Intraday Variation in Systematic Risks and Information Flows

Andrew J. Patton
Duke University

Yasin Simsek
Duke University

SoFiE Summer School

July 2024

Motivation

Factor models in asset pricing

- Factor models are fundamental tools for understanding risks in asset prices:

$$r_{i,t} = \beta_i^T \mathbf{f}_t + \varepsilon_{i,t}$$

- Most common approach is to use "*observable*" factors
  - e.g., Sharpe (1964, JF), Lintner (1965, JF), Fama and French (1993, JFE), Carhart (1997, JF), Fama and French (2015, JFE) and many more
  - *factor zoo* by Cochrane (2011, JF)
- An alternative to using a pre-specified set of factors is to extract them from the data using principal component analysis (PCA)
  - Connor and Korajczyk (1986, JFE), Kelly et al. (2019, JFE), Pelger (2020, JF), Lettau and Pelger (2020, RFS)

Motivation

Fixed vs. time-varying factor models

- Standard implementations of factor models, both observable and latent, assume that the factor exposures are **fixed** over time.
- A battery of evidence for
  - time-varying volatility; Bollerslev et al.(1994)
  - time-varying correlation; Engle(2009)
  - time-varying risk-premia; Jagannathan and Wang (1996, JF)

★ These motivate allowing for **time-varying** factor exposures.

What we do 1/2

High-frequency variation in factor structures

- Time-varying factor exposures have already received some attention:
  - Connor et al. (2012, *ECMA*), Gagliardini et al. (2016, *ECMA*, Kelly et al. (2019, *JFE*)
  - These papers consider variation at **lower** frequencies, daily and monthly.
- Recent works suggest that factor exposures vary **within the trading day.**
  - Andersen et al.(2021, *QE*; 2023 *JoE*); Liao and Todorov (2024, *QE*)

★ Motivated by this empirical evidence, we propose a new framework to allow for *intraday variation in latent factor models*.

- We combine PCA with non-parametric kernel methods.
- Generalizing the standard PCA of Connor and Korajczyk (1986, JFE)

$\Longrightarrow$ **Intraday PCA**

What we do 2/2

Summary of main findings

1. We estimate new factor model (*Intraday PCA*) on 15-minute data for more than 400 US stock returns over the period from Jan 1996 to Dec 2020.
   - superior explanatory power relative to well-known observable factor models and standard PCA
2. Then, we present a stylized model of asset prices and information flows that reflects our key empirical results.
3. We further investigate new factor model around
   - Earnings announcements
   - FOMC announcements

   and argue that information flows can generate this intraday variation.

Related Literature

- **Principal component analysis for economic data:**
  - Chamberlain and Rothschild (1983, *ECMA*), Connor and Korajczyk (1986, *JFE*), Stock and Watson (2002, *JASA*), Bai (2003, *ECMA*), Bai and Ng (2013, *JoE*)
- **High-frequency factor models:**
  - Ait-Sahalia and Xiu (2019, *JASA*), Pelger (2020, *JF*), Andersen et al. (2021, *QE*), Andersen et al. (2023, *JoE*), Liao and Todorov (2023, *wp*)
- **Information flows and asset prices:**
  - Fleming, et al. (1998, *JFE*), Kodres and Pritsker (2002, *JF*), Yuan (2005, *JF*), Pasquariello (2007, *RFS*), Patton and Verardo (2012, *RFS*), Ben-Rephael et al. (2020, *JF*), Andrei et al. (2023, *JFE*)

Outline

1. **Local PCA for high-frequency asset returns**
   - Evidence of time-varying factor structures
2. A stylized model of information flows and factor structures
   - Empirical predictions
3. Do information flows cause changes in factor structures?
   - Earnings announcements
   - FOMC announcements
4. Conclusion

PCA for asset returns

- Suppose that we have N assets observed in T days and M intraday periods per day.
- We assume that returns follow an approximate factor structure with K factors:

$$r_{i,t-1+\tau} = \underset{(1\times K)}{\boldsymbol{\beta}_i^\top} \underset{(K\times 1)}{\mathbf{f}_{t-1+\tau}} + \varepsilon_{i,t-1+\tau}$$

$$\underset{(TM\times N)}{\mathbf{R}} = \underset{(TM\times K)}{\mathbf{F}} \underset{(K\times N)}{\mathbf{B}^T} + \underset{(TM\times N)}{\varepsilon}$$

where $i \in \{1, ..., N\}$, $t \in \{1, ..., T\}$ and $\tau \in \{1, ..., M\}$

PCA for asset returns

- PCA provides estimates for factor and loadings:

$$\widehat{\mathbf{F}}, \widehat{\mathbf{B}} = \arg\min_{\mathbf{F},\beta} \frac{1}{TMN} \sum_{i,t,\tau} (r_{i,t-1+\tau} - \mathbf{f}_{t-1+\tau}^{T} \beta_i)^2$$

$$\text{s.t. } \mathbf{B}^T \mathbf{B} = I_K$$

- $\widehat{\mathbf{B}}$ can be obtained from the eigenvectors of $\mathbf{R}^T\mathbf{R}$.
- Factors are obtained via

$$\widehat{\mathbf{F}} = \mathbf{R}\widehat{\mathbf{B}}(\widehat{\mathbf{B}}^T\widehat{\mathbf{B}})^{-1} = \mathbf{R}\widehat{\mathbf{B}}$$

- Predicted returns are then

$$\widehat{\mathbf{R}} = \widehat{\mathbf{F}}\widehat{\mathbf{B}}^T = \mathbf{R}\widehat{\mathbf{B}}\widehat{\mathbf{B}}^T$$

Local PCA for high-frequency asset returns

- Rather than apply PCA to all days and all intraday periods, we can consider each intraday period separately:

$$\widehat{\mathbf{F}}(\tau), \widehat{\beta}(\tau) = \arg\min_{\mathbf{F},\beta} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (r_{i,t-1+\tau} - \boldsymbol{\beta}_i^\top \mathbf{f}_{t-1+\tau})^2$$

  for each $\tau \in \{1/M, 2/M, ..., 1\}$.

- As PCA can be done via EV decomposition, this is fast even for large M
- But adjacent periods through the day are likely to be similar
  - Can improve accuracy by using information in adjacent periods
  - Combine PCA with non-parametric kernel estimation:

Local PCA for high-frequency asset returns

- For a some kernel function, ϕ, and bandwidth, $h > 0$, local PCA solves:

$$\widetilde{\mathbf{F}}(\tau), \widetilde{\beta}(\tau) = \arg\min_{\mathbf{F},\beta} \frac{1}{TMN} \sum_{i,t,\tau} \phi\left(\frac{\tau - s}{h}\right) \left(r_{i,t-1+s} - \beta_i^\top \mathbf{f}_{t-1+s}\right)^2$$

  for each $\tau \in \{1/M, 2/M, ..., 1\}$
  - The kernel downweights periods s that are far from the target period τ.
- This is equivalent to standard PCA on *re-weighted* returns, $\widetilde{\mathbf{R}}(\tau)$, with rows

$$\widetilde{\mathbf{r}}_{t-1+s}(\tau) = \phi\left(\frac{\tau - s}{h}\right)^{1/2} \mathbf{r}_{t-1+s}$$

- So this is again available in "closed" form and fast.
  - But requires choosing a kernel and bandwidth.

Kernel and bandwidth selection

- We consider the familiar the Gaussian kernel in our local PCA estimation:

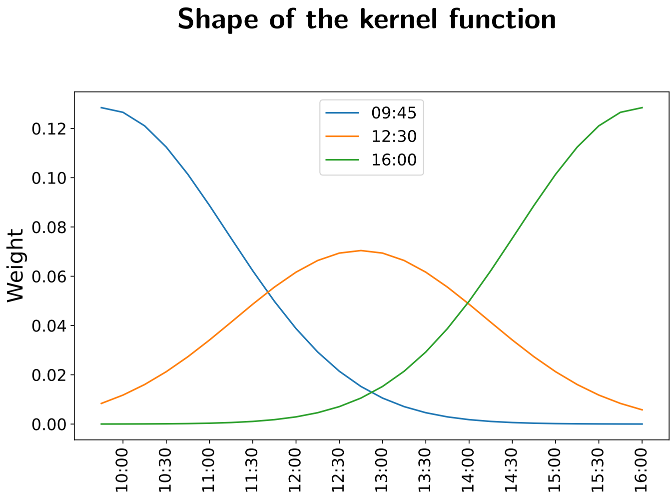
$$\phi\left(\frac{\tau - s}{h}\right) = \exp\left\{-\frac{1}{2}\left(\frac{\tau - s}{h}\right)^2\right\}$$

  - We use a truncated kernel so that 9:30 am is not treated as adjacent to 4 pm.

- The bandwidth parameter, h, controls how much info we draw from neighbouring periods
  - $h \to \infty$ leads us to classic PCA, treating all periods jointly
  - $h \to 0$ leads to PCA on each periods separately
  - We select the optimal h using a validation sample.

Kernel and bandwidth selection

Example kernels used in our estimation, $h^* = 0.17$



Selecting the number of factors

- A critical ingredient in latent factor models is the number of factors, K, to consider.
- We adopt widely-used information criteria proposed by Bai and Ng (2002, *ECMA*), recently extended for use on high frequency data by Liao and Todorov (2024, *QE*).

$$\hat{K} = \text{argmin}_{K \leq K_{max}} \log \sum_{k>K} \lambda_k + K \frac{T+N}{TN} \log \left(\frac{TN}{T+N} \right)$$

where $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_N$ denote the sorted eigenvalues of $\frac{1}{TN}\mathbf{R}^\top \mathbf{R}$.

- For standard PCA, we allow $\hat{K}$ to vary each year.
- For Intraday PCA, we additionally allow to vary across intradaily periods.

Estimating factor models

Data

- We study every stock that was ever listed in S&P 500 index between January 1996 and December 2020, and which traded for the full sample period:
  - $N = 407$, balanced panel.
- We use data from TAQ during this period. Also, we drop the days with incomplete trading hours.
  - $T = 6254$ days.
- We consider 15-minute sampling and discard overnight period.
  - $M = 26$ observations per day.

Estimating factor models

Estimation strategy

- Consider a range of factor models:
  - Intraday PCA; Standard PCA (Pelger (2020, JF)); CAPM; FF3 (Fama and French (1993, JFE)); FF6 (Fama and French (2015, JFE) + Momentum)
- We use one-year windows of data for training, validation and testing. eg:
  1. Estimate models across range of hyperparameters using 1996
  2. Select the optimal hyperparameters using 1997
  3. Compare the models using 1998
- Then repeat step 1-3 on next block of data: $\{1997, 1998, 1999\}$ and so on.

Estimating factor models

Comparing Intraday PCA with other factor models

- To compare models out-of-sample, we use "approximate leave-one-out" forecast:

$$\tilde{r}_{i,t} = \hat{\boldsymbol{\beta}}_i^\top \hat{\mathbf{f}}_t - \hat{\boldsymbol{\beta}}_i^\top \hat{\boldsymbol{\beta}}_i (r_{i,t} - \hat{\mu}_i) \quad (1)$$

which leads to "approximate leave-one-out" forecast R^2:

$$R^2_{\mathrm{ALOO}} = 1 - \frac{\sum_i \sum_{(t,\tau)\in\mathcal{T}} (r_{i,t-1+\tau} - \tilde{r}_{i,t-1+\tau})^2}{\sum_i \sum_{(t,\tau)\in\mathcal{T}} r^2_{i,t-1+\tau}} \quad (2)$$

- Lying between the "predictive" and "total" R^2 metrics considered in Kelly et al. (2019, JFE)
- Improvements in fit assessed via a panel Diebold-Mariano (1996, *JBES*)

Estimating factor models

Performance evaluation

	Factor model				
	CAPM	FF3	FF6	PCA	Intraday PCA
OOS R^2	18.27	20.53	20.77	22.49	23.76
(DM t-stat)	(16.43)	(12.12)	(12.07)	(18.69)	

Estimating factor models

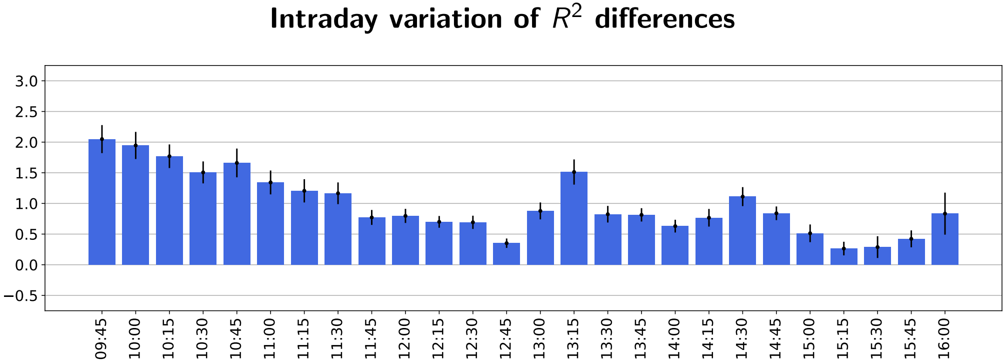
Performance evaluation

	Factor model				
	CAPM	FF3	FF6	PCA	Intraday PCA
OOS R^2	18.27	20.53	20.77	22.49	23.76
(DM t-stat)	(16.43)	(12.12)	(12.07)	(18.69)	

- Intraday PCA shows superior performance comparing to both observable factor models and standard PCA.
- Improvement over CAPM is 5%.
- Improvement over standard PCA is 1.3%.
    - The t-stat of 18.69 is strong evidence that the factor structure of asset returns changes over the trade day.

Estimating factor models

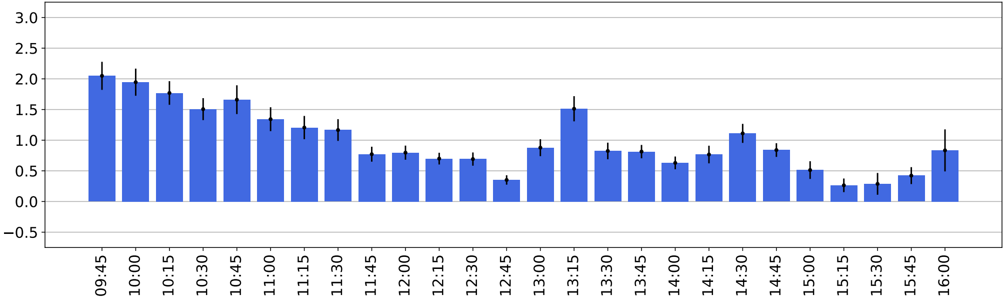
Performance evaluation



Estimating factor models

Performance evaluation

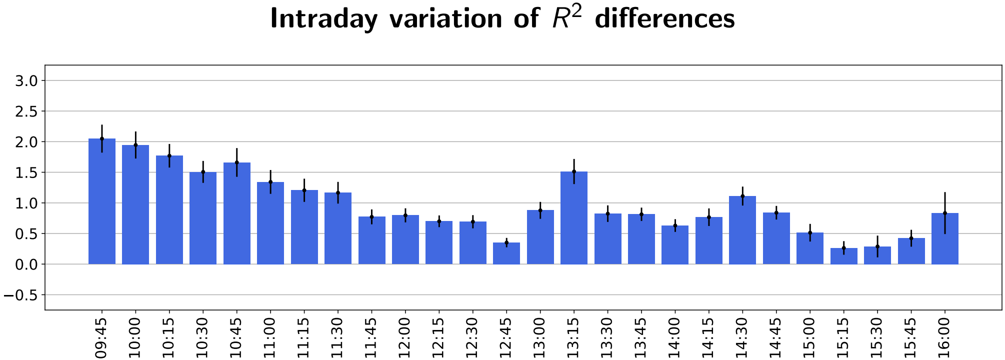
Intraday variation of R^2 differences



- Improved performance of Intraday PCA is not uniform.
- Gains are larger and more significant near the market open and smaller and less significant at the market close.

Estimating factor models

Performance evaluation



- The factor structure of asset returns varies over the trade day.
  - Why?

Outline

1. Local PCA for high-frequency asset returns
   - Evidence of time-varying factor structures
2. A stylized model of information flows and factor structures
   - Empirical predictions
3. **Do information flows generate changes in factor structures?**
   - Earnings announcements
   - FOMC announcements
4. Conclusion

Do information flows generate changes in factor structures?

- We argue that information flows can lead to variation in factor structures.
- We now consider empirical exercises to study that channel:
  1. Surprising earnings announcement days
  2. FOMC announcement days
- Constant factor structure $\implies$ Inraday PCA $\approx$ standard PCA

Surprising earnings announcements

- Earnings announcements have significant impacts on stock market.
  - Livnat and Mendenhall (2006, JAR); Savor and Wilson (2016, JF); Patton and Verardo (2012, RFS)
- We narrow down our evaluation sample to the earnings days:
  - Days following an overnight period in which the market share of announcing firms exceeds 1% of the total market equity of our sample of firms.
  - $\approx$ 2000 days in total
  - The data source is IBES for earnings date-time info

Surprising earnings announcements

- Measure earning surprises as the difference between the realized earnings per share and the median analyst forecast in the 90 days before the announcement, scaled by the price (Livnat and Mendenhall 2006; JAR)

$$ES_{i,t} = \frac{EPS_{i,t} - \widehat{EPS}_{i,t}}{P_{i,t-1}}$$

- We sort earnings days into quintiles using the absolute value of market-cap weighted average of earnings surprises.

Surprising earnings announcements

	Quintile of absolute earnings surprise					
	Low	2	3	4	High	High-Low
OOS R^2 Diff.	0.89	1.26	1.46	1.30	2.01	1.12
(DM t-stat)	(15.78)	(16.16)	(14.04)	(16.02)	(13.71)	(8.03)

Surprising earnings announcements

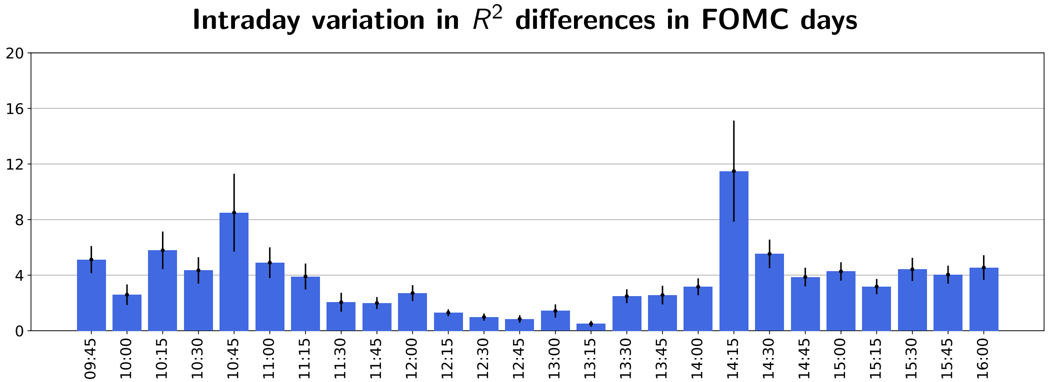
	Quintile of absolute earnings surprise					
	Low	2	3	4	High	High-Low
OOS R^2 Diff.	0.89	1.26	1.46	1.30	2.01	1.12
(DM t-stat)	(15.78)	(16.16)	(14.04)	(16.02)	(13.71)	(8.03)

- The improvement of Intraday PCA over standard PCA increases from 0.89% to 2.01%.
- High-Low difference is also significant.

FOMC announcements

- We compare factor models in FOMC days.
  - 8 times per year, roughly.
  - 196 announcements in our sample.
- We restrict full sample to only FOMC days.
- Given the relatively small sample,
  - adopt an expanding window estimation strategy
  - set $h = 0$ for Intraday PCA

FOMC announcements



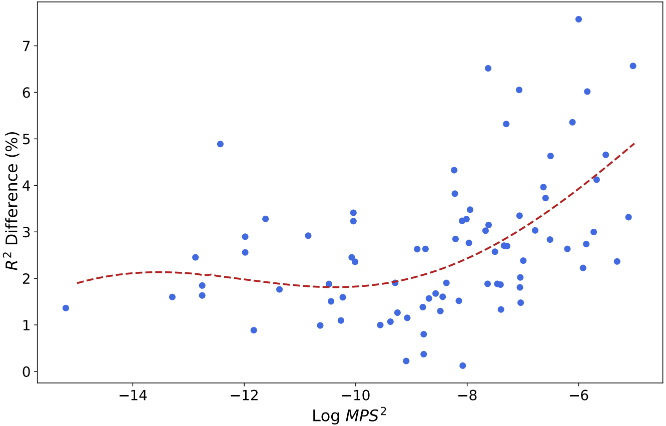
- The gains from using Intraday PCA are particularly large, $> 5\%$, in the first two 15-minute windows after the FOMC announcement.

FOMC announcements

- Similar to earnings announcements study, we also analyze the information content of FOMC announcements.
- We measure the information content by monetary policy surprise series of Bauer and Swanson (2023, NBERMA)
  - The first principal component of changes in first four quarterly Eurodollar futures contracts around announcement.
- We compute the R^2 differences for each FOMC day in the sample and plot them against monetary policy surprises.

FOMC announcements

R^2 differences across monetary policy surprises



Summary

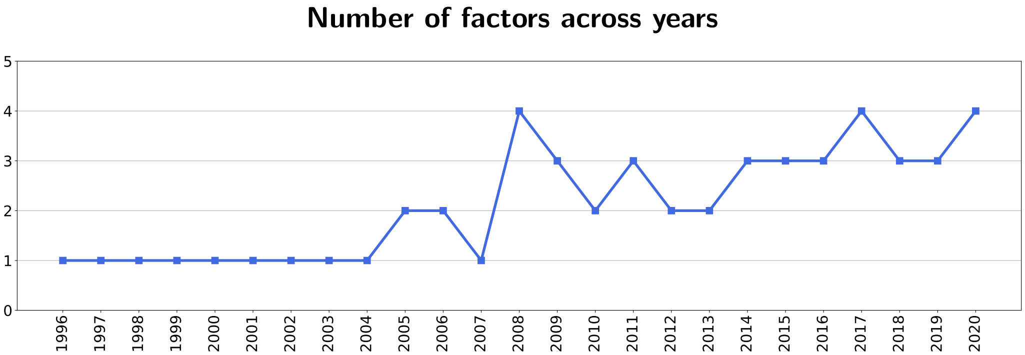
Three main contributions

1. We propose a method to allow for changes in factor structure of asset returns over the trade day: *Intraday PCA*
   - Combines with nonparametric kernel methods
   - Gains relative to standard PCA indicate significant changes in the factor structure
2. We present a stylized model of information flows and asset prices
   - Information flows can change the factor structure of prices
   - Changes predicted to be larger when information flow is larger
3. We investigate the information flow channels that can generate intraday variation.
   - The gains are strongest in the days with surprising earnings and FOMC announcements.

Additional Materials

Selecting the number of factors

Optimal number of factors is broadly increasing over the sample period



Selecting the number of factors

Optimal number of factors is higher at the open and broadly increasing over the sample period