiod 2007-2018, results show that only one difference is significant at the 5% level. For standard deviation equality tests, the

---

[7] The results are not reported here, but are available from the authors.
[8] The results are also not reported here, but are available from the authors.
[9] Number of pairwise comparisons among the 10 return time series (5 daytime and 5 overnight weekdays) obtained separately for each ETF.

results show that for SPY and QQQ during the subperiod 1996-2006, of the 45 pairwise comparisons, there are 28 and 30 significant differences at the 5% level, respectively, with standard deviations of the overnight returns being significantly lower than those of daytime returns. For the subperiod 2007-2018, and for SPY and QQQ, results show that there are 28 and 26 significant differences at the 5% level, respectively. Indeed, almost all the differences in the standard deviations are significant at the 1% level. Thus, as shown in the Tables 2 and 3, in SPY and QQQ, over the subperiods 1996-2006 and 2007-2018, we observe 28, 30, 26, and 26 pairs of MV dominance, respectively.

The results presented in Tables 2 and 3 exhibit significant pairs of MV dominance based on the pairwise error rate with a false-positive probability of $\alpha = 5\%$. Because the null hypothesis involves a large number of pairwise comparisons in mean and variance equality tests (i.e., 45 tests), *p*-values are corrected using multiple comparison procedures to control for the FDR (i.e., the fraction of tests called significant that are actually true nulls). Because tests involve some degree of dependence (i.e., for each ETF, we test for the equality of means and standard deviations among daytime and overnight periods across weekdays), for our FDR procedure, we use Benjamini and Hochberg's (1995) method, with an overall FDR value of $\delta = 5\%$, which is less stringent than Bonferroni and Holm's (1979) methods.

In both ETFs and subperiods, after correction of the original pairwise *p*-values in the mean equality tests using Benjamini and Hochberg's method, we find no significant differences at the 5% level. For SPY and QQQ and over the first subperiod, in the variance equality tests, all corrected *p*-values of the previously observed significant differences remain lower than the *p*-value cut-off of 5%. Over the second subperiod, we find that two previous significant differences in variances for SPY are no longer significant, and the same previous significant differences, i.e. 26, remain for QQQ.

Thus, in general, in both ETFs and subperiods, almost all the MV dominances exhibited in Table 2 and Table 3 are due to significant differences in standard deviations between daytime and overnight risk premiums of weekdays. Only in the first subperiod and in both ETFs are there significant differences in average returns at the 5% level, but these are very few. After we apply multiple testing procedures to pairwise *p*-values, all previous significant differences in average returns disappear, but almost all previous significant differences in variances remain. Thus, using daytime and overnight risk premium series, results do not support the occurrence of the day and night effect across weekdays, i.e., overnight mean returns are not significantly positive and higher than daytime mean returns. Hence, according to the MV criterion, risk-averse investors will prefer overnight to daytime periods just because the standard deviation of the overnight is significantly lower than that of the daytime period. However, as shown in Table 1, across weekdays and daytime and overnight periods, the non-normality of returns is pervasive. Applying the MV criterion as a decision rule for the preferences could lead to paradoxical results as higher-order moments are not taken into account.

Table 3 Mean-Variance Dominance Results of Overnight and Daytime Returns by Days of The Week for QQQ

ETF – daily period	*Day_T*	*Day_W*	*Day_Th*	*Day_F*	*Night_M*	*Night_T*	*Night_W*	*Night_Th*	*Night_F*
					1999– 2006				
Day(time) M(onday)	ND	ND	ND	ND	≺	≺	≺	≺	≺
T(uesday)		ND	ND	ND	≺	≺	≺	≺	≺
W(ednesday)			ND	≺	≺	≺	≺	≺	≺
Th(ursday)				ND	≺	≺	≺	≺	≺
F(riday)					≺	≺	≺	≺	≺
(Over)night M(onday)						ND	ND	ND	≻
T(uesday)							≻	≻	≻
W(ednesday)								ND	ND
Th(ursday)									ND
					2007– 2018				
Day(time) M(onday)	ND	ND	ND	ND	≺	≺	≺	≺	≺
T(uesday)		ND	ND	ND	≺	≺	≺	≺	≺
W(ednesday)			ND	ND	≺	≺	≺	≺	≺
Th(ursday)				ND	≺	≺	≺	≺	≺
F(riday)					≺	≺	≺	≺	≺
(Over)night M(onday)						ND	ND	ND	ND
T(uesday)							ND	ND	ND
W(ednesday)								ND	≻
Th(ursday)									ND

Notes: The table shows the results of the MV tests for day of the week by daytime and overnight period returns for the QQQ. The sample period is from 03/11/1999 to 12/31/2018. The results in this table are read based on row versus column. Y ≻ [≺] X means that Y (row) dominates [is dominated by] X (column) by MV criteria if μ_Y > (≥)[< (≤)] μ_X and $\sigma_Y \le (<) [\ge (>)]$ σ_X at the 5% significance level. ND denotes that Y do not dominates and is not dominated by X by the MV criterion.

4.3 Sharpe Ratio (SR) Results

In this section, we present and discuss results of the formal inferences made about SRs. We calculated SR as Sharpe (1966, 1994) suggests, dividing average risk premium by risk premium volatility. This measure shows the magnitude of the risk premium obtained per unit of risk volatility incurred, which is the most adequate for comparing diversified portfolios such as the broadly diversified equity ETFs analysed in this study.

The SR test results are presented[10] in Table 4 and Table 5. In each table, the bottom row of each subperiod presents the SR value and the statistically significant values of daytime and overnight risk-adjusted returns across days of the week. The significant values are for the two-sided, one-sample test statistic of the null hypothesis that SR is not statistically different from zero. This test enables us to examine whether each daytime and overnight period of the days of the week obtained a positive or negative risk premium. The statistic of this test is based on the assumptions of stationary and ergodic returns, properties whose risk premiums the series of the ETFs verify.

For SPY and over the subperiod 1996-2006, there is only one significant SR, which occurs during the Tuesday overnight period and that is significantly positive at the 5% level. Over the subperiod 2007-2018, none of the SRs is statistically significant. The QQQ, over the subperiod 1999-2006, exhibit two SRs that are significantly different from zero; the significantly positive SR also occurs during Tuesday overnight and the significantly negative SR during Tuesday daytime. Over the subperiod 2007-2018 there is only one significant SR, which occurs during the Wednesday overnight period. For both ETFs and all other daytime and overnight periods, SRs are not significantly different from zero. In short, in both ETFs over the subperiod 1996-2006, the results suggest the existence of a positive and significant SR only during the Tuesday overnight period. From the first to the second subperiod, results suggest that the overnight and daytime period exhibits reversal in the signal and magnitude of SR values towards equality of these values across days of the week. This reversal tends to be more marked in the QQQ. These results would point to a decrease in the hypothetical day and night effect in the US equity ETF market.

With the two-sample SR test, we test whether the SR of portfolio Y is statistically different from the SR of portfolio Z. Significant results from the null hypothesis test statistic for equality between SRs appear in Table 4 and Table 5 for SPY and QQQ, respectively. For SPY, over the first and second subperiod, we obtain 7 and 2 significant pairwise comparisons, although two and one of these are only marginally significant, respectively.

[10] The Excel spreadsheet formulas for the calculus of the variance of the asymptotic distributions of the one- and two-sample Sharpe Ratio tests and the results are available upon request from the authors.

Table 4 Sharpe Ratio Results of Overnight and Daytime Returns by Days of the Week for SPY

	Day_T	*Day_W*	*Day_Th*	*Day_F*	*Night_M*	*Night_T*	*Night_W*	*Night_Th*	*Night_F*
					1996– 2006				
Day(time) M(onday)	ND	ND	ND	ND	ND	$\prec_{**}$	ND	ND	ND
T(uesday)		ND	ND	ND	ND	$\prec_{**}$	ND	ND	ND
W(ednesday)			ND	ND	ND	ND	ND	ND	ND
Th(ursday)				ND	ND	$\prec_{*}$	ND	ND	ND
F(riday)					$\prec_{*}$	$\prec_{***}$	ND	ND	ND
(Over)night M(onday)						ND	ND	ND	ND
T(uesday)							$\succ_{**}$	$\succ_{**}$	ND
W(ednesday)								ND	ND
Th(ursday)									ND

Sharpe ratio values

Day_M	*Day_T*	*Day_W*	*Day_Th*	*Day_F*	*Night_M*	*Night_T*	*Night_W*	*Night_Th*	*Night_F*
-0.023	-0.034	0.009	-0.013	-0.062	0.052	0.104**	-0.053	-0.028	0.012

	Day_T	Day_W	Day_Th	Day_F	Night_M	Night_T	Night_W	Night_Th	Night_F
					2007– 2018				
Day(time) M(onday)	$\prec_{*}$	ND	ND	ND	ND	ND	ND	ND	ND
T(uesday)		ND	ND	ND	ND	ND	ND	ND	$\succ_{**}$
W(ednesday)			ND	ND	ND	ND	ND	ND	ND
Th(ursday)				ND	ND	ND	ND	ND	ND
F(riday)					ND	ND	ND	ND	ND
(Over)night M(onday)						ND	ND	ND	ND
T(uesday)							ND	ND	ND
W(ednesday)								ND	ND
Th(ursday)									ND

Sharpe ratio values

Day_M	**Day_T**	**Day_W**	**Day_Th**	**Day_F**	**Night_M**	**Night_T**	**Night_W**	**Night_Th**	**Night_F**
-0.063	0.063	-0.017	-0.001	-0.012	0.021	0.034	0.035	0.004	-0.043

Notes: The table shows the SR values and the results of the one-(hypothesis of SR value equal to zero) and two-sample Sharpe ratio tests for days of the week by daytime and overnight period returns for the SPY. The entire sample period is from 01/02/1996 to 12/31/2018. The results in this table are read based on row versus column. $Y \prec (\succ) X$ means that the SR of Y (row) is significantly lower (greater) than the SR of X (column) based on the two-sample test statistics in equation (4). ND denotes that the SR of Y is not significantly different from the SR of X. *, **. *** denote values that are statistically significant at the 10%, 5% and 1% levels based on two-tailed tests, respectively.

Table 5 Sharpe Ratio Results of Overnight and Daytime Returns by Days of the Week for QQQ

Daily period	*Day_T*	*Day_W*	*Day_Th*	*Day_F*	*Night_M*	*Night_T*	*Night_W*	*Night_Th*	*Night_F*
					1999– 2006				
Day(time) M(onday)	ND	ND	ND	ND	$\prec_{*}$	$\prec_{**}$	ND	$\prec_{*}$	ND
T(uesday)		ND	$\prec_{*}$	ND	$\prec_{**}$	$\prec_{***}$	ND	$\prec_{**}$	$\prec_{*}$
W(ednesday)			ND	ND	ND	ND	ND	ND	ND
Th(ursday)				ND	ND	ND	ND	ND	ND
F(riday)					$\prec_{**}$	$\prec_{**}$	ND	$\prec_{**}$	ND
(Over)night M(onday)						ND	ND	ND	ND
T(uesday)							$\succ_{*}$	ND	ND
W(ednesday)								ND	ND
Th(ursday)									ND
Sharpe ratio values									
Day_M	**Day_T**	**Day_W**	**Day_Th**	**Day_F**	**Night_M**	**Night_T**	**Night_W**	**Night_Th**	**Night_F**
-0.063	-0.127**	-0.010	0.012	-0.089	0.066	0.103**	-0.042	0.073	0.004
					2007– 2018				
Day(time) M(onday)	$\prec_{*}$	ND	ND	ND	ND	ND	$\prec_{**}$	ND	ND
T(uesday)		ND	ND	ND	ND	ND	ND	ND	ND
W(ednesday)			ND	ND	ND	ND	$\prec_{*}$	ND	ND
Th(ursday)				ND	ND	ND	ND	ND	ND
F(riday)					ND	ND	$\prec_{**}$	ND	ND
(Over)night M(onday)						ND	ND	ND	ND
T(uesday)							ND	ND	ND
W(ednesday)								ND	ND
Th(ursday)									ND
Sharpe ratio values									
Day_M	**Day_T**	**Day_W**	**Day_Th**	**Day_F**	**Night_M**	**Night_T**	**Night_W**	**Night_Th**	**Night_F**
-0.050	0.071	-0.010	-0.003	-0.033	0.024	0.051	0.093**	0.017	0.008

Notes: The table shows the SR values and the results of the one-(hypothesis of SR value equal to zero) and two-sample Sharpe ratio tests for days of the week by daytime and overnight period returns for QQQ. The entire sample period is from 03/11/1999 to 12/31/2018. The results in this table are read based on row versus column. $Y \prec (\succ) X$ means that the SR of Y (row) is significantly lower (greater) than the SR of X (column) based on the two-sample test statistics in equation (4). ND denotes that the SR of Y is not significantly different from the SR of X. *, **. *** denote values that are statistically significant at the 10%, 5% and 1% levels based on two-tailed tests, respectively.

For QQQ, over the first and second subperiod, we obtain 12 and 4 significant pairwise comparisons, albeit five and one of these are marginally significant, respectively. For both ETFs and over the first subperiod, significant pairwise comparisons are obtained mostly in cases where the overnight period SRs are significantly higher than the daytime period SRs. On the other hand, pairwise comparisons across days of the week among daytime periods and among overnight periods are not significant. Thus, for both ETFs, the decrease in the number of significant pairwise comparisons from the first to the second subperiod is consistent with a decrease and the disappearance of the day-and-night effect in the US equity ETF market.

In short, in both ETFs and subperiods, the SR approach reduces considerably in relation to the MV criterion, the number of significant dominance relationships. For both ETFs and the first subperiod, the results, in terms of significant dominances based on the pairwise *p*-values of the SR statistics, only seem to support the existence of a consistent and positive effect during the Tuesday overnight period. Over the second subperiod, some other significant dominance relationships in SRs, although few, are exhibited in QQQ but they are not consistent across ETFs.

The significant results of the one-sample and two-sample SR tests presented in Table 4 and Table 5 are based on the *p*-values of the pairwise comparisons. Because hypotheses involve many tests, we also apply multiple testing procedures to control for FDR. After we correct the original *p*-values using Benjamini and Hochberg's method, no significant one- and two-sample SR test statistics are recorded in both ETF and subperiods.

In short, when SR statistic *p*-values of the two-sample tests are corrected to control for FDR, all the previous significant SR dominance relations disappear. In general, these results tend to contrast with those of Kelly and Clark (2011) using the same set of ETFs, although these authors did not examine SRs differences between overnight and daytime returns across weekdays but only SR differences between overall overnight and overall daytime returns. Thus, across ETFs and weekdays and over the entire sample period, SR results after, and even before, correction of pairwise *p*-values do not support the existence of a pervasive and consistent overnight and daytime effect but, conversely, lend support to the decrease and disappearance of this effect in the U.S. equity market.

4.4 Stochastic Dominance (SD) Results

The results obtained for the MV and SR approaches, although they point to the decrease and disappearance of the day-and-night effect in the U.S. equity market over the last decade, do not allow us to conclude whether investors' preferences between the different portfolios for the daytime and overnight periods lead to an increase in wealth or, in the case of risk-averse investors, whether their preference will lead to an increase in utility without an increase in wealth. The SD approach enables us to answer these questions. To formally perform inference, we apply the DD test to the risk premium series of the daytime and overnight returns.

Recall that the DD test rejects the null hypothesis if none of the DD statistics is significantly positive (negative) and at least one of the DD statistics is significantly negative (positive) (Davidson and Duclos, 2000). Leshno and Levy (2002) show that in some situations, investment X stochastically dominates Y in a small range, but the vast majority of risk-averse investors prefer Y to X. In this case, it is said that Y almost stochastically dominates X. Thus, decision rules on which the DD test is based are very restrictive; that is, it is sufficient that only one DD statistic is significantly positive (negative) and the vast majority of other DD statistics significantly negative (positive) for the null hypothesis of equality between CDFs and high-order integrals not to be rejected. To minimize the Type II error of finding dominance when it does not exist and to account for the "almost" SD effect, we use a conservative 5% cut-off point. When using a 5% cut-off point, a given daytime or overnight portfolio dominates others if at least 5% of the $T_j(x_k)$, $j = 1,2,3$ and $k = 1,2,3,..,K$, are significantly negative (positive) and no portion of $T_j(x_k)$ is significantly positive (negative).

The results of the DD tests of the three SD orders, for SPY and QQQ, appear in Table 6 and Table 7, respectively.[11] Pairwise comparisons between daytime and overnight are delimited by dotted lines. For comparison purposes with our obtained SD findings, in Table 6 and Table 7, the last two bottom lines in each panel show the average return and the standard deviation of the risk premium series. The SD results in the tables are read based on row versus column. For example, in Table 6 and over the first subperiod, for the Day_M (Monday daytime) row, the $\prec_{2;3}$ under Night_M (Monday overnight) column means that Night_M dominates Day_M in the sense of the second– and third-order SD.

Examining comparisons of the complete return distributions between daytime and overnight across weekdays and during the two subperiods, we observe that $T_1(x_k)$ statistics are significant and entirely positive in the downside risk (i.e., in the lower range of the distribution) and significant and largely negative in the upside profit (i.e., in the upper range of the distribution).

This suggests that investors with increasing utility functions prefer to invest during the daytime when facing an upward trend in prices and overnight when facing downside risk. Thus, across both ETFs and pairwise comparisons between daytime and overnight, and given the pattern of the $T_1(x_k)$ statistics, the results lead us to accept the null hypothesis that overnight returns do not dominate nor are stochastically dominated by, in the sense of FSD, daytime returns. In no ETF does any overnight return display FSD over a daytime return. Bawa (1978) and Jarrow (1986) report that if there is no FSD, there are no arbitrage opportunities, and investors cannot increase their wealth and expected utility by switching from one investment to another. Thus, across both ETFs, these results indicate that there are no arbitrage opportunities between daytime and overnight returns in the U.S. equity ETF market, suggesting that these ETFs are efficient in impounding information, particularly in opening and closing prices. It may happen that arbitrage opportunities between daytime and overnight occur for short periods of time. However, since there is no FSD for a long time period in these ETFs, our FSD results do not reject that the US ETF equity market is efficient and investors are rational.

[11] The computation algorithm and the results of the three stochastic dominance orders are available upon request from the authors.

Table 6 Stochastic Dominance Relationships for Overnight and Daytime Returns for SPY over Weekdays

daily subperiod	*Day_T*	*Day_W*	*Day_Th*	*Day_F*	*Night_M*	*Night_T*	*Night_W*	*Night_Th*	*Night_F*
1996– 2006									
Day(time) M(onday)	ND	ND	ND	$\succ_{1;2}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$
T(uesday)		$\prec_{1;2;3}$	$\prec_{2;3}$	$\succ_{1}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$
W(ednesday)			ND	$\succ_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$
Th(ursday)				$\succ_{1;2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$
F(riday)					$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$
(Over)night M(onday)						$\prec_{2;3}$	$\succ_{1;2;3}$	$\succ_{1;2;3}$	$\succ_{2;3}$
T(uesday)							$\succ_{1;2;3}$	$\succ_{1;2;3}$	$\succ_{2;3}$
W(ednesday)								$\prec_{2}$	ND
Th(ursday)									$\succ_{2;3}$

Descriptive statistics

	Day_M	*Day_T*	*Day_W*	*Day_Th*	*Day_F*	*Night_M*	*Night_T*	*Night_W*	*Night_Th*	*Night_F*
Mean ret.(%)	-0.025	-0.038	0.009	-0.013	-0.062	0.029	0.052	-0.029	-0.015	0.007
Std.Dev. (%)	1.079	1.108	1.067	0.997	0.997	0.570	0.507	0.550	0.549	0.621

daily subperiod	*Day_T*	*Day_W*	*Day_Th*	*Day_F*	*Night_M*	*Night_T*	*Night_W*	*Night_Th*	*Night_F*
2007– 2018									
Day(time) M(onday)	$\prec_{1;2;3}$	$\prec_{1;2}$	ND	$\prec_{2}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$
T(uesday)		$\succ_{1;2}$	$\succ_{2}$	$\succ_{1;2}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	ND
W(ednesday)			$\succ_{2}$	$\succ_{1}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$
Th(ursday)				$\prec_{2}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$
F(riday)					$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2}$
(Over)night M(onday)						$\prec_{2}$	$\prec_{1;2}$	ND	$\succ_{1;2}$
T(uesday)							$\prec_{2;3}$	$\succ_{1;2}$	$\succ_{1;2}$
W(ednesday)								$\succ_{1;2;3}$	$\succ_{1;2}$
Th(ursday)									$\succ_{1;2}$

Descriptive statistics

	Day_M	*Day_T*	*Day_W*	*Day_Th*	*Day_F*	*Night_M*	*Night_T*	*Night_W*	*Night_Th*	*Night_F*
Mean ret.(%)	-0.064	0.066	-0.017	-0.001	-0.010	0.016	0.022	0.021	0.003	-0.034
Std.Dev. (%)	1.007	1.039	0.950	1.085	0.863	0.773	0.664	0.600	0.639	0.805

Notes: The table shows the results of the SD tests (i.e., DD; Davidson and Duclos, 2000) for day of the week by daytime and overnight period returns for the SPY. The sample period is from 01/02/1996 to 12/31/2018. The DD test statistics are computed over a grid of 100 daily returns of two portfolios (two daily subperiods: daytime and overnight). The returns of the two portfolios are pooled and then ranked, the 100 percentiles are identified, and the DD test statistics are calculated at each percentile. The results in this table are read based on row versus column. DN denotes that FSD, SSD, and TSD do not exist. $Y \succ_{1,2,3} (\prec_{1.2.3}) X$ means that the Y portfolio (row) stochastically dominates (is dominated by) the X portfolio (column) at FSD, SSD, and TSD, respectively, at the 5% significance level. The two bottom lines at the end of each subperiod are average return and standard deviation of the risk premium series used in the computation of the DD statistics.

Table 7 Stochastic Dominance Relationships for Overnight and Daytime Returns for QQQ over Weekdays

daily subperiod	*Day_T*	*Day_W*	*Day_Th*	*Day_F*	*Night_M*	*Night_T*	*Night_W*	*Night_Th*	*Night_F*
					1999– 2006				
Day(time) M(onday)	$\succ_{1;2;3}$	$\prec_{1};\succ_{2;3}$	$\prec_{1}$	$\succ_{1};\prec_{3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$
T(uesday)		$\prec_{1;2;3}$	$\prec_{1;2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$
W(ednesday)			$\prec_{2;3}$	$\succ_{1};\prec_{3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$
Th(ursday)				$\succ_{1};\prec_{3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$
F(riday)					$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$
(Over)night M(onday)						$\prec_{2;3}$	$\succ_{1;2;3}$	$\succ_{2}$	$\succ_{2;3}$
T(uesday)							$\succ_{1;2;3}$	$\succ_{2;3}$	$\succ_{2;3}$
W(ednesday)								$\prec_{1;2;3}$	$\prec_{1;2}$
Th(ursday)									$\succ_{1;2;3}$

Descriptive statistics

	Day_M	*Day_T*	*Day_W*	*Day_Th*	*Day_F*	*Night_M*	*Night_T*	*Night_W*	*Night_Th*	*Night_F*
Mean ret.(%)	-0.114	-0.263	-0.023	0.024	-0.160	0.067	0.093	-0.046	0.080	0.004
Std.Dev. (%)	1.813	2.064	2.183	2.039	1.820	1.021	0.908	1.114	1.097	1.172

	Day_T	*Day_W*	*Day_Th*	*Day_F*	*Night_M*	*Night_T*	*Night_W*	*Night_Th*	*Night_F*
					2007– 2018				
Day(time) M(onday)	$\prec_{1;2;3}$	$\prec_{1;2}$	ND	$\succ_{1}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$
T(uesday)		$\succ_{1;2;3}$	$\succ_{2}$	$\succ_{1;2}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2}$
W(ednesday)			$\succ_{2}$	$\succ_{1}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$
Th(ursday)				$\prec_{3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$
F(riday)					$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2;3}$	$\prec_{2}$
(Over)night M(onday)						$\prec_{1}$	$\prec_{1;2}$	$\succ_{1}$	$\succ_{2;3}$
T(uesday)							$\prec_{1;2;3}$	$\succ_{1;2;3}$	$\succ_{2}$
W(ednesday)								$\succ_{1;2;3}$	$\succ_{2}$
Th(ursday)									$\prec_{1}$

Descriptive statistics

	Day_M	*Day_T*	*Day_W*	*Day_Th*	*Day_F*	*Night_M*	*Night_T*	*Night_W*	*Night_Th*	*Night_F*
Mean ret. (%)	-0.054	0.082	-0.011	-0.004	-0.034	0.019	0.033	0.057	0.011	0.007
Std.Dev. (%)	1.080	1.155	1.126	1.154	1.019	0.793	0.663	0.618	0.672	0.850

Notes: The table shows the results of the SD tests (i.e., DD; Davidson and Duclos, 2000) for day of the week by daytime and overnight period returns for the QQQ. The sample period is from 03/11/1999 to 12/31/2018. The DD test statistics are computed over a grid of 100 daily returns of two portfolios (among daytime and overnight periods over weekdays). The returns of the two portfolios are pooled and then ranked, the 100 percentiles are identified, and the DD test statistics are calculated at each percentile. The results in this table are read based on row versus column. DN denotes that FSD, SSD, and TSD do not exist. $Y \succ_{1,2,3} (\prec_{1.2.3})\ X$ means that the Y portfolio (row) stochastically dominates (is dominated by) the X portfolio (column) at FSD, SSD, and TSD, respectively, at the 5% significance level. The two bottom lines at the end of each subperiod are average return and standard deviation of the risk premium series used in the computation of the DD statistics.

To further examine whether the market is efficient, we need to examine higher orders of SD. If there is no SSD between daytime and overnight, this suggests that risk-averse investors will be indifferent to these daily subperiods. Likewise, if there is no TSD, this implies that investors who are risk averse with decreasing absolute risk aversion preferences will also be indifferent and the market will be efficient. However, Falk and Levy (1989) consider that given two investments, X and Y, if switching from X to Y increases the investor's expected utility, the market is inefficient. Although SSD and TSD orders do not imply any arbitrage opportunity (i.e., an increase in expected wealth), it nevertheless implies risk-averse investors' preference for one investment over another, thus theoretically allowing investors to increase their expected utility.

From the analysis of higher-order SD relations between overnight and daytime, we observe that almost all $T_2(x_k)$ and $T_3(x_k)$ statistics are largely negative and significant throughout the distribution. In both ETFs, overnight exhibit SSD and TSD over daytime returns at the 5% significance level of the SMM distribution. It follows that any second- and third-order risk-averse investors prefer overnight rather than daytime returns because this preference, while it does not increase expected wealth, increases expected utility. Given that this effect is consistent across both ETFs, it follows that the results suggests the effect is present in the U.S. equity market. Thus, considering, however, that overnight returns only stochastically dominate daytime returns in the sense of SSD and TSD, adopting a long-short trading strategy to take advantage of these differences in expected utility would not be profitable. This is because, in addition to no SD in the sense of FSD, and thus no existing arbitrage opportunities, the practical implementation of such a strategy, repeated daily, would incur significant transaction costs (i.e., the full bid-ask spread), in addition to the brokerage fees to be paid to firms that intermediate orders with the stock exchange. These costs would nullify any theoretical benefit that might have been provided by SSD and TSD. Studies reported in Pettengill (2003) concerning the implementation of trading strategies to take advantage of differences in average returns between days of the week suggest that such differences would be offset by transaction costs. Thus, across ETFs, these results suggest that the market is efficient in impounding information into prices.

In Table 6 and Table 7, for each ETF, dominance relationships between daytime and between overnight periods over weekdays appear in the upper-left and lower-right corner of the triangle, respectively. The results show that across ETFs and weekdays, there is no common pattern regarding the direction and order of observed dominance relationships.

Another feature, mainly observed in dominance relationships between daytime periods, is that the results contradict the hierarchical property of SD (i.e., that FSD implies SSD, which in turn implies TSD), recommending that only the lower dominance order is reported (Levy, 1992). On the contrary, our results show that there are pairwise comparisons in which FSD does not imply SSD and TSD but in which portfolio Y first-order stochastically dominates portfolio X and, simultaneously, portfolio X second-order SD portfolio Y. The SD hierarchical property is indeed mathematically verified, but the data do not support it statistically. This could result from Y marginally dominating X in the sense of FSD and X marginally dominating Y in the sense of SSD and/or TSD. However, given that the three SD tests used a 5% cut-off point, why the SD hierarchical relationships do not hold in some cases is an open question. In short, across ETFs, dominance relationships among daytime periods of the weekdays show wide heterogeneity and inconsistency, thus providing no support for the existence of SD common patterns of any order.

Overall, results in Table 6 and Table 7 do not support those obtained by Cliff et al. (2008). Using the SD approach, we find that the evidence does not support the overnight effect. The results suggest that overnight returns stochastically dominate daytime returns, but only in the sense of SSD and TSD. These results do not support the hypothesis of arbitrage opportunities, but instead suggest that the U.S. equity market became more informationally efficient.

5. Conclusions

This study examines the presence of day-of-the-week effects in overnight and daytime period returns in a group of actively traded broad-index ETFs that track the major stock market indices in U.S. markets over the period spanning 1996 to 2018: the SPY (S&P 500) and the QQQ (NASDAQ 100 index). The analysis is decomposed into two subperiods: 1996-2006 and 2007-2018. We employ three approaches: the mean-variance (MV), the Sharpe ratio (SR), and the stochastic dominance (SD) approaches. Given its limitations, we use the MV approach for comparative purposes. Formal inferences about the SR approach rely on asymptotic distributions that are valid under very general conditions (i.e., stationary and ergodic returns). The SD approach has the advantage of not being distribution type-dependent, of considering information on the entire distribution, and of incorporating higher-order moments into test statistics.

Across ETFs and the MV and SR approaches, results do not exhibit a marked day-and-night effect on returns decomposed by days of the week. During the first subperiod, the results of the MV and SR approaches exhibit a few significant overnight effects, with the positive Tuesday overnight effect being the most salient. According to the MV and SR approaches, Tuesday overnight mean returns are significantly higher than a small number of daytime mean returns over days of the week and Tuesday risk-adjusted overnight returns are significantly higher than a small number of risk-adjusted daytime returns, respectively. However, from the first to the second subperiod, the results suggest a decrease and the disappearance of the previously found significant MV and SR dominances between overnight and daytime returns.

As both approaches involve a large number of pairwise comparisons, multiple testing procedures were applied to control for the false discovery rate. After applying multiple testing correction procedures in both subperiods, the MV approach does not exhibit any significant differences between overnight and daytime mean returns across days of the week. For the SR approach, the previous significant SR differences also disappear.

The results of the SD approach suggest that overnight returns do not dominate nor are they stochastically dominated by daytime returns, in the sense of first-order SD (FSD). In none of the ETF does any overnight exhibit FSD over daytime returns during weekdays. These results suggest that no arbitrage opportunities exist in the U.S. equity market and that investors cannot increase their wealth or expected utility by switching from any daytime to overnight periods, or vice versa, throughout the week. The results, however, suggest that overnight periods stochastically dominate daytime period returns, in the sense of second-order SD and third-order SD, suggesting that risk-averse investors would prefer overnight periods.

Overall, our results suggest the decrease and disappearance of the day-and-night effect previously reported by empirical studies on the U.S. equity market, i.e., our results suggest a mean return difference reversion toward zero between overnight and daytime returns in recent decades as evidenced by the change in mean returns and SR values from the first to the second subperiod. These results support the notion that the price discovery mechanism in U.S. equity markets has become more efficient.

These results are also consistent with the nature of this asset class; that is, these ETFs are broadly diversified portfolios with a diversification of private information, higher liquidity, and lower transaction costs. The regulatory changes introduced by the Securities and Exchange Commission (2005) could also have contributed to improvements in information impounding at the open and close of the markets and reduced trading activity leeway by specialists in the NYSE and market-makers in the NASDAQ in open and close price discovery. Other factors may have contributed to the reduction of the day-and-night effect. On the one hand, this group of ETFs, which represent the most important global equity indices, have become the most popular group of trading vehicles in the last decade by easily allowing establishing a diversified exposure to large-cap US stocks. The growing demand for this group of ETFs will have led to an unprecedented increase in liquidity and extremely low bid-ask spreads. On the other hand, its increasing liquidity and demand by the various investor groups may have led possible causes associated with the day-and-night effect, proposed by Berkman et al. (2012) and Lou et al. (2018), to disappear. Following these authors, the day-and-night effect would be significantly caused by the demand pressure by individual (uninformed) investors on opening prices that would later be reversed by institutional investors who would trade in the opposite direction throughout the day and especially at the close.

For market participants, our results imply that they will not have an advantage in timing their trades to benefit from significant overnight minus daytime return differences. For U.S. self-regulatory bodies of securities exchanges, our results suggest no evidence that might indicate deviant behaviour on the part of specialists and market-makers in price discovery mechanisms at the open and close of markets.

APPENDIX

Asymptotic Variance of the Difference between Two Sharpe Ratios

We obtain the variance of the difference as follows.[12] Consider that SR_y and SR_z are the respective Sharpe Ratios for the returns of portfolios y and z. Using the "delta method", Opdyke (2007) obtains the asymptotic variance of $\left(\widehat{SR}_y - \widehat{SR}_z\right) - (SR_y - SR_z)$:

$$Var_{diff} = \left(1 + \frac{SR_y^2}{4}\left[\frac{\mu_{4y}}{\sigma_y^4} - 1\right] - SR_y \frac{\mu_{3y}}{\sigma_y^3}\right) + \left(1 + \frac{SR_z^2}{4}\left[\frac{\mu_{4z}}{\sigma_z^4} - 1\right] - SR_z \frac{\mu_{3z}}{\sigma_z^3}\right) \text{-}2\left(\rho_{y,z} + \frac{SR_y\, SR_z}{4}\left[\frac{\mu_{2y,2z}}{\sigma_y^2 \sigma_z^2} - 1\right] - \frac{1}{2} SR_y \frac{\mu_{1z,2y}}{\sigma_z \sigma_y^2} - \frac{1}{2} SR_z \frac{\mu_{1y,2z}}{\sigma_y \sigma_z^2}\right),$$

where $\frac{\mu_3}{\sigma^3}$ is the skewness of portfolio, $\frac{\mu_4}{\sigma^4}$ is the kurtosis of portfolio, $\mu_{2y,2z} = E\left[\left(y - E(y)\right)^2 \left(z - E(z)\right)^2\right]$ is the joint central moment of the joint distribution of y and z and $\mu_{1y,2z} = E\left[\left(y - E(y)\right)\left(z - E(z)\right)^2\right]$ and $\mu_{1z,2y} = E\left[\left(z - E(z)\right)\left(y - E(y)\right)^2\right]$. Minimum variance unbiased estimators for these last three terms are the respective *h*-statistics $h_{2y,2z}$, $h_{1y,2z}$ and $h_{1z,2y}$, where $h_{1,2} = \left[2s_{0,1}^2 s_{1,0} - n s_{0,2} s_{1,0} - 2s_{0,1} s_{1,1} + n^2 s_{1,2}\right]/[n(n-1)(n-2)]$, and

$$\begin{aligned} h_{2,2} = \big[& -3s_{0,1}^2 s_{1,0}^2 + n s_{0,2} s_{1,0}^2 + 4n s_{0,1} s_{1,0} s_{1,1} - 2(2n-3) s_{1,1}^2 \\ & - 2(n^2 - 2n + 3) s_{1,0} s_{1,2} + s_{0,1}^2 s_{2,0} - (2n-3) s_{0,2} s_{2,0} \\ & - 2(n^2 - 2n + 3) s_{0,1} s_{2,1} + n(n^2 - 2n + 3) s_{2,2}\big] \\ & /[n(n-1)(n-2)(n-3)] \end{aligned}$$

and where $s_{a,b}$ are the simple power sums of $s_{a,b} = \sum_{i=1}^{n} y_i^a z_i^b$.

[12] The Excel spreadsheet formulas for the calculus of the variance are available upon request from the authors.

REFERENCES

Abhyankar A, Ghosh D, Levin E, Limmack R (1997): Bid-Ask Spreads, Trading Volume and Volatility: Intraday Evidence from the London Stock Exchange. *Journal of Business Finance and Accounting*, 24(3&4):343-362.

Amihud Y (2002): Illiquidity and Stock Returns: Cross-Section and Time Series Effects. *Journal of Financial Markets*, 5(1):31-56.

Anderson G (1996): Nonparametric Tests of Stochastic Dominance in Income Distributions. *Econometrica*, 64(5):1183-1193.

Bagnoli M, Clement M, Watts S, (2005): Around-the-Clock Media Coverage and the Timing of Earnings Announcements. Available at http://ssrn.com/abstract=570247.

Bai Z, Li H, McAleer M, Wong W (2015): Stochastic Dominance Statistics for Risk Averters and Risk Seekers: An Analysis of Stock Preferences for USA and China. *Quantitative Finance*, 15(5):889-900.

Barclay M, Hendershott T (2003): Price Discovery and Trading after Hours. *Review of Financial Studies*, 16(4):1041-1073.

Barrett G S, Donald S (2003): Consistent Tests for Stochastic Dominance. *Econometrica*, 71(1):71-104.

Bawa V (1978): Safety-First, Stochastic Dominance, and Optimal Portfolio Choice. *Journal of Financial and Quantitative Analysis*, 13(2):255-271.

Benjamini Y, Hochberg Y (1995): Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B*, 57(1):289-300.

Berkman H, Koch P, Tuttle L, Zhang Y (2012): Paying Attention: Overnight Returns and the Hidden Cost of Buying at the Open. *Journal of Financial and Quantitative Analysis*, 47(4):715-741.

Branch B, Ma A (2006): The Overnight Return, One More Anomaly. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=937997.

Chiappori P, Paiella M (2011): Relative Risk Aversion Is Constant: Evidence from Panel Data. *Journal of the European Economic Association*, 9(6):1021-1052.

Chow E, Lee Y, Liu Y (2004): Intraday Information, Trading Volume, and Return Volatility: Evidence from the Order Flows on the Taiwan Stock Exchange. *Academia Economic Papers*, 32(1):107-148.

Christie S (2005): Is the Sharpe Ratio Useful in Asset Allocation? MAFC Research Papers No. 31, Applied Finance Centre, Macquarie University. Available at https://ssrn.com/abstract=720801.

Cliff, MT, Cooper M, Gulen H (2008): Return Differences between Trading and Non-Trading Hours: Like Night and Day. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1004081.

Damodaran A (1989): The Weekend Effect in Information Releases: A Study of Earnings and Dividend Announcements. *Review of Financial Studies*, 2(4):607-623.

Davidson R, Duclos J-Y (2000): Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality. *Econometrica*, 68(6):1435-1464.

Doyle J, Magilke M (2009): The Timing of Earnings Announcements: An Examination of the Strategic Disclosure Hypothesis. *The Accounting Review*, 84(1):157-182.

Eling M, Schuhmacher F (2007): Does the Choice of Performance Measure Influence the Evaluation of Hedge Funds? *Journal of Banking and Finance*, 31(9):2632-2647.

Falk H, Levy H (1989): Market Reaction to Quarterly Earnings' Announcements: A Stochastic Dominance Based Test of Market Efficiency. *Management Science*, 35(4):425-446.

Feldstein, MS (1969): Mean Variance Analysis in the Theory of Liquidity Preference and Portfolio Selection. *Review of Economics Studies*, 36(1):5-12.

Fong, WM, Wong WK, Lean HH (2005): International Momentum Strategies: A Stochastic Dominance Approach. *Journal of Financial Markets*, 8(1):89-109.

Foster F, Viswanathan S (1993): Variations in Trading Volume, Return Volatility and Trading Costs: Evidence on Recent Price Formation Models. *Journal of Finance*, 48(1):187-211.

French K (1980): Stock Returns and the Weekend Effect. *Journal of Financial Economics*, 8(1):55-69.

French K, Roll R (1986): Stock Return Variances: The Arrival of Information and the Reaction of Traders. *Journal of Financial Economics*, 17(1):5-26.

George T, Hwang C (2001): Information Flow and Pricing Errors: A Unified Approach to Estimation and Testing. *Review of Financial Studies,* 14(4):979-1020.

Graves PE (1979): Relative Risk Aversion: Increasing or Decreasing? *Journal of Financial and Quantitative Analysis*, 14(2):205-2014.

Guiso L, Paiella M (2008): Risk Aversion, Wealth, and Background Risk. *Journal of the European Economic Association*, 6(6):1109-1150.

Hadar J, Russell W (1969): Rules for Ordering Uncertain Prospects. *American Economic Review*, 59(1):25-34.

Hakansson, N (1972): Mean Variance Analysis in a Finite World. *Journal of Financial and Quantitative Analysis*, 7(4):1873-1880.

Holm S (1979): A Simple Sequentially Rejective Bonferroni Test Procedure. *Scandinavian Journal of Statistics*, 6(2):65-70.

Hong H, Wang J (2000): Trading and Returns under Periodic Market Closures. *Journal of Finance*, 55(1): 297-354.

Jarrow R (1986): The Relationship between Arbitrage and First Order Stochastic Dominance. *Journal of Finance*, 41(4):915-921.

Jobson J, Korkie B (1981): Performance Hypothesis Testing with the Sharpe and Treynor Measures. *Journal of Finance*, 36(4):889-908.

Kelly M, Clark S (2011): Returns in Trading versus Non-Trading Hours: The Difference is Day and Night. *Journal of Asset Management*, 12(2):132-145.

Lachance M (2015): Night Trading: Lower Risk but Higher Returns? San Diego State University, Available at https://ssrn.com/abstract=2633476.

Lean HH, Wong W, Zhang X (2008): The Sizes and Powers of Some Stochastic Dominance Tests: A Monte Carlo Study for Correlated and Heteroskedastic Distributions. *Mathematics and Computers in Simulation*, 79(1):30-48.

Lean, HH, Smyth R, Wong WK (2007): Revisiting Calendar Anomalies in Asian Stock Markets Using a Stochastic Dominance Approach. *Journal of Multinational Financial Management*, 17(2):125-141.

Leshno M, Levy H (2002): Preferred by "All" and Preferred by "Most" Decision Makers: Almost Stochastic Dominance. *Management Science*, 48(8):1074-1085.

Levy H (1992): Stochastic Dominance and Expected Utility: Survey and Analysis. *Management Science*, 38(4):555-593.

Li CK, Wong WK (1999): Extensions of Stochastic Dominance Theory to Random Variables. *RAIRO–Operations Research*, 33(4):509-524.

Linton O, Maasoumi E, Whang YJ (2005): Consistent Testing for Stochastic Dominance under General Sampling Schemes. *Review of Economic Studies*, 72 (3):735-765.

Lo A (2002): The Statistics of Sharpe Ratios. *Financial Analysts Journal*, 58(4):36-52.

Lockwood L, Linn S (1990): An Examination of Stock Market Return Volatility during Overnight and Intraday Periods, 1964-1989. *Journal of Finance*, 45(2):591-601.

Longstaff F (1995): How Much Can Marketability Affect Security Values? *Journal of Finance*, 50(5):1767-1774.

Lou D, Polk C, Skouras S (2018): A Tug of War: Overnight versus Intraday Expected Returns, *Journal of Financial Economics*, forthcoming.

Markowitz, HM (1952): Portfolio Selection. *Journal of Finance*, 7(1):77-91.

McInish H, Wood R (1992): An Analysis of Intraday Patterns in Bid/Ask Spreads for NYSE Stocks. *Journal of Finance*, 47(2):753-764.

Mertens E (2002): Comments on Variance of the IID Estimator in Lo (2002, FAJ), Research Note, November. Available at http://www.elmarmertens.com/research/discussion.

Opdyke JD (2007): Comparing Sharpe Ratios: So Where are the *p*-Values? *Journal of Asset Management*, 8(5):308-336.

Patell J, Wolfson M (1982): Good News, Bad News, and the Intraday Timing of Corporate Disclosures. *Accounting Review*, 57(3):509-527.

Pettengill G (2003): A Survey of the Monday Effect Literature. *Quarterly Journal of Business and Economics*, 42(3/4):3-28.

Qiu, M, Cai T (2013): On Overnight Return Premiums of International Stock Markets. Available at http://www.nzfc.ac.nz/archives/2013/programme/.

Richmond J (1982): A General Method for Constructing Simultaneous Confidence Intervals. *Journal of the American Statistical Association*, 77(378):455-460.

Rogalski R (1984): New Findings Regarding Day-of-the-Week Returns over Trading and Non-Trading Periods: A Note. *Journal of Finance*, 39(5):1603-1614.

Schwert, G.W (1990): Indexes of US Stock Prices from 1802 to 1987. *Journal of Business*, 63(3):399-426.

SEC (2005): Securities and Exchange Commission, Regulation NMS: Final Rules and Amendments to Joint Industry Plans. Washington, D.C., 2005. (http://www.sec.gov/rules/final/34-51808.pdf).

Sharpe W (1966): Mutual Fund Performance. *Journal of Business*, 39(1):119-138.

Sharpe W (1994): The Sharpe Ratio. *Journal of Portfolio Management*, 21(1):49-58.

Stoline MR, Ury HK (1979): Tables of the Studentized Maximum Modulus Distribution and an Application to Multiple Comparisons among Means. *Technometrics*, 21(1):87-93.

Stoll H, Whaley R (1990): Stock Market Structure and Volatility. *Review of Financial Studies*, 3(1):37-71.

Tobin, J (1958): Liquidity Preference as Behavior Towards Risk. *Review of Economic Studies*, 25(2):65-86.

Tse YK, Zhang X (2004): A Monte Carlo Investigation of Some Tests for Stochastic Dominance. *Journal of Statistical Computation and Simulation*, 74(5):361-378.

Whitmore GA (1970): Third-Degree Stochastic Dominance. *American Economic Review*, 60(3):457-459.

Tompkins R, Wiener Z (2008): Bad Days and Good Nights: A Re-Examination of Non-Traded and Traded Period Returns, Available at https://ssrn.com/abstract=1102165.

Wong, WK (2007): Stochastic Dominance and Mean-Variance Measures of Profit and Loss for Business Planning and Investment. *European Journal of Operational Research*, 182(2):829-843.

Wood R, McInish T, Ord J (1985): An Investigation of Transaction Data for NYSE Stocks. *Journal of Finance*, 40(3):723-739.