Text-Based Linkages and Local Risk Spillovers in the Equity Market

Shuyi Ge

University of Cambridge

4th Nov, 2020

Introduction

- Cross-sectional co-movement in equity returns
- The co-movement are the combined effects of two types of cross-sectional dependence (CSD)
  - strong CSD (factors, non-diversifiable)
  - weak (local) CSD (local interactions, diversifiable)
- Factor models have been used to model strong CSD: CAPM by Sharpe (1964), APT by Ross (1976), Fama and French (1993).
- Connectivity in the de-factored returns are found to be non-negligible (Gabaix 2011, Acemoglu 2012, Barigozzi and Hallin 2017, Kou et al. (2018))
- Challenge of studying weak CSD: The network architecture of firms is hard to get.

This Paper

- Identify business links based on article co-mentioning using business news from Business Wire (Source: LexisNexis Academics)
  - Features: wide range of links that facilitate risk spillovers (strategic partner, joint venture, outsourcing, financing, customer-supplier, *M&A*, business lines acquisitions, legal, etc), public sources, granular, real time, high-frequency
- Integrated measure of connectivity. Quantify the strength of local dependency (contemporaneous & dynamic) among linked firms

Related Literature

- Economic Links and Spillover Effect: Moskowitz and Grinblatt (1999), Engelberg et al. (2018), Pirinsky and Wang (2006), and Parsons et al. (2020), Cohen and Frazzini (2008)

- Textual Analysis and Economics:
  - Sentiment: Garcia (2013), Baker et.al (2007)
  - Economic Policy Uncertainty: Baker et.al (2007)
  - Quantify Fundamentals: Tetlock et al. (2008)
  - Link Mining: Hoberg and Phillips (2016), Scherbina and Schlusche (2015), Schwenkler and Zheng (2019)

- Identify Links from Panel Data
  - Low-dimensional: Billio et al. (2012), Diebold and Yilmaz (2014), Hale and Lopez (2019)
  - High-dimensional: Hautsch et al. (2014, 2015), Barigozzi and Hallin (2017), Demirer et al.(2018)

The Rest of My Talk

- Identification of links from text data and measure of connectivity
- Model local dependencies in the equity market using spatial-temporal model. Application to S&P 500 stocks

A Typical Business News from LexisNexis Academics

- A general case: NLP

American Express and Regis Corporation Announce Strategic Partnership; Hair Care Industry's Global Leader to Roll-out Card Acceptance at all of its U.S. Locations

Business Wire

February 24, 2005 Thursday 2:00 PM GMT

Distribution: Business Editors

Length: 438 words

Dateline: NEW YORK Feb. 24, 2005

Body

American Express and Regis Corporation today announced a plan for nationwide card acceptance at Regis' U.S. salons. With thousands of locations currently accepting the American Express Card, the companies expect all corporate owned Regis U.S. locations to be accepting American Express by the end of calendar year 2005.

"Our partnership with American Express is in direct response to our customers. Over the last several years we have seen an increasing demand by our customers to accept American Express," commented Kyle Didier, vice president, finance at Regis Corporation. "In addition, American Express' willingness to extend the partnership benefits to our franchisees demonstrates their commitment to drive value throughout the entire Regis Corporation network."

"Working with the world's largest operator of hair salons reinforces our commitment to the hair salon industry overall and demonstrates our ability to drive value to hair salon owners," said Elizabeth Langwith, vice president, American Express Establishment Services. "We're delighted to provide our Cardmembers with another opportunity to earn rewards, cash or miles for their everyday purchases."

The companies will work together to develop marketing programs that deliver value to consumers using American Express-branded cards. In addition, Regis franchise owners will qualify for special discounts on a variety of business expenses ranging from shipping, technology, car rentals and cellular phone service through the American Express Business Savings Program.

Classification

Language: ENGLISH

Publication-Type: Newswire

Subject: FRANCHISING (85%); PRESS RELEASES (75%); CONSUMERS (69%); FRANCHISEES (62%); Contract/Agreement (%)

Company: REGIS CORP (94%); AMERICAN EXPRESS CO (94%); HAIR CLUB FOR MEN INC (52%); NY-AMEX/REGIS

Ticker: RGS (NYSE) (94%); ***AXP*** (NYSE) (94%); RGS (NYSE)

Identification of Links and Construction of W

- Identification Assumption: if a piece of business news reports only two companies together, they have a link.
- Use $N \times N$ adjacency matrix $W = (w_{ij})$ to store all the links identified in the sample news.
  - N: # companies
  - w_{ij} (strength of the link): # distinct news items that co-mention i and j
- Alternative identification assumptions (narrower definitions) and specifications of W (weighting schemes) explored

The Rest of My Talk

- Identification of links and measure of connectivity
- Model local dependencies in equity market using spatial-temporal model. Application to S&P 500 stocks
  1. Strength of local dependencies? heterogeneity?
  2. Systemically important companies?
  3. Evolution of two forms of co-movement?
  4. How do we benefit from the novel dataset?

"Factor + Spatial" Model of Strong and Weak CSD

- Model strong CSD using hierarchical factor model

$$r_{it} = \alpha_i + \sum_{k=1}^{K} b_{ik} f_{kt} + \gamma_i f_{gt} + \epsilon_{it} \quad \text{for } i \in g$$

- Model (local/weak CSD) use heterogenous spatio-temporal (HSAR) model (Bailey et.al 2016, Aquaro et.al 2020)
  -

$$\epsilon_{it} = a_i + \underbrace{\sum_{k=1}^{L_1} \lambda_{i,k} \epsilon_{i,t-k}}_{\text{temporal dependence}} + \underbrace{\sum_{k=0}^{L_2} \psi_{i,k} \Big(\sum_{j=1}^{N} w_{ij} \epsilon_{j,t-k} \Big)}_{\text{cross-sectional dependence (CSD)}} + \upsilon_{it}$$

  - Need large T for consistent estimation (for any N)

Application to S&P 500 Stocks (Full Sample)

stage 1: Remove strong CSD $\quad r_{it} = \alpha_i + \sum_{k=1}^{K} b_{ik} f_{kt} + \gamma_i f_{gt} + \epsilon_{it} \quad$ for $i \in g$

stage 2: Model weak CSD $\quad \epsilon_{it} = a_i + \sum_{k=1}^{L_1} \lambda_{i,k} \epsilon_{i,t-k} + \sum_{k=0}^{L_2} \psi_{i,k} (\sum_{j=1}^{N} w_{ij} \epsilon_{j,t-k}) + \upsilon_{it}$

- Panel of daily excess returns r_{it} of $S\&P$ 500 stocks from 03/01/2006 to 31/12/2013. $T = 2014$, $N = 413$. ($g = 1, \ldots, 5$ FF industry groups)
- Stage 1: OLS for each i. f_{kt}: FF5+momentum, f_{gt}: FF5 industry factors. Reduce average pairwise correlations from $\hat{\bar{\rho}}_{N,r} = 0.4308$ to $\hat{\bar{\rho}}_{N,\epsilon} = 0.008$
- Stage 2: Quasi-Maximum Likelihood Estimation(QMLE)
  $L_1 = L_2 = 5$, W full sample links (apply row normalization)

The Rest of My Talk

- Identification of links and measure of connectivity
- Model local dependencies in equity market using spatial-temporal model. Application to S&P 500 stocks
  1. Strength of local dependencies? heterogeneity?
  2. Systemically important companies?
  3. Evolution of two forms of co-movement?
  4. How do we benefit from the novel dataset?

Mean Group (MG) Estimator

$$\epsilon_{it} = a_i + \sum_{k=1}^{L_1} \lambda_{i,k}\epsilon_{i,t-k} + \sum_{k=0}^{L_2} \psi_{i,k}(\sum_{j=1}^{N} w_{ij}\epsilon_{j,t-k}) + \upsilon_{it}$$

- Assume $\psi_{k,i} = \psi_k + \varsigma_{k,i}$ and $\varsigma_i \sim IID(0, \Omega_\varsigma)$
  MG Estimator $\hat{\psi}_k^{MG} = \frac{1}{N}\sum_{i=1}^{N} \hat{\psi}_{k,i}$ (Bailey at el. 2016, 2020)
- Summarise by sector:
  Assume $\psi_{k,i,g} = \psi_{k,g} + \varsigma_{k,i,g}$ and $\varsigma_{i,g} \sim IID(0, \Omega_\varsigma)$
- Any other grouping goes

HSAR Estimation Results (Whole Sample)

$$\epsilon_{it} = a_i + \sum_{k=1}^{L_1} \lambda_{i,k}\epsilon_{i,t-k} + \sum_{k=0}^{L_2} \psi_{i,k}\left(\sum_{j=1}^{N} w_{ij}\epsilon_{j,t-k}\right) + \upsilon_{it}$$

	(1)AR coefs	(2) Cross-Sectional Dependence coefs					
	λ	ψ_0	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5
MG estimates	...	0.307***	0.037***	0.009	-0.008	0.008	0.011**
	...	(0.021)	(0.007)	(0.006)	(0.005)	(0.005)	(0.006)
% Signif (at 5%)	...	81.4%	21.6%	20.3%	16.2%	17.2%	14.7%

- Local shocks diffuse over time and space.
- $\hat{\psi}_0^{MG} = 0.307$. Considerable contemporaneous local dependency. 81.4% individual parameters are sig (success at mining links!)

HSAR Estimation Results (by sector)

	(1)AR coefs	(2) Cross-Sectional Dependence terms					
	λ	ψ_0	ψ_1	ψ_2	ψ_3	ψ_4	ψ_5
Panel A: Consumer (N=77)							
MG Estimates	...	0.232***	0.033***	0.026**	-0.001	-0.002	0.005
	...	(0.039)	(0.009)	(0.011)	(0.010)	(0.010)	(0.012)
% Sig(at 5%)	...	79.2%	15.6%	11.7%	14.3%	11.7%	7.8%
Panel B: Finance (N=75)							
MG Estimates	...	0.345***	0.056***	-0.010	-0.018	0.023	0.050***
	...	(0.057)	(0.026)	(0.019)	(0.020)	(0.017)	(0.017)
% Sig(at 5%)	...	82.7%	32.0%	34.7%	30.7%	30.7%	29.7%
Panel C: Health (N=35)							
MG Estimates	...	0.061	0.020	0.001	-0.001	0.029**	0.041**
	...	(0.061)	(0.016)	(0.015)	(0.013)	(0.016)	(0.020)
% Sig(at 5%)	...	68.6%	14.3%	11.4%	8.6%	5.7%	14.7%
Panel D: Hitech (N=73)							
MG Estimates	...	0.229***	0.018	-0.004	-0.001	0.004	-0.014
	...	(0.048)	(0.011)	(0.013)	(0.009)	(0.010)	(0.014)
% Sig(at 5%)	...	72.6%	11.0%	13.7%	6.8%	12.3%	11.0%
Panel E: Manufacturing (N=110)							
MG Estimates	...	0.446***	0.032***	0.018**	-0.008	0.004	0.005
	...	(0.033)	(0.009)	(0.008)	(0.008)	(0.007)	(0.007)
% Sig(at 5%)	...	85.5%	18.2%	18.2%	13.6%	16.4%	13.0%

The Rest of My Talk

- Identification of links and measure of connectivity
- Model local dependencies in equity market using spatial-temporal model. Application to S&P 500 stocks
  1. Strength of local dependencies? heterogeneity?
  2. Systemically important companies?
  3. Evolution of two forms of co-movement?
  4. How do we benefit from the novel dataset?

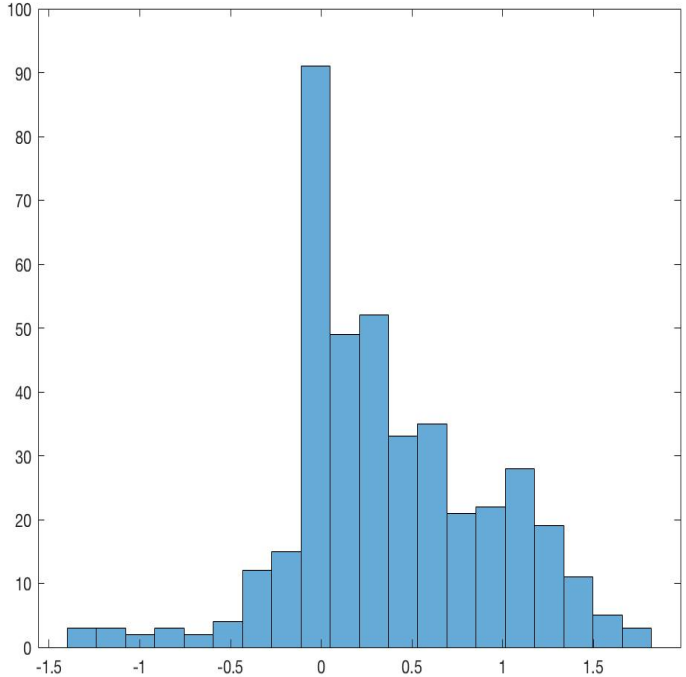
Spatial-temporal Spillover Matrix

- Generalized impulse response function (GIRF) to trace the effect of a primitive shock to firm k at t on the system h period ahead:

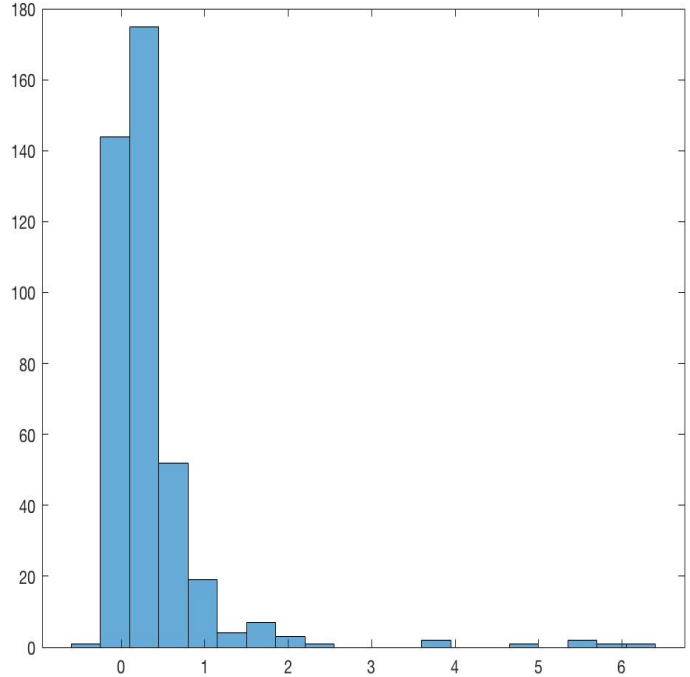
$$GI(h, \delta_k, \Omega_{t-1}) = E(\epsilon_{t+h} \mid \upsilon_{k,t} = \delta_k, \Omega_{t-1}) - E(\epsilon_{t+h} \mid \Omega_{t-1})$$

- For horizon h, compute $GI(h, \delta_k = 1, \Omega_{t-1})$ for $k = 1, \ldots, N$.
- Spatial-temporal spillover matrix $D_h = [d_{ij}^h]$ where $GI(h, \delta_k = 1, \Omega_{t-1})$ is the kth column.
- For h, summarize individual level cross effect,
  1. In-degree: $C_{i,in}^h = \sum_{j \neq i}^N d_{ij}^h \rightarrow$ total spillovers to i
  2. Out-degree: $C_{j,out}^h = \sum_{i \neq j}^N d_{ij}^h \rightarrow$ total spillovers from j

Degree Distribution (contemporaneous spillovers $h = 0$)



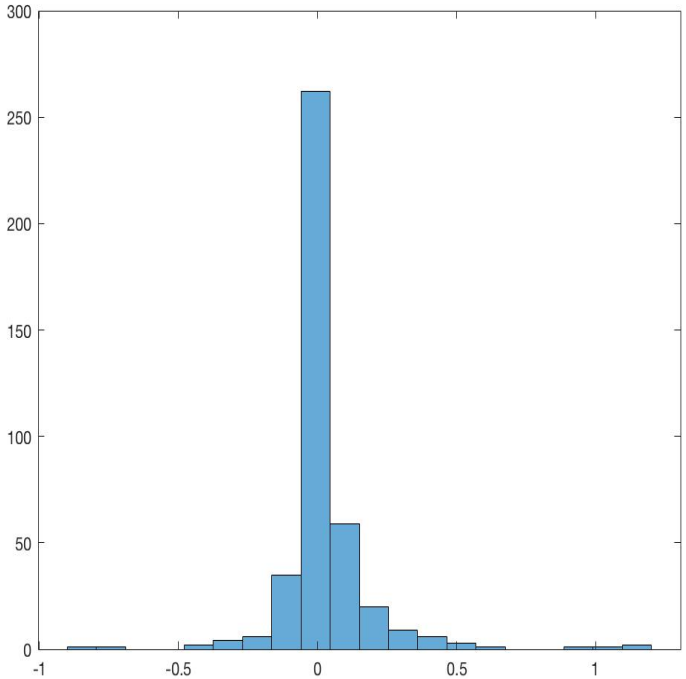
(a) In-Degree ($C^h_{i,in} = \sum_{j \neq i}^{N} d^0_{ij}$)



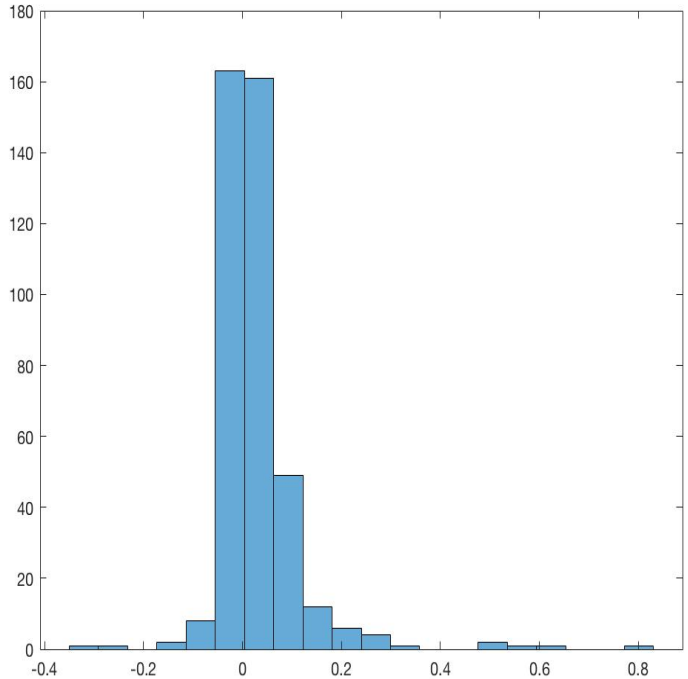
(b) Out-Degree ($C^h_{j,out} = \sum_{i \neq j}^{N} d^0_{ij}$)

- Right-skewed
- Out-degree distribution shows heavy right tail

Degree Distribution (dynamic spillovers $h = 1$)



(a) In-Degree ($C^h_{i,in} = \sum_{j \neq i}^{N} d^1_{ij}$)



(b) Out-Degree ($C^h_{j,out} = \sum_{i \neq j}^{N} d^1_{ij}$)

- Out-degree right-skewed
- Dynamic spillovers are smaller in magnitude
- Shock decay over time dimension quickly

Systemically Important (SI) Companies

		Company Ticker
h=0	In-degree	LEN, EIX, PCG, DUK, DHI, NOC, GD, RIG, RTN, LNC, ETR, ALTR, SO, LRCX, CSX, PNW, UNH, HBAN, PFG, POM
	Out-degree	BAC, MSFT, GE, GS, JPM, XOM, C, CVX, LNC, WFC APPL, USB, BA, FITB, VZ, JNJ, PG, AET, UNH, PFE
h=1	In-degree	GNW, FITB, HBAN, GE, WY, STT, LEN, LNC, CI, COF FLR, PG, ATI, AES, RIG, JEC, PH, CAG, HD, HUM
	Out-degree	BAC, MSFT, GE, GS, JPM, XOM, LM, C, CVX, LNC, DUK, WFC, APPL, USB, BA, FITB, BZ, JNJ, PG, HCP

Table: The 20 firms with highest in-degree and out-degree for $h = 0, 1$

- SI risk contributors : large cap financial institutions, hitech and manufacturers
- SI risk receivers: periphery manufacturing and financial firms

The Rest of My Talk

- Identification of links and measure of connectivity
- Model local dependencies in equity market using spatial-temporal model. Application to S&P 500 stocks
  1. Strength of local dependencies? heterogeneity?
  2. Systemically important companies?
  3. Evolution of two forms of co-movement?
  4. How do we benefit from the novel dataset?

Evolution of Weak CSD

- Rolling window analysis with 251-day rolling sample from 03/01/2006 to 31/12/2013 (on average, $N = 447$ for each window)
- For window $[t - 125, t + 126]$, do the two-stage procedure. W is constructed using all the news published within the window

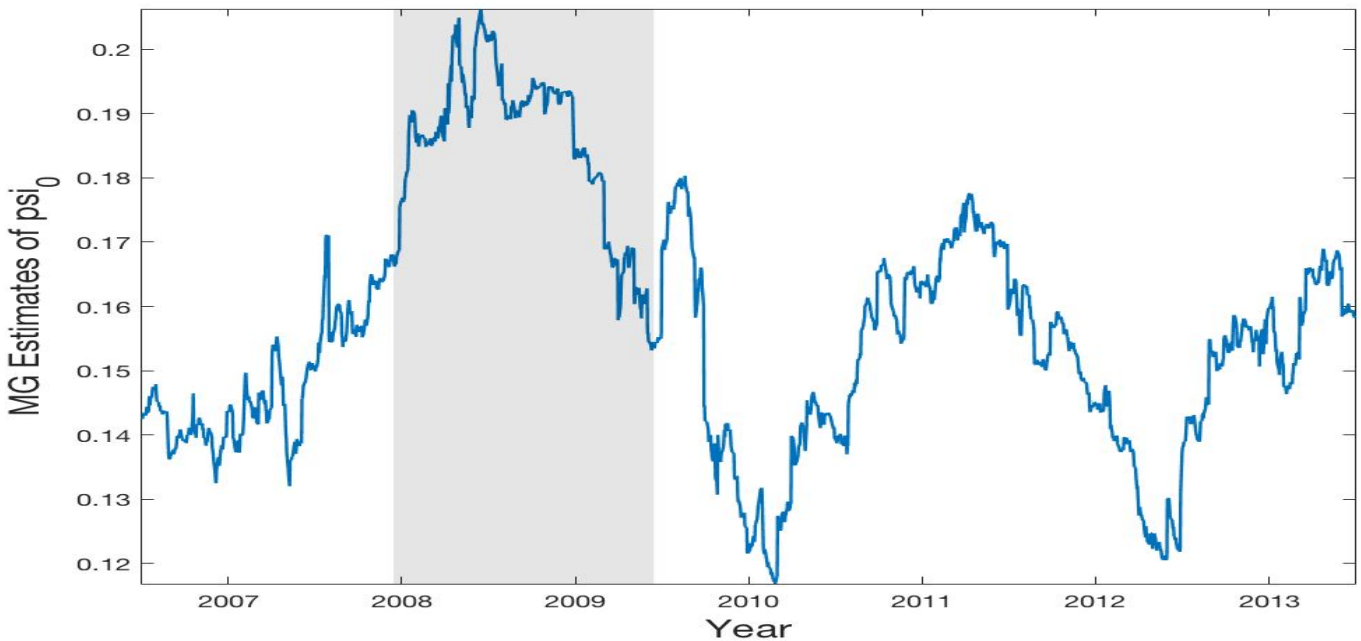
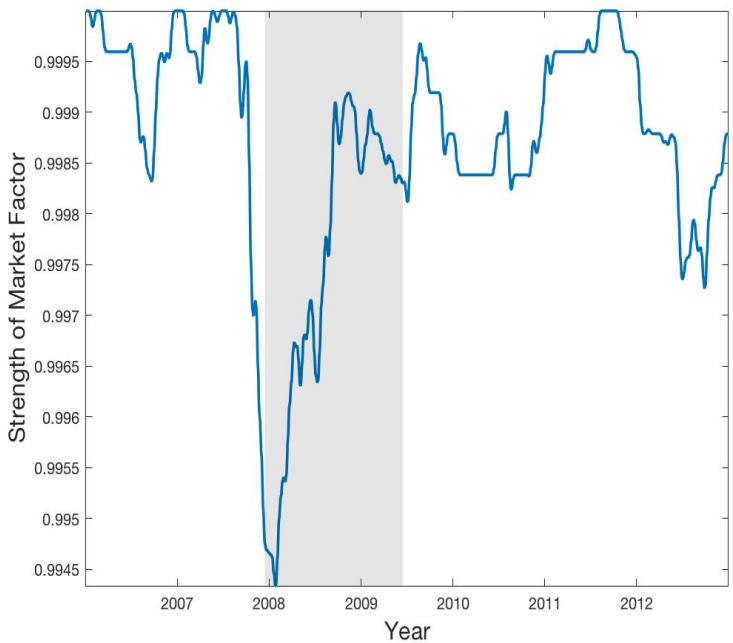


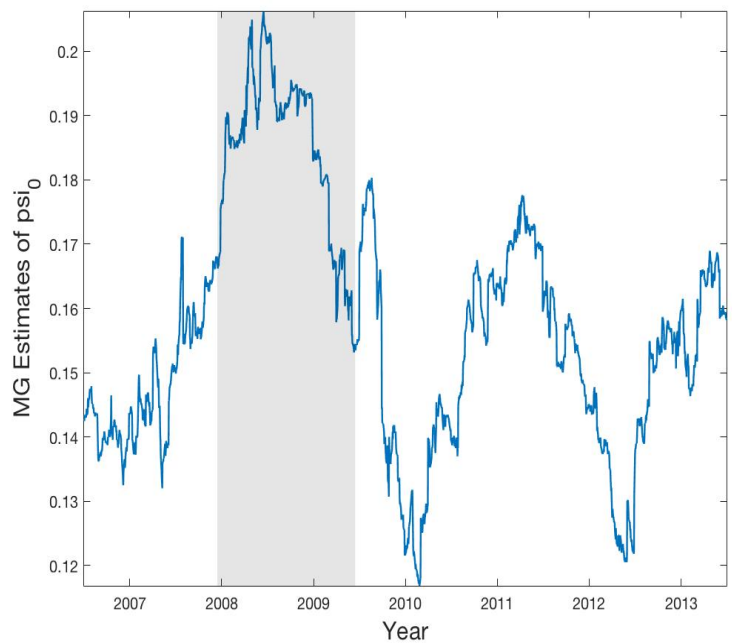
Figure: Time series of $\hat{\psi}_{0,t}^{MG}$, the 251-day rolling mean group estimates of the the contemporaneous dependence parameter.

Strength of Factor vs Strength of Local Spillovers

- Measure the strength of factors by the exponent of cross sectional dependence $\alpha \in [0, 1]$ from Bailey et.al (2019, 2020)
- α departs from 1 during financial crisis period
- Local dependencies gain importance as market factor loses importance $\rightarrow$ market decoupling



(a) rolling α_t of market factor



(b) rolling $\hat{\psi}_{0,t}^{MG}$

The Rest of My Talk

- Identification of links and measure of connectivity
- Model local dependencies in equity market using spatial-temporal model. Application to S&P 500 stocks
  1. Strength of local dependencies? heterogeneity?
  2. Systemically important companies?
  3. Evolution of two forms of co-movement?
  4. How do we benefit from the novel dataset?

Model Comparisons (In-sample & Out-of-sample)

- Compare mean squared error (MSE) alternative models
- Alternatives:
  1. Benchmark: Naive estimator 0
  2. High-dimensional VAR in Barigozzi and Hallin 2017 (BH-VAR)

$$y_{it} = \sum_{k=1}^{p} \sum_{j=1}^{N} \beta_{ij,k} y_{jt-k} + \sum_{h \neq i}^{N} \gamma_{ih} y_{ht} + e_{it}$$

  3. Spatial-temporal model with different W: (1)empty; (2)sectoral block diagonal matrix (Fan et al.2016); (3)geographic network (Pirinskyand Wang 2006); (4)compustat customer-supplier links; (5)news-based links

Model Comparison (In-sample & Out-of-sample)

	Naive	BH-VAR	W_{empty}	W_{sector}	$W_{geographic}$	$W_{compustat}$	W_{news}
In-Sample MSE							
(1)Heterogeneous coef	-	-	2.907	**2.829**	2.876	2.903	**2.764**
(2)Sectoral-heterogeneous coef	-	-	2.912	2.902	2.921	2.929	2.863
(3)Homogeneous coef	-	-	2.804	2.918	2.920	2.926	2.865
(4)	2.935	**2.211**	-	-	-	-	-
Out-of-Sample MSE							
(1)Heterogeneous coef	-	-	1.353	1.332	1.371	1.353	**1.287**
(2)Sectoral-heterogeneous coef	-	-	1.350	1.336	1.368	1.347	**1.302**
(3)Homogeneous coef	-	-	1.351	1.338	1.370	1.348	**1.309**
(4)	1.348	1.423	-	-	-	-	-

Table: In-sample and out-of-sample MSE (in basis point) of alternative models. For each panel, the best 3 (smallest MSE) cases are in red.

- The training sample 03/01/2006 to 31/12/2013 (2014 days), testing sample spans from 03/01/2014 to 31/12/2014 (252 days).
- Spatial-temporal model with W_{news}, has the best out-of sample performance, under any parameter heterogeneity assumption.
- Rolling three-year sub-samples. Robust

Alternative Specifications and Robustness Checks

▸ Alternative de-factoring method: PCA, Dynamic factor model

▸ Alternative specifications of W

- Weighting schemes
- Narrower definitions of links: (1)non-competitor links (2) persistent links (3) inter-sector (FF5) links (4) inter-industry (four-digit SIC) links

▸ 2-W specification

$$\epsilon_t = \mathrm{a}_\epsilon + \sum_{k=1}^{L_1} \Lambda_{\boldsymbol{k}} \epsilon_{t-k} + \sum_{k=0}^{L_2} \Psi_{1,\boldsymbol{k}} W_1 \epsilon_{t-k} + \sum_{k=0}^{L_2} \Psi_{2,\boldsymbol{k}} W_2 \epsilon_{t-k} + \upsilon_t$$

- W_1 inter-industry news-based network, W_2 industry block matrix
- 2 Specifications from broad to granular
  1. FF 5 industry
  2. Four-digit SIC code

Empirical Results Wrap Up

- Use novel text-based links to investigate local dependencies among linked firms.
- Stocks linked via news co-mentioning exhibit excess co-movement beyond predicted by standard asset pricing models.
- HSAR model provides a flexible framework to various specifications
- Text-based network as a promising alternative to existing network data. Competitive in modeling of local spillovers (wider contexts?)

General Case: Named-Entity-Recognition (NER)

Back

American Express [ORG] and Regis Corporation [ORG] today announced a plan for nationwide card acceptance at Regis [ORG] ' U.S. [LOC] salons . With thousands of locations currently accepting the American Express Card [ORG] , the companies expect all corporate owned Regis U.S. [LOC] locations to be accepting American Express [ORG] by the end of calendar year 2005 . " Our partnership with American Express [ORG] is in direct response to our customers . Over the last several years we have seen an increasing demand by our customers to accept American Express [ORG] , " commented Kyle Didier [PER] , vice president , finance at Regis Corporation [ORG] . " In addition , American Express [ORG] ' willingness to extend the partnership benefits to our franchisees demonstrates their commitment to drive value throughout the entire Regis Corporation [ORG] network . "

- Named entity recognition (NER) to identify organizations
- Match CRSP stocks by name (string distance) or ticker

Alternative Specifications of W: Mean Group Estimates

- Weighting schemes (1)weighted by # of monthly windows *i&j* get co-mentioned (2)weighted by the # news *i&j* get co-mentioned (3)unweighted
- Narrower definitions of links (1)non-competitor links (2) persistent links (3) inter-sector (FF5) links (4) inter-industry (four-digit SIC) links

	$W_{baseline}$	$W_{weighted}$	$W_{unweighted}$	$W_{noncomp}$	$W_{persistent}$	$W_{intersector}$	$W_{interindustry}$
MG Estimates	0.292***	0.281***	0.286***	0.282***	0.200***	0.060***	0.143***
	(0.021)	(0.020)	(0.023)	(0.20)	(0.018)	(0.014)	(0.143)
% Sig(at 5%)	77.0%	77.5%	74.8%	75.8%	69.8%	46.2%	62.7%

Table: MG estimates of ψ_0

2-W specification

Back

$$\epsilon_t = \mathrm{a}_\epsilon + \sum_{k=1}^{L_1} \Lambda_{\boldsymbol{k}} \epsilon_{t-k} + \sum_{k=0}^{L_2} \Psi_{1,\boldsymbol{k}} W_1 \epsilon_{t-k} + \sum_{k=0}^{L_2} \Psi_{2,\boldsymbol{k}} W_2 \epsilon_{t-k} + \upsilon_t$$

- W_1 inter-industry news-based network, W_2 industry block matrix
- 2 Specifications from broad to granular
  1. FF 5 industry
  2. Four-digit SIC code
- Conditional on W_2, ψ_1 continue to be sig
- Intra-industry effect is stronger, hetero