Appendix – Additional Results for
Empirical Investigation of an Equity Pairs Trading Strategy

Huafeng (Jason) Chen*

Shaojun (Jenny) Chen**

Zhuo Chen***

Feng Li****

A1 Additional Robustness Checks

We also consider whether conditional CAPM can explain the return spread of *RetDiff* sorted portfolios. We select four information variables, including the lagged T-bill rate, lagged dividend yield on the stock market, lagged yield spread between ten-year Treasury bond and three-month T-bill, and lagged yield spread between Moody's BAA and AAA bonds. We find that beta loadings conditional on those information variables cannot explain the return difference across *RetDiff* sorted portfolios (Table A.1).

The other possibility is that our results are caused by microstructure induced noises. To examine this possibility, we construct the *RetDiff* sorted portfolios using stocks with price greater than $5 and size greater than 10% NYSE cutoff. We find that the hedged value- and equal-weighted portfolios still earn positive and significant returns (Table A.2).

We examine the abnormal returns to the pairs trading strategy over time. We find returns from both the value-weighted and equal-weighted portfolios appear to be smaller over time. A regression of pairs trading returns on the calendar year confirms this observation (Table A.3).

* Associate Professor of Finance, PBC School of Finance, Tsinghua University, 43 Chengfu Road, Beijing, China 100083. ** Quantitative Analyst, Connor, Clark and Lunn Investment Management, 2200-1111 West Georgia Street, Vancouver, BC V6E 4M3, Canada. *** Assistant Professor of Finance, PBC School of Finance, Tsinghua University, 43 Chengfu Road, Beijing, China 100083. **** Professor of Accounting, Shanghai Advanced Institute of Finance, 211 West Huaihai Road, Shanghai, China 200030. Please email inquiries to chenhf@pbcsf.tsinghua.edu.cn, jchen@cclgroup.com, chenzh@pbcsf.tsinghua.edu.cn, or fli@saif.sjtu.edu.cn.

After the first draft is written, we also have extended the sample period to over the 2008-2015 period. This constitutes a true out-of-sample test. We find that the trading profits have declined in the recent sample period. However, the hedged portfolios are still statistically and economically significant for the equal-weighted portfolios (Appendix Table A.4).

A2 Evidence from the Determinants of the Pairwise Stock Return Correlations

Prior studies find that many economic variables are related to the return comovement. In this section, we ask the following question: are the abnormal returns to the pairs trading strategy driven by the pairs correlations that are related to the known determinants of return comovement?

A2.1 Determinants of Pairwise Stock Return Correlations

We start out by examining an array of variables that are likely to determine pairs correlations. Specifically, we consider variables that are potentially related to the comovement of firms' cash flows and discount rates. For example, accruals (Sloan 1996) can predict cross-sectional future stock returns. By definition, the variable is related to the comovement of stock returns. Firms with low (high) accruals tend to experience more positive (negative) stock returns in the future. This suggests that firms with similar past accruals tend to comove together in the future and form "pairs". We also consider variables that are related to investors' trading behavior that could cause stock comovement (Barberis, Shleifer, and Wurgler (2005)). To reduce computation burden, we require that firms have market equity of at least $500 million in this section. The variables we use include the following.

Earnings correlation. If the earnings of two stocks show a strong correlation, then their stock returns are more likely to be correlated. We compute the earnings correlations as follows using data from the Compustat monthly Price, Dividends, and Earnings file. For each quarter, we construct the return on equity, *ROE*, as ratio of the 12-month earnings per share to the book value of equity per share (*ERN* over *BKV*). To mitigate the influence of outliers, we set *ROEs* that are greater than 10 or less than -10 to be 10 and -10, respectively. We then compute the correlation between 20 quarterly *ROE* series of any two stocks for the period between year *t-4* and year *t*. We call this variable earnings correlation, *Earncorr*. To be included in our sample, firms must have all 20 quarterly *ROE* data between year *t-4* and year *t*. Alternatively, we construct correlations for changes in *ROEs*, and we label that variable as *Earncorr_ch*.

Earn_surprise_corr. To capture the correlations of two stocks' cash flow news, we calculate the correlation coefficient of their earnings surprises.[1] For each quarter, we measure the earnings surprise for a stock as the I/B/E/S actual quarterly earnings minus the most recent analyst forecast of the earnings divided by the book value of equity. *Earn_surprise_corr* is then computed as the Pearson correlation of the quarterly earnings surprises between a pair of stocks. We require firms to have 20 quarters of earnings surprises between year *t-4* and year *t*.

Revision_corr. We also compute the correlation between two stocks' analyst revisions of future earnings. If two stocks tend to experience similar analyst revisions around the same time (i.e., their earnings forecasts are revised upward or downward around the same time and are of similar magnitude), then they are more likely to be pairs stocks. Every month, we calculate the revision of one-quarter-ahead earnings forecasts by analysts as the current consensus earnings forecast minus last month's consensus forecast divided by the book value of equity. *Revision_corr* is the Pearson correlation of the forecast revision between a pair of stocks calculated using data from year *t-4* to year *t* for firms with at least 36 months of data in this period.

Growth. While earnings correlation captures the way in which historical earnings comove between two stocks, firm growth captures the difference in expected future earnings and therefore the difference in the growth expectation could capture the difference in the movement of stock prices. To measure the growth difference of two firms, we compute the sales growth using data in the last five years for each firm-year and calculate the absolute value of the difference in log growth rates for a pair of stocks.

[1] Earnings correlation and earnings change correlation can be viewed as parsimonious measures of earnings surprise correlations, when the expected earnings are assumed to be a constant, or the lagged earnings, respectively.

Industry. If firms are in the same industry, then they face similar business conditions, and therefore they are likely to experience the same cash-flow and discount-rate shocks. We construct dummy variables for a stock pair that equals 1 if they are from the same industry and 0 otherwise. Bhojraj, Lee, and Oler (2003) document that the Global Industry Classifications Standard (GICS) system is significantly better at explaining stock return comovements than the Standard Industrial Classifcation (SIC) or the North American Industry Classification System (NAICS). Therefore, we use GICS to construct this variable. For any two stocks that have the same eight-digit GICS industry code, we construct a dummy variable *Sgics8* that equals one, and zero otherwise. The prediction is that a pair of stocks with *Sgics8*=1 is more likely to be a pair than two stocks with *Sgics8*=0. We also set dummy variables *Sgics6*, *Sgics4*, *Sgics2* to one for stocks that have the same six-, four-, or two-digit GICS codes.

Size. If firms are similar in size, they may have similar exposures to risk factors (Fama and French (1993)). Therefore, they may have similar expected returns. Alternatively, investors may categorize assets into different styles and move funds among these styles, depending on the styles' relative performance, as suggested by Barberis and Shleifer (2003). Because size is a common style, investors' trading behavior can induce stocks of similar size to move together. Regardless of the interpretation of size as a risk factor or a mispricing factor, prior literature suggests that firms with similar size tend to comove in stock returns. We construct the absolute value difference in size for a pair of stocks as *Dsize*, which is the absolute difference in the logarithm of the market value of equity.

Book-to-market ratios. Similarly, if two firms have similar book-to-market ratios, then they may have similar exposure to some fundamental risks; alternatively, investors may view stocks with similar book-to-market ratios as having the same style (Fama and French (1993) and Boyer (2010)). We therefore construct the absolute value difference in log book-to-market ratios *Dlogbtm* and predict that a pair of firms is more likely to comove together in prices when *Dlogbtm* is smaller.

Accruals. Sloan (1996) shows that firms with low (high) accruals tend to have high (low) stock returns in the future. We construct accruals as earnings minus cash flows from operating activities divided by lagged book value of assets using Compustat annual data following Sloan (1996). We then calculate *Daccrual* as the absolute value of the difference in accruals between two firms and firms with smaller *Daccrual* should be more likely to be a pair.

Geographic location. If firms are located near one another geographically, then they may be subject to common shocks and have similar returns (Pirinsky and Wang (2006)). We use Compustat's state codes to construct a dummy variable *Sstate*, which equals one if two firms are located in the same state, and zero otherwise.

Firm age. Firms of similar ages are likely to be in similar stages of life cycle, and therefore have comovement in stock returns. We measure firm age as one plus the difference between the current year and the first year that the firm appears in the CRSP monthly data file. We then construct the absolute value difference between the logarithm of the ages of two firms i and j, as $Dage=|\log(Age_i)-\log(Age_j)|$.

Exchange listing. If two stocks are listed on the same exchange, then similar market microstructure issues may create stock return comovement between the two stocks. We set the variable *Slisting* to one if two stocks are listed in the same stock exchange, and zero otherwise.

S&P Index. Two stocks that belong to the same S&P major Index (e.g., Utilities, Transportation, and Financial Index) are more likely to move together (Barberis, Shleifer and Wurgler (2005)). This comovement could be driven by similar demand from index funds or due to investors' sentiment. We use the historical S&P index file and create the dummy variable *Sindex*, which equals one if two firms belong to the same S&P major index, S&P mid cap index or S&P small cap index in a given year, and zero otherwise.

Price difference. Green and Hwang (2009) show that firms with similar share prices comove with each other. Therefore, we construct a dummy variable *Dprice*, the absolute difference in the log price at the end of a given year, as a determinant of pairwise return correlations.

Duration. Firms with similar equity duration are likely to move in similar fashions because of correlations between expected future cash flows. We follow Dechow, Sloan, and Soliman (2004) and construct a variable that measures the difference of duration of the cash flows of two stocks, *Dduration*.

Financial leverage. Firms with similar financial leverage ratios (Long-Term Debt/Total Assets) may respond to changes in economic conditions in similar ways. We construct the absolute value difference in two firms' financial leverage ratios. We label this variable as *DLeverage*.

Upstream-downstream industry. Industries that have customer-supplier relations often benefit from one another's survival. Successes in the customers should lead to more order flows to the suppliers. If the market fully incorporates this information, then good news for the customer should be interpreted as good news for the supplier as well, thus creating comovement between the customer and the supplier. We follow Menzly and Ozbas (2010) and use the information from the Bureau of Economic Analysis' Input-Output Benchmark Survey to capture the upstream-downstream industry link between a pair of stocks. We construct the dummy variable *Indlink* as one if two firms' industries have customer-supplier relations, and zero otherwise.

Trading volume. Two stocks' returns could comove with each other because of similar liquidity situations. Therefore, we examine trading volume that is likely to be related to liquidity factors and investor trading behavior that may affect return comovement. For instance, Kumar and Lee (2006) find that firms with similar retail investor trading patterns tend to comove. We calculate the abnormal trading volume of a given stock by using the residual from a regression of monthly trading volume on annual trend and monthly dummies with data from the last 36 months. We calculate the correlation between two firms' abnormal trading volumes, *VolumeCorr*, and include it as a determinant of pairwise return correlation.

Common Analyst Coverage. Israelsen (2014) document that stocks with similar analysts tend to exhibit more excess comovement; Hameed, Morck, Shen, and Yeung (2015) find that stocks followed by few analysts comove significantly with firm-specific fluctuations in the prices of highly followed stocks in the same industry, but do not observe the converse. We therefore construct a variable *Rou_an* as the number of common analysts between two stocks divided by the square root of the product of the numbers of analysts that follow the two stocks. This variable captures the standardized level of overlapping analysts between two firms. We also construct a dummy variable *Drou_an* that equals one for any pair of two stocks that do not share any common financial analyst (i.e., when *Rou_an* is zero), and zero otherwise.

Table A.5 provides summary statistics for the pairwise return correlations. Our sample consists of almost 1.5 million pair-years of stocks over the sample period. On average, the Pearson correlation between any two stocks' monthly returns is 0.22 and the standard deviation of this correlation is 0.18. The most negative correlation between any two stocks is -0.562, and the most positive is 0.964.

The table also presents the summary statistics of the potential determinants of the pairwise correlations. The mean value of *Sgics2* is 0.13, indicating that 13% of our sample pairs are from the same two-digit GICS code industry. The mean of the *Sstate* variable is 0.054, which suggests that about 5% of the firm-pairs are from the same state. The mean value of *Slisting* (0.74) suggests that about 74% of our sample pairs are from the same exchange. About 49% of the stock pairs belong to the same S&P 500, S&P mid cap, or S&P small cap indexes.

We estimate the regression of return correlations on the variables we hypothesize as predicting pairwise correlations. In Table A.6, we present the results for the OLS pooled regression of stock return correlations on the determinants using data from 1987 to 2005. We start in 1987 because many of the determinants variables (e.g., common analyst coverage) are available only after 1982 and we need to five years of data to estimate them. To account for autocorrelation and the correlation between pairs that share a common stock, we adjust the standard errors by three-way clustering by the permno of the first stock, the permno of the second stock, and year. The three-way clustering method is based on Cameron, Gelbach, and Miller (2010).[2]

Most of the variables explain the pairwise return correlations in the ways we expect, and they have statistically significant coefficients. In Table A.6, Column 1, all the four variables that capture the correlation in earnings load up statistically significantly. Earnings correlation has a *t*-statistic of 4.42 and the correlation of changes in earnings has a *t*-statistics of 2.04. Earnings surprise correlation and earnings revision correlations are more significant with *t*-statistics of 6.47 and 7.43. However, the coefficients on these variables are only 0.024, 0.015, 0.017, and 0.091, which seem small in economic magnitude: given that both the dependent variable and independent variable are correlation coefficients, the perfect model would produce coefficients much closer to one. The R^2 of the regression is only 2.48%. This indicates that pairwise earnings correlations and differences in growth explain less than 3% of the total variation in the pairwise return correlations.

In Table A.6, Column 2, we include the four same-industry dummies (*Sgics2* to *Sgics8*) in the regression. All the industry dummies show up as positive and statistically significant. For instance, the coefficient on *Sgics8* is 0.057 (t = 5.05), which suggests that if two firms are within the same eight-digit GICS industry, then compared with a pair of firms that are in the same six-digit GICS industry but not in the same eight-digit GICS industry, their return correlation is higher by 0.057. Compared with two firms that do not share any industry membership, a pair of firms that are in the same eight-digit GICS industry have a return correlation coefficient higher by 0.202 (0.057+0.024+0.041+0.080). The R^2 also increases significantly when we include the industry dummies (the R^2 in Column 2 is 7.23%.)

In Column 4, to further explain pairwise return correlations, we include the distances in

[2] As an alternative statistical method, we estimate Fama-MacBeth regressions with Newey-West procedure to compute the standard errors, the results remain qualitatively similar.

firm size and book-to-market ratio (*Dsize* and *Dlogbtm*), the distance in accruals (*Daccrual*), the same-state dummy (*Sstate*), the distance in firm age (*Dage*), the same exchange listing dummy (*Slisting*), the same S&P index membership (*Sindex*); the distance in log prices (*Dprice*), durations (*Dduration*), and financial leverage ratios (*Dleverage*); the industrial customer-supplier link (*Indlink*), abnormal volume correlations (*Volumecorr*), common analyst coverage (*Rou_an*), and a dummy variable for missing common analyst coverage (*Drou_an*). Not surprisingly, the further the distance in size, book-to-market equity, accrual, age, and financial leverage, the lower is the correlation in stock returns of any two firms. All else equal, two firms located in the same state, listed in the same exchange list, in the same stock index, with similar trading volume patterns, and with common analyst coverage tend to comove in stock returns. Three factors show up as statistically insignificant: *Dprice*, the absolute value difference in log prices, *Dduration*, the distance in durations, and *Indlink*, the industrial customer-supplier link.

Overall, even when we include all the determinants in Column 4, there is substantial variation in the return comovement that cannot be explained; when we include all the explanatory variables, the R^2 is still less than 12%. This R^2 seems low. But we stress that the linear model of correlation is likely mis-specified and the interpretation should be taken with caution.

A2.2 Portfolios Formed on Fitted and Residual Correlations

To explore the economic drivers of the pairs trading profits, we now examine the trading profits based on different criteria in constructing the pairs portfolio. Specifically, we explore constructing the pairs portfolio according to the total pairwise return correlation, the pairwise return correlation that can be explained by the common factors in Column 4 of Table A.6 (fitted correlation), or the correlation that cannot be explained (residual correlation). Decomposing the pairs into those that can be explained by common factors (such as size, book-to-market, and accruals) versus those that cannot sheds light on the sources of the abnormal returns.

The results are reported in Table A.7. This table reports the value-weighted returns for portfolios that we form on the return difference (*RetDiff*) on the sample of firms that have valid observations to estimate the regression in Column 4 of Table A.6. In Panel A, the pairs portfolio is the 50 stocks with highest correlations. In Panel B, the pairs portfolio is the 50 stocks with highest fitted correlations in Column 4 of Table A.6. In Panel C, the pairs portfolio is the 50 stocks with highest residual correlations in Column 4 of Table A.6.

In Panel A of Table A.7, the six-factor alpha of returns based on total return correlation is 0.48%, and is close to being statistically significant. The trading profits are lower than that reported in Table 1, which is mostly due to the earlier sample period.

In Panel B of Table A.7, we find that the six-factor alpha of returns based on fitted return correlation is 0.67% and is statistically significant (t=1.74). Interestingly, in Panel C, the six-factor alpha of returns based on the residual return correlation is only 0.26% and is statistically insignificant (t=0.82). These results indicate that the pairs trading profits are driven by the correlation structure that is captured by our correlation prediction model in Column 4 of Table A.6.

The results suggest that "economic" factors help identify peer stocks that are "fundamentally" similar. Any price deviation from the resulting peer (based on "economic" factors) is therefore more likely due to "non-fundamental" factors and thus more likely to be reverted in the near future. In other words, the pairs trading strategy enhances short-term reversal by better isolating past price movement that is due to "non-fundamental" factors. This is consistent with the notion that pairs trading strategy is due to non-risk-based stories.

A3 Variable Selections

We apply the latest variable selection technique proposed by Harvey and Liu (2016) to evaluate whether *RetDiff* explains the cross section of expected returns given the existence of a large group of return predictors. Using the bootstrap approach for Fama-MacBeth regression described in Section 2.3 of Harvey and Liu (2016), we conduct variable selection tests for our *RetDiff* along with other cross-sectional return predictors. Tests are conducted on individual stocks' monthly returns. Two test statistics, the absolute value of Fama-MacBeth coefficient's *t*-statistics and R^2 of Fama-MacBeth regression, are used to measure the cross-sectional goodness-of-fit. We bootstrap 10,000 times to get the empirical distribution of |*t*-stat| and R^2. We evaluate the realized test statistics of all 11 return predictors against these empirical distributions.

Table A.8 reports the results. For the test statistics |*t*-stat|, four characteristics including *RetDiff*, size, book-to-market ratio, and past twelve months' cumulative return have *p*-value less than 5% and thus are individually significant in explaining the cross-section of expected returns. *RetDiff* has the largest |*t*-stat| and should be selected as the most important cross-sectional return predictor. For the test statistics R^2, only *RetDiff* is selected as an individually significant cross-sectional return predictor with a *p*-value of 0.000. Over all, both test statistics suggest that *RetDiff* plays an important role in explaining cross-sectional returns.

A4 Comparison with Other Residual Type Predictors

In the paper we show that the return predictability of *RetDiff* survives after we control for two types of residual predictors studied in Da et al. (2014) and Collin-Dufresne and Daniel (2015). In this section, we provide more discussions on the relation between *RetDiff* and these two types of residuals.

A4.1 Comparison with the Short-Term Reversal Residual of Da et al. (2014)

To investigate the difference between pairs residual *RetDiff* and short-term reversal (STR) residual in Da et al. (2014), we first check how correlated these two are. The time series average of cross-sectional correlation between *RetDiff* and STR residual is -0.67. The relative large magnitude is mainly mechanic because both terms have lagged return: *RetDiff* is defined as pairs return minus lagged return while STR residual is defined as lagged return minus cash flow news and expected return. We further calculate the correlation between STR residual and the two components of *RetDiff*: the number is 0.73 for lagged return while only 0.20 for pairs return, suggesting that the pairs term provides other information than STR residual.

We next examine the return predictability of the variation in *RetDiff* orthogonal to the STR residual. Table A.9 reports the returns and alphas of the value-weighted portfolios sorted by the *RetDiff* orthogonalized w.r.t. STR residual. The hedged Decile 10-Decile 1 portfolio generates a monthly return of 0.62% with a *t*-statistic of 2.47. The monthly 3-factor, 5-factor, and 6-factor alphas are 0.50% (*t*-statistic = 1.80), 1.18% (*t*-statistic = 4.25), and 1.20% (*t*-statistic = 4.15). Economically and statistically significant returns suggest that *RetDiff* contains information for future returns in addition to the STR residual.

Lastly, given the empirical findings, a natural question is to understand the economics of these two return predictors. Note that if we compare the definition of these two residuals, the difference lies on the pairs return *Cret* in *RetDiff*, and the cash flow news *CF* and expected returns μ in the STR residual. Vuolteenaho (2002) shows that while expected return news is highly correlated across firms, cash flow news can be diversified away in a portfolio. Based on our empirical findings and Vuolteenaho (2002), we conjecture that *Cret* could be a better proxy for expected return news, which is also likely to be shared by stocks within the same industry, than μ used in Da et al. (2014). If that is the case, the alternative STR residual ($\widetilde{\text{residual}}$) defined as lagged return minus cash flow news and *Cret* should predict future cross-sectional return better than either STR residual or *RetDiff*. Moreover, as cash flow news may be diversified away in a portfolio, replacing *CF* with *Cret* in the STR residual ($\widehat{\text{residual}}$) cannot improve return predictability. Table A.10 reports the monthly returns and alphas for hedged portfolios sorted by various residual-type predictors. While monthly raw returns of portfolios with sorting variables *RetDiff*, the STR residual, and $\widehat{\text{residual}}$ are similar in magnitude of around 0.85%, the number is 1.04% for $\widetilde{\text{residual}}$. The performance improvement is 22% from 0.85% to 1.04% per month. The result supports our conjecture that *Cret* could be a better proxy for expected return news that attributes to the return predictability.

A4.2 Comparison with the CAPM Residual of Collin-Dufresne and Daniel (2015)

Collin-Dufresne and Daniel (2015) attribute the return predictability of CAPM residual to slow moving capital. We find that the proposed *RetDiff* is different from CAPM residual. Table A.11 reports the results for value-weighted portfolios sorted on CAPM residual along. Even the hedged portfolio delivers a significant monthly return (-0.68% with a *t*-statistic of -2.92) and a three-factor alpha (-0.49% with a *t*-statistics of -1.96), the five-factor and six-factor alphas are not statistically significant after the short-term reversal factor is included, implying that the predictability of market residual is not beyond the short-term reversal effect. However, the return spread of the two extreme decile portfolios sorted on *RetDiff* cannot be explained by the short-term reversal factor.

In addition, Table A.12 reports the results of single sorted portfolios on market residual and *RetDiff* using the largest 100 stocks, which have the best liquidity and are the sample examined in Collin-Dufresne (2015). The hedged Quintile 5 minus Quintile 1 portfolio sorted on market residual does not generate statistically significant return or alpha, but the return spread of *RetDiff* sorting is still economically and statistically significant. Overall, both sets of results suggest that pairs residual contains different predictive power than market residual.

References

Barberis, Nicholas, and Andrei Shleifer, 2003, Style investing, *Journal of Financial Economics* 68, 161-199.

Barberis, Nicholas, Andrei Shleifer, and Jeffery Wurgler, 2005, Comovement, *Journal of Financial Economics* 75, 283-317.

Bhojraj, Sanjeev, Charles M. Lee, and Derek K. Oler, 2003, What's my line? A comparison of industry classification schemes for capital market research, *Journal of Accounting Research* 41, 745-774.

Boyer, H. Brian, 2010, Style related comovement: fundamentals or labels?, *Journal of Finance*, forthcoming.

Cameron, A. Colin, Jonah Gelbach and Douglas Miller, 2010, Robust inference with multi-way clustering, *Journal of Business and Economic Statistics*, forthcoming.

Campbell, John Y., and Jianping Mei, 1993, Where do betas come from? Asset price dynamics and the sources of systematic risk, *Review of Financial Studies* 6(3), 567-592.

Collin-Dufresne Pierre and Kent Daniel, 2015, Liquidity and return reversals, Working Paper.

Da, Zhi, Qianqiu Liu, and Ernst Schaumburg, 2014, A closer look at the short-term return reversal, Management Science 60 (3), 658-674.

Dechow, Patricia, Richard Sloan, and Mark Soliman, 2004, Implied equity duration: A new measure of equity risk, *Review of Accounting Studies* 9, 197-228.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3-56.

Green, T. Clifton, and Byoung-Hyoun Hwang, 2009, Price-based return comovement, *Journal of Financial Economics* 93, 37-50.

Hameed, Allaudeen, Randall Morck, Jianfeng Shen, Bernard Yin Yeung, 2015, Information, analysts, and stock return comovement, *Review of Financial Studies* 28 (11), 3153-3187.

Harvey, Campbell, and Yan Liu, 2016, Lucky factors, Working Paper, Duke University.

Israelsen, Ryan, 2014, Does correlated analyst coverage explain excess comovement? *Journal of Financial and Quantitative Analysis*, forthcoming.

Kumar, Alok, and Charles M.C. Lee, 2006, Retail investor sentiment and return comovements, *Journal of Finance* 61, 2451-2486.

Menzly, Lior, and Oguzhan Ozbas, 2010, Market segmentation and cross-predictability of returns, *Journal of Finance* 65, 1555-1580.

Pirinsky, Christo A. and Qinghai Wang, 2006, Does corporate headquarters location matter for stock returns?, *Journal of Finance* 61, 1991-2015.

Sloan, Richard, 1996, Do stock prices fully reflect information in accruals and cash flows about future earnings, *The Accounting Review* 71, 289-315.

Vuolteenaho, Tuomo, 2002, What Drives Firm-Level Stocks Returns?, Journal of Finance 57 (1), 233-264.

Table A.1: Betas of Conditional CAPM

This table reports the time series average of conditional betas of the ten portfolios formed on return difference (*RetDiff*). The conditional $beta_{i,t}^{cond}$ is estimated in a three-factor unconditional model: $R_{i,t} = alpha_i + beta_i^Z Z_{t-1} + beta_i^{MKT} MKT_t + beta_i^{Z,MKT} Z_{t-1} MKT_t + \epsilon_{i,t}$, where $beta_{i,t}^{cond} = beta_i^{MKT} + beta_i^{Z,MKT} Z_{t-1}$ and Z_{t-1} is the information variable. We choose four information variables, including lagged one-month Treasury bill rate, lagged dividend yield on the stock market, lagged yield spread between ten-year Treasury bond and the three-month T-bill, and the lagged yield spread between Moody's BAA bond and AAA bond. Panels A and B report the results for value-weighted and equal-weighted portfolios, respectively.

Panel A: Value-weighted

IV	Decile 1	2	3	4	5	6	7	8	9	Decile 10
Lagged one-month T-bill	1.15	1.07	0.99	0.95	0.93	0.93	0.98	1.04	1.17	1.39
Lagged dividend yield	1.16	1.01	0.96	0.93	0.92	0.92	0.98	1.04	1.15	1.38
Lagged term spread	1.12	1.01	0.90	0.88	0.89	0.88	0.96	1.06	1.16	1.41
Lagged credit spread	1.14	0.98	0.95	0.91	0.90	0.93	0.99	1.05	1.14	1.35

Panel B: Equal-weighted

IV	Decile 1	2	3	4	5	6	7	8	9	Decile 10
Lagged one-month T-bill	1.26	1.20	1.15	1.11	1.09	1.09	1.15	1.21	1.34	1.59
Lagged dividend yield	1.16	1.01	0.96	0.93	0.92	0.92	0.98	1.04	1.15	1.38
Lagged term spread	1.06	0.95	0.88	0.82	0.81	0.82	0.87	0.96	1.08	1.29
Lagged credit spread	1.23	1.11	1.07	1.04	1.02	1.05	1.10	1.16	1.26	1.40

Table A.2: Pairs Trading Strategy for Stocks with High Price and Large Size

This table reports the pairs trading strategy return for stocks with high price/large size. The portfolios are formed on the return difference (*RetDiff*). *Cret* is the previous month's pairs portfolio return. For each month in year *t+1*, the pairs portfolio is the equal-weighted portfolio of the 50 stocks that have the highest return correlations with a given stock between year *t-4* and year *t*. *Lret* is the previous month's stock return. *RetDiff* is $beta^{C}*(Cret\text{-}Rf) – (Lret – Rf)$, where $beta^{C}$ is the regression coefficient of a firm's monthly return on its pairs portfolio return in the most recent five years. Stocks with formation month price lower than $5 are dropped or stocks with formation month market cap smaller than the 10% NYSE cutoff are dropped. Panels A and B report returns of value-weighted and equal-weighted portfolios, respectively. The results are reported in percentage with Newey-West six-lag adjusted *t*-statistics in parentheses.

Panel A: Value-weighted portfolio		
	P>=$5	Size>=10% NYSE cutoff
Decile 1	0.16	0.13
	(0.71)	(0.57)
2	0.32	0.40
	(1.59)	(1.99)
3	0.45	0.42
	(2.37)	(2.07)
4	0.56	0.59
	(2.99)	(3.24)
5	0.70	0.65
	(3.96)	(3.54)
6	0.81	0.79
	(4.46)	(4.54)
7	0.80	0.84
	(4.30)	(4.34)
8	1.04	1.01
	(5.60)	(5.46)
9	1.15	1.15
	(5.74)	(5.34)
Decile 10	1.35	1.52
	(5.31)	(5.56)
Decile 10-Decile 1	1.19	1.39
	(8.76)	(9.93)

Panel B: Equal-weighted portfolio		
	P>=$5	Size>=10% NYSE cutoff
Decile 1	-0.02	0.04
	(-0.09)	(0.18)
2	0.40	0.37
	(1.71)	(1.56)
3	0.53	0.55
	(2.33)	(2.31)
4	0.73	0.75
	(3.47)	(3.35)
5	0.82	0.91
	(4.01)	(4.09)
6	0.97	0.94
	(4.76)	(4.44)
7	1.09	1.14
	(5.06)	(5.09)
8	1.33	1.35
	(6.02)	(5.76)
9	1.51	1.60
	(6.33)	(6.11)
Decile 10	1.95	2.15
	(7.04)	(6.18)
Decile 10-Decile 1	1.97	2.11
	(14.23)	(10.19)

Table A.3: Regressions of the Hedged Portfolio Returns on Year

This table presents the regression results of the annual value-weighted (Column 1) and equal-weighted (Column 2) hedge portfolios based on the pairs trading strategy on the calendar year. The sample period is 1931 to 2007. T-statistics are in parentheses.

	1	2
Intercept	8.620	25.278
	(4.41)	(3.24)
Year	-0.004	-0.013
	(-4.31)	(-3.17)
Obs.	77	77
Adjusted R^2	0.19	0.11

Table A.4: Out-of-Sample Portfolios Formed on Return Difference (2008-2015)

This table reports the value- and equal-weighted returns for portfolios that we form on the return difference (*RetDiff*). *Cret* is the previous month's pairs portfolio return. For each month in year *t+1*, the pairs portfolio is the equal-weighted portfolio of the 50 stocks that have the highest return correlations with a given stock between year *t-4* and year *t*. *Lret* is the previous month's stock return. *RetDiff* is $beta^C$*(*Cret-Rf*) – (*Lret* – *Rf*), where $beta^C$ is the regression coefficient of a firm's monthly return on its pairs portfolio return in the most recent five years. The three factors are excess market return, SMB, and HML. The five factors are the three factors, plus the momentum factor, and the short-term reversal factor. The five factors are the three factors, plus the momentum factor, and the short-term reversal factor. The six factors are the five factors, plus Pastor-Stambaugh's liquidity factor. Panel A reports value-weighted portfolios formed using all stocks with 60 monthly returns in the last five years. Panel B reports equal-weighted portfolios formed using all stocks with 60 monthly returns in the last five years. Monthly returns and alphas are reported in percentage with Newey-West six-lag adjusted *t*-statistics in parentheses. The sample period is January 2008 to December 2015.

Panel A: Value-weighted portfolios				
	Raw Returns	3-Factor Alpha	5-Factor Alpha	6-Factor Alpha
Quintile 1	0.45	-0.31	-0.33	-0.34
	(0.56)	(-1.94)	(-2.17)	(-2.25)
2	0.74	0.11	0.09	0.08
	(1.29)	(0.94)	(0.89)	(0.92)
3	0.72	0.05	0.05	0.04
	(1.18)	(0.64)	(0.61)	(0.53)
4	1.20	0.40	0.41	0.41
	(0.89)	(3.21)	(3.46)	(3.49)
Quintile 5	0.86	-0.11	-0.08	-0.07
	(1.15)	(-0.39)	(-0.33)	(-0.31)
Quintile 5- 1	0.41	0.21	0.25	0.26
	(0.95)	(0.63)	(0.82)	(0.90)

Panel 2: Equal-weighted portfolios				
	Raw Returns	3-Factor Alpha	5-Factor Alpha	6-Factor Alpha
Quintile 1	0.53	-0.44	-0.47	-0.47
	(0.38)	(-2.32)	(-2.76)	(-2.83)
2	0.79	0.00	0.00	0.00
	(1.03)	(-0.06)	(-0.01)	(0.01)
3	1.07	0.25	0.26	0.26
	(1.43)	(2.08)	(2.54)	(2.57)
4	1.39	0.46	0.47	0.47
	(1.72)	(3.58)	(4.89)	(5.00)
Quintile 5	1.91	0.63	0.67	0.66
	(1.55)	(1.68)	(2.85)	(2.82)
Quintile 5-1	1.37	1.07	1.14	1.13
	(3.56)	(3.34)	(4.91)	(4.94)

Table A.5: Summary Statistics

This table reports the summary statistics for the determinants of stock return correlations between 1987 and 2005. Each observation represents a stock pair year. *Corr* is the stock return correlation between two stocks using monthly returns between year *t-4* and *t*. We calculate earnings correlation (*Earncorr*) from the Compustat Price, Dividends, and Earnings data set as follows. In each quarter, we construct the return on equity *ROE* as the ratio of the 12-month earnings per share to the book value of equity per share (ERN over BKV). To mitigate outliers, we set *ROEs* that are greater than 10 or less than -10 to be 10 and -10, respectively. We then compute the correlation between 20 quarterly *ROEs* of any two stocks A and B, between year *t-4* and year *t*. *Earncorr_ch* is the correlation between two firms' changes in *ROE*s. *Earn_surprise_corr* is the Pearson correlation of the quarterly earnings surprises, measured as the I/B/E/S actual quarterly earnings minus the most recent analyst forecast of the earnings divided by the book value of equity. *Revision_corr* is the pairwise correlation between the monthly revisions in analyst forecasts of next quarter's earnings, where revisions are calculated as the change in mean consensus earnings forecasts scaled by lagged book value of equity. *Dgrowth* is the absolute value difference in five-year log sales growth rates. For any two stocks A and B, we construct a dummy variable *Sgics8* to be one if they have the same eight-digit GICS industry code, and zero otherwise. We set dummy variables *Sgics6*, *Sgics4*, *Sgics2* to one for stocks that have the same six-digit GICS codes, four-digit GICS codes, two-digit GICS codes, respectively. Market equity is the product of Compustat total shares outstanding and the fiscal year-end price (25*199). Size is the logarithm of the market equity at the fiscal year end in year *t*. *Dsize* is the absolute value difference in size. The book value of equity is the total assets minus total liabilities. Book-to-market equity is then the ratio of the book equity to the market value of equity at the fiscal year end in year *t*. *Dlogbtm* is the absolute value difference in log book-to-market ratios. *Daccrual* is the absolute value of the difference in accruals between a pair of stocks, where accruals are calculated as operating income after depreciation scaled by lagged book value of assets. *Sstate* is a dummy variable that is one if two firms are located in the same state, and zero otherwise. Firm age is the difference between the current year and the first year that a firm appears in CRSP, plus one. *Dage* is the absolute value difference between logarithm of ages of two firms *i*, and *j*, $|\log(Age_i)-\log(Age_j)|$. *Slisting* is one if two stocks are listed on the same exchange, and zero otherwise. *Sindex* is one if two stocks belong to the same S&P major, mid cap, or small cap indexes, and zero otherwise. *Dprice* is the absolute value difference in log prices per share. *Dduration* is the absolute value difference in cash flow durations. *Dleverage* is the absolute value difference in financial leverage ratios (long-term debt/total assets). *Indlink* is one if two firms belong to industries that have customer-supplier links, and zero otherwise. *Volumecorr* is the correlations between two firms' abnormal trading volumes. We include a stock in the sample only if it has 60 valid monthly returns, 20 quarterly earnings, and market equity of at least $500 million. *Rou_an* is the number of common financial analysts between two stocks divided by the square root of the product of the numbers of analysts that follow the two stocks. *Drou_an* is a dummy variable that equals one for any pair of two stocks that do not share any common financial analyst (i.e., when *Rou_an* is zero), and zero otherwise. There are 1,407,466 observations.

Variable	Mean	Std Dev	Minimum	Maximum
Corr	0.216	0.182	-0.562	0.964
Earncorr	0.078	0.453	-0.991	0.998
Earncorr_ch	0.062	0.390	-0.996	0.996
Earn_surprise_corr	0.015	0.248	-0.995	0.998
Revision_corr	0.018	0.153	-0.970	1.000
Dgrowth	0.409	0.373	0	5.607
Sgics8	0.015	0.123	0	1
Sgics6	0.029	0.167	0	1
Sgics4	0.059	0.236	0	1
Sgics2	0.126	0.332	0	1
Dsize	1.324	1.050	0	6.906
Dlogbtm	0.712	0.585	0	6.394
Daccrual	0.063	0.062	0	1.159
Sstate	0.054	0.226	0	1
Dage	0.681	0.505	0	2.315
Slisting	0.744	0.436	0	1
Sindex	0.494	0.500	0	1
Dprice	0.619	0.564	0	10.448
Dduration	1.555	1.541	0	28.810
Dleverage	0.155	0.118	0	0.887
Indlink	0.744	0.436	0	1
Volumecorr	0.094	0.200	-0.778	0.852
Rou_an	0.016	0.074	0	1
Drou_an	0.896	0.306	0	1

Table A.6: Determinants of Stock Return Correlations

This table reports OLS regression results of the pairwise stock return Pearson correlation coefficient (*Corr*) on the determinants between 1987 and 2005. Each observation represents a stock pair year. *Corr* is the stock return correlation between two stocks using monthly returns between year *t-4* and *t*. We calculate earnings correlation (*Earncorr*) from the Compustat Price, Dividends, and Earnings dataset as follows. In each quarter, we construct the return on equity *ROE* as the ratio of the 12-month earnings per share to the book value of equity per share (*ERN* over *BKV*). To mitigate outliers, we set *ROEs* that are greater than 10 or less than -10 to be 10 and -10, respectively. We then compute the correlation between 20 quarterly *ROEs* of any two stocks A and B, between year *t-4* and year *t*. *Earncorr_ch* is the correlation between two firms' changes in *ROE*s. *Earn_surprise_corr* is the Pearson correlation of the quarterly earnings surprises, measured as the I/B/E/S actual quarterly earnings minus the most recent analyst forecast of the earnings divided by the book value of equity. *Revision_corr* is the pairwise correlation between the monthly revisions in analyst forecasts of next quarter's earnings, where revisions are calculated as the change in mean consensus earnings forecasts scaled by lagged book value of equity. *Dgrowth* is the absolute value difference in five-year log sales growth rates. For any two stocks A and B, we construct a dummy variable *Sgics8* to be one if they have the same eight-digit GICS industry code, and zero otherwise. We set dummy variables *Sgics6*, *Sgics4*, *Sgics2* to one for stocks that have the same six-digit GICS codes, four-digit GICS codes, two-digit GICS codes, respectively. Market equity is the product of Compustat total shares outstanding and the fiscal year-end price (25*199). Size is the logarithm of the market equity at the fiscal year end in year *t*. *Dsize* is the absolute value difference in size. The book value of equity is the total assets minus total liabilities. Book-to-market equity is then the ratio of the book equity to the market value of equity at the fiscal year end in year *t*. *Dlogbtm* is the absolute value difference in log book-to-market ratios. *Daccrual* is the absolute value of the difference in accruals between a pair of stocks, where accruals are calculated as operating income after depreciation scaled by lagged book value of assets. *Sstate* is a dummy variable that is one if two firms are located in the same state, and zero otherwise. Firm age is the difference between the current year and the first year that a firm appears in CRSP, plus one. *Dage* is the absolute value difference between logarithm of ages of two firms *i*, and *j*, $|\log(Age_i)-\log(Age_j)|$. *Slisting* is one if two stocks are listed on the same exchange, and zero otherwise. *Sindex* is one if two stocks belong to the same S&P major, mid cap, or small cap indexes, and zero otherwise. *Dprice* is the absolute value difference in log prices per share. *Dduration* is the absolute value difference in cash flow durations. *Dleverage* is the absolute value difference in financial leverage ratios (long-term debt/total assets). *Indlink* is one if two firms belong to industries that have customer-supplier links, and zero otherwise. *Volumecorr* is the correlations between two firms' abnormal trading volumes. *Rou_an* is the number of common financial analysts between two stocks divided by the square root of the product of the numbers of analysts that follow the two stocks. *Drou_an* is a dummy variable that equals one for any pair of two stocks that do not share any common financial analyst (i.e., when *Rou_an* is zero), and zero otherwise. We report three-way clustered *t*-statistics by permno of stock *i*, permno of stock *j*, and year in parentheses below the coefficients.

	1	2	3	4
Earncorr	0.024	0.022	0.020	0.020
	(4.42)	(4.24)	(3.97)	(3.56)
Earncorr_ch	0.015	0.013	0.014	0.014
	(2.04)	(1.79)	(1.99)	(1.89)

Earn_surprise_corr	0.017	0.015	0.015	0.014
	(6.47)	(5.70)	(5.74)	(5.82)
Revision_corr	0.091	0.078	0.076	0.071
	(7.43)	(6.96)	(6.71)	(5.91)
Dgrowth	-0.044	-0.041	-0.039	-0.028
	(-4.14)	(-3.88)	(-3.71)	(-3.35)
Sgics8		0.057	0.055	0.007
		(5.05)	(4.92)	(0.76)
Sgics6		0.024	0.025	-0.011
		(2.76)	(2.78)	(-1.36)
Sgics4		0.041	0.038	0.016
		(4.48)	(4.23)	(1.88)
Sgics2		0.080	0.079	0.069
		(11.20)	(11.07)	(8.91)
Dsize			-0.010	-0.008
			(-3.84)	(-3.78)
Dlogbtm			-0.018	-0.019
			(-4.80)	(-4.26)
Daccrual				-0.051
				(-1.96)
Sstate				0.010
				(2.47)
Dage				-0.014
				(-3.49)
Slisting				0.035
				(3.81)
Sindex				0.016
				(2.67)
Dprice				-0.004
				(-1.0)
Dduration				0.003
				(1.19)
Dleverage				-0.034
				(-2.78)
Indlink				0.003
				(0.58)
Volumecorr				0.072
				(3.53)
Rou_an				0.180
				(8.33)
Drou_an				-0.046
				(-3.68)
Obs.	1,407,466	1,407,466	1,407,466	1,407,466
Adjusted R^2	0.025	0.072	0.080	0.117

Table A.7: Returns to Portfolios on Fitted and Residual Correlations

This table reports the value-weighted returns for portfolios that we form on the return difference (*RetDiff*) on the sample of firms that have valid observations to estimate the regression in Column 4 of Table A.6. The sample period is January 1988 to December 2006. *Cret* is the previous month's pairs portfolio return. For each month in year *t+1*, the pairs portfolio is the equal-weighted portfolio of the 50 stocks that have the highest return correlations (total, fitted, or residual correlations) with a given stock between year *t-4* and year *t*. In Panel A, the pairs portfolio is the 50 stocks with highest correlations. In Panel B, the pairs portfolio is the 50 stocks with highest fitted correlations in Column 4 of Table A.6. In Panel C, the pairs portfolio is the 50 stocks with highest residual correlations in Column 4 of Table A.6. In all three panels, *Lret* is the previous month's stock return. *RetDiff* is $beta^C*(Cret\text{-}Rf) - (Lret - Rf)$, where $beta^C$ is the regression coefficient of a firm's monthly return on its pairs portfolio return in the most recent five years. The three factors are excess market return, SMB, and HML. The five factors are the three factors, plus the momentum factor, and the short-term reversal factor. The six factors are the five factor, plus the Pastor-Stambaugh liquidity factor.

Panel A: Stocks with highest correlations as comovers

	Raw Return	3-Factor Alpha	5-Factor Alpha	6-Factor Alpha
Decile 1	0.77	-0.17	-0.12	0.26
2	0.62	-0.41	-0.44	-0.10
3	1.10	0.16	0.20	0.59
4	1.07	0.07	-0.02	0.35
5	0.95	-0.07	-0.06	0.34
6	1.17	0.25	0.26	0.66
7	1.81	0.85	0.74	1.12
8	1.44	0.48	0.42	0.80
9	1.36	0.33	0.40	0.80
Decile 10	1.19	0.10	0.37	0.74
Decile 10-1	0.42	0.26	0.50	0.48
t-statistics	(1.29)	(0.78)	(1.65)	(1.60)

Panel B: Stocks with highest fitted correlations as comovers

	Raw Return	3-Factor Alpha	5-Factor Alpha	6-Factor Alpha
Decile 1	0.66	-0.30	-0.22	0.15
2	0.71	-0.26	-0.29	0.06
3	0.92	-0.09	-0.08	0.30
4	1.14	0.21	0.15	0.53
5	0.87	-0.09	-0.05	0.31
6	1.26	0.27	0.21	0.63
7	1.59	0.65	0.70	1.08
8	1.42	0.43	0.27	0.68
9	1.24	0.28	0.32	0.71
Decile 10	1.37	0.23	0.44	0.81
Decile 10-1	0.71	0.53	0.66	0.67
t-statistics	(2.27)	(1.66)	(2.13)	(1.74)

Panel C: Stocks with highest residual correlations as comovers				
	Raw Return	3-Factor Alpha	5-Factor Alpha	6-Factor Alpha
Decile 1	0.96	0.02	0.10	0.48
2	0.83	-0.16	-0.18	0.16
3	0.88	-0.11	-0.07	0.31
4	1.06	0.11	0.00	0.41
5	1.06	0.11	0.12	0.51
6	1.35	0.38	0.42	0.83
7	1.58	0.66	0.69	1.09
8	1.28	0.22	0.13	0.49
9	1.23	0.19	0.20	0.59
Decile 10	1.24	0.20	0.35	0.73
Decile 10-1	0.28	0.17	0.25	0.26
t-statistics	(0.80)	(0.47)	(0.81)	(0.82)

Table A.8: Bootstrap Results of Fama-MacBeth Regression

This table reports the bootstrap results of Fama-MacBeth regression using individual stocks as test assets. The candidate return predictors include RetDiff, log size, log book-to-market ratio, past one-year momentum, Amihud illiquidity measure, idiosyncratic volatility w.r.t. Fama-French three factors, maximum daily return within a month, betas of the market, SMB, HML, WML factors. The procedure for bootstrap follows Harvey and Liu (2016). Two metrics, absolute value of beta coefficient's t-statistics and R^2, are used for variable selection. The sample period is July 1951 to December 2007.

		Single test			Single test	
	|t-stat|	5^{th}-percentile	p-value	R^2	5^{th}-percentile	p-value
RetDiff	**18.339**	**4.314**	**0.000**	**0.010**	**0.006**	**0.000**
Logsize	4.070	2.534	0.000	0.017	0.019	0.337
Logbm	6.514	2.689	0.000	0.011	0.011	0.206
$R^{-12,-2}$	5.997	5.914	0.043	0.018	0.020	0.295
Amihud	3.691	11.933	1.000	0.010	0.035	1.000
Idio vol	1.468	8.858	1.000	0.019	0.023	0.875
Max	3.615	8.247	0.998	0.015	0.018	0.757
$Beta^{MKT}$	0.076	2.157	0.945	0.012	0.013	0.489
$Beta^{SMB}$	1.503	2.239	0.192	0.019	0.021	0.482
$Beta^{HML}$	2.322	3.454	0.266	0.009	0.011	0.455
$Beta^{WML}$	1.800	2.385	0.149	0.005	0.006	0.469
		Multiple test			Multiple test	
	min	**11.932**	**0.000**		0.035	1.000

Table A.9: Portfolios Formed on Orthogonalized *RetDiff* over Return Residual

This table reports the value-weighted returns for portfolios formed on orthogonalized *RetDiff* over return residual. *Cret* is the previous month's pairs portfolio return. For each month in year t+1, the pairs portfolio is the equal-weighted portfolio of the 50 stocks that have the highest return correlations with a given stock between year t-4 and year t. *Lret* is the previous month's stock return. *RetDiff* is $beta^C$*(Cret-Rf) – (Lret – Rf)*, where $beta^C$ is the regression coefficient of a firm's monthly return on its pairs portfolio return in the most recent five years. The return residual is defined as the lagged return after taking off the components of expected return and cash flow news (Da et al. (2014)). The three factors are excess market return, SMB, and HML. The five factors are the three factors, the momentum factor, and the short-term reversal factor. The six factors are the five factors, plus Pastor-Stambaugh's liquidity factor. Monthly raw returns and alphas are reported in percentage with Newey-West six-lag adjusted *t*-statistics in parentheses. The sample period is January 1982 to December 2007.

	Raw Return	3-Factor Alpha	5-Factor Alpha	6-Factor Alpha
Decile 1	0.82	-0.35	-0.55	-0.57
	(2.77)	(-2.29)	(-3.54)	(-3.76)
2	0.66	-0.44	-0.60	-0.61
	(2.80)	(-3.60)	(-4.37)	(-4.56)
3	0.97	-0.20	-0.32	-0.31
	(3.77)	(-1.23)	(-2.31)	(-2.22)
4	1.08	-0.01	-0.02	-0.04
	(4.63)	(-0.09)	(-0.18)	(-0.33)
5	1.12	0.05	-0.02	0.01
	(5.18)	(0.53)	(-0.20)	(0.08)
6	1.35	0.23	0.20	0.21
	(5.33)	(2.35)	(1.76)	(1.90)
7	1.42	0.31	0.43	0.44
	(5.61)	(2.88)	(3.48)	(3.54)
8	1.37	0.21	0.36	0.37
	(5.28)	(1.92)	(3.30)	(3.20)
9	1.58	0.39	0.68	0.72
	(5.46)	(2.97)	(4.97)	(5.10)
Decile 10	1.45	0.15	0.63	0.63
	(4.05)	(0.73)	(3.13)	(2.97)
Decile 10-Decile 1	0.62	0.50	1.18	1.20
	(2.47)	(1.80)	(4.25)	(4.15)

Table A.10: Portfolios Formed on Alternative Residual Measures

This table presents the value-weighted returns and alphas for portfolios formed on various residual measures. *Cret* is the previous month's pairs portfolio return. For each month in year t+1, the pairs portfolio is the equal-weighted portfolio of the 50 stocks that have the highest return correlations with a given stock between year t-4 and year t. *Lret* is the previous month's stock return. *RetDiff* is $beta^{C}*(Cret\text{-}Rf) - (Lret - Rf)$, where $beta^{C}$ is the regression coefficient of a firm's monthly return on its pairs portfolio return in the most recent five years. The return residual is defined as the lagged return after taking off the components of expected return and cash flow news (Da et al. (2014)), which is $(Lret - Rf)$ -μ- *CF news*. The $\widehat{Residual}$ is defined as the lagged return after taking off the components of expected return and pairs portfolio return, which is $(Lret - Rf)$ -μ- $beta^{C}*(Cret\text{-}Rf)$. The $\widetilde{Residual}$ is defined as the lagged return after taking off the component cash flow news and pairs portfolio return, which is $(Lret - Rf)$ –*CF news*- $beta^{C}*(Cret\text{-}Rf)$. The three factors are excess market return, SMB, and HML. The five factors are the three factors, the momentum factor, and the short-term reversal factor. The six factors are the five factors, plus Pastor-Stambaugh's liquidity factor. Monthly raw returns and alphas are reported in percentage with Newey-West six-lag adjusted *t*-statistics in parentheses. The sample period is January 1982 to December 2007.

	RetDiff	Residual	$\widehat{Residual}$	$\widetilde{Residual}$
Raw Return	0.86	0.80	0.88	1.04
	(3.45)	(3.21)	(3.60)	(4.71)
3-Factor Alpha	0.64	0.59	0.64	0.96
	(2.27)	(1.91)	(2.51)	(3.51)
5-Factor Alpha	0.76	0.03	0.72	0.75
	(2.98)	(0.16)	(3.18)	(3.10)
6-Factor Alpha	0.79	0.06	0.76	0.78
	(3.06)	(0.33)	(3.31)	(3.14)

Table A.11: Portfolios Formed on CAPM Residual

This table reports the value-weighted returns for portfolios formed on CAPM residual. A stock's CAPM residuals are estimated using a rolling window of past 24 months' returns. The three factors are excess market return, SMB, and HML. The five factors are the three factors, the momentum factor, and the short-term reversal factor. The six factors are the five factors, plus Pastor-Stambaugh's liquidity factor. Monthly raw returns and alphas are reported in percentage with Newey-West six-lag adjusted *t*-statistics in parentheses. The sample period is July 1972 to December 2015.

	Raw Return	3-Factor Alpha	5-Factor Alpha	6-Factor Alpha
Decile 1	1.23	0.05	0.06	0.09
	(3.85)	(0.30)	(0.36)	(0.51)
2	1.37	0.34	0.27	0.27
	(5.32)	(2.99)	(2.49)	(2.46)
3	1.16	0.16	0.09	0.09
	(5.07)	(1.88)	(1.05)	(1.10)
4	1.23	0.30	0.23	0.23
	(5.67)	(4.08)	(2.85)	(2.73)
5	0.93	0.01	0.02	0.04
	(4.70)	(0.13)	(0.30)	(0.56)
6	1.01	0.12	0.10	0.10
	(5.07)	(1.72)	(1.28)	(1.17)
7	0.91	0.01	0.04	0.05
	(4.66)	(0.16)	(0.60)	(0.71)
8	0.87	-0.01	0.11	0.11
	(4.13)	(-0.14)	(1.40)	(1.29)
9	0.70	-0.23	-0.08	-0.08
	(3.30)	(-2.65)	(-1.01)	(-0.99)
Decile 10	0.55	-0.44	-0.16	-0.17
	(1.97)	(-2.95)	(-1.23)	(-1.25)
Decile 10-Decile 1	-0.68	-0.49	-0.22	-0.26
	(-2.92)	(-1.96)	(-1.15)	(-1.33)

Table A.12: Portfolios of the Largest 100 Stocks

This table reports the value-weighted returns for portfolios formed on CAPM residual and *RetDiff* using the largest 100 stocks. A stock's CAPM residuals are estimated using a rolling window of past 24 months' returns. *RetDiff* is defined as the lagged pair stock return minus lagged return. The largest 100 stocks are selected at the end of each year. The three factors are excess market return, SMB, and HML. The five factors are the three factors, the momentum factor, and the short-term reversal factor. The six factors are the five factors, plus Pastor-Stambaugh's liquidity factor. Monthly raw returns and alphas are reported in percentage with Newey-West six-lag adjusted *t*-statistics in parentheses. The sample period is July 1972 to December 2015.

Panel A: Portfolios sorted by CAPM residual				
	Raw Return	3-Factor Alpha	5-Factor Alpha	6-Factor Alpha
Quintile 1	1.02	0.09	0.01	0.05
	(4.01)	(0.63)	(0.08)	(0.38)
2	1.02	0.20	0.17	0.19
	(4.98)	(2.10)	(1.64)	(1.85)
3	0.98	0.17	0.14	0.14
	(5.10)	(1.78)	(1.37)	(1.34)
4	0.69	-0.10	0.00	0.00
	(3.42)	(-1.03)	(-0.02)	(0.01)
Quintile 5	0.70	-0.10	0.08	0.08
	(3.24)	(-0.93)	(0.87)	(0.89)
Quintile 6-Quintile 1	-0.31	-0.18	0.07	0.04
	(-1.51)	(-0.91)	(0.44)	(0.24)

Panel B: Portfolios sorted by *RetDiff*				
	Raw Return	3-Factor Alpha	5-Factor Alpha	6-Factor Alpha
Quintile 1	0.59	-0.20	-0.14	-0.13
	(2.81)	(-2.08)	(-1.42)	(-1.33)
2	0.73	-0.09	-0.02	0.00
	(3.73)	(-1.06)	(-0.25)	(-0.01)
3	0.92	0.09	0.05	0.07
	(4.67)	(0.94)	(0.54)	(0.67)
4	1.08	0.28	0.24	0.24
	(5.35)	(3.29)	(2.76)	(2.75)
Quintile 5	1.02	0.13	0.24	0.25
	(4.12)	(1.24)	(2.12)	(2.18)
Quintile 6-Quintile 1	0.43	0.33	0.38	0.38
	(2.59)	(2.05)	(2.38)	(2.33)