A Replication Study of Text-Based Industry Momentum

Group members: Junpeng Li, Ruiyang Liu

Abstract

This report presents a replication study of the text-based industry momentum as identified in the paper, "Text-Based Industry Momentum." The primary objective of this study is to examine the validity of the key findings provided in the original research paper. This replication study focuses on the recreation of Table 1, Table 11, Figure 2, and Figure 3 from the original paper. This report will discuss the data, methods, results, and the implications of the findings.

1. Introduction

For this quant final, our group chose the article of Text–Based Industry Momentum written by Gerard Hoberg and Gordon Phillips. The main idea of this article is to investigate the role of inattention in explaining industry momentum profits. The authors argue that investors may be inattentive to certain industries, leading to mispricing and momentum profits. They develop a text-based measure of industry visibility and show that firms with less visible peers generate higher momentum profits by comparing 4 different momentum strategies. Overall, the paper provides evidence that inattention is an important driver of industry momentum profits

Our group will use the data provided by Professor Alan Kwan. Then imply them to the original strategy and our idea to replicate Table 1: Summary Statistics, Table 11: Calendar–Time Portfolios, Figure 2: Equal Weighted Cumulative Portfolio Returns, Figure 3: Equal Weighted Cumulative Portfolio Returns. All the four original figures are in the reference. After that, we will compare our results with the original figures. In addition, we will discuss the pros and cons of these strategies in the real world.

2. Methods and Data Collection

2.1. Dataset Overview

There are totally 7 parquet files used in this replication. The Crsp.msf_delisting_adjusted.parquet (referred as crsp.msf.parquet) is a monthly database that contains information on all common stocks listed on the NYSE, AMEX and NASDAQ exchanges with uploaded delisting values. The Comp.finratios.parquet contains the book to market ratios and other data that are required in table 1. The five_factor_month.parquet, which shows the coefficient of Fama French's five factors from each month. The hoberg_cross_firm_similarity_permno.parquet (referred as Hoberg.parquet) shows the 10-K text-based similarity between two stocks in each year. The ff49.parquet file shows the Fama-French 49 industry portfolios with their corresponding SIC codes. It should be noticed that all the above datasets are almost the same as the original paper except we use the ff49 instead of ff48. While in the original

paper, the authors indicate that ff48 has a better return than ff49, for which FF–49 momentum loses its significance in the above–median market capitalization sample.

2.2 Filter conditions

In this paper, the condition that only mentions to use is to let stock price greater than one dollar. Therefore, we assume that the momentum filters in this paper are the same condition as the *Momentum Crashes* explained (written by Daniel / Moskowitz): abs(PRC) > 1, SHRCD IN c is between(10,11), with non-missing values for $MCAP_{t-1}$ which is for the later-on value weighted return. All observations are required to be valid to compute momentum variables.

2.3 Four different momentum strategies

In this replication, there are totally four types of strategies with two equal weighed and value weighed methods being used. The four strategies are the own firm momentum, the TNIC-3 industry momentum, the FF-49 momentum and the high disparity momentum strategy. The own firm strategy is the original momentum strategy for which we used the in *Momentum Crashes,* for which we calculate long the top 10% return momentum stocks and short the bottom 10% momentum stocks in each month.

In terms of TNIC-3 industry momentum, we calculate the momentum in a different way. Since there are two columns of stock permnos, we choose permno1 unchanged and merge permno2 with its return from msf.crsp files on the same date. Then average the permno2 return as the peer return for permno1. After that we get a table with permno1 and its new peer return in each month. Finally following the same procedure as what we have done in the own firm strategy and plot them into the diagram.

For the FF-49 momentum, we merge the FF-49 files and the crsp.msf.parquet files on the SIC integers, we can identify which industry each stock belongs to. Then link each stock to every other stock in the same industry and combine with msf.crsp files to get the peer return. Long the top 10% return momentum stocks and short the bottom 10% momentum stocks in each month in each industry.

In addition, we try the disparity momentum strategy. The word "disparity" is defined as the extent to which a given focal firm's less visible TNIC peers disagree with highly visible SIC peers. In particular, disparity is equal to one minus the ratio of total sales of peers in the intersection of TNIC–3 and SIC–3 industry peer groups, divided by the total combined sales of peers in the union of TNIC–3 and SIC–3 peer groups overall. The sales measure investors' attention. The lower the intersection of TNIC-3 and SIC-3 sales the higher the disparity which means higher disagreement.

There are 4 circumstances in total: 1) Both SIC-3 peers and TNIC-3 peers have high sales. 2) SIC-3 peers have high sales and TNIC-3 peers have low sales. 3) TNIC-3 peers have high sales and SIC-3 peers have low sales. 4) Both SIC-3 peers and TNIC-3 peers have low sales. The 1st and 4th circumstance indicates that TNIC-3 is aligned with SIC

and we cannot benefit from any information discreteness. The 3rd circumstance is less likely because according to the paper, SIC-3 is a more noticeable stock relationship observed by investors, which is proven to have a shorter momentum period than TNIC-3. The 2nd circumstance gives an intuition that the investors' overreaction in longing the economic/systematic/global shocks, we could potentially benefit from the investors' underreaction by longing the industrial/idiosyncratic/local shocks. Because there is an evident push on the more visible SIC-3 peers and a high overall push in the market (as the union of peer sales), the momentum gradually passes to the focal firm from its less visible TNIC-3 peers. Finally, follows the normal momentum procedure to long top 10% momentum and short bottom 10%.

3. Results

3.1. Replication of Table 1: Summary Statistics

In the first table, it is the summary statistics, which corresponds to table 1 in the original article from July 1997 to December 2012. This table consists of four panels.

	Mean	Std. Dev.	Minimum	Median	Maximum
Monthly Return	0.009902	0.177449	-0.981295	0.001603	13.495050
Log B/M Ratio	-0.662594	0.869036	-13.167480	-0.581258	12.449392
Log Market Cap	12.601365	2.036410	5.534969	12.473953	20.255740
Month t-1 Past Return	0.015444	0.187805	-0.877551	0.003125	13.495050
Month t-2 to t-12 Past Return	0.119426	0.733067	-0.994822	0.025160	49.980391

Panel A: Data from the Existing Literature

	Mean	Std. Dev.	Minimum	Median	Maximum
Month t-1 Past Return	0.009318	0.099987	-0.828767	0.009285	3.958333
Month t-1 to t-3 Past Return	0.030171	0.195525	-0.915644	0.024612	14.111111
Month t-1 to t-6 Past Return	0.064551	0.295888	-0.955208	0.044487	15.238023
Month t-1 to t-12 Past Return	0.137521	0.465917	-0.971670	0.080256	17.118022

Panel B shows the data from 10-K based TNIC–3 industries

	Mean	Std. Dev.	Minimum	Median	Maximum
Month t-1 Past Return	0.005319	0.069555	-0.467422	0.008962	0.677791
Month t-1 to t-3 Past Return	0.015973	0.123037	-0.710451	0.024132	1.168943
Month t-1 to t-6 Past Return	0.033873	0.176778	-0.795158	0.037159	1.736185
Month t-1 to t-12 Past Return	0.071962	0.257587	-0.781766	0.074980	2.657445

Panel C shows the data from FF–49 industries

	Month t Own-Firm Return	Log Book to Market Ratio	Log Mkt Capitalization	Month t − 1 Own-Firm Return	Month t − 1 FF−48 Return
B/M Ratio	0.000051	0.067025	0.007362	-0.000113	0.000415
Log Mkt Capitalization	0.062343	-0.315100	1.000000	0.020256	0.001821
Month t − 1 Own-Firm Return	0.008949	0.037040	0.020256	1.000000	0.003044
Month t − 1 FF−48 Return	0.001607	-0.001077	0.001821	0.003044	1.000000
Month t − 1 TNIC−3 Return	0.000944	0.000364	0.000078	0.001011	-0.003254

Panel D: Pearson correlations among these factors

3.2. Replication of Table 11: Table 11: Calendar–Time Portfolios

	Alpha	MKT	HML	SMB	UMD	Sharpe Ratios	R square	Obs.
All Months t-1 to t-12 Momentum long	0.004427	0.873981	0.218546	0.812841	-0.132004	0.496861	0.903567	174.0
All Months t-1 to t-12 Momentum short	-0.002583	-0.684151	-0.455376	-0.345562	0.095290	-0.462924	0.823949	174.0
Pre-2008 t-1 to t-12 Momentum long	0.000062	0.978685	-0.004547	0.812537	-0.127196	0.293338	0.966307	60.0
Pre-2008 t-1 to t-12 Momentum short	-0.000302	-0.764196	-0.285733	-0.316648	0.099253	-0.242233	0.864829	60.0

Panel A: 10-K Based TNIC–3 Momentum (High Disparity Quintile), long/short quintiles

	Alpha	MKT	HML	SMB	UMD	Sharpe Ratios	R square	Obs.
All Months t-1 to t-12 Momentum long	0.008736	1.147118	0.000124	1.139249	0.452692	0.712998	0.876087	174.0
All Months t-1 to t-12 Momentum short	-0.004979	-0.977795	-0.247200	-0.925255	0.763414	-0.268396	0.904808	174.0
Pre-2008 t-1 to t-12 Momentum long	0.001016	1.197548	-0.073984	0.778812	0.285701	0.256569	0.898254	60.0
Pre-2008 t-1 to t-12 Momentum short	0.002355	-0.948593	-0.173558	-1.194339	0.699814	-0.247933	0.930659	60.0

Panel B: 10-K Based TNIC–3 Momentum, long/short quintiles

	Alpha	MKT	HML	SMB	UMD	Sharpe Ratios	R square	Obs.
All Months t-1 to t-12 Momentum long	0.010416	1.083574	0.156234	1.016587	0.220644	0.789195	0.821199	174.0
All Months t-1 to t-12 Momentum short	-0.004915	-0.844989	-0.199980	-0.775486	0.536337	-0.333282	0.860747	174.0
Pre-2008 t-1 to t-12 Momentum long	0.002660	1.069743	-0.157921	0.840063	-0.018018	0.381799	0.843599	60.0
Pre-2008 t-1 to t-12 Momentum short	-0.000792	-0.703760	-0.619184	-0.805351	0.237620	-0.314024	0.943429	60.0

Panel C: FF-48 Momentum, long/short quintiles

|	Alpha	MKT	HML	SMB	UMD	Sharpe Ratios	R square	Obs.
All Months t-1 to t-12 Momentum long	0.005735	1.268709	-0.045040	0.962604	0.453655	0.593936	0.935687	174.0
All Months t-1 to t-12 Momentum short	-0.008885	-1.091646	-0.090277	-1.287967	0.918076	-0.344773	0.864912	174.0
Pre-2008 t-1 to t-12 Momentum long	-0.002386	1.286884	-0.006968	0.722608	0.358366	0.087643	0.964252	60.0
Pre-2008 t-1 to t-12 Momentum short	0.000632	-0.992281	-0.297577	-1.168989	0.856872	-0.294067	0.895676	60.0

Panel D: Own-Momentum, long/short quintiles

In both figure 2 and figure 3, we plot the equal weighted and value weighed cumulative returns of all four momentum strategies that has been mentioned, for which the x-axis is the date and the y-axis is the cumulative portfolio returns.

3.3. Replication of Figure 2: Equal Weighted Cumulative Portfolio Returns
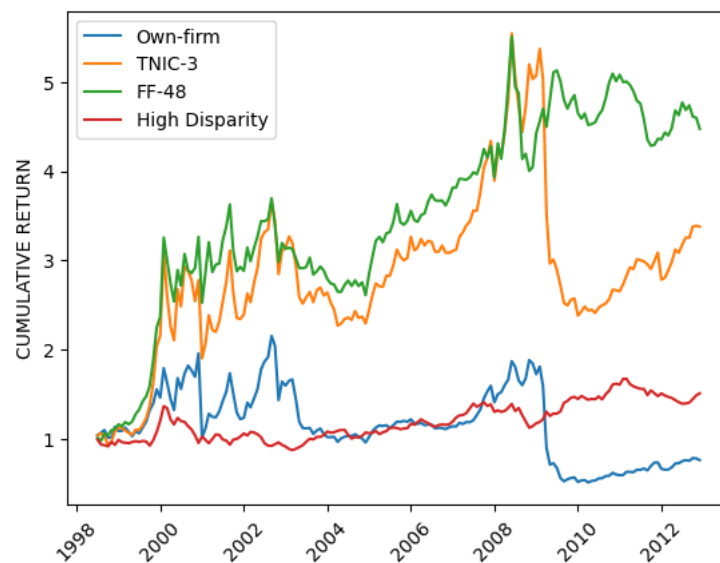
3.4. Replication of Figure 3: Value Weighted Cumulative Portfolio Returns
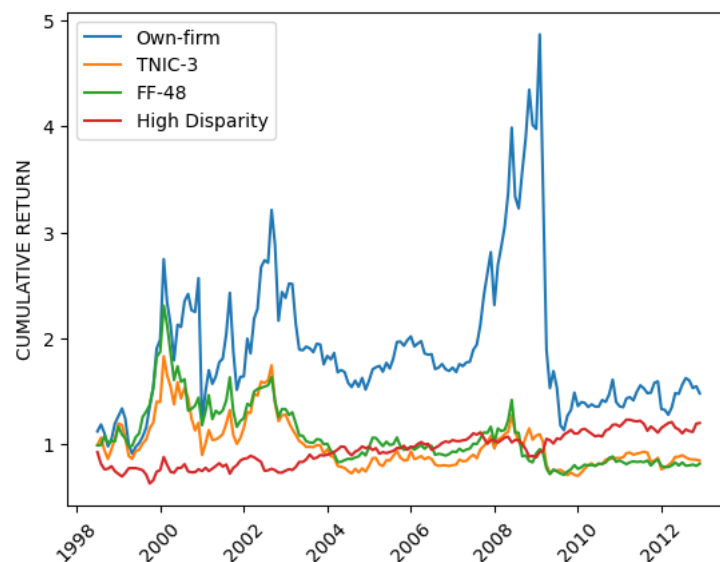
4. Discussion
4.1. Comparison of Findings with the Original Paper

For table 1, there are a few differences between the original paper and our replication. In panel A, both the log b/m ratios and maximum t-2 to t-12 past returns are not similar to the paper. These may be due to the unknown filters that authors use in their article. In panel B, both the minimum and maximum returns are different. This might be caused by the different raw data we use.

In panel C, there are differences among the mean, minimum, maximum and median data. This might be due to different raw data preparation of TNIC and the peer evaluation method. In panel D, many data are different since the majority of row variable data are quite different from the original. After the Pearson correlation, the

difference between our work and the original replication becomes larger. But overall our table 1 I quite similar as the original paper.

In table 11, the main focus is on UMD. We can observe that the UMD for disparity momentum is the smallest among four of the panels as the UMD for remaining three momentum depends on return.

For figure 2, the situation is quite similar as the original paper, where TNIC-3 has one of the highest rate of return. Apart from that, both figure 2 and 3 show that TNIC-3 and SIC have better ability to withstand market risks in equal weighted value strategy as cross momentum strategy include the systematic risk factor. Howevert, there are still many differences exist between our replication and the original paper. For instance, in the replicated figure 2 and 3, the disparity diagram has the worst return instead of the highest. This might due to the unclear explanation in the paper so that we can not replicate them. What's more, in the figure 3, the pattern is completely different from the one on the paper, where the own-firm strategy has the highest yield instead of the TNIC-3. But overall, our work verifies the message of this paper to some extent.

4.2. Implications of the Findings

In terms of realistic situation, for the large pension fund manager with $50 billion in assets who is trying to exceed the standard smart beta allocation to the FF5, any equal weighted value strategies will not suitable for their investment. Since if they spread their investment equally in each industry, the price of stocks will rise rapidly and make their competitors known.

Therefore, the value weighted strategies might more suitable for them. For the all four momentum strategies they used, based on our replication, the own-firm momentum has the highest field among four of them. However everyone knows the own-firm momentum stategy, this will not work in the future especially for large firms. It would be possible for large firms to invest in high dispartity strategy if this strategies can be replicated successfully.

By contrast, for small quant equity hedge fund who just needs to beat the market and wants a Sharpe ratio of at least 0.4, the equal weighted TNIC-3 momentum can achieve this target.

4.3. Limitations and Areas for Future Research

However, this is a back-testing. During the real trading, there is an uncertainty in this strategy. As the internet trolls who are being paid to post certain online comments with particular content might affect the quality of TNIC raw data updated every day. In this case, both the TNIC-3 momentum and the disparity momentum might not available to use. Firms might need to do another data filtering to ensure the quality of TNIC data.

5. Conclusion

Overall, after replicating the text-based industry momentum as identified in the paper, we can ensure that this strategy performs better than other momentum strategies in most of the cases. The validity of the key findings provided in the original research in the table 1, table 11, figure 2 and figure 3 are true. But it is questionable that whether this strategy can still work very well in nowadays due to so much rumors or trash information within the internet.
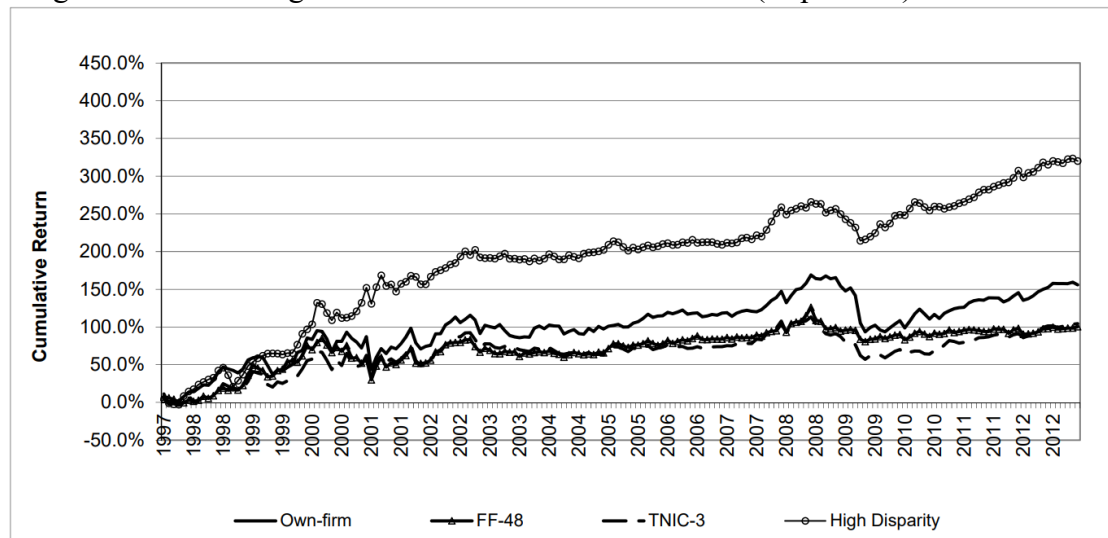
References

List of figures in the original paper

- Figure 2: Equal Weighted Cumulative Portfolio Returns (Replicated)
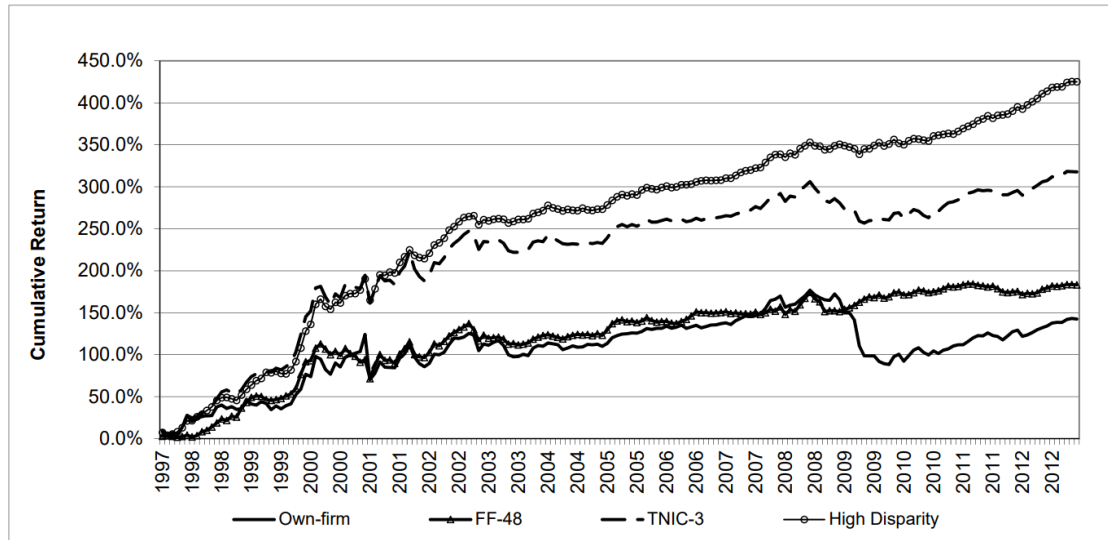
- Figure 3: Value Weighted Cumulative Portfolio Returns (Replicated)

List of tables in the original paper

- Table 1: Summary Statistics (Replicated)

Variable	Mean	Std. Dev.	Minimum	Median	Maximum
Panel A: Data from the Existing Literature					
Monthly Return	0.009	0.172	-0.981	0.002	9.374
Log B/M Ratio	-7.577	0.931	-16.164	-7.496	-1.223
Log Market Cap	12.664	2.009	6.233	12.575	20.121
Month t-1 Past Return	0.012	0.172	-0.878	0.003	13.495
Month t-2 to t-12 Past Return	0.158	0.811	-0.989	0.050	98.571
Panel B: Data from FF–48 industries					
Month t-1 Past Return	0.008	0.070	-0.437	0.011	0.622
Month t-1 to t-3 Past Return	0.027	0.126	-0.684	0.033	1.141
Month t-1 to t-6 Past Return	0.059	0.182	-0.770	0.059	1.806
Month t-1 to t-12 Past Return	0.158	0.315	-0.715	0.133	6.018
Panel C: Data from 10-K based TNIC–3 industries					
Month t-1 Past Return	0.012	0.094	-0.780	0.012	9.374
Month t-1 to t-3 Past Return	0.038	0.189	-0.952	0.034	10.202
Month t-1 to t-6 Past Return	0.075	0.291	-0.995	0.054	16.692
Month t-1 to t-12 Past Return	0.157	0.461	-0.997	0.097	26.500

Panel D: Pearson Correlations

Row	Variable	Month t Own-Firm Return	Log Book to Market Ratio	Log Mkt Capital-ization	Month $t-1$ Own-Firm Return	Month $t-1$ FF–48 Industry Return
(1)	B/M Ratio	0.024				
(2)	Log Mkt Cap	-0.012	-0.308			
(3)	Month $t-1$ Own-Firm Return	0.010	0.026	-0.022		
(4)	Month $t-1$ FF–48 Return	0.076	0.012	-0.012	0.325	
(5)	Month $t-1$ TNIC–3 Return	0.083	0.015	-0.017	0.402	0.622

- Table 11: Calendar–Time Portfolios (Equal Weighted BJS Alpha Tests) (Replicated)

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	0.023	-0.259	-0.280	0.530		1.432	0.195	186
	t-1 to t-12 Momentum	(6.18)	(-3.28)	(-2.50)	(4.91)		1.603		
(2)	All Months	0.019	0.039	-0.073	0.393	0.653	1.432	0.613	186
	t-1 to t-12 Momentum	(7.29)	(0.66)	(-0.92)	(5.20)	(13.99)	1.908		
(3)	Pre-2008	0.027	-0.333	-0.272	0.638		1.424	0.242	126
	t-1 to t-12 Momentum	(5.22)	(-2.54)	(-1.57)	(4.67)		1.680		
(4)	Pre-2008	0.018	0.025	-0.095	0.407	0.835	1.424	0.722	126
	t-1 to t-12 Momentum	(5.68)	(0.29)	(-0.90)	(4.81)	(14.46)	1.868		
	Panel B: 10-K Based TNIC–3 Momentum, long/short quintiles								
(5)	All Months	0.017	-0.386	-0.285	0.611		0.777	0.146	186
	t-1 to t-12 Momentum	(3.30)	(-3.50)	(-1.82)	(4.05)		0.857		
(6)	All Months	0.011	0.099	0.051	0.389	1.062	0.777	0.746	186
	t-1 to t-12 Momentum	(3.73)	(1.53)	(0.59)	(4.68)	(20.69)	0.975		
(7)	Pre-2008	0.024	-0.597	-0.380	0.678		0.898	0.192	126
	t-1 to t-12 Momentum	(3.39)	(-3.28)	(-1.58)	(3.57)		1.091		
(8)	Pre-2008	0.011	-0.045	-0.107	0.322	1.288	0.898	0.825	126
	t-1 to t-12 Momentum	(3.14)	(-0.50)	(-0.95)	(3.57)	(20.92)	1.033		
	Panel C: FF–48 Momentum, long/short quintiles								
(9)	All Months	0.010	-0.151	-0.247	0.410		0.649	0.132	186
	t-1 to t-12 Momentum	(2.68)	(-1.92)	(-2.22)	(3.82)		0.695		
(10)	All Months	0.006	0.126	-0.055	0.283	0.605	0.649	0.525	186
	t-1 to t-12 Momentum	(2.24)	(2.02)	(-0.65)	(3.54)	(12.23)	0.587		
(11)	Pre-2008	0.013	-0.271	-0.264	0.439		0.700	0.156	126
	t-1 to t-12 Momentum	(2.56)	(-2.11)	(-1.56)	(3.28)		0.824		
(12)	Pre-2008	0.004	0.092	-0.085	0.206	0.846	0.700	0.729	126
	t-1 to t-12 Momentum	(1.37)	(1.20)	(-0.88)	(2.65)	(16.00)	0.452		
	Panel D: Own-Momentum, long/short quintiles								
(13)	All Months	0.010	-0.470	-0.323	0.097		0.399	0.106	186
	t-1 to t-12 Momentum	(2.08)	(-4.38)	(-2.13)	(0.66)		0.540		
(14)	All Months	0.003	0.058	0.044	-0.145	1.157	0.399	0.895	186
	t-1 to t-12 Momentum	(1.96)	(1.46)	(0.82)	(-2.85)	(36.87)	0.512		
(15)	Pre-2008	0.016	-0.511	-0.253	0.157		0.653	0.082	126
	t-1 to t-12 Momentum	(2.52)	(-3.11)	(-1.17)	(0.92)		0.810		
(16)	Pre-2008	0.003	0.014	0.006	-0.181	1.224	0.653	0.886	126
	t-1 to t-12 Momentum	(1.42)	(0.24)	(0.08)	(-2.93)	(29.16)	0.466		