Internet Appendix to Factor Timing

Valentin Haddad
University of California, Los Angeles and NBER

Serhiy Kozak
University of Maryland

Shrihari Santosh
University of Colorado Boulder

The optimal factor timing portfolio is equivalent to the stochastic discount factor. We propose and implement a method to characterize both empirically. Our approach imposes restrictions on the dynamics of expected returns, leading to an economically plausible SDF. Market-neutral equity factors are strongly and robustly predictable. Exploiting this predictability leads to substantial improvement in portfolio performance relative to static factor investing. The variance of the corresponding SDF is larger, is more variable over time, and exhibits different cyclical behavior than estimates ignoring this fact. These results pose new challenges for theories that aim to match the cross-section of stock returns. (*JEL* G12, G14)

We appreciate helpful comments from Stijn Van Nieuwerburgh (editor) and two anonymous referees. We also thank John Campbell, Mikhail Chernov, John Cochrane, Julien Cujean, Robert Dittmar, Kenneth French, Stefano Giglio, Bryan Kelly, Ralph Koijen, Hugues Langlois, Lars Lochstoer, Mark Loewenstein, Tyler Muir, Stefan Nagel, Nikolai Roussanov, Avanidhar Subrahmanyan, and Michael Weber and seminar participants at AFA, Chicago, FIRS, LSE, Maryland, Michigan, University of Washington, and NBER for helpful comments and suggestions. This paper was previously circulated under the title "Predicting Relative Returns." Send correspondence to Serhiy Kozak, University of Maryland, 4453 Van Munching Hall, College Park, MD 20742; telephone: (301) 405-0703. E-mail: skozak@rhsmith.umd.edu.

Published by Oxford University Press on behalf of The Society for Financial Studies 2020.
doi:10.1093/rfs/Sample Advance Access publication February, 2020

Internet Appendix

I. Key formulas

We derive key results. First, start with a set of returns $R_{t+1}=[r_{1,t+1}\cdots r_{N,t+1}]'$. The maximum conditional squared Sharpe ratio is

$$\mathrm{SR}_t^2=\mathbb{E}_t[R_{t+1}]'\Sigma_{R,t}^{-1}\mathbb{E}_t[R_{t+1}]. \tag{A1}$$

I.A Expected conditional squared Sharpe ratio, $\mathbb{E}[\mathrm{SR}_t^2]$

If the assets are conditionally uncorrelated, then $\Sigma_{R,t}$ is diagonal and the formula becomes

$$\mathrm{SR}_t^2=\sum_{i=1}^{N}\frac{\mathbb{E}_t[r_{i,t+1}]^2}{\sigma_{i,t}^2}, \tag{A2}$$

where σ_i^2 is the conditional variance of the return on asset i. Assuming returns are homoskedastic and taking unconditional expectations we obtain

$$\mathbb{E}\left[\mathrm{SR}_t^2\right]=\sum_{i=1}^{N}\frac{\mathbb{E}\left(\mathbb{E}_t[r_{i,t+1}]^2\right)}{\sigma_i^2}. \tag{A3}$$

Substituting in the identity $\mathbb{E}\left(\mathbb{E}_t[r_{i,t+1}]^2\right)=\mathbb{E}[r_{i,t+1}]^2+\mathrm{var}(\mathbb{E}_t[r_{i,t+1}])$ we have

$$\mathbb{E}\left[\mathrm{SR}_t^2\right]=\sum_{i=1}^{N}\frac{\mathbb{E}[r_{i,t+1}]^2}{\sigma_i^2}+\sum_{i=1}^{N}\frac{\mathrm{var}(\mathbb{E}_t[r_{i,t+1}])}{\sigma_i^2}. \tag{A4}$$

Using the definition of R-squared, $R_i^2=1-\frac{\sigma_i^2}{\mathrm{var}\left(\mathbb{E}_t\left[r_{i,t+1}\right]\right)+\sigma_i^2}$, we obtain $\frac{R_i^2}{1-R_i^2}=\frac{\mathrm{var}\left(\mathbb{E}_t\left[r_{i,t+1}\right]\right)}{\sigma_i^2}$, which can be substituted in to get the formula in the paper:

$$\mathbb{E}[\mathrm{var}_t(m_{t+1})]=\mathbb{E}\left(\mathrm{SR}_t^2\right)=\sum_i\frac{\mathbb{E}[r_{i,t+1}]^2}{\sigma_i^2}+\sum_i\left(\frac{R_i^2}{1-R_i^2}\right). \tag{A5}$$

I.B Expected utility

We consider the perspective of an investor with mean-variance utility risk aversion parameter equal to γ. We maintain the assumption that $\Sigma_{R,t}$ is diagonal, constant, and known. Consider two portfolio strategies: (1) the static (s) strategy which does not use conditioning information and (2) the dynamic (d) strategy which can condition on $\mu_t=\mathbb{E}_t(R_{t+1})$. The optimal portfolio weights are given by

$$w_{t,s}=\frac{1}{\gamma}\Sigma_{R,t}^{-1}\mu \tag{A6}$$

$$w_{t,d}=\frac{1}{\gamma}\Sigma_{R,t}^{-1}\mu_t \tag{A7}$$

where $\mu=\mathbb{E}(R_{t+1})$ is the unconditional mean return. The date t expected utility for the agent under the two strategies are given by

$$U_{t,s}=\frac{1}{\gamma}\mu'\Sigma^{-1}\mu_t-\frac{1}{2\gamma}\mu'\Sigma^{-1}\mu \tag{A8}$$

$$U_{t,d}=\frac{1}{2\gamma}\mu_t'\Sigma^{-1}\mu_t, \tag{A9}$$

where expectations are taken under the same measure. Computing unconditional expectations we obtain

$$\mathbb{E}[U_{t,s}]=\frac{1}{2\gamma}\mu'\Sigma^{-1}\mu=\frac{1}{2\gamma}\sum_i\frac{\mathbb{E}[r_{i,t+1}]^2}{\sigma_i^2} \tag{A10}$$

$$\mathbb{E}[U_{t,d}]=\frac{1}{2\gamma}\mathbb{E}\left[\mu_t'\Sigma^{-1}\mu_t\right]=\frac{1}{2\gamma}\sum_i\frac{\mathbb{E}[r_{i,t+1}]^2}{\sigma_i^2}+\frac{1}{2\gamma}\sum_i\left(\frac{R_i^2}{1-R_i^2}\right). \tag{A11}$$

Therefore the "timing" term $\sum_i\left(\frac{R_i^2}{1-R_i^2}\right)$ exactly captures the increase in average utility obtained by using conditioning information.

I.C Total $\mathbf{R^2}$

Again start with a set of returns $R_{t+1}=[r_{1,t+1}\cdots r_{N,t+1}]'$ with arbitrary cross-correlations. Define the total R^2 as

$$R^2_{\text{total}}\equiv\frac{\text{tr}[\text{cov}(\mathbb{E}_t[R_{i,t+1}])]}{\text{tr}[\text{cov}(R_{i,t+1})]}, \tag{A12}$$

where tr is the trace function. By similarity invariance of trace, this is equal to

$$R^2_{\text{total}}\equiv\frac{\text{tr}[Q'\text{cov}(\mathbb{E}_t[R_{i,t+1}])Q]}{\text{tr}[Q'\text{cov}(R_{i,t+1})Q]}, \tag{A13}$$

where and Q is *any* orthogonal matrix ($Q'=Q^{-1}$). Next assume returns are homoskedastic, that is, $\Sigma_{R,t}$ is constant. This leads to the eigendecomposition $\Sigma_{R,t}=Q\Lambda Q'$. Denoting $PC_{t+1}=Q'R_{t+1}$ and substituting in we have

$$R^2_{\text{total}}=\frac{\text{tr}[\text{cov}(\mathbb{E}_t[PC_{t+1}])]}{\text{tr}[\Lambda]+\text{tr}[\text{cov}(\mathbb{E}_t[PC_{t+1}])]}, \tag{A14}$$

where we use $\text{cov}(PC_{t+1})=\Lambda+\text{cov}(\mathbb{E}_t[PC_{t+1}])$ and additivity of trace. Next use $\frac{R_i^2}{1-R_i^2}=\frac{\text{var}\left(\mathbb{E}_t\left[PC_{i,t+1}\right]\right)}{\lambda_i}$ to obtain

$$R^2_{\text{total}}=\sum_{i=1}^{K}\left(\frac{R_i^2}{1-R_i^2}\right)\frac{\lambda_i}{\lambda}, \tag{A15}$$

where

$$\lambda=\text{tr}[\Lambda]+\text{tr}[\text{cov}(\mathbb{E}_t[PC_{t+1}])] \tag{A16}$$

$$=\sum_{i=1}^{K}\frac{\lambda_i}{1-R_i^2}\approx\sum\lambda_i. \tag{A17}$$

I.D Number of PCs

Start with prior beliefs on the maximum squared Sharpe ratio $\mathbb{E}[\text{SR}_t^2]\leq s^\star$ and the total R^2 $R^2_{\text{total}}\geq r^\star$.[31] Given these beliefs how many PCs should we include? Under

[31] Here we ignore the static component of the Sharpe ratio.

the view that all included k PCs contribute equally to the total R^2, using Equation (7) and Equation (8), we can equivalently write

$$r^{\star} \leq k\left(\frac{R_i^2}{1-R_i^2}\right)\frac{\lambda_i}{\lambda} \tag{A18}$$

$$s^{\star} \geq \sum_{i=1}^{k}\left(\frac{R_i^2}{1-R_i^2}\right). \tag{A19}$$

Note this is analogous to the setup in Kozak et al. (2018) who assume that all included PCs contribute equally to cross-sectional heterogeneity in expected returns when determining the number of PCs to include. Combining these expressions we obtain the final formula:

$$\frac{r^{\star}}{s^{\star}} \leq \left[\frac{1}{k}\sum_{i=1}^{k}\frac{\lambda}{\lambda_i}\right]^{-1}. \tag{A20}$$

By inspection, the weaker the factor for a given set of assets the fewer PCs one may include given prior beliefs.

II. Additional Results

We report supplemental empirical results.

II.A Out-of-sample

Our main out-of-sample analysis uses a sample split where all parameters are estimated using the first half and used to construct OOS forecasts in the second half of the data. We consider two alternatives, expanding and rolling window analysis. For both, the OOS analysis begins on the same date as the main estimation, but predictive regression coefficients are reestimated each month. For rolling window, we use a twenty year (240 month) sample. Table A.1 presents results from the alternative OOS methods. The first row shows the coefficient estimate. The second row shows asymptotic t-statistics. The third and fourth rows show coefficients estimated from the first and second half data, respectively. The fifth shows the in-sample R^2. The next three rows give OOS R^2 based on split sample, expanding window, and rolling window analysis. PCs 1 and 4 show remarkable stability of estimated coeffcients and substantial OOS R^2 using all three methods. This stability reflects the precision of the coefficient estimates documented by the t-statistics. Finally, the last row reports a reverse OOS R^2 where estimation is conducted in the second half of the sample and we evaluate the performance of the prediction in the first half of the sample. Here again, PC1 and PC4 have sizable R^2. Interestingly, PC3 and PC5 have larger R^2 than in the baseline, while PC2 does poorly, consistent with the instability of the predictive coefficient across periods. Overall, these various OOS approaches lead to similar conclusions to our baseline in terms of predictability of the dominant components.

II.B Finite Sample Bias

The relative lack of Stambaugh-type bias for the PCs may be surprising given that bias for the aggregate market is large. However, this difference arises for two reasons. Assuming an AR(1) process for the predictor, Stambaugh (1999) shows that

$$E\left(\hat{\beta}-\beta\right) = c\rho_{xy}E\left(\hat{\rho}_x-\rho_x\right),$$

Table A.1
Predicting Dominant Equity Components with BE/ME ratios

	MKT	PC1	PC2	PC3	PC4	PC5
Own *bm* Full	0.76	4.32	1.62	1.80	4.86	1.56
	(1.24)	(4.31)	(1.81)	(2.01)	(3.74)	(0.78)
Own *bm* 1st	1.46	3.77	1.37	2.62	5.66	2.74
Own *bm* 2nd	2.79	4.91	7.68	2.83	4.31	2.14
R^2 Full	0.29	3.96	0.74	0.56	3.59	0.50
Split R^2	1.00	4.82	0.97	0.47	3.52	0.55
Expanding R^2	-0.53	4.43	-0.32	-0.30	2.59	-1.31
Rolling R^2	-0.28	3.04	-0.50	-1.10	2.47	-1.61
Reverse R^2	0.09	2.40	-7.30	1.17	2.65	1.62

We report results from predictive regressions of excess market returns and five PCs of long-short anomaly returns. The market is forecast using the log of the aggregate book-to-market ratio. The anomaly PCs are forecast using a restricted linear combination of anomalies' log book-to-market ratios with weights given by the corresponding eigenvector of pooled long-short strategy returns. The first row shows the coefficient estimate. The second row shows asymptotic t-statistics estimated using the method of Newey and West (1987). The third and fourth rows show coefficients estimated from the first and second half data, respectively. The fifth shows the in-sample R^2. The next three rows give OOS R^2 based on split sample, expanding window, and 240 month rolling window analysis. The last row reports a reverse OOS R^2 where estimation is conducted in the second half of the sample and performance is measured in the first half.

Table A.2
Stambaugh Bias

	MKT	PC1	PC2	PC3	PC4	PC5
Persistence	0.87	0.52	0.61	0.53	0.27	0.44
Error correlation	-0.84	-0.67	-0.33	-0.34	-0.30	-0.15

The first row reports annualized AR(1) coefficients of *bm* ratios, estimated from monthly data $\left(\rho_{annual} = \rho_{monthly}^{12}\right)$. The second row reports the contemporaneous correlation of innovations to returns and *bm* ratios assuming a VAR(1) data-generating process.

where $\hat{\beta}$ is the estimated predictive coefficient, ρ_{xy} is the contemporaneous correlation of innovations to x and y, and ρ_x is the autocorrelation of x. Marriott and Pope (1954) show that the bias in $\hat{\rho}_x$ is approximately proportional to ρ_x itself. Hence, the overall bias in $\hat{\beta}$ is proportional to $\rho_x \rho_{xy}$. We empircally estimate these quantities for the market and the anomaly PC portfolios to help decompose the lower simulated bias for PCs shown in Table 2.

In the first row of Table A.2 we report annualized AR(1) coefficients estimated by OLS from monthly data, $\rho_{annual} = \rho_{monthly}^{12}$. Unsurprisingly, they are much smaller for PCs relative to the aggregate market. By estimating the restricted VAR(1) assumed in Stambaugh (1999), we obtain estimates of the error correlation, shown in the second row of Table A.2. The error correlation is also substantially smaller for the PCs, further reducing the bias in estimated predictive coefficients.

II.C Macro Predictors

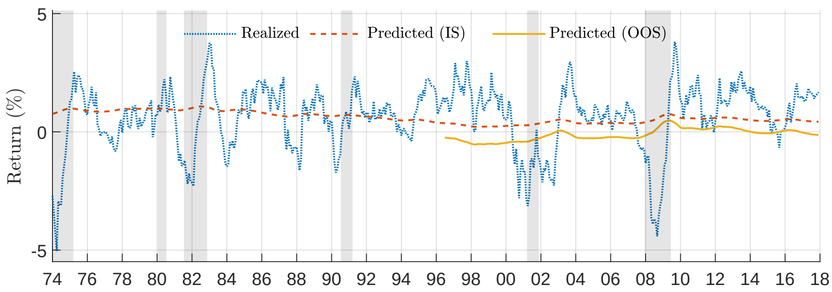
It is possible that price ratios are useful return forecasters of anomaly returns, but their predictive ability is subsumed by standard aggregate return predictors. We explore this by including the aggregate dividend-to-price ratio (D/P), cyclically-adjusted earnings-to-price (CAPE), lagged realized volatility, the term premium,

Table A.3
Including Macro Predictors

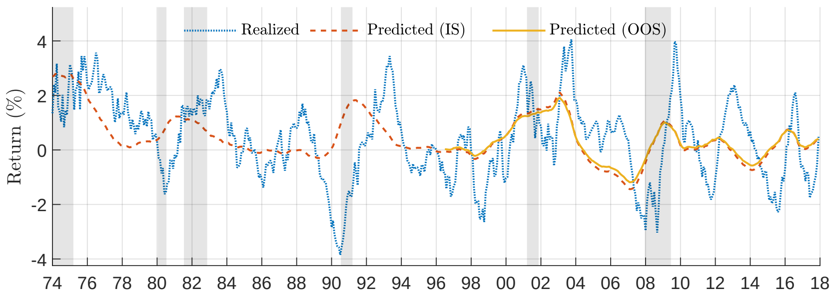
	MKT	PC1	PC2	PC3	PC4	PC5
Baseline	0.76	4.32	1.62	1.80	4.86	1.56
	(1.24)	(4.31)	(1.81)	(2.01)	(3.74)	(0.78)
	0.29	3.96	0.74	0.56	3.59	0.50
D/P	1.09	4.17	6.82	3.02	4.48	2.45
	(0.50)	(4.36)	(3.40)	(2.49)	(3.29)	(1.47)
	0.30	4.02	4.55	0.83	3.79	0.87
CAPE	1.21	4.31	1.97	2.64	4.44	1.89
	(1.42)	(4.29)	(1.90)	(2.96)	(3.65)	(1.06)
	0.39	3.97	0.77	0.87	4.14	0.77
Volatility	0.74	4.57	1.03	1.82	4.88	1.47
	(1.17)	(4.09)	(1.40)	(2.05)	(3.89)	(0.74)
	0.48	4.09	2.04	0.84	3.66	0.59
Term Premium	0.82	4.32	1.61	1.79	5.02	1.36
	(1.33)	(4.24)	(1.82)	(1.85)	(3.60)	(0.66)
	0.53	4.06	0.74	0.56	3.83	0.58
Corp. Spread	0.68	4.02	1.35	1.77	4.64	1.55
	(0.99)	(3.38)	(1.62)	(1.96)	(3.57)	(0.77)
	0.36	4.13	1.42	0.58	4.07	0.52
CAY	0.75	4.30	1.61	1.87	5.19	1.68
	(1.25)	(4.17)	(1.74)	(1.78)	(4.32)	(0.79)
	0.57	3.97	0.74	0.57	4.55	0.57
GDP growth	0.69	4.62	1.39	1.81	4.85	1.51
	(1.09)	(4.36)	(1.75)	(2.06)	(3.78)	(0.75)
	0.44	4.09	1.55	0.59	3.59	0.70
Sentiment	0.65	4.15	1.88	1.49	3.85	1.80
	(1.15)	(4.12)	(2.40)	(1.75)	(2.71)	(0.89)
	0.37	4.06	2.28	1.17	4.62	1.29

We we report the multivariate coefficients, t-statistics on the bm ratios and full sample R^2 values. The first row repeats the baseline estimates from Table 2.

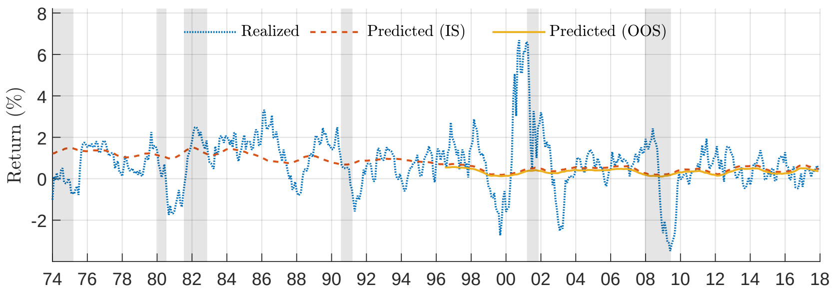
corporate bond yield spread, consumption-to-wealth ratio from Lettau and Ludvigson (2001) (CAY), GDP growth, and aggregate sentiment from Baker and Wurgler (2006). We include each of these additional predictors one at a time to the regressions of the market and PC returns on their own bm. In Table A.3 we report the multivariate coefficients, t-statistics on the bm ratios and full sample R^2 values. The first row repeats the baseline estimates from Table 2. The remaining rows show that macro variables do not even partially drive out the price ratios when predicting returns. This is not surprising. Even if we knew the "true" macro variables that drive time-variation in expected returns, the empirically measured values are likely extremely noisy since quantities like consumption, wealth, and gdp are not directly observable. Price ratios, by contrast, are likely much better measured expected return proxies. Daniel and Moskowitz (2016) find that aggregate market volatility predicts returns on the momentum portfolio. Among the largest five PCs, we find that market volatility predicts only PC2, which has a large loading on momentum.



(a) Market

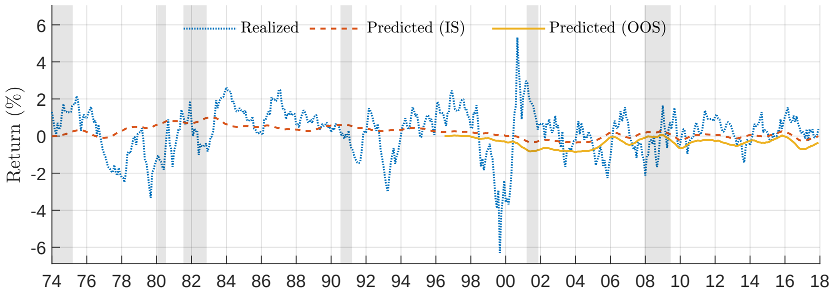


(b) PC1

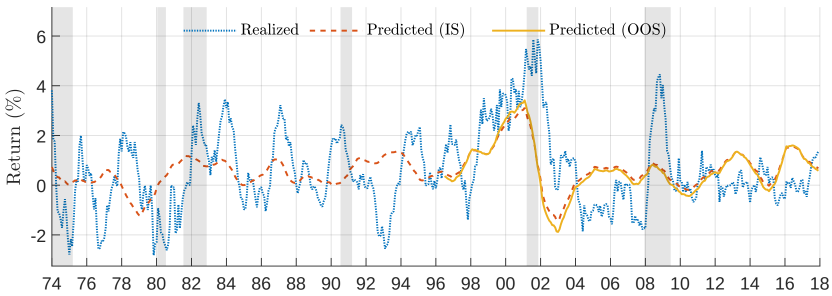


(a) PC2

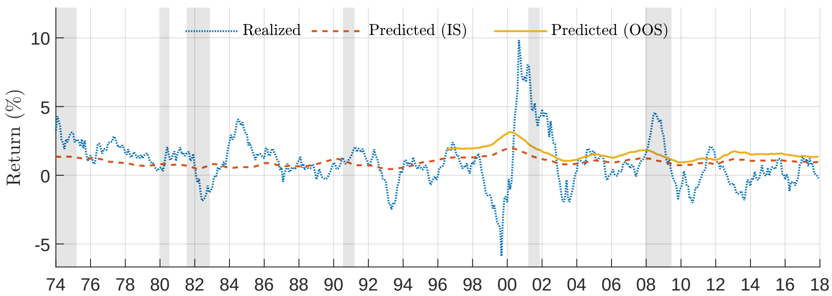
Figure A.1
Realized and Predicted Return (Part I)
The plot shows realized returns along with full sample and out-of-sample forecasts of returns on the aggregate market and first five PC portfolios of the fifty anomalies.



(b) PC3



(a) PC4



(a) PC5

Figure A.1
Realized and Predicted Return (Part II)
The plot shows realized returns along with full sample and out-of-sample forecasts of returns on the aggregate market and first five PC portfolios of the fifty anomalies.

II.D Forecast and Realized Returns

Figure A.1 shows realized returns along with full sample and out-of-sample forecasts of returns on the aggregate market and first five PC portfolios of anomalies.

II.E Anomaly Return Properties

Table A.3 shows annualized mean excess returns on the fifty anomaly long-short portfolios as well as the underlying characteristic-sorted decile portfolios.

Table A.3
Part I: Anomaly portfolios mean excess returns, %, annualized

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P10-P1
1. Size	6.5	8.5	9.0	9.7	9.4	10.1	9.7	10.5	9.9	9.6	3.1
2. Value (A)	5.8	8.0	9.0	7.6	8.8	9.1	9.0	9.2	9.0	12.2	6.4
3. Gross Profitability	6.0	6.0	6.9	6.4	8.5	7.6	8.2	7.1	7.9	9.8	3.8
4. Value-Profitablity	4.7	6.4	4.9	7.1	9.1	8.6	11.1	11.9	12.1	13.7	9.0
5. F-score	6.9	-	-	-	-	-	-	-	-	7.9	1.0
6. Debt Issuance	7.0	-	-	-	-	-	-	-	-	8.7	1.7
7. Share Repurchases	7.0	-	-	-	-	-	-	-	-	8.4	1.4
8. Net Issuance (A)	3.5	5.8	9.4	8.8	7.8	7.8	7.1	9.1	8.9	11.8	8.3
9. Accruals	5.0	6.7	6.1	7.5	7.8	7.8	8.7	7.7	10.3	9.0	4.0
10. Asset Growth	5.8	7.4	7.9	7.8	8.1	7.7	7.9	9.3	10.6	10.0	4.2
11. Asset Turnover	4.8	7.3	6.8	7.0	8.2	9.1	9.6	7.6	10.2	9.8	5.0
12. Gross Margins	6.9	7.5	8.7	7.7	8.7	7.2	8.1	7.5	6.5	7.5	0.6
13. Earnings/Price	4.6	5.8	7.2	7.9	7.7	7.9	10.4	9.3	9.7	12.3	7.6
14. Cash Flows/Price	5.3	8.1	6.7	8.6	8.7	9.1	8.4	9.7	11.4	11.2	5.9
15. Net Operating Assets	3.8	7.0	7.5	4.5	8.3	8.1	8.5	8.3	9.4	9.1	5.2
16. Investment/Assets	5.2	5.7	8.3	7.0	8.9	7.2	8.1	9.2	9.0	11.0	5.8
17. Investment/Capital	7.0	7.3	6.9	8.0	7.6	9.0	7.8	8.2	9.0	9.9	2.9
18. Investment Growth	5.4	8.7	7.4	7.1	7.0	7.9	8.6	8.4	10.3	9.1	3.7
19. Sales Growth	7.9	7.6	7.9	7.0	8.1	9.1	7.3	8.4	9.3	7.3	-0.5
20. Leverage	6.2	7.3	7.4	10.8	7.9	8.6	9.2	9.2	9.4	8.9	2.7
21. Return on Assets (A)	4.5	8.8	7.9	8.1	7.7	7.6	7.9	8.3	7.1	7.7	3.2
22. Return on Book Equity (A)	6.4	7.3	7.0	8.2	7.0	8.1	7.1	8.0	6.9	8.4	2.0

Columns P1 through P10 show mean annualized returns (in %) on each anomaly portfolio net of risk-free rate. The column P10-P1 lists mean returns on the strategy which is long portfolio 10 and short portfolio 1. Excess returns on beta arbitrage portfolios are scaled by their respective betas. F-score, Debt Issuance, and Share Repurchases are binary sorts; therefore only returns on P1 and P10 are reported for these characteristics. Portfolios include all NYSE, AMEX, and NASDAQ firms; however, the breakpoints use only NYSE firms. Monthly data from January 1974 to December 2017.

Table A.3
Part II: Anomaly portfolios mean excess returns, %, annualized

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P10-P1
23. Sales/Price	5.5	6.7	7.4	8.9	9.5	9.4	9.9	11.4	11.5	13.1	7.7
24. Growth in LTNOA	7.4	6.9	7.1	9.1	6.5	7.8	7.3	8.6	8.6	8.4	1.0
25. Momentum (6m)	9.9	9.6	9.1	9.1	8.3	8.6	7.5	6.1	7.9	11.2	1.3
26. Value-Momentum	6.8	8.6	7.4	8.1	8.9	9.7	10.1	9.1	8.4	11.5	4.7
27. Value-Momentum-Prof.	6.4	8.3	8.2	8.8	7.6	5.9	8.2	9.3	11.9	14.8	8.4
28. Short Interest	7.1	6.6	9.1	9.4	8.7	7.1	7.7	6.5	5.1	5.8	-1.4
29. Momentum (12m)	-0.3	5.6	7.0	8.1	6.6	7.4	7.6	9.6	9.4	12.7	13.0
30. Industry Momentum	6.6	6.2	8.5	5.8	8.2	10.4	8.2	7.4	9.6	9.5	2.9
31. Momentum-Reversals	5.6	7.5	8.0	7.6	8.0	9.7	7.8	9.8	9.5	12.5	6.9
32. Long Run Reversals	7.4	7.4	8.2	9.0	8.6	9.0	8.8	9.9	10.6	11.8	4.4
33. Value (M)	6.4	7.0	7.3	7.4	8.7	7.9	9.7	7.8	12.8	12.3	5.9
34. Net Issuance (M)	4.6	6.1	10.8	8.8	9.2	7.7	7.9	8.7	10.6	11.2	6.6
35. Earnings Surprises	5.0	5.2	5.8	7.7	7.4	8.3	7.6	8.0	8.9	11.5	6.4
36. Return on Book Equity (Q)	2.4	6.3	7.4	5.4	6.4	7.1	8.2	8.2	7.8	9.8	7.5
37. Return on Market Equity	1.3	2.2	7.1	6.4	7.8	7.7	8.7	11.1	12.1	15.8	14.4
38. Return on Assets (Q)	2.8	5.3	8.1	7.9	7.9	7.6	8.8	8.1	7.6	8.6	5.8
39. Short-Term Reversals	4.0	5.0	7.2	7.3	7.4	8.5	9.5	9.9	10.3	8.4	4.4
40. Idiosyncratic Volatility	0.9	8.9	11.4	8.5	10.6	9.1	8.3	8.2	7.9	7.5	6.7
41. Beta Arbitrage	3.9	4.0	5.1	7.3	8.6	10.2	11.2	11.8	14.6	17.2	13.3
42. Seasonality	4.0	4.4	6.7	6.3	8.5	7.4	7.9	7.6	9.8	13.2	9.2
43. Industry Rel. Reversals	2.6	4.2	4.9	6.3	6.8	8.2	9.6	11.6	13.3	13.1	10.6
44. Industry Rel. Rev. (L.V.)	1.7	5.2	5.2	6.8	6.6	7.4	9.8	10.8	13.8	15.6	13.9
45. Ind. Mom-Reversals	4.1	5.3	6.2	6.3	8.4	7.9	8.4	9.5	10.4	14.7	10.6
46. Composite Issuance	4.7	6.5	6.6	7.1	8.0	8.0	7.5	8.1	10.3	10.8	6.1
47. Price	6.1	9.5	9.2	10.8	9.2	8.9	7.9	7.9	7.9	6.5	0.5
48. Share Volume	7.2	8.7	7.3	7.6	8.1	6.8	8.4	7.3	6.9	6.8	-0.4
49. Duration	5.4	7.5	9.0	8.3	9.3	10.1	9.8	9.5	11.0	11.8	6.5
50. Firm age	7.0	9.1	6.0	9.8	6.4	8.8	10.0	8.5	7.3	7.7	0.7

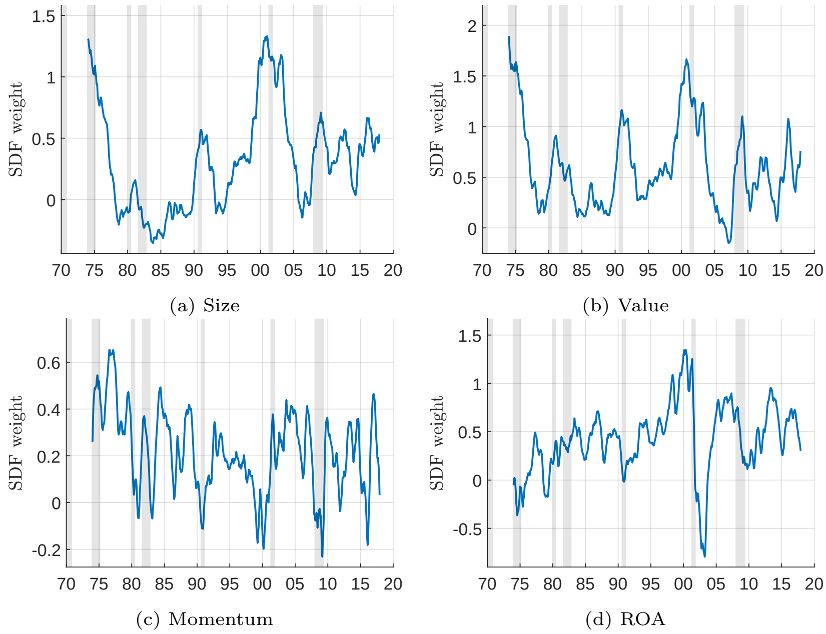


Figure A.2
Anomaly SDF Weights
The plot shows implied SDF coefficients on the size, value, momentum and ROA anomaly portfolios.

Figure A.2 gives the time-series of implied SDF coefficients on the size, value, momentum and ROA anomaly portfolios.

Figure A.3 shows the annualized unconditional mean return and standard deviation of conditional mean return on the fifty anomaly portfolios. Unconditional mean returns are computed as sample average returns. Standard deviation of conditional mean return are model implied based on the expected returns of five PC portfolios. The cross-sectional correlation of these two quantities is -20%.

II.F Role of the Rebalancing Frequency

Table A.4 studies the role of the rebalancing frequency for factor timing. To focus on anomaly timing, we report statistics for the "pure anomaly timing" strategy which always has zero weight in the market and has zero weight on average in each of the fifty anomaly portfolios.[32] We change portfolio weights on the anomalies monthly, quarterly, semi-annually, or annually. The first two rows report the Sharpe ratio and expected utility performance measures. The third row reports portfolio turnover, which we construct as follows. Given portfolio weights $w_{i,t}$ on anomaly i at date t, we construct period t turnover as $\sum_i |w_{i,t} - w_{i,t-1}| / \sum_i |w_{i,t-1}|$ which measures absolute trading scaled by gross exposure.[33] We report the average of this monthly measure over our sample. The last row of Table A.4 reports the correlation of

[32] The factor timing strategy has about 25-30% lower turnover at all rebalancing frequencies.

[33] Since our portfolios are all zero cost excess returns, the standard definition which divides by portfolio equity makes little sense in this context.

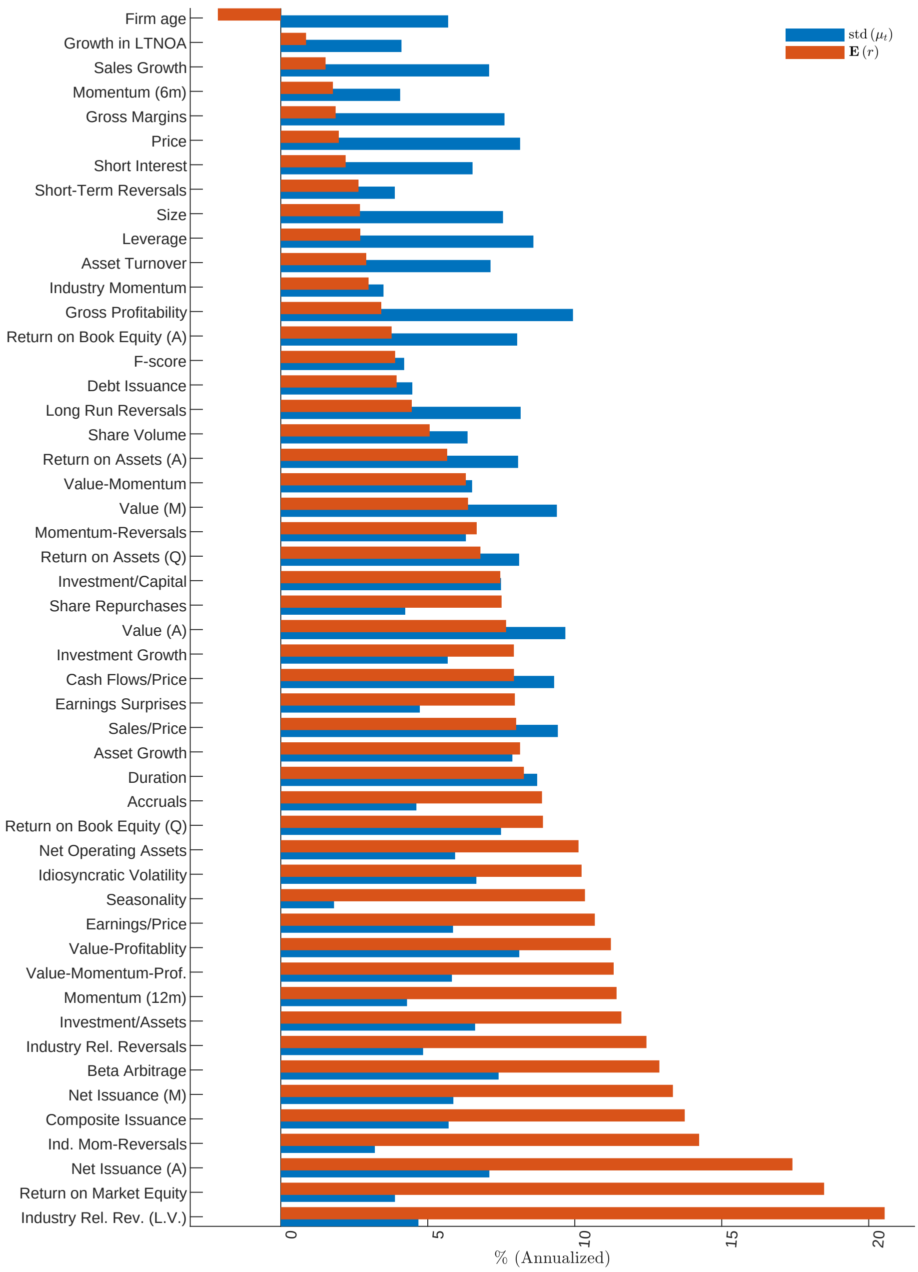


Figure A.3
Anomaly Expected Returns
The plot shows the annualized unconditional mean return and standard deviation of conditional mean return on the fifty anomaly portfolios.

Table A.4
Rebalancing frequency

	Monthly	Quarterly	Semi-annual	Annual
Sharpe ratio	0.71	0.72	0.82	0.79
Expected utility	1.26	1.08	0.88	0.81
Turnover	0.41	0.20	0.13	0.08
Correlation	1.00	0.98	0.92	0.88

We report the average unconditional Sharpe ratio, expected utility for a mean-variance investor, monthly portfolio turnover, and correlation with the monthly strategy with various rebalancing frequencies. Turnover is measured as the sum of absolute changes in portfolio weights divided by the sum of absolute initial portfolio weights.

each of the portfolios with the baseline monthly rebalanced return. Interestingly, the performance of the portfolios does not deteriorate meaningfully. The unconditional Sharpe ratio actually increases from 0.71 to 0.79 with annual rebalancing. Expected utility declines from 1.26 to 0.81, still a substantial value. For comparison, the static factor investing strategy yields an expected utility of 1.66 so even with annual rebalancing, timing benefits are economically meaningful. The correlation of the slower strategies with our baseline further confirms that the strategies are not that different. Even with annual rebalancing the correlation drops to only 0.88, showing that lowering the rebalancing frequency does not generate substantial tracking error. The signaling value of the predictors we use is sufficiently persistent to be used without continuous tracking.

Our anomaly timing strategy has a monthly turnover of 41%. Changing nearly half of positions each month might seem large, but it is important to remember that nothing in the construction of our strategy imposes a smooth trading path. The lower rebalancing frequencies drastically lower the turnover rate down to 8% with annual rebalancing. It is tempting to conclude that that these strategies are implementable in practice. Indeed, these numbers are in line with usual trading activity of investment funds. Griffin and Xu (2009) show that the median hedge fund has 8.5% monthly turnover and even the median mutual fund has 5% turnover. However, to reach a firm conclusion in terms of implementability, one would need a clear model of transaction costs. In addition, the transaction costs would likely depend of the scale at which the strategies are implemented.

II.G Conditional Variance of the SDF

Figures A.4 and A.5 show conditional variance of SDFs, as well as the relationship between SDF variance and inflation.

II.H Volatility timing

As discussed in Moreira and Muir (2017), optimal timing strategies rely not only on estimates of conditional expected returns, but also conditional volatilities. Going back to our one-asset example at the beginning of Section 3.1, consider the situation where volatility changes independently from expected returns. Then the average squared Sharpe ratio becomes

$$\left(\mathbb{E}[\mu_t]^2 + \mathrm{var}[\mu_t]\right)\left(\mathbb{E}\left[\frac{1}{\sigma_t}\right]^2 + \mathrm{var}\left[\frac{1}{\sigma_t}\right]\right),$$

the gains from timing returns and volatility are multiplicative.

In our multivariate setting, we need to construct estimates of $\Sigma_{Z,t}$. We proceed as follows. For each of our five principal components and the market returns, we compute the realized volatility of daily returns during the previous month. We use

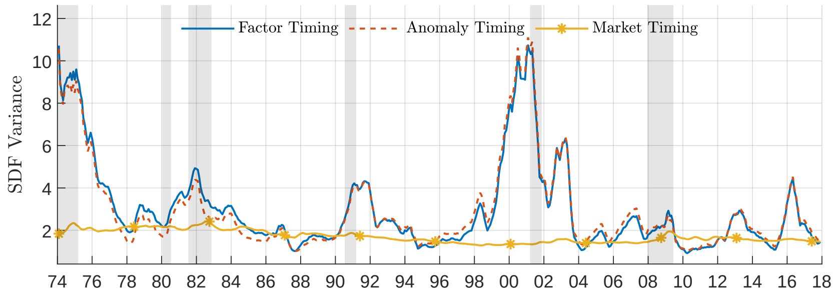


Figure A.4
Conditional Variance of SDFs
This figure plots the conditional variance of the SDF constructed under various sets of assumptions. "Factor timing" (solid blue line) is our full estimate, which takes into account variation in the means of the PCs and the market. "Anomaly timing" (dashed red line) imposes the assumption of no market timing: the conditional expectation of the market return is replaced by its unconditional counterpart. Conversely, "Market timing" (starred yellow line) allows for variation in the mean of the market return, but not the means of the factors.

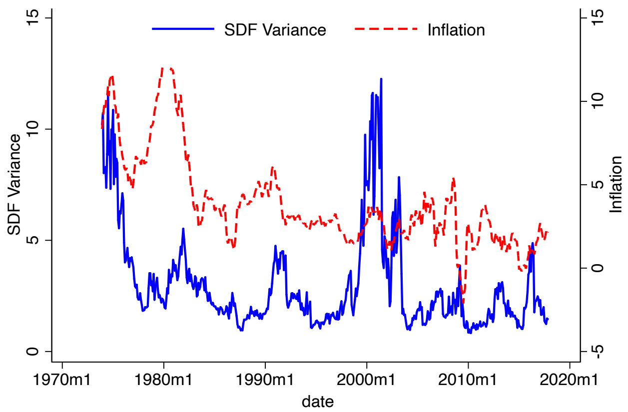


Figure A.5
Variance of the SDF and inflation
This figure plots the conditional variance of the SDF (solid blue line), and inflation rate over the previous year (dashed rate line). The SDF variance is constructed using the predictive regressions reported in Table 2. The inflation rate is the annual log change in the CPI.

these realized variances to create a forecast of the squared monthly prediction errors in the following month using a simple regression for each return series. These forecasts constitute the diagonal elements of $\Sigma_{Z,t}$. We confirmed that using GARCH(1,1) volatility forecasts leads to similar conclusions. We further assume that the five principal components and the market are *conditionally* orthogonal, and set the off-diagonal elements to 0.[34]

---

[34] The five components and the market are *unconditionally* orthogonal by construction.

Table A.5
Volatility timing

	Means	Variances	Both
$E[\text{var}_t(m_{t+1})]$	2.96	2.19	3.54
$\text{std}[\text{var}_t(m_{t+1})]$	2.17	0.74	2.06

We report the mean and standard deviation of the conditional variance of the SDF based on three estimates. The first column uses the SDF variance shown in Figure 2 based on return forecasts in Table 2 and assumes returns are homoskedastic. The second column assumes returns are not predictable and uses estimates of conditional return variances constructed from a regression of squared forecast errors on lagged realized variance. The final column combines both mean and variance forecasting.

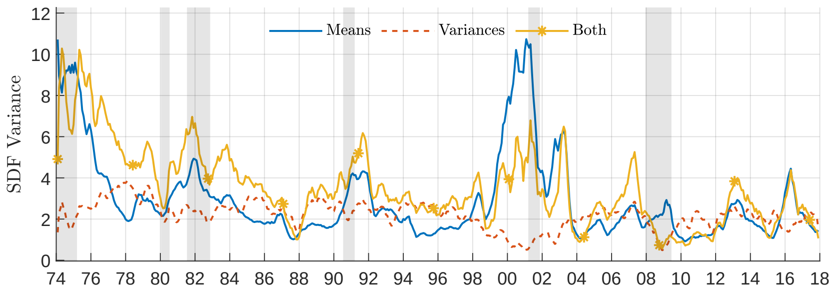


Figure A.6
Conditional Variance of SDFs
This figure plots the model-implied conditional variance of the SDF constructed in three ways. The solid blue line uses only timing of conditional means but constant variance. The red dashed line ignores predictability of returns but times variances. The yellow starred line times both means and variances.

In Section 4.4 we report statistical properties of estimated stochastic discount factors which incorporate time-varying means, variances, or both. We compute the mean and standard deviation of the corresponding SDF variances. Figure A.6 shows the time-series of the conditional SDF variance implied by each of these three estimates. Examining the two time series, we note that the largest volatility spikes tend to mitigate the effect of high expected returns on SDF variance. For example this coincidence occurs during the Internet boom and bust, and also during the financial crisis of 2008.

III. Statistical Approach

We first discuss an alternative statistical motivation behind our methodology, then derive some useful statistical properties.

III.A An Alternative Statistical Motivation

Another way to approach our empirical exercise is to look for common sources of variation in risk premia across base assets or factors. For example, starting from a vector of candidate predictors X_t, we want to assess their usefulness to forecast the returns. In a linear setting, this corresponds to studying the vector of coefficients b_i' in the panel regression:

$$R_{i,t+1} = a_i + b_i' X_t + \varepsilon_{i,t+1}, \tag{A21}$$

where one can replace $R_{i,t+1}$ by $F_{j,t+1}$ if focusing on factors. There are multiple ways to aggregate the information in the estimated coefficients of interest, b_i, to judge the success of X_t as a predictor.

One can ask if X_t predicts "something": is there a linear combination of the coefficients $b=[b_1 \cdots b_n]$ that is statistically distinct from zero? This corresponds exactly to a standard Wald test. This notion of predictability, is intuitively too lax. For instance, our conclusion about the predictive value of X_t could be driven by its ability to predict only a few assets or the lowest variance PC portfolios. A small amount of noise in measured returns can lead to significant spurious predictability of the smallest PC portfolios, even in population. This issue is exacerbated in small samples.

The other extreme is to ask whether elements of X_t predicts "everything", or that all coefficients in a row of b are statistically distinct from zero. For instance Cochrane and Piazzesi (2005) obtain such a pattern predicting Treasury bond returns of various maturities using the cross-section of yields, concluding to the presence of a single common factor in expected returns. While this approach can uncover interesting patterns, it is likely to be too stringent. We show in Section III.B that such a test is often equivalent to testing whether X_t predicts the first principal component of realized returns. In other words, finding uniform predictability across all assets simply finds predictability of the "level" factor in returns. In contrast, we show in Section III.C that if a predictor is useful for forecasting index-neutral factor returns, captured by a long-short portfolio, but not for aggregate returns, individual asset predictive regressions are unlikely to uncover such predictability.

Our approach strikes a balance between these two extremes by asking whether X_t predicts the largest principal components of returns. In other words, we focus on common predictability along the few dimensions explaining a large fraction of realized returns. Focusing on components with a large explanatory power avoids the issue of the Wald test. Entertaining multiple dimensions avoids the other extreme of only focusing on the first component of returns, and allows us to study time-series predictability of cross-sectional strategies.

Our approach to the predictability of cross-sections of returns is focused on predicting important dimensions of the data rather than considering regressions at the individual asset level. In this section, we study more systematically the relation between predicting important components of returns and predicting individual returns.

We consider three features that were relevant in our empirical applications and provide ways to quantify them more generally. First, there is a strong link between predicting the first principal component of returns and predicting each individual return. Second, it is difficult to detect predictability of the second or higher components of returns in individual regressions when the first component is large. Third, joint tests of significance in individual regressions are susceptible to picking up small unimportant patterns of predictability.

III.B First Principal Component and Individual Regressions

A common empirical situation is that a family of returns $\{R_{i,t+1}\}_{i \in I}$ has a strong common component F_{t+1}. When this component is predictable by a variable X_t, does this imply that the individual returns are predictable by X_t? We answer this question quantitatively by deriving a series of bounds linking the predictability of F_{t+1} with the individual predictability of asset returns. We first zoom in on one particular return before considering properties for an entire family of returns.

One individual return: a purely statistical bound. Define $\mathrm{R}^2_{1,i}$ as the population R-squared of the contemporaneous regression of an individual asset on the common component,

$$R_{i,t+1} = \lambda_i F_{t+1} + \varepsilon_{i,t+1}, \tag{A22}$$

and R^2_X as the R-squared of the predictive regression of the factor,

$$F_{t+1} = \beta_1 X_t + u_{t+1}. \tag{A23}$$

We are interested in $\mathrm{R}^2_{X,i}$, the R-squared of the predictive regression

$$R_{i,t+1} = b_i X_t + v_{t+1}. \tag{A24}$$

The following proposition characterizes a lower bound on this quantity.[35]

Proposition 2. If a variable X_t predicts a factor F_{t+1} with R-squared R^2_X and an individual return is explained by this factor with R-squared $\mathrm{R}^2_{1,i}$, then a lower bound for the R-squared $\mathrm{R}^2_{X,i}$ of predicting this return using X_t is given by:

$$\mathrm{R}^2_{X,i} \geq \max\left(\sqrt{\mathrm{R}^2_{1,i}\mathrm{R}^2_X} - \sqrt{(1-\mathrm{R}^2_{1,i})(1-\mathrm{R}^2_X)}, 0\right)^2. \tag{A25}$$

Proof. By the definition of a regression R^2 we have $\mathrm{R}^2_{1,i} = \frac{\lambda_i^2 \mathrm{var}(F_{t+1})}{\mathrm{var}(R_{i,t+1})}$, $\mathrm{R}^2_X = \frac{\beta_1^2}{\mathrm{var}(F_{t+1})}$, and $\mathrm{R}^2_{X,i} = \frac{b_i^2}{\mathrm{var}(R_{i,t+1})}$. The the linearity of regression we have $b_i = \lambda_i\beta_1 + \mathrm{cov}(X_t, u_{i,t+1})$. We can bound the second term in this expression:

$$\begin{aligned}|\mathrm{cov}(X_t,u_{i,t+1})| &= |\mathrm{corr}(X_t,u_{i,t+1})|\sqrt{\mathrm{var}(u_{i,t+1})}\\ &\leq \sqrt{1-\mathrm{R}^2_{1,i}}\sqrt{\mathrm{var}(u_{i,t+1})},\end{aligned}$$

where the bound comes from the fact that the correlation matrix of $u_{i,t+1}$, F_{t+1} and X_{t+1} has to be semidefinite positive and therefore have a positive determinant.

If $|\lambda_i\beta_i| \leq \sqrt{1-\mathrm{R}^2_{1,i}}\sqrt{\mathrm{var}(u_{i,t+1})}$, then 0 is a lower bound for $\mathrm{R}^2_{X,i}$. In the other case, we obtain the following bound:

$$\begin{aligned}\mathrm{R}^2_{X,i} &\geq \frac{\left(\lambda_i\beta_1 - \sqrt{1-\mathrm{R}^2_{1,i}}\mathrm{var}(u_{i,t+1})\right)^2}{\mathrm{var}(R_{i,t+1})}\\ &\geq \left(\sqrt{\frac{\lambda_i^2\beta_1^2}{\mathrm{var}(R_{i,t+1})}} - \sqrt{1-\mathrm{R}^2_{1,i}}\sqrt{\frac{\mathrm{var}(u_{i,t+1})}{\mathrm{var}(R_{i,t+1})}}\right)^2\\ &\geq \left(\sqrt{\mathrm{R}^2_{1,i}\mathrm{R}^2_X} - \sqrt{(1-\mathrm{R}^2_{1,i})(1-\mathrm{R}^2_X)}\right)^2\end{aligned}$$

Putting the two cases together gives Equation (A25). ■

---

[35] Without loss of generality, we assume that the predictor X_t has unit variance.

Intuitively, if X_t strongly predicts the common factor, and the factor has high explanatory power for individual returns, then X_t should predict the individual returns as well. The bound is indeed increasing in the R-squared of these two steps. However, it is lower than the product of the two R-squared — a naive guess that assumes "transitivity" of predictability. This is because the predictor X_t might also predict the residual $\varepsilon_{i,t+1}$ in a way that offsets the predictability coming from the factor. The orthogonality of F_{t+1} and $\varepsilon_{i,t+1}$ limits this force, but does not eliminate it.

To get a quantitative sense of the tightness of this bound, consider the case of bond returns. The level factor explains about 90% of the variation in individual returns, and it can be predicted with an R-squared around 25%. Plugging into our bound, this implies a predictive R-squared of at least 4% for a typical individual bond return. This is a sizable number, but also much less than the 22.5% implied by a naive approach.

One individual return: a bound with an economic restriction. One reason this bound is relatively lax is that it does not take into account the nature of the variable $\varepsilon_{i,t+1}$. Indeed, if, as is the case in our setting, the component F_{t+1} is itself an excess return, the residual $\varepsilon_{i,t+1}$ is one too. It is therefore natural to make the economic assumption that it cannot be "too"' predictable by the variable X_t. This corresponds to imposing an upper bound $\mathrm{R}^2_{\max}$ on the R-squared of the predictive regression of $\varepsilon_{i,t+1}$ by X_{t+1}.[36] In this case, our bound becomes:

$$\mathrm{R}^2_{X,i} \geq \max\left(\sqrt{\mathrm{R}^2_{1,i}\mathrm{R}^2_X} - \sqrt{\mathrm{R}^2_{\max}(1-\mathrm{R}^2_X)}, 0\right)^2. \tag{A26}$$

Such an approach can considerably tighten the bound. For instance, in our example for treasuries, one could impose an upper bound of 25% for predicting the residual. This yields a lower bound on predicting the return $R_{i,t+1}$ of 10%, a much larger number, statistically and economically.

Family of returns: the symmetric case. Another reason that predictability of the common factor must transmit to predictability of individual returns is that by design it absorbs common variation across all those returns. To highlight this point, we consider the following simple symmetric case. We assume that the factor is the average of all the individual returns, $F_{t+1} = \frac{1}{N}\sum_i R_{i,t+1}$. We further assume that all assets have the same loading on the factor and the factor has the same explanatory power for each return. This corresponds to constant λ_i, and R^2_{1i} across assets. We then immediately have:

$$\sum_i u_{i,t+1} = 0$$

$$\sum_i \mathrm{cov}(X_t, u_{i,t+1}) = 0.$$

[36] One way to determine a reasonable bound on $\mathrm{R}^2_{\max}$ is to note that the standard deviation of an asset's conditional Sharpe ratio equals $\sqrt{\frac{\mathrm{R}^2_{\mathrm{x,i}}}{1-\mathrm{R}^2_{x,i}}}$.

Letting $\gamma_i = \text{cov}(X_t, u_{i,t+1})$ we then obtain an expression for an individual asset:

$$\begin{aligned} \mathrm{R}^2_{X,i} &= \frac{(\lambda_i \beta_1 + \gamma_i)^2}{\text{var}(R_{i,t+1})} \\ &= \mathrm{R}^2_1 \mathrm{R}^2_X + \frac{\gamma_i^2}{\text{var}(R_{i,t+1})} + 2\gamma_i \frac{\lambda_i \beta_1}{\text{var}(R_{i,t+1})} \end{aligned}$$

Finally, taking averages across assets we have:

$$\mathbb{E}_i \left[\mathrm{R}^2_{X,i}\right] = \mathrm{R}^2_1 \mathrm{R}^2_X + \text{var}_i \left(\mathrm{R}^2_{X,i}\right), \tag{A27}$$

where $\mathbb{E}_i(\cdot)$ and $\text{var}_i(\cdot)$ are the mean and variance in the cross section of individual returns and we use the fact that we use the fact that $\mathbb{E}_i[\gamma_i] = 0$. This formula implies that the average explanatory power is now at least as large as given by the transitive formula. This would correspond to 22.5% in our example, almost the same value as the predictive R-squared for the common factor. Furthermore, the more unequal this predictive power is across assets, the stronger it must be on average. That is, if the variable X_t does less well than the transitive R-squared for some particular returns, it must compensate more than one-to-one for the other assets.

From predicting "everything" to aggregate returns. Maintaining the same assumptions, we can rearrange Equation (A27) to see what the predictability of "everything" implies for predictability of the common factor. We have:

$$\mathrm{R}^2_X = \frac{\mathbb{E}_i \left[\mathrm{R}^2_{X,i}\right] - \text{var}_i \left(\mathrm{R}^2_{X,i}\right)}{\mathrm{R}^2_1}.$$

At first this may not seem very powerful since $\text{var}_i \left(\mathrm{R}^2_{X,i}\right)$ could be large. This maximal variance, however, is related to the average $\mathbb{E}_i \left[\mathrm{R}^2_{X,i}\right]$. Consider the simple example of only two assets. Then, if the average $\mathbb{E}_i \left[\mathrm{R}^2_{X,i}\right]$ is 10%, the maximal variance is only 1%, which obtains when $\mathrm{R}^2_{X,1} = 0\%$ and $\mathrm{R}^2_{X,2} = 20\%$. In general with two assets we have

$$\text{var}_i \left(\mathrm{R}^2_{X,i}\right) \le \left(0.5 - \left|\mathbb{E}_i \left[\mathrm{R}^2_{X,i}\right] - 0.5\right|\right)^2$$

which gives the bound

$$\mathrm{R}^2_X \ge \frac{\mathbb{E}_i \left[\mathrm{R}^2_{X,i}\right] - \left(0.5 - \left|\mathbb{E}_i \left[\mathrm{R}^2_{X,i}\right] - 0.5\right|\right)^2}{\mathrm{R}^2_1}.$$

For large N, the Bhatia-Davis inequality gives:

$$\mathrm{R}^2_X \ge \frac{(1 - \mathrm{R}^2_{\max}) \mathbb{E}_i \left[\mathrm{R}^2_{X,i}\right] + \mathbb{E}_i \left[\mathrm{R}^2_{X,i}\right]^2}{\mathrm{R}^2_1},$$

where $\mathrm{R}^2_{\max}$, as before, is the maximum $\mathrm{R}^2_{X,i}$ from any individual asset forecasting regression. For reasonable values of $\mathrm{R}^2_{\max}$, such as 0.5 or less, the bound implies that ˜22% average R^2 we obtain for individual bonds implies at least 18% R^2_X, the R-squared when predicting the aggregate portfolio return.

III.C Low Power of Individual Tests

While individual regressions are strongly related to predicting the first common component of returns, they can face challenges in detecting predictability of other factors. We provide a way to quantify this issue by characterizing the statistical power of a test of significance for a predictor that only predicts one particular component of returns. We illustrate this idea in the simple case of an i.i.d. predictor. Simulations confirm these ideas extend to a situation with persistent predictors.

I.i.d. predictor. Consider first the case where the forecasting variable X_{t+1} has i.i.d. draws.[37] Suppose that X_t forecasts only one particular principal component j with population R-squared R_X^2 and the remaining principal component returns are i.i.d. Gaussian with known mean.[38] For power analysis, we consider repeated samples of length T.[39]

When directly forecasting the principal component return, $F_{j,t+1}$, the power to correctly reject the null with test of nominal size α is

$$\text{power}(F_2) = G(-t_{\alpha/2,T} - z) + [1 - G(t_{\alpha/2,T} - z)], \tag{A28}$$

where G is the CDF of a t-distribution with T degrees of freedom, $z = \sqrt{R_X^2}\sqrt{T}(1-R_X^2)^{-\frac{1}{2}}$, and $t_{\alpha/2,T}$ is the $\frac{\alpha}{2}$ critical value from the t-distribution.

In contrast, when directly forecasting an individual return, $R_{i,t+1}$, the power is

$$\text{power}(R_i) = G(-t_{\alpha/2,T} - \zeta) + (1 - G(t_{\alpha/2,T} - \zeta)), \tag{A29}$$

where $\zeta = \sqrt{R_X^2}\sqrt{T}\left((1-R_X^2) + \frac{1-R_{j,i}^2}{R_{ji}^2}\right)^{-\frac{1}{2}}$. By symmetry of the t-distribution and because $\zeta \le z$, we immediately obtain that power(F_2) is larger than power(R_i) for all assets. Therefore, there is always more information about predictability of the important component by studying it directly.

[37] The formulas hereafter admit simple generalizations to multivariate prediction.

[38] More generally, the components need not be principal components. They must be uncorrelated and only one particular component must be forecastable by our predictor. If the mean is unknown, the results below are unchanged except that the degrees of freedom are $T-1$ instead of T.

[39] The analysis treats X as stochastic. With fixed X the distribution is normal instead of a Student t.