EJW
A JOURNAL OF THE AMERICAN INSTITUTE FOR ECONOMIC RESEARCH

Econ Journal Watch
Volume 7, Number 1
January 2010, pp 91-98

Outliers and the Halloween Effect: Comment on Maberly and Pierce

H. Douglas Witte[1]

Abstract

In an article in the *American Economic Review*, Bouman and Jacobsen (2002) document a curious stock return pattern. They analyzed monthly returns in 37 countries from January 1970 through August 1998. For 36 of the 37 countries, mean monthly returns were lower over the period May to October than over the period November to April. The finding has been dubbed the Halloween effect and coincides with the older market adage, "Sell in May and go away."

In an *Econ Journal Watch* comment, Maberly and Pierce (2004) contend that "Bouman and Jacobsen's documentation of significant Halloween effects for U.S. equity returns appears to be driven by two outliers" (31). Maberly and Pierce identify the two outliers, without formalizing criteria, as October 1987, in which the U.S. and world equity markets declined markedly, and August 1998, in which the U.S. market declined amid the collapse of the hedge fund Long-Term Capital Management. They contend that the Halloween effect in the United States is rendered insignificant after an adjustment is made for the impact of these two outliers.

The primary contention of this paper is that Maberly and Pierce deal with outliers in an unsatisfactory way and that better methods of confronting influential data produce results very similar to those first reported in Bouman and Jacobsen. We use robust regression methods to estimate the Halloween effect for the same January 1970 through August 1998 monthly U.S. stock returns data analyzed by Maberly and Pierce. Contrary to the Maberly and Pierce findings, our results indicate statistical significance of a Halloween effect at levels similar to those

1. Assistant Professor of Finance and General Business, Missouri State University, Springfield, MO 65897.

originally reported in Bouman and Jacobsen. We find that the four biggest outliers aside from October 1987 and August 1998 all work against finding a Halloween effect. The effect of these additional outliers is, then, to obscure, rather than to drive, the Halloween effect.

Lucey and Zhao (2008) also examine Bouman and Jacobsen's work. They analyze monthly U.S. stock returns from 1926 to 2002 to determine the persistence of the Halloween effect over time. Using sub-period analysis, they find that the Halloween effect is not consistently significant over time for value weighted returns. Although the Lucey and Zhou study is not the focus of our comment, we have applied robust regression to the extended CRSP returns they examine. In untabulated results, the Halloween effect is statistically significant over the 1926 to 2002 time frame at a similar level to that found by Bouman and Jacobsen over the 1970 to 1998 time frame.[2]

Maberly and Pierce use a regression framework in which October 1987 and August 1998 are given a separate dummy variable in order to eliminate the impact of these two observations on the estimate of the Halloween effect. However compelling the case may be for Maberly and Pierce to control for October 1987 and August 1998, their framework is somewhat arbitrary in the number of outliers to control for. As we'll see in Table 2, November 1973 has very nearly the same degree of influence on the Halloween effect estimate as August 1998. During this time, President Nixon imposed price and marketing allocation controls on oil products in response to an oil boycott Arab nations had placed on nations that supported Israel. What makes one an outlier and the other not? Additionally, March 1980 is associated with widespread financial market uncertainty brought on by the precipitous fall in silver prices and concern over the large bank-financed silver positions of the Hunt brothers. Using the Maberly and Pierce framework, it is difficult to imagine a definitive way in which to determine the appropriate number of influential observations that are to be treated specially.

Bouman and Jacobsen employ the typical dummy variable regression technique which equates the regression equation to a simple means test. To determine whether the higher mean return over the November-April period might merely be the result of high January returns (the well known "January effect"), they modify their original regression specification to the following:

$$r_t = \mu + \alpha_1 S_t^{adj} + \alpha_2 Jan_t + \varepsilon_t \qquad \textit{Model 1}$$

2. The Huber and Tukey bisquare regression estimates of the Halloween effect over the 1926 to 2002 time frame are .52% (t = 1.69), and .49% (t=1.66), respectively.

The continuously compounded monthly return in month t is denoted by r_t. The adjusted Sell in May dummy, S_t^{adj}, takes the value of 1 in the period November to April, except in January, and 0 otherwise. The January dummy, Jan_t, takes the value of 1 in January and 0 otherwise. The intercept, μ, represents the average return over the May to October period. The coefficients α_1 and α_2 represent return relative to the May to October period. The size and statistical significance of α_1 relate to the question of a Halloween effect: Are mean returns over the period November-April, excluding January, significantly higher than during the period May-October?

Table 1. Regression estimates of the Halloween effect. Halloween effect represented by α_1. Monthly Returns Data for 1970:01-1998:08

(t-statistics in parentheses, returns in basis points)

Model 1: $r_t = \mu + \alpha_1 S_t^{adj} + \alpha_2 Jan_t + \varepsilon_t$ Model 2: $r_t = \mu + \alpha_1 S_t + \alpha_2 Jan_t + \alpha_3 D_t + \varepsilon_t$

	Regression Model					
	1	1	2	1	1	1
	Estimator					
	OLS	OLS	OLS	Huber	Tukey	Median
μ	49.9	42.3	68.0	67.2	71.5	47.6
	(1.39)	(1.21)	(2.08)	(2.16)	(2.2 9)	(.94)
α1	77.1	87.7	62.0	79.1	79.7	104.2
	(1.62)	(1.69)	(1.28)	(1.71)	(1.72)	(1.83)
α2	168.5	181.5	93.8	145.2	142.7	230.9
	(1.62)	(1.98)	(1.08)	(1.77)	(1.73)	(1.71)
α3			-2205.7			
			(-7.27)			

Notes: OLS estimates in Column 1 are based on continuously compounded MSCI index returns for the U.S. The t-statistics are derived from White (1980) heteroskedasticity-adjusted standard errors.

OLS estimates in Columns 2 and 3 are based on continuously compounded CRSP value weighted returns. The t-statistics are calculated from traditional coefficient standard errors.

Huber (Column 4) and Tukey bisquare (Column 5) estimation results are based on continuously compounded CRSP value weighted returns. The t-statistics are calculated from the estimated asymptotic covariance matrix (see Fox (1997) for details). The tuning constant in the Huber regression is 1.345. The tuning constant in the Tukey regression is 4.685.

Median regression estimates in Column 6 are based on continuously compounded CRSP value weighted returns. The t-statistics are derived from a bootstrapped sample covariance estimate.(see Koenker, et.al. (2000) for details)

The evidence Bouman and Jacobsen report for a Halloween effect in the United States is, at best, of marginal statistical significance. Although they only report the t-statistic of α_1, its value can be determined by using the current version of the value-weighted MSCI reinvestment index they used. The results are reported in Column 1 of Table 1. The estimate of α_1 is .771% and has a t-statistic of 1.62. The Bouman and Jacobsen paper reports a t-statistic of 1.61 for α_1, just outside the 10% significance level. Although Maberly and Pierce do not report the results of an identical regression run on the CRSP value weighted returns they analyze, we can easily do so. The results are reported in Column 2 of Table 1. The estimate of α_1 is .877% and has a t-statistic of 1.69, significant just inside the 10% level. Given the very similar t-statistics for α_1, inferences regarding the Halloween effect from the two different datasets are likely similar.

Some readers will surely view the U.S. market results reported by Bouman and Jacobsen as statistically insignificant given that a t-statistic of 1.61 does not indicate marginal significance at the 10% level much less significance at the more conventional 5% level. We do not disagree. However, the Maberly and Pierce critique does not focus on significance level issues. Given the nature of the Maberly and Pierce critique, it is more useful in the present context to use t-statistics and their associated p-values as gradated measures of the strength of the regression evidence as opposed to using them to make a formal rejection decision of a null hypothesis.[3] Thus, our analysis concerns whether consideration of outliers materially affects the strength of the regression evidence.

In order to determine the extent to which outliers drive the results, Maberly and Pierce formulate a regression specification augmented with a separate outlier dummy to accommodate October 1987 and August 1998:

$$r_t = \mu + \alpha_1 S_t + \alpha_2 Jan_t + \alpha_3 D_t + \varepsilon_t \qquad \textit{Model 2}$$

The dummy variable, D_t, takes the value of 1 on the two outlier dates and a value of 0 otherwise. The Sell in May dummy, S_t, takes the value of 1 in the period November to April, including January. The results of this alternative specification are reported in Column 3 of Table 1. The Halloween effect estimate is .62% with an associated t-statistic of 1.28 (p = .201). According to Maberly and Pierce, this recognition of outliers solves the Halloween effect puzzle and allows "stock market efficiency" to withstand another challenge.

3. Using p-values as a flexible measure of significance or relevance is often associated with the views of pioneering statistician R.A. Fisher. The use of p-values to make decisions about a hypothesis is more consistent with the views of Fisher's contemporaries Jerzy Neyman and Egon Pearson.

To identify the observations which most influence the Halloween-effect estimate and to determine the broader effect of unusual returns, we calculate an influence vector. An influence vector measures the influence of an observation by calculating OLS coefficient estimates with that observation omitted. Influence on coefficients is, heuristically speaking, the product of an observation's leverage (the degree to which the explanatory variable(s) is unconditionally unusual) and the observation's discrepancy (the degree to the response variable is unusual conditional on the explanatory variable(s)). For *Model 1*, all of the observations in months other than January have similar leverage given that six out of the 12 calendar months fall in the May-October period and five out of the 12 calendar months fall in the November-April period excluding January. The similar leverage of the pertinent observations (i.e., all those save January) implies we can attribute an observation's influence to the extent to which the observation is a *regression* outlier—an unusually large return in absolute value terms conditional on the explanatory variables. We focus on the change in the coefficient estimate on the Halloween indicator in regressions of *Model 1* using the same CRSP returns Maberly and Pierce use. We report the 10 biggest outliers in Table 2.

Table 2. The Impact of Outliers on the Halloween Effect. Change in α_1 Resulting from Omitting Outliers in Descending Order of Magnitude. Monthly CRSP VW Returns Data for 1970:01-1998:08.

				$\Delta\alpha_1$	
	Month	Year	Outlier Return	Outlier Omitted	Cumulative
Value Weighted Index Returns. α_1 = .877% in original regression					
1	October	1987	-25.5%	-.152%	-.152%
2	August	1998	-17.2%	-.103%	-.256%
3	November	1973	-12.9%	.100%	-.156%
4	March	1980	-12.8%	.099%	-.055%
5	April	1970	-11.1%	.088%	.035%
6	October	1974	15.3%	.087%	.122%
7	October	1978	-11.8%	-.071%	.048%
8	September	1974	-11.6%	-.071%	-.026%
9	August	1982	11.3%	.064%	.038%
10	October	1982	11.2%	.063%	.102%

Notes: α_1 is the estimated coefficient on the Halloween indicator from *Model 1*. The outlier return is the continuously compounded return to the CRSP value weighted index.

We find that, although the October 1987 and August 1998 observations have the biggest impact on α_1, the omission of any of the next four most influential returns, two of which are alluded to above, would all result in a higher estimated Halloween effect. Cumulatively, dropping the six most influential observations would add over 12 basis points to the Halloween effect estimate. Table 2 shows that while exclusion of the two biggest outliers would reduce the Halloween effect estimate by nearly one-third, omission of the 6 biggest (or even the 10 biggest) outliers would materially augment the estimate.

The results of omitting outliers demonstrated in Table 2 raise a question: How many outliers are appropriate to control for? One approach is to use formal statistical methods to indentify outliers, throw out the offending observations, and re-estimate the model using only the "clean" data. In our regression framework doing so would give outlying observations increased influence on the coefficient estimates up until some threshold point at which they are discarded and subsequently have no influence. When data are suspected of being contaminated or incorrectly coded, the approach may be appropriate. An alternative indirect approach is to include all of the data, but continuously downweight outlying data rather than simply discarding them. That is accomplished by robust regression. Given that our objective in this paper is to determine how unusual returns influence the Halloween-effect estimates rather than to determine an exact number of outliers, robust regression is a reasonable approach to resolving the issue.

The most common general method of robust regression is *M-estimation*, introduced by Huber (1964). Such methods are deemed robust because they produce estimates that are not as sensitive to outliers as OLS estimates. M-estimators and OLS typically display a similar sensitivity to observations with a high degree of leverage. In such high-leverage cases, bounded-influence estimators, such as least-trimmed-squares regression, are more effective. However, as noted above, the differential influence of the observations in our sample is almost entirely due to different degrees of discrepancy. M-estimation is effective in dampening the influence of observations with extreme discrepancy.

We use two robust M-regression methods to investigate the Halloween effect. First, we use the Huber (1964) estimator and we then apply the Tukey (1960) bisquare (or biweight) estimator. Additionally, we apply the median regression estimator that minimizes the sum of absolute errors (also known as least absolute deviation regression). All of these techniques limit the influence of outliers on regression results without arbitrarily selecting ex-ante the outliers to control for.

In Columns 4 through 6 of Table 1 we report the results of the three robust estimators of *Model 1* using the CRSP returns. The Huber and Tukey methods both produce Halloween-effect estimates that are slightly lower than the OLS estimate, but the associated t-statistics are slightly higher. This is often the case with robust regression. Outliers are a two-edged sword; extreme returns influence the OLS estimates but also tend to inflate the standard error of the regression estimates. The lower estimates do suggest, however, that there are a number of "semi-outliers" (those outside the 10 largest we report in Table 2) that nudge the OLS estimates of the Halloween effect upward. The median regression estimate suggests that the Halloween effect has a larger impact on the return distribution's median than on its mean. The α_1 estimate from median regression has a t-statistic of 1.83, the highest of any we find for the Halloween indicator.

All three robust regression methods indicate that the Halloween effect is significant at a level similar to their OLS counterpart and the original Bouman and Jacobsen results. Our robust regression analysis suggests that outliers do not drive the results of Bouman and Jacobsen. Marginal as the original results are, they remain marginally statistically significant using methods more resilient to outliers.

References

Bouman, Sven, and Ben Jacobsen. 2002. The Halloween Indicator, "Sell in May and Go Away": Another Puzzle. *American Economic Review* 92(5) (December):1618–1635.

Fox, John. 1997. *Applied Regression Analysis, Linear Models, and Related Methods*. Thousand Oaks, California: Sage Publications.

Huber, Peter J. 1964. Robust Estimation of a Location Parameter. *Annals of Mathematical Statistics* 35(1)(March):73-101.

Koenker, Roger, and Kevin Hallock. 2000. Quantile Regression: An Introduction. **Link**.

Lucey, Brian M., and Shelly Zhao. 2008. Halloween or January? Yet Another Puzzle. *International Review of Financial Analysis* 17(5)(December):1055–1069.

Maberly, Edwin D., and Raylene M. Pierce. 2004. Stock Market Efficiency Withstands another Challenge: Solving the "Sell in May/Buy after Halloween" Puzzle. *Econ Journal Watch* 1(1):29–46. **Link**.

Tukey, J.W. 1960. A survey of sampling from contaminated distributions. In *Contributions to Probability and Statistics*, ed. I. Olkin, S. Ghurye, W. Hoeffding, W. Madow and H. Mann, p. 448–485. Stanford: Stanford University Press.

White, Halbert. 1980. A Heteroskedasticity-consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica* 48: 817-838.

About the Author

H. Douglas Witte is Assistant Professor of Finance and General Business at Missouri State University in Springfield, Missouri. He earned a Ph.D. in Finance from the University of Arizona. Witte's work has been published in *The Journal of Finance*, *Studies in Economics and Finance*, and *Review of Business Research*. His email is witte@missouristate.edu.

Go to January 2010 Table of Contents

Go to Archive of Comments section