Are return seasonalities due to risk or mispricing?

Internet Appendix

Matti Keloharju[a,b,c], Juhani T. Linnainmaa[d,e,*], Peter Nyberg[a]

[a] *Aalto University School of Business, Ekonominaukio 1, 02150 Espoo, Finland*
[b] *Center for Economic Policy Research, 33 Great Sutton Street, London, EC1V 0DX, United Kingdom*
[c] *Research Institute of Industrial Economics, Grevgatan 34, SE-102 15 Stockholm, Sweden*
[d] *Dartmouth College, Tuck School of Business, 100 Tuck Hall, Hanover NH 03755, USA*
[e] *National Bureau of Economic Research, 1050 Massachusetts Avenue, Cambridge MA 02138, USA*

[*] Corresponding author. Tel.: +1-603-646-3160. E-mail address: Juhani.T.Linnainmaa@tuck.dartmouth.edu (J. T. Linnainmaa).

Appendix A. Robustness checks

A.1 Portfolio sorts based on same- and other-month returns

Table A1 measures seasonalities and seasonal reversals by sorting stocks into decile portfolios by their average same- and other-month returns. We cross-sectionally demean the data before computing the same- and other-month returns using up to 20 years of historical data. We report average returns for the decile portfolios and average returns and three alphas for hedge portfolios that are long the top decile and short the bottom decile. The three asset pricing models are the Fama and French (1993) three-factor model, the Carhart (1997) four-factor model, and the Carhart model augmented with the long-term reversals factor.

Table A1
Average returns and alphas for portfolios sorted by same- and other-month returns.

This table presents average monthly returns for decile portfolios sorted by the average same- and other-month returns. We form portfolios at the end of each month t and compute their value-weighted returns in month $t + 1$. We cross-sectionally demean the data before computing the same- and other-month returns using up to 20 years of historical data. The bottom of the table reports average returns and alphas from three asset pricing models: Fama and French (1993) three-factor model (FF3), Carhart (1997) four-factor model (FF4), and the Carhart model augmented with the long-term reversals (FF4 + LTREV) factor. We compute returns for the portfolios from January 1963 through December 2016.

	Sorting variable	
Decile	Same-month return	Other-month return
Low	0.46	1.64
2	0.70	1.28
3	0.63	1.17
4	0.80	1.15
5	0.78	0.98
6	0.94	0.93
7	1.18	0.83
8	1.18	0.87
9	1.22	0.92
High	1.49	0.77
High − low hedge portfolio		
Average return	1.03	−0.87
	(7.19)	(−5.55)
FF3 $\hat{\alpha}$	1.08	−0.53
	(7.74)	(−4.91)
FF4 $\hat{\alpha}$	1.09	−0.54
	(7.21)	(−4.73)
FF4 + LTREV $\hat{\alpha}$	1.09	−0.53
	(7.19)	(−4.74)

A.2 Robustness check: excluding the month of January

Table 2 reports estimates from regressions that predict the cross section of monthly returns using average returns computed over different horizons and decomposed into the same- and other-month parts. Table A2 reports estimates from similar regressions, which use data from all months except January.

Table A2
Average annual and non-annual returns in Fama-MacBeth regressions excluding January.

This table presents average Fama and MacBeth (1973) regression slopes and their t-values from cross-sectional regressions that predict monthly returns in all other months except for January. The regressions predict returns using the average of all past returns, average of same-month returns, average of other-month returns, and the difference between the average same- and other-month returns. We estimate each specification as a univariate regression. Row "1" uses data from $t-12$ through $t-1$; "2–5," from $t-60$ through $t-13$; and so forth. The regressions use data from January 1963 through December 2016 for all stocks, all-but-microcaps, and the 48 value-weighted Fama-French industries. Microcaps are stocks with market values of equity below the 20th percentile of the NYSE market capitalization distribution. We cross-sectionally demean returns when computing the averages of past returns. Regression estimates are multiplied by one hundred.

	Construction of historical average return							
	All		Same-month return		Other-month return		Same-month − other-month	
Years	$\hat{b}$	$t(\hat{b})$	$\hat{b}$	$t(\hat{b})$	$\hat{b}$	$t(\hat{b})$	$\hat{b}$	$t(\hat{b})$
All stocks								
1	6.69	3.98	1.39	5.61	5.55	3.44	0.80	3.31
2–5	−7.17	−2.48	1.59	3.31	−9.35	−3.50	2.24	5.24
6–10	−8.25	−3.09	2.69	5.22	−10.91	−4.43	3.19	6.74
11–15	2.22	0.90	3.11	5.05	−1.36	−0.58	2.97	5.12
16–20	−2.16	−0.86	2.18	3.38	−4.61	−1.92	2.31	3.75
All-but-microcaps								
1	9.93	4.27	1.56	4.10	8.86	3.99	0.69	1.83
2–5	−8.19	−2.47	2.06	3.33	−10.82	−3.41	2.86	4.87
6–10	−6.64	−2.12	3.26	5.21	−9.92	−3.38	3.76	6.41
11–15	0.19	0.07	2.81	4.13	−2.45	−0.96	2.78	4.28
16–20	−3.44	−1.31	2.47	3.36	−5.79	−2.26	2.76	3.90
Industries								
1	26.30	5.70	4.80	4.15	24.06	5.46	1.78	1.55
2–5	1.33	0.19	1.78	0.86	−0.34	−0.05	3.42	1.79
6–10	−22.30	−3.01	5.04	2.39	−27.48	−3.78	7.15	3.45
11–15	−2.13	−0.33	6.37	3.23	−6.39	−1.03	6.86	3.48
16–20	−9.98	−1.51	3.77	1.66	−11.48	−1.84	4.61	2.17

A.3 Lists of commodities and countries

Table A3 lists the 24 commodities and 15 countries used in Panels B and C of Table 4.

Table A3
Lists of commodities and countries.

The regressions in Panels B and C of Table 4 use monthly and daily data on commodities and country equity indices. This table lists the 24 commodities (Panel A) and 15 countries (Panel B) used in this analysis. The commodity data are the S&P GSCI index returns, retrieved from Bloomberg, and the country equity data are the MSCI indices retrieved from Datastream.

Panel A: Commodities

Sector	Commodity	Bloomberg index ticker
Energy	Crude oil	SPGCCLP
	Brent crude	SPGCBRP
	Heating oil	SPGCHOP
	Gasoil	SPGCGOP
	Unleaded gasoline	SPGCHUP
	Natural gas	SPGCNGP
Softs	Sugar	SPGCSBP
	Cocoa	SPGCCCP
	Coffee	SPGCKCP
	Cotton	SPGCCTP
Grains	Wheat	SPGCWHP
	Kansas wheat	SPGCKWP
	Corn	SPGCCNP
	Soybeans	SPGCSOP
Livestock	Live cattle	SPGCLCP
	Feeder cattle	SPGCFCP
	Lean hogs	SPGCLHP
Industrial Metals	Copper	SPGCICP
	Zinc	SPGCIZP
	Aluminium	SPGCIAP
	Nickel	SPGCIKP
	Lead	SPGCILP
Precious Metals	Gold	SPGCGCP
	Silver	SPGCSIP

Panel B: Countries	
Country	MSCI index ticker
Austria	MSASTRL
Belgium	MSBELGL
Canada	MSCNDAL
Denmark	MSDNMKL
France	MSFRNCL
Germany	MSGERML
Italy	MSITALL
Japan	MSJPANL
Netherlands	MSNETHL
Norway	MSNWAYL
Spain	MSSPANL
Sweden	MSSWDNL
Switzerland	MSSWITL
United Kingdom	MSUTDKL
United States	MSUSAML

A.4 Robustness check: seasonalities, seasonal reversals, and mood betas in select months

Table A4 reports estimates from regressions that are similar to those in Panel A of Table 5, Columns 4 and 8. Whereas the regressions in Panel A of Table 5 use the full sample to estimate the regressions, Table A4 regressions use either January and March or September and October returns. Hirshleifer, Jiang and Meng (2020) identify January and March as high-mood months and September and October as low-mood months.

Table A4
Seasonalities and seasonal reversals in monthly Fama-MacBeth regressions: association with high- and low-mood betas.

This table presents average Fama and MacBeth (1973) regression slopes and their t-values from cross-sectional regressions that predict monthly returns. The samples and the main explanatory variables are the same as in Tables 1 and 3. The regressions control for the historical mood betas of Hirshleifer, Jiang and Meng (2020). These regressions are similar to those reported as Regressions 4 and 8 in Panel A of Table 5 except that the sample is restricted either to January and March or September and October returns. Regression estimates are multiplied by one hundred.

Explanatory	All stocks		All-but-microcaps	
variable	Jan and Mar	Sep and Oct	Jan and Mar	Sep and Oct
log(ME)	−0.43	−0.01	−0.17	−0.05
	(−5.72)	(−0.15)	(−2.23)	(−0.76)
log(BE/ME)	0.17	−0.03	0.38	−0.16
	(1.51)	(−0.25)	(3.41)	(−1.50)
r_1	−11.41	−5.03	−8.18	−3.50
	(−10.23)	(−5.82)	(−7.03)	(−3.23)
$r_{12,2}$	−1.22	1.15	−0.80	0.92
	(−2.87)	(3.47)	(−1.68)	(2.29)
$r_{60,13}$	−0.46	−0.02	−0.33	−0.10
	(−4.92)	(−0.27)	(−3.50)	(−1.63)
$\bar{r}_{\text{same-month}}$	8.97	5.14	9.81	8.65
	(4.60)	(3.34)	(4.90)	(4.36)
$\bar{r}_{\text{other-month}}$	−35.59	4.94	−15.93	1.89
	(−3.77)	(0.61)	(−1.68)	(0.20)
$\hat{\beta}_{\text{mood}}$	1.13	−0.81	1.04	−0.98
	(3.02)	(−2.92)	(2.94)	(−2.88)

A.5 Seasonalities and limits to arbitrage: conditional sorts

Table A5 reports value-weighted returns for portfolios sorted by return seasonality and three proxies for limits to arbitrage. We first assign stocks into portfolios by firm size (Panel A), idiosyncratic volatility (Panel B), and size-adjusted institutional ownership (Panel C) and then, conditional on this assignment, by return seasonality. Return seasonality is the difference between the average same-month and other-month returns over the prior 20-year period.

Table A5
Seasonalities and proxies for limits to arbitrage: conditional sorts.

This table assigns stocks into portfolios by firm size (Panel A), idiosyncratic volatility (Panel B), and size-adjusted institutional ownership (Panel C) and then, conditional on this assignment, by return seasonality. Return seasonality is the difference between the average same-month and other-month returns over the prior 20-year period. At the bottom of the table, we report the average fraction of stocks that belong to either the high or low return-seasonality portfolio. Because the portfolio sorts in this table are conditional, a total of 40% of stocks are always, by construction, in the extreme quintiles.

Panel A: Firm Size

Return	Firm size		
seasonality	Microcap	Small	Large
Low	0.69	0.83	0.58
2	1.15	1.14	0.75
3	1.28	1.32	0.88
4	1.46	1.45	1.10
High	1.70	1.53	1.34
High − low	1.01	0.71	0.76
	(9.90)	(7.04)	(7.47)
Fraction of stocks in the extreme quintiles	40%	40%	40%

Panel B: Idiosyncratic volatility

Return	Idiosyncratic volatility				
seasonality	Low	2	3	4	High
Low	0.70	0.67	0.47	0.44	−0.28
2	0.72	0.84	0.82	0.67	0.29
3	0.91	0.93	0.96	1.09	0.46
4	1.14	1.17	1.34	1.37	0.75
High	1.43	1.52	1.65	1.58	0.83
High − low	0.73	0.85	1.18	1.15	1.11
	(6.15)	(6.21)	(7.08)	(6.19)	(4.80)
Fraction of stocks in the extreme quintiles	40%	40%	40%	40%	40%

Panel C: Size-adjusted institutional ownership

Return	Institutional ownership				
seasonality	Low	2	3	4	High
Low	0.32	0.46	0.66	0.73	0.64
2	0.76	0.90	0.94	1.00	0.93
3	1.13	1.00	1.06	1.40	1.16
4	1.25	1.28	1.46	1.30	1.45
High	1.32	1.57	1.45	1.66	1.77
High − low	1.01	1.11	0.78	0.93	1.12
	(4.01)	(5.40)	(4.18)	(4.46)	(4.70)
Fraction of stocks in the extreme quintiles	40%	40%	40%	40%	40%

A.6 Measuring the speed of seasonal reversals

In the Section 2 model, we assume that seasonalities add up to zero over the calendar year. This assumption, in turn, implies that seasonal reversals are evenly distributed over the calendar year: a high seasonal return in one month is offset, on average, by lower returns over the other 11 months. Seasonalities and seasonal reversals in the actual data perhaps are not consistent with this assumption. Seasonalities and seasonal reversals could, for example, be clustered. If a stock's return is seasonally "too high" one month, its return could be "too low" by the same amount already the next month. In this section, we use simulations to examine how different assumptions about the timing of seasonalities and seasonal reversals alter the pattern of Fama-MacBeth regression coefficients.

Fig. A1 simulates data using three different assumptions about the timing of seasonal reversals. The first specification, evenly distributed reversals, maintains the assumption from Eq. (10) that return seasonalities are unrelated save for the additional assumption that they satisfy the adding-up constraint in Eq. (11). We report its results in Panel A. The second specification (Panel B), immediate reversals, assumes that seasonalities always completely reverse in one month. Here, each stock has one month, drawn randomly, in which its seasonal return is high or low, and this seasonal return fully offsets the following month. For example, if a stock earns a seasonal return of 5% in April, it earns a seasonal reversal of -5% in May. The third specification (Panel C), long-lasting seasonalities with evenly distributed reversals, assumes that seasonalities can last up to three months but then reverse over the rest of the year. We assume that one-third of stocks earn a high or low seasonal return in one month; another one-third of stocks earn the seasonal return over a two-month period; and the remaining one-third of stocks earn the seasonal return over a three-month period. For example, a stock that belongs to the three-month group could earn seasonal returns of +1% in July, August, and September. This stock then earns a seasonal return of $-\frac{1}{3}\%$ in each of the remaining nine months of the year. In all cases, we draw each stock's seasonal period at random.

We create 50 simulated samples for each specification using the same parameters as in Fig. 2 except that we set momentum and long-term reversals to zero and alter the method by which we generate the return seasonalities. Each sample has the same dimensions as the actual return data. Using the simulated data, we estimate Fama-MacBeth regressions that predict month $t+1$ returns

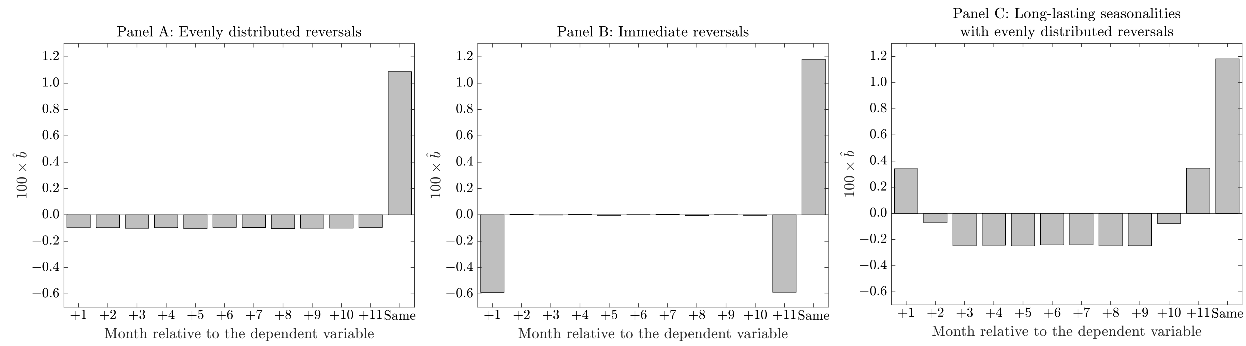

Fig. A1. The timing of seasonal reversals: Fama-MacBeth regressions using simulated data. This figure shows how different assumptions about the timing of seasonal reversals alter the pattern of Fama-MacBeth regression coefficients. We simulate data from the model described and calibrated in Sections 2.2 and 2.3. Specification "Evenly distributed reversals" (Panel A) maintains the assumption in Eq. (10) that return seasonalities are unrelated but add up to zero. Specification "Immediate reversals" (Panel B) assumes instead that each stock earns a high or low seasonal return in one month that then fully reverses the next month. Specification "Long-lasting seasonalities with evenly distributed reversals" (Panel C) assumes that seasonalities can last up to three months. One-third of stocks earn a high or low seasonal return in one month; another one-third of stocks earn the seasonal return over a two-month period; and the remaining one-third of stocks earn the seasonal return over a three-month period. These seasonalities reverse over the rest of the year. In Panel B and Panel C, we randomly assign the seasonal periods to each stock. We generate 50 simulated samples for each specification using the same parameters as in Fig. 2 except that we set the momentum and long-term reversals effects to zero. These simulated data have the same dimensions as the actual return data. Using the simulated data, we estimate Fama-MacBeth regressions that predict month $t+1$ returns using the average historical return in each month: average January return, average February return, and so forth. We order these 12 right-hand-side variables relative to the month of the dependent variable. For example, in a regression that predicts March returns, month +1 is the average historical April return; month +2 is the average historical May return; and "Same" is the historical average March return. We plot the average Fama-MacBeth regression coefficients across the 50 simulated samples.

using average historical returns in different calendar months: average January return, average February return, and so forth. We compute each average return using up to 20 years of data. We order the 12 right-hand-side variables in these regressions by their offset relative to the month that the regressions predict. For example, in a regression that predicts March returns, month +1 is the average historical April return; month +2 is the average historical May return; and "Same" is the historical average March return. We plot the average Fama-MacBeth regression coefficients across the 50 simulated samples.

The average same-month return positively predicts the cross section of returns in the same way in all three specifications of Fig. A1. The assumption about the behavior of seasonal reversals affects the pattern of the 11 other-month coefficients. In Panel A, all other-month returns predict returns with coefficients that are of approximately equal magnitude. This pattern results from the assumption that the seasonalities are unrelated to each other except for satisfying the adding-up constraint. A high seasonal return in one month implies that the return in at least one of the other

11 months must be seasonally low.

In the "Immediate reversals" specification of Panel B, the +1 and +11 returns predict returns with negative signs, and the remaining coefficients are close to zero. If a stock's return is unusually high in March its return in either February or April must be unusually low. The returns that the stock earns over the remaining year, from May through January, would be uninformative about its March returns because seasonalities and seasonal reversals always cancel out over a two-month period.

In Panel C, the average returns in the nearest months positively predict returns and the returns in the second nearest months predict returns negatively but not as much as the remaining seven months of the year. The nearest-month coefficients are positive because seasonalities in this specification last, on average, for two months. If a stock earns a high seasonal return in February, there is a two-thirds probability that this stock's return in January or March is also high. The coefficients farthest away from the current month are negative because seasonalities still reverse. A high seasonal return in February, for example, is still offset by low returns at least from May through November because seasonalities in this specification cannot last for more than three months.

The coefficient patterns in Fig. A1 are symmetric because seasonalities and seasonal reversals themselves are symmetric. Without additional information about a stock's fundamental value, we cannot say that a stock has, for example, a 2% seasonality in February and a −2% seasonal reversal in some other month. The −2% return could, just as well, be the initial seasonality and the 2% return the reversal that follows. This symmetry property is the same as that with autocovariances. If a stock's return today positively predicts returns k months later, $\text{cov}(\tilde{r}_{i,t}, \tilde{r}_{i,t+k}) > 0$, then a stock's return k months ago positively predicts returns today, $\text{cov}(\tilde{r}_{i,t-k}, \tilde{r}_{i,t}) > 0$. In the actual data, the pattern is not perfectly symmetric because samples are finite. Returns early in the sample are used only to predict future returns and returns late in the sample are used only on the regression's left-hand side.

A.7 Correlations between noisy factors: a simulation analysis

Table 7 shows that the correlation between the seasonality and seasonal reversals factors is negative, $\hat{\rho}_{\text{ann,nann}} = -0.15$. This correlation is negative because these factors load on the market with different signs. Panel A of Table 8 shows that the seasonality factor's market beta is 0.06 (t-value = 2.34) and that of the seasonal reversal factor is -0.18. Even though the factors are unconditionally negatively correlated, they correlate positively on a market-adjusted basis. In the spanning regression of the seasonality factor on the seasonal reversal factor (see Column 3 in Panel A of Table 8), the coefficient for the seasonal reversal factor is 0.15 (t-value = 2.51). This positive partial correlation is consistent with our hypothesis that these two factors are noisy versions of the same signal: seasonal variation in expected returns.

In this section, we use a simulation analysis to quantify the extent to which the noise in the two signals (the average same-month return and the average other-month return) pulls the factors' correlation toward zero. That is, even though both signals are about $\text{E}_t(\tilde{r}_{i,t+1})$, they are both noisy. What determines the correlation between the signals and, by extension, the factors, is the amount of noise. Because we compute the two signals using non overlapping data, same months and other months, we expect the noise terms in these signals to be approximately uncorrelated. We simulate factors that, similar to the actual seasonality and seasonal reversal factors, capture differences in expected returns. We create these factors as follows.

1. We create two random signals by adding uncorrelated and normally distributed noise to month $t+1$ returns: $x^1_{i,t} = r_{i,t+1} + \tilde{n}$ and $x^2_{i,t} = r_{i,t+1} + \tilde{n}'$. We set the standard deviation of $\tilde{n}$ to 10 and its mean to zero. The resulting signal-to-noise ratio, as measured by the R^2 from a regression of $r_{i,t+1}$ on $x_{i,t}$, is 0.03%.
2. We construct two HML-style factors, F^1 and F^2, by sorting stocks into six portfolios by market capitalization and either x^1 or x^2. We then compute value-weighted returns on these portfolios and define the factors as the average return on the two high portfolios minus the average return on the two low portfolios.
3. We regress F^1 against the three factors of the Fama and French (1993) three-factor model and against this model augmented with the other noisy factor, F^2. The two regression models are

$$F^1_t = a + b_{\text{mkt}}\text{MKTRF}_t + b_{\text{smb}}\text{SMB}_t + b_{\text{hml}}\text{HML}_t + e_t \quad \text{(A1)}$$

and

$$F_t^1 \quad = \quad a' + b'_{\text{mkt}}\text{MKTRF}_t + b'_{\text{smb}}\text{SMB}_t + b'_{\text{hml}}\text{HML}_t + b'_{\text{f}}F_t^2 + e'_t. \qquad \text{(A2)}$$

4. We record the intercepts, slope, and the t-values associated with these estimates from regressions (A1) and (A2).

We repeat these steps one thousand times. In these simulations, the two signals are, by construction, noisy measures of expected returns. The two factors therefore earn significant average returns and alphas. The average three-factor model alpha, $\hat{a}$, in Eq. (A1) is 25 basis points per month, and the average t-value associated with this estimate is 8.27. In the regression that controls for the other factor, the average alpha is 23 basis points per month and the average t-value is 7.39. Similar to the actual data, the two factors do not correlate significantly. The average slope $\hat{b}'_{\text{f}}$ in Eq. (A2) is 0.06 with an average t-value of 1.60.[1]

These simulated factors therefore resemble the actual seasonality and seasonal reversal factors. Even if both factors capture seasonalities in expected returns, the correlation between the resulting factors is low as they are infused with noise.

[1] Because our simulations add normally distributed noise, the resulting estimates are well behaved: the distributions of $\hat{a}$, $\hat{a}'$, $\hat{b}'_{\text{f}}$, and the t-values associated with these estimates are centered around the averages. The median alphas are 25 and 23 basis points; the median t-values associated with these alphas are 8.27 and 7.39, respectively; the median slope on F_t^2 is 0.06; and the median t-value associated with this slope is 1.62.