Taming the Factor Zoo - A Test of New Factors: Feng, Giglio, Xiu (JF 2020)

Presented by: Adarsh Kuthuru

December 8, 2020

Overview

- About the paper
- Theoretical Background
- Empirical approach
- Data
- Results

About the paper

Research Question:

- Evaluating the marginal contribution of new factors relative to the myriad of existing factors, and conducting appropriate statistical inference in high-dimensional setting

Contribution:

- Propose a regularized two-pass cross-sectional regression (DS LASSO) to establish the contribution of new factors g_t relative to a set of control factors h_t (benchmarks the factors against a large-dimensional set of existing ones - Belloni et al. (2014))
- Takes into account model selection mistakes
- The procedure leverages information from the cross section of the test assets in addition to the times-series of the factors

About the paper

Contribution:

- Demonstrate how their results differ starkly from using the risk premia of the factors or the standard Fama-French three factor model as control as opposed to its role in driving marginal utility

Findings:

- Several newly proposed factors (especially different versions of profitability) are useful in explaining asset prices, even after accounting for the large set of existing factors proposed up to 2012
- the SDF loadings' estimates for several factors are robust to changes in the tuning parameters, despite the fact that the models selected vary substantially when the tuning parameters are changed
- Applying their test recursively over time would have deemed only a small number of factors proposed in the literature significant

Model setup

Linear specification for the SDF:

$$m_t := \gamma_0^{-1} - \gamma_0^{-1}\lambda_v^\top v_t := \gamma_0^{-1}\left(1 - \lambda_g^\top g_t - \lambda_h^\top h_t\right) \tag{1}$$

where λ_g and λ_h are the SDF loadings of the factors g_t $(d \times 1)$ and h_t $(p \times 1)$

Expected returns:

$$\mathrm{E}\left(r_t\right) = \iota_n\gamma_0 + C_v\lambda_v = \iota_n\gamma_0 + C_g\lambda_g + C_h\lambda_h \tag{2}$$

where ι_n is a $n \times 1$ vector of 1s, $C_a = \mathrm{Cov}(r_t; a_t)$, for a = g; h; or v. For the estimation of λ_g, it is essential to characterize the cross-sectional dependence between C_g and C_h, so we write the cross-sectional projection of C_g onto C_h as:

$$C_g = \iota_n\xi^\top + C_h\chi^\top + C_e \tag{3}$$

where ξ is a $d \times 1$ vector, χ is a $d \times p$ matrix, and C_e is a $n \times d$ matrix of cross-sectional regression residuals.

Empirical approach

The regularized two-pass estimation proceeds as follows:
(1) Two-Pass Variable Selection
(1.a) Run a cross-sectional LASSO regression of average returns on sample covariances between factors in h_t and returns

$$\min_{\gamma,\lambda} \left\{ n^{-1} \left\| \bar{r} - \iota_n \gamma - \widehat{C}_h \lambda \right\|^2 + \tau_0 n^{-1} \|\lambda\|_1 \right\} \tag{4}$$

where, $\widehat{C}_h = \widehat{\mathrm{Cov}}\,(r_t, h_t)$
This step selects among the factors in ht those that best explain the cross section of expected returns. Denote $\widehat{I}_1$ as the set of indices corresponding to the selected factors in this step.

Empirical approach

(1.b) For each factor j in g_t (with j = 1;...; d), run a cross-sectional LASSO regression of $\widehat{C}_{g,\cdot,j}$ (the covariance between returns and the jth factor of g_t) on $\widehat{C}_h$ (the covariance between returns and all factors h_t):

$$\min_{\xi_j,\chi_{j,\cdot}} \left\{ n^{-1} \left\| \left(\widehat{C}_{g,\cdot,j} - \iota_n \xi_j - \widehat{C}_h \chi_{j,\cdot}^{\top} \right) \right\|^2 + \tau_j n^{-1} \left\| \chi_{j,\cdot}^{\top} \right\|_1 \right\} \tag{5}$$

This step identifies factors whose exposures are highly correlated to the exposures to g_t in the cross-section. This is the crucial second step in the double-selection algorithm, that searches for factors that may be missed by the first step but that may still induce large omitted variable bias in the estimation of λ_g if omitted. Denote $\widehat{I}_{2,j}$ as the set of indices corresponding to the selected factors in the jth regression, and $\widehat{I}_2 = \cup_{j=1}^{d} \widehat{I}_{2,j}$.

Empirical approach

(2) Post-selection Estimation Run an OLS cross-sectional regression using covariances between the selected factors from both steps and returns:

$$\left(\widehat{\gamma}_0, \widehat{\lambda}_g, \widehat{\lambda}_h\right) = \arg\min_{\gamma_0,\lambda_g,\lambda_h} \left\{ \left\| \bar{r} - \iota_n \gamma_0 - \widehat{C}_g \lambda_g - \widehat{C}_h \lambda_h \right\|^2 : \tag{6}$$

$$\lambda_{h,j} = 0, \quad \forall j \notin \widehat{I} = \widehat{I}_1 \bigcup | \widehat{I}_2 \right\}$$

Data

- Factor library contains 150 risk factors
- Monthly frequency for the period from July 1976 to December 2017
- Obtained from multiple sources (Ken French's data library, AQR data library, & respective author's websites)

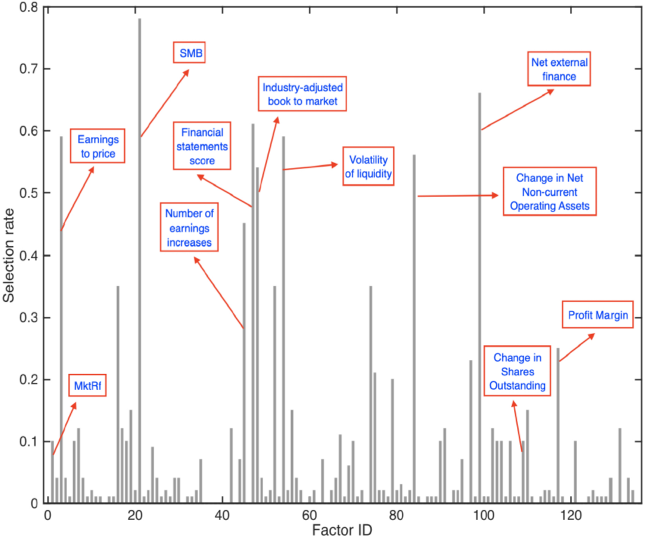
Results

Table 1: Testing for Factors Introduced in 2012-2016

		(1) DS		(2) SS		(3) FF3		(4) No Selection		(5) Avg. Ret.	
id	Factor Description	λ_s (bp)	tstat (DS)	λ_s (bp)	tstat (SS)	λ_s (bp)	tstat (OLS)	λ_s (bp)	tstat (OLS)	avg.ret. (bp)	tstat
136	Cash holdings	-34	-0.42	15	0.17	10	0.54	-18	-0.16	13	0.98
137	HML Devil	54	1.04	-13	-0.25	-100	-2.46**	68	0.84	23	1.46
138	Gross profitability	20	0.48	3	0.06	23	2.00**	13	0.26	15	1.45
139	Organizational Capital	28	0.92	-1	-0.03	20	1.91*	16	0.41	21	2.05**
140	Betting Against Beta	35	1.45	38	1.50	36	2.25**	49	1.49	91	5.98***
141	Quality Minus Junk	73	2.03**	4	0.11	39	3.10***	50	1.04	43	3.87***
142	Employee growth	43	1.36	-4	-0.12	-12	-0.89	18	0.37	8	0.83
143	Growth in advertising	-12	-1.18	0	0.03	12	1.32	-2	-0.13	7	0.84
144	Book Asset Liquidity	40	1.07	5	0.12	20	1.59	20	0.42	9	0.79
145	RMW	160	4.45***	15	0.41	20	1.80*	74	1.48	34	3.21***
146	CMA	38	1.10	0	0.01	3	0.28	7	0.14	26	3.02***
147	HXZ IA	51	2.11**	5	0.21	21	1.94*	40	1.08	34	4.17***
148	HXZ ROE	77	3.37***	23	0.83	33	2.92***	104	2.87***	57	4.99***
149	Intermediary Risk Factor	112	2.21**	60	1.19	4	0.08	22	0.32		
150	Convertible debt	-15	-1.36	-39	-3.22***	26	3.32***	17	1.01	11	1.70*

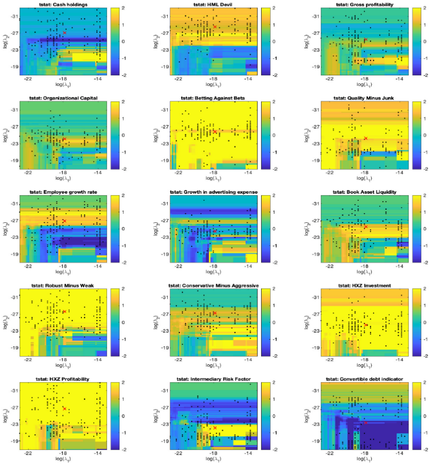
Results

Figure 1: Subsamples: Factor 1st Selection Rate



Results

Figure 2: Factors Introduced in 2012-2016: Robustness to Tuning Parameters (t-statistics)



Results

Table 2: Testing Factors Recursively by Year of Publication

Year	(1) # Assets	(2) # Controls	(3) New factors (IDs)											
1994	138	25	26	27										
1995	150	27	28	29	30									
1996	150	30	31	32	33									
1997	168	33	34											
1998	174	34	35	36	37	38	39	40	41	42	43	44		
1999	228	44	45	46										
2000	234	46	47	48	49	50	51							
2001	252	51	52	53	54	55	56	57	58					
2002	294	58	59	60	61									
2003	312	61	62	63	64	65	66							
2004	336	66	67	68	69	70	71	72	73	74				
2005	372	74	75	76	77	78	79	80	81	82	83	84	85	86
			87	88	89	90								
2006	456	90	91	92	93	94	95	96	97	98	99	100	101	102
2007	516	102	103	104	105	106	107	108						
2008	552	108	109	110	111	112	113	114	115	116	117	118	119	120
2009	618	120	121	122	123	124								
2010	636	124	125	126	127	128	129							
2011	666	129	130	131	132	133	134	135						
2012	702	135	136											
2013	708	136	137	138	139									
2014	720	139	140	141	142	143	144							
2015	738	144	145	146	147	148								
2016	750	148	149	150										

Results

Table 3: Robustness for Factors Introduced in 2012-2016

		(1) Bivariate 3 × 2		(2) Bivariate 5 × 5		(3) 202 Portfolios		(4) Elastic Net		(5) PCA	
id	Factor Description	λ_s (bp)	tstat (DS)	λ_s (bp)	tstat (DS)	λ_s (bp)	tstat (DS)	λ_s (bp)	tstat (DS)	λ_s (bp)	tstat (DS)
136	Cash holdings	-34	-0.42	34	0.40	131	0.89	-13	-0.14	-65	-0.62
137	HML Devil	54	1.04	15	0.29	56	0.57	62	1.23	-27	-0.51
138	Gross profitability	20	0.48	28	0.66	88	1.42	-11	-0.26	16	0.35
139	Organizational Capital	28	0.92	23	0.75	6	0.16	12	0.38	21	0.57
140	Betting Against Beta	35	1.45	43	1.94*	31	1.03	28	1.12	59	2.56***
141	Quality Minus Junk	73	2.03**	58	1.67	123	2.45**	74	2.13**	71	1.89*
142	Employee growth	43	1.36	12	0.34	54	1.34	51	1.49	-4	-0.09
143	Growth in advertising	-12	-1.18	6	0.57	17	1.30	9	0.74	-6	-0.57
144	Book Asset Liquidity	40	1.07	-24	-0.61	37	0.77	26	0.68	24	0.63
145	RMW	160	4.45***	104	3.13***	112	1.98**	125	3.43***	88	2.11**
146	CMA	38	1.10	19	0.59	33	0.52	32	0.85	18	0.44
147	HXZ IA	51	2.11**	44	1.87*	-45	-1.42	69	2.77***	36	1.31
148	HXZ ROE	77	3.37***	72	2.62***	116	2.22***	103	3.85***	41	1.46
149	Intermediary Risk Factor	112	2.21**	38	0.73	-16	-0.33	-16	-0.33	103	1.92*
150	Convertible debt	-15	-1.36	-6	-0.56	68	5.13***	-12	-1.08	-9	-0.88

References

Belloni, A., Chernozhukov, V., and Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81(2):608–650.