Do Daylight-Saving Time Adjustments Really Impact Stock Returns?

Douglas G. Steigerwald
Economics Department
University of California, Santa Barbara
E-mail: doug@econ.ucsb.edu

Marc N. Conte
Bren School of Environmental Science and Management
University of California, Santa Barbara

July 24, 2007

Abstract

We study the possible impact of daylight-saving time adjustments on stock returns. Previous work reveals that average returns tend to decline following an adjustment. As averages are sensitive to outliers, more recent work focused on the entire distribution of returns and found little impact following adjustments. Unfortunately, the general nature of the alternative hypothesis reduces the power of the distribution test to detect an effect of adjustments on the location of the distribution. We construct robust tests that are designed to have power to detect a time-adjustment effect on the location of returns. We also develop a more novel test of exponential tilting that is designed to accommodate possible heterogeneity in the return distribution over time. When we apply these tests to S&P 500 stock returns, we are unable to rigorously detect a time adjustment effect on stock returns.

Acknowledgements: We thank the Santa Barbara Econometrics Group and David Hinkley for helpful comments.

Key Words: Asset price anomalies, daylight-saving time, exponential tilting, order and rank statistics, robust test
Subject Classification: C14, C15, G12

1 Introduction

The returns to assets that trade on sophisticated markets would seem to be immune to calendar regularities. Yet the search for such effects has a long history, and the relevant evidence continues to provoke spirited debate. A recent chapter in this debate concerns the impact of daylight-saving time on asset returns. Kamstra, Kramer and Levi (2000) find that daylight-saving time adjustments, driven largely by the fall adjustment, are associated with declines in average returns on stock exchanges. As the fall adjustment occurs very near several large exchange losses, Pinegar (2002) employs a test that is robust to these outliers and finds no impact from the time adjustment. In response, Kamstra, Kramer and Levi (2002) argue that because the test employed by Pinegar studies the entire distribution of returns, the test lacks the power needed to detect the adjustment impact. In this paper we construct robust statistics that more closely focus on the location of the return distribution. As these statistics do not allow time variation in the return distribution, we also develop a robust statistic that allows for general forms of such time variation. With these statistics we focus on the potential impact of daylight-saving time adjustments and find little evidence of adjustment impacts.

In their 2000 paper Kamstra, Kramer and Levi appeal to psychological evidence that (daylight-saving) time adjustments alter the performance of individuals in complex tasks. As trade analysis can be viewed as a complex task, they turn to analysis of asset returns on the days on which traders are most likely to be affected by the time adjustment. Because time adjustments are made on Sundays, the returns over the weekend are the subject of analysis. They find that asset returns are on average lower immediately following the time adjustments. They further find that the adjustment effect holds in three of the four countries that they study, although the countries do make their adjustments simultaneously (hence contagion might be the cause).[1]

The broad nature of hypotheses about calendar effects allows one to model the impact quite generally. Of course, the breadth of potential hypotheses

[1] Worldwide, 38 countries make daylight-saving time adjustments.

also lends weight to concerns about finding one statistic out of many that supports the hypothesis. These concerns are especially pronounced in assessing the impact of time adjustments on asset markets. It is not transparent what aspect of returns should be affected by loss of sleep. For example, should returns be less tightly clustered as sleepy traders are affected by potentially inaccurate calculations? The very nature of the preceding question belies intuition that may not be valid about which time adjustment has an effect. In spring, individuals potentially have one less hour of sleep on Sunday, while in fall there is potentially one more hour of sleep. From this logic the effect should be driven by the spring adjustment. Yet the estimates show the effect to be largely driven by the fall adjustment.

With such questions in mind, Pinegar revisits the original analysis. Given the strong correlation in asset return volatility, one is immediately concerned that the large decline in October 1987, which occurred over a weekend neighboring the time adjustment, is driving the estimated adjustment impact. Given the broad nature of the original hypothesis, Pinegar conjectures that the adjustment impact should affect more than simply the mean return. He examines the entire return distribution, with the Kolmogorov-Smirnov statistic, and is unable to distinguish a time adjustment impact. Yet the Kolmogorov-Smirnov statistic has low power to distinguish any one departure between distributions, as it is a test for general departures.

We propose an alternative approach that also mitigates the effect of outliers but concentrates power on the hypothesis of low returns following time adjustments. As the presence of outliers indicates that the median may be a more reliable measure of location than the mean, we focus on the hypothesis that time adjustment weekends (both spring and fall) have a lower median return than do all other weekends. To detect this effect we use the Wilcoxon rank-sum statistic together with the sign and the Wilcoxon signed-rank statistics. Because each of these statistics are less sensitive to an outlier than is the return mean, the tests are robust to events such as the October 1987 crash.

All of the statistics described above, including those used in the previous analyses, are developed under the assumption that the return distribution is invariant over time.[2] Yet the last 40 years have seen a number of stock

[2] Kamstra, Kramer and Levi (2000) fit a GARCH(1,1) model to returns and find that this specific form of variation in the return distribution does not alter their conclusion that the mean return is lower following a time adjustment.

market changes that have lowered the cost of trading. It is possible that these changes have not only altered the return distribution, but have done so in ways that mask the time adjustment impact. To address the possibility we develop a test statistic that allows the return distribution to vary over time. For our test we rank the weekend returns for each year separately, yielding a sequence of annual ranks. As the time adjustment effect would imply a large number of time adjustment weekends with low annual ranks, we then estimate the degree of exponential tilting in the histogram of adjustment weekend annual ranks. Because this method does not depend upon a specific model of time variation, we may be able to unmask a time adjustment impact in the presence of a wide range of possible patterns of time variation in the return distribution.

The remainder of the paper is organized as follows. In Section 2 we review the initial t-tests with the inclusion of more recent exchange returns. We also report a counterfactual analysis, based on the weekends immediately preceding the adjustment weekends. In Section 3, we discuss the impact of outliers and present the standard robust test statistics. We then develop the test statistic based on annual ranks and present the estimates for both the adjustment weekends and the counterfactual sample.

2 Time Adjustments and Average Returns

We begin our analysis of the impact of time adjustments on stock returns by focusing on average returns for the Standard and Poor's (S&P) 500 index, in accord with the earlier work of Kamstra and his coauthors. With daily returns from the Center for Research in Security Prices, we construct weekend returns for the period June 1962 through December 2006.[3] From these weekend returns we construct average returns for both the spring and fall adjusted weekends and for all other weekends (the unadjusted weekends). We also construct a (joint t) test statistic of the null hypothesis that the average return for unadjusted weekends equals the average return for adjusted weekends. In detail, if $\bar{r}_a$ is the sample mean of the n_a adjusted weekends and $\bar{r}_u$ is the sample mean of the n_u unadjusted weekends, then the test

[3]Weekend returns are measured as the price change between the Friday close and the closing price on the first trading day of the following week, which most often is Monday.

statistic for the null hypothesis of no adjustment effect ($H_0 : \bar{r}_a = \bar{r}_u$) is

$$t_{stat} = \frac{\bar{r}_a - \bar{r}_u}{s \cdot \sqrt{\frac{1}{n_a} + \frac{1}{n_u}}},$$

where s is the square root of the pooled variance, $s^2 = \frac{(n_a-1)s_a^2+(n_u-1)s_u^2}{n_a+n_u-2}$ formed from the sample variances for adjusted (s_a^2) and unadjusted (s_u^2) weekends. A maintained assumption in constructing this test statistic is that the variance of returns is not affected by the time adjustment (nor is the variance changing in any other way over time) as the variance of all weekend returns is treated as constant.

In Table 1, we report the constructed averages and the associated test statistics. For the adjusted weekends, we report the overall average as well as separate averages for the spring and fall adjustments. The first row mirrors the data sample reported by Kamstra, Kramer and Levi (2000) (hereafter, Kamstra et al.), for which there are thirty adjusted weekends for both spring and fall (no time adjustments were made in 1974). The second row extends the analysis through the most recent data, and so contains 39 adjusted weekends for both spring and fall. The large negative fall return, which gives rise to a value of the test statistic that finds significant evidence of time adjustment effects, is apparent in the first row.[4] For the updated sample, the negative fall return is reduced (in magnitude) as is the test statistic, which no longer indicates such strong evidence of time adjustment effects.

Table 1
Mean of S&P 500 Weekend Returns

Time Period	Unadjusted	Adjusted	Spring	Fall	Joint t-test
1967-1997	$-.0004$ (.0107)	$-.0035$ (.0153)	$-.0014$ (.0078)	$-.0055$ (.0202)	-2.13
1967-2006	$-.0003$ (.0111)	$-.0025$ (.0143)	$-.0004$ (.0081)	$-.0046$ (.0184)	-1.71
1962-1966	$-.0011$ (.0067)	$+.0018$ (.0080)	$-.0019$ (.0022)	$+.0048$ (.0100)	1.26

With the attenuation of the adjustment effect in more recent data, we construct two additional measures of the effect. The first relies on the fact

[4]The sample standard deviation appears below each sample mean. The large discrepancy between the Fall and Spring standard deviations calls into question the maintained assumption of constant variance, a point we address in the following section.

that daylight-saving time adjustments occurred only sporadically prior to the Uniform Time Act that took effect in 1967. Thus, although such time adjustments first occurred during World War I (in an effort to save energy), the application of time adjustments varied widely across states and over time.[5] In consequence we calculate the average returns for the period prior to 1967, in which we expect to see little impact. From the third row of Table 1, we find mixed evidence. We find that fall weekends are associated with higher returns, although the small sample size results in a lack of precision for the test statistic. If the time adjustment began precisely in 1967, then such evidence would certainly weaken the claim of a time adjustment effect, as there is evidence of an adjustment effect prior to the adjustment. But, as Kamstra et al. note in their original article, time adjustments did occur in various states over the early period and the observed return effect may be due to these adjustments.[6]

To more cleanly isolate the impact of time adjustments, we return to the post-1967 sample, in which time adjustments were nearly uniform across the country. We perform a counterfactual analysis in which we construct average returns for the weekends that immediately precede the time adjustment weekends. We focus on the week preceding the time adjustment, rather than the week following the time adjustment, to avoid the possible lingering effects of time adjustment on succeeding returns. We refer to these counterfactual (CF) weekends as the CF-Spring weekends (the weekends that immediately precede the spring adjustment weekend), the CF-Fall weekends and the CF-Unadjusted weekends. Note that the actual adjustment weekends are now included in the unadjusted category, so if time adjustments do cause return declines then it would be possible to have lower average returns for CF-Unadjusted weekends.

The results of this counter-factual construction, which are contained in Table 2, are striking. For the sample period that is the focus of earlier

[5]For example, the 1918 law establishing daylight-saving time for the entire United States was later repealed and the adoption of daylight-saving time became a decision made at the state level. A brief history of daylight-saving time in the US, along with details of implementation, is presented in Appendix Table A1.

[6]Kamstra et al. report a significant negative adjustment effect for the period 1928-1966. Our reporting period differs, as we use the returns reported directly by Standard and Poors, which are available beginning in June 1962. Kamstra et al. use returns constructed by the Center for Research in Securities Prices, which attempt to mimic Standard and Poors and begin earlier. As the Standard and Poors index weights are proprietary, the two returns series differ slightly.

studies, 1967-1997, the evidence for a "fall effect" is quite strong. As these fall weekends are not characterized by time adjustments, time adjustment cannot be the cause of the sharp decline in average returns.

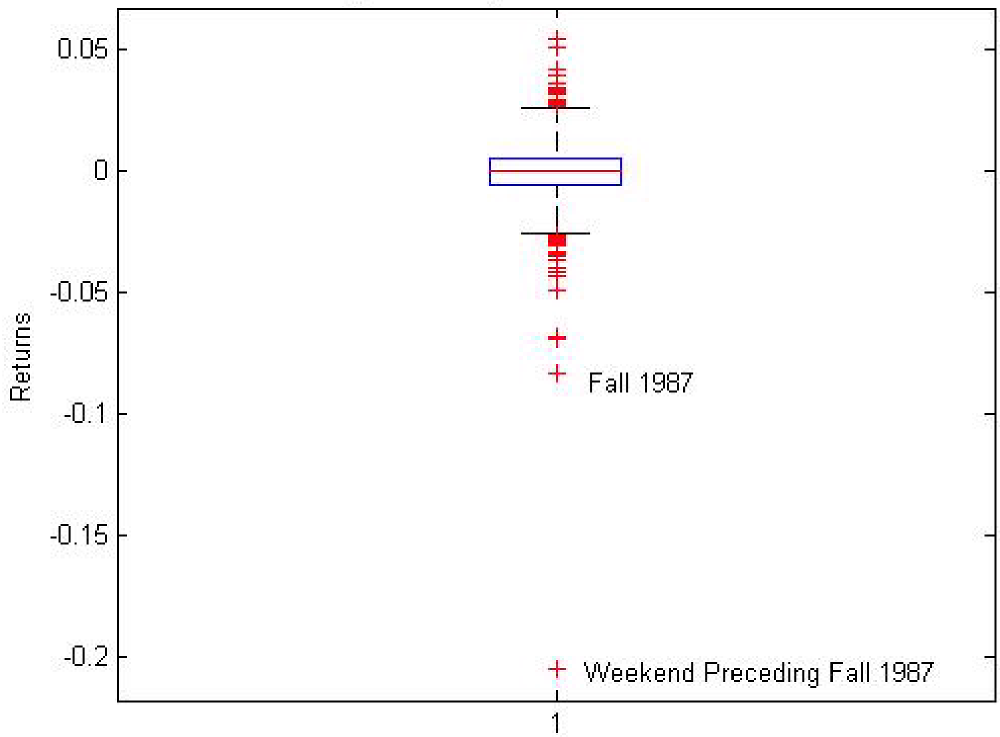
Table 2
Mean of S&P 500 Weekend Returns: Counterfactual Analysis

Time Period	CF-Unadjusted	CF-Adjusted	CF-Spring	CF-Fall	Joint t-test
1967-1997	$-.0004$ (.0096)	$-.0047$ (.0279)	$-.0003$ (.0089)	$-.0092$ (.0382)	-3.06
1967-2006	$-.0003$ (.0103)	$-.0026$ (.0251)	$+.0004$ (.0094)	$-.0056$ (.0343)	-1.80

3 Outlier Influence

While the counterfactual analysis provides some evidence against the conclusion that time adjustments lead to lower returns, the entire analysis of means could be affected by outliers. To determine the impact of outliers on the data sample from 1967 through 2006, we turn to the compact description in Figure 1. The rectangular box reflects the interquartile range for returns, with the median bisecting the box. The black lines above and below the interquartile box mark a further spread of twice the interquartile range. Observations lying beyond these lines are often referred to as outliers. Two important points emerge from the figure. First, both the adjusted and counterfactual adjusted weekend return means are likely impacted by outliers as the most dramatic (negative) outliers correspond to October 1987, in which both the fall adjustment weekend and the weekend preceding the fall adjustment saw sharp declines. Second, the large number of outliers emphasizes the point that returns are not well characterized by a Gaussian distribution. As the joint t-test (reported in the first two tables) may not be correctly sized if returns come from a non-Gaussian distribution, accurate testing of the proposed time adjustment effect may require robust test statistics.

Figure 1: Boxplot of Weekend Returns



To reduce the impact of outliers, and so more accurately measure any time adjustment effect, we turn to nonparametric test statistics that are robust to outliers. We focus on tests designed to detect an adjustment effect that reduces the size of returns.[7] These tests involve the rank of weekend returns, which are obtained as follows. For the combined sample vector of returns $\{r_t\}_{t=1}^{n}$ (our sample of 40 years contains $n = 2087$ weekend returns), order the returns from smallest to largest yielding the order statistics $\left\{r_{(t)}\right\}$.[8] The rank (order statistic) is then

$$\rho\left(r_{(t)}\right) = t.$$

[7] Pinegar constructs the Kolomogorov-Smirnov statistic, which as Kamstra, Kramer and Levi (2002) note, is not designed specifically to measure a reduction in the size of returns and so may not provide as precise a measure of the proposed adjustment effect.

[8] Seven years in the sample have 53 weekends.

Thus $r_{(1)}$ corresponds to the return of -20.4% for the weekend ending October 19, 1987, which immediately precedes an adjustment weekend and has rank $\rho = 1$.

In the presence of such substantial outliers, the median replaces the mean as the robust measure of return location. The proposed adjustment effect is then that the median of adjusted weekends (med_a) is smaller than the median of unadjusted weekends (med_u). To test $H_0 : med_a = med_u$ against $H_1 : med_a < med_u$, we use the Wilcoxon rank-sum test statistic

$$W_r = \sum_{t=1}^{n} \rho\left(r_{(t)}\right) \cdot z_t,$$

where $z_t = 1$ if $r_{(t)}$ is an adjustment weekend and $z_t = 0$ otherwise.[9] If n_a is the number of adjustment weekends (and $n_u = n - n_a$, is the number of unadjusted weekends), then the value of W_r is simply the sum of the ranks of the n_a adjustment weekends. If time adjustment has no impact on the size of weekend returns, then the ranks of the adjusted weekends should be randomly scattered among the ranks of the unadjusted weekends and the expected value of the test statistic is $E\left(W_r\right) = \frac{n_a(n+1)}{2}$ (with standard deviation $\sigma_{W_r} = \sqrt{\frac{n_a n_u (n+1)}{12}}$). If, however, time adjustment does lower weekend returns, then the ranks of the adjusted weekends should tend to be smaller than the ranks of the unadjusted weekends and the sum of the ranks for the adjusted weekends should be smaller than $E\left(W_r\right)$. For the sample size at hand, in which we have $n_a = 78$ adjustment weekends,

$$E\left(W_r\right) = 81,432 \text{ and } \sigma_{W_r} = 5,222,$$

so the critical value for a one-tailed test is 72,816 ($= 81,432 - 1.65 * 5,222$).[10] As the observed value of 79,049 exceeds the critical value by a substantial margin, we do not find evidence to support the hypothesis that adjustment weekends have a lower (median) return.

The analysis of both means and medians compares adjustment weekends with all other weekends. As adjustment weekends comprise only 4 percent (=

[9] The rank-sum test statistic dates to Wilcoxon (1945) for data in which $n_1 = n_2$. Mann and Whitney (1947) extend the analysis to $n_1 < n_2$ in developing their U-statistic. As the rank-sum statistic is a linear transformation of the U-statistic, the analysis in Mann and Whitney establishes the theory for the rank-sum statistic with $n_1 < n_2$. Gaussian critical values are recommended if $n_1 > 15$.

[10] Because n_1 (=78) exceeds 15, we use Gaussian limit theory.

78/2087) of all weekends, the discrepancy in relative sample sizes may hamper the ability to detect an adjustment effect.[11] To address the issue of uneven sample size, we construct matched pairs, for which $n_a = n_u$. Specifically, we pair every adjustment weekend with the weekend immediately preceding the adjustment weekend, yielding $\{(r_{a,t-1}, r_{a,t})\}_{t=1}^{n_a}$ where $r_{a,t-1}$ is the return for the weekend immediately preceding adjustment weekend t. We then form the sequence of differences $\{d_t\}$ where $d_t = r_{a,t-1} - r_{a,t}$. If adjustment weekends have lower returns, then these differences should be positive.

The sign test counts the number of positive differences, to determine if there are more positive differences than would be expected under the null hypothesis that $med\,(d_t) = 0$. As the sign test statistic S is formed from the sum of n_a Bernoulli random variables, each of which is positive with probability $\frac{1}{2}$ under the null hypothesis: $E\,(S) = 39$ $(= \frac{n_a}{2})$ and $\sigma_S = 4.4$ $(= \sqrt{\frac{n_a}{4}})$ for the sample size at hand. The observed value of 38 positive differences is *less* than the expected value under the null, and so falls far below the critical value of 46.3 $(= 39 + 1.65 * 4.4)$.

It may be the case that although fewer than half the adjustment weekends have positive differences, those that do are consistently large in magnitude. To discern such an effect we need to track both the sign and the magnitude of the differences. The magnitudes are tracked through the vector of ordered absolute differences $\left\{\tilde{d}_{(t)}\right\}$ where $\tilde{d}_t = |d_t|$. Under the null hypothesis, time adjustment affects neither the sign nor the magnitude of the differences, which is expressed as $H_0 : d_t$ is distributed symmetrically about 0 (which implies $med\,(d_t) = 0$). A natural test statistic for the symmetry hypothesis is the Wilcoxon signed-rank test

$$W_s = \sum_{t=1}^{n_a} \rho\left(\tilde{d}_{(t)}\right) \cdot \tilde{z}_t,$$

where $\tilde{z}_t = 1$ if $d_{(t)}$ is positive. Under the null hypothesis that there is no adjustment effect, the ranks of the positive differences should be equally likely among any of the ranks and $E\,(W_s) = 1,541$ $(= \frac{n_a(n_a+1)}{4})$ with $\sigma_{W_s} = 201$ $(= \sqrt{\frac{n_a(n_a+1)(2n_a+1)}{24}})$ for our sample size. If there is a pronounced adjustment effect, then the ranks of the positive differences should exceed the null expected value. As the observed value of 1,456 is *less* than the null

[11] There are only 78 adjustment weekends over the 40 year span because there was no time adjustment in 1974.

expected value, it falls well short of the critical value of 1,872 ($= 1,541 + 1.65 * 201$).

Table 3 contains the standardized values of the three nonparametric test statistics together with their critical values. As none of the estimated values lie close to, much less beyond, their critical values, there is little evidence to support the claim that time adjustment impacts the median of returns.

Table 3
Nonparametric Test Statistics (Critical Values in Parentheses)

Time Period	Wilcoxon Rank Sum	Sign	Wilcoxon Signed Rank
1967-2006	-0.46	-0.23	$-$ 0.42
	(-1.65)	$(+1.65)$	$(+1.65)$

3.1 Time Variation

The statistics reported in Table 3 are constructed under the assumption that the return distribution is unchanging over time. In practice, there have been a number of substantive changes over the past 40 years to the exchange markets from which the returns are computed. Dramatic growth in trading volume, the development of electronic trading, declining bid-ask spreads and the related decision to decimalize price quotes could all have led to changes in the return distribution over time.[12] Such changes might mask the adjustment effect, as significant adjustment effects in one year might be obscured by the magnitude of returns in other years.

Suppose, for example, that the adjustment effect is present but that the variance of the return distribution increases at some point in the sample, perhaps due to a development in electronic trading. We would then expect to see that adjustment weekends have low annual ranks, but that not all adjustment weekends have low ranks in the combined sample. In particular, the adjustment effect before the increase in return variance may be swamped by the magnitude of unadjusted weekend return changes after the increase in return variance.

To shed light on this possibility we develop a test statistic that accommodates very general time variation in the return distribution. To form this test statistic, we first construct annual ranks. The annual ranks are obtained from the vectors of ordered returns for each of the 40 years in our sample $\{r_{y,(t)}\}$, where $y = 1, \ldots, 40$ indexes the sample year. If there are n

[12] The decision to quote prices in a unit value of one cent rather than 1/8th of a dollar.

years in the sample, then there are n weekends (one in each year) that have an annual rank of 1.[13] In Table 4 we present the 11 adjustment weekends with an annual rank of 5 or lower (which corresponds to the lowest decile of annual ranks), together with their overall rank in the combined sample. If the return distribution is invariant over time, we would expect adjustment weekends with an annual rank of 1 to (generally) have an overall rank less than 40. Similarly, we would expect adjustment weekends with an annual rank of $j = 2, \ldots, 5$ to have an overall rank less than $j * 40$. As Table 4 reveals, this expected correspondence between annual ranks and overall ranks is more pronounced in the latter half of the sample. Indeed, the large overall ranks for the early part of the sample are indicative of time variation in the return distribution.

Table 4
Adjustment Weekends: Annual and Overall Ranks

Date	Annual	Overall	Date	Annual	Overall
1970, Spring	3	113	1988, Spring	4	230
1971, Fall	1	123	1994, Spring	1	121
1972, Spring	5	301	1996, Spring	2	87
1977, Spring	1	174	1997, Fall	1	3
1985, Spring	4	342	2001, Fall	4	39
1987, Fall	2	2			

To test for an adjustment effect with time variation in the return distribution, we examine the histogram of annual ranks for the adjustment weekends. Because 7 years in the sample have 53 weekends, we first divide the annual ranks by the number of weekends in the given year to form the standardized ranks

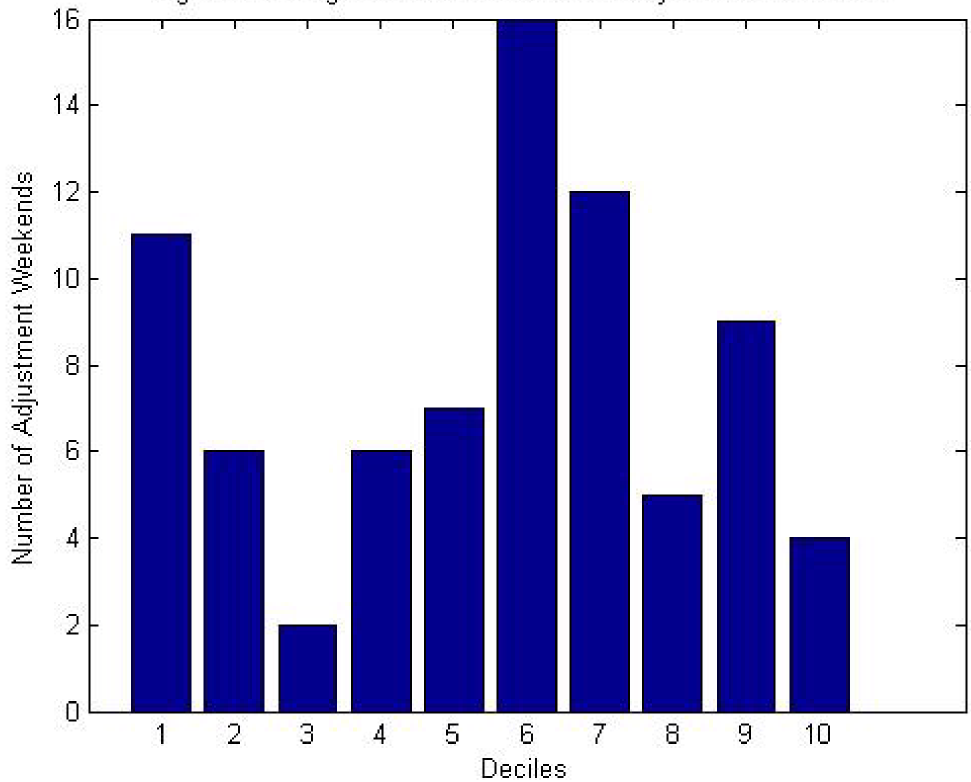
$$\tilde{\rho}\left(r_{y,(t)}\right) = \frac{t}{n_y},$$

where n_y is the number of weekends in year y. We then sort the standardized annual ranks for adjustment weekends into deciles, which yields the histogram in Figure 2.[14]

[13] The annual ranks of all adjustment weekends are presented in Appendix Table A2.

[14] As $\tilde{\rho}$ is an element of $\left(\frac{1}{n_y}, \ldots, 1\right)$ rather than $(1, \ldots, n_y)$, we construct the bins from the deciles over the unit interval. The first decile corresponds to $\tilde{\rho} \in (0, .1]$, which implies $\frac{(t)}{n_y} \leq .1$ or $(t) \leq 5$. Appendix Table A3 contains the complete list of the correspondence between bins and annual ranks.

Figure 2: Histogram of Annual Ranks for Adjustment Weekends



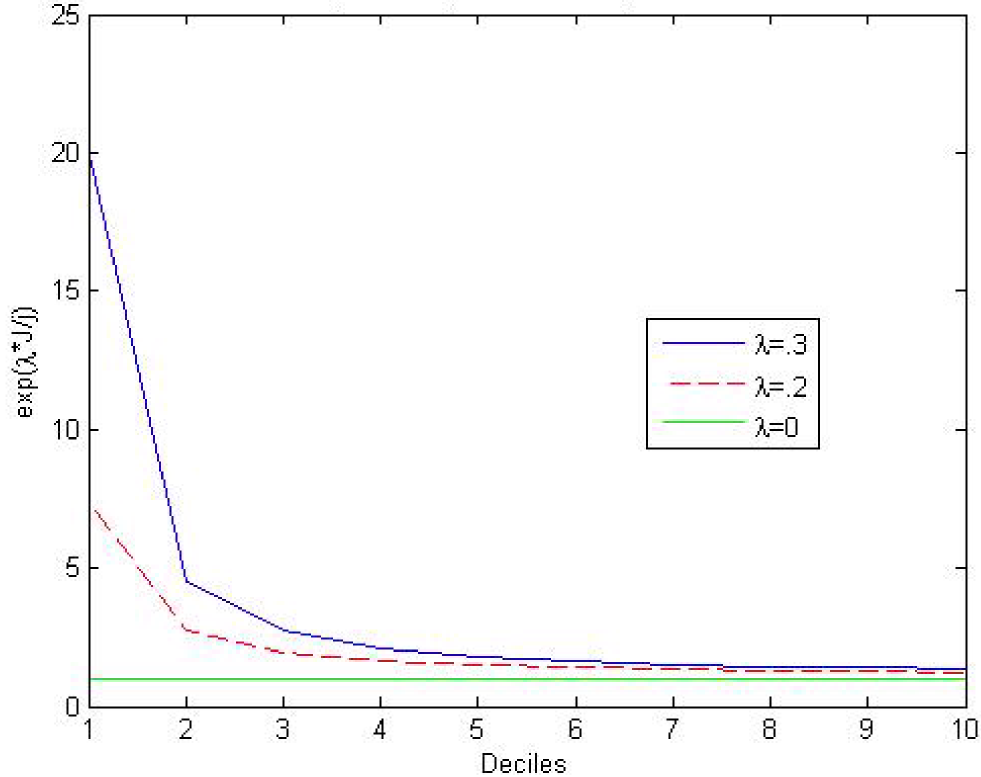
Under the null hypothesis of no adjustment effect, the histogram should be approximately uniform, while under the alternative hypothesis that there is an adjustment effect, the annual ranks of adjustment weekends should cluster in the lowest bins. We search for this pattern of clustering by testing for exponential tilting. If we let p_j be the number of adjustment weekends whose standardized ranks place them in bin j, then an exponential model for the histogram is

$$p_j \propto \exp\left(\lambda_0 \frac{J}{j}\right),$$

where $J = 10$ is the number of bins. Under the null hypothesis the histogram should be uniform and so $\lambda_0 = 0$ (p_j is proportional to a constant). If the ranks of adjustment weekends tend to cluster in the lowest bins, then $\lambda_0 > 0$.

As we can see from Figure 3, which plots the exponential curve of bin heights, the size of λ_0 indicates the degree of clustering.

Figure 3: Exponential Tilting Function



To form a test of $H_0 : \lambda_0 = 0$ against $H_1 : \lambda_0 > 0$, we use the nonlinear least-squares estimator

$$\hat{\lambda} = \arg\min_{\lambda} \sum_{j=1}^{J} \left(P_j - \exp\left(\lambda \frac{J}{j} \right) \right)^2 .$$

For the sample at hand $\hat{\lambda} = .25$. We obtain an appropriate confidence bound from Monte Carlo simulation under the null hypothesis of no adjustment effect. We generate 40 rank pairs for each simulation sample. The first element of each pair is drawn from the full range of ranks (the values 1 through 52 for 33 of the pairs and the values 1 through 53 for the remaining 7 pairs). The second element of each pair is drawn without replacement from the appropriate range of ranks.[15] With the simulated pairs of annual ranks, we then follow the estimation strategy for the original sample. That is, we standardize the ranks, bin them by deciles and construct the nonlinear least-squares estimator $\hat{\lambda}$. We perform 10,000 simulations and obtain a 95% upper confidence bound of .26. As the sample estimate is below the upper confidence bound, we are unable to reject the null hypothesis of no adjustment effect.

4 Conclusion

We attempt to discern a negative impact on stock returns arising from daylight-saving time adjustments. While the mean return shows a negative impact from time adjustment in the sample originally analyzed, extension of the sample lessens the significance of the impact. Curiously, the impact is driven by fall weekends, in which participants gain an hour of sleep and so are not sleep deprived. Moreover, the counterfactual returns constructed for the weekends immediately preceding the time adjustments also show a significant negative mean.

As the mean analysis is likely sensitive to the large number of outliers in the data, we also perform three nonparametric statistical tests that are robust to outliers. None of the three statistics reveals evidence of a negative impact of time adjustment on returns. Indeed, for two of the statistics the observed value provides (slight) evidence of a *positive* impact of time adjustment on returns. As all of the statistics assume that returns are i.i.d. draws from one distribution, we next examine the possible impact of time variation in the return distribution. To do so, we develop a test of exponential tilting that may prove useful in other applications with general time variation in the underlying distribution. While we do find evidence that the return distribution varies over the sample, when we allow for this variation we are still unable to detect an adjustment effect. Together, these results call into question a causal link between daylight-saving time adjustment and stock returns.

[15] The drawing is without replacement to ensure that no simulated annual pair has the same rank for both adjustment weekends.

5 Appendix: A Brief History of Calendar Time

The deadline-driven culture of today's business climate is the most recent manifestation of a connection between time and the functioning of human society. Human society's fascination with time and time-keeping extends back to the origins of man. Few examples remain from pre-historic times, but it is known that hunters in Europe over 20,000 years ago carved lines and bored holes in sticks and bones, potentially in efforts to track the phases of the moon. Among earlier civilizations, time-keeping methods were essential to survival, so that adequate preparations could be made for important events, such as animal migrations and river flooding.

Calendars were important developments among many of the ancient civilizations around the globe. The first Egyptian calendar followed the lunar cycle, but the observation that the 'Dog Star" rose next to the sun every 365 days, nearly coinciding with the flooding of the Nile, led to the development of a 365 day calendar sometime during the years 2937-2821 B.C. (Richards, 1999). Before 1800 B.C., the Babylonians had developed a lunar calendar made up of 12 months, which were between 29 and 30 days in length. The Maya created a 365-day calendar based on the motion of planets through the sky. The current 365-day solar calendar incorporates a leap year in every fourth year (excepting century years not divisible by 400) and is derived from the Gregorian calendar, which was introduced in 1582 based on the mean tropical year.

The next major development in the standardization of time arose in response to a desire for scheduled railroad operation. Sanford Fleming was the chief engineer of the Canadian Pacific Railroad in 1879 and he was primarily concerned with running his railway on a specified schedule, which was very difficult using local times. His paper "The Selection of a Prime Meridian" is the first record of a public presentation of the standard time system as used today (Curran and Taylor, 1935). The result of his efforts was the establishment of standardized time put into effect in Canada and the United States on November 18, 1883 at noon.

The implementation of daylight-saving time first occurred as an energy-saving means during World War I. In the United States, a law signed on March 31, 1918 established daylight-saving time in the United States. The act was repealed just one year later and the adoption of daylight-saving time became a decision made at the state level until World War II. President Roosevelt re-established national daylight-saving time on a year-round basis during World War II for energy conservation purposes.

The Uniform Time Act of 1966 became law under President Johnson on April 12, 1966 and established a uniform period of daylight-saving time, spanning from the last Sunday of April until the last Sunday of October. The grandeur of its name belied the fact that any state that chose not to follow daylight-saving time could do so using state law. In response to the energy crisis spawned by a reduction in oil production by OPEC, national year-round daylight-saving time was adopted for a fifteen month period beginning in January of 1974. According to a study of consumption for 1974 and 1975 by the U.S. Department of Transportation, which has jurisdiction over daylight-saving time, observing daylight-saving time in March and April saved an amount of energy equivalent to 10,000 barrels of oil each day. In 1987 federal legislation extended the duration of daylight-saving time. As Table A1 reveals, the duration of daylight-saving time (DST) has been extended again in 2007.

Table A1: daylight-saving Time Legislation

Year	DST Observed	Spring Adjustment	Fall Adjustment
1918	Nationally	Last Sunday in March	Last Sunday in October
1919	Locally		
1942[a]	Nationally	Year-Round DST	
1945[b]	Locally		
1967	Nationally	Last Sunday in April	Last Sunday in October
1974	Nationally	Year-Round DST	
1975	Nationally	Last Sunday in April	Last Sunday in October
1987	Nationally	First Sunday in April	Last Sunday in October
2007[c]	Nationally	Second Sunday in March	First Sunday in November

[a] Year-round DST began on the first Sunday in February, 1942
[b] Year-round DST ended on the last Sunday in September, 1945
[c] Enacted in 2005

Table A2: Adjustment Weekend Annual Ranks

Year	Spring	Fall	Year	Spring	Fall
1967	26	27	1987	37	2
1968*	38	18	1988	4	31
1969	45	30	1989	44	26
1970	3	19	1990*	15	9
1971	27	1	1991	42	46
1972	5	32	1992	48	47
1973*	32	33	1993	29	31
1974	none	none	1994	1	13
1975	18	24	1995	33	44
1976	29	32	1996*	2	10
1977	1	18	1997	31	1
1978	45	39	1998	19	27
1979*	24	30	1999	50	8
1980	36	9	2000	31	47
1981	35	19	2001*	9	4
1982	39	45	2002	34	21
1983	10	28	2003	23	24
1984*	37	26	2004	43	26
1985	4	33	2005	33	45
1986	32	31	2006	37	24

Thus the spring adjustment weekend in 1967 has annual rank 26. The seven years with 53 weeks are denoted with an asterisk.

Table A3: Annual Rank and Histogram Bin Correspondence

Bin	Annual Ranks		Bin	Annual Ranks	
	52 Weekends	53 Weekends		52 Weekends	53 Weekends
1	1-5	1-5	6	27-31	27-31
2	6-10	6-10	7	32-36	32-37
3	11-15	11-15	8	37-41	38-42
4	16-20	16-21	9	42-46	43-47
5	21-26	22-26	10	47-52	48-53

Thus adjustment returns fall in the second bin if $.1 < \frac{(t)}{n_y} \leq .2$ or $6 \leq (t) \leq 10$.

Bibliography

Curran, G. and I. Taylor, 1935, *World Daylight Saving Time*, Chicago: Curran Publishing.

Kamstra, M., L. Kramer and M. Levi, 2000, "Losing Sleep at the Market: The Daylight Saving Anomaly", *American Economic Review* 90, 1005-1011.

Kamstra, M., L. Kramer and M. Levi, 2002, "Losing Sleep at the Market: The Daylight Saving Anomaly: Reply", *American Economic Review* 92, 1257-1263.

Mann, H. and D. Whitney, 1947, "On a Test of Whether One or Two Random Variables is Stochastically Larger than the Other", *Annals of Mathematical Statistics* 18, 50-60.

Pinegar, J., 2002, "Losing Sleep at the Market: Comment", *American Economic Review* 92, 1251-1256.

Richards, E., 1999, *Mapping Time: The Calendar and Its History*, New York: Oxford University Press.

Wilcoxon, F., 1945, "Individual Comparisons by Ranking Methods", *Biometrics Bulletin* 1, 80-83.