Aggregation of Information about the Cross Section of Stock Returns: a Latent Variable Approach

Nathaniel Light, Denys Maskov, Oleg Rytchkov

Presented by Nuoya Hong
For Fall 19 MGMTMFE-431-3 Statistical Arbitrage

October 21, 2019

Outline

Introduction

Overview

- Light et al. (2017) propose a new approach for predicting stock returns on individual stocks from a large number firm characteristics.
- The expected returns are treated as latent variables, and the firm characteristics are assumed to link to expected returns through one or more common latent factors.
- They use partial least squares (PLS) estimator to estimate expected returns and considered 26 firm characteristics.
- They argue that their PLS approach outperforms alternative techniques (Factor Analysis, PCA, Fama-MacBeth Regression, Rank-based approach).

Introduction

List of 26 Firm Characteristics (4+7+9+2+2+2=26!)

Value Related (4)	Return Related (7)	Growth Related (9)	Profitability (2)	Distress Related (2)	Others (2)
◇ size (S) ◇ book-to-market ratio (B/M) ◇earnings-to-price ratio (E/P) ◇cashflow to price ratio (C/P)	◇market beta (B) ◇momentum (MOM) ◇long-term reversal (LTR) ◇short-term reversal (STR) ◇idiosyncratic volatility (idVOL) ◇Max daily return over past month (MAX) ◇ expected idiosyncratic skewness (EIS)	◇ total asset growth (AG) ◇abnormal capital investments (CI) ◇investment growth (IG) ◇investment-to-capital ratio (I/K) ◇investment-to-assets ratio (I/A) ◇accurals (ACC) ◇net operating assets (NOA) ◇net stock issues (NS) ◇composite stock issuance (ι)	◇returns on equity (ROE) ◇returns on assets (ROA)	◇leverage (LV) ◇O-score (O)	◇turnover (TO) ◇analysts' forecasts dispersion (D)

Methodology

Baseline Specification: Single Latent Variable

Consider a set of N stocks whose characteristics are observed in at least two periods. The best predictor of returns on stock i at time t is expected return $\mu_{it} = E\left[R_{it+1}|\mathcal{F}_t\right]$ where $\mathcal{F}_t$ denotes public information.

The realized return on stock i can be written as

$$R_{it+1} = \underbrace{\mu_{it}}_{\text{unobserved factor in the firm characteristic space}} + \varepsilon_{it+1}$$

where $E\left[\varepsilon_{it+1}|\mathcal{F}_t\right] = 0$ and unexpected returns ε_{it+1} are assumed to be orthogonal to $\mathcal{F}_t$. ε_{it+1} and ε_{jt+1} can be correlated.

Methodology

Baseline Specification: Single Latent Variable

Econometricians lack some information contained in $\mathcal{F}_t$ and do not know μ_{it}. Instead they observe firm characteristics $X_{it}^a, a = 1, \ldots, A$ such that $\{X_{it-s}, s \geq 0, a = 1, \ldots, A, i = 1, \ldots, N\} \subseteq \mathcal{F}_t$. Assume at each time period X_{it}^a have zero cross-sectional means and unit variances (demean, standardize). Assume

$$X_{it}^a = \delta_t^a \left(\mu_{it} - \bar{\mu}_t\right) + u_{it}^a$$

where δ_t^a measures the sensitivity of characteristics a to expected returns, and $\bar{\mu}_t$ is the cross-sectional average of expected returns at time t. Denote the sample cross-sectional variance and covariance as $\overline{Var}$ and $\overline{Cov}$ respectively, and reserve the sample cross-sectional variance and covariance in the characteristic space as $\widetilde{Var}$ and $\widetilde{Cov}$.

Methodology

Identifying Assumptions

Assumption 1 (Distribution of expected returns) In each period t,

$$\overline{\mu}_t = \frac{1}{N}\sum_{i=1}^{N}\mu_{it} \xrightarrow{p} \mu_t \text{ and } \overline{Var}\left(\mu_{it}\right) = \frac{1}{N}\sum_{i=1}^{N}\left(\mu_{it} - \overline{\mu}_t\right)^2 \xrightarrow{p} V_t \text{ as } N \to \infty \text{ where } V_t > 0.$$

Assumption 2 (Distribution of characteristic loadings) In each period t,

$$\tilde{\delta}_t = \frac{1}{A}\sum_{a=1}^{A}\delta_t^a \xrightarrow{p} \delta_t \text{ and } \widetilde{Var}\left(\delta_t^a\right) = \frac{1}{A}\sum_{a=1}^{A}\left(\delta_t^a - \tilde{\delta}_t\right)^2 \xrightarrow{p} \Lambda_{t,t} \text{ as } A \to \infty \text{ where } \Lambda_{t,t} > 0.$$

Also, for consecutive periods $t-1$ and t

$$\widetilde{Cov}\left(\delta_{t-1}^a, \delta_t^a\right) = \frac{1}{A}\sum_{a=1}^{A}\left(\delta_{t-1}^a - \tilde{\delta}_{t-1}\right)\left(\delta_t^a - \tilde{\delta}_t\right) \xrightarrow{p} \Lambda_{t-1,t} \text{ as } A \to \infty, \text{ where } \Lambda_{t-1,t} > 0.$$

Methodology

Identifying Assumptions

Assumption 3 (Orthogonality of errors and expected returns) In each period t and for each characteristic a, $a = 1, \ldots, A$,

$$\overline{Cov}\left(\mu_{it}, u_{it}^{a}\right) = \frac{1}{N}\sum_{i=1}^{N}\left(\mu_{it} - \overline{\mu}_{t}\right)\left(u_{it}^{a} - \overline{u}_{t}^{a}\right) \xrightarrow{p} 0 \text{ as } N \to \infty.$$

Assumption 4 (Orthogonality of errors and characteristic loadings) In each period t and for each stock i, $i = 1, \ldots, N$,

$$\widetilde{Cov}\left(\delta_{t-1}^{a}, u_{it}^{a}\right) = \frac{1}{A}\sum_{a=1}^{A}\left(\delta_{t-1}^{a} - \tilde{\delta}_{t-1}\right)\left(u_{it}^{a} - \tilde{u}_{it}\right) \xrightarrow{p} 0 \text{ as } A \to \infty.$$

Methodology

Estimation Procedure: Motivation

- The main problem of an econometrician is to estimate μ_{it} by filtering them from the observable characteristics X_{it}^{a}.
- Light et al. (2017) use a procedure that belongs to a family of partial least squares (PLS) algorithms.
- The PLS framework is actively used in computational chemistry and behavioral sciences, and recently in applications in finance:
  - Kelly and Pruitt (2013): extract market expectations from a cross-section of variation ratios;
  - Huang et al. (2015): construct an investor sentiment index from several sentiment proxies;
  - Giglio et al. (2016): extend PLS to quantile regressions for aggregation of 19 measures of systemic risk.

Methodology

Estimation Procedure: Implementation

The PLS-based estimation of baseline model at time t can be implemented in two steps:

- **Step 1**: Run separate cross-sectional regressions of $R_{it}, i = 1, \ldots, N$, on each individual firm characteristic $X^a_{it-1}, i = 1, \ldots, N$ for $a = 1, \ldots, A$ and denote the obtained slopes a λ^a_t.
- **Step 2**: For each firm i, $i = 1, \ldots, N$ run a regression of X^a_{it} on $\lambda^a_t, a = 1, \ldots, A$, and denote the obtained slopes as $\hat{\mu}_{it}$.

Methodology

Extension: Multiple Latent Variables - Model

Instead of a single latent factor, the firms characteristics are linked to expected returns through L common latent factors $\beta_{lit}, l = 1, \ldots, L$. The expected return on stock i at time t is related to the latent factor as

$$\mu_{it} = \sum_{l=1}^{L} \beta_{lit} \gamma_l$$

where $\gamma_l, l = 1, \ldots, L$ are loadings on various factors. Introducing a $1 \times L$ matrix of factor values B_{it} and an $L \times 1$ matrix of loadings Γ, the realized return on stock i at time $t+1$ can be written as

$$R_{it+1} = B_{it}\Gamma + \varepsilon_{it+1}.$$

Unexpected returns ε_{it+1} are independent from all variables available at time t, $E\left[\varepsilon_{it} | \mathcal{F}_t\right] = 0$ and matrices B_{it} and Γ are unobservable to econometricians.

Methodology

Extension: Multiple Latent Variables - Model

The firm characteristics are aligned with stock returns because they are proxies for the unobservable latent factors: the demeaned and standardized characteristics $X_{it}^{a}, a = 1, \ldots, A$ of firm i at time t are related to B_{it} as

$$X_{it}^{a} = \left(B_{it} - \overline{B}_t\right)\delta^{a} + u_{it}^{a}$$

where $\Delta^{a}, a = 1, \ldots, A$ are $L \times 1$ matrices that determine the sensitivity of characteristics to the factors, $\overline{B}_t$ is the cross-sectional average of the factors, and u_{it}^{a} are components of characteristics that are unrelated to stock returns and that can be arbitrary correlated among themselves.

Methodology Extension: Multiple Latent Variables

Identifying assumptions

Assumption 1' (Distribution of latent factors) In each period t, $\overline{Var}\left(B_{it}\right) \xrightarrow{p} V_t$ as $N \to \infty$ and the matrix V_t is positive definite.

Assumption 2' (Distribution of characteristic loadings) $\widetilde{Var}\left(\Delta^a\right) \xrightarrow{p} \Lambda$ as $A \to \infty$ and the matrix Λ is positive definite.

Assumption 3' (Orthogonality of errors and latent factors) In each period t and for each characteristic a, $a = 1, \ldots, A$, $\overline{Cov}\left(B_{it}, u_{it}^a\right) \xrightarrow{p} 0$ as $N \to \infty$.

Assumption 4' (Orthogonality of errors and characteristic loadings) In each period t and for each stock i, $i = 1, \ldots, N$, $\widetilde{Cov}\left(\Delta^a, u_{it}^a\right) \xrightarrow{p} 0$ as $A \to \infty$.

Methodology Extension: Multiple Latent Variables

Estimation Procedure

The three-pass regression filter (3PRF) procedure is adopted from Kelly and Pruitt (2015):

- **Step 1:** Run cross-sectional regressions of the characteristics $X^a_{it-1}, a = 1, \ldots, A$ on the proxy variables $Z^l_{it}, l = 1, \ldots, L$ and retain the $L \times 1$ matrices of slopes λ^a_t.
- **Step 2:** For each stock $i, i = 1, \ldots, N$, run regressions of X^a_{it-1} and X^a_{it} on λ^a_t in the characteristics space and denote the obtained matrices of slopes are $\hat{B}_{it-1}$ and $\hat{B}_{it}$.
- **Step 3:** Run a cross-sectional regression of R_{it} on $\hat{B}_{it-1}, i = 1, \ldots, N$ and denote the obtained matrix of slopes as $\hat{\Gamma}_t$. Compute an estimate of expected returns as $\hat{\mu}_{it} = \hat{B}_{it}\hat{\Gamma}_t$.

Empirical Analysis
Data

- standard sources: CRSP monthly files, Compustat Fundamentals annuals files
- exchanges: NYSE, AMEX, NASDAX common stocks
- exclude financial firms, adjust for delisting, exclude stocks below \$5 at the end of the previous month
- use accounting data in calendar year t from financial statements with the fiscal year end in year $t-1$
- construct 26 firm characteristic variables

Empirical Analysis

Individual characteristics: individual characteristic are informative about future stock returns.

Table 1
Individual characteristics

Panel A: equal-weighted portfolios

	S	B/M	E/P	C/P	B	MOM	LTR	STR	IdVol	MAX	EIS	ROE	ROA
Means	−0.16	0.98	0.84	0.95	−0.46	1.59	−0.55	−1.04	−1.02	−0.96	−0.65	−0.07	1.54
Stds	3.71	4.55	3.75	4.26	7.33	5.86	3.59	4.97	6.43	5.97	3.47	2.16	4.30
t-stats	−1.01	4.88	5.11	5.08	−1.43	6.15	−3.50	−4.76	−3.59	−3.65	−3.26	−0.69	7.63

	AG	CI	IG	I/K	I/A	ACC	NOA	NS	ι	LV	O	TO	D
Means	−0.73	−0.25	−0.43	−0.55	−0.68	−0.52	−0.60	−0.34	−0.70	0.70	−0.64	−0.56	−0.72
Stds	2.86	1.60	1.98	4.59	2.53	2.32	3.62	2.30	3.62	5.17	3.11	4.79	3.92
t-stats	−5.82	−3.54	−4.95	−2.75	−6.08	−5.15	−3.79	−3.40	−4.40	3.09	−4.33	−2.65	−3.83

Panel B: value-weighted portfolios

	S	B/M	E/P	C/P	B	MOM	LTR	STR	IdVol	MAX	EIS	ROE	ROA
Means	−0.17	0.58	0.62	0.49	−0.19	1.34	−0.26	−0.20	−0.99	−0.62	−0.73	−0.07	0.90
Stds	4.02	4.99	4.97	5.08	7.46	7.34	5.07	6.19	7.18	7.13	5.19	3.53	5.29
t-stats	−0.94	2.66	2.83	2.20	−0.57	4.16	−1.17	−0.72	−3.12	−1.98	−2.43	−0.44	3.62

	AG	CI	IG	I/K	I/A	ACC	NOA	NS	ι	LV	O	TO	D
Means	−0.37	−0.32	−0.29	−0.35	−0.49	−0.38	−0.28	−0.57	−0.59	0.51	−0.12	−0.27	−0.40
Stds	3.70	3.27	3.43	6.43	3.15	3.95	4.04	2.97	4.19	5.74	4.22	5.15	5.15
t-stats	−2.28	−2.23	−1.93	−1.25	−3.55	−2.17	−1.57	−4.35	−3.19	2.00	−0.58	−1.19	−1.61

This table reports time-series averages, standard deviations, and *t*-statistics of hedge returns defined as a difference in monthly returns on top and bottom decile portfolios formed individually by 26 firm characteristics. The decile portfolios are equal-weighted in panel A and value-weighted in panel B. All variables are named in Section 2.1 and their construction is described in Appendix B. Average returns and their standard deviations are reported in percentage points. The sample is from 1970 to 2012 for all characteristics except for *EIS*, *ROA*, *O*, and *D*, for which it starts in 1987, 1975, 1976, and 1977, respectively.

Empirical Analysis

Filtered expected returns (*AFER*): high expected returns are generated by *AFER* without a commensurate increase in the volatility of returns.

Table 2
Returns on decile *AFER* portfolios

Panel A: no averaging of λ_t^a

	EW portfolios											VW portfolios										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**(10-1)**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**(10-1)**
Means	−0.10	0.43	0.70	1.00	1.06	1.33	1.38	1.46	1.57	1.59	1.69	0.10	0.62	0.74	0.78	0.90	1.06	0.97	1.10	1.15	1.31	1.21
Stds	8.02	7.09	6.45	5.98	5.66	5.49	5.48	5.63	5.91	6.82	7.05	8.00	6.93	6.15	5.88	5.46	5.27	5.22	5.50	5.82	6.76	7.60
t-stats	−0.28	1.38	2.45	3.79	4.25	5.48	5.71	5.89	6.04	5.30	5.45	0.28	2.04	2.72	3.00	3.73	4.56	4.23	4.54	4.49	4.39	3.61

Panel B: averaging of λ_t^a over past 5 years

	EW portfolios											VW portfolios										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**(10-1)**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**(10-1)**
Means	−0.30	0.49	0.82	1.07	1.15	1.32	1.39	1.41	1.48	1.58	1.88	−0.27	0.34	0.66	0.74	0.96	1.05	1.00	1.16	1.23	1.14	1.41
Stds	7.85	7.11	6.55	6.09	5.78	5.54	5.32	5.13	5.16	5.88	4.68	8.60	7.46	6.76	6.03	5.69	5.20	4.90	4.81	4.72	5.27	6.53
t-stats	−0.87	1.55	2.84	4.00	4.51	5.41	5.95	6.24	6.50	6.10	9.12	−0.72	1.05	2.21	2.79	3.83	4.57	4.65	5.47	5.93	4.91	4.91

Panel C: averaging of λ_t^a over past 10 years

	EW portfolios											VW portfolios										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**(10-1)**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**(10-1)**
Means	−0.35	0.48	0.84	1.09	1.18	1.34	1.35	1.43	1.46	1.59	1.94	−0.17	0.24	0.49	0.79	0.99	0.94	1.02	1.03	1.06	1.20	1.37
Stds	8.19	7.35	6.72	6.20	5.83	5.45	5.13	4.90	4.81	5.58	4.49	8.84	7.56	6.69	6.14	5.49	5.10	4.56	4.31	4.17	4.79	6.51
t-stats	−0.97	1.47	2.84	3.99	4.61	5.58	5.95	6.62	6.89	6.49	9.82	−0.44	0.73	1.66	2.92	4.08	4.17	5.09	5.41	5.77	5.68	4.78

Panel D: averaging of λ_t^a over all previous months

	EW portfolios											VW portfolios										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**(10-1)**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**(10-1)**
Means	−0.38	0.43	0.81	1.04	1.16	1.33	1.39	1.42	1.53	1.68	2.06	−0.28	0.27	0.55	0.85	0.95	1.00	1.06	1.14	1.10	1.28	1.56
Stds	8.49	7.60	6.91	6.23	5.84	5.42	5.09	4.81	4.64	5.16	5.02	8.93	7.58	6.58	5.85	5.36	4.93	4.47	4.14	4.20	4.94	6.52
t-stats	−1.02	1.29	2.65	3.81	4.50	5.58	6.21	6.69	7.48	7.38	9.31	−0.72	0.80	1.89	3.29	4.01	4.62	5.39	6.25	5.96	5.89	5.45

This table shows time-series averages, standard deviations, and *t*-statistics of monthly equal-weighted (EW) and value-weighted (VW) stock returns on decile portfolios formed by sorting firms on the aggregate filtered expected returns *AFER*. The columns (10–1) report the statistics for the difference in returns on the top and bottom portfolios. The variable *AFER* is constructed without time-series averaging of λ_t^a (in panel A), with averaging of λ_t^a over the most recent 5 years (in panel B), with averaging of λ_t^a over the most recent 10 years (in panel C), and with averaging of λ_t^a over all previous months (in panel D). The sample is from January 1970 to December 2012. The portfolios are rebalanced monthly. All returns and standard deviations are reported in percentage points.

Empirical Analysis

Filtered expected returns in subsamples

Table 5
Filtered expected returns in subsamples

	Sample periods					
	Hedge returns			***t*-stats**		
	(1)	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**
EW	2.06	2.15	1.92	9.31	10.12	4.21
VW	1.56	1.55	1.58	5.45	5.11	2.83

	Size portfolios						**Idiosyncratic volatility portfolios**					
	Hedge returns			***t*-stats**			**Hedge returns**			***t*-stats**		
	(1)	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**
EW	2.52	1.75	1.15	11.31	6.81	4.56	0.95	1.94	2.66	6.30	10.37	11.68
VW	2.45	1.68	0.93	10.11	6.50	3.53	1.03	1.68	1.94	4.57	6.12	5.28

This table reports the average differences in monthly equal-weighted (EW) and value-weighted (VW) returns on top and bottom decile *AFER* portfolios and their *t*-statistics in various subsamples. To denote subsamples we use the following notation. Sample periods: (1) full sample (January 1970–December 2012), (2) early sample (January 1970–December 1995), (3) late sample (January 1996–December 2012); size portfolios: (1) microcap, (2) small, (3) large; idiosyncratic volatility portfolios: (1) low volatility, (2) medium volatility, (3) high volatility. In all panels *AFER* is constructed by averaging λ_t^a from all previous months. All returns are reported in percentage points.

Alternative aggregation techniques

Advantage of PLS-based regression

- Factor analysis vs PLS: common factor in the characteristics produces much smaller hedge returns than *AFER* does.
- PCA vs PLS: the first principle component is a poor predictor of future returns, *AFER* statistically produce higher hedge returns.
- Fama-MacBeth vs PLS: regression based estimators of expected returns are likely to be less efficient; they are imprecise and even nonexistent when the number of characteristics is large; potential multicollinearity problems when characteristics are highly correlated.
- Rank-based approach vs PLS: theoretical and empirical advantage.
- Large cap stocks: the superior performance of the PLS-based aggregation technique is robust in the subsample of liquid large cap stocks.

Conclusion

Light et al. (2017) They find that the hedge returns produced by expected returns filtered from 26 characteristics are large, highly statistically significant, and exceed the hedge returns produced by each characteristic individually, which suggests:

1. there is commonality in the considered asset pricing anomalies
2. the relations between characteristics and returns are not a statistical fluke
3. individual characteristics are likely to be diverse signals about expected returns

Reference

Giglio, S., Kelly, B., and Pruitt, S. (2016). Systemic risk and the macroeconomy: An empirical evaluation. *Journal of Financial Economics Volume 119, Issue 3*, pages 457–471.

Huang, D., Jiang, F., Tu, J., and Zhou, G. (2015). Investor sentiment aligned: A powerful predictor of stock returns. *The Review of Financial Studies, Volume 28, Issue 3*, pages 791–837.

Kelly, B. and Pruitt, S. (2013). Market expectations in the cross section of present values. *Journal of Finance 68*, pages 1721–1756.

Kelly, B. and Pruitt, S. (2015). The three-pass regression filter: A new approach to forecasting using many predictors. *Journal of Econometrics, Volume 186, Issue 2*, pages 294–316.

Light, N., Maskov, D., and Rytchkov, O. (2017). Aggregation of information about the cross section of stock returns: a latent variable approach. *The Review of Financial Studies, Volume 30, Issue 4*, pages 1339–1381.