June 27, 2005

Professor Michael J. Cooper  
Purdue University  
West Lafayette, IN 47907  
mcooper@purdue.edu

Dear Professor Cooper:

Enclosed is a copy of the referee’s report on your paper “The other January effect (#2005-335),” with John J. McConnell and Alexei V. Ovtchinnikov. As you can see from his report, the referee feels that the current version of your paper is not publishable in the Journal of Financial Economics. The referee feels that you are analyzing a potentially interesting question, but he raises many questions about your analysis. He has made several suggestions that may help you improve the paper. In particular, I agree with him that the paper seems too long in its current form.

If you feel that you can respond to the referee’s criticisms, I will allow you to revise and resubmit your paper. However, there is no assurance that your paper will eventually be publishable in the JFE. Alternatively, you might decide to send the revised paper to another journal.

If you decide to resubmit this paper to the JFE, it would have to be refereed, so I would require a submission fee. Of course, if your paper is accepted for publication in the JFE, the last submission fee is refunded.

I hope these comments are useful to you. Thank you for submitting your paper to the Journal of Financial Economics.

Sincerely,

G. William Schwert  
Managing Editor
Review of “The Other January Effect”

Overall, this paper is not bad. There looks to be a robust effect here, and that is the paper’s main contribution. I am not as impressed with the hand wringing over some of the suspects such as Presidential election cycle and sentiment, but I suppose that jumping through such hoops does no great harm other than to drag out the story. Showing that the effect works across all types of stocks is more important, but not too surprising. Being able to forecast the FF risk premiums related to SMB is of some interest. But I have no idea what that means; the paper is short on interpretation of that. What does it mean? That requires interpreting what SMB itself means.

My two main comments below relate to tests of significance and to alternative ways to forecast the 11 month holdout return. I am not so sure that the January return itself is a unique predictor. The paper needs to establish that more convincingly by doing its best to forecast 11 month returns by other means, such as an autoregressive model using past annual returns combined with interest rate level, or add in term structure spread too if thou wilt..

I am not persuaded that the effect was there from 1825 to 1928, because of having to leave out 1929-1939, but others may disagree so I will not press that point.

I do have some quite strong statistical reservations because of multiple comparisons and non-normality of buy and hold returns.

1. I think that the tests of significance give a misleading impression of the level of significance. I’ll explain why. The January/Feb-Dec sequence has been observed after the fact. There were other trial possibilities, namely, Feb/March-January, March/April-February, etc. So there are really multiple trials occurring. So it is pretty easy to apply a Bonferroni correction. That may lead to understating the level of significance - I do not know. However, I think some of the tests will survive well enough. With a 5% level and 12 tests (for the 12 possibilities), there is a 46% chance of observing the result. The Bonferroni test lowers the 0.05 level to 0.0042 in this circumstance, and some of the tests in the tables survive this. In fact, the main ones of 1940-2003 are at 0.004 and 0.003, so in my mind they just pass the test. As a consequence, the results survive but in my mind they are not as impressive as they might seem at first blush. On the other hand, these 12 tests are not all independent because the months overlap, so the Bonferroni test may be too strict.

2. Another problem with the significance tests is the assumption of normality behind them. These 11 month returns are compounded returns, and their distribution is not normal. It will be skewed, and that’s going to confound the test in a second way. So there is going to be real doubt now as to significance, in my opinion.

3. These problems can be overcome by devising a better test or tests. For example, it might be somewhat persuasive if you did the other 11 tests and found that none showed anything, that March-January was independent of the sign of the prior February, etc. These tests would have the problems of points 1 and 2 but they would still help. I see that this has been done in the paper. However, my intuition is that there exists a bootstrapping or simulation test that will overcome all objections. Maybe some of the things done in the long-horizon literature will help
in this respect. Just thinking out loud and I could be way off base, you might take all the Feb-Dec monthly returns in 11 bins and the January returns in another bin 2. Then, you sample 11 times and construct a buy and hold return, and select 1 from bin 2 at random. This gives one observation. Repeat this 10,000 times. Then check the holdout returns when Jan is + versus when Jan is -. Even though this experiment randomizes, some fraction of times, the Jan + conditioning will produce a higher return than the Jan - conditioning, so you can get an experimental alpha level. I leave the design to you.

Another alternative is to repeat the tests using a CAR instead of a buy and hold return.

These procedures help on distributional issues, but do not resolve the multiple comparison problem. What we need to know is under the null, how many times you would see the posited relation. The null, in my opinion, involves looking at all the 12 month sequences. That’s not hard. Again set up bins for each month as earlier. But this time also sample the other 11 sequences and out of all the sequences, find out how many times any of them produce a higher 11 month return conditional on a starting + month than on a starting - month. I leave the design to you, because I might be wrong here. But I think I’m conveying the idea of how to do a proper test.

And again, doing this with a CAR is feasible.

4. Now this is a no-lose proposition for you. If the other January effect survives, that’s a win. If it does not survive, then you write the paper to show that a seeming success using the tests in the current paper is really not a success at all, and everyone must be careful in the future, etc. 5. I recollect a simple test that Fama (and French?) once did in which annual returns were regressed on annual returns of the prior 3 years, and there was a good amount of negative correlation. This test should be one of your robustness tests. Sorry I do not have the reference. If there have been some bad years, you are more likely to get a good January and also a good Feb-Dec. Hence, the other January effect could be subsumed by this negative autocorrelation of long-run returns. In the same vein, a down period will produce more tax-loss selling at the end of the year, so that the following January (and next few months) are more likely both to be positive. Anything involving January is likely to be proxying for prior down markets and tax-loss selling, in my opinion. Therefore, you really should not see the effect before 1928 or so, when the tax laws altered (again my memory on this is rusty.) If you do see it, then maybe there is a harvest explanation. A good crop at the year ned led to a good January and rest of year.

Also, after good years, in which capital gains are deferred into the following year, January tends to be bad as the selling is realized.

In all of these cases, you should see diminishing effects as the base month is moved from Jan. to Feb. to March, etc. A graph showing the effect as the base month advances might be instructive. (Looking at your data, I think this is there but is only very noisily there.)

6. Although the streetlore quotes are accurate, if memory serves, Hirsch and others used to compute the full year return including January in getting the effect. They did not realize how much this biased the test in favor of the other January effect. So their percentages are overstated.
is my guess.

7. Although the tests need to be built up statistically speaking, the data pass the eyeball test. The counts of years bad after good and bad Januaries is impressive: 14 down years after the 23 down Januaries, and only 5 down after the 41 up Januaries. Very lopsided. It’s almost too lopsided. And I suspect (as in comment 5) that the January return is a signal like the annual returns of the prior 1 to 3 years. It may be partly subsumed by that alternative predictor, but the tax-loss effects could give it some oomph of its own.

8. I find puzzling the increase in spread (11 month conditional on +Jan minus 11 month conditional on -Jan) from raw to excess returns. It’s almost 3%. Since the +Jan one falls by about 3%, this means that the risk-free rates average almost 6% higher following the -Jan. Since rates are autocorrelated and very low from 1840 to 1953, a lot of this must be concentrated in the high interest rate period. And it means that the bad years, which are the years after the - Januaries, are also very high interest years. This means that the bad returns may reflect high initial interest rates. Fama showed this effect a long time ago, so now we have another hypothesis to contend with, namely, that the negative Januaries occur in high interest years and the rest of the year has low returns too. This should be looked at. But I would not use a linear regression predictive model to do this. It asks too much of the data and relation. I’d use the January interest rate or perhaps its level compared to the average over the prior 12 months as a +, - variable.

9. I do not understand why it is legitimate to remove the 1929-1939 interval in assessing the earlier period. Investors did not know at the time that this period would later be regarded by Bernanke as unusual. Had they speculated on the basis of the other January effect, they would have not done well. This could be like writing puts. Most of the time you make some money, but every so often you lose a lot.

10. The way that the multifactor model is used bothers me. It is used as a predictive model to get excess returns and is updated each month. A simpler way more in accord with the categorical nature of the other January effect is to create a +,- indicator based on the values provided in the base month by the macro variables. Earlier, I suggested that Fama had a simple effective model that used the past 3 years of stock returns, and also a model that used the rate of interest in a base month. These could be combined into one model perhaps. Then, if these variables forecast a + forward 11 month return, the indicator is a +, and if they forecast a negative return, the indicator is a -. The model has future 11 month return as a function of past three year annual returns and short-term interest rate. Combine the fitted coefficient values with the most recent actual values to forecast the holdout 11 month return. Both his and the model in the paper share the YLD variable.

Actually, as said earlier, using 3 past annual stock returns as a predictor is not problematic. Using interest rate in there has a problem because the yield is not stationary, and so its coefficient is really not stable. Maybe YLD(t-1)/average yield over prior 12 months will be better. I’d think about this and try some different models.