Do Daylight-Saving Time Adjustments Impact Human Performance?

Marc N. Conte and Douglas G. Steigerwald*

December 28, 2007

Abstract

We study the possible impact of daylight-saving time adjustments on human performance. Researchers have detected potential impacts, which arise from behavioral effects due to sleep desynchronicity, on investment decisions and driver performance. The impacts have been detected in the mean: for investment decisions the mean stock return potentially decreases following adjustments; for driver performance the mean number of accidents potentially increases following adjustments. These findings have generated controversy, as both stock returns and accidents are subject to outliers and possible nonstationarity. To help resolve the controversy, we search for adjustment impacts with statistics other than the mean. We use rank statistics to account for outliers and develop a novel test of exponential tilting to account for nonstationarity. When these statistics are applied to S&P 500 stock returns and accident data for Oregon, we are able to detect a time adjustment impact only for accidents.

Acknowledgements: We thank the Santa Barbara Econometrics Group and David Hinkley for helpful comments. We thank Greg DeAngelo and Benjamin Hansen for assistance in obtaining the traffic accident dataset.
Key Words: Accidents, asset price anomalies, daylight-saving time, exponential tilting, order and rank statistics, robust test
Subject Classification: C14, C15, G12

*Bren School of Environmental Science & Management and Economics Department, University of California, Santa Barbara, respectively. mconte@bren.ucsb.edu and doug@econ.ucsb.edu.

1 Introduction

Sleep deprivation has been shown to impair a person's ability to complete complex tasks. It has also been suggested that sleep desynchronicity, or anxiety stemming from the interruption of normal sleeping rhythms, impacts human performance. The time adjustments associated with daylight saving are potential sources of sleep desynchronicity. Potential time-adjustment impacts have been detected for investment decisions, through an analysis of mean stock returns, and on driver performance, through an analysis of the mean number of accidents. When the mean is used to depict human performance, inference may be inaccurate due to the presence of outliers as well as possible nonstationarity. We utilize tests that are robust to outliers and have power to detect a time-adjustment effect on the location of the performance measures. In addition, we develop a more novel test of exponential tilting that is designed to accommodate nonstationarity in the distribution of our variables of interest over time.

The potential adjustment impacts on stock returns and accidents have been controversial. In a 2000 paper, Kamstra, Kramer and Levi detect a negative adjustment impact on the mean of weekend stock returns. Pinegar (2002) utilizes the Kolmogorov-Smirnov statistic to test for adjustment impacts on the entire distribution of weekend returns and is unable to detect an adjustment impact. Kamstra, Kramer and Levi (2002) reply that Pinegar's test has low power to distinguish an adjustment impact specifically on the location of weekend returns. The adjustment impact on accidents is also unresolved. Coren (1996) detects a positive impact on the mean of daily accidents in Canada following the spring adjustment (and a corresponding negative impact following the fall adjustment). Vincent (1998) extends the data and is unable to detect either impact. Coren (1998) again detects the spring adjustment impact with daily accident data from the United States.

The uncertainty clouding the impact of time adjustments on human performance is, at least in part, due to reliance on the mean as the measure of location. Both weekend returns and daily accidents are characterized by outliers. Weekend returns have disproportionately large (in magnitude) negative returns. Daily accident totals have several large values, arising from inclement weather. As the mean is sensitive to such outliers, we use the median to detect an adjustment effect on the location of the performance measures.[1]

Our attempt to identify the behavioral impacts of time adjustments utilizes an approach that mitigates the effect of outliers but concentrates power on the hypothesis of low returns and more accidents following time adjustments. For the question of stock returns, we focus on the hypothesis that time-adjustment weekends (both spring and fall) have a lower median return than do all other weekends. On the topic of daily traffic accidents, we focus on the hypothesis that Mondays following time adjustments will have a higher median number of accidents than neighboring Mondays. To detect these effects we use the Wilcoxon rank sum statistic together with the sign and the Wilcoxon signed-rank statistics. Because each of these statistics is less sensitive to outliers than is the mean, the tests are robust to events such as the October 1987 crash in the stock market or a particularly fierce ice storm that dramatically degrades driving conditions.

Further, the distributions of both weekend returns and daily accidents are likely to change over time. For weekend returns, nonstationary influences include the dramatic increase in trading volume over the past 40 years and the decision to decimalize stock prices. For daily accidents, nonstationary influences include an increase in the number of vehicles on the road over the past 20 years and traffic-related policy changes such as alteration of laws governing teenage drivers or variation in policing levels. To account for such general forms of nonstationarity, we construct the rank statistics for each year separately and then test for exponential tilting in the histogram of ranks.

It is possible that these changes have not only altered the distributions of our focal variables, but have done so in ways that mask the time adjustment impact.To address such a possibility, we develop a test statistic that allows the return distribution to vary over time.[2] For our test, we rank the weekend returns and Monday traffic accidents for each year separately, yielding

[1] If the effort to detect adjustment impacts on stock returns is equated to a test of market efficiency, then the appropriate measure of location depends on the behavior of investors. If investors are assumed to be risk-neutral, then the mean is the appropriate location measure. If, however, investors are risk averse, then the median is an appropriate location measure, as the variance and other features of the return distribution impact investment decisions.

[2] Kamstra et al. (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.

sequences 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 for both stock returns and traffic accidents. 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. We present our analysis of stock returns first and then turn our focus to traffic accidents. 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 the adjustment weekends. Sections 4 and 5 present similar analysis for the accident data.

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

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

[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.

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 et al. (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 that 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

[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.

[5] For example, the 1918 law establishing daylight-saving time for the entire United

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 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.

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.

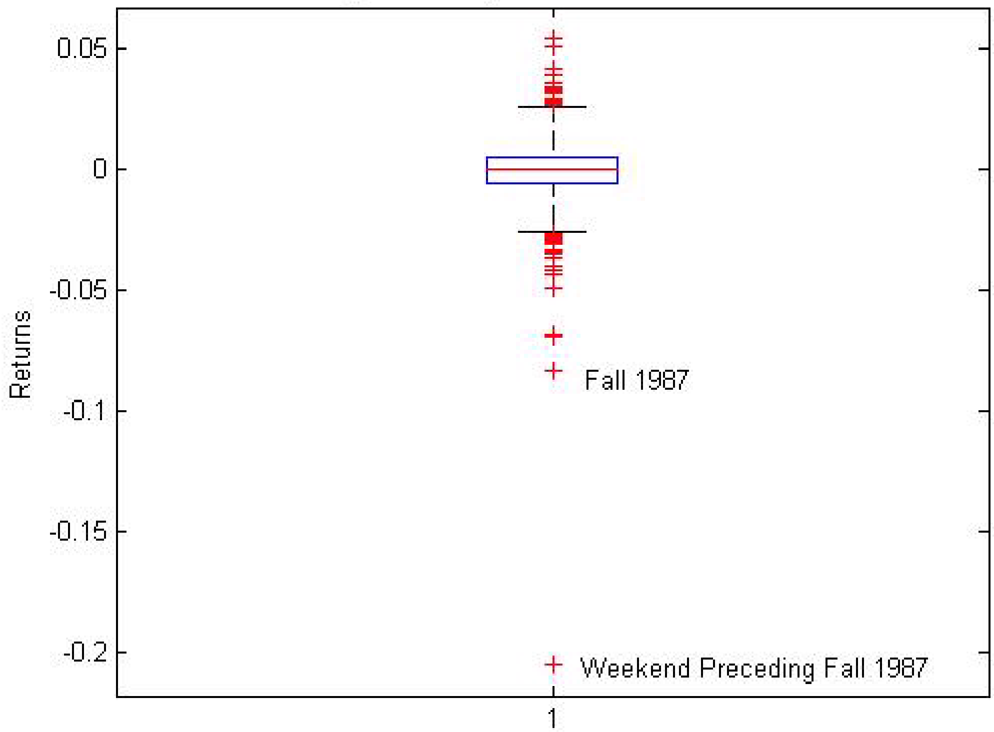
Table 2
Mean of S&P 500 Weekend Returns: Counterfactual Analysis

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.[9] 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.[10] 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 ($= 81,432 - 1.65 * 5,222$).[11] 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.

[9] With risk-averse investors, tests of market efficiency are not based (solely) on the mean return.

[10] 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 $n_1 > 15$.

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

The analysis of both means and medians compares adjustment weekends with all other weekends. As adjustment weekends comprise only 4 percent (= 78/2087) of all weekends, the discrepancy in relative sample sizes may hamper the ability to detect an adjustment effect.[12] 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

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

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 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.[13] 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

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

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 years in the sample, then there are n weekends (one in each year) that have an annual rank of 1.[14] 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

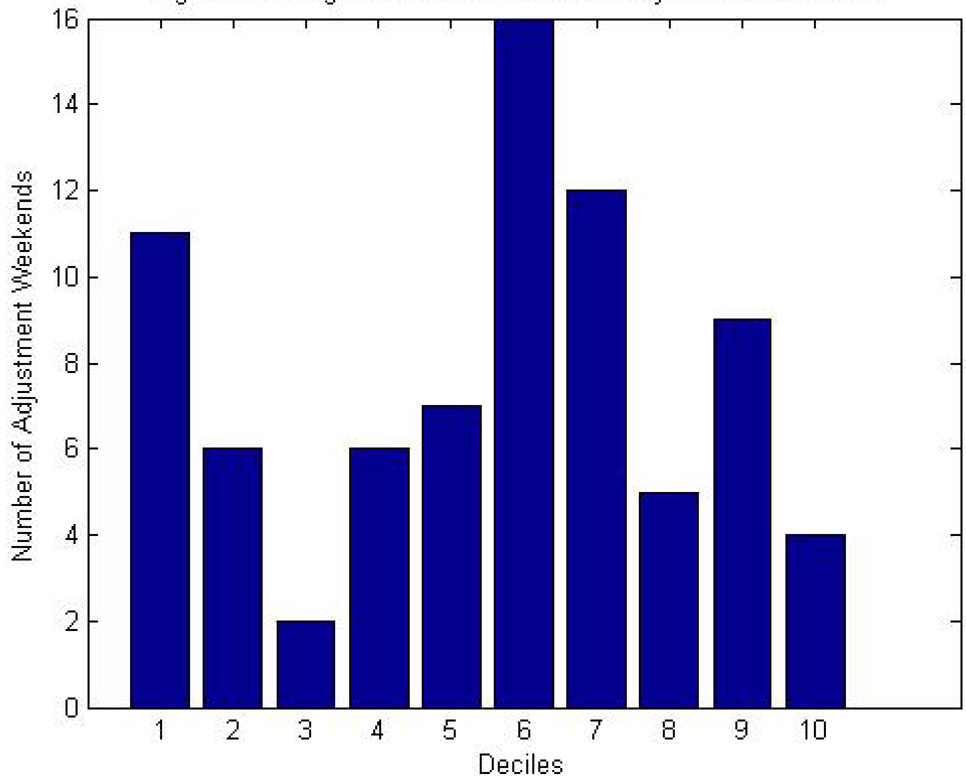
$$\tilde{\rho}\left(r_{y,(t)}\right) = \frac{t}{n_y},$$

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.[15]

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

[15] 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

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

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

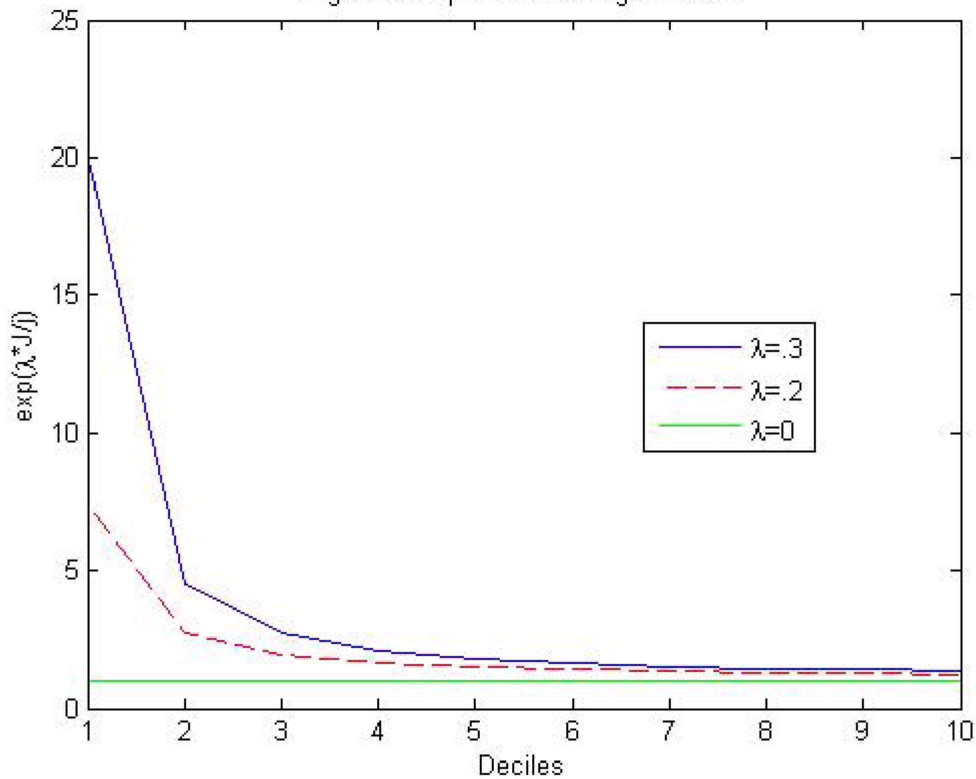
where $J = 10$ is the number of bins. Under the null hypothesis the histogram

---

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 A4 contains the complete list of the correspondence between bins and annual ranks.

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

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

where $\bar{P}$, the average number of ranks per bin, captures the factor of proportionality.

For the sample at hand $\hat{\lambda} = 0.0147$. 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.[16] 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 0.0269. As the sample estimate is below the upper confidence bound, we are unable to reject the null hypothesis of no adjustment effect.

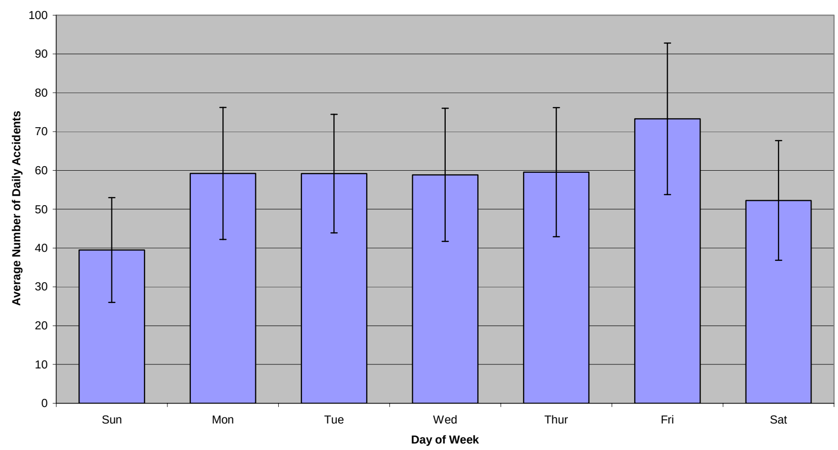
4 Traffic Accidents

4.1 Time Adjustment and Average Returns

We utilize data from the Oregon Department of Transportation in order to test the hypothesis that time adjustments impact driver performance. The dataset consists of daily traffic accident data (those not involving pedestrians) from 1987-2006. As Figure 4 indicates, the number of accidents varies greatly by day of the week in the sample from Oregon, with inherent differences across days of the week making estimates of average accidents on unadjusted days more noisy than if the comparison is made across Mondays. We concentrate on changes in the number of highway vehicle accidents that occur on Mondays, comparing Mondays after time adjustments with neighoring Mondays.

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

Figure 4. Average number of highway traffic accidents in Oregon by day of week (1987-2006)



Traffic accidents differ systematically from stock returns in one fundamental sense: variables that impact traffic accidents, such as weather and road conditions, do not vary uniformly across the year.[17] In order to best isolate the impact of time adjustments on performance from other potential heterogeneity, we identify two subsamples of Mondays in each year to act as our unadjusted controls. First, we utilize Mondays within one week of the adjustment Mondays (both one week prior and one week after), a strategy used previously (Coren, 1996; Vincent 1998). We also create an unadjusted sample of Mondays within two weeks of adjustment Mondays. The use of this control group is driven by Monk and Folkard (1976), who show that the effects of sleep desynchronicity can persist for up to one week.

Using the same methodology as was applied to the stock returns, we begin our analysis of the impact of time adjustments on traffic accidents by focusing on average daily highway accidents. We construct the average number of accidents for both the spring and fall adjusted Mondays and for unadjusted (neigbhoring) Mondays. We also construct a (joint t) test statistic of the null hypothesis that the average number of accidents for adjusted Mondays equals the average number of accidents on unadjusted Mondays.

[17] All but one of the outliers from a boxplot of all Mondays are observations from winter months. Furthermore, the road conditions during the accidents on these days have a far greater percentage of icy and wet conditions than other days in the sample.

Table 5 reports the constructed averages and the associated test statistics.[18] For the adjusted weekends, we report the overall average as well as separate averages for the spring and fall adjustments. The higher mean and larger standard deviation for the one-week control group indicate the presence of contamination due to the lingering effects of sleep desynchronicitiy noted by Monk and Folkard. For this reason, we focus on the two-week control group for the remainder of this section.

Table 5
Mean of Daily Highway Traffic Accidents in Oregon

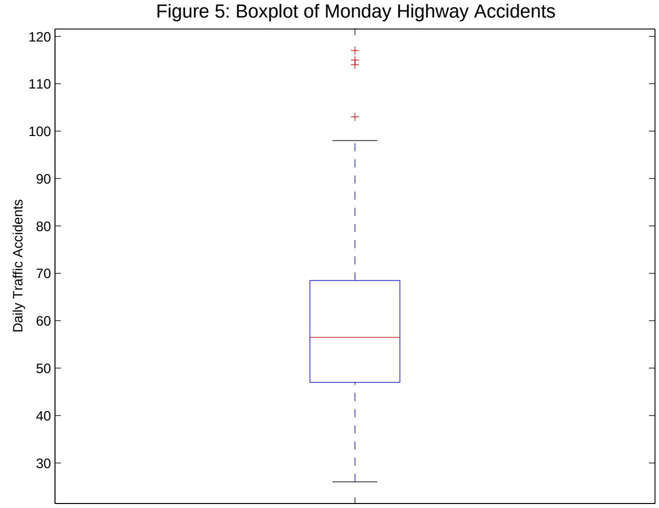
Dataset	Unadjusted	Adjusted	Spring	Fall	Joint t-test
Mondays +/-2 weeks of adjustment	57.47	64.05	63.35	64.75	2.23
	(17.47)	(14.75)	(14.94)	(14.91)	
Mondays +/-1 week of adjustment	60.78	64.05	63.35	64.75	0.93
	(19.61)	(14.75)	(14.94)	(14.91)	

4.2 Outlier Influence

Even for a two-week control group, which spans 5 weeks around each time adjustment, infrequent storms can lead to outliers in the number of recorded accidents. To determine the potential impact of outliers, we turn to the description provided in Figure 5. The presence of outliers above the median indicates that accurate testing of the proposed adjustment effect may require robust test statistics.

[18] The sample standard deviation appears below each sample mean.

Figure 5: Boxplot of Monday Highway Accidents



For the accident data, we utilize the rank of 1 to symbolize the Monday with the greatest number of highway traffic accidents. In the context of the Wilcoxon rank sum test, the expected value of the test statistic, given $n_a = 40$ and $n = 200$, is $4,020$, with a standard deviation of 327. Because we have utilized a rank of 1 to indicate the Monday with the highest number of highway traffic accidents, we will reject the null hypothesis that time adjustment has no impact on the size of Monday traffic accidents if the sum of adjustment ranks is less than the critical value. Under the alternative hypothesis that there will be more accidents following time adjustments the critical value is $3,481$. Our observed value of $3,144$ is well below the critical value. Accordingly, we are able to reject the null hypothesis that the time adjustments do not impact traffic accidents. This result holds across the subsample on Mondays within one week of time adjustments as well (the value of the statistic is -2.68).

The Wilcoxon rank sum test compares adjustment weekends with all other weekends, and the discrepancy in relative sample sizes may distort inference regarding the presence or absence of an adjustment effect. To account for this possibility, we form a sequence of differences between daily traffic accidents on Mondays two weeks prior to a time adjustment with the daily traffic accidents on the Monday immediately following a time adjustment.[19] If

[19] For the one-week control group, we construct the one week difference.

adjustment Mondays tend to have higher accidents, then these differences should be negative. The sign test counts the number of negative differences, to determine if there are more negative differences than would be expected under the null hypothesis that $med(d_t) = 0$. Given our twenty-year sample, the expected number of negative differences is 20, with a sample standard deviation of 3.16. The observed value of 27 negative differences is greater than the critical value of 25 ($= 20 + 1.65 * 3.16$) and so we are able to reject the null hypothesis that time adjustments do not increase traffic accidents.[20]

4.2.1 Time Variation

The statistics reported above are constructed under the assumption that the distribution of highway traffic accidents is unchanging over time. In practice, there have been a number of substantive policy changes over the past 20 years that might have impacted the number of traffic accidents within the state of Oregon. Table 6 provides examples of such changes and the year in which the change occurred. These changes might obfuscate the impact of the time adjustment on traffic accidents, as significant effects in one year could be washed out by the magnitude of accidents in another year.

Table 6
Policy Changes with Consequences for Highway Traffic Accidents

Year	Action
1990	Seatbelt rules enforced at state level
1997	Increase in threshold damage required to qualify as an accident
2000	Teen driving policy (may not drive with other teenagers as passengers)
2003	35% of state troopers are laid off from Oregon State Police
2004	Increase in threshold damage required to qualify as an accident

The variability in traffic-related law enforcement during the years of our sample, in addition to the increase in population and the overall number of vehicles associated with higher population levels is clearly depicted in Table 7.[21] The nonstationarity of the data is made obvious by the fact that identical annual ranks are associated with lower overall ranks as the years

[20] For the one-week control group, we find 25 negative differences, which lies at the critical value.

[21] To clearly display the variation in the accident distribution over time, the annual ranks in Table 7 are calculated over all Mondays in the year.

approach the present. The increase in accidents throughout time is also apparent in comparing higher annual ranks in more recent years with older, lower ranks and noticing that the more recent data tends to be ranked lower overall. The one exception to this trend is the observation from 2004, which has a low annual rank but higher overall rank than older data, a result that is certainly driven by the policy change of that year. These observations support the use of a test that allows for general distributional changes over time. To accommodate this need, we again turn to the idea of exponential tilting.

Table 7
Adjustment Weekends: Annual and Overall Ranks

Date	Annual	Overall	Date	Annual	Overall
1987, Spring	1	39	1998, Spring	6	80
1991, Fall	2	28	1994, Fall	7	189
2005, Fall	2	22	1995, Spring	7	108
1999, Spring	3	58	2002, Fall	7	81
2001, Spring	3	40	2001, Fall	10	123
2004, Fall	3	190			

Utilizing the same methodology as described in the section on stock returns, we develop estimates of the tilting parameter, λ_0 for our adjustment Mondays. We develop the ranks of adjustment Mondays based on a control group consisting of Mondays within two weeks of adjustment Mondays. For our adjustment Mondays, the tilting parameter is calculated based on the histogram presented in Figure 6. From this data, we find that $\lambda = 0.0423$. We again rely on Monte Carlo simulation for the development of a pertinent critical value, which is found to be 0.0184 after $10,000$ simulations. As such, we are able to reject the null hypothesis that time adjustment has no effect on daily highway traffic accidents.

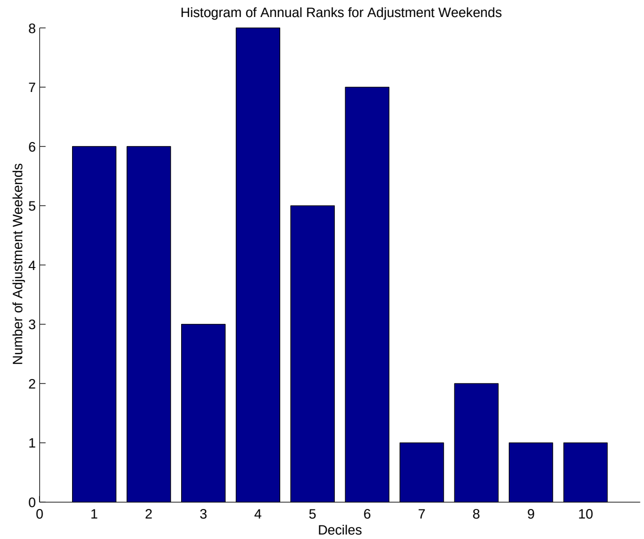


Figure 6. Histogram of highway traffic accidents in Oregon on adjustment Mondays.

We also develop estimates of the tilting parameter for a set of counterfactual Mondays, which are defined as those Mondays that occur one week before a time adjustment. From our counterfactual sample of Mondays, we estimate the tilting parameter based on the histogram shown in Figure 7. Using this dataset, we find that $\lambda = -0.0091$, which is not significantly different from zero. Not surprisingly, we are unable to reject the null hypothesis that time adjustments do not impact daily traffic accidents when using our counterfactual Mondays as the treatment group.

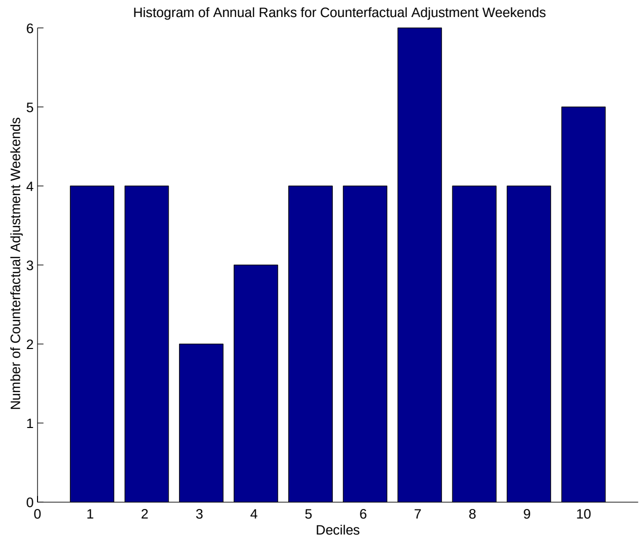


Figure 7. Histogram of highway traffic accidents in Oregon on counterfactual Mondays.

Allowing for nonstationarity in the distribution of traffic accidents, provides identification of the impact of time adjustments on traffic accidents. The results from our test statistic that is robust to nonstationarity confirm the findings suggested by our prior analysis. Daylight saving time seems to increase the number of traffic accidents, *ceteris paribus*.

5 Conclusion

We aim to identify the impact of time adjustments on human performance through an analysis of stock returns and traffic accidents. These variables have been studied in previous attempts to identify behavioral impacts of sleep desynchronicity. These two variables are also similar in that they are both subject to the presence of outliers driven by exogenous events and by nonstationarity.

Our study of stock returns attempts to discern a negative impact on stock returns arising from time adjustments. While the mean return shows a negative impact from time adjustment in the sample orginally analyzed, extension of the sample lessens the significance of the impact. 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 identical and independent draws from one distribution, we next examine the possible impact of time variation in the return 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.

With regard to the study of daylight saving time and traffic accidents, our results suggest that there is an impact of time adjustment on driver performance. While the effect is not detected by an analysis of means throughout the year, due to the presence of weather-related outliers amongst the unadjusted Mondays, when limiting the unadjusted weekends to represent days that are more similar to the time adjustment days, we do detect significant differences in mean daily accidents between the adjusted and unadjusted Mondays. Furthermore, an analysis focused on the median number of accidents provides consistent evidence of an impact. The impact of time adjustments is significant enough to withstand the nonstationarity of traffic accidents over our sample period.

Overall, our analysis seems to indicate that the impact of time adjustment on human performance may not be as severe as previous studies claim. The suggested impact on investor behavior is not detected once we control for the impacts of outliers and nonstationarity. However, we are able to confirm the previous finding that time adjustments do impact performance as manifested in driver performance. This finding should be of use to policymakers as decisions are made about further extending daylight-saving time in the United States.

6 Appendix: A Brief History of Daylight Saving Time

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 for Stock Returns

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: Adjustment Weekend Annual Ranks for Traffic Accidents

Year	Spring	Fall	Year	Spring	Fall
1987	1	6	1997	4.5	6
1988	5	1.5	1998	1	6
1989	5	6	1999	2	5.5
1990	6	4	2000	2	4.5
1991	6	2	2001	2	3
1992	4	5	2002	7	1
1993	4	5	2003	4	3
1994	3	1	2004	9	2
1995	2	10	2005	4.5	1
1996	8	4.5	2006	8	6.5

Table A4: 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 $6 \leq (t) \leq 10$.

References

1. Coate, D. and S. Markowitz, 2004. "The effect of daylight and daylight saving time on US pedestrian fatalities and motor vehicle occupant fatalities", *Accident Analysis and Prevention* 36, 351-357.

2. Coren, S., 1996. "Daylight Savings Time and Traffic Accidents", *The New England Journal of Medicine* 334, 924-925.

3. Coren, S., 1998. Reply to: "Effects of Daylight Saving Time on Collision Rates", *The New England Journal of Medicine* 339, 1169-1170.

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

5. Ferguson, S. A., D. F. Preusser, A. K. Lund, P. L. Zador, and R. G. Ulmer, 1995. "Daylight Saving Time and Motor Vehicle Crashes: The Reduction in Pedestrian and Vehicle Occupant Fatalities", *American Journal of Public Health* 85, 92-97.

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

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

8. 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.

9. Monk, T. H. and S. Folkard, 1976, "Adjusting to the changes to and from Daylight Saving Time", *Nature* 261, 688-689.

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

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

12. Vincent, A., 1998, "Effects of Daylight Savings Time on Collision Rates", *The New England Journal of Medicine* 339, 1167-1168.

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