Understanding Systematic Risk - A High-Frequency Approach

Markus Pelger*

January 25, 2019

Abstract

Based on a novel high-frequency data set for a large number of firms, I estimate the time-varying latent continuous and jump factors that explain individual stock returns. The factors are estimated with a principal component analysis applied to a local volatility and jump covariance matrix. I find four stable continuous systematic factors, which can be approximated very well by a market, oil, finance and electricity portfolio, while there is only one stable jump market factor. The risk exposure to these factors varies substantially over time. The four continuous factors carry an intraday risk premium that reverses overnight.

Keywords: Systematic Risk, High-dimensional Data, High-Frequency Data, Latent Factors, PCA, Jumps, Cross-Section of Returns, Time-Varying Risk, Industry Factors

JEL classification: C14, C52, C58, G12

*Department of Management Science & Engineering, Stanford University, Stanford, CA 94305, Email: mpelger@stanford.edu. I thank Jason Zhu for excellent research assistance. I particularly thank Robert M. Anderson, Martin Lettau, Michael Jansson and Lisa Goldberg. I am very grateful for comments and helpful discussions with Yacine Aït-Sahalia, Torben Andersen, Svetlana Bryzgalova, Mikhail Chernov, John Cochrane, Frank Diebold, Darrell Duffie, Noureddine El Karoui, Steve Evans, Jianqing Fan, Kay Giesecke, Valentin Haddad, Ulrike Malmendier, Olivier Scaillet, Ken Singleton, George Tauchen, Viktor Todorov, Neil Shephard, Dacheng Xiu, and the audience participants at UC Berkeley, Stanford, University of Pennsylvania, University of Bonn and the SoFiE, Informs, FERM, Econometric society and NBER Time-Series meetings. This work was supported by the Center for Risk Management Research at UC Berkeley.

1 Introduction

One of the most popular methods for modeling and estimating systematic risk are factor models. Finding the "right" systematic factors has become the central question of asset pricing. This paper enhances our understanding about systematic risk by answering the following questions: (1) What are the factors that explain the systematic co-momvent in individual stocks? (2) How does the systematic factor structure for stocks change over time? (3) What are the asset pricing implications of the systematic factors?

This paper considers three key elements concurrently: First, I do not use a pre-specified (and potentially miss-specified) set of factors. Instead I estimate the statistical factors, which can explain most of the common comovement in a large cross-section of stock returns. Second, using high-frequency data allows me to study the time-variation in the factor structure under minimal assumptions as I can analyze very short time horizons independently. Allowing for the time-variation in the factor structure is crucial as individual stocks do not have a constant risk exposure in contrast to some characteristic sorted portfolios.[1] Third, by separating the high frequency returns into continuous intraday returns and intraday and overnight jumps I can decompose the systematic risk structure into its smooth and rough component, which have different asset pricing implications.

The statistical theory underlying my estimations is very general and developed in Pelger (2019). It combines the two fields of high-frequency econometrics and large-dimensional factor analysis. Under the assumption of an approximate factor model it estimates an unknown factor structure for general continuous-time processes based on high-frequency data. Using a truncation approach, I can separate the continuous and jump components of the price processes, which I use to construct a "jump covariance" and a "continuous risk covariance" matrix. The latent continuous and jump factors can be separately estimated by principal component analysis.

My empirical investigation is based on a novel high-frequency data set of 5-minutes returns of the stocks in the WRDS TAQ Millisecond Trades database for the time period 2004 to 2016. My findings are sixfold. First, I find four high-frequency factors. There is time-variation in the exposure to these risk factors and the amount of variation that they explain, e.g. one factor is only systematic during the financial crisis. Surprisingly, the portfolio weights to construct the statistical factors are stable over time. Second, these four factors have an economically meaningful interpretation and can be closely approximated by a market, oil, finance and electricity factor. The size, value and momentum factors of the

[1] See Lettau and Pelger (2018)

Fama-French Carhart model cannot span the statistical factors. Third, the factor structure for smooth continuous movements is different from rough jump movements. There seems to be only one intraday jump market factor. The continuous factors have the same composition as the high-frequency factors including the jumps. Fourth, my high-frequency factors carry an intraday risk premium that reverses overnight. Decomposing returns into their intraday and overnight component, I show a strong reversal pattern in individual stock returns, which is captured by the high-frequency factors. Fifth, the high-frequency factors explain the expected return structure of industry portfolios while the Fama-French-Carhart factors can better explain size- and value sorted portfolios. This finding suggests that time-varying factors that explain the co-movement in individual stocks do indeed have cross-sectional pricing information, but are not necessarily related to characteristics.

My paper contributes to the central question in empirical and theoretical asset pricing what constitutes systematic risk. There are essentially three common ways of selecting which factors and how many describe the systematic risk. The first approach is based on theory and economic intuition. The capital asset pricing model (CAPM) of Sharpe (1964) with the market as the only common factor falls into this category. The second approach bases factors on firm characteristics with the three-factor model of Fama and French (1993) as its most famous example. My approach falls into the third category where factor selection is statistical. This approach is motivated by the arbitrage pricing theory (APT) of Ross (1976). Factor analysis can be used to analyze the covariance structure of returns. This approach yields estimates of factor exposures as well as returns to underlying factors, which are linear combinations of returns on underlying assets. The notion of an "approximate factor model" was introduced by Chamberlain and Rothschild (1983), which allowed for a non-diagonal covariance matrix of the idiosyncratic component. Connor and Korajczyk (1988, 1993) study the use of principal component analysis in the case of an unknown covariance matrix, which has to be estimated.[2]

One distinctive feature of the factor literature described above is that they work with a constant factor model that does not allow for time-variation. The object of interest are sorted portfolios based on previously established knowledge about the empirical behavior of average returns. Lettau and Pelger (2018) show that the characteristic-sorted portfolios are well described by a constant factor model. However, a shortcoming of the characteristic sorted approach is that the results depend on the choice of conditioning variables to generate the portfolios (Nagel (2013)).

[2]The general case of a static large dimensional factor model is treated in Bai (2003) and Bai and Ng (2002). Fan et al. (2013) study an approximate factor structure with sparsity.

My study works directly on a large cross-section of individual stocks. However, Lettau and Pelger (2018) show that a constant loading model is not appropriate to model individual stock returns. The dominant approach is to model time-variation in the risk exposure through time-varying characteristics. Kelly et al. (2017) and Fan et al. (2016b) follow this approach. Fundamentally, these approaches apply a version of PCA to characteristic-managed portfolios. Thus, their results depends on the choice of characteristics and the basis functions to model the functional relationship between characteristics and loadings. I propose an alternative to account for the time-variation which is completely general and does not depend on the choice of basis functions. Therefore, I can find the factors that can actually explain most of the variation in individual stocks without using any (potentially wrong) prior about the characteristics which drive the variation. Interestingly, I show that factors that explain the variation in individual stocks are not related to popular characteristic-based factors such as the Fama-French Carhart factors.[3]

I combine tools from high-frequency econometrics with a large-dimensional panel data set. The apparent advantage of high-frequency observations is to estimate a time-varying factor model without any prior assumption about the time-variation. Furthermore, the significantly larger amount of time-series observations also allows for a more precise estimation of the risk structure. So far most of the empirical literature that utilizes the tools of high-frequency econometrics to analyze a factor structure is limited to a pre-specified set of factors. For example, Bollerslev et al. (2016) estimate the betas for a continuous and jump market factor. Fan et al. (2016a) estimate a large-dimensional covariance matrix with high-frequency data for a given factor structure. My work goes further as I estimate the unknown continuous and jump factor structure in a large cross-section. My method uses a purely statistical criterion to derive factors, and has the advantage of requiring no ex ante knowledge of the structure of returns.[4] In addition, the high-frequency data allows me to study separately smooth continuous factor risk and rough intraday and overnight jump risk, which cannot be estimated from observing only daily returns. Empirical evidence suggests that for a given factor structure, systemic risk associated with discontinuous price movements is different

[3]For a pre-specified set of factors studies have already shown that time-varying systematic risk factors capture the data better. The idea of time-varying systematic risk factors contains the conditional version of the CAPM as a special case, which seems to explain systematic risk significantly better than its constant unconditional version. Contributions to this literature include for example Jagannathan and Wang (1996) and Lettau and Ludvigson (2001). Bali et al. (2014) have also shown that GARCH-based time-varying conditional betas help explain the cross-sectional variation in expected stock returns.

[4]Aït-Sahalia and Xiu (2017a) apply nonparametric principal component analysis to a low-dimensional cross-section of high-frequency data. Aït-Sahalia and Xiu (2017b) also estimate a large-dimensional factor model based on continuous high-frequency observations. Their studies focus on estimating the continuous covariance matrix, while my work tries to explain the factor structure itself.

from continuous systematic risk.[5] I confirm and extend these results to a latent factor structure.

I show different risk-return patterns during trading and non-trading hours. While Branch and Ma (2012), Cliff et al. (2008) and Berkman et al. (47) also show distinctly different return patterns during trading and non-trading hours, I link these patterns to the underlying factor structure. The risk premium for characteristic based factors like size and value is mainly earned overnight. My high-frequency factors are only based on intraday information and hence are expected to capture the intraday risk premium. Indeed, these factors have a positive intraday risk premium which reverses overnight. Therefore the daily risk premium of the high-frequency factors is lower in absolute value than in its intraday and overnight component. The reversal pattern in factor risk is in line with anecdotal evidence, that traders want to close certain positions before the end of the day in order to avoid risk exposure during non-trading hours.

My paper contributes to an emerging literature that uses new econometric techniques in asset pricing for high-dimensional data. Lettau and Pelger (2018) generalize PCA by including a penalty on the pricing error in expected returns. Kozak et al. (2017) apply mean-variance optimization with an elastic net to PCA based factors. The 3-pass model in Giglio and Xiu (2017) corrects for missing factors by including PCA based factors. All these approaches have in common that they use a constant loading model. Kelly et al. (2017) and Fan et al. (2016b) apply a version of PCA to projected portfolio data. All these paper argue that the stochastic discount factor based on some version of dominant principal components can explain the cross-section of expected returns well.

The rest of the paper is organized as follows. Section 2 introduces the factor model setup. In Section 3 I explain the estimation method. Section 4 reports the empirical findings and section 5 concludes. The supplementary appendix contains additional empirical results.

2 Methodology

2.1 Factor Model

This paper assumes that log asset prices can be modeled by a local approximate factor model. Hence most co-movements in asset prices are due to a systematic factor component. In more detail I have N assets with log prices denoted by $X_i(t)$. I assume that the N-dimensional

[5] Empirical studies supporting this hypothesis include Bollerslev et al. (2016), Pan (2002), Eraker et al. (2003), Eraker (2004), Bollerslev and Todorov (2011) and Gabaix (2012).

log price process $X(t)$ can be explained by a local factor model, i.e.

$$X_i(t) = \Lambda_{i,s}^\top F(t) + e_i(t) \qquad i = 1, ..., N,\ t \in [T_s, T_{s+1}] \text{ and } s = 1, ..., S$$

where $\Lambda_{i,s}$ is a $K \times 1$ dimensional vector and $F(t)$ is a K-dimensional stochastic process. The loadings $\Lambda_{i,s}$ describe the exposure to the systematic factors F, while the residuals e_i are stochastic processes that describe the idiosyncratic component. The assumption of constant loadings Λ_s only needs to hold locally, i.e. for the time interval $[T_s, T_{s+1}]$. In my empirical analysis I study local time windows ranging from one week (five trading days), one month, three months, one year to 13 years. The loadings, factors and residual structure is allowed to be completely different for different local time windows. This model is very general and includes most popular factor models as a special case, e.g. time-varying factor models where the time-variation in loadings is due to monthly varying characteristic variables as in Kelly et al. (2017).

As my analysis is applied only locally and in order to simplify notation, I will drop the subscript s and formulate the model and results in terms of the locally constant model:

$$X_i(t) = \Lambda_i^\top F(t) + e_i(t) \qquad i = 1, ..., N \text{ and } t \in [0, T].$$

For the fixed (and potentially short) time interval $[0, T]$ I only observe the stochastic process X at discrete time observations $t_0 = 0, t_1 = \Delta_M, t_2 = 2\Delta_M, ..., t_M = M\Delta_M$, where the time increment is defined as $\Delta_M = t_{j+1} - t_j = \frac{T}{M}$. Hence, I observe the log returns

$$X(t_j) = \Lambda F(t_j) + e(t_j) \qquad j = 1, ..., M.$$

with $\Lambda = (\Lambda_1, ..., \Lambda_N)^\top$ and $X(t) = (X_1(t), ..., X_N(t))^\top$. In my setup the number of cross-sectional observations N and the number of high-frequency observations M is large, while the time horizon T is short and the number of systematic factors K is fixed. The loadings Λ, factors F, residuals e and number of factors K are unknown and have to be estimated.

The asset price dynamics are completely non-parametric and very general. The log prices are modeled as Itô-semimartingales, which is the most general class of stochastic processes, for which the general results of high-frequency econometrics are available.[6]:

$$X(t) = X(0) + \int_0^t \mu_s^X ds + \int_0^t \sigma_s^X dW_s^X + \int_{\mathbb{R}} x\nu^X(ds, dx).$$

[6]Locally bounded special Itô semimartingales are specified in Pelger (2019)

The process consists of a predictable drift term, a continuous martingale with N-dimensional Brownian motion W_t^X and volatility process σ_t and a jump martingale described by a compensated jump counting measure accounting for the discontinuous price movements. These particular semimartingales are standard in high-frequency econometrics, see e.g. Aït-Sahalia and Jacod (2014). They allow for correlation between the volatility and asset price processes. The model includes many well-known continuous-time models as special cases: for example stochastic volatility models like the CIR or Heston model, the affine class of models in Duffie et al. (2000), the Ornstein-Uhlenbeck stochastic volatility model with jumps of Barndorff-Nielsen and Shephard (2002) or a stochastic volatility model with log-normal jumps generated by a non-homogenous Poisson process as in Andersen et al. (2002).

The factors in an approximate factor model are systematic in the sense that they explain most of the co-movement in the data. A large fraction of the correlation between stock returns is explained by exposure to a small number of factors. As I will discuss in Section 2.2 this correlation structure has direct implications for asset pricing. Under essentially the same model assumptions the pricing kernel has to be spanned by the systematic factors and the differences in risk premia of assets should be explained by differences in factor exposure.

My approach requires only very weak assumptions which are stated formally in Pelger (2019). First, the dependence between the assets is modeled by an approximate factor structure similar to Chamberlain and Rothschild (1983). The idiosyncratic risk can be serially correlated and weakly cross-sectionally correlated and hence allows for a very general specification. The main identification criterion for the systematic risk is that the quadratic covariation matrix of the idiosyncratic risk has bounded eigenvalues, while the quadratic covariation matrix of the systematic factor part has unbounded eigenvalues. For this reason the principal component analysis can relate the eigenvectors of the exploding eigenvalues to the loadings of the factors. Second, in order to separate continuous systematic risk from jump risk, I allow only finite activity jumps, i.e. there are only finitely many jumps in the asset price process of each stock.[7] This still allows for a very rich class of models and for example general compound poisson processes with stochastic intensity rates can be accommodated. Last but not least, I work under the simultaneous limit of a growing number of high-frequency and cross-sectional observations. I do not restrict the path of how these two parameters go to infinity. The high number of cross-sectional observations makes the large dimensional covariance analysis challenging, but under the assumption of a general approximate factor structure the "curse of dimensionality" turns into a "blessing" as it becomes necessary for

[7]Many of my results work without this restriction and it is only needed for the separation of these two components.

estimating the systematic factors.

I am also interested in estimating the continuous, jump and overnight component of the factor return. I can separate the factors into continuous factors that have only a continuous martingale and predictable finite variation part and into jump factors consisting of a jump martingale and predicable finite variation term but no continuous martingale. The continuous and jump component constitute the intraday return. The overnight component is the difference between the logarithm of the closing and opening price which I will model as in Bollerslev et al. (2016) as an overnight jump. Without loss of generality I can formulate the model as

$$X(t) = \Lambda^C F^C(t) + \Lambda^D F^D(t) + \Lambda^N F^N(t) + e(t),$$

where Λ^C, Λ^D and Λ^N denote the continuous, jump and overnight loadings, while $F^C(t)$, $F^D(t)$ and $F^N(t)$ are the corresponding components of the factors. The model allows the loadings to jumps and overnight movements to be different from the continuous loadings. The model also allows the factors themselves to be different. For example, if a certain risk factor only contributes to continuous co-movements but not to jumps, I set the jump loadings for this factor to zero.

A special case is the CAPM model of Bollerslev et al. (2016) which allows for different continuous, jump and overnight betas for the market factor. Under the assumption of this one factor model I would estimate one factor when studying separately the continuous returns, the intraday jumps or the overnight jumps. If the market betas were different for continuous, jump and overnight returns, then I would expect to estimate a three factor model for the combined returns with a pure continuous, an intraday jump and an overnight jump market factor. If the market betas were the same then the combined returns are described by one market factor with continuous and jump movements.

2.2 Asset Pricing Implications

Under minimal no-arbitrage assumptions on the approximate factor model (Chamberlain (1988) and Reisman (1992)), there exist an economy-wide pricing kernel which is spanned by the risk factors. In more detail, the drift term of the stock excess returns corresponding

to the risk premium equals the exposure to risk times the price of risk:[8]

$$E_t[dX_t - r_t dt] = \mu_t^X dt - r_t dt = \Lambda E_t[dF_t - r_t dt] = \Lambda \left(\mu_t^F dt - r_t dt\right),$$

where r_t denotes the instantaneous risk-free interest rate and μ_t^F the drift term of the factors.[9] The stochastic discount factor (SDF) is spanned by the systematic risk factors. The potentially different exposure to continuous, jump and overnight risk is reflected by different weights for these components in the SDF (Bollerslev et al. (2016) and Duffie et al. (2000)).

As shown in Back (1991) the risk premium can be decomposed into its continuous, jump and overnight component.[10] However, using only return data I cannot in general separately estimate the continuous risk premium, the jump risk premium and the overnight risk premium. If the span of the loadings Λ^C, Λ^D and Λ^D were different, I could separate the different risk premia by a cross-sectional regression on expected excess returns. However, I will show empirically that Λ^D is spanned by Λ^C. This finding rules out an estimation of the jump risk premium with a two stage Fama and MacBeth (1973) type regression as proposed in Alexeev et al. (2017). Instead I focus on the intraday and overnight component of returns.

Separating returns into their intraday and overnight component, I study how systematic risk affects the expected intraday and overnight excess returns. The daily returns are composed of the intraday and overnight price movements:

$$X = X^{intra} + X^{night} = \Lambda(F^{intra} + F^{night}) + e^{intra} + e^{night}$$

By the no-arbitrage condition the risk-premium can also be separated into an intraday and overnight component:

$$\mu_t^{X,intra} - r_t^{intra} = \Lambda \left(\mu_t^{F,intra} - r_t^{intra}\right), \quad \mu_t^{X,night} - r_t^{night} = \Lambda \left(\mu_t^{F,night} - r_t^{night}\right)$$

where $\mu_t^{F,intra}$ and $\mu_t^{F,intra}$ are the intraday respectively overnight drift terms of the factors

[8]The approximate no-arbitrage condition in Chamberlain (1988) and Reisman (1992) results in only an approximate pricing statement, i.e. only the risk premia of well-diversified portfolios is explained by the systematic factor risk. Connor (1984) provides conditions under which the approximate pricing equation also holds exactly for non-well-diversified portfolios.

[9]The factors can be represented as $dF_t = d\mu_t^F + dM_t^F$ where $dM_t = \sigma^F dW_t^F + \int_{\mathbb{R}} x\nu^F(dt, dx)$ is a local martingale consisting of a Brownian motion W_t^F and a compensated jump martingale with compensated jump measure ν^F (see Pelger (2019) for the technical details). In this formulation the drift term minus the risk free rate correspond to the risk premium for continuous and jump risk. Note that all factors under consideration are traded portfolios of the underlying assets.

[10]Back (1991) proves the decomposition into a continuous and jump component but the same arguments can be applied to obtain the overnight component as well.

and r_t^{intra} and r_t^{night} are the intraday and overnight risk-free interest rates. The time-series average of X^{intra} estimates the intraday expected return $\int_0^T \mu_t^{intra} dt$ and similarly for the overnight part. As factors estimated from high-frequency data only incorporate intraday price information, I want to study if the average intraday return differs from the average overnight return.

The stochastic discount factor (SDF) implies a tangency portfolio based only on the factors that has the maximal conditional Sharpe-ratio. Based on the different set of factors in my analysis, the Sharpe-ratio of the corresponding tangency portfolio will serve as a measure for the pricing performance of the factors. In particular, I will also decompose the returns of the tangency portfolio into an intraday and overnight component and study when the risk premium for different factors is earned. As without further model assumptions I cannot estimate the time-varying drift term, I can only use factor means $\bar{\mu}^F - \bar{r} = \int_0^T (\mu_t^F - r_t) dt$ to construct the tangency portfolio weights $w^{SDF} = [F, F]^{-1}(\bar{\mu}^F - \bar{r})$ and study the overall return Fw^{SDF}, intraday return $F^{intra}w^{SDF}$ and overnight return $F^{night}w^{SDF}$ of the tangency portfolio. I will also analyze the weights of the tangency portfolio based only on a specific component of the factors returns, e.g. optimal overnight weights $w^{SDF,night} = [F^{night}, F^{night}]^{-1}(\bar{\mu}^{F,night} - \bar{r}^{night})$.

3 Estimation

3.1 Factor Estimation

I employ the estimation technique developed in Pelger (2019) which is essentially Principal Component Analysis applied to a spot volatility and jump covariance matrix. There are M observations of the N-dimensional stochastic process X in the time interval $[0, T]$. For the time increments $\Delta_M = \frac{T}{M} = t_{j+1} - t_j$ I denote the increments of the stochastic processes by

$$X_{j,i} = X_i(t_{j+1}) - X_i(t_j) \qquad F_j = F(t_{j+1}) - F(t_j) \qquad e_{j,i} = e_i(t_{j+1}) - e_i(t_j).$$

In matrix notation I have

$$\underset{(M\times N)}{X} = \underset{(M\times K)}{F}\underset{(K\times N)}{\Lambda^\top} + \underset{(M\times N)}{e}.$$

For a given K my goal is to estimate Λ and F. As in any factor model where only X is observed, Λ and F are only identified up to invertible transformations. I impose the standard normalization that $\frac{\hat{\Lambda}^\top \hat{\Lambda}}{N} = I_K$ and that $\hat{F}^\top \hat{F}$ is a diagonal matrix.

The estimator for the loadings $\hat{\Lambda}$ is defined as the eigenvectors associated with the K largest eigenvalues of $\frac{1}{N}X^\top X$ multiplied by $\sqrt{N}$. The estimator for the factor increments is $\hat{F} = \frac{1}{N}X\hat{\Lambda}$. Note that $\frac{1}{N}X^\top X$ is an estimator for the quadratic covariation $\frac{1}{N}[X,X]$ for a finite N. The asymptotic theory is applied for $M, N \to \infty$. The systematic component of $X(t)$ is the part that is explained by the factors and defined as $C(t) = \Lambda F(t)$. The increments of the systematic component $C_{j,i} = F_j\Lambda_i^\top$ are estimated by $\hat{C}_{j,i} = \hat{F}_j\hat{\Lambda}_i^\top$.

Intuitively under some assumptions I can identify the jumps of the process $X_i(t)$ as the big movements that are larger than a specific threshold. I set the threshold identifier for jumps as $\alpha\Delta_M^{\bar{\omega}}$ for some $\alpha > 0$ and $\bar{\omega} \in \left(0, \frac{1}{2}\right)$ and define $\hat{X}^C_{j,i} = X_{j,i}\mathbb{1}_{\{|X_{j,i}| \leq \alpha\Delta_M^{\bar{\omega}}\}}$ and $\hat{X}^D_{j,i} = X_{j,i}\mathbb{1}_{\{|X_{j,i}| > \alpha\Delta_M^{\bar{\omega}}\}}$.[11] The estimators $\hat{\Lambda}^C$, $\hat{\Lambda}^D$, $\hat{F}^C$ and $\hat{F}^D$ are defined analogously to $\hat{\Lambda}$ and $\hat{F}$, but using $\hat{X}^C$ and $\hat{X}^D$ instead of X. [12] Overnight returns are modeled as separate jumps.

The quadratic covariation of the factors can be estimated by $\hat{F}^\top\hat{F}$ and the volatility component of the factors by $\hat{F}^{C\top}\hat{F}^C$. The estimated increments of the factors $\hat{F}, \hat{F}^C$ and $\hat{F}^D$ can be used to estimate the quadratic covariation with any other process, i.e. I can use them in a high-frequency regression to consistently estimate the loadings for the different components of the different high-frequency factors.

The estimated loadings $\hat{\Lambda}, \hat{\Lambda}^C$ and $\hat{\Lambda}^D$ measure the risk exposure to the factors as well as serve as portfolio weights to construct the factors. For example the portfolio weights for the continuous factors are $w^C = \frac{1}{\sqrt{N}}\hat{\Lambda}^C$. Based on the portfolio weights I will not only study the continuous returns of the continuous factors $X^C w^C$ but also their overall daily $X^{day}w^C$, intraday $X^{intra}w^C$ and overnight returns $X^{night}w^C$. In the following I refer to a continuous factor if it is constructed with the continuous portfolio weights and will specify the return component under consideration. The loadings that I estimate on a local time window coincide with the regression coefficients on the same local time window, i.e. the rescaled eigenvectors Λ^C are equal to $X^{C\top}F^C(F^{C\top}F^C)^{-1}$. As I will also study the regression coefficients of different return components of the continuous factors, I label as loadings the regression coefficients for these factors and specify the return component and time window under consideration.

[11] I set $\omega = 0.49$ (See Pelger (2019), Aït-Sahalia and Xiu (2017b) and Bollerslev et al. (2013).) ω is typically chosen between 0.47 and 0.49 and the results are insensitive to this choice. Intuitively I classify all increments as jumps that are beyond α standard deviations of a local estimator of the stochastic volatility with $\alpha = 3, 4, 4.5$ or 5.

[12] For the jump threshold I use the TOD specification of Bollerslev et al. (2013), which takes into account time-of-the-day pattern in the spot volatility estimation.

3.2 Number of Factors

In Pelger (2019) I develop a new diagnostic criterion for the number of factors, that can also distinguish between the number of continuous and jump factors. This estimator uses only the same weak assumptions that are needed for the consistency of my factor estimator. In simulations it outperforms the existing estimators while maintaining weaker assumptions. Intuitively the large eigenvalues are associated with the systematic factors and hence the problem of estimating the number of factors is roughly equivalent to deciding which eigenvalues are considered to be large with respect to the rest of the spectrum. Under the approximate factor model assumptions the first K "systematic" eigenvalues of $X^\top X$ are $O_p(N)$, while the nonsystematic eigenvalues are $O_p(1)$. A straightforward estimator for the number of factors considers the eigenvalue ratio of two successive eigenvalues and associates the number of factors with a large eigenvalue ratio. However, without very strong assumptions the small eigenvalues cannot be bounded from below, which could lead to exploding eigenvalue ratios in the nonsystematic spectrum. I propose a perturbation method to avoid this problem. As long as the eigenvalue ratios of the perturbed eigenvalues cluster around 1, we are in the nonsystematic spectrum. As soon as we do not observe this clustering any more, but a large eigenvalue ratio of the perturbed eigenvalues, we are in the systematic spectrum.

The number of factors can be consistently estimated through the perturbed eigenvalue ratio statistic and hence, I can replace the unknown number K by its estimator $\hat{K}$. Denote the ordered eigenvalues of $X^\top X$ by $\lambda_1 \geq ... \geq \lambda_N$. I choose a slowly increasing sequence $g(N,M)$ such that $\frac{g(N,M)}{N} \to 0$ and $g(N,M) \to \infty$. Based on simulations a good choice for the perturbation term g is the median eigenvalue rescaled by $\sqrt{N}$, but the results are very robust to different choices of the perturbation. [13] Then, I define perturbed eigenvalues $\hat{\lambda}_k = \lambda_k + g(N,M)$ and the perturbed eigenvalue ratio statistic

$$ER_k = \frac{\hat{\lambda}_k}{\hat{\lambda}_{k+1}} \qquad \text{for } k = 1, ..., N-1.$$

The estimator for the number of factors is defined as the first time that the perturbed

[13] I estimate the number of factors using the perturbed eigenvalue ratio estimator with $g(N,M) = \sqrt{N} \cdot median\{\lambda_1, ..., \lambda_N\}$. For robustness I also use an unperturbed eigenvalue ratio test and $g(N,M) = log(N) \cdot median\{\lambda_1, ..., \lambda_N\}$ and the Onatski (2010) eigenvalue difference estimator. The results are the same and available upon request.

eigenvalue ratio statistic does not cluster around 1 any more:

$$\hat{K}(\gamma) = \max\{k \leq N-1 : ER_k > 1+\gamma\} \qquad \text{for } \gamma > 0.$$

The definitions of $\hat{K}^C(\gamma)$ and $\hat{K}^D(\gamma)$ are analogous but using λ_i^C respectively λ_i^D of the matrices $\hat{X}^{C\top}\hat{X}^C$ and $\hat{X}^{D\top}\hat{X}^D$. The results in my empirical analysis are robust to a wide range of values for the threshold γ as illustrated by the eigenvalue ratio plots.

3.3 Comparison between Factors

One of the major problems when comparing two different sets of factors is that a factor model is only identified up to invertible linear transformations. Two sets of factors represent the same factor model if the factors span the same vector space. When trying to interpret estimated factors by comparing them with economic factors, I need a measure to describe how close two vector spaces are to each other. As proposed by Bai and Ng (2006) the generalized correlation is a natural candidate measure.[14] Intuitively, I calculate the correlation between the latent and candidate factors after rotating them appropriately. Generalized correlations close to 1 measure of how many factors two sets have in common.

Let F be my K-dimensional set of factor processes and G be a K_G-dimensional set of economic candidate factor processes. I want to test if a linear combination of the candidate factors G can replicate some or all of the true factors F. The first generalized correlation is the highest correlation that can be achieved through a linear combination of the factors F and the candidate factors G. For the second generalized correlation I first project out the subspace that spans the linear combination for the first generalized correlation and then determine the highest possible correlation that can be achieved through linear combinations of the remaining $K-1$ respectively K_G-1 dimensional subspaces. This procedure continues until I have calculated the $min(K, K_G)$ generalized correlation. If $K = K_G = 1$ it is simply the correlation as measured by the quadratic covariation. [15] If two matrices span the same vector spaces, the generalized correlations are all equal to 1. Otherwise they denote the highest possible correlations that can be achieved through linear combinations of the subspaces. If for example for $K = K_G = 3$ the generalized correlations are $\{1, 1, 0\}$, it implies that there exists a linear combination of the three factors in G that can replicate two

[14] The generalized correlation is also called canonical correlation.

[15] Mathematically the generalized correlations are the square root of the $min(K, K_G)$ largest eigenvalues of the matrix $[F,G]^{-1}[F,F][F,G][G,G]^{-1}$. Similarly the distance between two loading matrices Λ and $\tilde{\Lambda}$ with dimension $N \times K$ respectively $N \times \tilde{K}$ is measured as the square root of the $\min(K, \tilde{K})$ largest eigenvalues of $(\Lambda^\top\Lambda)^{-1}\Lambda^\top\tilde{\Lambda}(\tilde{\Lambda}^\top\tilde{\Lambda})^{-1}\tilde{\Lambda}^\top\Lambda$.

of the three factors in F. Note, that the number of candidate factors K_G can be different from the number of factors K that I want to explain.[16]

In order to interpret latent factor models Pelger and Xiong (2018) propose the use of proxy factors. The proxy factors use only the largest portfolio weights of the latent factors and set the smaller portfolio weights to zero. Pelger and Xiong (2018) show that the largest factor portfolio weights already contain most of the information signal to construct the latent factor even if the true factor itself is not sparse.

4 Empirical Results

4.1 Data

I combine data from the WRDS TAQ Millisecond trades database, WRDS CRSP Daily Security database and WRDS Compustat from January 2004 to December 2016. This is the earliest time period for which the TAQ Millisecond data is available.[17] I calculate high-frequency, daily and overnight returns for all assets included in the S&P 500 index at any time between 1993 to 2012.[18] In order to strike a balance between the competing interests of utilizing as much data as possible and minimizing the effect of microstructure noise and asynchronous returns, I choose to use 5-minute prices. As I will show in the in section 4.7 my results are robust to this choice. More details about the data selection and cleaning procedures are in Appendix 6.1. Each trading day contains 79 price observations. As for a significant number of stocks I do not observe trading at the opening at 9:30am but only some minutes later, I start the intraday sample at 9:35am resulting in 77 log returns for each day and each asset.[19] For each of the 13 years I have on average 250 trading days with a cross-section between 555 to 667 firms. The intersection of all firms for the whole time horizon forms my balanced panel and consists of 332 firms.

The daily returns are downloaded from the WRDS CRSP Daily Security database and adjusted for dividends and stock splits. Overnight log returns are calculated as the difference between intraday log returns and daily adjusted log returns. I extend my data set to

[16] In Pelger (2019) I provide the statistical arguments for consisteny.

[17] In a previous version of this paper I used data from the WRDS TAQ Second trades database. The Second trades database goes back further, but includes a smaller number of trades for less stocks.

[18] By working only with stocks that have been in the S&P 500 index at some point, I avoid that my results are driven by small, illiquid stocks. In addition, the large stocks have typically a longer time series. When creating a balanced panel based on either all available stocks or the stocks in the S&P 500 I obtain roughly the same data set after the data cleaning procedure.

[19] The main results are not affected by this choice and the results with 78 daily log returns are available upon request.

all stocks that are available at any day from January 2004 to December 2016 in the WRDS TAQ Millisecond trades database and calculate their high-frequency, daily and overnight returns and the firm characteristics size, book-to-market ratio and momentum. Following the standard procedure of Fama and French (1992) and using the breakpoints from Kenneth French website, I create 6 portfolios formed on size and book-to-market ratio and a value weighted market, size, value and momentum factor for a high-frequency version of the Fama-French-Carhart model. Using the daily interest rates from Kenneth French website, I calculate a high-frequency and overnight interest rate under the assumption that the interest rate stays constant over the day.[20]

When identifying jumps, I face the tradeoff of finding all discontinuous movements against misclassifying high-volatility regimes as jumps. Therefore, the threshold should take into account changes in volatilities and intra-day volatility patterns. I use the TOD estimator of Bollerslev et al. (2013) for separating the continuous from the jump movements. Hence the threshold is set as $a \cdot 77^{-0.49}\hat{\sigma}_{j,i}$, where $\hat{\sigma}_{j,i}$ estimates the daily volatility of asset i at time j by combining an estimated Time-of-Day volatility pattern with a jump robust bi-power variation estimator for that day. Intuitively I classify all increments as jumps that are beyond a standard deviations of a local estimator of the stochastic volatility. For my analysis I use $a = 3$, $a = 4$, $a = 4.5$ and $a = 5$.

I have applied the factor estimation to the quadratic covariation and the quadratic correlation matrix, which corresponds to using the covariance or the correlation matrix in long-horizon factor modeling. For the second estimator I rescale each asset for the time period under consideration by the square-root of its quadratic covariation. Of course, the resulting eigenvectors need to be rescaled accordingly in order to obtain estimators for the loadings and factors. All my results are virtually identical for the covariation and the correlation approach, but the second approach seems to provide slightly more robust estimators for shorter time horizons. Hence, all results reported in this paper are based on the second approach.

Table 1 lists the fraction of increments identified as jumps for different thresholds for the balanced and unbalanced data where I use the total cross-section available for each year. Depending on the year for $a = 3$ more than 99% of the observations are classified as continuous, while less than 1% are jumps. In 2012, 99.4% of the movements are continuous and explain around 87-88% of the total quadratic variation, while the 0.6% jumps explain the remaining 12-13% of the total quadratic covariation. Increasing the threshold less movements

[20] As interest rate are significantly smaller than stock returns over the time horizon under consideration, this assumption has a neglible effect.

are classified as jumps.[21] All the results for the continuous factors are extremely robust to this choice. However, the results for the jump factors are sensitive to the threshold. If not noted otherwise, the threshold is set to $a = 3$ in the following.

One of the main contributions of the paper is to understand the time-variation in the factor structure. I start by estimating the factor structure within each year, i.e. I apply PCA based estimators to each year independently. In section 4.5 I refine the analysis and study the time-varying structure on a weekly and monthly time window. The first step is to determine the number of high-frequency, continuous and jump factors in section 4.2. My main argument is based on the perturbed eigenvalue ratio diagnostic criterion, but I complement it by showing that the including more factors than indicated by my estimator creates an unstable pattern either when I consider different samples of stocks or different time periods. Second, I show that the factor structure is essentially identical on the larger unbalanced panel and the balanced panel. Hence, without loss of generality I can study the factor structure on the representative balanced panel which is important for the asset pricing applications which require long-term means. Third, I show that the factors constructed with the portfolio weights estimated on the whole horizon are essentially identical to the factors constructed with locally estimated portfolio weights. This is relevant as I it allows me to use the portfolio weights estimated on the whole horizon to interpret the factors. Furthermore, it gives me a benchmark "rotation" of the factors to study the time-varying loadings and long-term means of the factor returns.

As a first step Table 1 lists for each year the fraction of the total continuous variation explained by the first four continuous factors and the fraction of the jump variation explained by the first jump factor.[22] The choice of four continuous and one jump factor is motivated by the results in the subsequent sections. As expected systematic risk varies over time and is larger during the financial crisis. The systematic continuous risk with 4 factors accounts for around 40-47% of the total correlation from 2008 to 2011, but explains only around 20-36% in the other years.[23] A similar pattern holds for the jumps where the first jump factor explains up to 10 times more of the correlation in 2010 than in the years before the financial crisis.

[21] There is no consensus on the number of jumps in the literature. Christensen et al. (2014) use ultra high-frequency data and estimate that the jump variation accounts for about 1% of total variability. Most studies based on 5 minutes data find that the jump variation should be around 10 - 20% of the total variation. My analysis based on different thresholds considers both cases.

[22] Each year I apply PCA to the yearly continuous respectively jump quadratic correlation matrix to estimate to the underlying factor structure.

[23] The percentage of correlation explained by the first four factors is calculated as the sum of the first four eigenvalues divided by the sum of all eigenvalues of the continuous quadratic correlation matrix.

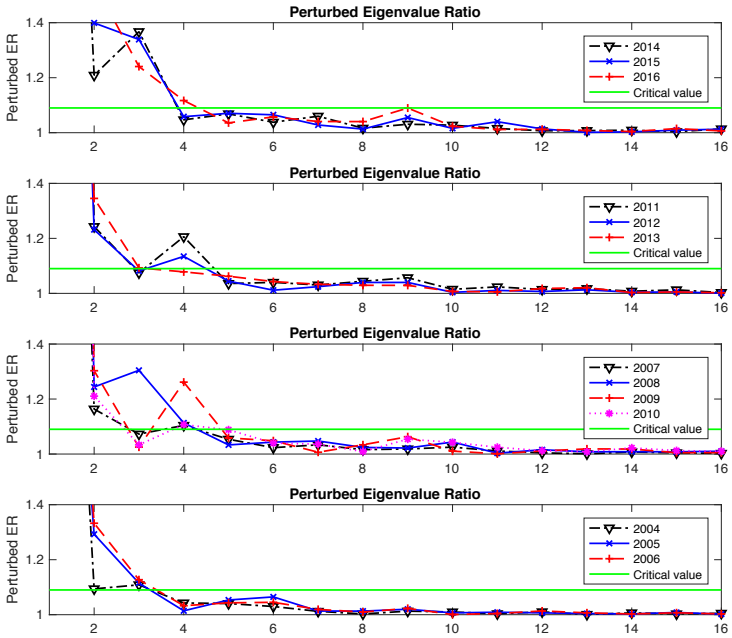
Table 1: Summary statistics for continuous and jump returns

	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
	Intersection of balanced panel ($N = 332$)												
	Fraction of increments identified as jump for different thresholds												
$a = 3$	0.89	0.87	0.79	0.84	0.71	0.60	0.60	0.56	0.63	0.60	0.56	0.53	0.53
$a = 4$	0.20	0.19	0.17	0.19	0.14	0.11	0.12	0.09	0.14	0.13	0.11	0.10	0.10
$a = 4.5$	0.10	0.10	0.09	0.11	0.07	0.06	0.06	0.04	0.07	0.07	0.06	0.05	0.05
$a = 5$	0.05	0.05	0.05	0.06	0.04	0.03	0.03	0.02	0.04	0.04	0.03	0.03	0.02
	Fraction of total quadratic variation explained by jumps												
$a = 3$	0.16	0.16	0.14	0.19	0.15	0.14	0.13	0.09	0.13	0.13	0.12	0.25	0.11
$a = 4$	0.06	0.06	0.06	0.09	0.06	0.06	0.06	0.03	0.05	0.05	0.05	0.19	0.04
$a = 4.5$	0.04	0.04	0.04	0.07	0.04	0.05	0.04	0.02	0.04	0.03	0.03	0.17	0.03
$a = 5$	0.03	0.03	0.03	0.06	0.03	0.04	0.03	0.01	0.03	0.03	0.03	0.17	0.02
	Fraction of jump correlation explained by first 1 jump factor												
$a = 3$	0.03	0.03	0.03	0.05	0.07	0.08	0.20	0.11	0.06	0.10	0.05	0.08	0.05
$a = 4$	0.02	0.02	0.04	0.07	0.05	0.07	0.28	0.07	0.07	0.18	0.07	0.09	0.06
$a = 4.5$	0.02	0.01	0.05	0.08	0.05	0.07	0.29	0.07	0.09	0.22	0.10	0.09	0.07
$a = 5$	0.02	0.01	0.05	0.09	0.06	0.08	0.30	0.07	0.11	0.22	0.12	0.09	0.07
	Fraction of continuous correlation explained by first 4 continuous factors ($a = 3$)												
	0.20	0.21	0.22	0.29	0.44	0.41	0.40	0.47	0.30	0.27	0.32	0.36	0.33
	All observations (unbalanced panel)												
N	555	607	649	649	674	674	667	659	609	656	668	650	643
	Fraction of increments identified as jump for different thresholds												
$a = 3$	0.87	0.82	0.75	0.80	0.67	0.58	0.57	0.53	0.59	0.57	0.53	0.52	0.52
$a = 4$	0.19	0.17	0.16	0.17	0.13	0.10	0.11	0.09	0.13	0.12	0.10	0.10	0.10
$a = 4.5$	0.10	0.09	0.08	0.10	0.06	0.05	0.06	0.04	0.06	0.06	0.05	0.05	0.05
$a = 5$	0.05	0.05	0.05	0.06	0.03	0.02	0.03	0.02	0.04	0.04	0.03	0.03	0.02
	Fraction of total quadratic variation explained by jumps												
$a = 3$	0.17	0.15	0.14	0.21	0.13	0.11	0.12	0.09	0.12	0.12	0.11	0.36	0.11
$a = 4$	0.06	0.06	0.05	0.11	0.05	0.03	0.05	0.03	0.05	0.04	0.04	0.31	0.04
$a = 4.5$	0.04	0.04	0.04	0.09	0.03	0.02	0.04	0.02	0.03	0.03	0.03	0.30	0.03
$a = 5$	0.03	0.03	0.03	0.08	0.02	0.01	0.03	0.01	0.03	0.02	0.03	0.30	0.03
	Fraction of jump correlation explained by first 1 jump factor												
$a = 3$	0.03	0.03	0.03	0.24	0.07	0.08	0.24	0.10	0.06	0.11	0.06	0.82	0.05
$a = 4$	0.02	0.03	0.04	0.36	0.05	0.07	0.36	0.08	0.07	0.19	0.10	0.92	0.09
$a = 4.5$	0.03	0.04	0.05	0.45	0.05	0.07	0.41	0.12	0.08	0.23	0.13	0.94	0.11
$a = 5$	0.04	0.05	0.06	0.51	0.05	0.06	0.46	0.17	0.09	0.23	0.15	0.95	0.14
	Fraction of continuous correlation explained by first 4 continuous factors ($a = 3$)												
	0.22	0.23	0.24	0.31	0.47	0.42	0.42	0.49	0.33	0.29	0.34	0.39	0.36

4.2 Number of Factors

I estimate four high frequency factors for each of the years from 2007 to 2012 and in 2016 and three factors for the years 2004 to 2006 and 2013 to 2015. Figures 1 and 2 show the estimation results for the perturbed eigenvalue ratio diagnostic criterion for the balanced and unbalanced panel.[24] Starting from the right I am looking for a visible strong increase in the perturbed eigenvalue ratio. Asymptotically any critical value larger than 1 should indicate the beginning of the systematic spectrum. However, for my finite sample I need to choose a critical value. In the plots I set the critical value equal to 1.08. Fortunately there are very visible humps at 4 for the years 2007 to 2012 and strong increases at 3 for the years 2004 to 2006 and 2013 to 2016, which can be detected for a wide range of critical values. Therefore, my estimator strongly indicates that there are 4 high frequency factors from 2007 to 2012 and three high-frequency factors for the other years. The number of factors is the same for the balanced and unbalanced panel, which is in line with the results in Table 2 in the next section.

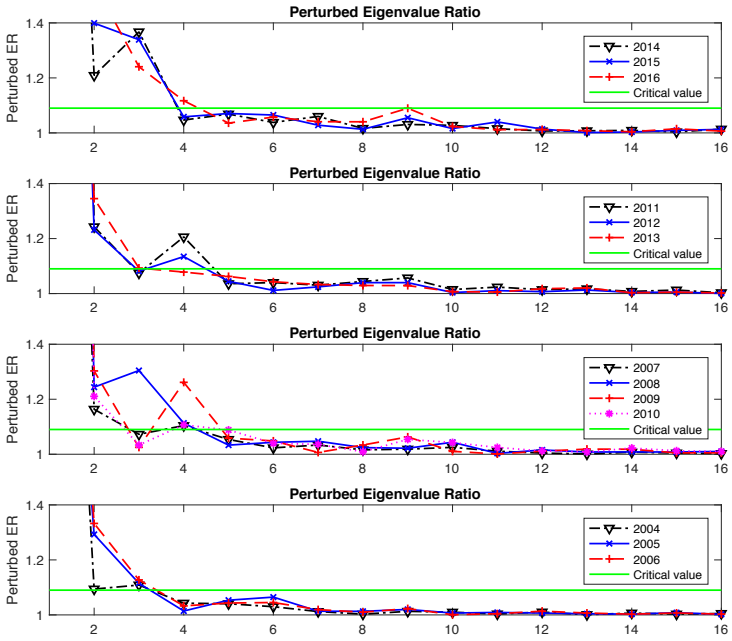
Figure 1: Number of HF factors



Note: Perturbed eigenvalue ratio statistic for unbalanced panel of all HF (continuous + jump) returns.

[24] I have conducted the same analysis for more perturbation functions and with the Onatski (2010) eigenvalue difference estimator with the same findings. The results are available upon request.

Figure 2: Number of HF factors



Note: Perturbed eigenvalue ratio statistic for balanced panel of HF (continuous + jump) returns.

The number of continuous and jump factors should be bounded by the number of total high-frequency factors. Figures 15 and 16 apply the diagnostic criterion to the continuous movements. The number of continuous factors seems to be the same, i.e. four continuous factors from 2007 to 2012 and three continuous factors for the other years. The only outlier is the year 2015, where the number of continuous factors is estimated as 4 respectively 5. As this outlier result is sensitive to the cutoff threshold, it is likely due to estimation noise.

There seems to be a lower number of jump factors and in most years there is only one jump factor. The diagnostic criterion applied to jumps which are identified by three standard deviations ($a = 3$) in Figures 17 and 18 suggests one factor in 6 respectively 7 of the 13 years. Overall the estimation is much noisier. When classifying jumps based on more standard deviations ($a > 3$) there are too few jump observations to make a reliable prediction.[25] In particular, there is only a small number of co-jumps between different assets left. Thus, even if only a relatively small number of assets jump together, it will already lead to large eigenvalue in the jump covariance matrix although it might not be systematic from an economic point of view. Based on the instability in the estimation for jumps factors for $a > 3$, I will consider only $a = 3$ in the following.

[25]The results are collected in the supplementary appendix.

Table 2: Generalized correlation of factors from unbalanced and balanced panel.

	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
$\hat{K}$	3	3	3	4	4	4	4	4	4	3	3	3	4
N	555	607	649	649	674	674	667	659	609	656	668	650	643
	First 3 continuous PCA factors												
1. GC	1.00	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2. GC	0.99	0.99	0.99	0.98	0.98	0.97	0.96	0.97	0.97	0.97	0.98	0.99	0.98
3. GC	0.96	0.87	0.92	0.87	0.97	0.90	0.90	0.87	0.97	0.94	0.98	0.98	0.98
	First 4 continuous PCA factors												
1. GC	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2. GC	0.99	0.99	0.99	0.98	0.99	0.98	0.96	0.97	0.98	0.98	0.98	0.99	0.98
3. GC	0.96	0.96	0.93	0.96	0.97	0.97	0.96	0.96	0.97	0.96	0.98	0.98	0.98
4. GC	0.92	0.64	0.71	0.94	0.91	0.95	0.90	0.92	0.91	0.14	0.57	0.88	0.93
	First 3 jump PCA factors												
1. GC	0.97	0.97	0.97	0.99	0.98	0.99	1.00	0.99	0.99	0.99	0.99	1.00	0.98
2. GC	0.37	0.48	0.90	0.08	0.97	0.88	0.99	0.93	0.90	0.71	0.97	0.99	0.42
3. GC	0.07	0.05	0.16	0.00	0.93	0.51	0.94	0.03	0.45	0.02	0.07	0.96	0.01
	First 4 jump PCA factors												
1. GC	0.97	0.97	0.98	0.99	0.98	0.99	1.00	0.99	0.99	0.99	0.99	1.00	0.98
2. GC	0.52	0.53	0.92	0.94	0.97	0.89	0.99	0.94	0.94	0.91	0.97	0.99	0.96
3. GC	0.17	0.14	0.88	0.08	0.93	0.57	0.96	0.76	0.89	0.06	0.74	0.98	0.08
4. GC	0.06	0.06	0.00	0.00	0.68	0.02	0.83	0.04	0.02	0.02	0.08	0.03	0.01

Note: Generalized correlation between 3 or 4 PCA factors estimated on the continuous respectively jump ($a = 3$) returns for the balanced and unbalanced data ($N = 332$ for balanced data). The number of estimated high-frequency factors and the number of stocks in the unbalanced panel are indicated at the top.

4.3 Balanced panel

I find that the balanced subsample is representative for the whole data set, which justifies why I will focus on the balanced panel in the following. A first takeaway from Table 1 is that the balanced and unbalanced panel seem to have a very similar pattern. I confirm in Table 2 that the factor structure estimated from the unbalanced and balanced panel coincides. Table 2 lists the generalized correlations between the first three and four latent continuous factors estimated on the whole data and its intersection for each year. Generalized correlations equal to 1 indicate that the factors are the same. It is apparent that the first three continuous PCA factors are essentially identical on both data sets. The first four factors coincide for the years where I estimate a four factor structure. In the years where I estimate only three factors, the fourth PCA factor can be different for the two data sets. This makes sense as if

there are only three factors in a specific year, the fourth PCA is expected to fit only noise and hence should not be the same in the two panels. Therefore I view this as a confirmation for the estimator of the number of factors. The jump structure seems to share only one or two common jump factors. This is again in line with the results about the number of factors in the previous subsection.

4.4 High-Frequency Factors

The four continuous and high-frequency factors for 2004 to 2016 can be approximated very well by industry factors. I find a clear pattern in the factor portfolio weights that is linked to industry affiliations and allows me to put an economic label on the statistical factors. As I can show that the factor portfolio weights for the continuous and high-frequency (continuous and jumps) factors are identical, most of the following analysis is based on the continuous portfolio weights. In this section I will study the factor portfolio weights estimated on the whole time horizon of 13 years. In Table 8 I study the generalized correlations between the portfolio weights estimated on the whole time horizon and for each year. In section 4.5 I extend this analysis to shorter local windows. As it is apparent the factor composition is stable over time and the same pattern that holds for the whole time horizon also holds locally.

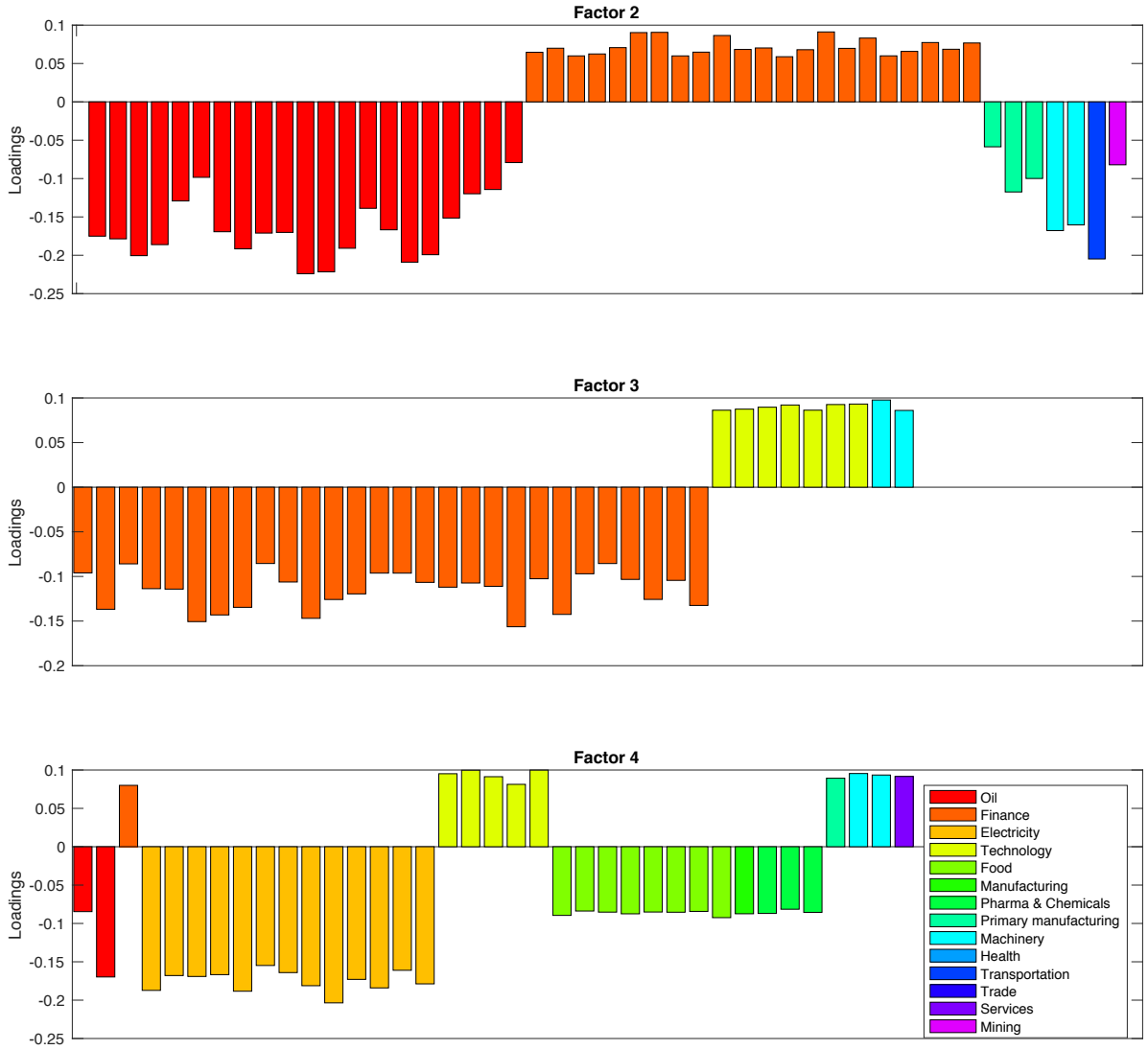
First, I will study the composition of the continuous factors. Pelger and Xiong (2018) suggests the use of proxy factors to interpret latent PCA factors. My first proxy factor is an equally weighted market portfolio. The second proxy factors has the 15% and third and fourth proxy factors have the 11% largest portfolios weights of the corresponding statistical factors.[26] Figure 3 depicts the portfolio weights of the proxy factors sorted according to industries. Appendix 6.1 provides the details for the industry classifications. The second proxy factor is a long-short factor in the oil and finance industry. The third proxy factor is a finance industry factor. The fourth proxy factor seems to be an electricity industry factor.

Figure 4 shows the factor portfolio weights of the continuous PCA factors without the proxy shrinking. The stocks are again sorted according to industries. The first factor has only long positions of similar magnitude which justifies the interpretation as an equally weighted market portfolio. The second continuous PCA factor has exactly the same interpretation as a long-short factor in the oil and finance industry with mostly negligible weights in the other industries. The third factor has the largest weights in the finance industry, but also

[26]The fraction or largest portfolio weights is chosen to obtain a large generalized correlation with the high-frequency PCA factors.

non-negligible weights in the oil and technology industry. The fourth factor has the largest weights in the electricity industry with some minor outliers in other industries.

Figure 3: Portfolio weights of proxy factors



Note: Proxy factors for the 4 continuous PCA factors. First proxy factor is an equally weighted market portfolio. Second proxy factors has 15% and third and fourth proxy factors have the 11% largest portfolios weights of the corresponding statistical factors.

Based on these insights I construct four industry factors: (1) an equally weighted market portfolio, (2) equally weighted oil industry factor, (3) equally weighted finance industry factor and (4) equally weighted electricity factor. In addition, I compare the latent factors to the four Fama-French-Carhart factors. Table 3 lists the generalized correlations for different

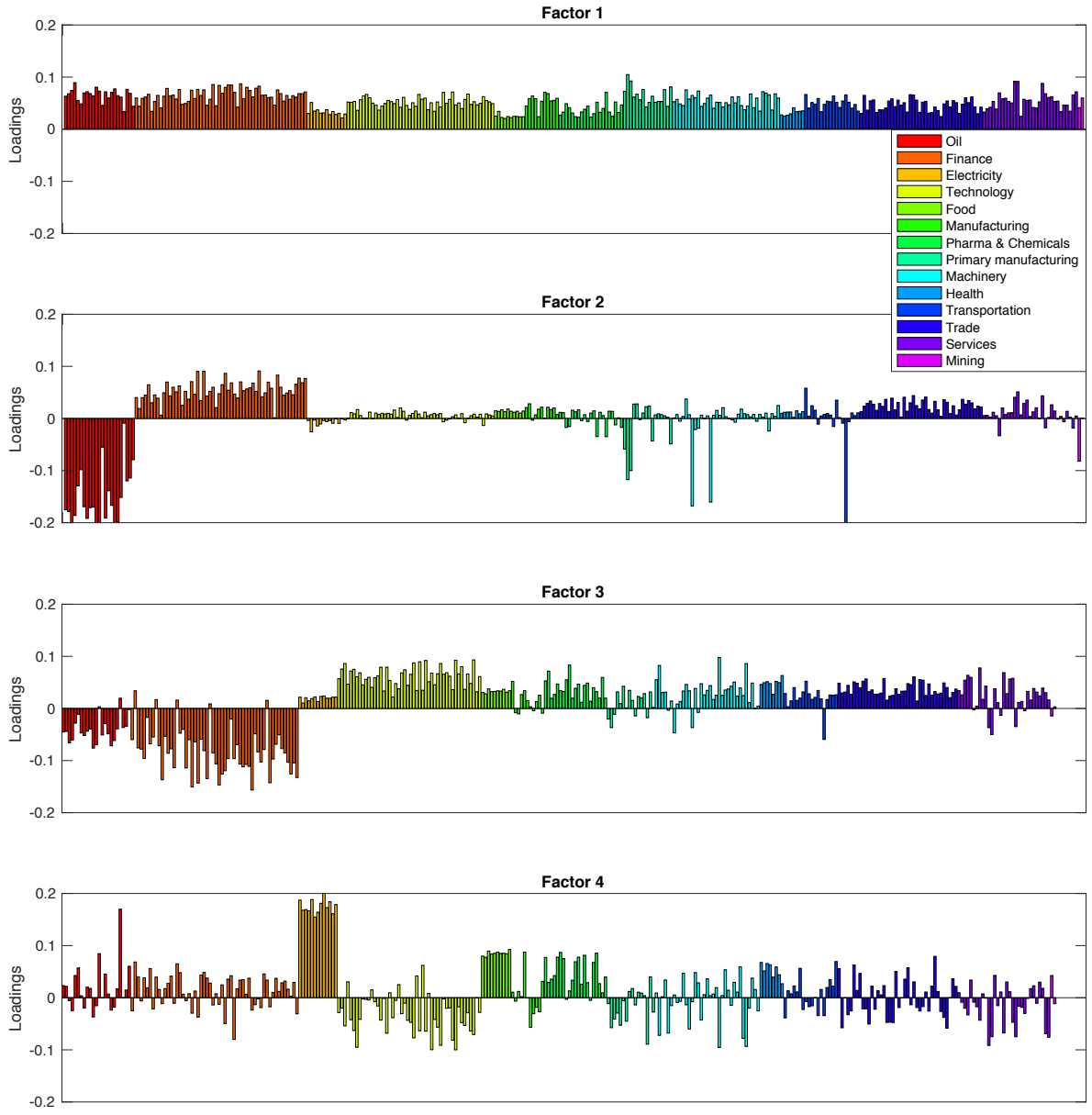
continuous factors with the four continuous PCA factors. The proxy factors have generalized correlations of $\{1, 0.99, 0.95, 0.91\}$ confirming that they provide an excellent approximation to the latent PCA factors. The four industry factors approximate three of the latent factors very well. The generalized correlation between the PCA factors and the market, oil and finance factor are $\{1, 0.98, 0.95\}$ indicating that these three factors do indeed capture three of the PCA factors. Adding the electricity factor provides a good but not perfect approximation of the fourth PCA factor. Using the factor portfolio weights based on all high frequency returns (continuous and jumps), I construct HF PCA factors with continuous returns, which are perfectly correlated with the continuous PCA factors. Figure 19 confirms that the high-frequency PCA factor portfolio weights have the same pattern as the continuous weights.

Next, I compare the portfolio weights based on jump, overnight and daily log returns in Table 3. As already noted, the high-frequency (continuous + jumps) factor portfolio weights have the same span as the continuous weights. As expected the first four PCA jump portfolio weights are different and share at only two common factors. Figure 20 suggests that the jumps pick up a market and oil factor. However, when increasing the jump threshold (i.e. $a > 3$) the commonality between continuous and jump factors shrinks until they have only a market factor in common as shown in Table 7. Overnight and daily data seem to extract a similar but noisier pattern than the high-frequency returns. Figure 22 and 21 show a similar pattern in the three industries and the market, but with more outliers.

The Fama-French-Carhart factors are different from the continuous PCA factors. They have the market factor in common, but the size, value and momentum factor do not replicate the other three PCA factors. In particular, the momentum factor is essentially orthogonal to the PCA factors. Table 3 shows that the market factor has the almost perfect correlation of 0.98 with the continuous PCA factors. The size and momentum factor only add a correlation of 0.65 respectively 0.43. The momentum factor adds essentially zero to the generalized correlations.

Figure 23 shows that the above findings are robust to the type of returns that I study. Based on the same portfolio weights I calculate the intraday, overnight and daily returns for the continuous PCA, high-frequency PCA, proxy PCA, industry and Fama-French-Carhart factor portfolio weights. The generalized correlations are identical to the continuous returns, i.e. the proxy and industry factors provide a good approximation while size, value and momentum factors are only weakly correlated with the PCA factors.

Figure 4: Portfolio weights of 4 continuous PCA factors.



Note: Portfolio weights of 4 continuous PCA factors over the whole time horizon 2004 to 2016. Stocks are sorted according to industry.

Table 3: Generalized correlations of continuous PCA factors with other factors

	1. GC	2. GC	3. GC	4. GC
Continuous generalized correlations with 4 continuous PCA factors				
HF PCA	1.00	1.00	1.00	1.00
PCA Proxy	1.00	0.99	0.95	0.91
Industry (M,O,F,E)	1.00	0.98	0.95	0.78
Industry (M,O,F)	1.00	0.98	0.95	0.00
Industry (M,F,E)	1.00	0.98	0.78	0.00
Industry (M,O,E)	1.00	0.96	0.78	0.00
Fama-French-Carhart	0.98	0.66	0.45	0.05
Fama-French 3	0.98	0.65	0.43	0.00
Market	0.98	0.00	0.00	0.00
Generalized correlations with factor portfolio weights of 4 continuous PCA factors				
Λ HF	1.00	1.00	1.00	1.00
Λ Jump	0.98	0.96	0.65	0.39
Λ Overnight	0.99	0.98	0.93	0.82
Λ Daily	1.00	0.98	0.97	0.94

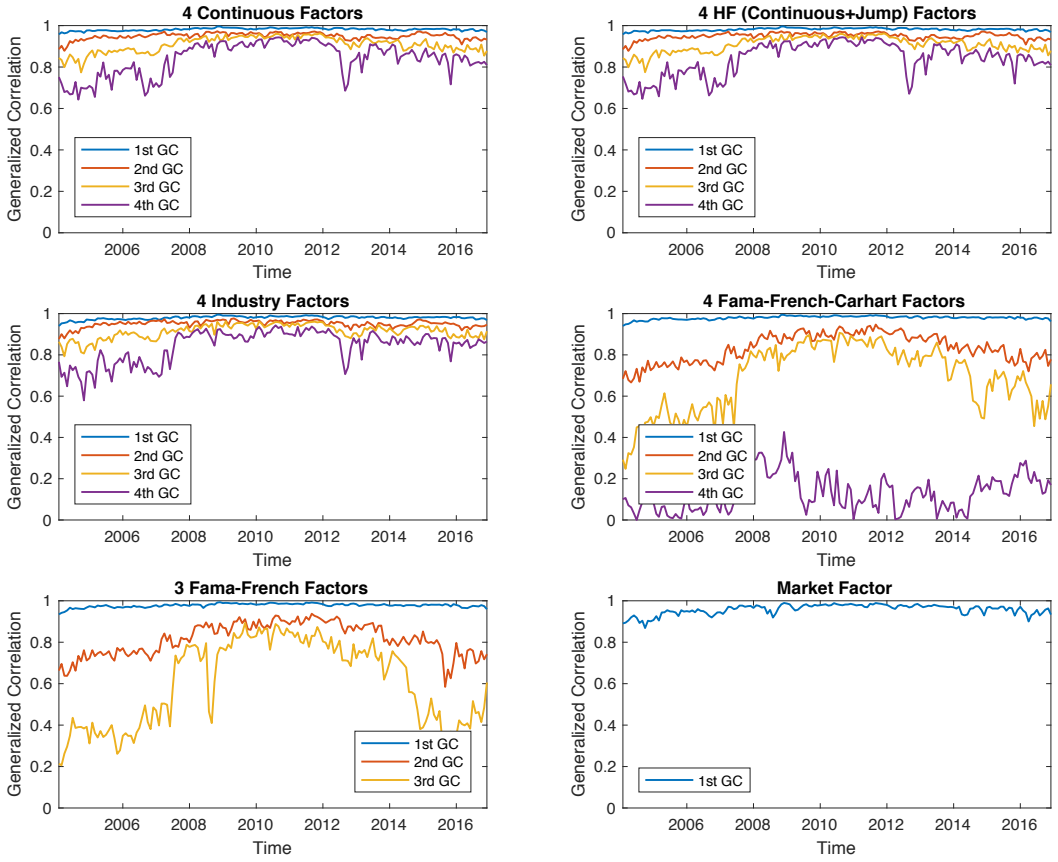
Note: (1) Continuous generalized correlations of the first four statistical continuous PCA factors with the 4 PCA factors based on all HF data (continuous+jumps), the 4 continuous PCA proxy factors, different combinations of industry factors (market, oil, finance and electricity), the 4 Fama-French-Carhart continuous factors, the 3 Frama-French factors and the continuous market factor. (2) Generalized correlations of factor portfolio weights of the 4 continuous PCA factors with portfolio weights of 4 PCA factors based on all HF data (continuous+jumps), 4 jump PCA factors, 4 overnight PCA factors and 4 daily PCA factors. Balanced panel from 2004 to 2016 ($M = 252,021$ HF increments, $T = 3,273$ days and $N = 332$).

4.5 Time-Variation

Using a short time horizon of one month (21 trading days) I study the time variation in the portfolio weights and loadings of different factors. The supplementary appendix collects the results for a one week horizon (5 trading days) and a three month window (63 trading days) with essentially identical results. The four continuous PCA factors are very stable over time while the Fama-French-Carhart factors have a time-varying factor structure. Given the continuous 4 PCA factors estimated over the whole time horizon, the industry factors and the Fama-French-Carhart factors I estimate their loadings on a rolling window of one months

with continuous log returns.[27]

Figure 5: Time-variation in loadings



Note: Generalized correlations of continuous loadings estimated on the whole time horizon and on a moving window of one month (21 trading days).

First, I calculate the generalized correlations of the loadings estimated on the whole time horizon with those estimated on the moving window as depicted in Figure 5. Here, I keep the factor portfolio weights constant, but allow for arbitrary monthly time variation in the loadings. Surprisingly, the loadings for the PCA and industry factors are very stable over time. However, the loadings for the size, value and momentum factors do not have the same span for different time periods. This finding does not imply that the loadings for the PCA factors are constant over time. Figure 31 plots the loadings for the PCA and Fama-French-Carhart factors for different times suggesting variation over time, but with more time-variation for the Fama-French-Carhart factors. The finding that the generalized

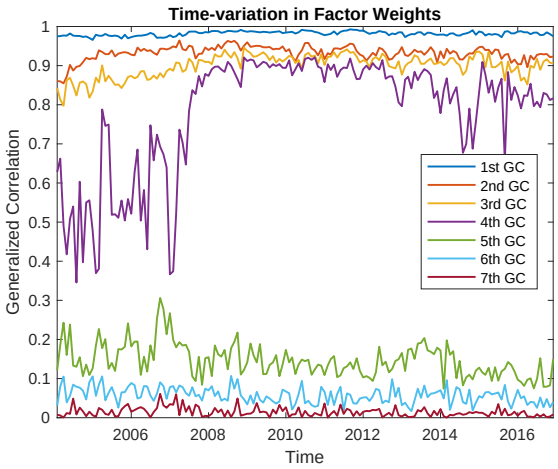
[27] As already noted before the high-frequency loadings are identical to the continuous loadings and I only report the continuous loadings here.

correlation of the PCA loadings is the same for different time periods corresponds to a model of the form

$$\Lambda(t) = \underbrace{\bar{\Lambda}}_{N\times K}\underbrace{H(t)}_{K\times K}.$$

This implies that the projection of the return space on the common component or residual space is the same with each $\Lambda(t)$, i.e. the cross-sectional relationship has a stable structure. This justifies why I can estimate the portfolio weights of PCA factors on the whole time horizon. However, the loadings for an individual asset can have very different values for different time periods. Importantly, this stable structure does not hold for the Fama-French-Carhart factors, implying that regressions of stocks on these factors are biased even for very short time horizons.

Figure 6: Time-variation in locally estimated continuous factors



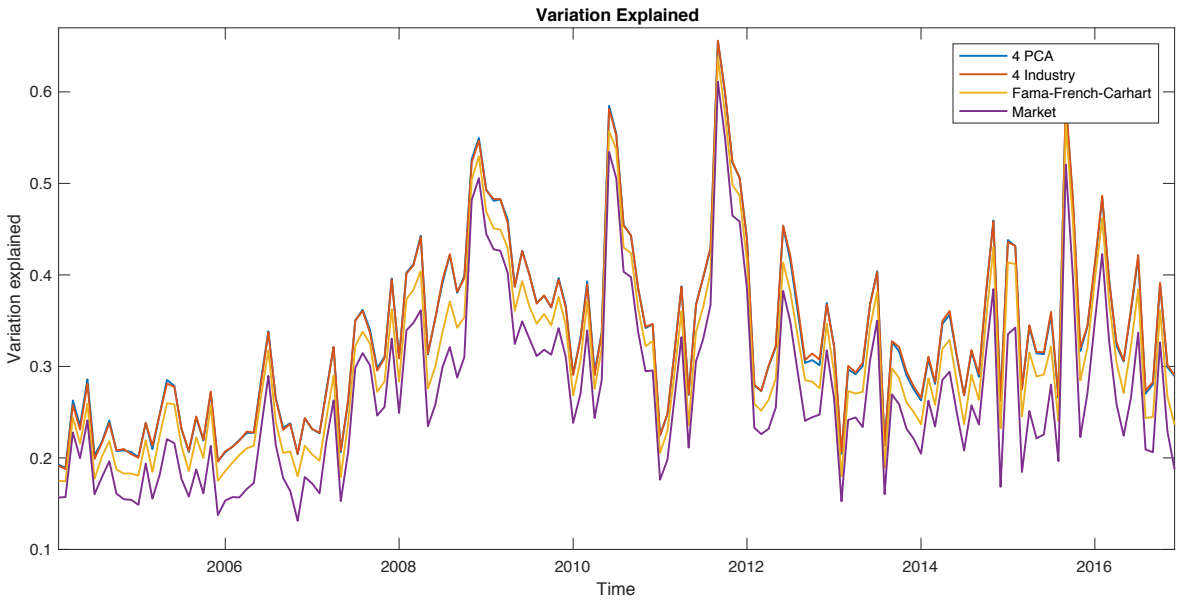
Note: Generalized correlations of factor portfolio weights of first 7 continuous PCA factors estimated on the whole time horizon and on a moving window of one month (21 trading days).

My previous analysis for the number of factors indicates that some time periods have more systematic factors than others. This is not in contradiction with the stable span of the loadings. It is possible that the loadings for a factor take very small values in a specific time period or that the volatility of a factor is locally very small and thus this factor does not explain a systematic portion of the correlation in the data for this period. In this case PCA does not detect this factor. Figures 6 and 30 estimate the PCA factor portfolio weights on the local one month window and the total time horizon and report the generalized correlations. Note, that in contrast to the previous analysis I apply a separate PCA to each of the local

one month windows to obtain the local factor portfolio weights. When estimating only the first four PCAs the highest generalized correlations are from 2007 to 2012 with a sharp drop for the other years. When adding the first seven PCAs the fifth to seven generalized correlations are close to zero which is a clear indication that they are fitting noise. Overall, these findings confirm the estimation for the number of factors.

Figures 27 and 28 take a closer look at the time variation in the locally estimated continuous PCA factors. They show the factor portfolio weights for November 2006 respectively April 2008. In 2006 there is no finance factor present resulting in the three factor structure. In 2008 the fourth PCA factor is clearly loading heavily on the finance industry. This suggests that the fourth factor from 2007 to 2012 is the finance factor. Figure 29 plots the generalized correlation between the locally estimated PCA factors and different combinations of the industry factors. The strong increase in the generalized correlation from the years 2007 to 2012 appears only when including the finance factor which supports the hypothesis.

Figure 7: Time-variation in the percentage of explained variation for different factors



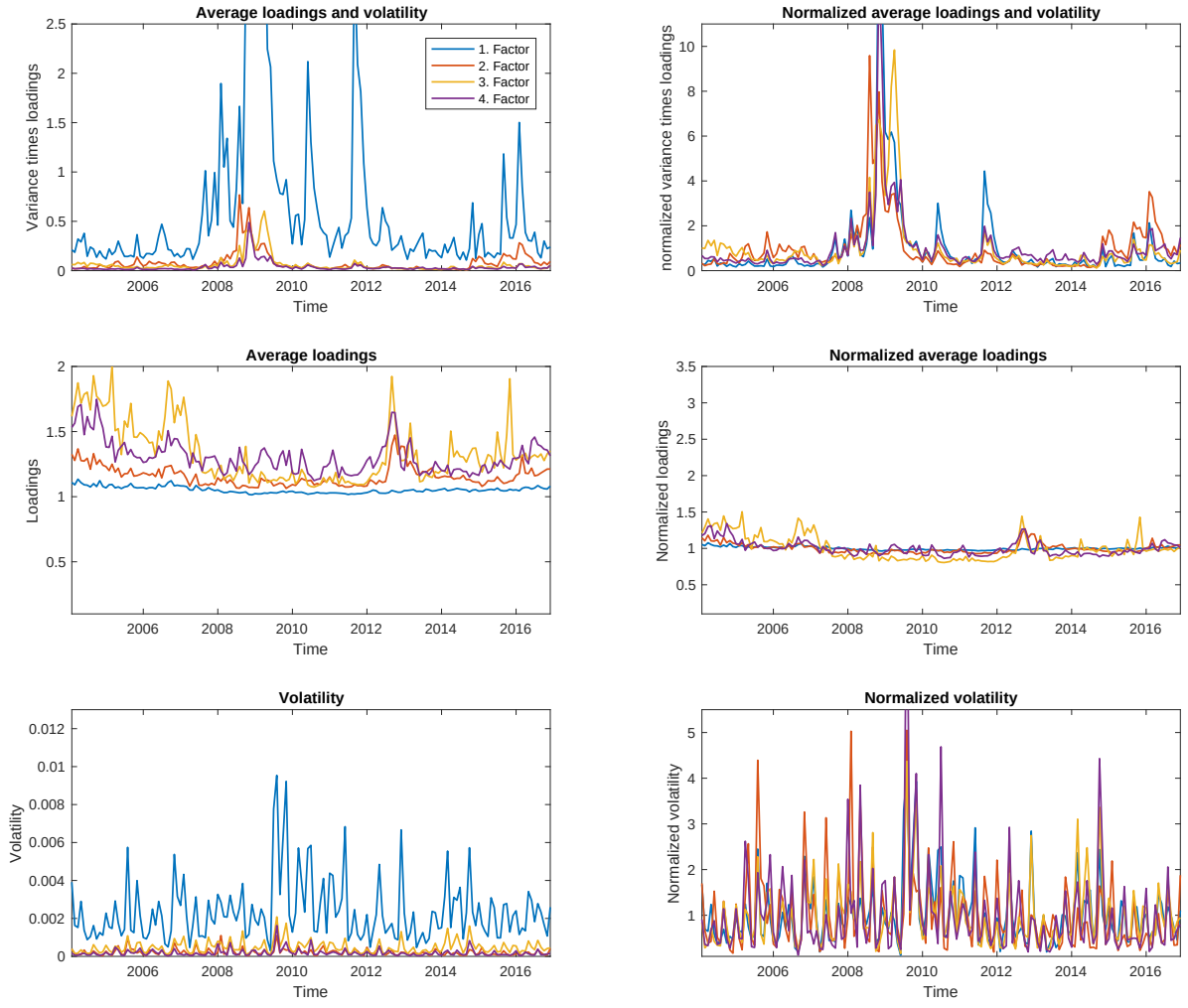
Note: The continuous variation is calculated on a moving window of one month (21 trading days).

The amount of variation explained by the factors changes over time. Table 1 already indicated that the proportion of variation explained by factors increased during the financial crisis.[28] Figure 7 provides a more refined analysis using the monthly window with local

[28]In the supplementary appendix I show a very similar pattern for the HF, daily and overnight variation.

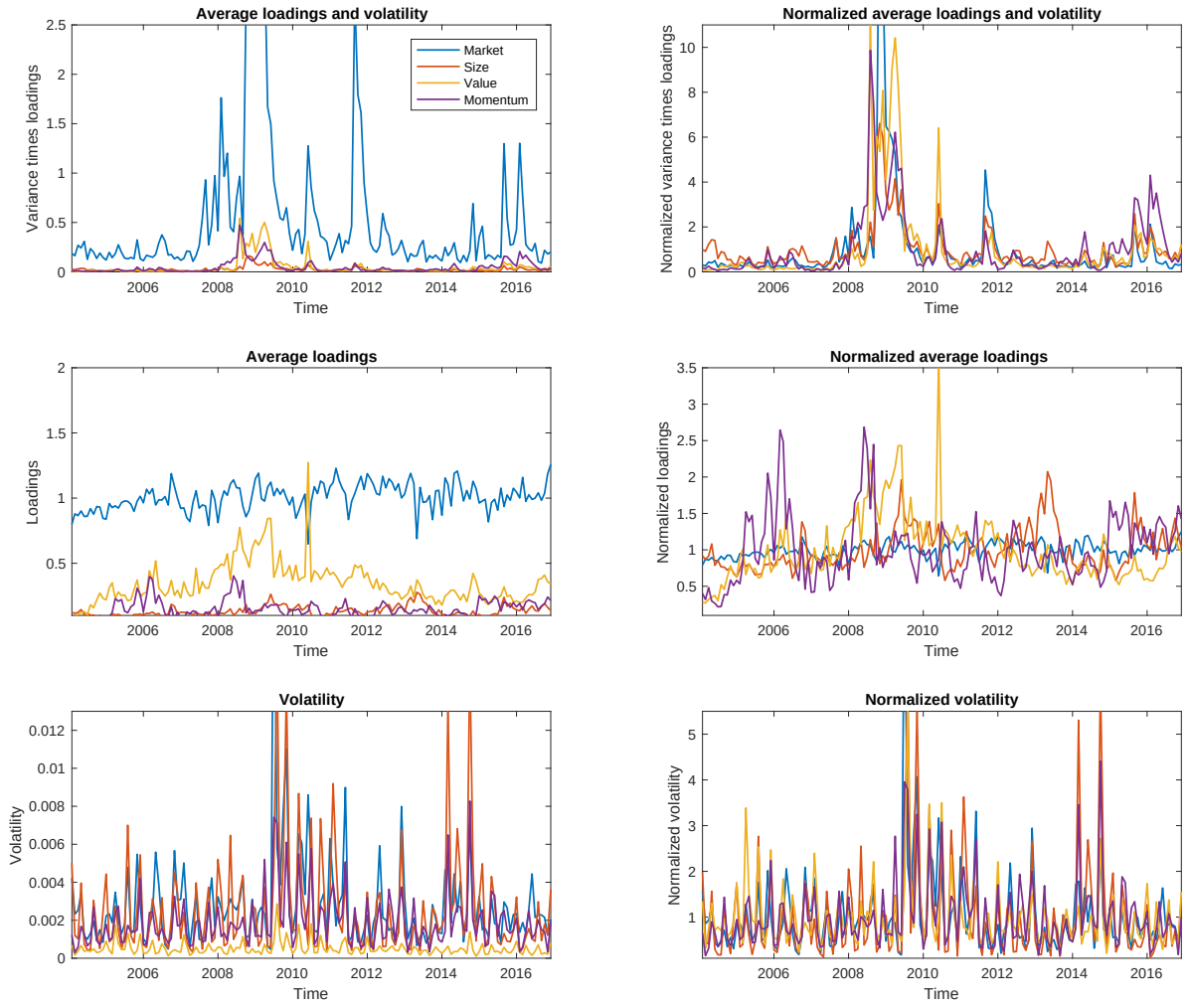
regressions and confirms this pattern. The proportion of systematic risk peaks from 2009 to 2012 and again at the end of 2015. The four continuous PCA and industry factors explain roughly the same proportion of risk while the Fama-French-Carhart factors capture a smaller portion. The market factor alone already explains a large part of the variation.

Figure 8: Decomposition of time variation in factor structure for 4 continuous PCA factors



Note: Time-varying loadings and volatilities are estimated on a moving window of one month (21 trading days) based on continuous returns. Left panel: (1) Systematic impact of factors $\frac{\Lambda_k(t)^\top \Lambda_k(t)}{N}\sigma_k^2(t)$ with $\Lambda_k(t)$ being the continuous loadings of factor k in month t and σ_k^2 is the continuous quadratic variation of factor k in month t. (2) Average loadings $\frac{\Lambda_k(t)^\top \Lambda_k(t)}{N}$, (3) Volatility $\sigma_k^2(t)$. Right panel: The same quantities as in the left panel but normalized by the time average of the quantity of interest.

Figure 9: Decomposition of time variation in factor structure for 4 continuous Fama-French Carhart factors



Note: Time-varying loadings and volatilities are estimated on a moving window of one month (21 trading days) based on continuous returns. Left panel: (1) Systematic impact of factors $\frac{\Lambda_k(t)^\top \Lambda_k(t)}{N}\sigma_k^2(t)$ with $\Lambda_k(t)$ being the continuous loadings of factor k in month t and σ_k^2 is the continuous quadratic variation of factor k in month t. (2) Average loadings $\frac{\Lambda_k(t)^\top \Lambda_k(t)}{N}$, (3) Volatility $\sigma_k^2(t)$. Right panel: The same quantities as in the left panel but normalized by the time average of the quantity of interest.

I take a closer look at the source of the time variation in the factor structure of continuous PCA and Fama-French-Carhart factors in Figures 8 and 9. I keep constant factor portfolio weights and estimate the regression loadings and volatilities locally on a one month horizon. First, I study the time variation in the systematic part which is the product of the loadings

and the corresponding factor volatility $\frac{\Lambda_k(t)^\top \Lambda_k(t)}{N}\sigma_k^2(t)$. This product corresponds roughly to the eigenvalues that are due to the individual factors. The left plots show the absolute values while the right plots are a normalization with the time average of the quantity. It is apparent that all factors have a larger systematic portion from 2008 to 2012. During the times when the eigenvalues due to the second and third factors, which are linked to finance, are small, I estimate only three factors. The middle plots show the average loadings $\frac{\Lambda_k(t)^\top \Lambda_k(t)}{N}$ over time. Here is the biggest difference between the PCA and Fama-French-Carhart factors. The average loadings for PCA factors are almost constant, while the average loadings for the Fama-French-Carhart factors fluctuate wildly which supports the finding that this factor structure is not stable. Finally, as expected the volatility for the different factors varies substantially over time, which is the reason that the number of PCA factors is time-varying.

Overall, the findings of this section suggest that PCA analysis applied to a 13 year horizon on intraday data provides a valid set of factors and it is not necessary to estimate the factors locally. In contrast, even for very short time horizons Fama-French-Carhart factors require an adjustment for time-variation. However, when working with monthly returns over the longer horizon of 40 years, Lettau and Pelger (2018) show that the estimated PCA portfolio weights are not stable anymore.

4.6 Asset Pricing

Arbitrage pricing theory postulates a connection between factors that explain the co-movement and the cross-section of expected returns. Under the assumption of arbitrage pricing theory the stochastic discount factor is spanned by the factors that explain the systematic co-movement. Hence, the tangency portfolio with the optimal (conditional) Sharpe-ratio has to be composed of only of these factors. Comparing the Sharpe-ratios of mean-variance efficient portfolios based on factors is therefore a test for the pricing performance of the factors.

Table 4 lists the maximum Sharpe ratio for tangency portfolios based on different factors. The four daily PCA factors have an annual Sharpe ratio of 0.4 which is the same as for a market factor.[29] In contrast the tangency portfolio based on the Fama-French-Carhart daily factor returns strongly outperforms the market portfolio for this time period. As before continuous PCA factors are constructed with the portfolio weights estimated from the continuous loadings over the whole time horizon. The corresponding tangency portfolio

[29]PCA factors based on monthly stock returns over long horizons do not outperform characteristics-based factors as shown in Lettau and Pelger (2018).

significantly outperforms the characteristics based factors on the daily horizon.[30]

Table 4: Sharpe-ratios of factors intraday and overnight

	Intraday	Overnight	Daily
Sharpe ratio of tangency portfolio			
Continuous PCA	1.45	1.02	0.89
Proxy PCA	1.09	1.09	0.65
Industry	0.84	0.71	0.61
Fama-French-Carhart	0.41	1.32	0.60
PCA Daily	0.16	0.47	0.40
PCA Overnight	0.43	1.01	0.42
Sharpe ratios of individual factors			
1. Continuous PCA Factor	0.33	-0.03	0.25
2. Continuous PCA Factor	0.43	-0.61	0.04
3. Continuous PCA Factor	0.62	0.09	0.57
4. Continuous PCA Factor	1.05	-0.79	0.52
Market	0.16	0.47	0.41
Size	-0.07	0.71	0.27
Value	-0.17	0.48	0.17
Momentum	-0.14	0.69	0.22

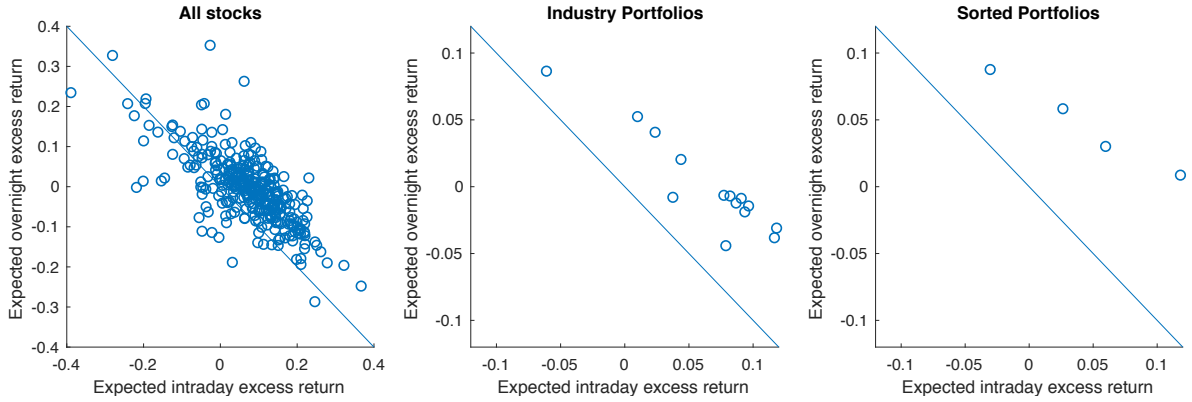
Note: (1) Maximum Sharpe ratio of optimal factor portfolio for intraday, overnight and daily (intraday + overnight) returns. (2) Sharpe ratios of individual factors for intraday, overnight and daily returns.

Interestingly the risk premium earned intraday and overnight differs significantly for characteristic based and statistical factor portfolios. The four statistical factors earn the

[30]The factors are excess returns of traded assets and hence their risk premium can be calculated by their time-series mean. The price of risk for each factor depends on the correlation of this factor with the stochastic discount factor. If factors are uncorrelated, e.g. PCA based factors, and span the SDF, then the price of risk and the risk premium coincide. The Sharpe-ratio of the tangency portfolio based on the factors directly takes into account the correlation structure between the factors and hence can be viewed as an aggregate risk premium. For the following reasons I am not using a Fama-MacBeth type cross-sectional regression to estimate the factor risk premia. First, the mean excess returns of individual stocks seem to be very noisy over this short time horizon and the mean of factor portfolios seems to be more reliable. Second, in particular the Fama-French-Carhart factors require time-varying loadings, which can only be estimated with the high-frequency returns for short time horizons. However, the results on portfolios suggest that overnight returns contain most of the relevant loading information for the Fama-French-Carhart factors but the overnight observation frequency is to low to obtain reliable estimates of time-varying loadings for the individual stocks for short time horizons. Third, as the PCA based factors are orthogonal (respectively close to orthogonal depending on the return component and horizon) the cross-sectional regressions yield similar estimates of the risk premia as the factor means.

largest part of the risk premium intraday, which is not surprising as they are estimated to explain the intraday variation. On the other hand the size, value and momentum factors earn their positive risk compensation overnight while the intraday compensation is minor and negative.

Figure 10: Expected intraday and overnight returns



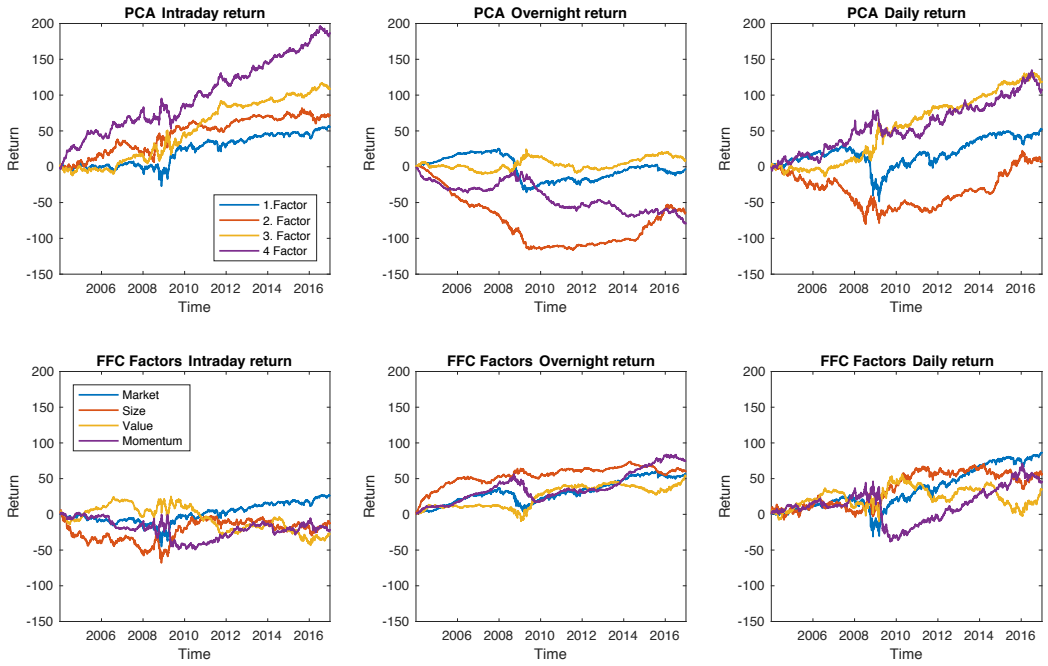
Note: Expected intraday and overnight excess returns from 2004 to 2016 ($T = 3273$) for balanced panel of all stocks ($N = 332$), 14 industry portfolios and 6 Fama-French size and value double-sorted portfolios (two size and three book-to-market quantiles)

The PCA factors have a significantly larger risk premium intraday which reverses overnight. In particular the second and fourth factor show a strong reversal pattern. Based on the intraday means and covariances respectively the overnight means and covariances, I construct the tangency portfolios for each time segment. The first two columns of Table 4 show that the intraday tangency portfolio of PCA factors has almost twice the Sharpe ratio as the daily return portfolio. On the other hand the overnight tangency portfolio of the Fama-French-Carhart factors has twice the Sharpe-ratio compared to the daily returns. This confirms again that the PCA factor are compensated for intraday risk while the characteristic based factors earn the risk compensation mainly overnight.

Figure 10 takes a closer look at the reversal pattern in stock returns and portfolios. The expected intraday excess returns are clearly negatively correlated with expected overnight returns. The same pattern holds for the expected returns of the 14 industry portfolios. The six size- and value-sorted portfolios only show a very weak reversal pattern. The overnight reversal pattern is captured by the continuous PCA factors. Figures 11 and **??** plot the cumulative return for the four statistical and characteristic based factors for daily returns and their intraday and overnight component. As expected the PCA factors show a strong reversal pattern, i.e. they earn higher average returns during the day but reverse the return

during the night resulting in a lower daily return. On the other hand the Fama-French-Carhart factors show a strongly increasing return pattern for size and value overnight which is lower for daily returns and close to zero during the day.

Figure 11: Normalized cumulative factor returns for intraday, overnight and daily returns



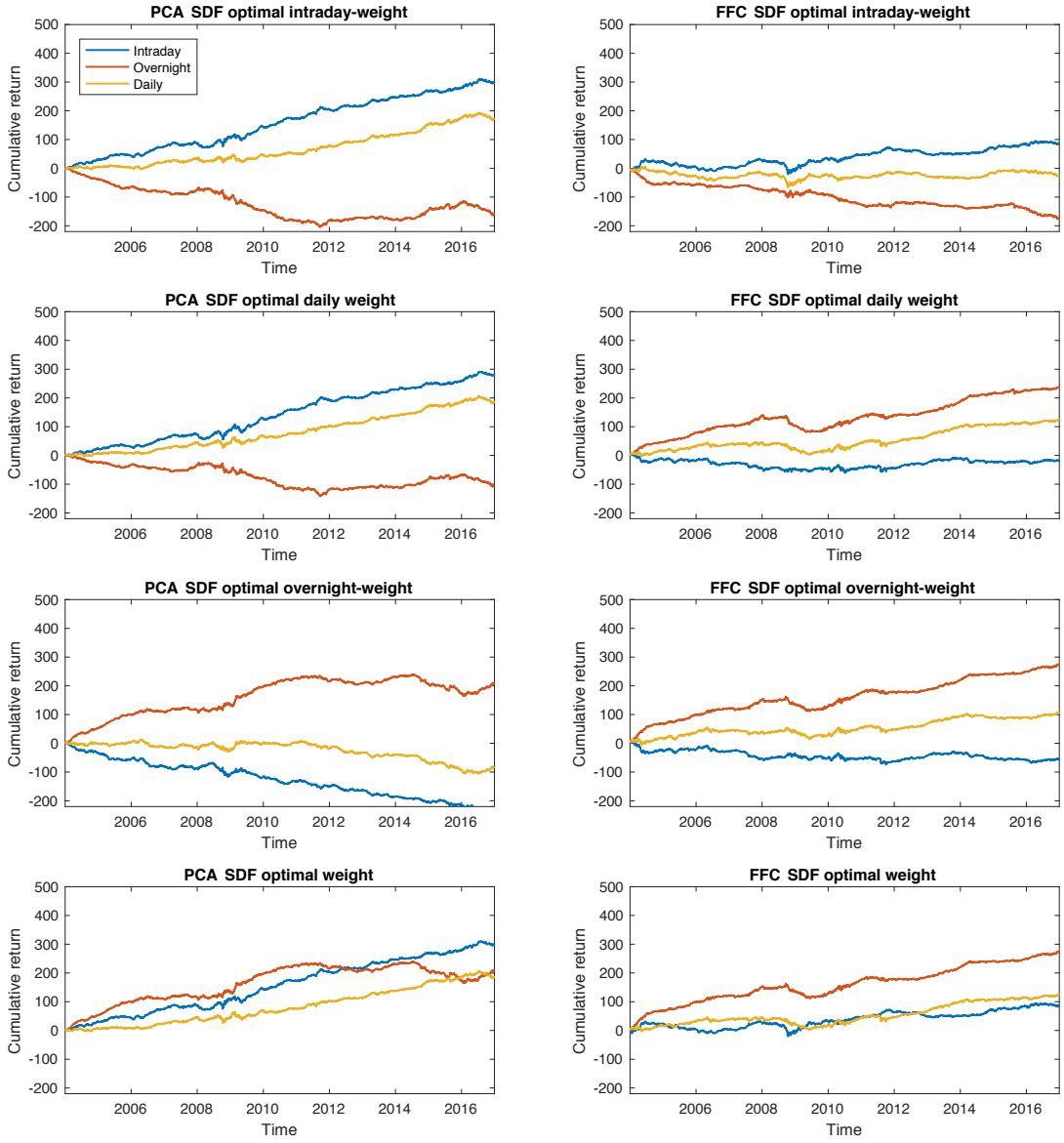
Note: Factors are normalized by their daily standard deviation.

In Figure 12 I study the realized returns for the optimal tangency portfolios for the two sets of factors. The optimal portfolio weights for the factors are either chosen based on the daily, intraday or overnight returns. The first plot shows the strong overnight reversal of the optimal PCA factor portfolio. The daily returns are lower because the overnight returns move into the opposite direction. On the other hand using only overnight returns yields the best portfolio for the Fama-French-Carhart factors. The two subplots in the fourth row show the optimal portfolio weights for each time horizon. As expected using daily returns is not optimal for either set of factors.

These observations suggest that a long-short strategy with intraday long position and overnight short position (respectively overnight long position and intraday short position for the characteristic based factors) in the tangency portfolios should perform better than keeping a simple long position in the optimal portfolios. Indeed, Figure 24 illustrates the impressive outperformance of this strategy for the PCA factors. The Fama-French-Carhart factors also outperform the daily returns if optimal overnight portfolio returns are chosen for

the long-short strategy. In the last row a long-short strategy is based on intraday returns with optimal intraday portfolio weights and overnight returns with optimal overnight portfolio weights, yielding the best possible risk-adjusted return for the factor portfolios.

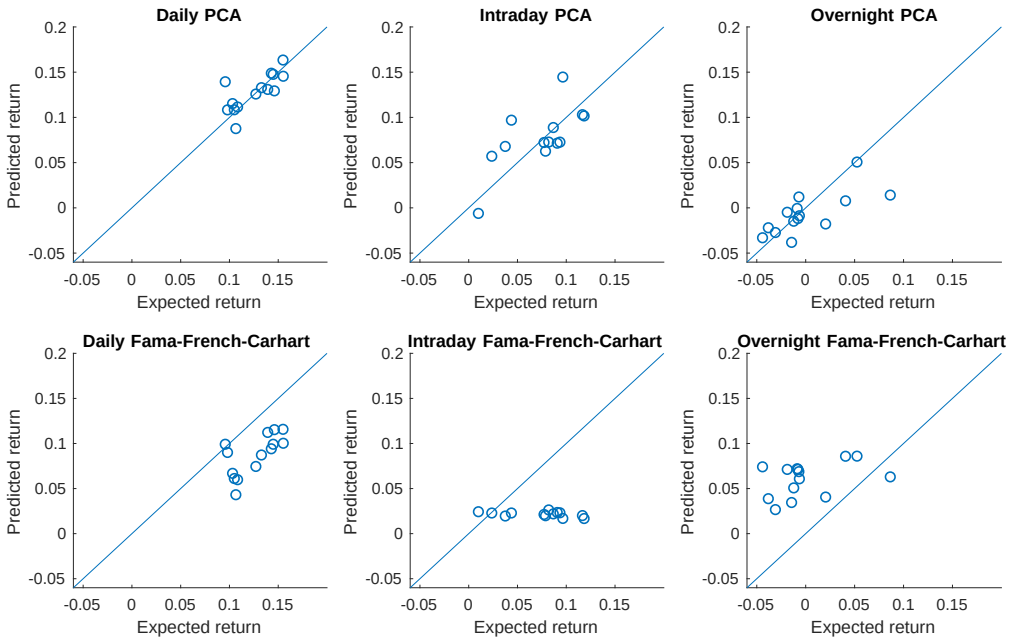
Figure 12: Cumulative returns of tangency portfolios



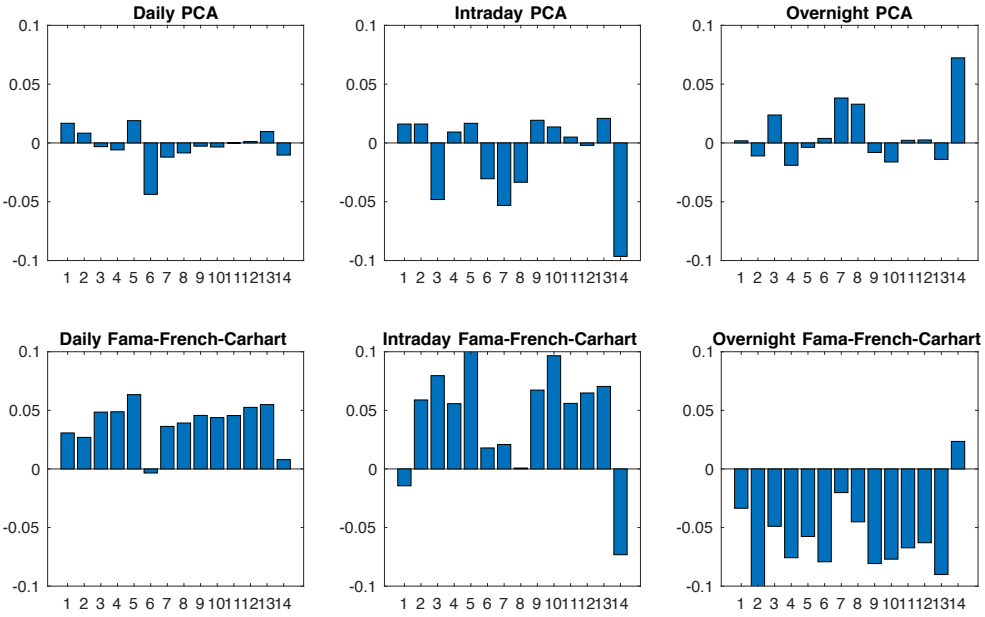
Note: Cumulative intraday, overnight and daily returns of optimal tangency portfolio implied by 4 continuous PCA factors and 4 Fama-French-Carhart factors. The returns are based on optimal portfolio weights calculated (1) based on intraday returns, (2) daily returns and (3) overnight returns. The last subplots show the intraday returns with optimal intraday portfolio weights, overnight returns with optimal overnight portfolio weights and similarly for daily returns. Returns are normalized by the daily standard deviation of the corresponding SDF portfolio.

Figure 13: Asset pricing of industry portfolios

Panel A: Predicted Returns



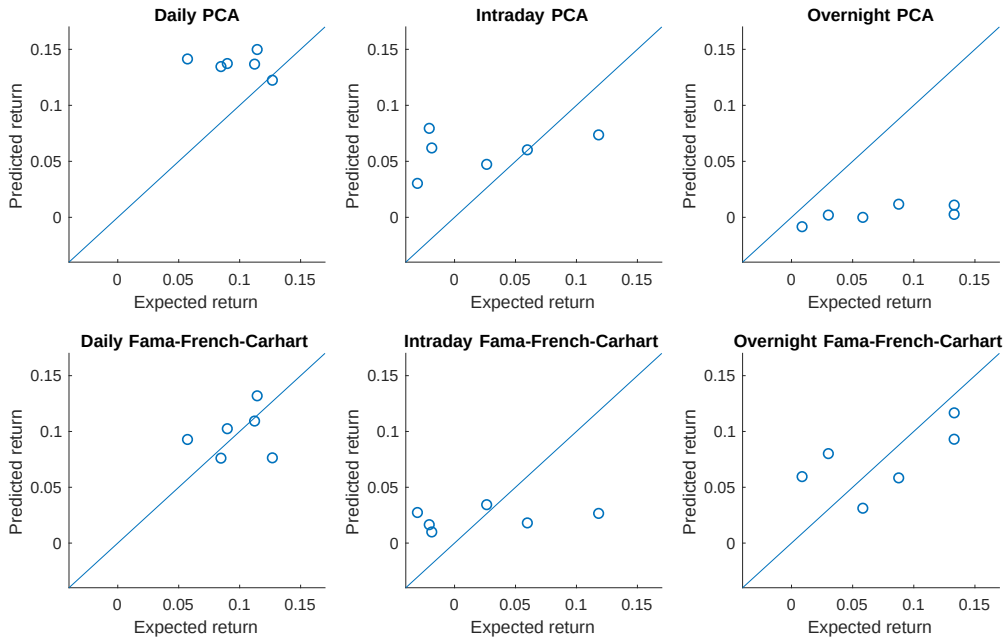
Panel B: Pricing Errors



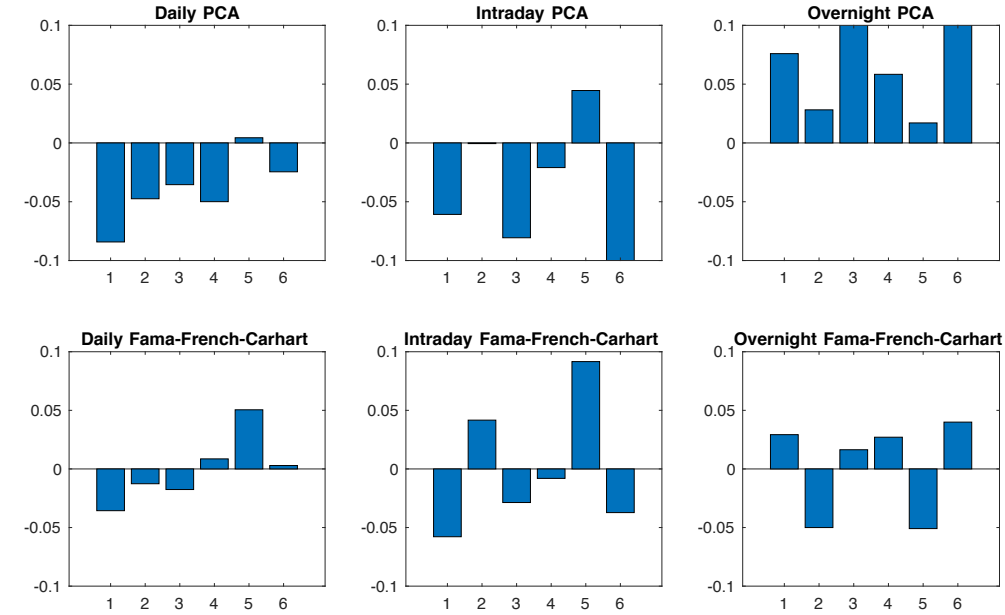
Note: Return prediction for 14 industry portfolios based on 4 continuous PCA factors and 4 Fama-French-Carhart factors. The predictive regression is based on daily, intraday or overnight data. Panel A shows the predicted and expected return. Panel B plots the time-series pricing errors.

Figure 14: Asset pricing of size- and value sorted portfolios

Panel A: Predicted Returns



Panel B: Pricing Errors



Note: Return prediction for 6 Fama-French size and value double-sorted portfolios (two size and three book-to-market quantiles) based on 4 continuous PCA factors and 4 Fama-French-Carhart factors. The predictive regression is based on daily, intraday or overnight data. Panel A shows the predicted and expected return. Panel B plots the time-series pricing errors.

How well can the statistical factors explain asset returns? I run time series regressions with intercept on different test assets for their daily, intraday and overnight returns for two sets of factors. As the means of individual stocks are estimated with a lot of noise I test 14 industry portfolios and the 6 size and value sorted portfolios. As the industry portfolios exhibit a strong overnight reversal, I expect the PCA factors to better price these assets. Indeed, Figure 25 shows that the four PCA factors explain the industry portfolios better than the 4 Fama-French-Carhart factors. Separating the returns into their intraday and overnight component, I show that the PCA factors succeed in explaining the intraday and overnight pattern. In contrast, the Fama-French-Carhart factors miss essentially the intraday risk-premium and assign a too large risk-premium to the overnight returns. As this over- and underestimation partially cancels out, the daily pricing errors are smaller than for the intraday and overnight component. In Figure 25 I repeat the same analysis with time-varying loadings and pricing errors using a 120 days moving window to estimate the intercepts and predicted returns which are then averaged over time. The result is virtually identical suggesting that time-variation in the loadings does not improve the weaker performance of the Fama-French-Carhart factors.

In Figure 25 I test the 6 size and value sorted portfolios. As expected these portfolios are better explained by the Fama-French-Carhart factors. In particular, the overnight returns are better explained by the anomaly factors compared to PCA. This suggests that the intraday co-movement does not capture the size and value information. Figure 26 estimates predicted returns and pricing errors with a 120 days moving window yielding similar results. The performance of the PCA factors does not improve with the time-varying loadings and pricing errors, while the Fama-French-Carhart factors have a slightly better pricing performance. This is in line with the results of section 4.5 showing that the characteristic based factors are less stable over time.

4.7 Microstructure Noise

Non-synchronicity and microstructure noise are two distinguishing characteristics of high-frequency financial data. First, the time interval separating successive trades can be random, or at least time varying. Second, the observations are subject to market microstructure noise, especially as the sampling frequency increases. The fact that this form of noise interacts with the sampling frequency distinguishes this from the classical measurement error problem in statistics. Inference on the volatility of a continuous semimartingale under noise contami-

nation can be pursued using smoothing techniques.[31] However, neither the microstructure robust estimators nor the non-synchronicity robust estimators can be easily extended to my large dimensional problem. The main results of my paper assume synchronous data with negligible microstructure noise. Using for example 5-minute sampling frequency as commonly advocated in the literature on realized volatility estimation, e.g. Andersen et al. (2001) and the survey by Hansen and Lunde (2006), seems to justify this assumption.

Table 5: Robustness against microstructure noise

	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
	Loadings based on all HF (continuous + jumps) data												
1. GC	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00
2. GC	1.00	0.98	0.99	0.99	0.99	1.00	0.97	0.99	0.99	0.99	1.00	1.00	1.00
3. GC	0.94	0.97	0.94	0.99	0.99	1.00	0.95	0.99	0.99	0.99	1.00	0.99	1.00
4. GC	0.94	0.97	0.94	0.98	0.99	0.99	0.95	0.99	0.99	0.99	0.98	0.99	1.00
	Loadings based on continuous data												
1. GC	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2. GC	1.00	1.00	0.99	0.99	0.99	1.00	1.00	0.99	1.00	0.99	0.99	1.00	1.00
3. GC	0.94	0.97	0.91	0.99	0.99	0.99	0.99	0.99	1.00	0.99	0.99	1.00	0.99
4. GC	0.94	0.97	0.91	0.98	0.99	0.99	0.99	0.99	0.98	0.99	0.99	1.00	0.99

Note: Generalized correlations of the loadings of the first four factors estimated on 5 minutes and 15 minutes log returns for $N = 332$ stocks.

My estimation results are robust to the sampling frequency. As lower sampling frequencies are less affected by microstructure noise, it suggests that microstructure noise does not affect my findings. Table 5 is a representative robustness test that compares the loadings estimated on 5 minutes and 15 minutes log returns. The loadings are virtually identical. The other main findings in this paper are the same for 15 minutes and 30 minutes log return and are available upon request.

5 Conclusion

My primary conclusions are four-fold. First, by estimating latent factors as opposed to relying on pre-specified observable factors I find a low dimensional model that explains the

[31] Several approaches have been developed, prominent ones by Aït-Sahalia and Zhang (2005) and Barndorff-Nielsen et al. (2008) in the one-dimensional setting and generalizations for a noisy non-synchronous multi-dimensional setting by Aït-Sahalia et al. (2010), Podolskij and Vetter (2009) and Barndorff-Nielsen et al. (2011) among others.

co-movement in intra-day stock returns. Time-variation in the factor structure matters, but is subtle. The span of the regression loadings for my high-frequency PCA factors is stable over time. This means projections on the systematic or idiosyncratic return component are the same if they are done locally or with loadings estimated over the whole time horizon. However, due to the time-varying volatility matrix of the factors, the contribution of a factor to the systematic component and the regression loadings on individual stocks vary over time. In contrast the Fama-French-Carhart factors have a time-varying span of the regression loadings and lead to biased pricing errors for stocks if they are not estimated locally. Second, the type of factors that are obtained by PCA based methods depends crucially on the underlying assets. Characteristic based factors are obtained by applying PCA to characteristic managed portfolios as in Kelly et al. (2017). However, most of the time variation in individual stocks returns is actually explained by industry factors. This raises the question how characteristics are related to covariances of individual stocks. Third, the systematic structure in smooth continuous intraday movements is different from rough intraday and overnight jumps. Fourth, the factors that explain the intraday co-movements in stocks earn a significant intraday risk premium. However, as this risk premium reverses overnight, the daily risk premium is comparatively low for these factors. This finding suggests that studying only the cross-section of daily returns might neglect risk-return trade-offs that have an overnight reversal. My findings have direct implications for investment strategies as there is a sizeable risk compensation from exploiting the reversal pattern.

References

Aït-Sahalia, Y., P. A. M., Zhang, L., 2005. How often to sample a continuous-time process in the presence of market microstructure noise. Review of Financial Studies 18, 351–416.

Aït-Sahalia, Y., Fan, J., Xiu, D., 2010. High-frequency estimates with noisy and asynchronous financial data. Journal of American Statistical Association 105, 1504–1516.

Aït-Sahalia, Y., Jacod, J., 2014. High-Frequency Financial Econometrics. New Jersey: Princeton University Press.

Aït-Sahalia, Y., Xiu, D., 2017a. Principal component analysis of high frequency data. Journal of American Statistical Association.

Aït-Sahalia, Y., Xiu, D., 2017b. Principal component estimation of a large covariance matrix with high-frequency data. Journal of Econometrics 201, 384–399.

Alexeev, V., Dungey, M., Yao, W., 2017. Time-varying continuous and jump betas: The role of firm characteristics and periods of stress. Journal of Empirical Finance 40, 1–19.

Andersen, T., Bollerslev, T., Diebold, F. X., Labys, P., 2001. The distribution of realized exchange rate volatility. Journal of the American Statistical Association 42, 42–55.

Andersen, T. G., Benzoni, L., Lund, J., 2002. An empirical investigation of continuous-time equity return models. Journal of Finance 57 (4), 1239–1284.

Back, K., 1991. Asset prices for general processes. Journal of Mathematical Economics 20, 371–395.

Bai, J., 2003. Inferential theory for factor models of large dimensions. Econometrica 71, 135–171.

Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191–221.

Bai, J., Ng, S., 2006. Evaluating latent and observed factors in macroeconomics and finance. Journal of Econometrics (131), 507–537.

Bali, T. T., Engle, R. F., Tang, Y., 2014. Dynamic conditional beta is alive and well in the cross-section of daily stock returns. Working paper.

Barndorff-Nielsen, O., Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society, 253–280.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., Shephard, N., 2008. Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. Econometrica 76, 1481–1536.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., Shephard, N., 2011. Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. Journal of Econometrics 162, 149–169.

Berkman, H., Koch, P., Ruttle, L., Zhang, Y., 47. Paying attention: Overnight returns and the hidden cost of buying at the open. Journal of Financial and Quantitative Analysis 715-741.

Bollerslev, T., Li, S. Z., Todorov, V., 2013. Jump tails, extreme dependencies and the distribution of stock returns. Journal of Econometrics 172, 307–324.

Bollerslev, T., Li, S. Z., Todorov, V., 2016. Roughing up beta: Continuous vs. discontinuous betas, and the cross section of expected stock returns. Journal of Financial Economics 120, 464–490.

Bollerslev, T., Todorov, V., 2011. Estimation of jump tails. Econometrica 79, 1727–1783.

Branch, B., Ma, A., 2012. Overnight return, the invisible hand behind intraday returns? Journal of Applied Finance 22 (3), 90–100.

Chamberlain, G., 1988. Asset pricing in multiperiod securities markets. Econometrica 56, 1283–1300.

Chamberlain, G., Rothschild, M., 1983. Arbitrage, factor structure, and mean-variance analysis on large asset markets. Econometrica 51, 1281–1304.

Christensen, K., Oomen, R. C. A., Podolskij, M., 2014. Fact or friction: Jumps at ultra high frequency. Journal of Financial Economics 114 (3), 576–599.

Cliff, M., Cooper, M., Gulen, H., 2008. Return differences between trading and non-trading hours: like night and day. Working paper.

Connor, G., 1984. A unified beta pricing theory. Journal of Economic Theory 34, 13–31.

Connor, G., Korajczyk, R., 1988. Risk and return in an equilibrium apt: Application to a new test methodology. Journal of Financial Economics 21, 255–289.

Connor, G., Korajczyk, R., 1993. A test for the number of factors in an approximate factor model,. Journal of Finance 58, 1263–1291.

Duffie, D., Pan, J., Singleton, K. J., 2000. Transform analysis and asset pricing for affine jump-diffusions. Econometrica 68 (6), 1343–1376.

Eraker, B., 2004. Do stock prices and volatility jump? Reconciling evidence from spot and option prices,. Journal of Finance 59, 1367–1404.

Eraker, B., Johannes, M., Polson, N., 2003. The impact of jumps in volatility and returns. Journal of Finance 58 (1269-1300).

Fama, E. F., French, K. R., 1992. The cross section of expected stock returns. The Journal of Finance 47 (2).

Fama, E. F., French, K. R., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

Fama, E. F., MacBeth, J. D., 1973. Risk, return and equilibrium: Empirical tests. THe Journal of Political Economy 81 (3), 607–636.

Fan, J., Furger, A., Xiu, D., 2016a. Incorporating global industrial classification standard into portfolio allocation: A simple factor-based large covariance matrix estimator with high frequency data. Journal of Business and Economic Statistics 34, 489–503.

Fan, J., Liao, Y., Wang, W., 2016b. Projected principal component analysis in factor models. The Annals of Statistics 44 (1), 219–254.

Fan, L., Liao, Y., Mincheva, M., 2013. Large covariance estimation by thresholding principal orthogonal complements. Journal of the Royal Statistical Society 75, 603–680.

Gabaix, X., 2012. Variable rare disasters: An exactly solved framework for ten puzzles in macrofinance. Quarterly Journal of Economics (127), 645–700.

Giglio, S., Xiu, D., 2017. Asset pricing with omitted factors. Working paper.

Hansen, P., Lunde, A., 2006. Realized variance and market microstructure noise. Journal of Business and Economic Statistics 24, 127–161.

Jagannathan, R., Wang, Z., 1996. The conditional capm and the cross-section of expected stock returns. Journal of Finance (51), 3–53.

Kelly, B., Pruitt, S., Su, Y., 2017. Instrumented principal component analysis. Working Paper.

Kozak, S., Nagel, S., Santosh, S., 2017. Shrinking the cross section. Working Paper, Chicago Booth.

Lettau, M., Ludvigson, S., 2001. Resurrecting the (c)capm: a cross-sectional test when risk premia are time-varying. Journal of Political Economy (109), 1238–1287.

Lettau, M., Pelger, M., 2018. Factors that fit the time-series and cross-section of stock returns. Working paper.

Nagel, S., 2013. Empirical cross-sectional asset pricing. Annual Review of Financial Economics, 5 (1), 167–199.

Onatski, A., 2010. Determining the number of factors from empirical distribution of eigenvalues. Review of Economic and Statistics 92, 1004–1016.

Pan, J., 2002. The jump risk premium implicit in options: Evidence from an integrated time-series study. Journal of Financial Economics (53), 3–50.

Pelger, M., 2019. Large-dimensional factor modeling based on high-frequency observations. Journal of Econometrics 208 (1), 23–42.

Pelger, M., Xiong, R., 2018. Interpretable proximate factors for large dimensions. Working paper.

Podolskij, M., Vetter, M., 2009. Bipower-type estimation in a noisy diffusion setting. Stochastic Processes and their Applications 11, 2803–2831.

Reisman, H., 1992. Intertemporal arbitrage pricing theory. The Review of Financial Studies 5 (1), 105–122.

Ross, S. A., 1976. The arbitrage theory of capital asset pricing. Journal of Economic Theory 13, 341–360.

Sharpe, W., 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. Journal of Finance 3 (19), 425–442.

6 Appendix

6.1 Data Description

I collect the price data from the WRDS TAQ Millisecond trades database for the time period 2004 to 2016. I construct the log-prices for 5 minutes sampling, which gives me on average 250 days per year with 77 daily increments.

The first observation is the volume-weighted trading prices in the exact second of 9:30:00. For the remaining 78 observations, the volume-weighted trading prices are calculated for each second, and then the last observations in each of the 300 seconds interval are taken. As for significant portion of the stocks there is not trade in the first seconds of the day, I start my sample at 9:35am.

I use the price of the trade at or immediately proceeding each 5-min mark. For each year I take the intersection of stocks traded each day with the stocks that have been in the S&P500 index at any point during 1993-2012 based on the Bloomberg terminal. This gives me a cross-section N of 555 to 667 firms for each year with an intersection of $N = 332$ for all 13 years. I apply standard data cleaning procedures:

- Delete all entries with a time stamp outside 9:30am-4pm
- Delete entries with a transaction price equal to zero
- Retain entries originating from a single exchange
- Delete entries with corrected trades and abnormal sale condition.
- Aggregate data with identical time stamp using volume-weighted average prices

In each year I eliminate stocks from my data set if any of the following conditions is true:

- All first 10 5-min observations are missing in any of the day of this year
- There are in total more than 50 missing values before the first trade of each day for this year

- There are in total more than 500 missing values in the year

Missing observations are replaced by interpolated values. If there is no available 5-min price of the same day before, the next available 5-min price is used. Otherwise, the last available 5-min price is used. As my estimators are based on increments, the interpolated values will result in increments of zeros, which do not contribute to the quadratic covariation.

Based on SIC codes I classify stocks into 14 different industries. Table 6 lists the classifications and number of stocks based on the balanced panel.

Table 6: Industry portfolios

Industry	Number of firms	SIC Codes
Oil and gas	23	1200, 1221, 1311, 1381, 1382, 1389, 2870, 2911, 3533, 4922, 4923, 4932, 4924
Finance	56	6020, 6021, 6029, 6035, 6036, 6099, 6111, 6141, 6159, 6162, 6189, 6199, 6282, 6311, 6331, 6351, 6798, 6022, 6330, 6711, 6211, 6321, 6324, 6411, 6722, 6200, 6726
Electricity	13	4911, 4931, 4991
Technology	48	3571, 3577, 3674, 3572, 2559, 3825, 3578, 3651, 3663, 3670, 3672, 3674, 3678, 3679, 3761, 3827, 7370, 7372, 7373, 7374, 7375, 7379
Food	10	2032, 2041, 2043, 2066, 2099, 2111, 2086
Manufacturing	11	2320, 2439, 2493, 2499, 2515, 2531, 2621, 2679, 2711, 2329
Pharmaceuticals and chemicals	21	2834, 2841, 2842, 2844, 2851, 2879, 2830, 2813, 2869, 2812
Primary manufacturing	16	3011, 3089, 3199, 3221, 3275, 3312, 3317, 3324, 3334, 3357, 3411, 3423, 3429, 3334, 3021
Machinery	35	3511, 3519, 3519, 3519, 3533, 3550, 3559, 3561, 3562, 3579, 3621, 3633, 3711, 3714, 3714, 3714, 3721, 3721, 3721, 3728, 3810, 3822, 3825, 3826, 3829, 3873, 3942, 3944, 3812
Health	8	3840, 3841, 3845
Transportation	18	4011, 4213, 4481, 4492, 4512, 4512, 4513, 4730, 4813, 4813, 4813, 4841, 4841, 4841, 4899
Trade	40	5013, 5047, 5063, 5122, 5149, 5200, 5211, 5231, 5311, 5330, 5331, 5411, 5521, 5531, 5650, 5651, 5700, 5731, 5810, 5812, 5812, 5912, 5912, 5944, 5990, 5999, 5661, 5940
Services	31	1531,1731,4950, 7011, 7311, 7322, 7323, 7359, 7363, 7389, 7513, 7812, 7841, 7990, 7999, 8062, 8062, 8071, 8093, 8200, 8700, 8711, 8731, 8741
Mining	2	1041,1442

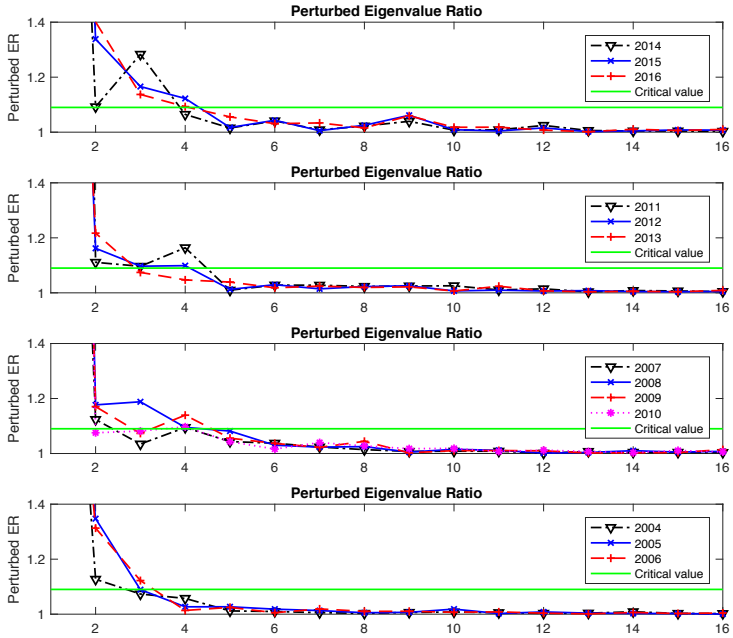
Note: Composition of 14 industry portfolios for balanced panel of $N = 332$ stocks.

I construct the Fama-French-Carhart factors and characteristics sorted portfolios based on all stocks in the WRDS TAQ Millisecond trades database that satisfy the above requirements, i.e. it includes also the stocks that have not been part of the S&P 500 index.[32]

[32]The characteristic sorted portfolio are labeled: 1=small growth, 2= small neutral, 3= small value, 4= big growth, 5= big neutral, 6=big value.

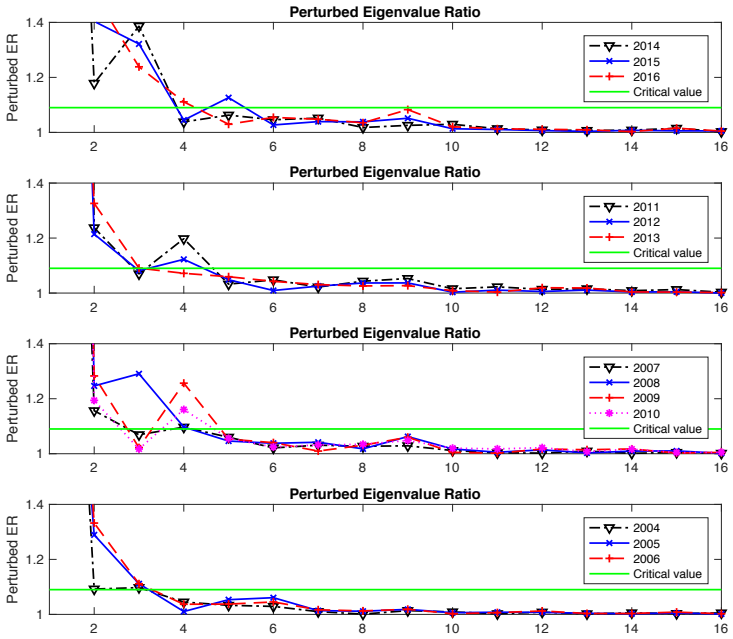
6.2 Number of Factors

Figure 15: Number of continuous factors



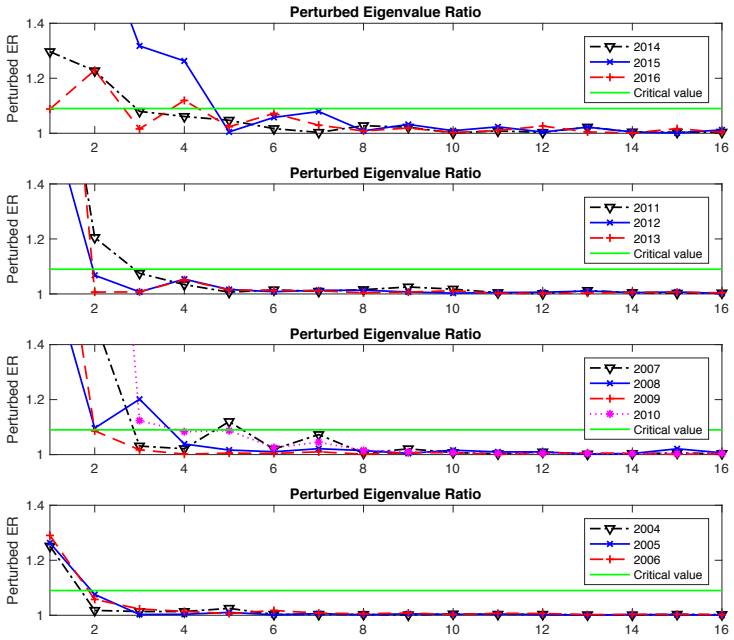
Note: Perturbed eigenvalue ratio test statistic for balanced panel of continuous returns.

Figure 16: Number of continuous factors



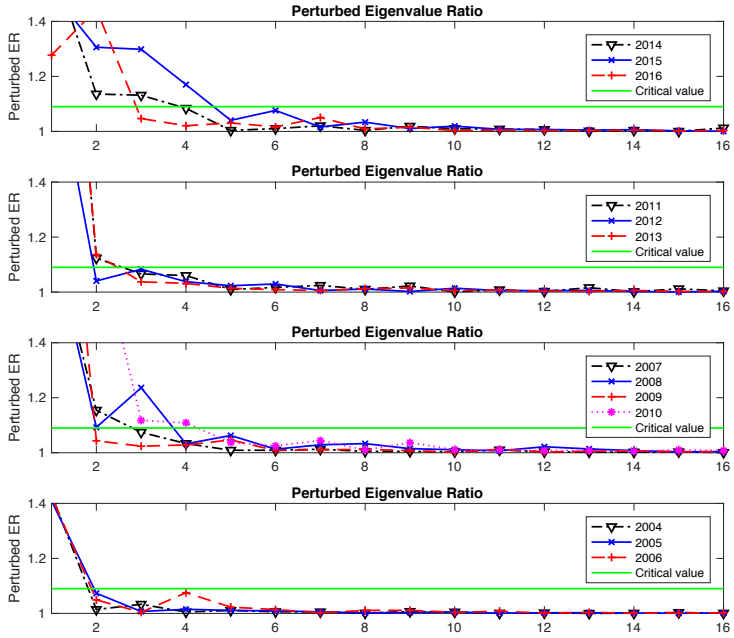
Note: Perturbed eigenvalue ratio statistic for unbalanced panel of all continuous returns.

Figure 17: Number of jump factors ($a = 3$)



Note: Perturbed eigenvalue ratio test statistic for balanced panel of jump returns ($a = 3$).

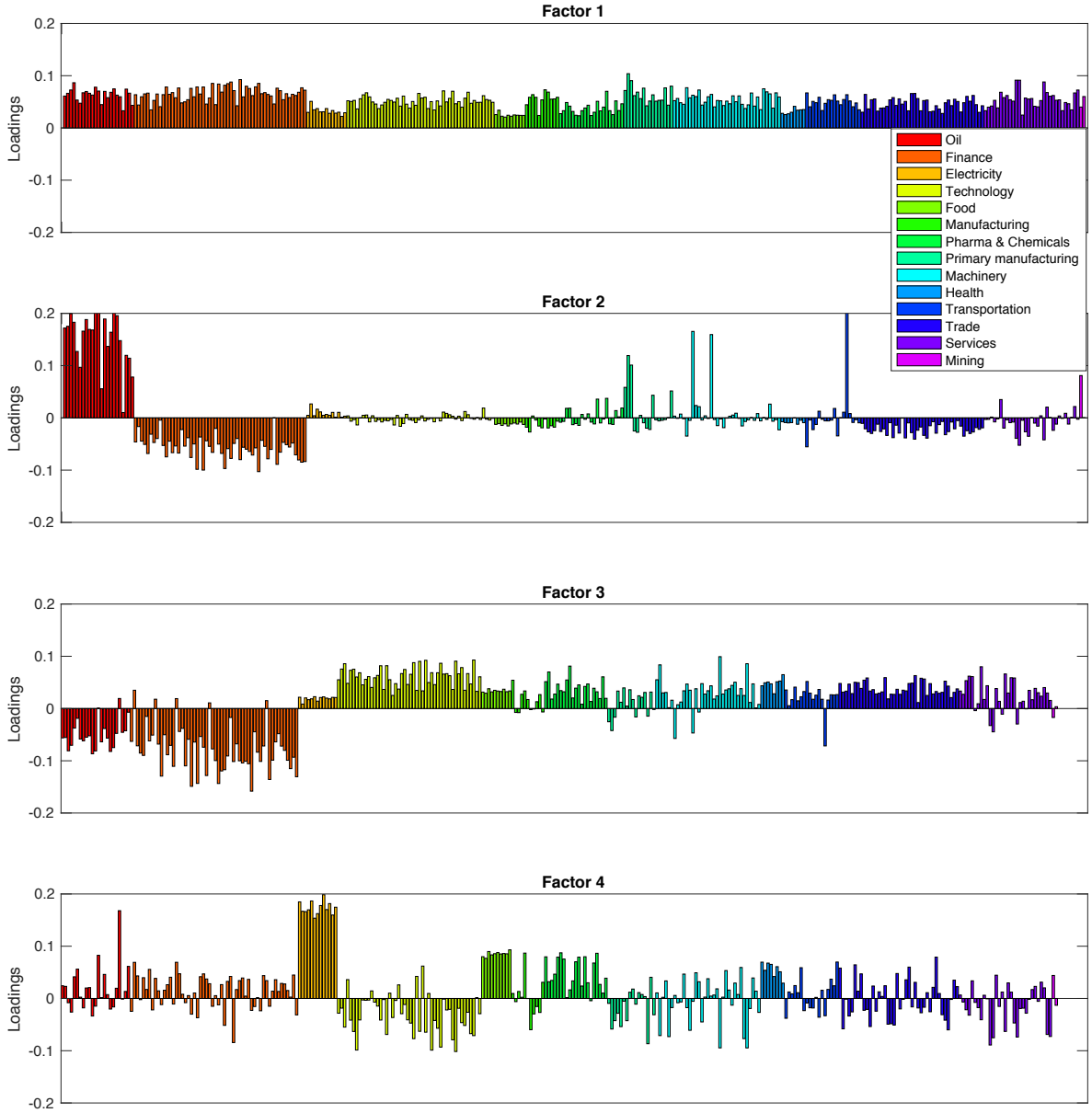
Figure 18: Number of jump factors ($a = 3$)



Note: Perturbed eigenvalue ratio statistic for unbalanced panel of all jump returns ($a = 3$).

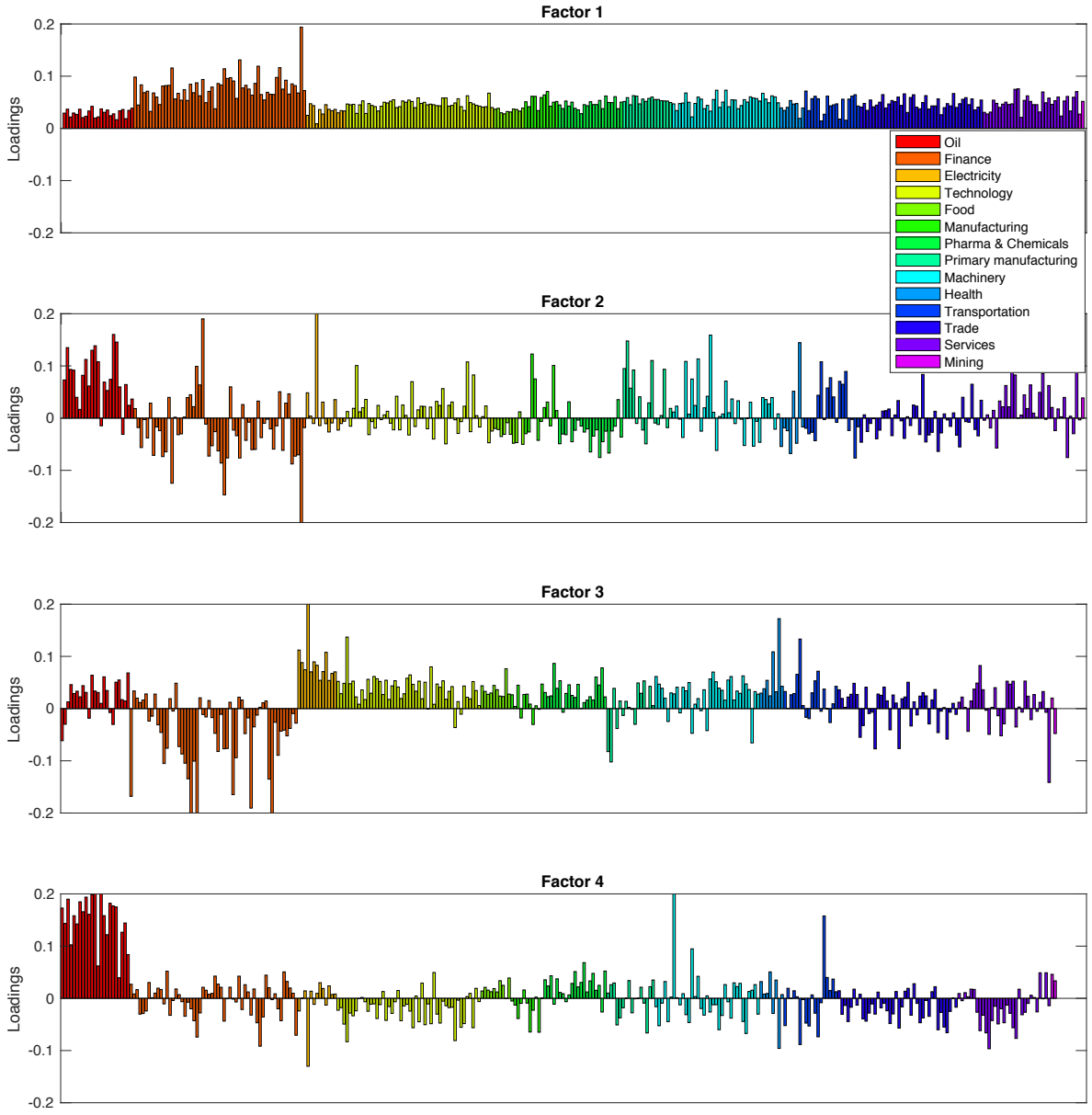
6.3 Composition of Factors

Figure 19: Portfolio weights of 4 high-frequency (continuous +jump) PCA factors



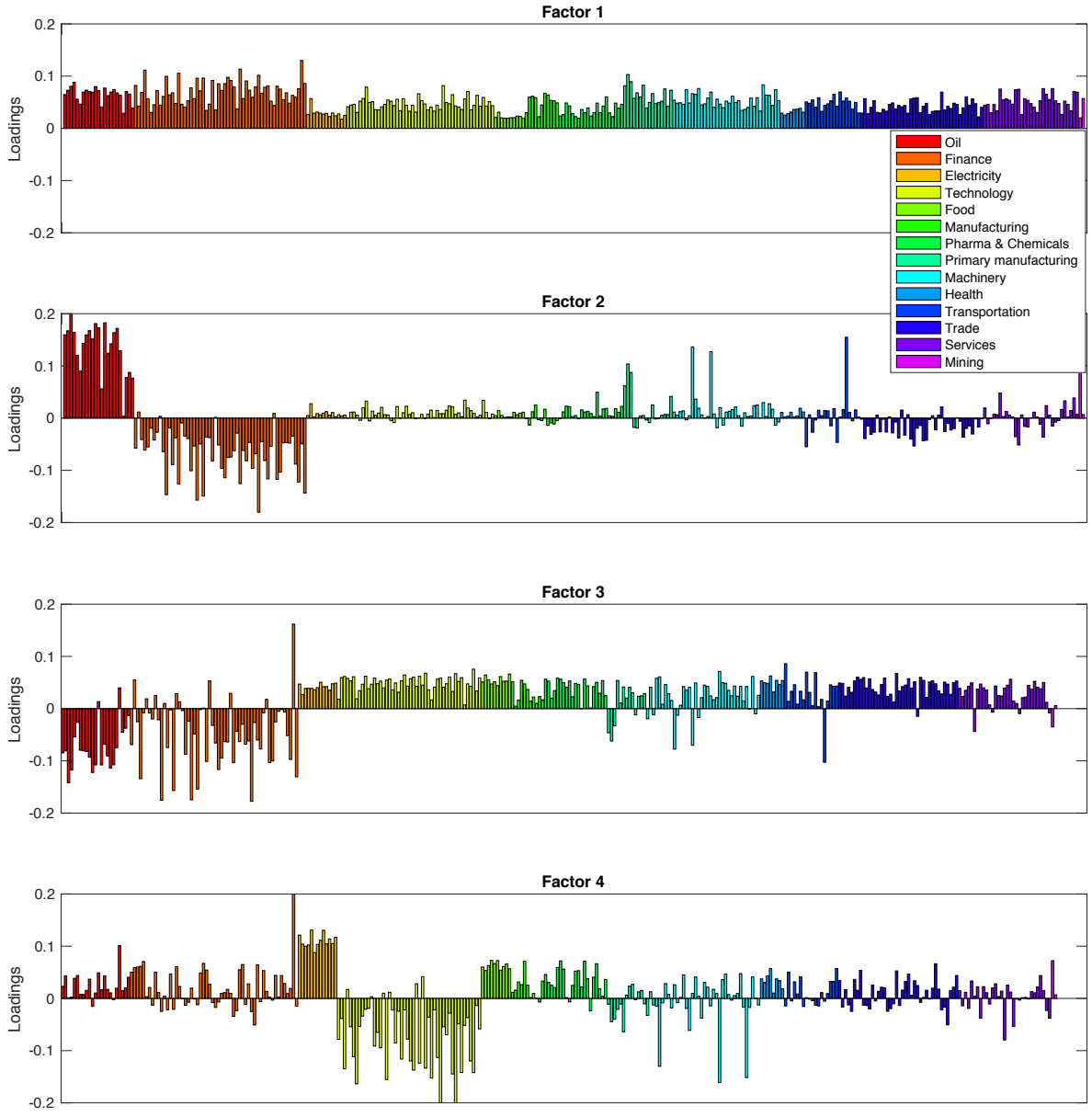
Note: Portfolio weights of 4 HF PCA factors over the whole time horizon 2004 to 2016. Stocks are sorted according to industry.

Figure 20: Portfolio weights of 4 jump PCA factors.



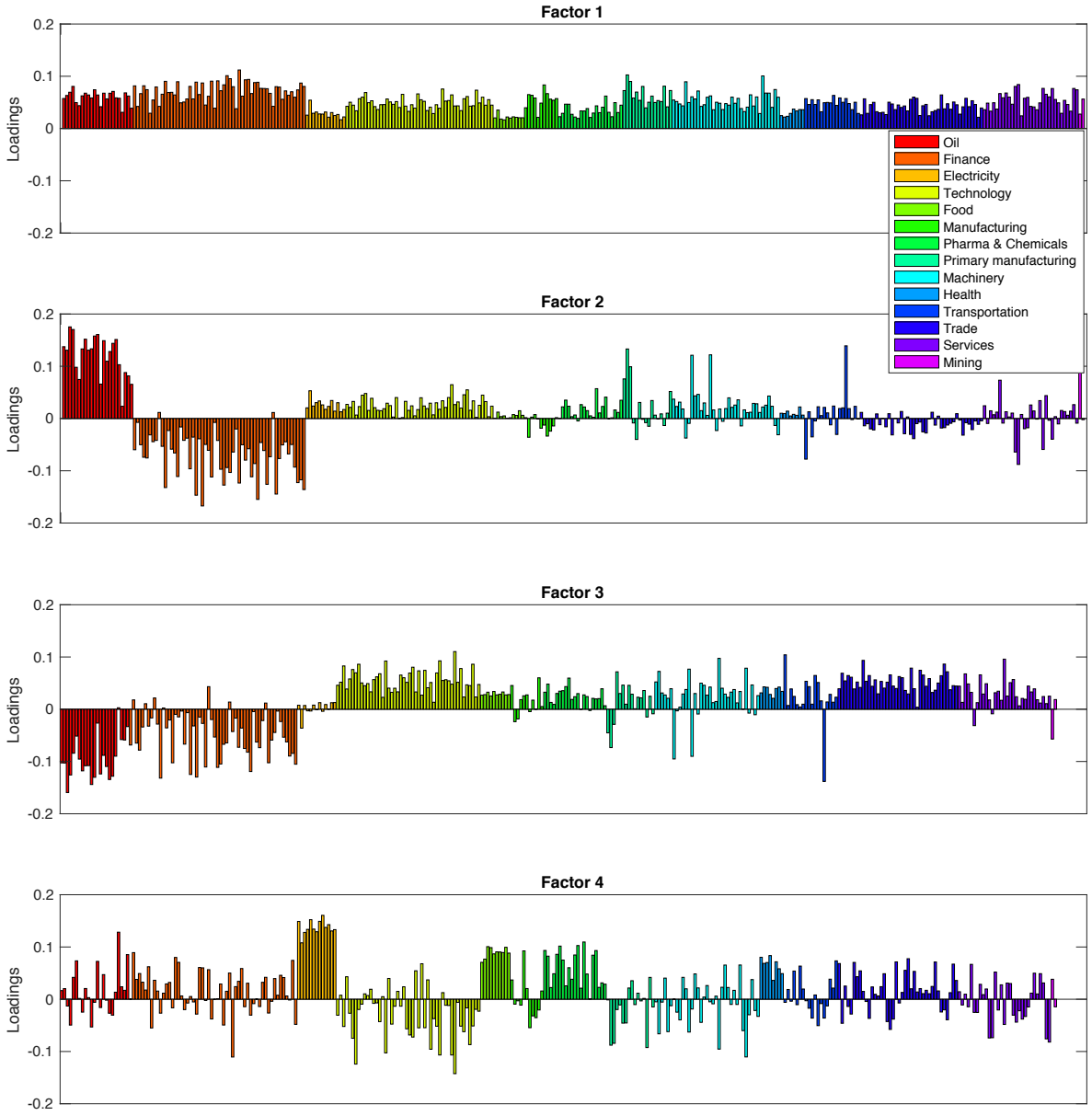
Note: Portfolio weights of 4 jump PCA factors ($a = 3$) over the whole time horizon 2004 to 2016. Stocks are sorted according to industry.

Figure 21: Portfolio weights of 4 overnight PCA factors.



Note: Portfolio weights of 4 overnight PCA factors over the whole time horizon 2004 to 2016. Stocks are sorted according to industry.

Figure 22: Portfolio weights of 4 daily PCA factors.



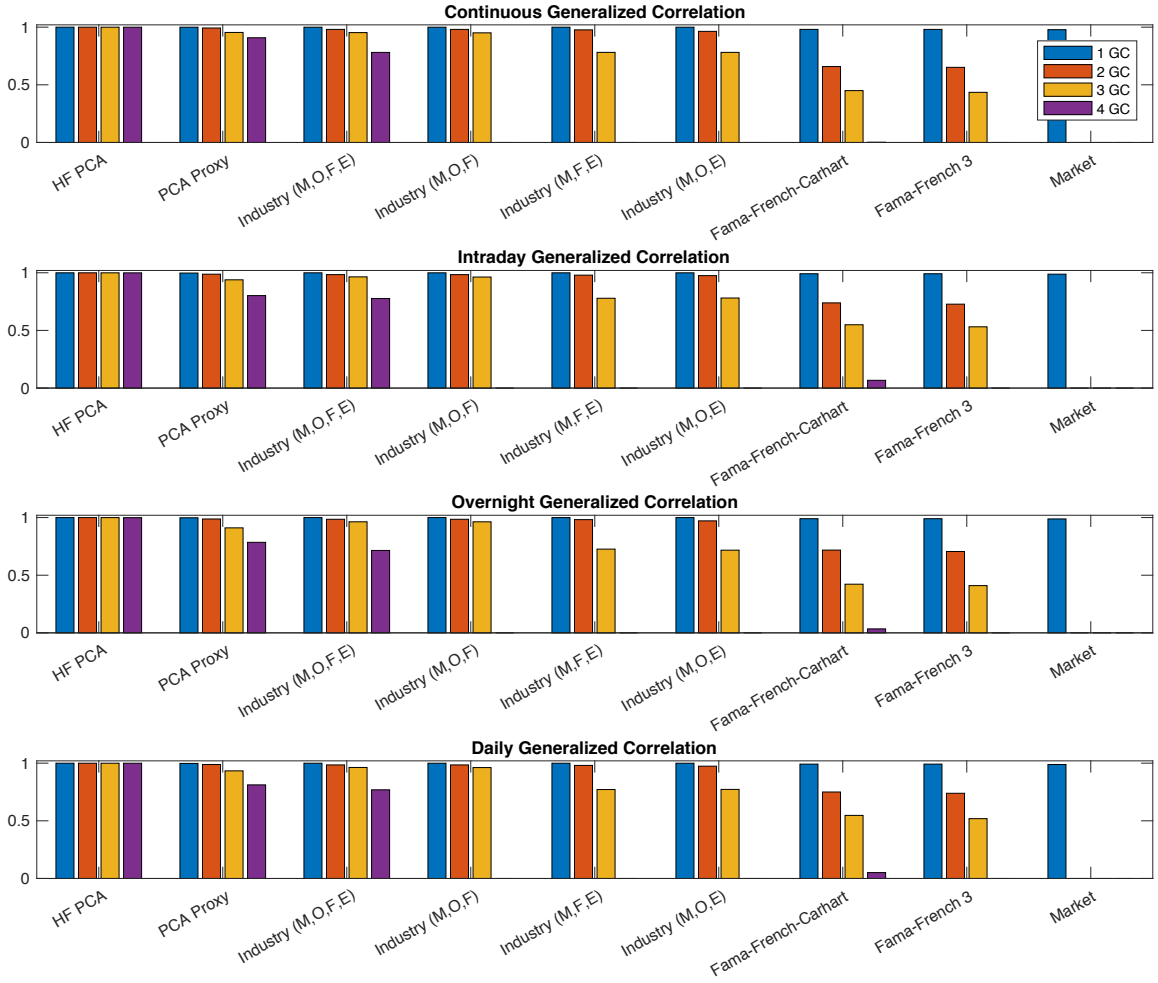
Note: Portfolio weights of 4 daily PCA factors over the whole time horizon 2004 to 2016. Stocks are sorted according to industry.

Table 7: Comparison of continuous factors with other statistical factors over time

	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
	High-frequency (continuous + jumps) factor portfolio weights												
1. GC	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2. GC	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
3. GC	1.00	0.99	1.00	1.00	1.00	1.00	0.97	1.00	1.00	1.00	1.00	1.00	1.00
4. GC	1.00	0.99	1.00	1.00	0.99	1.00	0.97	1.00	1.00	1.00	1.00	1.00	1.00
	Daily factor portfolio weights												
1. GC	0.98	0.98	0.97	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
2. GC	0.86	0.84	0.89	0.89	0.93	0.91	0.90	0.84	0.88	0.94	0.94	0.92	0.89
3. GC	0.86	0.84	0.89	0.57	0.86	0.91	0.86	0.84	0.88	0.77	0.94	0.89	0.89
4. GC	0.03	0.62	0.77	0.36	0.28	0.91	0.86	0.82	0.84	0.77	0.78	0.85	0.69
	Jump factor portfolio weights ($a = 3$)												
1. GC	0.94	0.94	0.96	0.96	0.90	0.95	0.91	0.96	0.96	0.89	0.96	0.95	0.92
2. GC	0.89	0.94	0.43	0.46	0.19	0.76	0.39	0.59	0.41	0.89	0.60	0.57	0.92
3. GC	0.89	0.84	0.43	0.26	0.19	0.76	0.18	0.59	0.41	0.65	0.51	0.36	0.81
4. GC	0.31	0.42	0.20	0.26	0.19	0.63	0.18	0.14	0.05	0.65	0.06	0.01	0.81
	Jump factor portfolio weights ($a = 4$)												
1. GC	0.78	0.84	0.61	0.89	0.71	0.78	0.79	0.85	0.74	0.92	0.69	0.88	0.87
2. GC	0.18	0.19	0.37	0.57	0.14	0.21	0.20	0.31	0.30	0.38	0.69	0.68	0.66
3. GC	0.06	0.10	0.11	0.06	0.14	0.21	0.14	0.16	0.30	0.38	0.09	0.07	0.66
4. GC	0.01	0.05	0.03	0.06	0.01	0.11	0.02	0.00	0.18	0.19	0.02	0.07	0.50
	Jump factor portfolio weights ($a = 4.5$)												
1. GC	0.59	0.59	0.42	0.80	0.20	0.61	0.78	0.77	0.68	0.90	0.64	0.78	0.78
2. GC	0.13	0.19	0.42	0.56	0.20	0.25	0.17	0.22	0.27	0.38	0.51	0.54	0.63
3. GC	0.03	0.04	0.13	0.12	0.16	0.15	0.10	0.15	0.27	0.38	0.16	0.09	0.33
4. GC	0.03	0.01	0.13	0.05	0.01	0.06	0.06	0.01	0.14	0.12	0.01	0.04	0.20
	Jump factor portfolio weights ($a = 5$)												
1. GC	0.47	0.29	0.54	0.82	0.26	0.23	0.59	0.66	0.59	0.84	0.41	0.60	0.57
2. GC	0.21	0.08	0.21	0.41	0.23	0.21	0.25	0.26	0.05	0.27	0.28	0.60	0.57
3. GC	0.07	0.04	0.14	0.04	0.23	0.21	0.08	0.08	0.05	0.27	0.15	0.11	0.40
4. GC	0.00	0.04	0.09	0.04	0.05	0.20	0.08	0.08	0.05	0.10	0.06	0.02	0.15

Note: Generalized correlations of the portfolio weights of the first four largest yearly continuous factors with the weights of the first four largest HF, daily and jump factors and jump factors estimated yearly from 2004 to 2016 for $N = 332$ stocks.

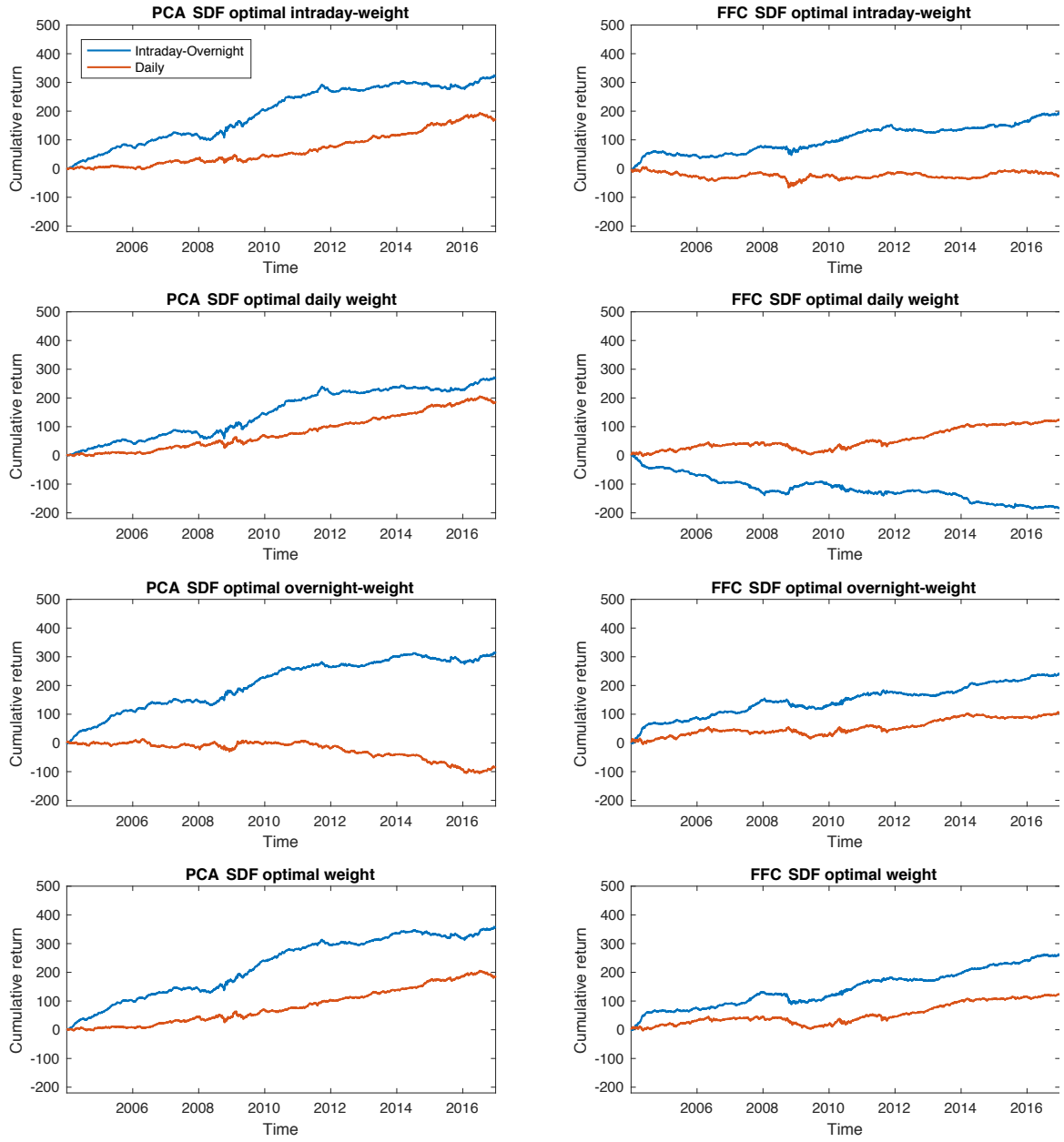
Figure 23: Generalized correlations of 4 continuous PCA factors with other factors.



Note: Generalized correlations of the first four statistical continuous PCA factors with the 4 PCA factors based on all HF data (continuous+jumps), the 4 continuous PCA proxy factors, different combinations of industry factors (market, oil, finance and electricity), the 4 Fama-French-Carhart factors, the 3 Fama-French factors and the market factor. Generalized correlations are based on continuous, intraday, overnight and daily returns, i.e. given the portfolio weights of a factor I calculate continuous, intraday, overnight and daily returns for this factor.

6.4 Asset Pricing

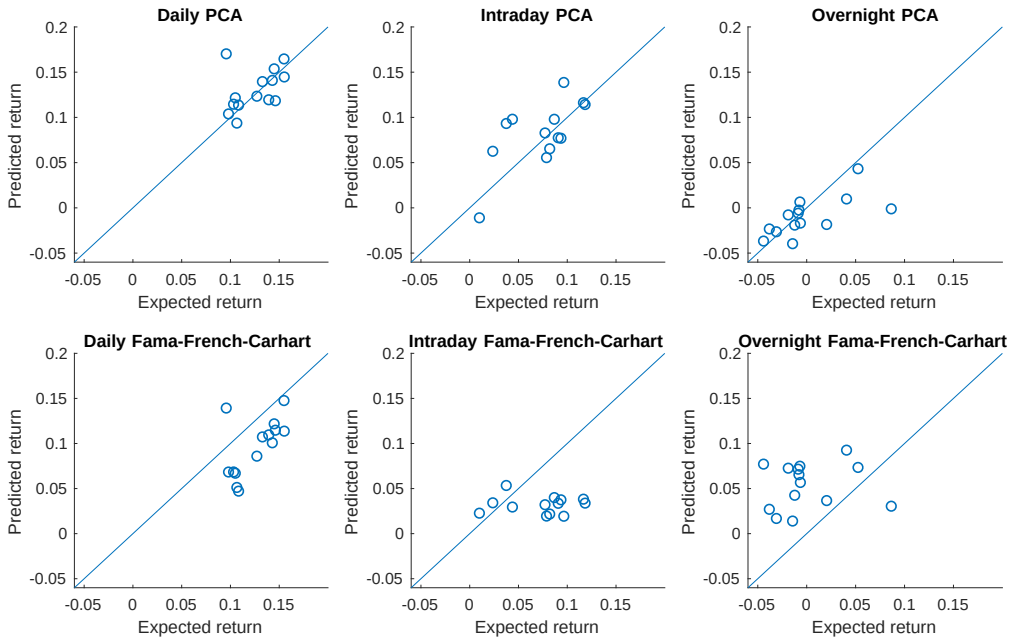
Figure 24: Cumulative returns of long-short intraday and overnight tangency portfolios



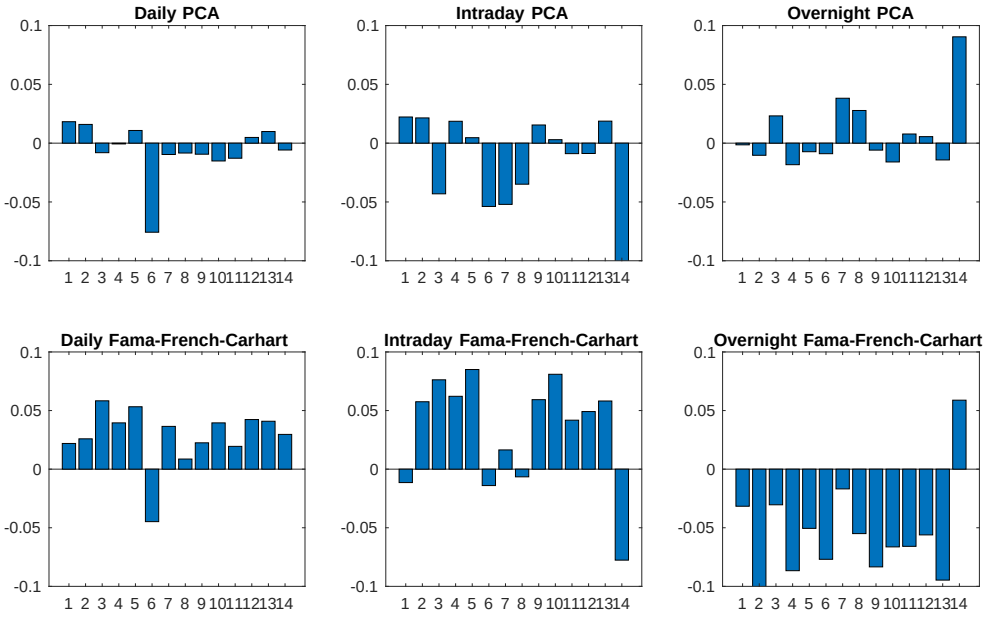
Note: Cumulative daily returns of optimal tangency portfolios and long-short strategy with intraday long position and overnight short position implied by 4 continuous PCA factors and 4 Fama-French-Carhart factors. The optimal portfolio weights are calculated (1) based on intraday returns, (2) daily returns and (3) overnight returns. The last subplots uses intraday returns with optimal intraday portfolio weights and overnight returns with optimal overnight portfolio weights. Returns are normalized by the standard deviation of the daily portfolio returns.

Figure 25: Asset pricing of industry portfolios on moving window

Panel A: Predicted Returns



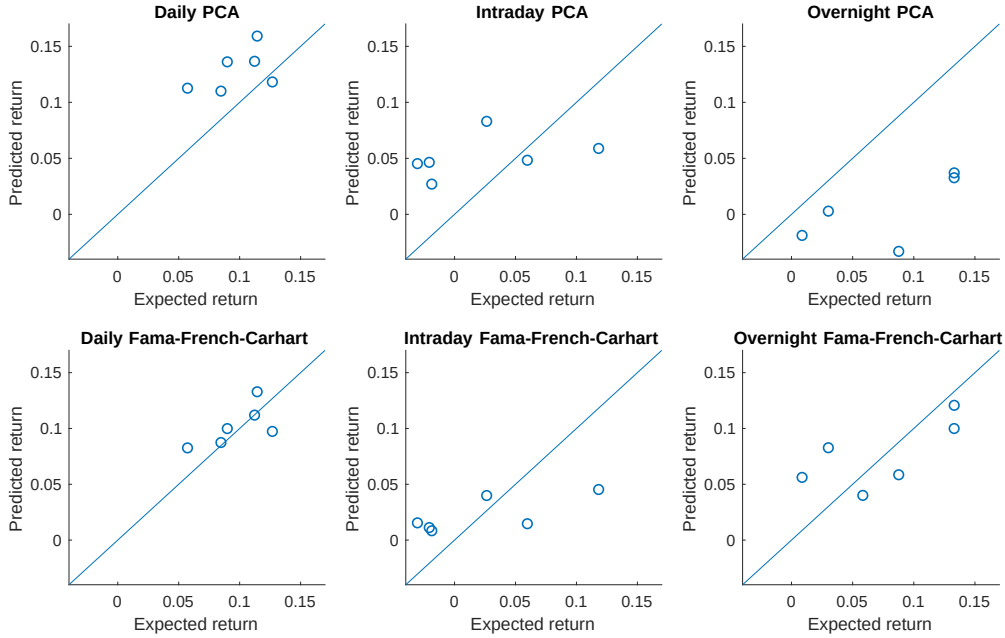
Panel B: Pricing Errors



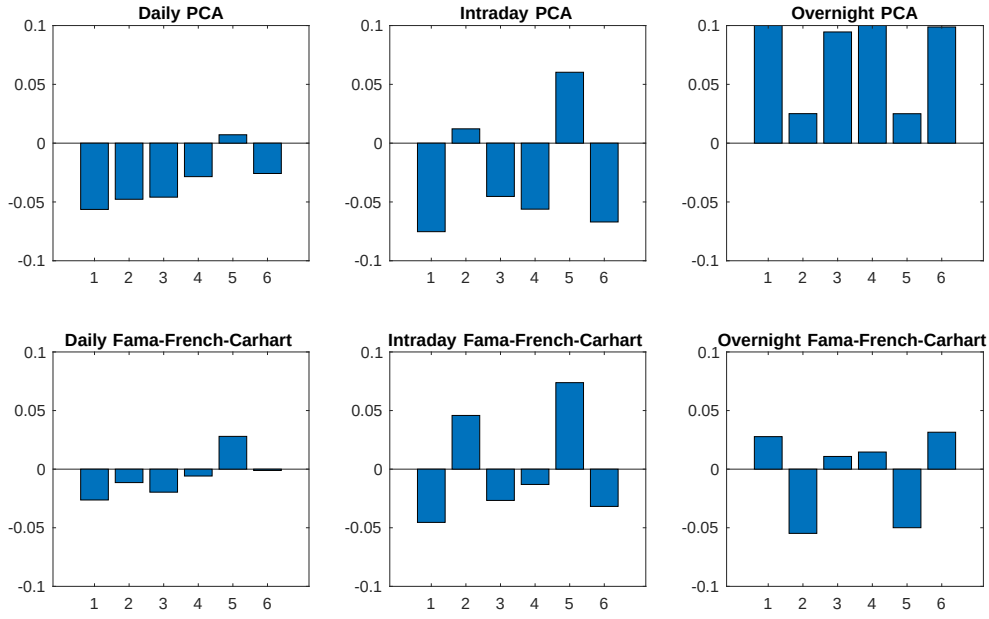
Note: Return prediction for 14 industry portfolios based on 4 continuous PCA factors and 4 Fama-French-Carhart factors. The predictive regression is based on daily, intraday or overnight data. Panel A shows the predicted and expected return. Panel B plots the time-series pricing errors. The time-series regression are run on a moving window of 120 days.

Figure 26: Asset pricing of size- and value sorted portfolios on moving window

Panel A: Predicted Returns



Panel B: Pricing Errors



Note: Return prediction for 6 Fama-French size and value double-sorted portfolios based on 4 continuous PCA factors and 4 Fama-French-Carhart factors. The predictive regression is based on daily, intraday or overnight data. Panel A shows the predicted and expected return. Panel B plots the time-series pricing errors. The time-series regression are run on a moving window of 120 days.

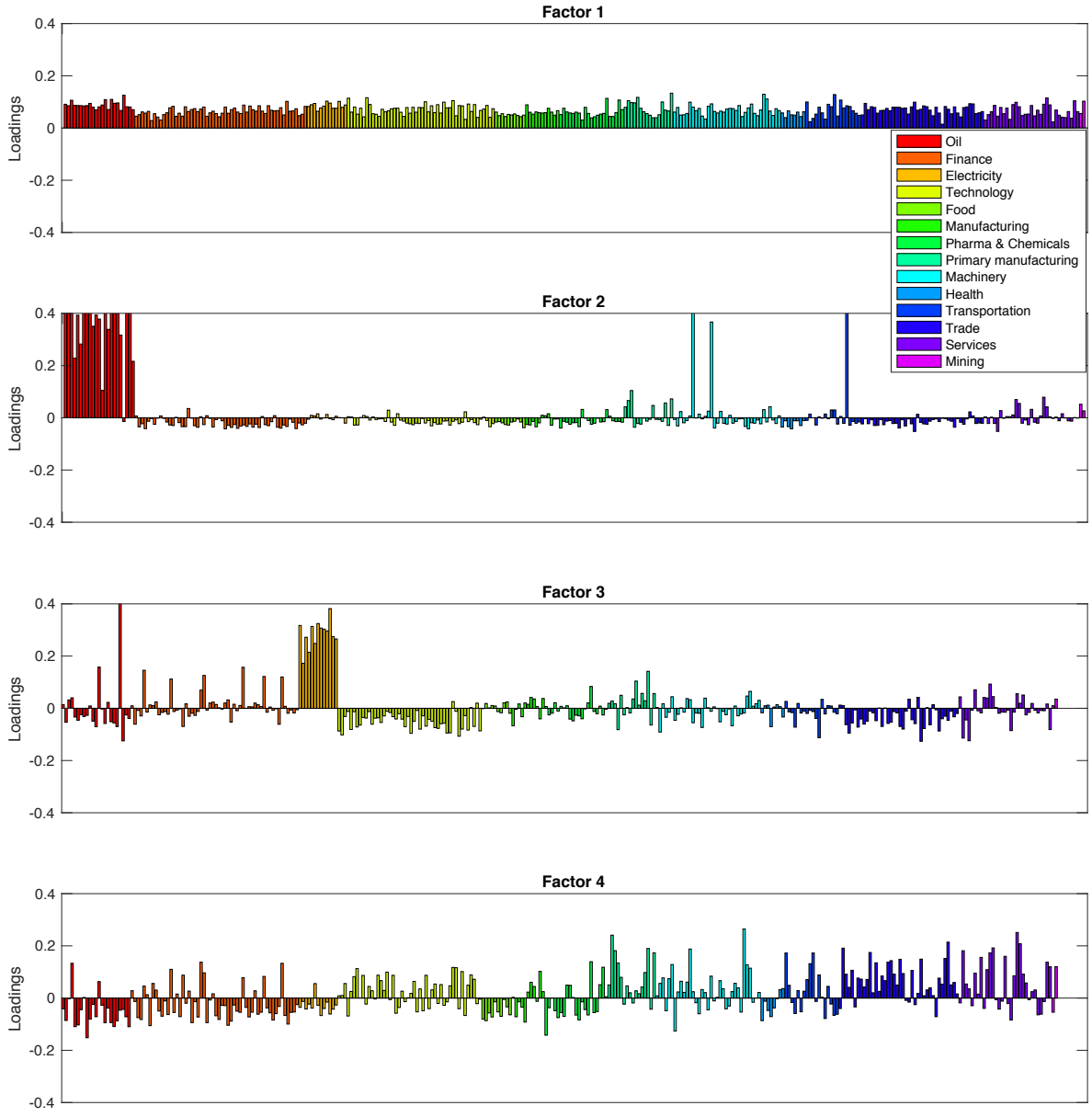
6.5 Time-Variation

Table 8: Stability of continuous and jump factors over time

	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
$\hat{K}$	3	3	3	4	4	4	4	4	4	3	3	3	3
	Continuous factors ($a = 3$)												
1. GC	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
2. GC	0.99	0.99	0.99	0.98	0.98	0.97	0.96	0.97	0.96	0.97	0.97	0.98	0.97
3. GC	0.95	0.95	0.92	0.95	0.96	0.95	0.95	0.95	0.95	0.94	0.97	0.96	0.96
4. GC	0.91	0.63	0.74	0.93	0.89	0.94	0.90	0.93	0.92	0.24	0.63	0.89	0.90
	Jump factors ($a = 3$)												
1. GC	0.93	0.94	0.96	0.97	0.96	0.97	1.00	0.99	0.97	0.99	0.97	1.00	0.95
2. GC	0.44	0.47	0.84	0.86	0.92	0.78	0.99	0.91	0.86	0.83	0.96	0.98	0.88
3. GC	0.15	0.12	0.78	0.13	0.88	0.42	0.94	0.63	0.78	0.12	0.70	0.96	0.07
4. GC	0.05	0.10	0.00	0.02	0.63	0.02	0.84	0.07	0.01	0.04	0.06	0.03	0.02
	Jump factors ($a = 4$)												
1. GC	0.84	0.85	0.96	0.97	0.94	0.95	1.00	0.95	0.96	0.99	0.98	1.00	0.50
2. GC	0.21	0.76	0.94	0.92	0.89	0.85	1.00	0.85	0.89	0.82	0.93	0.97	0.12
3. GC	0.04	0.04	0.84	0.07	0.86	0.79	0.98	0.36	0.49	0.17	0.67	0.88	0.11
4. GC	0.01	0.02	0.06	0.00	0.53	0.47	0.97	0.07	0.01	0.10	0.00	0.03	0.00
	Jump factors ($a = 4.5$)												
1. GC	0.75	0.94	0.97	0.98	0.89	0.96	1.00	0.93	0.95	0.99	0.98	0.97	0.24
2. GC	0.08	0.46	0.96	0.23	0.81	0.95	1.00	0.81	0.90	0.90	0.79	0.94	0.10
3. GC	0.02	0.05	0.87	0.11	0.53	0.87	0.99	0.44	0.36	0.10	0.59	0.16	0.00
4. GC	0.00	0.00	0.70	0.00	0.12	0.21	0.99	0.30	0.02	0.06	0.01	0.06	0.00
	Jump factors ($a = 5$)												
1. GC	0.16	0.95	0.98	0.98	0.87	0.97	1.00	0.93	0.95	0.99	0.98	0.99	0.02
2. GC	0.05	0.06	0.97	0.13	0.78	0.97	1.00	0.79	0.92	0.92	0.82	0.93	0.00
3. GC	0.01	0.00	0.81	0.05	0.64	0.80	0.99	0.45	0.42	0.23	0.08	0.07	0.00
4. GC	0.00	0.00	0.33	0.00	0.17	0.27	0.99	0.43	0.02	0.03	0.00	0.02	0.00

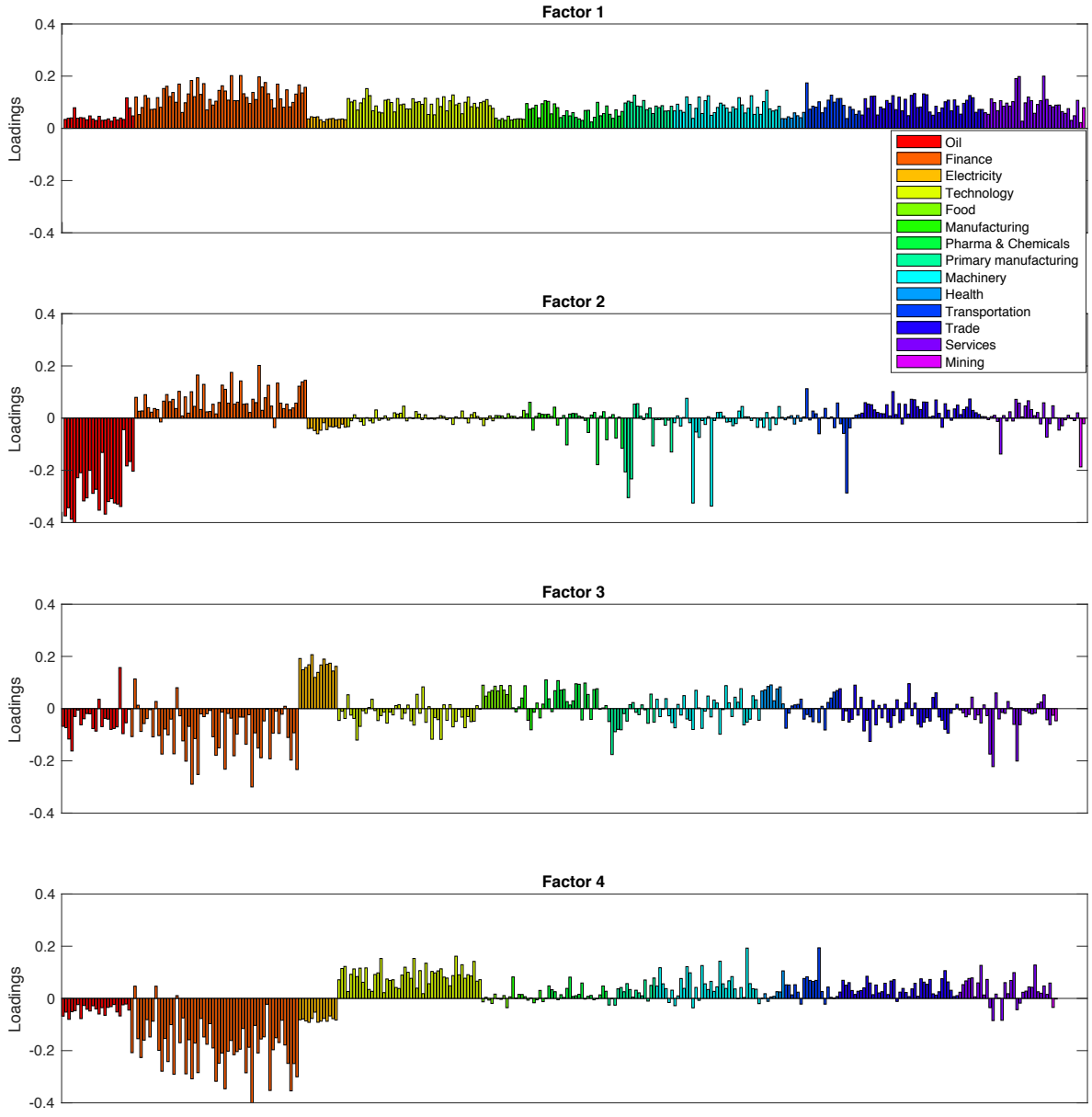
Note: Generalized correlations of the first four largest yearly continuous and jump factors with the first four statistical continuous and jump factors estimated from 2004 to 2016 for $N = 332$ stocks.

Figure 27: Time-varying loadings of 4 continuous PCA factors.



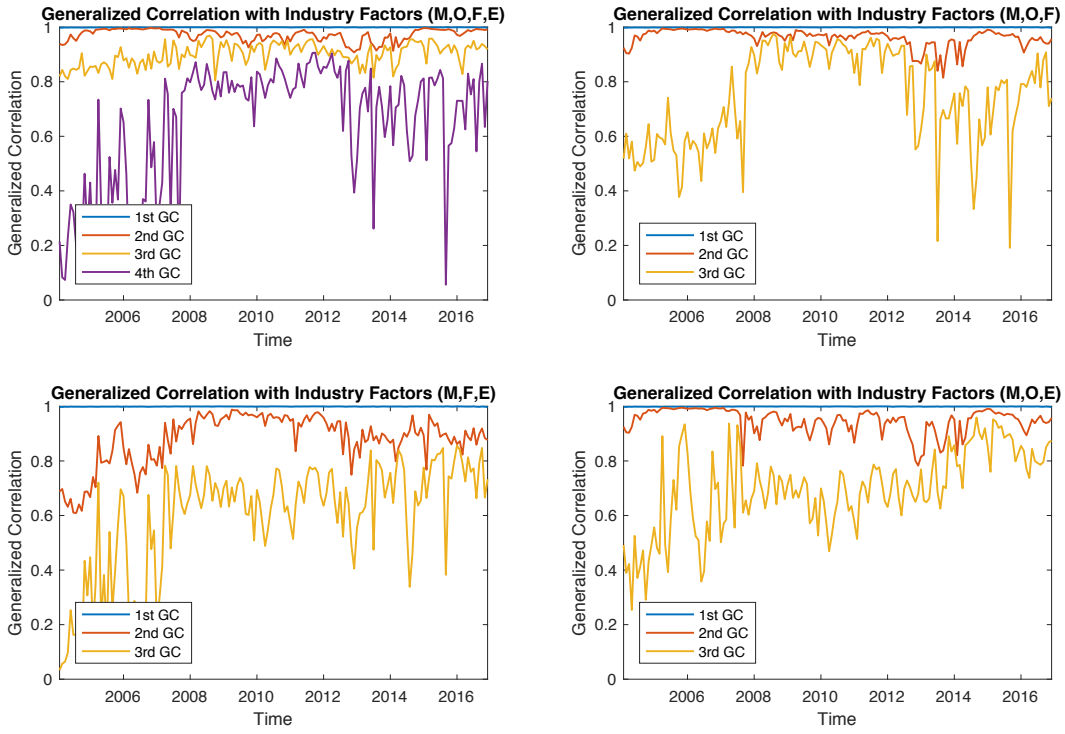
Note: Loadings of 4 continuous PCA factors in November 2006.

Figure 28: Time-varying loadings of 4 continuous PCA factors.



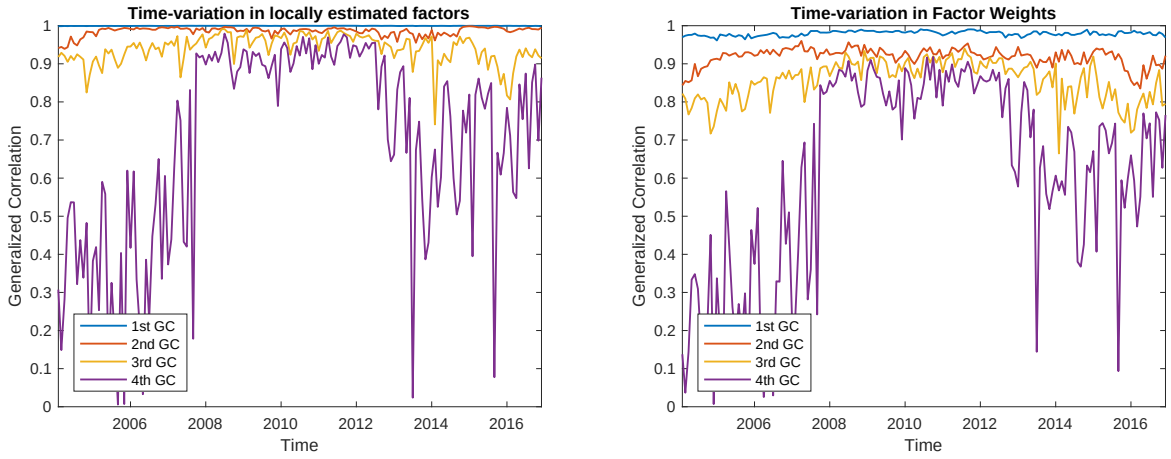
Note: Loadings of 4 continuous PCA factors in April 2008.

Figure 29: Time-variation for 4 continuous PCA factors



Note: Generalized correlations between the 4 continuous PCA factors estimated locally on a moving window of one month (21 trading days) and with continuous industry factors (market (M), oil (O), finance (F) and energy (E)).

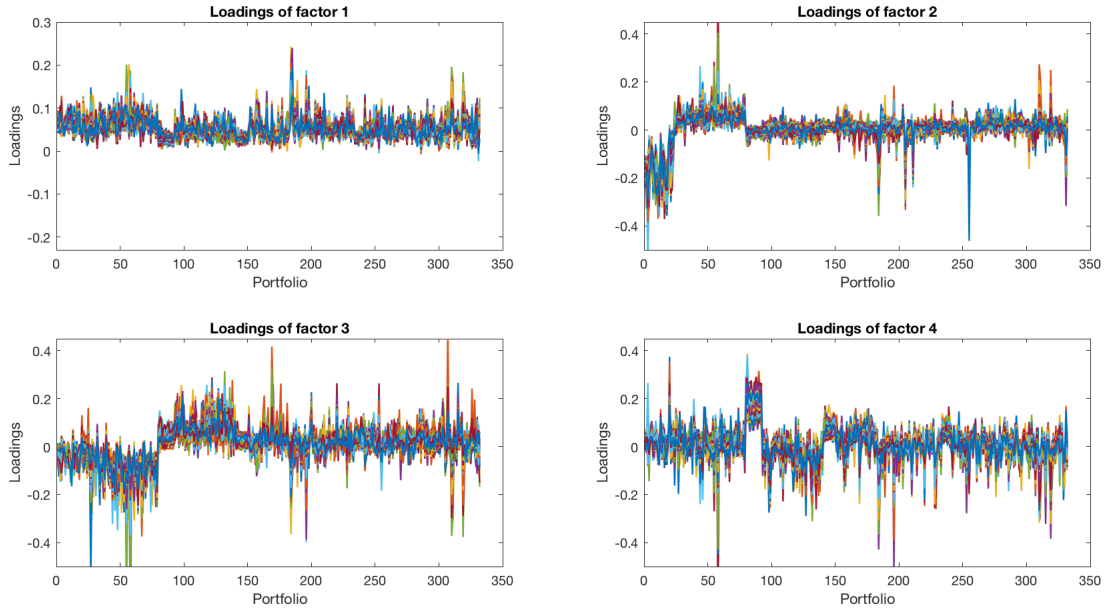
Figure 30: Time-variation in locally estimated continuous factors



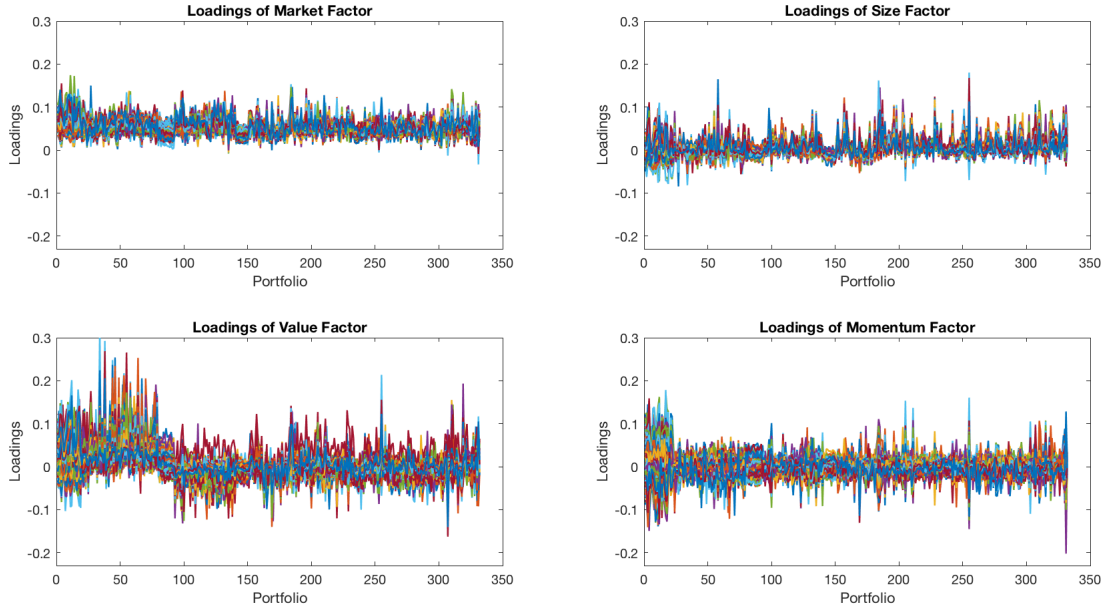
Note: Left panel: Generalized correlations of 4 continuous PCA factors estimated on the whole time horizon and on a moving window of one month (21 trading days). Right panel: Generalized correlations of factor portfolio weights estimated locally and the whole time horizon.

Figure 31: Time-variation in the continuous loadings for different factors

Panel A: 4 Continuous PCA factors



Panel B: 4 Fama-French-Carhart factors



Note: Loadings are estimated on a moving window of one month (21 trading days) based on continuous returns. Different lines corresponds to different months. Panel A: 4 Continuous PCA factors. Panel B: 4 Fama-French-Carhart factors.