Contents

List of Figures

List of Tables

1 Introduction and Motivation

Capital Asset Pricing Model (CAPM) (Sharpe (1964), Lintner (1965), and Black (1972)) introduces a risk pricing paradigm. By incorporating factors, the model divided asset's risk into two parts: systematic risk and asset specified idiosyncratic risk. In general, the market factor capture the systematic risk, and different risk factors price the idiosyncratic risk. Researches (see Fama and French (1992), Carhart (1997), Kelly, Pruitt, and Su (2019)) has shown that, adding different risk factors into the CAPM model can enhance the ability of risk pricing, Because of this, identify risk factors has becomes an important topic in finance. Numerous of researchers are devoted into this field, Harvey and Liu (2019) had collected over 500 factors from papers published in the top financial and economic journals, and they found the growth of new factors speed up since 2008.

But we should notice that not all factors can pass the significant test comfortably every time like factors in three-factor model (Fama & French, 1992). Pesaran and Smith (2019) provide a criteria called factor strength to measure such discrepancy. In general, if a factor can generate loading significantly different from zero for all assets, then we call such factor strong factor. And the less significant loading a factor can generate, the weaker the strength it has.

In his 2011 president address Cochrane emphasis the importance of finding factor which can provides independent information about average return and risk. With regard of this, a number of scholars had applied various methods to find such factor. For instance, Harvey and Liu (2017) provided a bootstrap methods to adjust the threshold of factor loading's significant test, trying to exclude some falsely significant factor caused by multiple-test problem. Some other scholars use machine learning methods to reduce the potential candidates, more precisely, a stream of them have used a shrinkage and subset selection method called Lasso (Tibshirani, 1996) and it's variations to find suitable factors. One example is Rapach, Strauss, and Zhou (2013). They applied the Lasso regression, trying to find some characteristics from a large group to predict the global stock market's return.

But an additional challenge is that factors, especially in the high-dimension, are commonly correlated. Kozak, Nagel, and Santosh (2020) point out that when facing a group of correlated factors, Lasso will only pick several highly correlated factors seemly randomly, and then ignore the other and shrink them to zero. In other word, Lasso fails to handle the correlated factor appropriately.

Therefore, the main empirical question in this project is: how to select useful factors from a large group of highly correlated candidates. We address this question from two different prospects.

From one side, we employs the idea of factor strength discuss above, trying to use this criteria to select those strong factors. From the other hand, we will use another variable selection method called Elastic Net (Zou & Hastie, 2005) to select factors. With regard of the first solution, Bailey, Kapetanios, and Pesaran (2020) provides a consistent estimates method for the factor strength, and we will use such method to exam the strength of each candidate factors, and filter out those spurious factors. For the second solution, elastic net fixes the problem of Lasso can not handle correlated variables by adding extra penalty term, which makes it suitable for our purpose. And we will compare two method to use will we have a consistent selection of the risk factors. What's more, we can also use the factor strength as a standard to reduce the dimension of our candidates factor and then applied elastic net to conduct further selection.

In the rest of this thesis, we will first go through some literatures relates with CAPM model and methods about factor selection. Then in the section 3, we will provides a detailed description of the concept of factor strength and the estimation method. Also, we will introduce the elastic net. In the section 4, we set up a simple Monte Carlo simulation experiment to exam the estimation of factor strength. Section 5 includes the empirical application, we estimates the factors' strength and applied the elastic net method to select factors.

2 Related Literature

This project is builds on papers devoted on risk pricing. Formularisedd by Sharpe (1964), Lintner (1965), and Black (1972), the CAPM model only contains the market factor, which is denotes by the difference between market return and risk free return. Fama and French (1992) develop the model into three-factors, and then it even extend it into four (Carhart, 1997), and five (Fama & French, 2015). Recent research created a six-factors model and claim it outperform all other sparse factor model. (Kelly et al., 2019). Harvey and Liu (2019)

This thesis also connects with papers about involving factors has no or weak correlation with assets' return into CAPM model. Kan and Zhang (1999) found that the test-statistic of FM two-stage regression (Fama & MacBeth, 1973) will inflate when incorporating factors which are independent with the cross-section return. Therefore, when factors with no pricing power was involved in the model, those factors may have the chance to pass the significant test. Kleibergen and Zhan (2015) found out that even when some factor-return relationship does not exist, the r-square and the t-statistic of the FM two stage regression would become in favour of the conclusion of such structure presence. Gospodinov, Kan, and Robotti (2017) show how the involving of a spurious factor will distort the statistical inference of parameters. And, Anatolyev and Mikusheva (2018) studied the behaviours of the model with the presence of weak factors under asymptotic settings, find the regression will lead to an inconsistent risk premia estimation result.

This project also relates to some researches effort to identify useful factors from a group of potential factors. Harvey, Liu, and Zhu (2015) exam over 300 factors published on journals, presents that the traditional threshold for a significant test is too low for newly proposed factor, and they suggest to adjust the p-value threshold to around 3. Methods like a Bayesian procedure introduced by Barillas and Shanken (2018) were used to compare different factor models. Pukthuanthong, Roll, and Subrahmanyam (2019) defined several criteria for ”genuine risk factor”, and base on those criteria introduced a protocol to exam does a factor associated with the risk premium.

More details about the previous effort of identifying useful factors

This thesis will attempt to address the factor selection problem by using machine learning techniques. Gu, Kelly, and Xiu (2020) elaborate the advantages of using emerging machine learning algorithms in asset pricing. Those advantages including more accurate predict result, and superior efficiency. Various machine learning algorithms have been adopted on selecting factors for the factor model, especially in recent years. Lettau and Pelger (2020) applying Principle Components Analysis on investigating the latent factor of model. Lasso method, since it’s ability to select features, is popular in the field of the factor selection. Feng, Giglio, and Xiu (2019) used the double-selected Lasso method (Belloni, Chernozhukov, & Hansen, 2014),and a grouped lasso method (Huang, Horowitz, & Wei, 2010) is used by Freyberger, Neuhierl, and Weber (2020) on picking factors from a group of candidates. Kozak et al. (2020) used a Bayesian-based method, combing with both Ridge and Lasso regression, argues that the sparse factor model is ultimately futile.

3 Factor Strength

The concept of factor strength employed by this project comes from Bailey et al. (2020), and it was first introduced by Bailey, Kapetanios, and Pesaran (2016). They defined the strength of factor from prospect of the cross-section dependences of large panel and connect it to the pervasiveness of the factor, which is captured by the factor loadings. In a latter paper, Bailey et al. (2016) extended the method by loosen some restrictions, and proved that their estimation can also be applied on the residuals or regression result. Thereafter, they focusing on the case of observed factors, and proposed the method we employed in this project (Bailey et al., 2020).

3.1 Definition

Consider the following multi-factor model for n different cross-section units and T observations with k factors.

$$x_{it} = a_t + \sum_{j=1}^{k} \beta_{ij} f_{jt} + \varepsilon_{it} \tag{1}$$

In the left-hand side, we have x_{it} denotes the cross-section unit i at time t, where $i = 1, 2, 3, \cdots, n$ and $t = 1, 2, 3, \cdots T$. In the other hand, a_i is the constant term. f_{jt} of $j = 1, 2, 3 \cdots k$ is factors included in the model, and β_{ij} is the corresponding factor loading. ε_{it} is the stochastic error term.

The factor strength is relates to how many non-zero loadings correspond to a factor. More precisely, for a factor f_{jt} with n different factor loading β_j, we assume that:

$$|\beta_j| > 0 \quad i = 1, 2, \ldots, [n^{\alpha_j}]$$
$$|\beta_j| = 0 \quad i = [n^{\alpha_j}] + 1, [n^{\alpha_j}] + 2, \ldots, n$$

The α_j represents strength of factor f_{jt} and $\alpha_j \in [0, 1]$. If factor has strength α_j, we will assume that the first $[n^{\alpha_j}]$ loadings are all different from zero, and here $[\cdot]$ is defined as integral operator, which will only take the integral part of inside value.The rest $n - [n^{\alpha_j}]$ terms are all equal to zero. Assume for a factor which has strength $\alpha = 1$, the factor's loadings will be non-zero for all cross-section units. We will refer such factor as strong factor. And if we have factor strength $\alpha = 0$, it means that the factor has all factor loadings equal to zero, and we will describe such factor as weak factor (Bailey et al., 2016). For any factor with strength in [0.5, 1], we will refer such factor as semi-strong factor. In general term, the more non-zero loading a factor has, the stronger the factor's strength is.

3.2 Estimation Under single factor setting

To estimate the strength α_j, Bailey et al. (2020) provides following estimation.

To begin with, we consider a single-factor model with only factor named f_t. β_i is the factor loading of unit i. v_{it} is the stochastic error term.

$$x_{it} = a_i + \beta_i f_t + v_{it} \tag{2}$$

Assume we have n different units and T observations for each unit: $i = 1, 2, 3, \cdots, n$ and $t = 1, 2, 3, \cdots T$. Running the OLS regression for each $i = 1, 2, 3 \cdots, n$, we obtain:

$$x_{it} = \hat{a}_{iT} + \hat{\beta}_{iT} f_t + \hat{v}_{it}$$

For every factor loading $\hat{\beta}_{iT}$, we can examming their significance by constructing a t-test. The t-test statistic will be $t_{iT} = \frac{\hat{\beta}_{iT} - 0}{\hat{\sigma}_{iT}}$. Then the test statistic for the corresponding $\hat{\beta}_i$ will be:

$$t_{iT} = \frac{(\mathbf{f}'\mathbf{M}_\tau \mathbf{f})^{1/2} \hat{\beta}_{iT}}{\hat{\sigma}_{iT}} = \frac{(\mathbf{f}'\mathbf{M}_\tau \mathbf{f})^{-1/2} (\mathbf{f}'\mathbf{M}_\tau \mathbf{x}_i)}{\hat{\sigma}_{iT}} \tag{3}$$

Here, the $\mathbf{M}_\tau = \mathbf{I}_T - T^{-1}\tau\tau'$, and the τ is a $T \times 1$ vector with every elements equals to 1. $\mathbf{f}$ and $\mathbf{x_i}$ are two vectors with: $\mathbf{f} = (f_1, f_2 \cdots, f_T)'$ $\mathbf{x_i} = (x_{i1}, x_{i2}, \cdots, x_{iT})$. The denominator $\hat{\sigma}_{iT} = \frac{\sum_{i=1}^{T} \hat{v}_{it}^2}{T}$.

Using this test statistic, we can then define an indicator function as: $\ell_{i,n} := \mathbf{1}[|\beta_i| > 0]$. If the factor loading is none-zero, $\ell_{i,n} = 1$. In practice, we use the $\hat{\ell}_{i,nT} =:= \mathbf{1}[|t_{it}| > c_p(n)]$ Here, if the t-statistic t_{iT} is greater than critical value $c_p(n)$, $\hat{\ell}_{i,n} = 1$, otherwise $\hat{\ell}_{i,n} = 0$ In other word, we are counting how many $\hat{\beta}_{iT}$ are significant. With the indicator function, we then defined $\hat{\pi}_{nT}$ as the fraction of significant factor loading amount to the total factor loadings:

$$\hat{\pi}_{nT} = \frac{\sum_{i=1}^{n} \hat{\ell}_{i,nT}}{n} \tag{4}$$

In term of the critical value $c_p(n)$, rather than use the traditional critical value from student-t distribution $\Phi^{-1}(1-\frac{p}{2})$, we use:

$$c_p(n) = \Phi^{-1}(1 - \frac{p}{2n^{\delta}}) \tag{5}$$

Suggested by Bailey, Pesaran, and Smith (2019), here, $\Phi^{-1}(\cdot)$ is the inverse cumulative distribution function of a standard normal distribution, P is the size of the test, and δ is a non-negative value represent the critical value exponent. This adjusted critical value, adopt helps to tackle the problem of multiple-test.

After obtain the $\hat{\pi}_{nT}$, we can use the following formula provided by Bailey et al. (2020) to estimate our strength indicator α_j:

$$\hat{\alpha} = \begin{cases} 1 + \frac{\ln(\hat{\pi}_{nT})}{\ln n} & \text{if } \hat{\pi}_{nT} > 0, \\ 0, & \text{if } \hat{\pi}_{nT} = 0. \end{cases} \tag{6}$$

Whenever we have $\hat{\pi}_{nT}$, the estimated $\hat{\alpha}$ will be equal to zero. From the estimation, we can find out that $\hat{\alpha} \in [0,1]$

3.3 Estimation Under Multi-Factor Setting

This estimation can also be extended into a multi-factor set up. Consider the following multi-factor model:

$$x_{it} = a_i + \sum_{j=1}^{k} \beta_{ij} f_{jt} + v_{it} = a_i + \boldsymbol{\beta_i}' \mathbf{f_t} + v_{it}$$

In this set up, we have $i = 1,2,\cdots,n$ units, $t = 1,2,\cdots,T$ time observations, and specially, $j = 1,2,\cdots,k$ different factors. Here $\beta_i = (\beta_{i1}, \beta_{i2}, \cdots, \beta_{ij})'$ and $\mathbf{f}_t = (f_{1t}, f_{2t} \cdots, f_{jt})$. We employed the same strategy as above, after running OLS and obtain the:

$$x_{it} = \hat{a}_{iT} + \hat{\boldsymbol{\beta}}_{\mathbf{ij}} \mathbf{f}_{jt} + \hat{v}_{it}$$

To conduct the significant test, we calculates the t-statistic: $t_{ijT} = \frac{\hat{\beta}_{ijT} - 0}{\hat{\sigma}_{ijT}}$. Empirically, the test statistic can be calculated using:

$$t_{ijT} = \frac{\left(\mathbf{f}'_{j\circ}\mathbf{M}_{F_{-j}}\mathbf{f}_{j\circ}\right)^{-1/2}\left(\mathbf{f}'_{j\circ}\mathbf{M}_{F_{-j}}\mathbf{x}_i\right)}{\hat{\sigma}_{iT}}$$

Here, $\mathbf{f}_{j\circ} = \left(f_{j1}, f_{j2}, \ldots, f_{jT}\right)'$, $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{iT})'$, $\mathbf{M}_{F_{-j}} = \mathbf{I} - \mathbf{F}_{-j}\left(\mathbf{F}'_{-j}\mathbf{F}_{-j}\right)^{-1}\mathbf{F}'_{-j}$, and $\mathbf{F}_{-j} = \left(\mathbf{f}_{1\circ}, \ldots, \mathbf{f}_{j-1\circ}, \mathbf{f}_{j+1\circ}, \ldots, \mathbf{f}_{m\circ}\right)'$ For the denominator's $\hat{\sigma}_{iT}$, it was from $\hat{\sigma}^2_{iT} = T^{-1}\sum_{t=1}^{T}\hat{u}^2_{it}$, the $\hat{u}_{it}$ is the residuals of the model. Then, we can use the same critical value from (5). Obtaining the correspond ratio $\hat{\pi}_{nTj}$ from (4), and after that use the function:

$$\hat{\alpha}_j = \begin{cases} 1 + \frac{\ln \hat{\pi}_{nT,j}}{\ln n}, \text{ if } \hat{\pi}_{nT,j} > 0 \\ 0, \text{ if } \hat{\pi}_{nT,j} = 0 \end{cases}$$

to estimates the factor loading.

4 Monte Carlo Design

4.1 Design

In order to study the finite sample property of factor strength $\hat{\alpha}_j$, we designed a Monte Carlo simulation. Through the simulation, we compare the property of the factor strength in different settings. We set up the experiments to reflect the CAPM model and it's extension. Consider the following data generating process (DGP):

$$r_{it} - r_{ft} = q_1(r_{mt} - r_f) + q_2\left(\sum_{j=1}^{k}\beta_{ij}f_{jt}\right) + \varepsilon_{it}$$

In the simulation, we consider a dataset has $i = 1, 2, \ldots, n$ different cross-section units, with $t = 1, 2, \ldots, T$ different observations. r_{it} is the unit's return, and r_{ft} represent the risk free rate at time t, therefore, the left hand side term $r_{it} - r_{ft}$ is the excess return of the unit i. For simplicity, we define $x_{it} := r_{it} - r_{ft}$. f_{jt} represents different risk factors, and the corresponding β_{ij} are the factor loadings. We use $r_{mt} - r_{ft}$ to denotes the market factor, and here r_{mt} is the average market return. Also, we use the term $f_{mt} := r_{mt} - r_{ft}$ to denotes the market factor. We expect the market factor will has strength equals to one all the time, so we consider the market factor has strength $\alpha_m = 1$. ε_{it} is the stochastic error term. Therefore, the simulation model can be simplified as:

$$x_{it} = q_1(f_{mt}) + q_2(\sum_{j=1}^{k} \beta_{ij} f_{jt}) + \varepsilon_{it}$$

$q_1(\cdot)$ and $q_2(\cdot)$ are two different functions represent the unknown mechanism of market factor and other risk factors in pricing asset risk. In the classical CAPM model and it's multi-factor extensions, for example the three factor model introduced by Fama and French (1992), both q_1 and q_2 are linear.

For each factor, we assume they follow a multinomial distribution with mean zero and a $k \times k$ variance-covariance matrix Σ.

$$\mathbf{f_t} = \begin{pmatrix} f_{i,t} \\ f_{2,t} \\ \vdots \\ f_{k,t} \end{pmatrix} \sim MVN(\mathbf{0}, \Sigma) \quad \Sigma := \begin{pmatrix} \sigma_{f_1}^2, & \rho_{12}\sigma_{f1}\sigma_{f2} & \cdots & \rho_{1k}\sigma_{f1}\sigma_{fk} \\ \rho_{12}\sigma_{f2}\sigma_{f1}, & \sigma_{f2}^2 & \cdots & \rho_{2k}\sigma_{f2}\sigma_{fk} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1k}\sigma_{fk}\sigma_{f1}, & \rho_{k2}\sigma_{fk}\sigma_{f2} & \cdots & \sigma_{fk}^2 \end{pmatrix}$$

The diagonal of matrix Σ indicates the variance of each factor, and the rest represent the covariance among all k factors.

4.2 Experiment Setting

Follow the general model above, we assume both $q_1(\cdot)$ and $q_2(\cdot)$ are linear function:

$$q_1(f_{mt}) = a_i + \beta_{im} f_{mt}$$

$$q_2(\sum_{j=1}^{k} \beta_{ij} f_{jt}) = \sum_{j=1}^{k} \beta_{ij} f_{jt}$$

To start the simulation, we consider a two factor model:

$$x_{it} = a_i + \beta_{i1} f_{1t} + \beta_{i2} f_{2t} + \varepsilon_{it} \tag{7}$$

The constant term a_i is generate from a uniform distribution, $a_{it} \sim U[-0.5, 0.5]$. For the factor loading β_{i1} and β_{i2}, we first use a uniform distribution $IIDU(\mu_\beta - 0.2, \mu_\beta + 0.2)$ to produce the values. Here we set $\mu_{beta} = 0.71$ to make sure every generated loading value is sufficiently larger than 0. Then we randomly assign $n - [n^{\alpha_1}]$ and $n - [n^{\alpha_2}]$ factor loadings as zero. α_1 and α_2 are the true factor strength of f_1 and f_2. In this simulation, we will start the factor strength from 0.7 and increase it gradually till unity with pace 0.05, say $(\alpha_1, \alpha_2) = \{0.7, 0.75, 0.8, \cdots, 1\}$. $[\cdot]$ is the integer operator defined at section (3.2). This step reflects the fact that only $[n^{\alpha}]$ factor loadings are non-zero. In terms of the factors, they comes from a multinomial distribution $MVN(\mathbf{0}, \Sigma)$, as we discuss before.

Currently, we consider three different experiments set up:

Experiment 1 (single factor, normal error, no correlation) *Set β_{i2} from (7) as 0, the error term ε_{it} and the factor f_{1t} are both standard normal.*

Experiment 2 (two factors, normal error, no correlation) *Both β_{i1} and β_{i2} are non-zero. Error term and both factors are standard normal. The correlation ρ_{12} between f_{1t} andf_{2t} is zero. The factor strength for the first factor $\alpha_1 = 1$ all the time, and α_2 various.*

Experiment 3 (two factors, normal error, weak correlation) *Both β_{i1} and β_{i2} are non-zero. Error term and both factors are standard normal. The correlation ρ_{12} between f_{1t} andf_{2t} is 0.3. The factor strength for the first factor $\alpha_1 = 1$ all the time, and α_2 various.*

The factor strength in each experiment is estimated using the method discussed in section (3.2), the size of significant test is $p = 0.05$, and the critical value exponent σ has been set as 0.5. For each of the experiment, we calculate the bias, the RMSE and the size of the test to justify the estimation performances. The bias is calculated as the difference between the true factor strength α and the estimate factor strength $\hat{\alpha}$. The Root Square Mean Error (RMSE) comes from:

$$RMSE = [\frac{1}{R}\sum_{r=1}^{R}(bias_r)^2]^{1/2}$$

Where the R represent the total replicate times. The size of the test is under the hypothesis that $H_0 : \hat{\alpha}_j = \alpha_j,\ j = 1,2$ against the alternative hypothesis $H_1 : \hat{\alpha}_j \neq \alpha_j,\ j = 1,2$. Here we employed the following test statistic from Bailey et al. (2020).

$$z_{\hat{\alpha}_j:\alpha_j} = \frac{(\ln n)\left(\hat{\alpha}_j - \alpha_j\right) - p\left(n - n^{\hat{\alpha}_j}\right)n^{-\delta-\hat{\alpha}_j}}{\left[p\left(n - n^{\hat{\alpha}_j}\right)n^{-\delta-2\hat{\alpha}_j}\left(1 - \frac{p}{n^{\delta}}\right)\right]^{1/2}} \quad j = 1,2 \tag{8}$$

Define a indicator function $\mathbf{1}(|z_{\hat{\alpha}_j:\alpha_j}| > c|H_0)$. For each replication, if this test statistic is greater than the critical value of standard normal distribution: $c = 1.96$, the indicator function will return value 1, and 0 otherwise. Therefore, we calculate the size of the test base on:

$$size = \frac{\sum_{r=1}^{R} \mathbf{1}(|z_{\hat{\alpha}_j:\alpha_j}| > 1.96|H_0)}{R} \quad j = 1,2, \tag{9}$$

In purpose of Monte Carlo Simulation, we consider the different combinations of T and n with $T = \{120, 240, 360\}$, $n = \{100, 300, 500\}$. The market factor, if included in the experiment, will have strength $\alpha_m = 1$ all the time, and the strength of the other factor will be $\alpha_x = \{0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1\}$. For every setting, we will replicate 2000 times independently, all the constant and variables will be re-generated for each replication.

4.3 Monte Carlo Discoveries

We report the results in Table (4) , (5) and (6) in Appendix A.

Table (4) provides the results under the experiment 1. The estimation method we applied tends to over-estimate the strength slightly most of the time when the true strength is relatively weak under the single factor set up. With the strength increase, the bias will turn to negative, represents a under-estimated results. Such bias, however, vanish quickly while observation t, unit amount n, and α increase. When we increase the time spam by including more data from the time dimensions, the bias, as well as the RMSE decrease significantly. Also, when including more cross-section unit n into the simulation, the performance of the estimation improves, showing by the decrease bias and RMSE values. An impressive result is that, the gap between estimation and true strength will goes to zero when we have $\alpha = 1$, the strongest strength we can have. With the strength approaching unity, the both bias and RMSE will converge to zero. We also presents the size of the test in the table. The size of the test will not variate too much when the strength increases, so as the unit increases, But we can observe that when observations for each unit increase, in other word, when t increases, the size will shrink dramatically. The size will become smaller than the 0.05 threshold after we extend the t to 240, or empirically speaking, when we included 20 years monthly return data into estimation. Notice that, from the equation (8), when $\hat{\alpha} = \alpha = 1$, the nominator will becomes zero. Therefore, the size will collapse into zero in all settings, so we do not report the size for $\hat{\alpha} = \alpha = 1$

For the two factors scenarios, we obtain similar conclusions in both the no correlation setting and weak correlation setting. The result of no correlation settings is shown in the table (5), and the table (6) shows the result when the correlation between two factors is 0.3. The estimation results will be improved by increasing either the observations amount t, or the cross-section units amount n. We also have the same unbiased estimation when true factor strength is unity under all unit-time combination. In some cases, even when the factor strength is relatively weak,we can have unbiased estimation if the n and t are big enough. (see table (6)). However, we should also notice that when we have a imbalanced panel data, like the scenario when $t > n$ The results of size of the test in two factors setting are performing similar to the single factor result. The size will shrink with the observation amount t increasing, and when we have t grater than 240, the size will be smaller than 0.05 threshold in all situations.

5 Empirical Application

In this section, we introduced the data prepared for the empirical application, discuss the results of factor strength estimations. Then we apply the elastic net method introduced before,

5.1 Data

In the empirical application part, we use the monthly U.S. stock return as the assets. The companies are selected from Standard Poor (S&P) 500 index component companies.[1] We prepared three data sets for different time spams: 10 years (January 2008 to December 2017), 20 years (January 1998 to December 2017), and 30 years (January 1989 to December 2017). Because of the components companies of the index are constantly changing, bankrupt companies will be moved out, and new companies will be added in. Also, some companies does not have enough observation. Therefore, for each of the datasets, the companies amount (n) are different, the dimensions of the data set is showing in the table (1) below.

[1] The data was obtained from the Global Finance Data, Osiris, and Yahoo Finance.

Table 1: Data Set Dimensions

	Time Spam	Companies Amount (n)	Observations Amount (T)
10 Years	January 2008 - December 2017	419	120
20 Years	January 1998 - December 2017	342	240
30 Years	January 1988 - December 2017	242	360

For the risk free rate, we use the one-month U.S. treasury bill return.[2] For company i, we calculates the companies return at month t (r_{it}) use the following formula:

$$r_{it} = \frac{p_{it} - p_{it-1}}{p_{it-1}} \times 100$$

and calculate the excess return $x_{it} = r_{it} - r_{ft}$. Here the p_{it} and p_{it-1} are the company's close stock price at the first day of month t and t-1. The price is adjusted for the dividends and splits.[3]

With regard of the factors, we use 145 different risk factors from Feng, Giglio, and Xiu (2020). The factor set also includes one market factor, represented by the difference between the average market return and risk free return. The average market return is a weighted average return of all stocks in U.S. market, incorporated by CSRP. Each factors contains monthly value from the January 1988 to December 2017.

5.2 Factor Strength Analysis

5.2.1 Regression model for single security and two facotrs

For the first part of the empirical application, we estimates the factor strength using the method discussed in the section 3. Precisely, we set the regression model base on the discussion of section 3.3.

$$x_{it} = a_i + \beta_{im}(r_{mt} - r_{ft}) + \beta_{ij} f_{jt} + v_{it}$$

where x_{it} is the excess return of asset i at time t, which is pre-defined in the section 5.1. $r_{mt} - r_{ft}$ represents the market factor, calculated by the difference between average market return and risk free return at the same time t. f_{jt} is the value of j^{th} risk factor at time t. Here $j = 1, 2, 3, \cdots 145$.

[2] The data was fetched from the Kenneth R. French website: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/

[3] The data is adjusted base on the Central for Research in Security Price (CRSP) method.

β_{mt} and β_{ij} are the factor loadings for market factor and risk factor, respectively.

5.2.2 Factor Strength Finding

The complete results of factor strength estimation is present in the appendix B and C. We estimates the factors' strength using three different data set discussed in the section 5.1, and list those strength from strong to weak, alongside the market factor strength, in the table (8). For different data sets, we obtains inconsistent factor strength estimations. In general, the ten-year data set provides a significantly weaker result, compares with the other two data sets results. Except the market factor, no factors shows strength above 0.8 from the ten-years result. The strongest factor beside the market factor is the beta factor which has strength around 0.75. The strongest risk factor in the twenty-year data set is the ndp (net debt-to-price),which has strength 0.904. In the thirty-year scenario, the salecash (sales to cash) is the strongest with strength 0.857. When comparing the proportion (see table (2)), we can find that nearly 40% factors from the ten year dataset shows strength less than 0.5, which is almost three times higher than the twenty and thirty years proportion. If we use 0.8 as a threshold, we can see that there are over forty percent factors in the twenty-year results excess this threshold, and the percentage for the thirty year results is 31%.

Table 2: Proportion of Strength (Excluded Market Factor)

Strength Level	10 Year Data Proportion	20 Year Data Proportion	30 Year Data Proportion
[0.9, 1]	0%	2.07%	0%
[0.85, 0.9)	0%	24.1%	4.14%
[0.8, 0.85)	0%	16.6%	27.6%
[0.75, 0.8)	0%	8.28%	12.4%
[0.7, 0.75)	7.59%	11.7%	9.66%
[0.65, 07)	15.9%	5.52%	15.9%
[0.6, 0.65)	17.9%	8.28%	5.52%
[0.55, 0.6)	13.1%	8.97%	5.52%
[0.5, 0.55)	8.97%	2.76%	4.83%
[0, 0.5)	36.6%	11.7%	14.5%

Another important finding is that from the twenty year data set, we obtained three factors: ndq (Net debt-to-price, $\hat{\alpha} = 0.904$), salecash (sales to cash, $\hat{\alpha} = 0.902$), and quick (quick ratio, $\hat{\alpha} = 0.901$) has strength greater than 0.9. We would expected when applying the elastic net method with the twenty-year data set, those three factors with the market factors would be selected.

Table 3: Selected Risk Factor with Strength

Ten Year			Twenty Yera			Thirty Year		
Rank	Factor	Strength	Rank	Factor	Strength	Rank	Factor	Strength
1	beta	0.749	1	ndp	0.904	1	salecash	0.857
2	baspread	0.730	2	salecash	0.902	2	ndp	0.852
3	turn	0.728	3	quick	0.901	3	quick	0.851
4	zerotrade	0.725	4	dy	0.897	4	age	0.851
5	idiovol	0.723	5	lev	0.897	5	roavol	0.850
6	retvol	0.721	6	cash	0.897	6	ep	0.849
7	std_turn	0.719	7	zs	0.896	7	depr	0.848
8	HML_Devil	0.719	8	cp	0.894	8	cash	0.847
9	maret	0.715	9	roavol	0.894	9	rds	0.843
10	roavol	0.713	10	age	0.894	10	currat	0.840
20	UMD	0.678	28	HML	0.874	38	HML	0.811
24	HML	0.672	76	SMB	0.745	69	SMB	0.721
87	SMB	0.512	88	UMD	0.703	95	UMD	0.672

We also pay attention to some famous factors, precisely the Fama-French size factor (Small Minus Big SMB), Fama-French Value factor (High Minus Low: HML) (Fama & French, 1992) and the Momentum factor (UMD) (Carhart, 1997). It is surprise that none of these three factors enter the top ten list for each data sets. Except the HML factor from the twenty and thirty year data set has strength above 0.8, none of the other factors in any data set shows strength higher than 0.75. The value factor SMB from the ten year data set only has strength 0.512, ranked no.87.

In order to see how factor strength evolve through the time, we decompose the thirty year data set into three small subsamples. For each subsamples, it contains 242 companies (n = 242). And for each companies, we obtained 120 observations (t = 120). The results is present in the table (12) and figure (2).

In general, we can conclude that for most of the factors, their strength gradually increased from the first decade (January 1988 to December 1997) to the second decade (January 1998 to December 2007), and then decreased in the third decade (January 2008 to December 2017). This pattern can also be seen in the figure (2). The drop of factor strength in the third decades has been regards as the main reasons why the ten-years data results shows a significantly weaker results than the twenty and thirty years data set.

5.2.3 Conclusion and Explanation

From the factor strength prospect, we would expect that for different time period, we will have different candidate factors for the CAPM model. For the ten-year data set, we would expected that only the market factor be useful, and therefore the elastic net method applied latter may only select the market factor. If we use the twenty and thirty year data, we will have a longer list for potential factors, 62 factors from the twenty-year estimation and 45 from the thirty years has strength greater than 0.8. Hence, we would expect the elastic net to select a less parsimonious model.

In terms of this discrepancy, there are several potential explanation. First if we consider the structure of our data set, we will find that the longer the time span, the less companies are included. This is because the S&P index will adjust the component, remove companies with inadequate behaviours, and add in new companies to reflect the market situation. Hence, those 242 companies in the thirty year data set can be viewed as survivals after series of financial and economic crisis.[4] We would expect those companies will have above average performances, such as better profitability and administration, comparing with other companies.

Notes: (But for how will those merits influence the factor's risk pricing ability is unclear for now)

Another possible explanation is the Global Financial Crisis in 2008. The financial market has been disturbed by this crisis, so therefore some mechanism may no longer working properly during that period.

We also need to notice that for some factors, their strength will decrease with the time. For instance, the gma (gross profitability) factor and convind (convertible debt indicator) factor (see figure 2) has consecutive strength decrease from the 1987-1997 period to 2007-2017 period. And for most of the factors, their strength will decrease significantly from the 1997-2007 period to 2007-2017 period.

[4]For instance, the dot-com bubble in the early 20th century, the 911 attack, and the 2008 Global Financial Crisis

References

Anatolyev, S., & Mikusheva, A. (2018, 7). Factor models with many assets: strong factors, weak factors, and the two-pass procedure. *CESifo Working Paper Series*. Retrieved from http://arxiv.org/abs/1807.04094

Bailey, N., Kapetanios, G., & Pesaran, M. H. (2016, 9). Exponent of cross-sectional dependence: Estimation and inference. *Journal of Applied Econometrics*, *31*, 929-960. Retrieved from http://doi.wiley.com/10.1002/jae.2476 doi: 10.1002/jae.2476

Bailey, N., Kapetanios, G., & Pesaran, M. H. (2020). Measurement of factor strength: Theory and practice. *CESifo Working Paper*.

Bailey, N., Pesaran, M. H., & Smith, L. V. (2019, 2). A multiple testing approach to the regularisation of large sample correlation matrices. *Journal of Econometrics*, *208*, 507-534. doi: 10.1016/j.jeconom.2018.10.006

Barillas, F., & Shanken, J. (2018, 4). Comparing asset pricing models. *The Journal of Finance*, *73*, 715-754. Retrieved from http://doi.wiley.com/10.1111/jofi.12607 doi: 10.1111/jofi.12607

Belloni, A., Chernozhukov, V., & Hansen, C. (2014, 4). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, *81*, 608-650. doi: 10.1093/restud/rdt044

Black, F. (1972). Capital market equilibrium with restricted borrowing. *The Journal of Business*, *45*, 444-455. Retrieved from www.jstor.org/stable/2351499

Carhart, M. M. (1997, 3). On persistence in mutual fund performance. *The Journal of Finance*, *52*, 57-82. Retrieved from http://doi.wiley.com/10.1111/j.1540-6261.1997.tb03808.x doi: 10.1111/j.1540-6261.1997.tb03808.x

Cochrane, J. H. (2011, 8). Presidential address: Discount rates. *The Journal of Finance*, *66*, 1047-1108. Retrieved from http://doi.wiley.com/10.1111/j.1540-6261.2011.01671.x doi: 10.1111/j.1540-6261.2011.01671.x

Fama, E. F., & French, K. R. (1992, 6). The cross-section of expected stock returns. *The Journal of Finance*, *47*, 427-465. Retrieved from http://doi.wiley.com/10.1111/j.1540-6261.1992.tb04398.x doi: 10.1111/j.1540-6261.1992.tb04398.x

Fama, E. F., & French, K. R. (2015, 4). A five-factor asset pricing model. *Journal of Financial Economics*, *116*, 1-22. doi: 10.1016/j.jfineco.2014.10.010

Fama, E. F., & MacBeth, J. D. (1973, 5). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy*, *81*, 607-636. doi: 10.1086/260061

Feng, G., Giglio, S., & Xiu, D. (2019, 1). *Taming the factor zoo: A test of new factors.* Retrieved from http://www.nber.org/papers/w25481.pdf doi: 10.3386/w25481

Feng, G., Giglio, S., & Xiu, D. (2020, 6). Taming the factor zoo: A test of new factors. *The Journal of Finance*, *75*, 1327-1370. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/jofi.12883 doi: 10.1111/jofi.12883

Freyberger, J., Neuhierl, A., & Weber, M. (2020, 4). Dissecting characteristics non-parametrically. *The Review of Financial Studies*, *33*, 2326-2377. Retrieved from https://doi.org/10.1093/rfs/hhz123 doi: 10.1093/rfs/hhz123

Gospodinov, N., Kan, R., & Robotti, C. (2017, 9). Spurious inference in reduced-rank asset-pricing models. *Econometrica*, *85*, 1613-1628. doi: 10.3982/ecta13750

Gu, S., Kelly, B., & Xiu, D. (2020, 2). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*, 2223-2273. Retrieved from https://doi.org/10.1093/rfs/hhaa009 doi: 10.1093/rfs/hhaa009

Harvey, C. R., & Liu, Y. (2017, 12). False (and missed) discoveries in financial economics. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3073799

Harvey, C. R., & Liu, Y. (2019, 3). A census of the factor zoo. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3341728

Harvey, C. R., Liu, Y., & Zhu, H. (2015, 10). … and the cross-section of expected returns. *The Review of Financial Studies*, *29*, 5-68. Retrieved from https://doi.org/10.1093/rfs/hhv059 doi: 10.1093/rfs/hhv059

Huang, J., Horowitz, J. L., & Wei, F. (2010, 8). Variable selection in nonparametric additive models. *Annals of Statistics*, *38*, 2282-2313. doi: 10.1214/09-AOS781

Kan, R., & Zhang, C. (1999, 2). Two-pass tests of asset pricing models with useless factors. *The Journal of Finance*, *54*, 203-235. Retrieved from http://doi.wiley.com/10.1111/0022-1082.00102 doi: 10.1111/0022-1082.00102

Kelly, B. T., Pruitt, S., & Su, Y. (2019, 12). Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, *134*, 501-524. doi: 10.1016/j.jfineco.2019.05.001

Kleibergen, F., & Zhan, Z. (2015, 11). Unexplained factors and their effects on second pass r-squared's. *Journal of Econometrics*, *189*, 101-116. doi: 10.1016/j.jeconom.2014.11.006

Kozak, S., Nagel, S., & Santosh, S. (2020, 2). Shrinking the cross-section. *Journal of Financial Economics*, *135*, 271-292. doi: 10.1016/j.jfineco.2019.06.008

Lettau, M., & Pelger, M. (2020, 2). Estimating latent asset-pricing factors. *Journal of Econometrics*. doi: 10.1016/j.jeconom.2019.08.012

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*, *47*, 13-37. doi: 10.2307/1924119

Pesaran, M. H., & Smith, R. P. (2019). The role of factor strength and pricing errors for estimation and inference in asset pricing models. *CESifo Working Paper Series*.

Pukthuanthong, K., Roll, R., & Subrahmanyam, A. (2019, 8). A protocol for factor identification. *Review of Financial Studies*, *32*, 1573-1607. Retrieved from https://doi.org/10.1093/rfs/hhy093 doi: 10.1093/rfs/hhy093

Rapach, D. E., Strauss, J. K., & Zhou, G. (2013, 8). International stock return predictability: What is the role of the united states? *The Journal of Finance*, *68*, 1633-1662. Retrieved from http://doi.wiley.com/10.1111/jofi.12041 doi: 10.1111/jofi.12041

Sharpe, W. F. (1964, 9). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, *19*, 425-442. Retrieved from http://doi.wiley.com/10.1111/j.1540-6261.1964.tb02865.x doi: 10.1111/j.1540-6261.1964.tb02865.x

Tibshirani, R. (1996, 1). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*, 267-288. Retrieved from http://doi.wiley.com/10.1111/j.2517-6161.1996.tb02080.x doi: 10.1111/j.2517-6161.1996.tb02080.x

Zou, H., & Hastie, T. (2005, 4). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *67*, 301-320. Retrieved from `http://doi.wiley.com/10.1111/j.1467-9868.2005.00503.x` doi: 10.1111/j.1467-9868.2005.00503.x

A Simulation Result Table

Table 4: Simulation result for experiment 1

	Single Factor								
	Bias × 100			RMSE × 100			Size × 100		
	$\alpha_1 = 0.7$								
n\T	120	240	360	120	240	360	120	240	360
100	0.256	0.265	0.227	0.612	0.623	0.560	7.85	7.7	5.55
300	0.185	0.184	0.184	0.363	0.338	0.335	8.9	4.45	4.5
500	0.107	0.124	0.109	0.259	0.248	0.234	6.9	2.5	1.6
	$\alpha_1 = 0.75$								
100	-0.178	-0.159	-0.168	0.490	0.465	0.450	2.5	0.85	0.4
300	0.154	0.156	0.143	0.281	0.258	0.234	9.4	3.7	3.35
500	0.024	0.033	0.263	0.171	0.155	0.148	7.8	2	1.25
	$\alpha_1 = 0.8$								
100	-0.270	-0.265	-0.258	0.434	0.409	0.411	71.4	72.05	71.45
300	-0.052	-0.044	-0.043	0.183	0.149	0.150	10.15	2.45	2.9
500	0.045	0.068	0.067	0.136	0.126	0.121	16.6	6.4	5.9
	$\alpha_1 = 0.85$								
100	0.053	0.062	0.058	0.253	0.228	0.221	6.05	2.95	2.5
300	-0.012	0.009	-0.001	0.124	0.104	0.095	10.55	1.8	1.15
500	-0.026	-0.007	-0.011	0.096	0.073	0.069	13.25	0.9	0.7
	$\alpha_1 = 0.9$								
100	0.025	0.038	0.360	0.191	0.163	0.157	6.85	2	1.65
300	-0.034	-0.018	-0.020	0.099	0.069	0.068	13.2	0.8	0.9
500	-0.025	-0.001	-0.001	0.072	0.044	0.044	22.3	1.95	1.8
	$\alpha_1 = 0.95$								
100	-0.099	-0.088	-0.090	0.156	0.125	0.126	5.6	0.3	0.55
300	-0.046	-0.025	-0.026	0.083	0.045	0.045	22.5	2.2	2.25
500	-0.030	-0.006	-0.006	0.061	0.026	0.025	33.1	4.4	3.8
	$\alpha_1 = 1$								
100	0	0	0	0	0	0	-	-	-
300	0	0	0	0	0	0	-	-	-
500	0	0	0	0	0	0	-	-	-

Notes: This table shows the result of experiment 1. Factors and error are generate from standard normal distribution. Factor loadings come form uniform distribution $IIDU(\mu_\beta - 0.2, \mu_\beta + 0.2)$, and $\mu_\beta = 0.71$. We keep $[n^{\alpha_j}]$ amount of loadings and assign the rest as zero. For each different time-unit combinations, we replicate 2000 times. For the size of the test, we use a two-tail test, under the hypothesis of $H_0, \hat{\alpha}_j = \alpha_j$ $j = 1,2$. Cause under the scenarios of $\alpha = 1$, the size of the test will collapse, therefore the table does not report the sizes for $\alpha_1 = 1$.

Table 5: Simulation result for experiment 2

	Double Factor with correlation $\rho_{12} = 0$								
	Bias $\times$ 100			RMSE $\times$ 100			Size $\times$ 100		
	$\alpha_1 = 1, \alpha_2 = 0.7$								
n\T	120	240	360	120	240	360	120	240	360
100	0.567	0.737	0.628	4.062	3.819	3.799	2.95	1.45	1.85
300	0.512	0.611	0.518	2.398	2.103	1.979	6.25	0.55	0.5
500	-0.149	0.08	-0.019	1.796	1.498	1.443	8	0.2	0.1
	$\alpha_1 = 1, \alpha_2 = 0.75$								
100	-3.051	-3.02	-3.092	4.582	4.245	4.248	2.45	0.1	0.10
300	0.491	-1.035	0.640	1.843	1.460	1.576	7.6	0.8	0.55
500	-0.611	-0.372	-0.393	1.520	1.136	1.125	11.35	0.15	0.1
	$\alpha_1 = 1, \alpha_2 = 0.8$								
100	-3.752	-3.630	-3.581	4.557	4.213	4.210	84.65	85.9	85.25
300	-1.218	-0.331	-1.021	1.812	0.792	1.438	9.35	0.2	0.3
500	-0.022	0.192	0.147	1.047	0.782	0.742	15.35	1.1	1.1
	$\alpha_1 = 1, \alpha_2 = 0.85$								
100	-0.075	0.127	0.088	1.996	1.697	1.606	5.4	1.15	0.95
300	-0.531	-0.406	-0.351	1.097	0.613	0.777	10.8	0.15	0.2
500	-0.647	-0.391	-0.391	1.020	0.643	0.630	19.1	0.15	0
	$\alpha_1 = 1, \alpha_2 = 0.9$								
100	-0.128	0.043	0.025	1.428	1.143	1.118	4.9	0.65	0.7
300	-0.651	-0.334	-0.394	1.002	0.435	0.617	17.1	0.6	0.2
500	-0.434	-0.168	-0.171	0.7435	0.367	0.368	25.2	0.4	0.3
	$\alpha_1 = 1, \alpha_2 = 0.95$								
100	-1.218	-1.043	-1.036	1.603	1.222	1.212	6.65	0.25	0.05
300	-0.611	-0.344	-0.356	0.881	0.435	0.434	23.35	0.6	0.45
500	-0.415	-0.123	-0.134	0.661	0.220	0.216	36.75	1.35	1.1
	$\alpha_1 = 1, \alpha_2 = 1$								
100	0	0	0	0	0	0	-	-	-
300	0	0	0	0	0	0	-	-	-
500	0	0	0	0	0	0	-	-	-

Notes: This table shows the result of experiment 2. Factors and errors are generate from standard normal distribution. Between two factors, we assume they have no correlation. Factor loadings come form uniform distribution $IIDU(\mu_\beta - 0.2, \mu_\beta + 0.2)$, and μ_β is set to 0.71. We keep $[n^{\alpha_j}]$ amount of loadings and assign the rest as zero. For each different time-unit combinations, we replicate 2000 times. For the size of the test, we use a two-tail test, under the hypothesis of $H_0, \hat{\alpha}_j = \alpha_j$ $j = 1,2$. Cause under the scenarios of $\alpha = 1$, the size of the test will collapse, therefore the table does not report the sizes for $\alpha_1 = \alpha_2 = 1$

Table 6: Simulation result for experiment 3

	Double Factor with correlation $\rho_{12} = 0.3$								
	Bias $\times$ 100			RMSE $\times$ 100			Size $\times$ 100		
	$\alpha_1 = 1, \alpha_2 = 0.7$								
n\T	120	240	360	120	240	360	120	240	360
100	0.038	0.064	0.072	0.421	0.382	0.389	4.6	1.75	1.95
300	0.021	0.058	0.056	0.253	0.206	0.198	9.95	0.9	0.25
500	-0.032	0.006	0	0.201	0.153	0	12.20	0.1	0.05
	$\alpha_1 = 1, \alpha_2 = 0.75$								
100	-0.325	-0.313	-0.310	0.488	0.419	0.420	4.75	0.1	0
300	0.028	0.063	0.065	0.253	0.157	0.159	9.95	0.55	0.5
500	-0.082	-0.037	-0.039	0.175	0.114	0.112	19.25	0.25	0.3
	$\alpha_1 = 1, \alpha_2 = 0.8$								
100	-0.393	-0.361	-0.368	0.477	0.418	0.421	85.45	85.2	86.4
300	0.029	-0.099	-0.100	0.192	0.145	0.145	12.2	0.65	0.5
500	-0.037	-0.016	0.016	0.129	0.074	0.074	27.8	0.25	1.2
	$\alpha_1 = 1, \alpha_2 = 0.85$								
100	-0.027	0.008	0.007	0.234	0.160	0.155	9.3	0.9	0.65
300	-0.147	-0.031	-0.037	0.219	0.079	0.077	16.75	0.3	0.2
500	-0.088	-0.039	-0.039	0.136	0.063	0.062	30.6	0.15	0
	$\alpha_1 = 1, \alpha_2 = 0.9$								
100	-0.033	0.003	0.002	0.173	0.111	0.110	9.4	0.6	0.55
300	-0.087	-0.040	-0.041	0.131	0.061	0.061	27.8	0.1	0.05
500	-0.070	-0.017	-0.018	0.111	0.037	0.037	41.15	0.6	0.35
	$\alpha_1 = 1, \alpha_2 = 0.95$								
100	-0.134	-0.101	-0.104	0.185	0.122	0.122	10.15	0.1	0.15
300	-0.083	-0.034	-0.034	0.118	0.043	0.044	39.35	0.6	0.6
500	-0.062	-0.013	-0.012	0.937	0.022	0.023	51.8	1.25	2.0
	$\alpha_1 = 1, \alpha_2 = 1$								
100	0	0	0	0	0	0	-	-	-
300	0	0	0	0	0	0	-	-	-
500	0	0	0	0	0	0	-	-	-

Notes: This table shows the result of experiment 2. Factors and errors are generate from standard normal distribution. Between two factors, we assume they have correlation $\rho_{12} = 0.3$ Factor loadings come form uniform distribution $IIDU(\mu_\beta - 0.2, \mu_\beta + 0.2)$, and μ_β is set to 0.71. We keep $[n^{\alpha_j}]$ amount of loadings and assign the rest as zero. For each different time-unit combinations, we replicate 2000 times. For the size of the test, we use a two-tail test, under the hypothesis of $H_0, \hat{\alpha}_j = \alpha_j$ $j = 1, 2$. Cause under the scenarios of $\alpha = 1$, the size of the test will collapse, therefore the table does not report the sizes when $\alpha_1 = \alpha_2 = 1$

B Comparison Table

Table 7: Ranked Three Data Set Comparison

	Ten Year Data			Twenty Year Data			Thirty Year Data		
	Factor	Market Factor Strength	Risk Factor Strength	Factor	Market Factor Strength	Risk Factor Strength	Factor	Market Factor Strength	Risk Factor Strength
1	beta	0.976	0.749	ndp	0.960	0.904	salecash	0.905	0.857
2	baspread	0.980	0.730	salecash	0.958	0.902	ndp	0.905	0.852
3	turn	0.983	0.728	quick	0.958	0.901	quick	0.905	0.851
4	zerotrade	0.983	0.725	dy	0.957	0.897	age	0.905	0.851
5	idiovol	0.981	0.723	lev	0.959	0.897	roavol	0.904	0.850
6	retvol	0.978	0.721	cash	0.958	0.897	ep	0.905	0.849
7	std_turn	0.983	0.719	zs	0.959	0.896	depr	0.905	0.848
8	HML_Devil	0.989	0.719	cp	0.960	0.894	cash	0.905	0.847
9	maxret	0.981	0.715	roavol	0.957	0.894	rds	0.905	0.843
10	roavol	0.985	0.713	age	0.959	0.894	currat	0.905	0.840
11	age	0.989	0.703	cfp	0.960	0.893	chcsho	0.905	0.840
12	sp	0.985	0.699	op	0.958	0.893	zs	0.903	0.839
13	ala	0.986	0.699	nop	0.958	0.893	nop	0.904	0.839
14	ndp	0.987	0.686	ebp	0.959	0.893	dy	0.905	0.838
15	orgcap	0.989	0.686	ep	0.958	0.891	lev	0.903	0.838
16	tang	0.990	0.683	rds	0.958	0.890	cfp	0.905	0.838
17	ebp	0.988	0.683	depr	0.958	0.889	stdacc	0.905	0.837
18	invest	0.986	0.683	sp	0.958	0.888	cp	0.905	0.836
19	dpia	0.986	0.681	currat	0.958	0.887	stdcf	0.905	0.836
20	UMD	0.989	0.678	kz	0.958	0.887	op	0.904	0.835
21	zs	0.986	0.675	chcsho	0.957	0.884	ebp	0.903	0.835
22	grltnoa	0.988	0.675	tang	0.960	0.884	tang	0.904	0.833
23	dy	0.988	0.672	ato	0.958	0.884	kz	0.903	0.831
24	HML	0.987	0.672	stdacc	0.958	0.883	ato	0.904	0.831
25	kz	0.986	0.669	adm	0.958	0.881	ww	0.904	0.827
26	ob_a	0.989	0.669	cashpr	0.959	0.878	std_turn	0.902	0.826
27	BAB	0.989	0.666	stdcf	0.956	0.878	adm	0.904	0.825
28	op	0.990	0.663	HML	0.958	0.874	idiovol	0.902	0.825
29	realestate_hxz	0.987	0.663	nef	0.956	0.873	maxret	0.902	0.825
30	ol	0.987	0.663	std_turn	0.956	0.870	baspread	0.902	0.820

Table 7: Ranked Three Data Set Comparison (Cont.)

	Ten Year Data			Twenty Year Data			Thirty Year Data		
	Factor	Market Factor Strength	Risk Factor Strength	Factor	Market Factor Strength	Risk Factor Strength	Factor	Market Factor Strength	Risk Factor Strength
31	adm	0.988	0.660	idiovol	0.955	0.870	IPO	0.905	0.818
32	lev	0.986	0.657	zerotrade	0.953	0.865	nef	0.902	0.818
33	nxf	0.989	0.651	turn	0.955	0.864	sp	0.903	0.817
34	nop	0.989	0.651	ww	0.959	0.863	turn	0.902	0.813
35	pm	0.986	0.648	maxret	0.956	0.863	retvol	0.902	0.813
36	pchcapx3	0.988	0.644	absacc	0.960	0.859	zerotrade	0.900	0.812
37	nef	0.988	0.644	baspread	0.955	0.854	absacc	0.905	0.812
38	cash	0.989	0.637	hire	0.959	0.851	HML	0.903	0.811
39	QMJ	0.978	0.637	IPO	0.960	0.850	lgr	0.905	0.810
40	rds	0.989	0.634	lgr	0.959	0.850	cashpr	0.903	0.808
41	LIQ_PS	0.988	0.634	nxf	0.956	0.849	dcol	0.905	0.807
42	ato	0.988	0.634	retvol	0.955	0.848	beta	0.900	0.806
43	salerec	0.992	0.630	salerec	0.957	0.847	RMW	0.904	0.806
44	currat	0.989	0.626	RMW	0.957	0.847	hire	0.905	0.805
45	acc	0.989	0.619	beta	0.954	0.846	salerec	0.905	0.803
46	stdcf	0.989	0.619	sin	0.959	0.844	nxf	0.903	0.801
47	HXZ_ROE	0.989	0.619	acc	0.960	0.843	acc	0.904	0.797
48	depr	0.988	0.615	bm_ia	0.960	0.843	dfin	0.902	0.791
49	noa	0.989	0.615	dcol	0.959	0.838	nincr	0.904	0.790
50	cashpr	0.987	0.615	dfin	0.959	0.838	noa	0.902	0.787
51	absacc	0.989	0.615	HML_Devil	0.953	0.838	HML_Devil	0.902	0.781
52	gma	0.987	0.615	HXZ_IA	0.960	0.838	HXZ_IA	0.904	0.780
53	dncl	0.986	0.611	nincr	0.959	0.834	rdm	0.904	0.778
54	ms	0.980	0.611	rna	0.958	0.826	rna	0.904	0.778
55	rna	0.989	0.611	noa	0.957	0.825	rd	0.903	0.774
56	STR	0.987	0.607	herf	0.957	0.824	bm_ia	0.904	0.772
57	rdm	0.988	0.607	rdm	0.958	0.823	sgr	0.904	0.769
58	chcsho	0.987	0.607	sgr	0.958	0.819	ps	0.904	0.769
59	sin	0.987	0.607	dnco	0.959	0.816	sin	0.904	0.769
60	salecash	0.989	0.602	ps	0.957	0.807	realestate_hxz	0.905	0.769
61	dnco	0.988	0.598	CMA	0.960	0.805	herf	0.902	0.766
62	quick	0.989	0.593	egr_hxz	0.958	0.803	dnco	0.904	0.761
63	stdacc	0.989	0.593	realestate_hxz	0.957	0.798	CMA	0.905	0.759
64	poa	0.988	0.593	gad	0.958	0.788	egr_hxz	0.904	0.750

Table 7: Ranked Three Data Set Comparison (Cont.)

	Ten Year Data			Twenty Year Data			Thirty Year Data		
	Factor	Market Factor Strength	Risk Factor Strength	Factor	Market Factor Strength	Risk Factor Strength	Factor	Market Factor Strength	Risk Factor Strength
65	cp	0.988	0.589	rd	0.958	0.787	ob_a	0.903	0.745
66	tb	0.988	0.589	ol	0.954	0.787	ol	0.902	0.741
67	HXZ_IA	0.987	0.584	cinvest_a	0.959	0.784	cinvest_a	0.903	0.739
68	saleinv	0.987	0.579	dolvol	0.960	0.774	gad	0.902	0.723
69	cfp	0.988	0.579	ob_a	0.955	0.764	SMB	0.902	0.721
70	egr	0.987	0.579	ala	0.958	0.762	dolvol	0.904	0.715
71	dnca	0.986	0.579	pchdepr	0.959	0.761	gma	0.902	0.715
72	egr_hxz	0.988	0.579	BAB	0.960	0.757	ala	0.904	0.715
73	os	0.984	0.569	gma	0.955	0.756	cto	0.902	0.710
74	pps	0.983	0.563	pchcapx3	0.957	0.752	aeavol	0.905	0.710
75	cto	0.987	0.563	dnca	0.958	0.747	BAB	0.905	0.710
76	grltnoa_hxz	0.986	0.563	SMB	0.957	0.745	convind	0.904	0.710
77	cei	0.988	0.563	poa	0.957	0.739	tb	0.902	0.708
78	CMA	0.988	0.563	aeavol	0.961	0.737	QMJ	0.903	0.708
79	em	0.989	0.552	tb	0.953	0.732	pricedelay	0.904	0.701
80	ww	0.990	0.546	grltnoa_hxz	0.958	0.730	egr	0.902	0.699
81	std_dolvol	0.987	0.539	cei	0.953	0.730	orgcap	0.902	0.699
82	grcapx	0.986	0.539	indmom	0.956	0.725	pchdepr	0.903	0.696
83	pctacc	0.989	0.539	egr	0.958	0.725	indmom	0.902	0.696
84	ep	0.989	0.533	moms12m	0.957	0.725	dcoa	0.902	0.696
85	pricedelay	0.989	0.533	dsti	0.957	0.723	moms12m	0.903	0.694
86	hire	0.988	0.519	orgcap	0.956	0.715	pchcapx3	0.902	0.691
87	SMB	0.987	0.512	pchcurrat	0.958	0.710	cei	0.902	0.691
88	pchcapx_ia	0.989	0.512	UMD	0.951	0.706	roic	0.902	0.691
89	aeavol	0.988	0.512	dcoa	0.959	0.706	pm	0.903	0.691
90	moms12m	0.987	0.512	roic	0.951	0.703	dnca	0.902	0.689
91	cashdebt	0.984	0.504	QMJ	0.951	0.703	saleinv	0.903	0.686
92	lgr	0.987	0.504	cinvest	0.958	0.701	grltnoa_hxz	0.903	0.683
93	cinvest	0.988	0.496	HXZ_ROE	0.957	0.699	poa	0.903	0.681
94	herf	0.987	0.496	cto	0.955	0.694	HXZ_ROE	0.905	0.678
95	bm_ia	0.988	0.487	pctacc	0.954	0.694	UMD	0.902	0.672
96	cfp_ia	0.987	0.479	pricedelay	0.958	0.691	pctacc	0.902	0.672
97	cinvest_a	0.989	0.479	pchcapx_ia	0.957	0.681	cinvest	0.903	0.660
98	chmom	0.989	0.469	convind	0.955	0.669	dsti	0.902	0.660

Table 7: Ranked Three Data Set Comparison (Cont.)

	Ten Year Data			Twenty Year Data			Thirty Year Data		
	Factor	Market Factor Strength	Risk Factor Strength	Factor	Market Factor Strength	Risk Factor Strength	Factor	Market Factor Strength	Risk Factor Strength
99	RMW	0.987	0.469	cdi	0.958	0.654	em	0.902	0.657
100	sue	0.987	0.459	rsup	0.957	0.651	pchcurrat	0.902	0.654
101	mom36m	0.986	0.459	chtx	0.958	0.644	ms	0.902	0.648
102	indmom	0.987	0.459	invest	0.957	0.644	invest	0.902	0.641
103	dcoa	0.988	0.459	em	0.952	0.644	pchcapx_ia	0.902	0.630
104	etr	0.986	0.448	pm	0.957	0.641	os	0.900	0.623
105	chinv	0.988	0.448	saleinv	0.955	0.637	chtx	0.902	0.623
106	ill	0.988	0.448	ta	0.958	0.634	dpia	0.902	0.623
107	roic	0.986	0.448	dpia	0.957	0.634	cdi	0.903	0.623
108	convind	0.988	0.448	pchquick	0.957	0.626	pps	0.902	0.611
109	sgr	0.988	0.437	os	0.948	0.626	roaq	0.900	0.602
110	IPO	0.989	0.437	ms	0.950	0.619	rs	0.902	0.584
111	dolvol	0.989	0.437	roaq	0.953	0.607	rsup	0.902	0.579
112	dcol	0.987	0.425	grcapx	0.955	0.593	chinv	0.902	0.569
113	nincr	0.989	0.411	pps	0.952	0.589	cfp_ia	0.902	0.563
114	chempia	0.987	0.411	ndf	0.957	0.589	ta	0.903	0.563
115	rs	0.988	0.411	cfp_ia	0.957	0.584	cashdebt	0.900	0.557
116	pchcapx	0.988	0.411	dncl	0.957	0.584	ndf	0.902	0.557
117	chtx	0.988	0.397	pchsale_pchrect	0.955	0.574	grcapx	0.902	0.552
118	ivg	0.988	0.381	mom6m	0.958	0.569	STR	0.902	0.546
119	LTR	0.985	0.364	rs	0.955	0.563	pchcapx	0.902	0.546
120	mom6m	0.987	0.364	pchcapx	0.958	0.563	pchquick	0.902	0.539
121	cdi	0.987	0.364	cashdebt	0.951	0.557	grltnoa	0.902	0.539
122	chatoia	0.987	0.364	pchsaleinv	0.955	0.557	pchsaleinv	0.902	0.519
123	gad	0.985	0.364	chempia	0.958	0.557	dncl	0.902	0.519
124	pchcurrat	0.988	0.297	LIQ_PS	0.956	0.557	ivg	0.902	0.504
125	pchgm_pchsale	0.988	0.297	dwc	0.955	0.546	mom6m	0.902	0.496
126	rd	0.986	0.297	grltnoa	0.956	0.533	chempia	0.902	0.496
127	dsti	0.989	0.297	STR	0.956	0.526	LIQ_PS	0.902	0.496
128	dfnl	0.987	0.297	dfnl	0.955	0.519	mom36m	0.902	0.479
129	roaq	0.986	0.297	mom36m	0.957	0.496	std_dolvol	0.903	0.459
130	pchdepr	0.988	0.266	std_dolvol	0.955	0.496	pchsale_pchinvt	0.902	0.448
131	dnoa	0.988	0.230	sue	0.956	0.487	pchsale_pchxsga	0.902	0.448
132	ta	0.988	0.230	LTR	0.954	0.487	dwc	0.902	0.448

Table 7: Ranked Three Data Set Comparison (Cont.)

	Ten Year Data			Twenty Year Data			Thirty Year Data		
	Factor	Market Factor Strength	Risk Factor Strength	Factor	Market Factor Strength	Risk Factor Strength	Factor	Market Factor Strength	Risk Factor Strength
133	chpmia	0.987	0.230	chmom	0.953	0.479	dfnl	0.902	0.437
134	pchquick	0.987	0.182	pchsale_pchinvt	0.955	0.448	chmom	0.902	0.437
135	dfin	0.988	0.182	chatoia	0.957	0.437	pchsale_pchrect	0.902	0.425
136	rsup	0.988	0.182	pchsale_pchxsga	0.957	0.425	sue	0.902	0.397
137	pchsaleinv	0.988	0.115	lfe	0.956	0.425	LTR	0.902	0.381
138	pchsale_pchinvt	0.988	0.115	chinv	0.956	0.397	pchgm_pchsale	0.902	0.322
139	pchsale_pchrect	0.988	0.115	ivg	0.957	0.397	lfe	0.902	0.297
140	ps	0.990	0.115	pchgm_pchsale	0.957	0.381	ill	0.902	0.297
141	dwc	0.989	0.115	etr	0.955	0.344	dnoa	0.902	0.182
142	pchsale_pchxsga	0.989	0.000	chpmia	0.957	0.344	ear	0.903	0.182
143	lfe	0.988	0.000	ill	0.955	0.266	chatoia	0.902	0.182
144	ndf	0.986	0.000	dnoa	0.955	0.266	chpmia	0.902	0.182
145	ear	0.988	0.000	ear	0.958	0.266	etr	0.902	0.115

Notes: This table presents the estimation results of factors' strength, ordered decreasingly by risk factor strength. For the estimation, we use the method from Section 3.3, with one market factor and one risk factor. The three data set is describe in the section 5.1

Table 8: Three Data Set Comparison

	Factor	10 Year Strength	20 Year Strength	30 Year Strength	Mean	Standard Deviation
1	ps	0.115	0.807	0.769	0.564	0.318
2	dfin	0.182	0.838	0.791	0.604	0.299
3	ndf	0.000	0.589	0.557	0.382	0.270
4	rd	0.297	0.787	0.774	0.619	0.228
5	pchdepr	0.266	0.761	0.696	0.574	0.219
6	rsup	0.182	0.651	0.579	0.471	0.206
7	pchsale_pchxsga	0.000	0.425	0.448	0.291	0.206
8	pchsaleinv	0.115	0.557	0.519	0.397	0.200
9	pchquick	0.182	0.626	0.539	0.449	0.192
10	pchsale_pchrect	0.115	0.574	0.425	0.371	0.191
11	nincr	0.411	0.834	0.790	0.678	0.190
12	dsti	0.297	0.723	0.660	0.560	0.188
13	dcol	0.425	0.838	0.807	0.690	0.188
14	IPO	0.437	0.850	0.818	0.702	0.188
15	gad	0.364	0.788	0.723	0.625	0.187
16	dwc	0.115	0.546	0.448	0.370	0.185
17	pchcurrat	0.297	0.710	0.654	0.554	0.183
18	lfe	0.000	0.425	0.297	0.240	0.178
19	ta	0.230	0.634	0.563	0.475	0.176
20	sgr	0.437	0.819	0.769	0.675	0.170
21	RMW	0.469	0.847	0.806	0.707	0.169
22	ep	0.533	0.891	0.849	0.758	0.160
23	pchsale_pchinvt	0.115	0.448	0.448	0.337	0.157
24	lgr	0.504	0.850	0.810	0.721	0.155
25	bm_ia	0.487	0.843	0.772	0.701	0.154
26	dolvol	0.437	0.774	0.715	0.642	0.147
27	hire	0.519	0.851	0.805	0.725	0.147
28	roaq	0.297	0.607	0.602	0.502	0.145
29	herf	0.496	0.824	0.766	0.695	0.143
30	ww	0.546	0.863	0.827	0.745	0.142
31	etr	0.448	0.344	0.115	0.302	0.139
32	cfp	0.579	0.893	0.838	0.770	0.137
33	quick	0.593	0.901	0.851	0.782	0.135
34	cinvest_a	0.479	0.784	0.739	0.667	0.135
35	cp	0.589	0.894	0.836	0.773	0.133
36	salecash	0.602	0.902	0.857	0.787	0.132
37	cdi	0.364	0.654	0.623	0.547	0.130
38	stdacc	0.593	0.883	0.837	0.771	0.127
39	chcsho	0.607	0.884	0.840	0.777	0.122
40	depr	0.615	0.889	0.848	0.784	0.121
41	indmom	0.459	0.725	0.696	0.627	0.119
42	roic	0.448	0.703	0.691	0.614	0.117
43	convind	0.448	0.669	0.710	0.609	0.115
44	dcoa	0.459	0.706	0.696	0.620	0.114
45	stdcf	0.619	0.878	0.836	0.778	0.114
46	currat	0.626	0.887	0.840	0.785	0.113
47	cash	0.637	0.897	0.847	0.794	0.113
48	chtx	0.397	0.644	0.623	0.555	0.112
49	rds	0.634	0.890	0.843	0.789	0.111
50	cashpr	0.615	0.878	0.808	0.767	0.111
51	ear	0.000	0.266	0.182	0.149	0.111

Table 8: Three Data Set Comparison (Cont.)

	Factor	10 Year Strength	20 Year Strength	30 Year Strength	Mean	Standard Deviation
52	HXZ_IA	0.584	0.838	0.780	0.734	0.109
53	ato	0.634	0.884	0.831	0.783	0.107
54	chatoia	0.364	0.437	0.182	0.328	0.107
55	absacc	0.615	0.859	0.812	0.762	0.106
56	SMB	0.512	0.745	0.721	0.659	0.105
57	CMA	0.563	0.805	0.759	0.709	0.105
58	nop	0.651	0.893	0.839	0.794	0.104
59	lev	0.657	0.897	0.838	0.798	0.102
60	aeavol	0.512	0.737	0.710	0.653	0.101
61	sin	0.607	0.844	0.769	0.740	0.099
62	nef	0.644	0.873	0.818	0.778	0.097
63	op	0.663	0.893	0.835	0.797	0.097
64	acc	0.619	0.843	0.797	0.753	0.097
65	egr_hxz	0.579	0.803	0.750	0.711	0.096
66	dy	0.672	0.897	0.838	0.803	0.095
67	moms12m	0.512	0.725	0.694	0.644	0.094
68	adm	0.660	0.881	0.825	0.789	0.094
69	salerec	0.630	0.847	0.803	0.760	0.094
70	zs	0.675	0.896	0.839	0.803	0.094
71	rdm	0.607	0.823	0.778	0.736	0.093
72	ndp	0.686	0.904	0.852	0.814	0.093
73	dnco	0.598	0.816	0.761	0.725	0.093
74	rna	0.611	0.826	0.778	0.738	0.092
75	kz	0.669	0.887	0.831	0.796	0.092
76	dfnl	0.297	0.519	0.437	0.418	0.092
77	noa	0.615	0.825	0.787	0.742	0.091
78	cinvest	0.496	0.701	0.660	0.619	0.089
79	ebp	0.683	0.893	0.835	0.804	0.088
80	tang	0.683	0.884	0.833	0.800	0.085
81	mom6m	0.364	0.569	0.496	0.476	0.085
82	nxf	0.651	0.849	0.801	0.767	0.084
83	HML	0.672	0.874	0.811	0.786	0.084
84	age	0.703	0.894	0.851	0.816	0.082
85	ill	0.448	0.266	0.297	0.337	0.080
86	sp	0.699	0.888	0.817	0.801	0.078
87	roavol	0.713	0.894	0.850	0.819	0.077
88	pricedelay	0.533	0.691	0.701	0.642	0.077
89	rs	0.411	0.563	0.584	0.519	0.077
90	chinv	0.448	0.397	0.569	0.471	0.072
91	cei	0.563	0.730	0.691	0.661	0.071
92	pchcapx_ia	0.512	0.681	0.630	0.608	0.071
93	grltnoa_hxz	0.563	0.730	0.683	0.659	0.070
94	dnca	0.579	0.747	0.689	0.671	0.070
95	pctacc	0.539	0.694	0.672	0.635	0.068
96	chpmia	0.230	0.344	0.182	0.252	0.068
97	pchcapx	0.411	0.563	0.546	0.507	0.068
98	cto	0.563	0.694	0.710	0.656	0.066
99	grltnoa	0.675	0.533	0.539	0.582	0.066
100	egr	0.579	0.725	0.699	0.668	0.064
101	std_turn	0.719	0.870	0.826	0.805	0.064
102	maxret	0.715	0.863	0.825	0.801	0.063

Table 8: Three Data Set Comparison (Cont.)

	Factor	10 Year Strength	20 Year Strength	30 Year Strength	Mean	Standard Deviation
103	tb	0.589	0.732	0.708	0.676	0.063
104	idiovol	0.723	0.870	0.825	0.806	0.061
105	poa	0.593	0.739	0.681	0.671	0.060
106	chempia	0.411	0.557	0.496	0.488	0.060
107	gma	0.615	0.756	0.715	0.695	0.059
108	realestate_hxz	0.663	0.798	0.769	0.743	0.058
109	zerotrade	0.725	0.865	0.812	0.801	0.058
110	turn	0.728	0.864	0.813	0.802	0.056
111	LIQ_PS	0.634	0.557	0.496	0.562	0.056
112	LTR	0.364	0.487	0.381	0.411	0.055
113	ivg	0.381	0.397	0.504	0.427	0.055
114	retvol	0.721	0.848	0.813	0.794	0.054
115	baspread	0.730	0.854	0.820	0.801	0.053
116	ol	0.663	0.787	0.741	0.731	0.051
117	HML_Devil	0.719	0.838	0.781	0.779	0.049
118	em	0.552	0.644	0.657	0.618	0.047
119	cfp_ia	0.479	0.584	0.563	0.542	0.046
120	pchcapx3	0.644	0.752	0.691	0.696	0.044
121	saleinv	0.579	0.637	0.686	0.634	0.044
122	ob_a	0.669	0.764	0.745	0.726	0.041
123	beta	0.749	0.846	0.806	0.800	0.040
124	dncl	0.611	0.584	0.519	0.571	0.038
125	sue	0.459	0.487	0.397	0.448	0.038
126	BAB	0.666	0.757	0.710	0.711	0.037
127	pchgm_pchsale	0.297	0.381	0.322	0.333	0.035
128	dnoa	0.230	0.266	0.182	0.226	0.035
129	STR	0.607	0.526	0.546	0.559	0.034
130	HXZ_ROE	0.619	0.699	0.678	0.665	0.034
131	std_dolvol	0.539	0.496	0.459	0.498	0.033
132	QMJ	0.637	0.703	0.708	0.683	0.032
133	ala	0.699	0.762	0.715	0.725	0.027
134	os	0.569	0.626	0.623	0.606	0.026
135	cashdebt	0.504	0.557	0.557	0.540	0.025
136	dpia	0.681	0.634	0.623	0.646	0.025
137	grcapx	0.539	0.593	0.552	0.561	0.023
138	pm	0.648	0.641	0.691	0.660	0.022
139	pps	0.563	0.589	0.611	0.587	0.019
140	invest	0.683	0.644	0.641	0.656	0.019
141	chmom	0.469	0.479	0.437	0.461	0.018
142	ms	0.611	0.619	0.648	0.626	0.016
143	mom36m	0.459	0.496	0.479	0.478	0.015
144	UMD	0.678	0.706	0.672	0.685	0.015
145	orgcap	0.686	0.715	0.699	0.700	0.012

Notes: This table presents the estimated factor strength, using data from three different data set. For the data description see Section 5.1. The table also presents the calculated mean and standard deviation of each factors. The table is ordered decreasingly by the standard deviation.

Table 9: Ten and Twenty Comparison

	Factor	10 Year Strength	20 Year Strength	Difference
1	ps	0.115	0.807	0.692
2	dfin	0.182	0.838	0.656
3	ndf	0.000	0.589	0.589
4	pchdepr	0.266	0.761	0.494
5	rd	0.297	0.787	0.490
6	rsup	0.182	0.651	0.469
7	pchsale_pchrect	0.115	0.574	0.459
8	pchquick	0.182	0.626	0.445
9	pchsaleinv	0.115	0.557	0.443
10	dwc	0.115	0.546	0.431
11	dsti	0.297	0.723	0.427
12	pchsale_pchxsga	0.000	0.425	0.425
13	lfe	0.000	0.425	0.425
14	gad	0.364	0.788	0.425
15	nincr	0.411	0.834	0.423
16	pchcurrat	0.297	0.710	0.414
17	dcol	0.425	0.838	0.414
18	IPO	0.437	0.850	0.413
19	ta	0.230	0.634	0.404
20	sgr	0.437	0.819	0.382
21	RMW	0.469	0.847	0.378
22	ep	0.533	0.891	0.358
23	bm_ia	0.487	0.843	0.356
24	lgr	0.504	0.850	0.346
25	dolvol	0.437	0.774	0.337
26	pchsale_pchinvt	0.115	0.448	0.334
27	hire	0.519	0.851	0.332
28	herf	0.496	0.824	0.328
29	ww	0.546	0.863	0.318
30	cfp	0.579	0.893	0.314
31	roaq	0.297	0.607	0.310
32	quick	0.593	0.901	0.308
33	cp	0.589	0.894	0.306
34	cinvest_a	0.479	0.784	0.306
35	salecash	0.602	0.902	0.300
36	cdi	0.364	0.654	0.290
37	stdacc	0.593	0.883	0.290
38	chcsho	0.607	0.884	0.278
39	depr	0.615	0.889	0.274
40	indmom	0.459	0.725	0.266
41	ear	0.000	0.266	0.266
42	cashpr	0.615	0.878	0.263
43	currat	0.626	0.887	0.260
44	cash	0.637	0.897	0.260
45	stdcf	0.619	0.878	0.259
46	rds	0.634	0.890	0.256
47	roic	0.448	0.703	0.255
48	HXZ_IA	0.584	0.838	0.254
49	ato	0.634	0.884	0.250
50	chtx	0.397	0.644	0.247
51	dcoa	0.459	0.706	0.247

Table 9: Ten and Twenty Comparison (Cont.)

	Factor	10 Year Strength	20 Year Strength	Difference
52	absacc	0.615	0.859	0.244
53	nop	0.651	0.893	0.242
54	CMA	0.563	0.805	0.241
55	lev	0.657	0.897	0.240
56	sin	0.607	0.844	0.238
57	SMB	0.512	0.745	0.233
58	op	0.663	0.893	0.230
59	nef	0.644	0.873	0.229
60	aeavol	0.512	0.737	0.226
61	dy	0.672	0.897	0.225
62	acc	0.619	0.843	0.225
63	egr_hxz	0.579	0.803	0.224
64	dfnl	0.297	0.519	0.222
65	convind	0.448	0.669	0.221
66	zs	0.675	0.896	0.221
67	adm	0.660	0.881	0.221
68	ndp	0.686	0.904	0.218
69	dnco	0.598	0.816	0.218
70	kz	0.669	0.887	0.217
71	salerec	0.630	0.847	0.217
72	rdm	0.607	0.823	0.216
73	rna	0.611	0.826	0.215
74	moms12m	0.512	0.725	0.214
75	noa	0.615	0.825	0.210
76	ebp	0.683	0.893	0.210
77	cinvest	0.496	0.701	0.205
78	mom6m	0.364	0.569	0.205
79	HML	0.672	0.874	0.202
80	tang	0.683	0.884	0.200
81	nxf	0.651	0.849	0.198
82	age	0.703	0.894	0.191
83	sp	0.699	0.888	0.189
84	roavol	0.713	0.894	0.182
85	ill	0.448	0.266	0.182
86	pchcapx_ia	0.512	0.681	0.169
87	dnca	0.579	0.747	0.168
88	grltnoa_hxz	0.563	0.730	0.166
89	cei	0.563	0.730	0.166
90	pricedelay	0.533	0.691	0.158
91	pctacc	0.539	0.694	0.154
92	rs	0.411	0.563	0.152
93	pchcapx	0.411	0.563	0.152
94	std_turn	0.719	0.870	0.151
95	maxret	0.715	0.863	0.149
96	idiovol	0.723	0.870	0.147
97	egr	0.579	0.725	0.147
98	poa	0.593	0.739	0.146
99	chempia	0.411	0.557	0.146
100	tb	0.589	0.732	0.143
101	grltnoa	0.675	0.533	0.142
102	gma	0.615	0.756	0.141

Table 9: Ten and Twenty Comparison (Cont.)

	Factor	10 Year Strength	20 Year Strength	Difference
103	zerotrade	0.725	0.865	0.140
104	turn	0.728	0.864	0.137
105	realestate_hxz	0.663	0.798	0.135
106	cto	0.563	0.694	0.131
107	retvol	0.721	0.848	0.127
108	baspread	0.730	0.854	0.125
109	LTR	0.364	0.487	0.124
110	ol	0.663	0.787	0.124
111	HML_Devil	0.719	0.838	0.119
112	chpmia	0.230	0.344	0.115
113	pchcapx3	0.644	0.752	0.108
114	cfp_ia	0.479	0.584	0.105
115	etr	0.448	0.344	0.104
116	beta	0.749	0.846	0.098
117	ob_a	0.669	0.764	0.095
118	em	0.552	0.644	0.093
119	BAB	0.666	0.757	0.091
120	pchgm_pchsale	0.297	0.381	0.085
121	STR	0.607	0.526	0.080
122	HXZ_ROE	0.619	0.699	0.080
123	LIQ_PS	0.634	0.557	0.076
124	chatoia	0.364	0.437	0.073
125	QMJ	0.637	0.703	0.066
126	ala	0.699	0.762	0.064
127	saleinv	0.579	0.637	0.059
128	os	0.569	0.626	0.058
129	grcapx	0.539	0.593	0.054
130	cashdebt	0.504	0.557	0.053
131	chinv	0.448	0.397	0.051
132	dpia	0.681	0.634	0.047
133	std_dolvol	0.539	0.496	0.043
134	invest	0.683	0.644	0.039
135	mom36m	0.459	0.496	0.037
136	dnoa	0.230	0.266	0.037
137	orgcap	0.686	0.715	0.029
138	sue	0.459	0.487	0.028
139	UMD	0.678	0.706	0.028
140	dncl	0.611	0.584	0.027
141	pps	0.563	0.589	0.026
142	ivg	0.381	0.397	0.016
143	chmom	0.469	0.479	0.009
144	ms	0.611	0.619	0.008
145	pm	0.648	0.641	0.007

Notes: This table presents the estimated factor strength, using data from the ten year data and twenty year data. For the data description see Section 5.1. The table also presents the difference between the two estimated strengths. The table is ordered decreasingly by the difference.

Table 10: Ten and Thirty Comparison

	Factor	10 Year Strength	30 Year Strength	Difference
1	ps	0.115	0.769	0.654
2	dfin	0.182	0.791	0.609
3	ndf	0.000	0.557	0.557
4	rd	0.297	0.774	0.477
5	pchsale_pchxsga	0.000	0.448	0.448
6	pchdepr	0.266	0.696	0.430
7	pchsaleinv	0.115	0.519	0.404
8	rsup	0.182	0.579	0.397
9	dcol	0.425	0.807	0.382
10	IPO	0.437	0.818	0.381
11	nincr	0.411	0.790	0.378
12	dsti	0.297	0.660	0.364
13	gad	0.364	0.723	0.360
14	pchquick	0.182	0.539	0.358
15	pchcurrat	0.297	0.654	0.358
16	RMW	0.469	0.806	0.337
17	pchsale_pchinvt	0.115	0.448	0.334
18	etr	0.448	0.115	0.334
19	dwc	0.115	0.448	0.334
20	ta	0.230	0.563	0.334
21	sgr	0.437	0.769	0.332
22	ep	0.533	0.849	0.316
23	pchsale_pchrect	0.115	0.425	0.310
24	roaq	0.297	0.602	0.306
25	lgr	0.504	0.810	0.306
26	lfe	0.000	0.297	0.297
27	hire	0.519	0.805	0.285
28	bm_ia	0.487	0.772	0.285
29	ww	0.546	0.827	0.282
30	dolvol	0.437	0.715	0.278
31	herf	0.496	0.766	0.270
32	convind	0.448	0.710	0.262
33	cinvest_a	0.479	0.739	0.261
34	cfp	0.579	0.838	0.259
35	cdi	0.364	0.623	0.259
36	quick	0.593	0.851	0.258
37	salecash	0.602	0.857	0.255
38	cp	0.589	0.836	0.247
39	stdacc	0.593	0.837	0.244
40	roic	0.448	0.691	0.243
41	indmom	0.459	0.696	0.237
42	dcoa	0.459	0.696	0.237
43	chcsho	0.607	0.840	0.234
44	depr	0.615	0.848	0.233
45	chtx	0.397	0.623	0.226
46	stdcf	0.619	0.836	0.217
47	currat	0.626	0.840	0.214
48	cash	0.637	0.847	0.210
49	SMB	0.512	0.721	0.210
50	rds	0.634	0.843	0.209
51	aeavol	0.512	0.710	0.199

Table 10: Ten and Thirty Comparison (Cont.)

	Factor	10 Year Strength	30 Year Strength	Difference
52	absacc	0.615	0.812	0.197
53	ato	0.634	0.831	0.197
54	CMA	0.563	0.759	0.196
55	HXZ_IA	0.584	0.780	0.196
56	cashpr	0.615	0.808	0.194
57	nop	0.651	0.839	0.188
58	moms12m	0.512	0.694	0.182
59	chatoia	0.364	0.182	0.182
60	ear	0.000	0.182	0.182
61	lev	0.657	0.838	0.181
62	acc	0.619	0.797	0.178
63	nef	0.644	0.818	0.174
64	salerec	0.630	0.803	0.173
65	rs	0.411	0.584	0.172
66	noa	0.615	0.787	0.172
67	rdm	0.607	0.778	0.172
68	op	0.663	0.835	0.172
69	egr_hxz	0.579	0.750	0.172
70	pricedelay	0.533	0.701	0.168
71	rna	0.611	0.778	0.167
72	ndp	0.686	0.852	0.166
73	dy	0.672	0.838	0.166
74	adm	0.660	0.825	0.165
75	cinvest	0.496	0.660	0.164
76	zs	0.675	0.839	0.164
77	dnco	0.598	0.761	0.163
78	sin	0.607	0.769	0.162
79	kz	0.669	0.831	0.161
80	ill	0.448	0.297	0.152
81	ebp	0.683	0.835	0.152
82	nxf	0.651	0.801	0.150
83	tang	0.683	0.833	0.150
84	age	0.703	0.851	0.148
85	cto	0.563	0.710	0.147
86	dfnl	0.297	0.437	0.140
87	HML	0.672	0.811	0.139
88	LIQ_PS	0.634	0.496	0.138
89	roavol	0.713	0.850	0.138
90	grltnoa	0.675	0.539	0.136
91	pchcapx	0.411	0.546	0.134
92	pctacc	0.539	0.672	0.133
93	mom6m	0.364	0.496	0.132
94	cei	0.563	0.691	0.128
95	ivg	0.381	0.504	0.123
96	chinv	0.448	0.569	0.120
97	grltnoa_hxz	0.563	0.683	0.120
98	egr	0.579	0.699	0.120
99	tb	0.589	0.708	0.119
100	pchcapx_ia	0.512	0.630	0.118
101	sp	0.699	0.817	0.118
102	maxret	0.715	0.825	0.110

Table 10: Ten and Thirty Comparison (Cont.)

	Factor	10 Year Strength	30 Year Strength	Difference
103	dnca	0.579	0.689	0.110
104	saleinv	0.579	0.686	0.107
105	std_turn	0.719	0.826	0.107
106	em	0.552	0.657	0.106
107	realestate_hxz	0.663	0.769	0.105
108	idiovol	0.723	0.825	0.102
109	gma	0.615	0.715	0.100
110	retvol	0.721	0.813	0.092
111	dncl	0.611	0.519	0.092
112	baspread	0.730	0.820	0.091
113	poa	0.593	0.681	0.087
114	zerotrade	0.725	0.812	0.087
115	turn	0.728	0.813	0.086
116	cfp_ia	0.479	0.563	0.085
117	chempia	0.411	0.496	0.085
118	std_dolvol	0.539	0.459	0.080
119	ol	0.663	0.741	0.078
120	ob_a	0.669	0.745	0.076
121	QMJ	0.637	0.708	0.071
122	sue	0.459	0.397	0.062
123	HML_Devil	0.719	0.781	0.062
124	STR	0.607	0.546	0.061
125	HXZ_ROE	0.619	0.678	0.059
126	dpia	0.681	0.623	0.058
127	beta	0.749	0.806	0.057
128	os	0.569	0.623	0.054
129	cashdebt	0.504	0.557	0.053
130	dnoa	0.230	0.182	0.048
131	chpmia	0.230	0.182	0.048
132	pps	0.563	0.611	0.048
133	pchcapx3	0.644	0.691	0.047
134	BAB	0.666	0.710	0.044
135	pm	0.648	0.691	0.043
136	invest	0.683	0.641	0.042
137	ms	0.611	0.648	0.037
138	chmom	0.469	0.437	0.032
139	pchgm_pchsale	0.297	0.322	0.026
140	mom36m	0.459	0.479	0.019
141	LTR	0.364	0.381	0.017
142	ala	0.699	0.715	0.016
143	orgcap	0.686	0.699	0.013
144	grcapx	0.539	0.552	0.012
145	UMD	0.678	0.672	0.006

Notes: This table presents the estimated factor strength, using data from the ten year data and thirty year data. For the data description see Section 5.1. The table also presents the difference between the two estimated strengths. The table is ordered decreasingly by the difference.

Table 11: Twenty and Thirty Comparison

	Factor	20 Year Strength	30 Year Strength	Difference
1	chatoia	0.437	0.182	0.255
2	etr	0.344	0.115	0.230
3	chinv	0.397	0.569	0.172
4	chpmia	0.344	0.182	0.162
5	pchsale_pchrect	0.574	0.425	0.149
6	lfe	0.425	0.297	0.128
7	ivg	0.397	0.504	0.107
8	LTR	0.487	0.381	0.106
9	dwc	0.546	0.448	0.097
10	sue	0.487	0.397	0.090
11	pchquick	0.626	0.539	0.087
12	dnoa	0.266	0.182	0.085
13	ear	0.266	0.182	0.085
14	dfnl	0.519	0.437	0.082
15	sin	0.844	0.769	0.075
16	mom6m	0.569	0.496	0.073
17	rsup	0.651	0.579	0.072
18	bm_ia	0.843	0.772	0.071
19	ta	0.634	0.563	0.071
20	sp	0.888	0.817	0.071
21	cashpr	0.878	0.808	0.070
22	gad	0.788	0.723	0.065
23	dncl	0.584	0.519	0.065
24	pchdepr	0.761	0.696	0.065
25	dsti	0.723	0.660	0.063
26	HML	0.874	0.811	0.063
27	chempia	0.557	0.496	0.062
28	LIQ_PS	0.557	0.496	0.062
29	pchcapx3	0.752	0.691	0.061
30	dy	0.897	0.838	0.059
31	lev	0.897	0.838	0.059
32	pchgm_pchsale	0.381	0.322	0.059
33	dolvol	0.774	0.715	0.059
34	poa	0.739	0.681	0.059
35	HXZ_IA	0.838	0.780	0.058
36	cp	0.894	0.836	0.058
37	dnca	0.747	0.689	0.058
38	herf	0.824	0.766	0.058
39	op	0.893	0.835	0.058
40	ebp	0.893	0.835	0.058
41	HML_Devil	0.838	0.781	0.057
42	zs	0.896	0.839	0.057
43	adm	0.881	0.825	0.056
44	kz	0.887	0.831	0.056
45	pchcurrat	0.710	0.654	0.056
46	dnco	0.816	0.761	0.055
47	nef	0.873	0.818	0.055
48	cfp	0.893	0.838	0.055
49	nop	0.893	0.839	0.054
50	zerotrade	0.865	0.812	0.053
51	ato	0.884	0.831	0.053

Table 11: Twenty and Thirty Comparison (Cont.)

	Factor	20 Year Strength	30 Year Strength	Difference
52	egr_hxz	0.803	0.750	0.053
53	ndp	0.904	0.852	0.052
54	turn	0.864	0.813	0.051
55	tang	0.884	0.833	0.051
56	sgr	0.819	0.769	0.050
57	pchcapx_ia	0.681	0.630	0.050
58	pm	0.641	0.691	0.050
59	cash	0.897	0.847	0.050
60	quick	0.901	0.851	0.050
61	nxf	0.849	0.801	0.049
62	saleinv	0.637	0.686	0.049
63	rna	0.826	0.778	0.048
64	ala	0.762	0.715	0.048
65	BAB	0.757	0.710	0.047
66	dfin	0.838	0.791	0.047
67	absacc	0.859	0.812	0.047
68	hire	0.851	0.805	0.047
69	acc	0.843	0.797	0.047
70	rds	0.890	0.843	0.047
71	currat	0.887	0.840	0.047
72	grltnoa_hxz	0.730	0.683	0.046
73	stdacc	0.883	0.837	0.046
74	ol	0.787	0.741	0.046
75	CMA	0.805	0.759	0.046
76	idiovol	0.870	0.825	0.045
77	salecash	0.902	0.857	0.045
78	cinvest_a	0.784	0.739	0.045
79	rdm	0.823	0.778	0.045
80	chcsho	0.884	0.840	0.044
81	std_turn	0.870	0.826	0.044
82	roavol	0.894	0.850	0.044
83	nincr	0.834	0.790	0.044
84	salerec	0.847	0.803	0.044
85	age	0.894	0.851	0.043
86	stdcf	0.878	0.836	0.042
87	chmom	0.479	0.437	0.042
88	grcapx	0.593	0.552	0.042
89	RMW	0.847	0.806	0.041
90	ep	0.891	0.849	0.041
91	convind	0.669	0.710	0.041
92	gma	0.756	0.715	0.041
93	depr	0.889	0.848	0.041
94	lgr	0.850	0.810	0.041
95	cinvest	0.701	0.660	0.041
96	beta	0.846	0.806	0.040
97	cei	0.730	0.691	0.038
98	pchsaleinv	0.557	0.519	0.038
99	maxret	0.863	0.825	0.038
100	ps	0.807	0.769	0.038
101	noa	0.825	0.787	0.038
102	std_dolvol	0.496	0.459	0.037

Table 11: Twenty and Thirty Comparison (Cont.)

	Factor	20 Year Strength	30 Year Strength	Difference
103	ww	0.863	0.827	0.036
104	retvol	0.848	0.813	0.035
105	baspread	0.854	0.820	0.034
106	UMD	0.706	0.672	0.033
107	IPO	0.850	0.818	0.032
108	moms12m	0.725	0.694	0.032
109	cdi	0.654	0.623	0.031
110	ndf	0.589	0.557	0.031
111	dcol	0.838	0.807	0.031
112	ill	0.266	0.297	0.030
113	indmom	0.725	0.696	0.029
114	realestate_hxz	0.798	0.769	0.029
115	ms	0.619	0.648	0.029
116	aeavol	0.737	0.710	0.027
117	egr	0.725	0.699	0.027
118	SMB	0.745	0.721	0.024
119	pchsale_pchxsga	0.425	0.448	0.024
120	tb	0.732	0.708	0.024
121	pps	0.589	0.611	0.022
122	chtx	0.644	0.623	0.022
123	pctacc	0.694	0.672	0.021
124	cfp_ia	0.584	0.563	0.021
125	rs	0.563	0.584	0.021
126	HXZ_ROE	0.699	0.678	0.021
127	STR	0.526	0.546	0.019
128	ob_a	0.764	0.745	0.019
129	pchcapx	0.563	0.546	0.017
130	mom36m	0.496	0.479	0.017
131	cto	0.694	0.710	0.017
132	orgcap	0.715	0.699	0.016
133	rd	0.787	0.774	0.013
134	em	0.644	0.657	0.013
135	roic	0.703	0.691	0.012
136	dpia	0.634	0.623	0.011
137	pricedelay	0.691	0.701	0.010
138	dcoa	0.706	0.696	0.010
139	grltnoa	0.533	0.539	0.006
140	QMJ	0.703	0.708	0.005
141	roaq	0.607	0.602	0.004
142	os	0.626	0.623	0.004
143	invest	0.644	0.641	0.003
144	cashdebt	0.557	0.557	0.000
145	pchsale_pchinvt	0.448	0.448	0.000

Notes: This table presents the estimated factor strength, using data from the twenty year data and thirty year data. For the data description see Section 5.1. The table also presents the difference between the two estimated strengths. The table is ordered decreasingly by the difference.

Table 12: Thirty Year Decompose

Factor	Factor Strength $\hat{\alpha}$ Full Sample	January 1988 to December 1997	January 1998 to December 2007	January 2008 to December 2017	Standard Deviation of Three sub-samples
1 salecash	0.857	0.539	0.823	0.519	0.139
2 ndp	0.852	0.574	0.823	0.557	0.121
3 quick	0.851	0.607	0.819	0.512	0.129
4 age	0.851	0.546	0.817	0.630	0.113
5 ep	0.850	0.504	0.826	0.425	0.174
6 roavol	0.850	0.593	0.826	0.654	0.099
7 depr	0.848	0.637	0.820	0.533	0.119
8 cash	0.847	0.584	0.819	0.552	0.119
9 rds	0.843	0.437	0.810	0.546	0.156
10 currat	0.840	0.626	0.810	0.574	0.101
11 chcsho	0.840	0.504	0.808	0.526	0.139
12 dy	0.839	0.634	0.820	0.626	0.090
13 zs	0.839	0.519	0.813	0.539	0.134
14 nop	0.839	0.615	0.824	0.598	0.103
15 lev	0.838	0.504	0.817	0.539	0.140
16 cfp	0.838	0.512	0.816	0.487	0.149
17 stdacc	0.838	0.322	0.803	0.526	0.197
18 stdcf	0.837	0.381	0.805	0.546	0.174
19 cp	0.836	0.557	0.820	0.512	0.136
20 op	0.835	0.626	0.819	0.615	0.094
21 ebp	0.835	0.557	0.816	0.557	0.122
22 tang	0.833	0.615	0.793	0.626	0.081
23 kz	0.831	0.519	0.813	0.539	0.134
24 ato	0.831	0.364	0.797	0.533	0.178
25 ww	0.827	0.397	0.791	0.496	0.167
26 std_turn	0.826	0.637	0.799	0.657	0.072
27 idiovol	0.826	0.644	0.799	0.657	0.070
28 adm	0.825	0.297	0.786	0.552	0.200
29 maxret	0.825	0.615	0.803	0.654	0.081
30 baspread	0.822	0.630	0.808	0.663	0.077
31 nef	0.819	0.657	0.799	0.589	0.088
32 IPO	0.818	0.230	0.777	0.364	0.233
33 sp	0.817	0.593	0.805	0.630	0.092
34 turn	0.815	0.666	0.798	0.663	0.063
35 retvol	0.815	0.634	0.801	0.660	0.073
36 absacc	0.813	0.469	0.756	0.579	0.118
37 zerotrade	0.812	0.651	0.803	0.663	0.069
38 HML	0.811	0.539	0.815	0.589	0.120
39 lgr	0.810	0.496	0.752	0.397	0.150
40 cashpr	0.808	0.557	0.790	0.519	0.120
41 beta	0.807	0.660	0.807	0.696	0.062
42 dcol	0.807	0.526	0.752	0.266	0.198
43 RMW	0.807	0.448	0.774	0.425	0.159
44 hire	0.805	0.569	0.749	0.448	0.123
45 salerec	0.803	0.557	0.754	0.557	0.093
46 nxf	0.802	0.623	0.762	0.584	0.077
47 acc	0.797	0.519	0.764	0.574	0.105
48 dfin	0.791	0.266	0.772	0.115	0.281

Table 12: Thirty Year Decompose (Cont.)

Factor	Factor Strength $\hat{\alpha}$ Full Sample	January 1988 to December 1997	January 1998 to December 2007	January 2008 to December 2017	Standard Deviation of Three sub-samples
49 nincr	0.790	0.519	0.736	0.364	0.152
50 noa	0.787	0.297	0.741	0.519	0.182
51 HML_Devil	0.781	0.557	0.774	0.619	0.091
52 HXZ_IA	0.780	0.563	0.713	0.519	0.083
53 rdm	0.778	0.519	0.725	0.512	0.099
54 rna	0.778	0.266	0.686	0.519	0.172
55 rd	0.774	0.364	0.715	0.115	0.246
56 bm_ia	0.772	0.584	0.756	0.364	0.160
57 ps	0.770	0.425	0.715	0.000	0.294
58 realestate_hxz	0.770	0.479	0.696	0.611	0.090
59 sgr	0.769	0.546	0.703	0.344	0.147
60 sin	0.769	0.182	0.770	0.469	0.240
61 herf	0.766	0.689	0.728	0.411	0.141
62 dnco	0.761	0.557	0.736	0.512	0.097
63 CMA	0.759	0.496	0.678	0.448	0.099
64 egr_hxz	0.750	0.563	0.666	0.512	0.064
65 ob_a	0.745	0.437	0.637	0.598	0.087
66 ol	0.743	0.644	0.657	0.619	0.016
67 cinvest_a	0.739	0.266	0.708	0.364	0.189
68 gad	0.723	0.000	0.584	0.182	0.244
69 SMB	0.721	0.607	0.699	0.437	0.108
70 ala	0.717	0.651	0.728	0.615	0.047
71 dolvol	0.715	0.397	0.703	0.397	0.144
72 gma	0.715	0.637	0.619	0.519	0.052
73 convind	0.713	0.589	0.563	0.297	0.132
74 cto	0.710	0.657	0.539	0.519	0.061
75 tb	0.710	0.115	0.651	0.448	0.221
76 aeavol	0.710	0.589	0.615	0.425	0.084
77 BAB	0.710	0.425	0.706	0.598	0.116
78 QMJ	0.710	0.546	0.660	0.552	0.053
79 pricedelay	0.699	0.593	0.563	0.411	0.080
80 egr	0.699	0.539	0.563	0.496	0.028
81 orgcap	0.699	0.611	0.615	0.607	0.003
82 pchdepr	0.696	0.000	0.666	0.182	0.281
83 indmom	0.696	0.479	0.630	0.364	0.109
84 dcoa	0.696	0.557	0.546	0.411	0.066
85 roic	0.694	0.563	0.641	0.364	0.117
86 moms12m	0.694	0.411	0.657	0.381	0.124
87 pm	0.694	0.611	0.479	0.602	0.060
88 pchcapx3	0.691	0.539	0.598	0.574	0.024
89 cei	0.691	0.487	0.666	0.448	0.095
90 dnca	0.689	0.487	0.584	0.487	0.045
91 saleinv	0.686	0.663	0.425	0.533	0.098
92 grltnoa_hxz	0.683	0.512	0.602	0.487	0.049
93 poa	0.681	0.552	0.660	0.533	0.056
94 HXZ_ROE	0.681	0.364	0.623	0.519	0.106
95 UMD	0.672	0.496	0.607	0.579	0.047
96 pctacc	0.672	0.512	0.593	0.469	0.052
97 cinvest	0.660	0.397	0.602	0.397	0.097

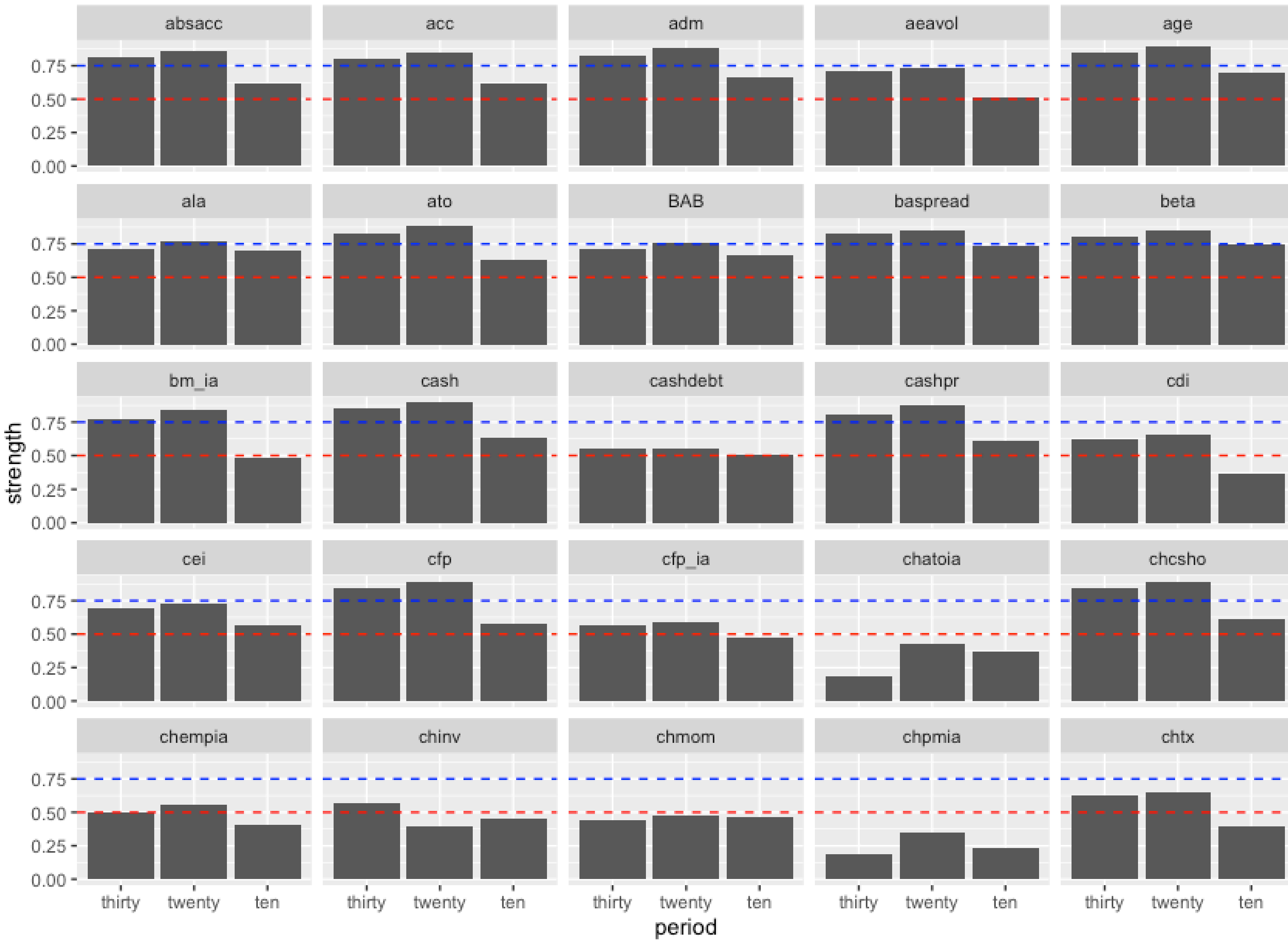
Table 12: Thirty Year Decompose (Cont.)

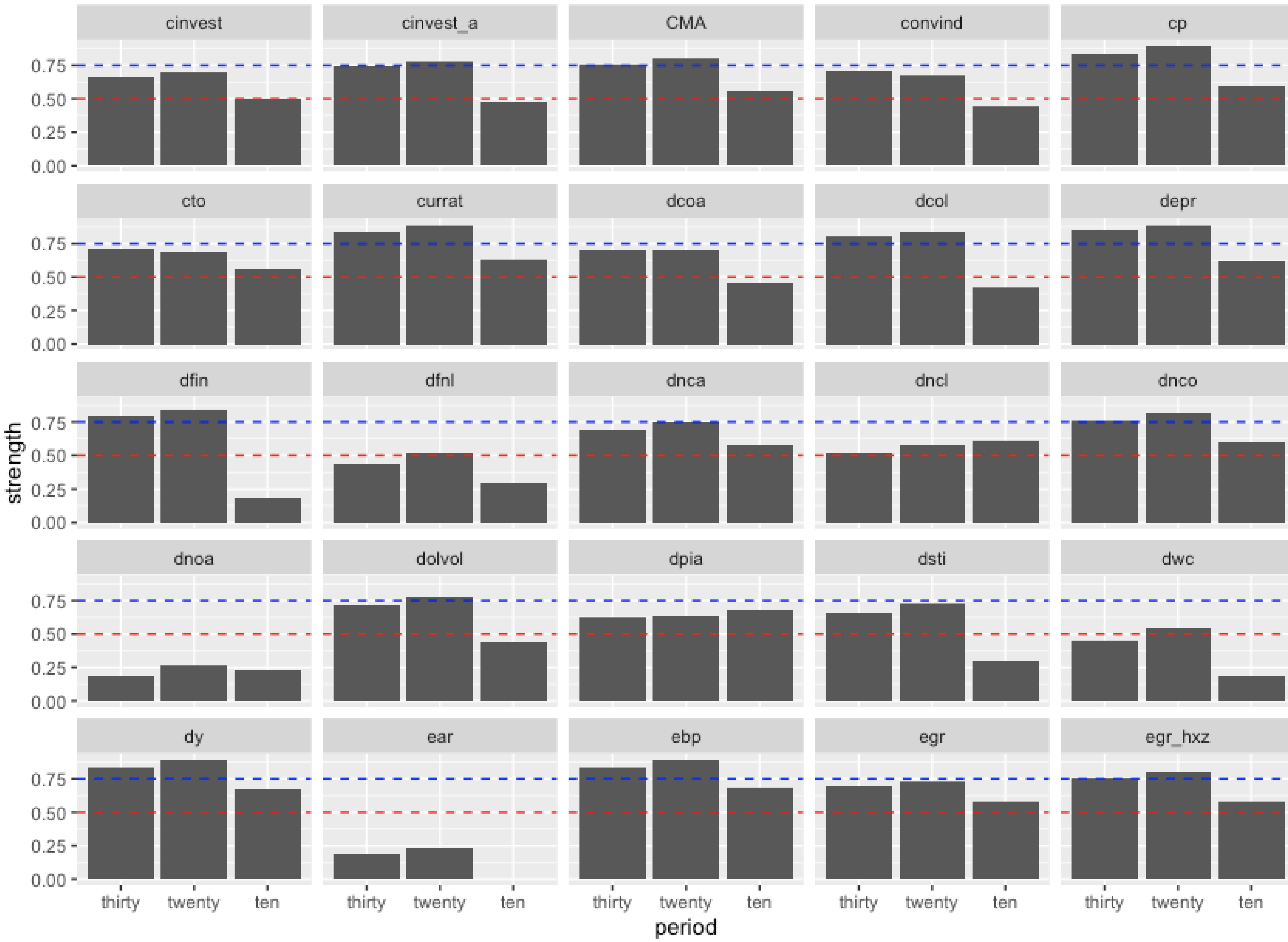
Factor		Factor Strength $\hat{\alpha}$				
		Full Sample	January 1988 to December 1997	January 1998 to December 2007	January 2008 to December 2017	Standard Deviation of Three sub-samples
98	dsti	0.660	0.115	0.563	0.115	0.211
99	em	0.660	0.546	0.504	0.411	0.056
100	pchcurrat	0.654	0.000	0.593	0.182	0.248
101	ms	0.651	0.437	0.557	0.512	0.050
102	invest	0.641	0.557	0.000	0.607	0.275
103	pchcapx_ia	0.630	0.000	0.589	0.437	0.250
104	os	0.626	0.437	0.469	0.469	0.015
105	chtx	0.623	0.364	0.504	0.266	0.098
106	dpia	0.623	0.546	0.000	0.607	0.273
107	cdi	0.623	0.546	0.552	0.266	0.133
108	pps	0.611	0.512	0.344	0.496	0.076
109	roaq	0.607	0.425	0.533	0.115	0.177
110	rs	0.584	0.411	0.344	0.381	0.027
111	rsup	0.579	0.519	0.611	0.115	0.215
112	chinv	0.569	0.557	0.266	0.397	0.119
113	cfp_ia	0.563	0.266	0.569	0.425	0.123
114	ta	0.563	0.344	0.533	0.000	0.221
115	cashdebt	0.557	0.344	0.533	0.344	0.089
116	ndf	0.557	0.230	0.678	0.000	0.281
117	grcapx	0.552	0.411	0.266	0.448	0.078
118	STR	0.546	0.469	0.230	0.533	0.131
119	pchcapx	0.546	0.322	0.448	0.230	0.090
120	pchquick	0.539	0.000	0.569	0.000	0.268
121	grltnoa	0.539	0.425	0.182	0.593	0.169
122	pchsaleinv	0.519	0.182	0.519	0.000	0.215
123	dncl	0.519	0.115	0.546	0.563	0.207
124	ivg	0.504	0.569	0.115	0.297	0.186
125	mom6m	0.496	0.182	0.479	0.230	0.130
126	chempia	0.496	0.437	0.437	0.364	0.034
127	LIQ_PS	0.496	0.115	0.397	0.552	0.181
128	mom36m	0.479	0.397	0.574	0.381	0.087
129	std_dolvol	0.459	0.381	0.297	0.437	0.058
130	pchsale_pchinvt	0.448	0.266	0.496	0.000	0.203
131	pchsale_pchxsga	0.448	0.230	0.469	0.000	0.192
132	dwc	0.448	0.479	0.437	0.182	0.131
133	dfnl	0.437	0.344	0.589	0.230	0.150
134	chmom	0.437	0.487	0.266	0.322	0.094
135	pchsale_pchrect	0.425	0.115	0.519	0.000	0.223
136	sue	0.397	0.230	0.182	0.230	0.022
137	LTR	0.381	0.182	0.297	0.364	0.075
138	pchgm_pchsale	0.322	0.000	0.381	0.266	0.160
139	lfe	0.297	0.230	0.322	0.000	0.135
140	ill	0.297	0.182	0.230	0.322	0.058
141	dnoa	0.182	0.000	0.182	0.115	0.075
142	ear	0.182	0.115	0.182	0.000	0.075
143	chatoia	0.182	0.411	0.182	0.230	0.099
144	chpmia	0.182	0.000	0.230	0.182	0.099
145	etr	0.115	0.230	0.115	0.297	0.075

Notes: This table presents the estimated factor strength, using the decomposed thirty years data. The thirty year data set is decomposed into three subsets: January 1988 to December 1997, January 1998 to December 2007, and January 2008 to December 2017. For each data set, it contains 120 observations ($t = 120$), and 242 units ($n = 242$) The table also contains the full sample estimation results of factor strength, and the standard deviation among the three sub samples results. The table is ordered decreasingly base on the full sample factor strength.

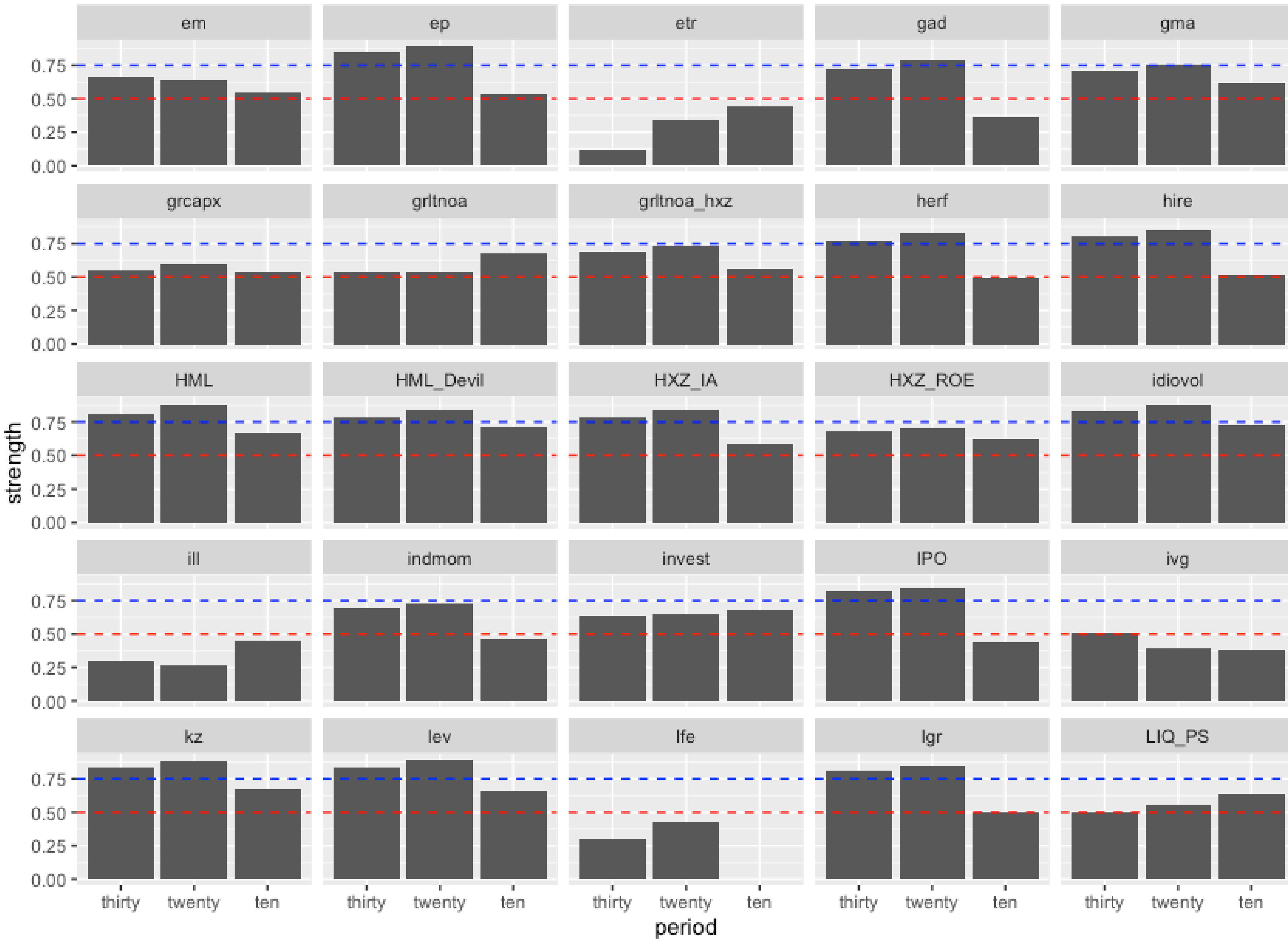
C Strength Comparison Figures

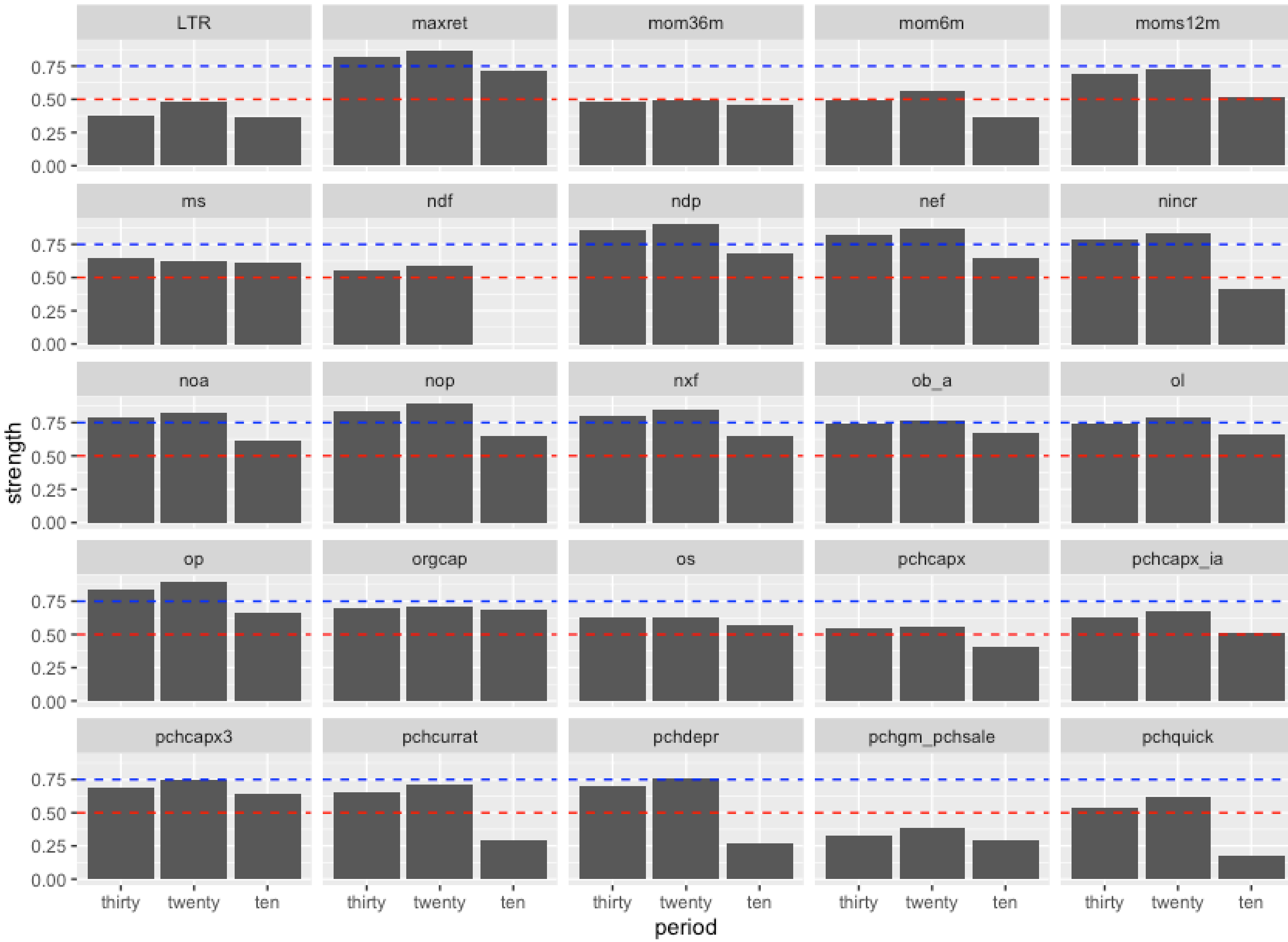
Figure 1: Strength Comparison

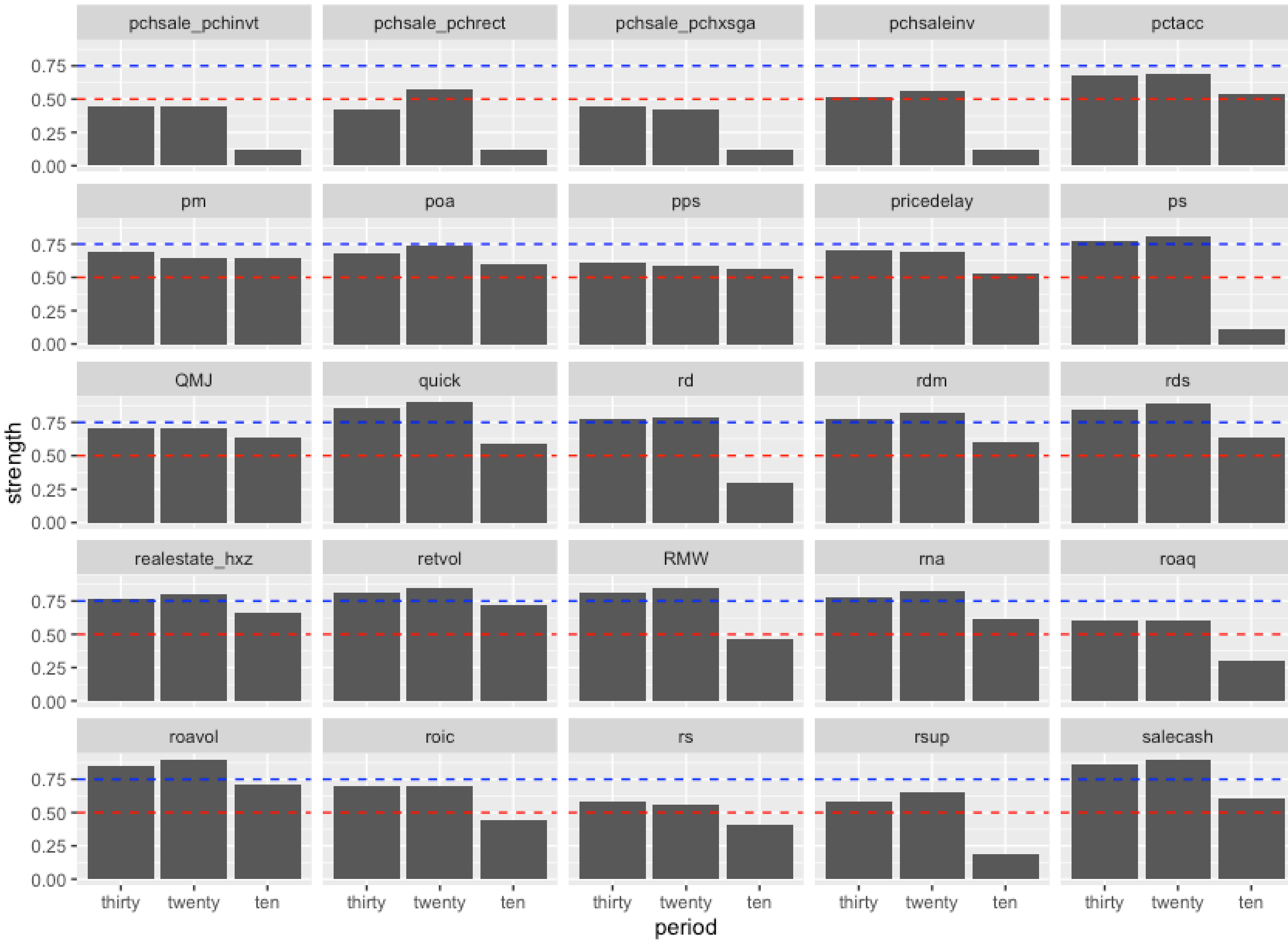


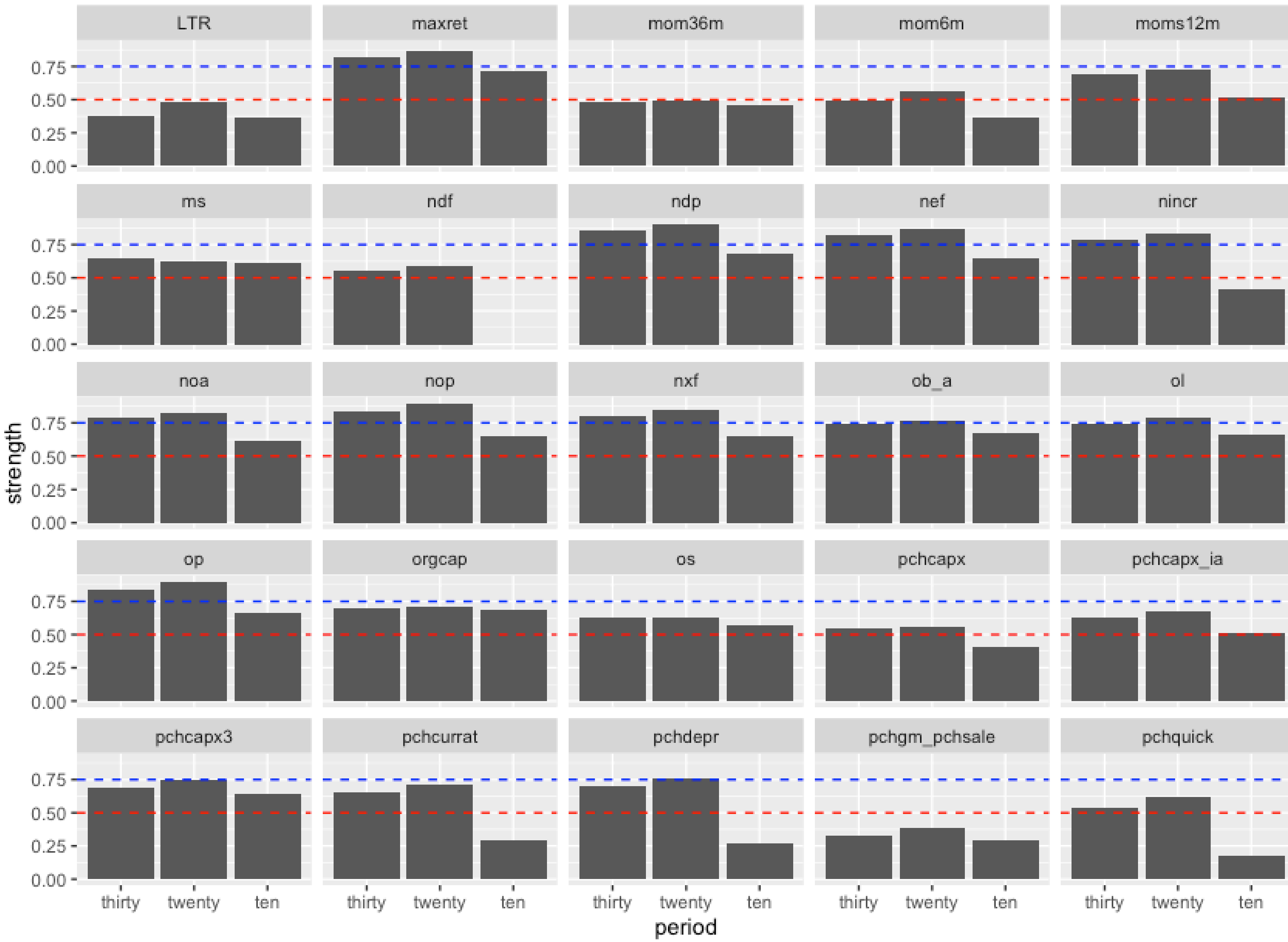


cinvest
cinvest_a
CMA
convind
cp
cto
currat
dcoa
dcol
depr
dfin
dfnl
dnca
dncl
dnco
dnoa
dolvol
dpia
dsti
dwc
dy
ear
ebp
egr
egr_hxz
strength
0.00
0.25
0.50
0.75
thirty
twenty
ten
period


em
ep
etr
gad
gma
grcapx
grltnoa
grltnoa_hxz
herf
hire
HML
HML_Devil
HXZ_IA
HXZ_ROE
idiovol
ill
indmom
invest
IPO
ivg
kz
lev
lfe
lgr
LIQ_PS
0.75
0.50
0.25
0.00
strength
thirty
twenty
ten
period

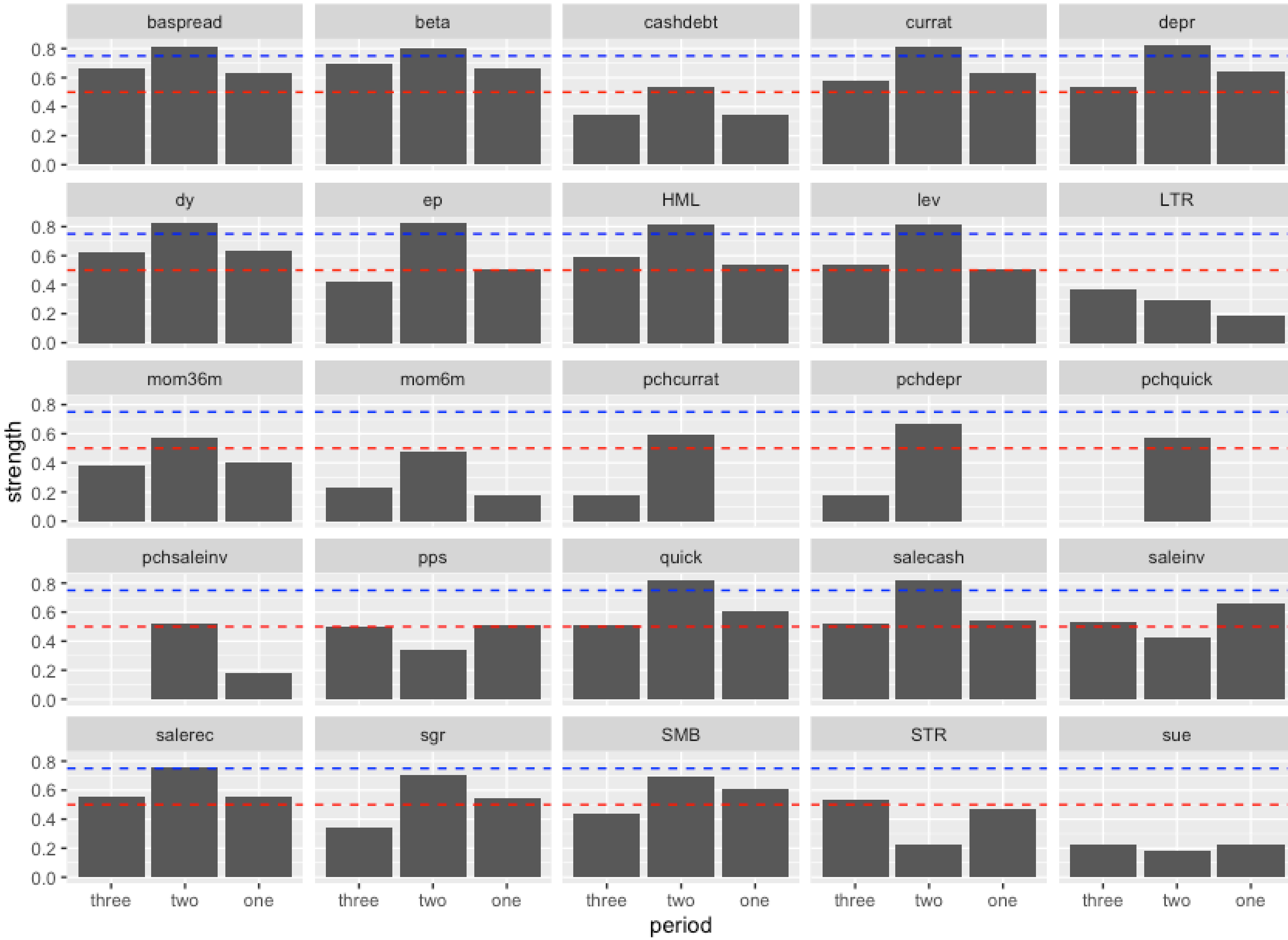

LTR
maxret
mom36m
mom6m
moms12m
ms
ndf
ndp
nef
nincr
noa
nop
nxf
ob_a
ol
op
orgcap
os
pchcapx
pchcapx_ia
pchcapx3
pchcurrat
pchdepr
pchgm_pchsale
pchquick
strength
0.00
0.25
0.50
0.75
thirty
twenty
ten
period

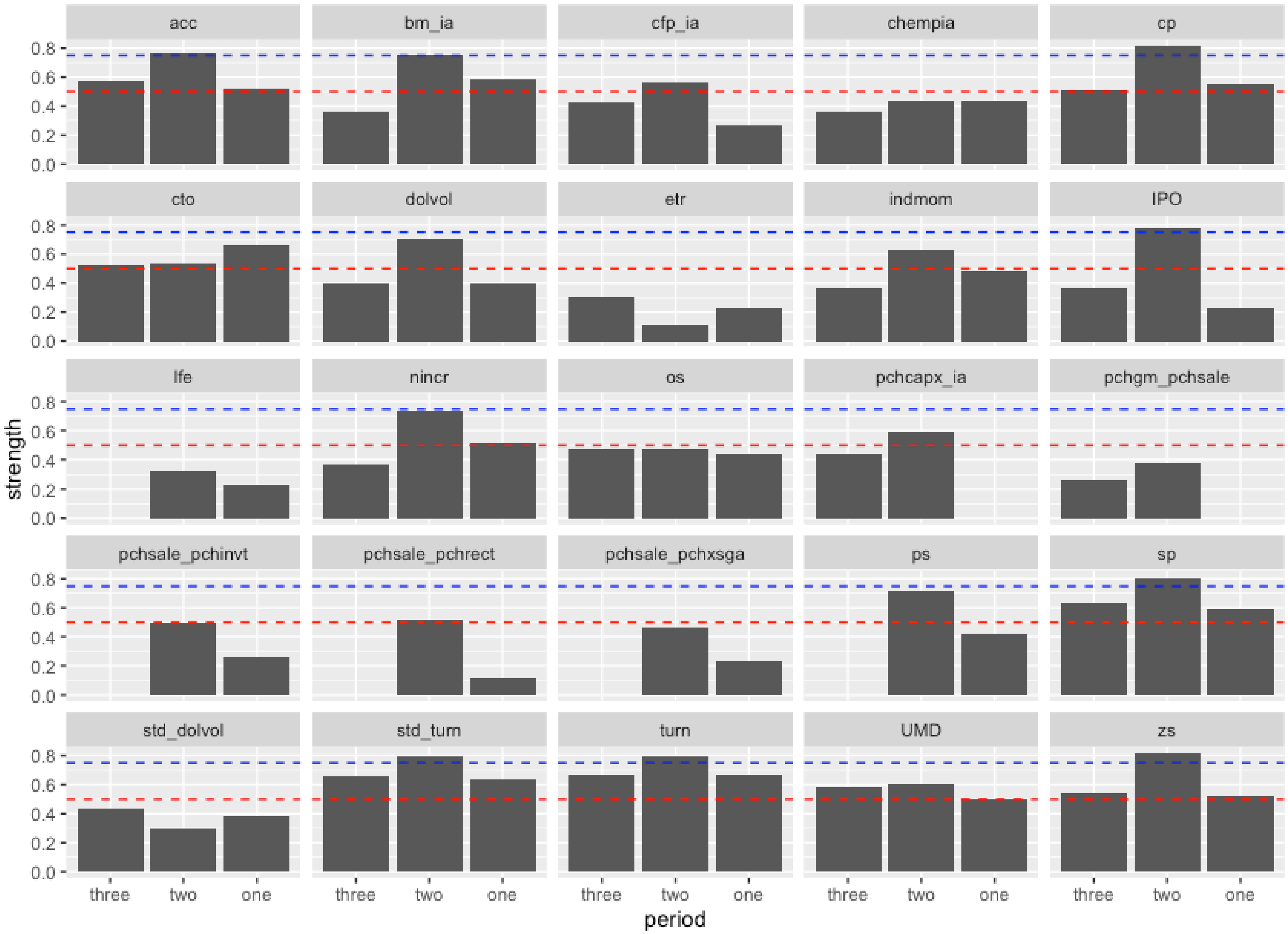

pchsale_pchinvt
pchsale_pchrect
pchsale_pchxsga
pchsaleinv
pctacc
pm
poa
pps
pricedelay
ps
QMJ
quick
rd
rdm
rds
realestate_hxz
retvol
RMW
rna
roaq
roavol
roic
rs
rsup
salecash
0.00
0.25
0.50
0.75
strength
thirty
twenty
ten
period

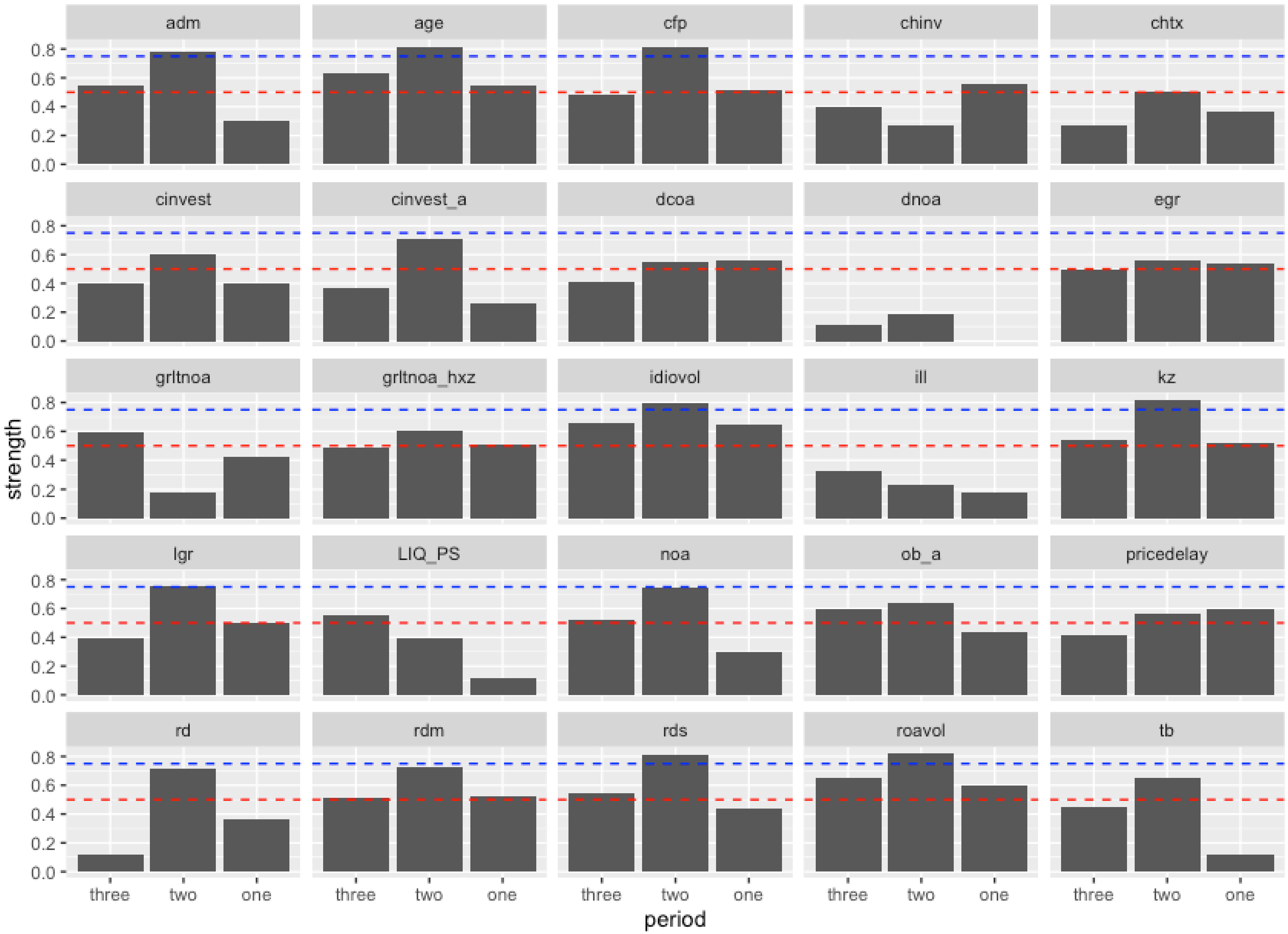


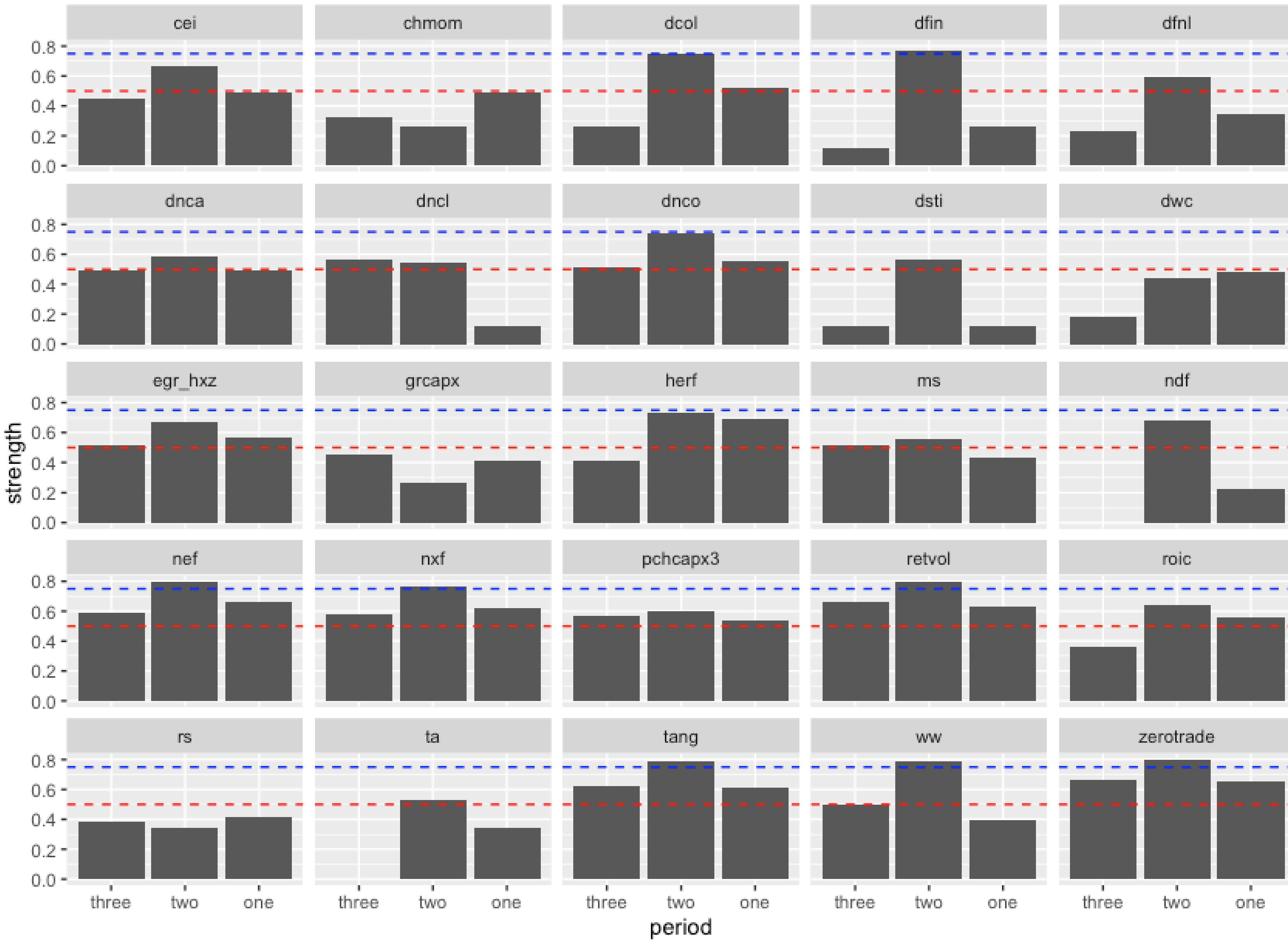
Notes: The figure compare the strength of every factor's strength in different data set. The x-axis indicates the data set: thirty is thirty years data set (January 1987 to December 2017), twenty is twenty year data set (January 1997 to December 2017), and ten is ten year data set (January 2007 to December 2017). The red dash line and blue dash line represent 0.5 and 0.75 threshold value respectively.

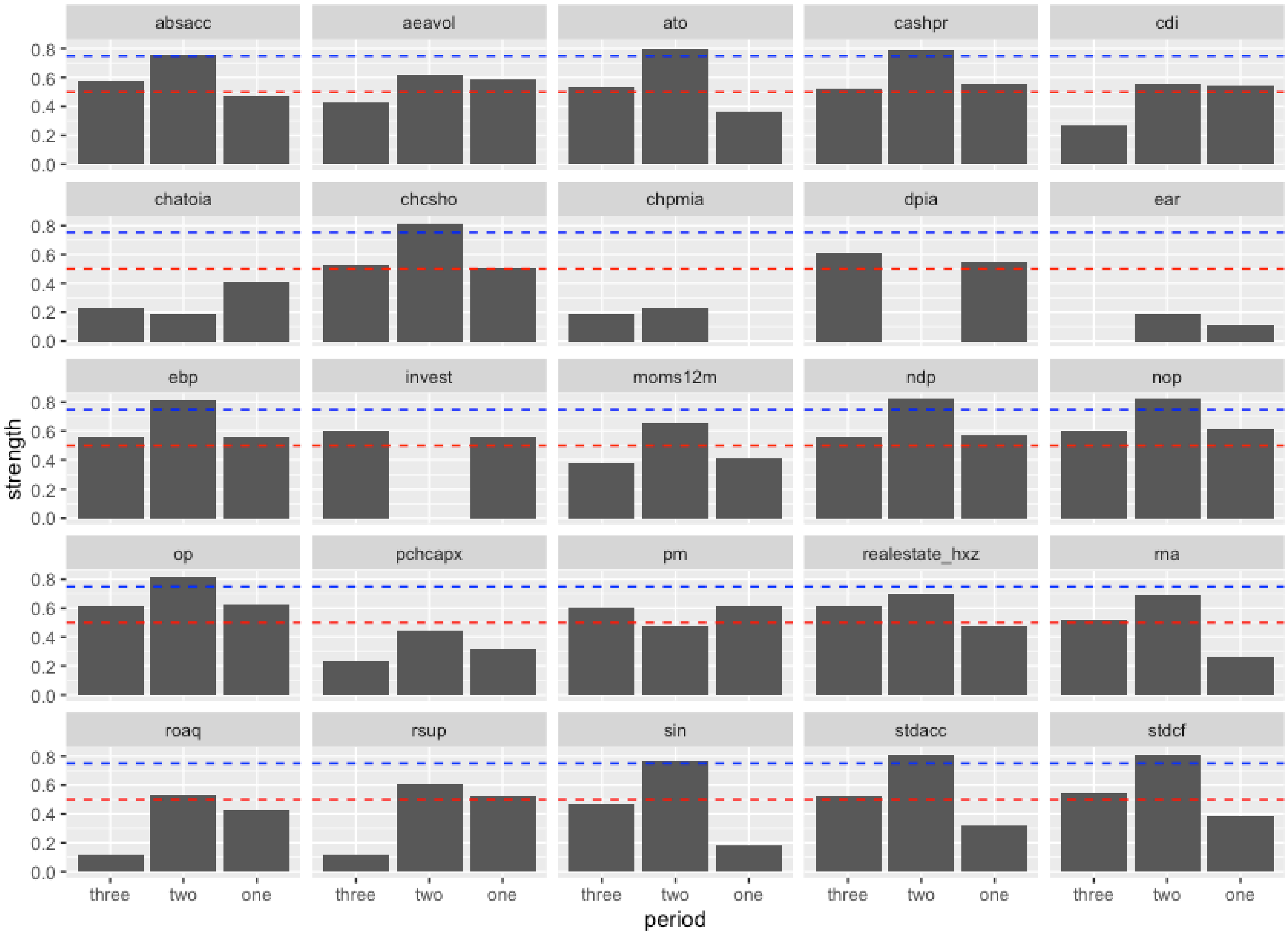
Figure 2: Thirty Year Decompose Comparison

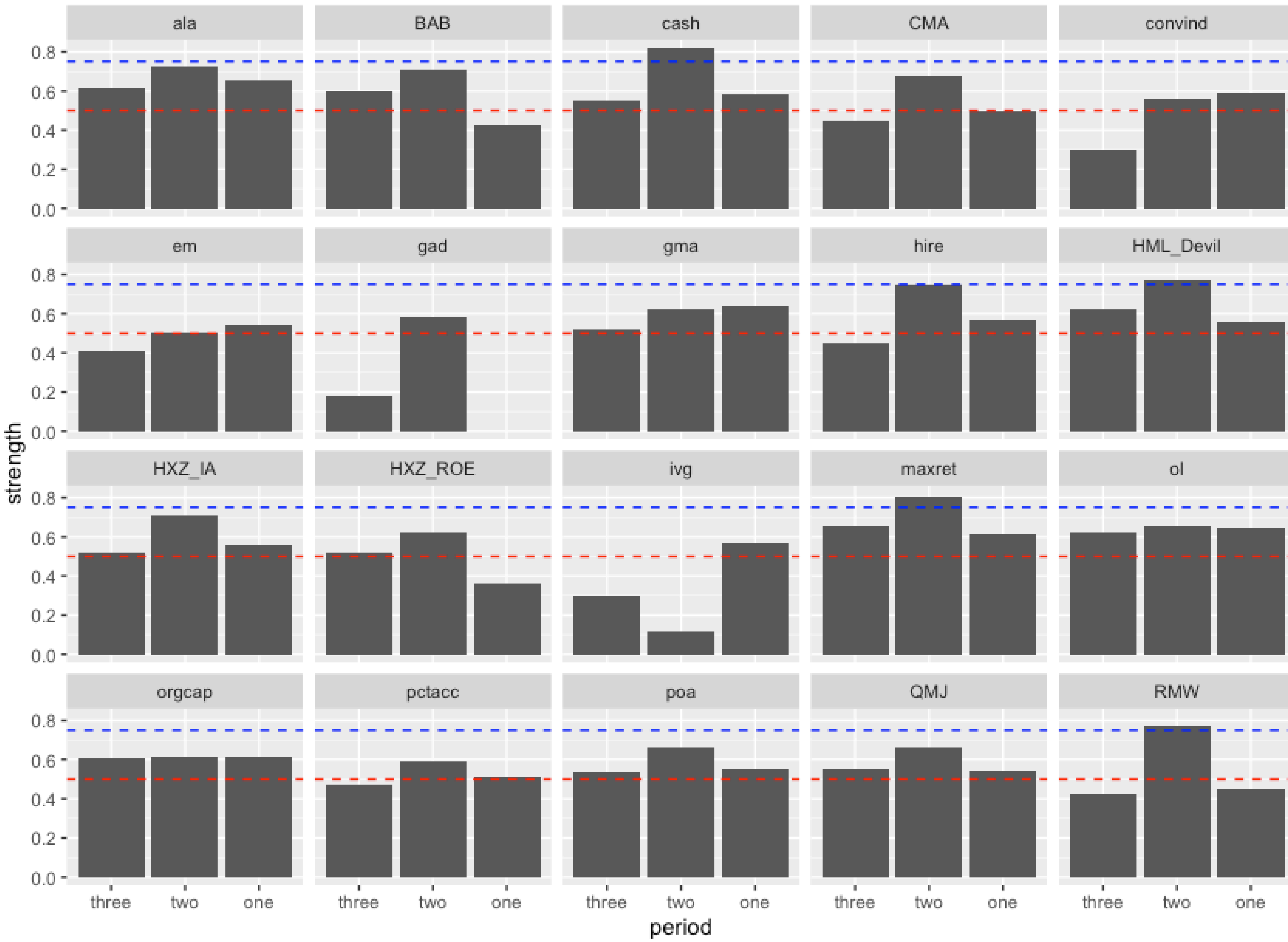



acc
bm_ia
cfp_ia
chempia
cp
cto
dolvol
etr
indmom
IPO
lfe
nincr
os
pchcapx_ia
pchgm_pchsale
pchsale_pchinvt
pchsale_pchrect
pchsale_pchxsga
ps
sp
std_dolvol
std_turn
turn
UMD
zs
strength
0.0
0.2
0.4
0.6
0.8
three
two
one
period


adm
age
cfp
chinv
chtx
cinvest
cinvest_a
dcoa
dnoa
egr
grltnoa
grltnoa_hxz
idiovol
ill
kz
lgr
LIQ_PS
noa
ob_a
pricedelay
rd
rdm
rds
roavol
tb
0.0
0.2
0.4
0.6
0.8
strength
three
two
one
period


cei
chmom
dcol
dfin
dfnl
dnca
dncl
dnco
dsti
dwc
egr_hxz
grcapx
herf
ms
ndf
nef
nxf
pchcapx3
retvol
roic
rs
ta
tang
ww
zerotrade
0.0
0.2
0.4
0.6
0.8
strength
three
two
one
period


absacc
aeavol
ato
cashpr
cdi
chatoia
chcsho
chpmia
dpia
ear
ebp
invest
moms12m
ndp
nop
op
pchcapx
pm
realestate_hxz
rna
roaq
rsup
sin
stdacc
stdcf
strength
0.0
0.2
0.4
0.6
0.8
three
two
one
period



Notes: The figure compare the strength of factor using subsample from the thirty year data.. The x-axis indicates the subsample data set: three is third decade (January 2007 to December 2017), two is second decade (January 1997 to December 2007), and one is the first decade (January 1987 to December 1997). The red dash line and blue dash line represent 0.5 and 0.75 threshold value respectively.