Online Appendix:
Text-Based Industry Momentum

Gerard Hoberg and Gordon Phillips

Online Appendix Table A1
Fama MacBeth Return Regressions (Various SIC-based Momentum Variables)

Fama-MacBeth regressions with the monthly stock return as the dependent variable for our full sample from July 1997 to December 2012. The independent variables are all measured ex-ante using the lag structure given by Fama and French. The key variables include industry momentum variables based on different mappings based on the SIC-code industry classification. In Panel A, we consider methods that value-weight the return of industry peers in the given group (excluding the focal firm itself). In Panel B, we consider methods that equal-weight the return of industry peers in the given group (excluding the focal firm itself). In each panel, we consider four specifications in increasing granularity, all of which are based on SIC codes: (1) the 20-industry group based on Moskowitz and Grinblatt (1997), (2) the Fama-French-48 industries, (3) 2-digit SIC codes, and (4) 3-digit SIC codes. In all specifications, we also include our baseline 10-K Based TNIC-3 momentum variables. Although not displayed to conserve spece, we also include controls for own-firm momentum (month $t-2$ to $t-11$)), own firm lagged return (month $t-1$), log book to market ratio and log market capitalization. In the SIC-based method column, we note the method used in the given test. All RHS variables are standardized prior to running the regression for ease of comparison. All standard errors are adjusted using Newey-West with two lags.

Row	SIC-based Method	t-1 to t-6 SIC-based Industry Past Ret.	t-7 to t-12 SIC-based Industry Past Ret.	t-1 to t-6 TNIC-3 Industry Past Ret.	t-7 to t-12 TNIC-3 Industry Past Ret.	# Obs. / R^2
		Panel A: Value-weighted SIC-based methods (full sample)				
(1)	MG-20	0.004	0.000	0.007	0.004	0.052
		(3.33)	(0.03)	(4.33)	(2.98)	805,089
(2)	FF-48	0.003	-0.001	0.007	0.005	0.051
		(3.25)	(-0.55)	(4.13)	(3.34)	805,089
(3)	SIC-2	0.003	-0.001	0.007	0.005	0.051
		(4.06)	(-0.84)	(3.99)	(3.25)	805,089
(4)	SIC-3	0.003	-0.000	0.007	0.005	0.051
		(3.78)	(-0.53)	(4.15)	(2.98)	805,089
		Panel B: Value-weighted SIC-based methods (above-median sized firms)				
(5)	MG-20	0.003	-0.000	0.005	0.004	0.083
		(1.97)	(-0.09)	(2.88)	(2.73)	403,074
(6)	FF-48	0.002	-0.001	0.005	0.005	0.081
		(1.99)	(-0.51)	(2.66)	(3.23)	403,074
(7)	SIC-2	0.002	-0.001	0.005	0.004	0.081
		(2.15)	(-0.56)	(2.64)	(2.84)	403,074
(8)	SIC-3	0.001	-0.000	0.005	0.004	0.080
		(1.77)	(-0.53)	(2.73)	(2.71)	403,074
		Panel C: Equal-weighted SIC-based methods (full sample)				
(9)	MG-20	0.003	-0.001	0.007	0.005	0.055
		(1.38)	(-0.40)	(4.55)	(3.30)	805,089
(10)	FF-48	0.003	-0.000	0.007	0.005	0.054
		(1.92)	(-0.14)	(4.31)	(3.20)	805,089
(11)	SIC-2	0.004	-0.000	0.007	0.005	0.053
		(2.33)	(-0.11)	(4.31)	(3.08)	805,089
(12)	SIC-3	0.004	0.000	0.007	0.005	0.052
		(3.08)	(0.23)	(4.20)	(2.89)	805,089
		Panel D: Equal-weighted SIC-based methods (above-median sized firms)				
(13)	MG-20	0.001	0.000	0.005	0.005	0.086
		(0.28)	(0.04)	(3.10)	(3.06)	403,074
(14)	FF-48	-0.000	0.000	0.005	0.004	0.084
		(-0.02)	(0.25)	(2.91)	(3.00)	403,074
(15)	SIC-2	0.000	0.000	0.005	0.004	0.083
		(0.13)	(0.21)	(2.90)	(2.79)	403,074
(16)	SIC-3	0.001	0.000	0.005	0.004	0.081
		(0.90)	(0.04)	(2.68)	(2.75)	403,074

Online Appendix Table A2
Return Comovement: High and Low Industry Disparity

Fama-MacBeth regressions with own-firm monthly stock return as the dependent variable. One observation is one firm from July 1997 to December 2012. The independent variables include one to six month lags of TNIC-based peer returns (excluding the firm itself) and Fama-French-48 (SIC-based) peer returns (also excluding the firm itself). We also include controls for own firm momentum (own-firm 11 month lagged return from month $t-12$ to $t-2$) and the Fama and French (1992) variables including log book to market ratio and log size, a dummy for negative book to market ratio stocks (these control variables are not displayed to conserve space). Panels A and B display results for firms with below-median and above median TNIC/SIC disparity, respectively. Disparity is one minus the total sales of firms in the intersection of TNIC and SIC peers divided by the total sales of firms in the union of the TNIC and SIC peers. All peer variables are standardized to have a standard deviation of one for ease of comparison and interpretation. All standard errors are adjusted using Newey-West with two lags.

Row	TNIC-3 t Return	TNIC-3 $t-1$ Return	TNIC-3 $t-2$ Return	TNIC-3 $t-3$ Return	TNIC-3 $t-4$ Return	TNIC-3 $t-5$ Return	TNIC-3 $t-6$ Return	FF-48 t Return	FF-48 $t-1$ Return	FF-48 $t-2$ Return	FF-48 $t-3$ Return	FF-48 $t-4$ Return	FF-48 $t-5$ Return	FF-48 $t-6$ Return	RSQ / # Obs.
								Panel A: Below Median Industry Disparity							
(1)	0.041	0.005	0.002	0.000	0.001	0.000	0.001	0.020	-0.001	0.000	0.001	0.000	0.000	-0.002	0.117
	(42.98)	(6.37)	(2.37)	(0.61)	(0.80)	(0.17)	(1.86)	(22.47)	(-1.90)	(0.33)	(1.37)	(0.01)	(0.15)	(-1.93)	376,235
(2)		0.009	0.003	0.002	0.001	-0.000	0.002		-0.000	-0.001	0.002	-0.001	0.001	-0.001	0.090
		(6.32)	(2.56)	(1.67)	(0.81)	(-0.09)	(1.34)		(-0.09)	(-0.77)	(1.81)	(-0.66)	(0.89)	(-0.83)	376,235
								Panel B: Above Median Industry Disparity							
(3)	0.023	0.005	0.003	0.003	0.001	0.002	0.002	0.014	0.001	-0.000	0.001	0.001	-0.001	-0.000	0.063
	(23.81)	(7.91)	(4.08)	(4.58)	(2.13)	(2.70)	(2.29)	(14.29)	(1.75)	(-0.00)	(1.02)	(1.00)	(-1.61)	(-0.30)	369,617
(4)		0.006	0.004	0.003	0.002	0.002	0.002		0.002	-0.000	0.001	0.001	-0.001	0.000	0.050
		(7.62)	(4.08)	(4.36)	(2.31)	(2.02)	(2.41)		(1.88)	(-0.30)	(1.19)	(0.94)	(-1.13)	(0.14)	369,617

Online Appendix Table A3
Return Comovement (Systematic versus Idiosyncratic Components)

Fama-MacBeth regressions with own-firm monthly stock return as the dependent variable. One observation is one firm from July 1997 to December 2012. The independent variables include one to six month lags of the systematic and idiosyncratic portions of the TNIC industry return portfolio. To compute the systematic portion, we first regress (for each month) daily stock returns for each firm onto the three Fama French factors and the momentum factor. The projection from this regression (excluding the projection from the intercept) is the systematic portion of a firm's daily return. These are then aggregated to monthly observations, and we compute the average of these systematic returns over each firm's text based peers to get the "Systematic Peer Return". The idiosyncratic Peer Return is the raw TNIC peer return minus the systematic TNIC peer return. Panels A to C display results for the full sample, firms with below-median, and firms with above median TNIC/SIC disparity, respectively. Disparity is one minus the total sales of firms in the intersection of TNIC and SIC peers divided by the total sales of firms in the union of the TNIC and SIC peers. All peer variables are standardized to have a standard deviation of one for ease of comparison and interpretation. All standard errors are adjusted using Newey-West with two lags.

	SYSTEMATIC COMPONENTS						***IDIOSYNCRATIC COMPONENTS***						
Row	TNIC-3 $t-1$ Return	TNIC-3 $t-2$ Return	TNIC-3 $t-3$ Return	TNIC-3 $t-4$ Return	TNIC-3 $t-5$ Return	TNIC-3 $t-6$ Return	TNIC-3 $t-1$ Return	TNIC-3 $t-2$ Return	TNIC-3 $t-3$ Return	TNIC-3 $t-4$ Return	TNIC-3 $t-5$ Return	TNIC-3 $t-6$ Return	RSQ / # Obs.
						Panel A: Full Sample							
(1)	0.004	0.001	0.001	0.002	0.001	0.001	0.005	0.003	0.003	0.001	0.001	0.002	0.060
	(2.66)	(0.82)	(0.54)	(1.09)	(0.40)	(0.75)	(8.20)	(4.79)	(4.83)	(2.39)	(1.68)	(2.68)	750,924
						Panel B: Below Median Industry Disparity							
(2)	0.004	0.000	0.002	0.001	0.000	0.000	0.006	0.003	0.002	0.001	0.000	0.002	0.090
	(2.01)	(0.16)	(0.86)	(0.43)	(0.12)	(0.06)	(6.95)	(3.61)	(2.90)	(1.11)	(0.48)	(2.15)	378,507
						Panel C: Above Median Industry Disparity							
(3)	0.004	0.002	0.000	0.002	0.001	0.002	0.005	0.003	0.003	0.002	0.002	0.002	0.051
	(2.70)	(1.02)	(0.18)	(1.22)	(0.97)	(1.08)	(8.43)	(5.06)	(5.40)	(2.64)	(2.41)	(3.02)	372,417

Online Appendix Table A4
Actual vs Random TNIC Industry Momentum Returns

The first two columns report average ex-post returns and the average past 11 month returns for deciles sorted based on the past 11 month return. The past 11 month return is lagged one extra month and is thus measured from month $t = -12$ to $t = -2$. The ex-post return is from month $t = 0$. Reported averages are the industry average value of the past 11 month return and the ex-post return following the approach in Moskowitz and Grinblatt (1997). The last two columns report the results of random industry portfolios also following the approach in Moskowitz and Grinblatt (1997). In particular, we sort all firms by their past 11 month return, and replace each firm in each industry with the other firm that had the closest 11 month return. This creates random industries of similar size and granularity as TNIC, and each random industry had almost identical past returns as each actual industry (as displayed below in the first and third data columns). Unlike the past returns, the ex-post returns are not mechanistically linked in this way. The ex-post returns are thus the variable of interest. The table shows that the upward pattern of ex-post returns for actual TNIC peer returns is significantly steeper than the trend for random TNIC peer returns. A statistical test of differences indicates that these patterns are statistically different at the 1% level. Hence, actual horizontal industry links are critical in generating our observed results.

Decile	Actual TNIC-3 t-2 to t-12 Past Return	Actual TNIC-3 t=0 ex-post Return	Random TNIC-3 t-2 to t-12 Past Return	Random TNIC-3 t=0 ex-post Return
Lowest Past Return	-25.60	0.64	-25.52	1.08
2	-12.82	0.86	-12.75	0.99
3	-6.34	1.11	-6.29	0.92
4	0.00	1.00	0.03	0.90
5	5.99	0.83	5.99	0.91
6	12.13	0.99	12.09	0.97
7	19.85	1.39	19.80	1.09
8	29.55	1.75	29.41	1.18
9	41.49	1.84	41.19	1.19
Highest Past Return	71.53	1.89	70.54	1.20

Online Appendix Table A5
Fama MacBeth Return Regressions (Robustness to Large Firms)

To examine if our results can be explained by the well-known lead-lag effect between large and small firms as shown in Hou (2007), we consider several increasingly stringent specifications and examine if our key one-year TNIC-3 momentum variable remains significant. Panels A to C display results for the full sample, firms with above-median, and firms in the highest tercile size (market capitalization), respectively. These size-restricted tests are particularly stringent and illustrate that our results cannot be explained by the well-known lead-lag effect because returns attributable to the lead-lag effect do not exist in samples of larger firms. Because our results are robust to these specifications, and also because we find long-lasting predictable returns (12 months), we conclude that our results are related to the industry momentum anomaly and not to the lead-lag anomaly. All RHS variables are standardized prior to running the regression for ease of comparison. All standard errors are adjusted using Newey-West with two lags.

Row	Sample	t-1 to t-12 TNIC-3 Industry Past Return	t-1 to t-12 FF-48 Industry Past Return	t-1 Own Firm Past Return	t-2 to t-12 Own Firm Past Return	Log Market Capital-ization	Log Book to Market Ratio	R^2/ # Obs.
				Panel A: Full Sample				
(1)	All Months	0.008	0.003	-0.004	-0.000	-0.000	0.002	0.047
		(4.36)	(1.65)	(-4.26)	(-0.18)	(-0.35)	(2.30)	805,090
(2)	Pre-2008	0.009	0.003	-0.004	0.002	-0.001	0.003	0.050
		(4.15)	(1.56)	(-3.40)	(1.36)	(-0.73)	(2.17)	591,241
				Panel B: Above median size only				
(3)	All Months	0.005	0.002	-0.003	-0.003	-0.001	0.001	0.071
		(2.73)	(1.23)	(-2.63)	(-0.77)	(-0.94)	(1.19)	402,592
(4)	Pre-2008	0.007	0.002	-0.003	0.000	-0.000	0.002	0.074
		(3.05)	(1.19)	(-2.35)	(0.10)	(-0.28)	(1.45)	295,652
				Panel C: Largest size tercile only				
(5)	All Months	0.003	0.002	-0.002	-0.003	-0.001	0.001	0.085
		(1.82)	(1.59)	(-1.95)	(-0.83)	(-1.31)	(0.92)	268,365
(6)	Pre-2008	0.005	0.002	-0.002	0.001	-0.000	0.002	0.089
		(2.57)	(1.35)	(-1.63)	(0.21)	(-0.47)	(1.20)	197,083

Online Appendix Table A6
Calendar Time Portfolios (Value Weighted BJS Alpha Tests)

We report OLS coefficients and factor loadings based on calendar time zero investment portfolios investing long in positive momentum stocks and short in negative momentum stocks. All portfolios are value weighted. The portfolios are constructed from varying definitions of momentum: TNIC-3 momentum (Panels A and B), Fama-French-49 (SIC-based) momentum (Panel C), and own momentum (Panel D). All tests are based on the full sample except Panel A, which is based on portfolios of stocks in the highest quintile of industry disparity (one minus the total sales of all firms in the intersection of TNIC-3 and SIC-3 peer groups, divided by the total sales of firms in the union of the given firm's TNIC-3 and SIC-3 industries). We consider a one year measurement period for past returns as noted in the first column. For own firm momentum, we skip the most recent month following the existing literature. Zero investment calendar time portfolios are constructed by first sorting firms into quintiles based on the given momentum variables in each month. We then compute equal weighted average returns of firms in the highest quintile, and subtract the equal weighted returns of firms in the lowest quintile. Annualized Sharpe ratios are computed as the square root of twelve times the monthly mean divided by the monthly standard deviation. We report the Sharpe ratio of the raw return (top) and the residual return (bottom) for each specification.

Row	Sample / Horizon	Alpha	MKT	HML	SMB	UMD	Sharpe Ratios	R^2	Obs.
	Panel A: 10-K Based TNIC-3 Momentum (High Disparity Quintile), long/short quintiles								
(1)	All Months	0.017	-0.439	-0.263	0.610		0.814	0.168	186
	t-1 to t-12 Momentum	(3.53)	(-4.22)	(-1.79)	(4.29)		0.916		
(2)	All Months	0.012	-0.080	-0.013	0.446	0.787	0.814	0.530	186
	t-1 to t-12 Momentum	(3.36)	(-0.95)	(-0.12)	(4.13)	(11.80)	0.879		
(3)	Pre-2008	0.021	-0.511	-0.278	0.739		0.864	0.224	126
	t-1 to t-12 Momentum	(3.23)	(-3.14)	(-1.30)	(4.37)		1.038		
(4)	Pre-2008	0.012	-0.149	-0.100	0.507	0.842	0.864	0.551	126
	t-1 to t-12 Momentum	(2.36)	(-1.15)	(-0.61)	(3.85)	(9.38)	0.776		
	Panel B: 10-K Based TNIC-3 Momentum, long/short quintiles								
(5)	All Months	0.006	-0.356	0.053	0.394		0.311	0.094	186
	t-1 to t-12 Momentum	(1.24)	(-3.69)	(0.39)	(2.99)		0.322		
(6)	All Months	0.000	0.052	0.336	0.208	0.894	0.311	0.683	186
	t-1 to t-12 Momentum	(0.06)	(0.85)	(4.08)	(2.63)	(18.37)	0.015		
(7)	Pre-2008	0.008	-0.385	0.173	0.522		0.415	0.126	126
	t-1 to t-12 Momentum	(1.27)	(-2.45)	(0.84)	(3.20)		0.409		
(8)	Pre-2008	-0.003	0.064	0.395	0.233	1.047	0.415	0.735	126
	t-1 to t-12 Momentum	(-0.92)	(0.70)	(3.43)	(2.53)	(16.67)	-0.302		
	Panel C: FF-48 Momentum, long/short quintiles								
(9)	All Months	0.005	-0.264	-0.186	0.292		0.278	0.073	186
	t-1 to t-12 Momentum	(1.25)	(-2.86)	(-1.42)	(2.31)		0.323		
(10)	All Months	0.001	0.100	0.067	0.125	0.798	0.278	0.598	186
	t-1 to t-12 Momentum	(0.18)	(1.53)	(0.77)	(1.49)	(15.39)	0.047		
(11)	Pre-2008	0.007	-0.190	0.011	0.425		0.393	0.079	126
	t-1 to t-12 Momentum	(1.19)	(-1.30)	(0.06)	(2.80)		0.384		
(12)	Pre-2008	-0.003	0.204	0.206	0.171	0.918	0.393	0.653	126
	t-1 to t-12 Momentum	(-0.80)	(2.17)	(1.73)	(1.80)	(14.14)	-0.262		
	Panel D: Own-Momentum, long/short quintiles								
(13)	All Months	0.008	-0.684	-0.395	0.151		0.216	0.150	186
	t-1 to t-12 Momentum	(1.43)	(-5.46)	(-2.23)	(0.88)		0.370		
(14)	All Months	0.000	-0.062	0.038	-0.134	1.365	0.216	0.916	186
	t-1 to t-12 Momentum	(0.02)	(-1.45)	(0.66)	(-2.45)	(40.54)	0.006		
(15)	Pre-2008	0.010	-0.579	-0.178	0.250		0.328	0.086	126
	t-1 to t-12 Momentum	(1.37)	(-3.07)	(-0.72)	(1.27)		0.442		
(16)	Pre-2008	-0.005	0.035	0.125	-0.146	1.432	0.328	0.913	126
	t-1 to t-12 Momentum	(-2.11)	(0.58)	(1.61)	(-2.35)	(33.85)	-0.693		

Online Appendix Table A7
Fama MacBeth Return Regressions (separately consider positive and negative past returns)

This table divides all momentum variables into their positive and negative components. This is done by defining the positive component as being equal to the past return itself if it is positive, and zero otherwise. The negative component is similarly defined. Results are based on TNIC-3 and Fama-French-48 (SIC-based) momentum variables. This test is considered to show that our results are robust to an alternative specification that is motivated by Hou (2007), who shows that the negative component of past returns is more significant than the positive component when explaining the lead-lag anomaly. Our goal is to explain industry momentum, and indeed our results below show that the positive component of past returns is more important than the negative component (which is not consistent with the lead-lag anomaly).

		POSITIVE COMPONENTS				*NEGATIVE COMPONENTS*				
Row	Sample	t-1 to t-12 TNIC-3 Industry Past Ret	t-1 to t-12 FF-48 Industry Past Ret	t-1 Self Past Ret	t-2 to t-12 Self Past Ret	t-1 to t-12 TNIC-3 Industry Past Ret	t-1 to t-12 FF-48 Industry Past Ret	t-1 Self Past Ret	t-2 to t-12 Self Past Ret	R^2 # Obs
(1)	All Months	0.008				0.004				0.034
		(2.94)				(2.62)				805,090
(2)	All Months	0.007	0.002	-0.002	0.002	0.003	-0.191	-0.003	0.002	0.060
		(3.73)	(0.75)	(-4.11)	(1.47)	(3.32)	(-0.96)	(-2.80)	(1.45)	805,090
(3)	Pre-2008	0.011				0.003				0.040
		(4.27)				(1.87)				591,241
(4)	Pre-2008	0.010	0.001	-0.003	0.003	0.002	-0.232	-0.002	0.003	0.063
		(4.60)	(1.01)	(-3.21)	(2.29)	(2.08)	(-0.98)	(-2.10)	(1.76)	591,241