An Efficient Factor from Basis Anomalies*

Bingzhi Zhao†

October 1, 2016

Click for most recent version‡

Abstract

A look-ahead-bias-free, ex-ante efficient portfolio from Size, B/M and Momentum anomalies has an ex-post Sharpe ratio of 2.3. It picks up the non-monotonic benefits from characteristics that cannot be captured by the multi-factors and eliminates 39 out of 42 unique anomalies. Using tests of cross-sectional regressions, mean-variance efficiency, miss-specification, model comparison and spurious factors, the 1-factor significantly out-perform the *combined (or separate)* 11 factors: MKT-Rf, SMB, HML, MOM, RMW, CMA, qME, qIA, qROE, QMJ, LIQ among combinations of 147 test assets. The efficient factor is priced at the firm-level with 12% per year spread. Optimal mix of new exotic characteristics can be engineered to pass existing testing tools as "unique anomalies", yet are completely manifested by the efficient factor.

A theory where assets are priced *recursively* w.r.t. the group-specific efficient factor shows that "anomalous" predictabilities are *equivalent* to 1-factor pricing, regardless of rational/behavioral cause. An implied Stochastic Discount Factor return deduced from the efficient factor is consistent with economic theory.

Keywords: Anomaly, efficient portfolio, beta-pricing
JEL Classification: G11, G12, G14

*I am gratefully indebted to my advisors Tim Bollerslev, George Tauchen, Andrew Patton, Jia Li and Federico Bugni, all errors are mine. I thank the seminar participants at the 2016 Triangle Financial Volatility Conference, Duke Financial Econometric Lunch Group. I give special thanks to Svetlana Bryzgalova for her critical comments and suggestions. I also thank Michael Brandt, Anna Cieslak, Dacheng Xiu, Federico Bandi, Brian Weller for helpful comments. The ETF or Inverse ETF construction for the efficient portfolio is patent pending. A Excel sanity check and reproducible package is available on my website: www.sites.duke.edu/bingzhizhao

†Department of Economics, Duke University; bingzhi.zhao@duke.edu.

‡http://ssrn.com/abstract=2841496

1 Introduction

This study:

This study attempts to find a ex-ante efficient return that potentially gets *closer* to ex-post efficiency, yet turns out to: price assets, drive out a massive number of other anomalies, can be used to engineer "exotic" anomalies and test against new anomalies. It reveals that Size, B/M and Momentum are 3 "basis" anomalies that spans many other acclaimed anomalies.

I use the 8 "corner" portfolios of 25 Size-B/M and 25 Size-Momentum sorts to construct a simple mean-variance efficient excess return portfolio, with no look-ahead-bias, and test the asset pricing performance of this portfolio. The results are:

(1) The portfolio is highly efficient ex-post, with a Sharpe Ratio of 2.3. It allocates time-stable weights to the base portfolios that exploits the non-monotonic benefits of characteristics, which cannot be captured by multi-factor model, resulting in a return performance that is virtually unaffected by market fluctuations through-out time.

(2) Out of the 42 acclaimed anomalies that I can find and claims to survive common multi-factor models, 39 of them cannot survive the 1-factor model, the rest 3 are high turnover or illiquid anomalies that may diminish after implementation costs.

(3) While testing various combinations of a total of 147 portfolios (with 3 other anomalous portfolio sorts) using cross-sectional regressions, mean-variance efficiency tests, performance comparison tests, miss-specification tests and spurious factor tests, the 1-factor significantly out-perform 11 acclaimed factors, separately or combined.

(4) A closer look at firm-level sorts yields that assets grouped by expected beta to the efficient portfolio has 12%/year significant alpha w.r.t the multi-factor models, but not w.r.t the 1-factor model.

(5) By constructing proxies for the expected efficient factor beta, one could engineer an optimal mix of new exotic characteristics that will significantly predict future returns, and pass all current testing tools as "unique anomalies", while being completely manifested by the efficient factor.

The efficient factor is designed to solve a portfolio optimization problem in order to break the ex-ante/ex-post barrier, yet turned out to have robust pricing performance even though it is not created for that particular purpose. To understand the theoretical implications of the efficient factor, I develop an recursive asset pricing model that points a direction to reconcile debates around the "anomalies". It has 2 main conclusions and 2 useful applications:

(a) Predictabilities are equivalent to 1-factor pricing. Whenever returns are predictable from a known characteristics, the predictive slope can be backed out as an excess return. Pricing power of multiple slopes are completely captured by one efficient portfolio, hence the common pricing factor. Conversely, whenever a pricing factor portfolio exists, the weights assigned to each asset by the portfolio are the predictive characteristics. The questions whether a particular predictability is consistent with a rational or behavioral investment decision making process, or, whether the return predictabilities from a particular characteristics are stable over time, are completely irrelevant to beta pricing.

(b) To price a given set of assets, one is not required to obtain good proxy to an "universal" factor, therefore by-passing the Roll (1977) Critique. In the presence of predictability, for each cascading asset classes, there is a recursively defined pricing factor obtainable *ex-ante* within each asset class, and will price all assets of that class relative to the class specific group average return.

The empirical and theoretical results of this paper can be useful for other research:

(i) The 1-factor model can help to filter out "anomalies" that are in fact manifested by a set of basis predictive characteristics. The discovery of new anomalies thus should be disciplined by whether the "new" reliably improve the efficiency of the "old", subject to implementation constraint. The hundreds of unique anomalies that emerged over the decades are possibly the results of in-adequate asset pricing models, rather than market in-efficiency.

(ii) The efficient factor gives rise to an implied Stochastic Discount Factor return via the Hansen and Richard (1987) decomposition. The conditional expectation of the Stochastic Discount Factor return can be deduced to inversely relate to the conditional risk-premium of the efficient excess return. Using cross-sectional data, estimates the conditional risk-premium of the efficient factor is consistent with economic theory that the Stochastic Discount Factor should be high during recessions and low during expansions. In comparison, the opposite is observed for the conditional risk-premium of the market, suggesting that market excess return is not the efficient excess return.

Related Studies

Empirically, this study relies on extracting efficient portfolios from the "anomalous" portfolios of Fama and French (1993, 2014); Carhart (1997). The discoveries in Asness, Frazzini, and Pedersen (2014); Frazzini and Pedersen (2014); Asness and Frazzini (2013) hints that some ways to combine known firm-characteristics seem to have better return predictabilities over-time than other ways; together with the "predicted beta" framework developed by Pástor and Stambaugh (2003) jointly inspired the "engineered anomalies" section. In a recent study, Malamud and Vilkov (2015) also extracts an ex-ante efficient portfolio (but from the multi-factor portfolios instead of the "anomalous" portfolios), and argue that the "next-period" beta significantly predicts future return and is generated by investor's non-myopic behavior. Brandt, Santa-Clara, and Valkanov (2009) implements portfolio optimization at stock level exploring size, B/M and momentum characteristics and found significant improvement over the market portfolio. Ahn, Conrad, and Dittmar (2009) explore basis assets using cluster analysis on correlations to generate return spreads. The recursive feature of asset pricing is related the the recent evidence in Gandhi and Lustig (2015) where a bank industry specific size factor governs bank specific stock returns and is distinct from the market wide size factor, suggesting a class-specific factor structure exists. Asness, Frazzini, Israel, Moskowitz, and Pedersen (2015) documents that size premium is weak in univariate sort, but strong when controlled for quality measures, confirming that non-monotonicity of benefits is pervasive among anomalies.

This study helps understand the massive collection of anomalies that have been discovered over time, and there is only a few anomalies that sufficiently span the space. Other anomalies existence is simple due to the lack of a better control. For example, Novy-Marx (2014) demonstrated in a sarcastic way how one could predict asset returns using weather, stars and sunspots.

Theoretically, the cross-sectional and time-series dependence structure is adapted from the thorough work of Gagliardini, Ossola, and Scaillet (2016) which also leads to easy estimation of time-varying risk-premium of the efficient factor. The recursive interpretation of asset pricing is closely related to the equivalence between efficiency and linear pricing established early on by Roll (1977). The efficient factor is one of the three orthogonal components in the Hansen and Richard (1987) decomposition, and I follow Cochrane (2009)'s succinct summary to derive the Implied Stochastic Discount Factor returns.

I structure the rest of the paper by first comprehensively present the empirical evidence, and then derive the full theoretical model. A 1-click reproducible package is available upon request.

2 Evidence

In this section, I (1) construct an efficient portfolio, then (2) test its asset-pricing performance using an armory of tools, and further (3) investigate the firm-level portfolio sorts, then (4) engineer an "anomaly". The appendix describes the details of the data used.

2.1 The Efficient Portfolio

2.1.1 Base Assets

To construct the efficient excess return portfolio m, I set the base excess returns, $\boldsymbol{x}$ to the portfolios exploiting 3 "anomalies" (Size, Book-to-Market and Momentum). I use the 4 corner portfolios of each of the 5×5 Size—Book-to-Market and 5×5 Size—Momentum sorts from Kenneth French's website, resulting in 8 portfolios of excess returns in total that utilize the Size, Book-to-Market and Momentum "anomalies".

2.1.2 Efficient Allocation

Given the monthly excess return of 8 "corner" portfolios $\boldsymbol{x}$, I construct the mean-variance efficient portfolios using weights proportional to the following:

$$\widehat{\boldsymbol{w}}_t^* = \left(\widehat{\boldsymbol{\Sigma}}_t^{HF}\right)^{-1} \widehat{\boldsymbol{\mu}}_t \tag{1}$$

$\widehat{\boldsymbol{\mu}}_t$ is the historical mean excess return and $\widehat{\boldsymbol{\Sigma}}_t^{HF}$ is historical covariance of anomaly portfolios, given information up to time t. To utilize higher frequency data in order to ensure precision, the covariance $\widehat{\boldsymbol{\Sigma}}_t^{HF}$ is estimated using a method in the same spirit of Frazzini and Pedersen (2014) that correlations and variances are estimated separately

$$\widehat{\boldsymbol{\Sigma}}_t^{HF} = \left\{\widehat{\sigma}_t^{ij}\right\} = \left\{\widehat{\rho}_t^{ij}\widehat{\sigma}_t^{i}\widehat{\sigma}_t^{j}\right\} \tag{2}$$

The variances are estimated via realized variance estimator (Andersen and Bollerslev, 1998) from daily returns using 1 year rolling window, and correlations are estimated from overlapping 3-day returns using 5 year rolling window, to account for asynchronous trading (Epps, 1979; Scholes and Williams, 1977).

The efficient portfolio is unique up to a multiplicative constant (details in later theoretical section), the weights do not necessarily sum up to 1, I normalize the weights $\widehat{\boldsymbol{w}}_t^*$ of the zero-cost portfolios by its L^1 norm (sum of absolute values):

$$\widehat{\boldsymbol{w}}_t^n = \frac{\widehat{\boldsymbol{w}}_t^*}{\|\widehat{\boldsymbol{w}}_t^*\|} \tag{3}$$

Table 1 summarizes the weights allocated to the 8 portfolios and their time-series properties. The efficient allocations are all highly statistically significant over time, even though the covariance matrix is calculated in rolling window. The efficient portfolio takes heavy long-short positions amongst the small size firms compared with large size firms, even though there is no significant long-short position across the size dimension.

Table 1: **Efficient Portfolio Allocation**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sort1	SZ1		SZ5		SZ1		SZ5	
Sort2	BM1	BM5	BM1	BM5	MOM1	MOM5	MOM1	MOM5
Avg.	-0.206	0.221	0.051	-0.004	-0.165	0.255	0.004	-0.054
Std.	(0.08)	(0.11)	(0.05)	(0.02)	(0.11)	(0.14)	(0.03)	(0.03)
T-stat.	[-55.0]	[44.8]	[21.8]	[-3.6]	[-35.4]	[42.5]	[2.7]	[-37.1]

This table reports the summary of efficient allocations assigned to each of the 8 portfolios. The "Sort1" indicates the double-sorted portfolio's first level sort, and "Sort2" indicates the double-sorted portfolios' second level sort. "Avg" reports the time-series sample average of the percentage weights allocated by the m^n factor. "Std." reports the sample standard deviation of the weights over time; "T-stat" reports the T-test for the hypothesis that sample average weight is zero. Sample period: 1967.06.30—2014.12.31, with the first 5 years used as starting sample;

2.1.3 Portfolio Performance

The efficient portfolio is constructed by investing in the zero-cost portfolios $\boldsymbol{x}_{t+1}$ using time t available weights:

$$m_{t+1}^n = \boldsymbol{x}'_{t+1}\widehat{\boldsymbol{w}}_t^n \tag{4}$$

The portfolio is rebalanced monthly, with weights re-estimated monthly. Table 2 presents the summary statistics and correlations of different factors. The efficient excess return portfolio m^n and the risk-parity scaled[1] efficient excess return portfolio m^p both exhibit high Sharpe Ratio. Modest correlations are reported between the efficient portfolio and other acclaimed factors, except that it seems to be uncorrelated with Mkt-Rf, SMB, QMJ and LIQ factors. With Sharpe ratios of 2.27 and 2.55, the return performance of efficient portfolios m^n and m^p is remarkably stable compared to the market portfolio whose Sharpe ratio is 0.38. Since the efficient portfolios m^n and m^p have no look-ahead bias by construction, they could be attractive investment strategies. Figure 1 plots the cumulative excess returns of m^n and m^p compared with the market excess return in log-scale. The robust performance of the efficient portfolio is not sensitive to either: (1) choice of base assets, (2) estimation window, (3) data frequency, (4) scaling. The appendix explore a total of 40 different variants[2] as robustness check as well as a basic sanity check, illustrated in Excel, to achieve a Sharpe ratio of 1.36 via an extremely simple 2-asset long-short portfolio.

[1]Scaled to match expected volatility of the market excess return, with no-forward-bias. See the appendix for detailed construction.

[2](2 variants: normalized weights, risk-parity weights)(2 variants: high frequency daily returns, low frequency monthly return) (2 variants: expanding window, rolling window) (5 variants: 8 portfolios, 16 portfolios, 20 portfolios variant a, 20 portfolios variant b, 24 portfolios) (2) (3) (4) . In the appendix, I present the portfolio performance exploring all 40 ($5 \times 2 \times 2 \times 2$) variants.

Table 2: **Summary Statistics**

Factors	(1) m^n	(2) m^p	(3) Mkt-Rf	(4) SMB	(5) HML	(6) RMW	(7) CMA	(8) MOM	(9) qME	(10) qIA	(11) qROE	(12) QMJ	(13) LIQ
#Obs.	511	511	571	571	571	571	571	571	571	571	571	571	564
Avg. Ret.	7.34	54.40	6.01	2.78	4.39	3.12	4.34	8.09	3.63	5.35	6.57	4.58	5.19
Std.	11.23	74.03	54.78	37.23	35.33	26.25	24.41	51.93	37.25	22.57	30.51	29.46	42.20
t-Stat	14.78	16.61	2.62	1.79	2.97	2.84	4.24	3.72	2.33	5.67	5.14	3.72	2.92
Sharpe	2.27	2.55	0.38	0.26	0.43	0.41	0.62	0.54	0.34	0.82	0.75	0.54	0.43
AutoCorr_1	19.0	28.6	7.8	5.5	16.0	18.2	13.9	7.0	3.8	9.4	9.8	17.7	8.0
AutoCorr_2	20.4	25.3	-3.2	2.6	4.5	3.5	2.3	-7.4	2.7	0.2	-9.8	-1.8	-8.4
Correlations													
$\rho(m^n, \cdot)$	100.0	89.6	8.6	7.2	29.3	17.8	27.5	38.5	10.9	30.2	25.7	10.3	1.3
$\rho(m^p, \cdot)$		100.0	14.0	6.8	21.5	11.7	17.7	32.5	9.2	22.3	21.2	4.0	-0.4

Observations are at monthly frequency. Sample period: 1967.06.30—2014.12.31. First 5 years are used as starting sample to calculate the efficient factors;

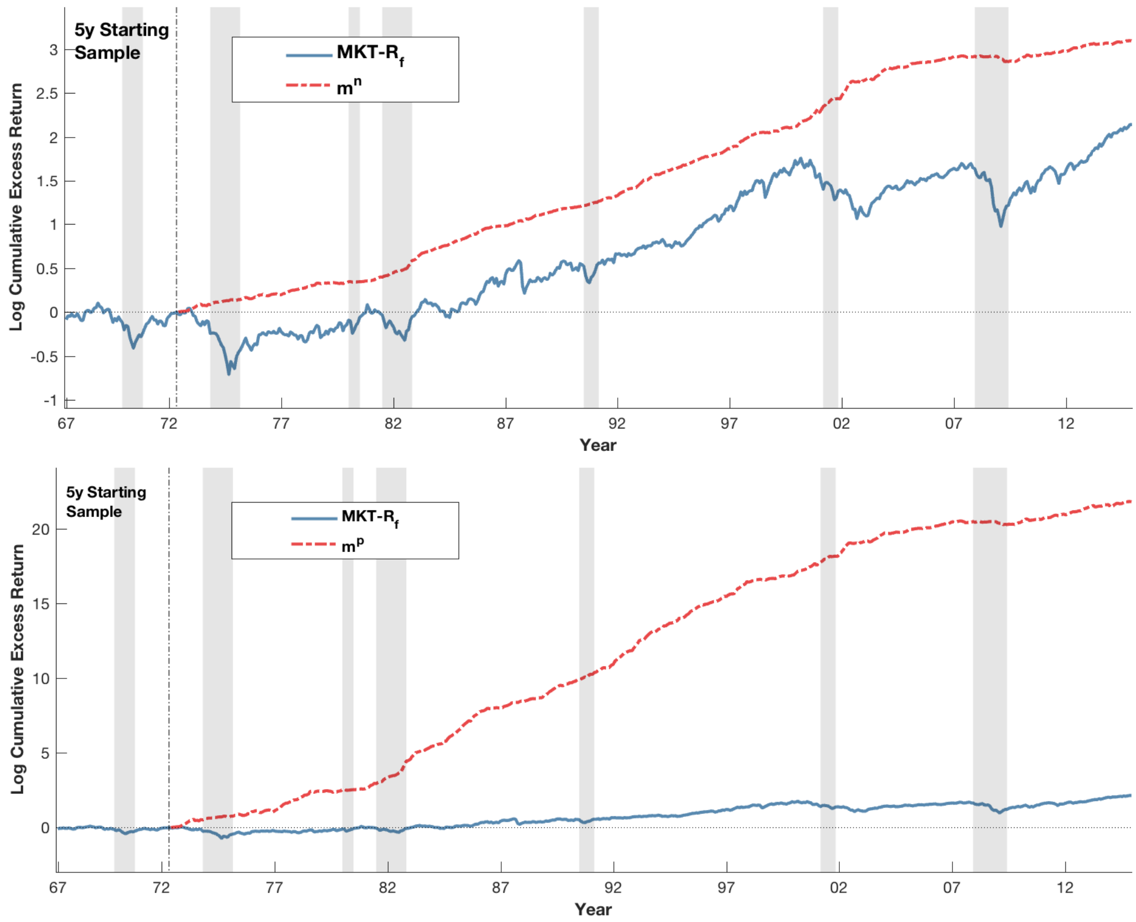
2.1.4 The Reversal of Benefit from "Cheap & Quality"

Leading industry practitioners as well as academic literature suggest that high value and momentum characteristics predict higher future return. For example, long-horizon investors like GMO's Jeremy Grantham advocates "high return, stable return, and low debt" (high momentum and low leverage) as measures of quality; while value characteristics with price in the denominator (Book-to-Market, Earnings-to-Price, Dividend-to-Price, Cash flow-to-Price etc.) has been widely treated as signals for cheapness and bargain. The "buy-low-sell-high" doctrine thereby leads to longing for "Cheap and Quality" and shorting the opposite.

In the literature, the large and positive value and momentum premiums are widely documented. Asness, Moskowitz, and Pedersen (2013) finds extensive evidence that "value and momentum" are "everywhere". Novy-Marx (2015) shows that price momentum is driven by the momentum of fundamentals, that signals the quality of firms' earnings. Moreover, prominent multi-factor models like the Fama and French (1993); Carhart (1997) 4-factor model uses the long-short portfolios to price assets and has become a common practice.

However, in this study, the efficient allocation suggests that this "Cheap and Quality" characteristic is not beneficial in all circumstances. If we define "Cheap and Quality" as high B/M and momentum, then it is easy to see that among smaller firms, the efficient allocation goes significantly short on low "Cheap and Quality" and significantly long on high "Cheap & Quality" firms. Yet, this direction of position on "Cheap and Quality" is *reversed* pertaining to larger firms. The efficient portfolio allocates long positions to low B/M and short positions to high Momentum firms, although the magnitude of the weights are about 4 times smaller compared with the weights on smaller firms, but nonetheless highly statistically significant in all sense

Figure 1: The Efficient Portfolios: Compared to Market



(t-statistics amounts to 21.8 and -37.1). This reversal of benefit is even more evident in the time-series plot shown in Figure 2. By combining the weights allocated to the firms that are "Small, Low 'Cheap & Quality' "(SZ1BM1, SZ1MOM1), "Small, High 'Cheap & Quality' "(SZ1BM5,SZ1MOM5), and comparing with the larger firms "Large, Low 'Cheap & Quality' "(SZ5BM1, SZ5BM1), "Large, High 'Cheap & Quality' "(SZ5BM5, SZ5MOM5), it is easy to see that an efficient portfolio goes long on "Cheap & Quality" for small firms but goes short on "Cheap & Quality" for large firms, and it consistently does so throughout the 40+ years of history.

A possible explanation for this subtle yet significant contrast can be that the large firms are easier to trade compared to small firms. A significant spread for the larger firms is like "low hanging fruits" compared with the smaller firms that usually have low liquidity and high trading costs. Yet, the "common sense" demand for "Cheap & Quality" is so high throughout time that the spread for the "low hanging fruits" have been squeezed so much so that it makes sense to do the opposite for the large firms from a mean-variance efficiency optimization standpoint.

This type of "reversal of benefit" from the value and momentum characteristics is an important reason why the efficient portfolio is able to get much closer to ex-post

Figure 2: **Efficient Allocations to 4 Types of Assets**

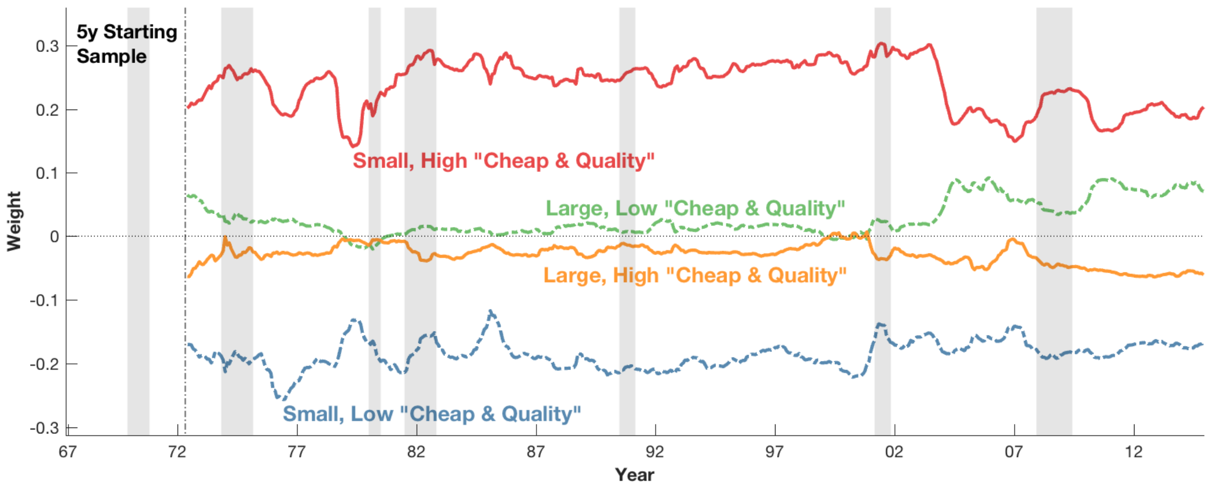


Figure illustrates the time-varying weights assigned to 4 types of assets by the efficient excess return portfolio m^n. "Small, Low 'Cheap & Quality' " represents the average of the efficient weights assigned to the base asset portfolios SZ1BM1 and SZ1MOM1; "Small, High 'Cheap & Quality' " for SZ1BM5 and SZ1MOM5, "Large, Low 'Cheap & Quality' " for SZ5BM1 and SZ5BM1, "Large, High 'Cheap & Quality' " for SZ5BM5 and SZ5MOM5; Sample period: 1967.06.30—2014.12.31, with the first 5 years used as starting sample;

efficiency as well as extract more pricing power than the commonly used multi-factors that are based on univariate or double sorted characteristics. In fact, the Sharpe ratio of the mean-variance efficient portfolio from all the 11 factors amounts to 1.74, not significantly higher than the simple long-short portfolio explored in the sanity check, 1.36.

It is also important to point out that the above is only one illustrative mechanism where efficient portfolio gets higher efficiency than the multi-factors combined, as there are more "non-monotonicity" in the benefits interacting among all available characteristic dimensions.

3 Asset Pricing Tests

After obtaining the ex-ante efficient portfolio m^n, which turns out to be highly efficient ex-post, it is natural to test its asset pricing performance. It has been established by Roll (1977) that mean-variance efficiency is equivalent to linear beta pricing. When a portfolio constructed ex-ante turns out to be ex-post efficient, such portfolio should automatically serve as a pricing factor. To test this theoretical prediction rigorously, I first discuss the results using standard cross-sectional regression methods, and then explore other tests involving efficiency, model comparison, miss-specification and spurious factors.

3.1 The Cross-Sectional Regression Method

Introduced by Fama and MacBeth (1973), the 2-step cross-sectional regression is an intuitive way to test the performance of an unconditional linear pricing model. Let i be the cross-sectional index for a given security, $r_t(i)$ be the holing period return of asset i at time t and r_t^f be the risk-free rate, and $\boldsymbol{f}_t$ be a set of factors at time t, the 2-step procedure first estimate the time-series model for each asset i:

$$r_t^e(i) = r_t(i) - r_t^f = \alpha(i) + \boldsymbol{\beta}(i)'\boldsymbol{f}_t + u_t(i) \tag{5}$$

After obtaining the factor loadings $\boldsymbol{\beta}(i)$ for each asset, perform the second step cross-sectional regression, where the LHS variable is the average returns of each asset, and RHS variables are the intercept and factor loadings $\boldsymbol{\beta}(i)$:

$$\mathbb{E}\left[r_t^e(i)\right] = \gamma_0 + \boldsymbol{\gamma}'\boldsymbol{\beta}(i) + v(i)$$

The robust standard errors of the risk-premia γ's in the second-step regression is estimated with Shanken (1992) errors-in-variables correction, Jagannathan and Wang (1998) hetero-skedasticity adjustment and Gospodinov, Kan, and Robotti (2014) miss-specification robust adjustment.

3.2 Pricing Assets of the Basis Anomalies

In this section, I present the evidence of a comprehensive series of tests to establish the robust pricing performance of the efficient 1-factor model using the "basis anomalies" portfolios, and further show that the 1-factor is able to drive-out a combined 11-factors.

To relieve from the concern that the 8 corner portfolios may be the main driver of the pricing results (even though it is 8 random variables summarized into 1), I purposefully remove the 8 corner portfolios from the test assets, resulting in 42[3] Size, B/M and Momentum portfolios (the 5×5 Size–B/M, 5×5 Size–Momentum less the 8 corners that were used in the efficient portfolio construction).

3.2.1 Risk-Premia, Specification and Mean-Variance Efficiency Tests

Table 3 reports OLS estimation results for the efficient 1 factor model compared with the Sharpe (1964) and Lintner (1965) CAPM 1-factor model, as well as acclaimed empirical multi-factor models commonly used in the literature, including Fama and French (1993) 3-factor (Mkt-Rf, SMB, HML), Carhart (1997) 4-factor (3+MOM), Asness, Frazzini, and Pedersen (2014) 4-factor (3+QMJ), Hou, Xue, and Zhang (2014) 4-factor (Mkt-

[3] These are 42 portfolios on 3 anomalies, not to be confused with the 42 "unique" anomalies that will be priced later.

Rf, qME, qIA, qROE), Pástor and Stambaugh (2003) 4-factor (3+LIQ), Fama and French (2014) (3+RMW, CMA) 5-factor models as well as a combined 11-factor model.

At a glance, the efficient 1-factor model (RAP1, Recursive Asset Pricing model, detailed in the theoretical section) achieves remarkable pricing performance. The efficient factor has a significant t statistics that surpasses all acclaimed risk-factors. The efficient 1-factor model has an R^2 of 82%. Pointed out by Lewellen, Nagel, and Shanken (2010), the cross-sectional R^2 as a random variable can be extremely unstable with variability grow proportional to the number of factors used. A model with 5-factors will need at least 69% R^2 to be even statistically significant with 95% confidence. With this in mind, I also employ the Kan, Robotti, and Shanken (2013) test to assess the statistical significance of the R^2 under different specification assumptions. Reported in column (16), with a high p-value of 0.50, one cannot reject the hypothesis that the 1-factor model is correctly specified, i.e. the true $R^2 = 1$, and one strongly reject the hypothesis that the model is miss-specified, i.e. the true $R^2 = 0$. However, the multi-factor models do not have strong supportive results. Although the R^2's of CAPM1, FF3, AFP4, FF5 have high point estimate, they are statistically indistinguishable from 0 with 95% confidence.

Moreover, as Gibbons, Ross, and Shanken (1989) (GRS) test shows, the hypothesis that all asset's pricing errors $\alpha(i) = 0$ is not rejected for the efficient 1-factor model, but rejected for all the other multi-factor models. Since GRS test is equivalent to testing the factor model's efficient frontier against the test assets, the result implies that the efficient 1-factor is close to the mean-variance efficient frontier spanned by the entire 42 portfolios. This result is expected as previously illustrated that although based on the same or less anomalies, (3 anomalies rather than 10), the efficient 1-factor is able to utilize the non-monotonic benefits that cannot be captured by the multi-factors. Put simply, the combined mean-variance efficient frontier of the 11 multi-factors is not "efficient" enough.

A good asset pricing model should have its performance robust to the choice of weighting matrix. The model's overall performance may not necessarily hold for various weighting matrix, if its pricing performance mainly comes from some particular observation in the sample. For example, some studies include the factor excess returns in the test assets to boost the significance of the factor's risk-premium estimate. Doing so will mechanically cause part of the test assets to be perfectly explained by the factor. It is also an interest to explore different weighting matrix depending on the particular economic application. Table 4 reports the results under GLS where the weighting matrix $\boldsymbol{W} = \widehat{\boldsymbol{\Sigma}}_R^{-1}$, the sample covariance of the returns following Kan, Robotti, and Shanken (2013), assuming potential miss-specification. Notice that under this weighting matrix, the performance of all models significantly drop to similar magnitudes reported by Kan, Robotti, and Shanken (2013) Table I. However, the R^2 of the 1-factor model is at statistically significant 40%, yet the combined 11-factor model has an

Table 3: **Pricing 42 Size, B/M, Momentum Portfolios, OLS**

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)
	γ_0	γ_{MKT}	γ_{SMB}	γ_{HML}	γ_{RMW}	γ_{CMA}	γ_{MOM}	γ_{qME}	γ_{qIA}	γ_{qROE}	γ_{QMJ}	γ_{LIQ}	γ_{m^n}	$R^2(\%)$	$p[R^2=1]$	$p[R^2=0]$	$p[\alpha(i)=0]$
RAP1	5.96												4.21	81.6	0.50	0.00	0.07
	(1.46)												(3.18)				
	[1.44]												[2.95]				
CAPM1	15.78	-6.62												12.1	0.00	0.22	0.00
	(3.92)	(-1.53)															
	[3.17]	[-1.23]															
FF3	21.76	-14.24	2.25	2.08										44.6	0.01	0.05	0.01
	(3.72)	(-2.34)	(1.32)	(1.01)													
	[4.81]	[-2.96]	[1.24]	[0.85]													
FFC4	5.09	1.93	2.79	4.34			7.84							84.4	0.07	0.00	0.01
	(1.60)	(0.52)	(1.60)	(2.19)			(2.94)										
	[1.59]	[0.54]	[1.60]	[2.15]			[2.94]										
AFP4	10.31	-3.67	3.12	1.36							6.06			55.8	0.08	0.06	0.01
	(2.23)	(-0.84)	(1.83)	(0.57)							(1.91)						
	[1.16]	[-0.42]	[1.58]	[0.53]							[2.44]						
HXZ4	1.80	3.97						5.23	4.43	6.13				82.1	0.28	0.01	0.02
	(0.50)	(1.08)						(2.55)	(2.24)	(2.22)							
	[0.46]	[0.95]						[2.58]	[2.04]	[2.31]							
FF5	0.13	5.89	3.89	2.24	-0.15	13.19								74.0	0.14	0.07	0.01
	(0.02)	(0.90)	(2.31)	(0.90)	(-0.06)	(2.32)											
	[0.02]	[0.81]	[2.17]	[0.79]	[-0.04]	[2.66]											
PS5	5.13	1.86	2.79	4.31			7.82					0.02		84.4	0.06	0.00	0.01
	(1.60)	(0.50)	(1.60)	(2.18)			(2.92)					(0.33)					
	[1.57]	[0.50]	[1.61]	[2.16]			[2.92]					[0.29]					
ALL11	10.67	-4.64	3.64	4.82	-1.95	6.47	6.85	4.00	4.18	6.97	-2.44	-0.10		88.5	0.17	0.01	0.01
	(1.65)	(-0.72)	(1.79)	(2.16)	(-0.50)	(1.39)	(2.92)	(1.78)	(1.29)	(2.09)	(-0.83)	(-1.07)					
	[1.55]	[-0.63]	[1.97]	[2.34]	[-0.46]	[1.37]	[2.77]	[1.43]	[0.97]	[1.83]	[-0.94]	[-0.94]					
$\mathbb{E}[f]$		6.01	2.78	4.39	3.12	4.34	8.09	3.63	5.35	6.57	4.58	5.19	7.34				

2-step OLS cross-sectional regression results. The second-step uses multiple regression beta as explanatory variables. The excess returns are the 42 portfolios (25SZBM and 25 SZMOM less the 8 corner portfolios used to construct the efficient factor). The first row reports the estimated risk-premium annualized, the second row reports the Jagannathan and Wang (1998) heteroskedasticity and error-in-variable corrected t-statistics in parenthesis, the third row reports the Gospodinov, Kan, and Robotti (2014) miss-specification robust t-statistics. Column (15) report the 2-step cross-sectional R^2; column (16) reports the p-value Kan, Robotti, and Shanken (2013) specification test under the null hypothesis $H_0 : R^2 = 1$ and column (17) for the null hypothesis $H_0 : R^2 = 0$; column (18) reports the p-value of Gibbons, Ross, and Shanken (1989) test for the hypothesis that all pricing errors are zero. The last row of the table reports the annualized mean of the factors. The sample period is 1967.06.30—2014.12.31 with 511 effective months after excluding the first 5-year starting sample.

Table 4: **Pricing 42 Size, B/M, Momentum Portfolios, GLS, Assuming Miss-Specification**

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)
	γ_0	γ_{MKT}	γ_{SMB}	γ_{HML}	γ_{RMW}	γ_{CMA}	γ_{MOM}	γ_{qME}	γ_{qIA}	γ_{qROE}	γ_{QMJ}	γ_{LIQ}	γ_{m^n}	R^2(%)	$p[R^2=1]$	$p[R^2=0]$
RAP1	3.53												5.29	39.4	0.21	0.00
	[1.31]												[3.70]			
CAPM1	10.44	-3.23												0.8	0.00	0.42
	[2.96]	[-0.81]														
FF3	9.48	-2.47	2.40	3.97										7.3	0.05	0.08
	[2.85]	[-0.60]	[1.46]	[1.96]												
FFC4	7.28	-0.27	2.41	4.08			7.21							17.8	0.08	0.00
	[2.10]	[-0.07]	[1.40]	[1.97]			[2.75]									
AFP4	7.64	-0.64	2.42	3.86							0.76			8.7	0.08	0.11
	[1.93]	[-0.15]	[1.34]	[1.81]							[0.40]					
HXZ4	6.44	0.27						3.42	4.05	3.76				15.0	0.07	0.01
	[1.83]	[0.06]						[1.95]	[2.37]	[1.63]						
FF5	7.66	-0.82	2.59	3.82	0.84	3.92								9.6	0.05	0.26
	[2.05]	[-0.20]	[1.43]	[1.80]	[0.32]	[1.19]										
PS5	7.88	-0.87	2.45	4.08			7.28					-0.05		19.0	0.08	0.01
	[2.37]	[-0.22]	[1.45]	[1.98]			[2.75]					[-0.90]				
ALL11	8.69	-1.63	2.44	4.04	-0.33	3.27	7.14	1.69	3.04	2.99	-0.50	-0.06		22.6	0.01	0.31
	[1.89]	[-0.39]	[1.21]	[1.93]	[-0.09]	[0.82]	[3.15]	[0.66]	[0.84]	[0.81]	[-0.19]	[-0.87]				
$\mathbb{E}[f]$		6.01	2.78	4.39	3.12	4.34	8.09	3.63	5.35	6.57	4.58	5.19	7.34			

This table reports the 2-step GLS cross-sectional regression with weighting matrix $\boldsymbol{W} = \widehat{\boldsymbol{\Sigma}}_R^{-1}$ following Kan, Robotti, and Shanken (2013) assuming model mis-specification. The second-step regression uses multiple regression beta's as explanatory variables. The LHS variables are the excess returns of the 42 portfolios (25SZBM and 25 SZMOM less the 8 corner portfolios used to construct the efficient factor). The first row of each model reports the estimated multiple regression beta risk-premium γ's annualized, the second row reports the Gospodinov, Kan, and Robotti (2014) miss-specification robust t-statistics. Column (15) report the 2-step cross-sectional R^2; column (16) reports the p-value of Kan, Robotti, and Shanken (2013) specification test under the null hypothesis $H_0 : R^2 = 1$ and column (17) for the null hypothesis $H_0 : R^2 = 0$; The last row of the table reports the annualized mean of the factors for comparison. The sample period is 1967.06.30—2014.12.31 with 511 effective months after excluding the first 5-year starting sample.

statistically insignificant R^2 of 22.6%. Moreover, the risk-premium estimate associated with the efficient factor remain highly statistical significant after adjusting for potential miss-specification, with a Gospodinov, Kan, and Robotti (2014) miss-specification robust t-statistics of 3.70, well pass the "new two is three" hurdle advocated by Harvey, Liu, and Zhu (2015).

3.2.2 Performance Comparison Tests

Given the robust asset pricing performances of the efficient factor, I conduct a series of statistical tests using the most recent tools and further show that: (1) The 1-factor model significantly out-perform the combined and separate multi-factor models in many ways (2) When the 1-factor model do not completely dominate a particular multi-factor model, it provides significant additional pricing power that virtually drives out the significance of all the 11-factors.

To draw a statistical conclusion on the pricing performances of different models, I employ the tools developed by recent study Kan, Robotti, and Shanken (2013) model comparison test based on R^2 measure, in the context of potential model miss-specification. Table 5 reports the R^2 difference and p-values for the non-nested models by comparing RAP1 directly against the multi-factor models and the combined 11-factor model. When a particular multi-factor model has higher R^2 estimate than the 1-factor model, the difference is small and easily rejected by the test. For example, in the OLS case, FFC4, HXZ4, PS4 and ALL11 has slightly higher point estimate of R^2, but the difference is insignificant at all common confidence levels (p-values of 0.18-0.22). On the other hand, except for FF5 under OLS has a smaller R^2 with statistically insignificant difference, every multi-factor model under GLS is statistically significantly dominated by the 1-factor model. This result is consistent with the previous tables that the multi-factor model has weak or borderline significant R^2 and have higher numbers of factors that increases the variability of the R^2 and are much farther away from the mean-variance efficient frontier than the efficient 1-factor.

Next I show that the efficient 1-factor provide significant incremental pricing power that virtually drives out the significance of all multi-factor models. To show the nested comparison, I add the efficient factor incrementally to each of the multi-factor models. Since the comparison is about competing factor specification, "what matters is whether the prices of *covariance risk* are non zero", as clarified by Kan, Robotti, and Shanken (2013). Therefore, in the second-stage, instead of using the multiple-regression-beta's as explanatory variables, we must use the covariance between the test asset returns and factors (simple beta scaled by factor's variance). The resulting slope in the second stage regression is the covariance risk λ (as oppose to γ, the risk-premium with multiple beta). Table 6 reports both OLS and GLS estimation with covariance risk. Evidently, when the efficient factor is added to each of the multi-factor models, the covariance risks of the multi-factor become statistically insignificant in all specifications, moreover, the appended model gains significant boost in the R^2. In the OLS case, the efficient factor brings all model R^2 to around 90% and for the GLS case, to about 40%. The covariance risk of the efficient factor λ_{m^n} is strongly statistically significant in all specifications with the harshest penalty[4], except in HXZ4 model in OLS case, it is significant only at 90% confidence level. This means that the pricing power of all the multi-factor models are essentially driven-out in the presence of the efficient 1-factor.

Similar conclusion is made if one employ the nested model comparison test, also developed in Kan, Robotti, and Shanken (2013). One advantage of the nested model comparison test is that it accommodates different assumptions on the potential model (miss-)specification, compared with the test of covariance risk. It tests the hypothesis that the multi-factor pricing power are jointly zero. Table 7 reports the nested model comparison with the R^2 difference and p-values under different assumptions about potential miss-specification. Shown in panel A, when the the multi-factors are added to the 1 factor model, the combined model is not statistically better than the original 1-factor alone. However, when the efficient factor is added to the multi-factor specification, the combined factor is significantly better than the original multi-factor models (p-values less than 0.02), with the exception of Hou, Xue, and Zhang (2014) 4-factor model in the OLS case (p-value less than 0.09), while still significant at 90% confidence.

Aside from comparison based on R^2 measure, Kan and Robotti (2009) develops a test for model comparison based on the Hansen and Jagannathan (1997) Distance Measure. As pointed out by Kan, Robotti, and Shanken (2013), the R^2 is a better model comparison if the interest is the expected return due to the zero-beta rate. Appendix includes the test comparison using Kan and Robotti (2009) methodology as another robustness check.

[4]One could more or less rank the penalty for statistical significance from the different estimators of standard errors as: Fama and MacBeth (1973) $\prec$ Shanken (1992) error-in-variable correction, which is, in majority of the cases, less harsh than Jagannathan and Wang (1998) heteroskedasticity robust adjustment and Gospodinov, Kan, and Robotti (2014) miss-specification robust adjustment

Table 5: **Model Comparison, R^2 Difference and p-Value**

"Row" Model - "Column" Model

	OLS, $\boldsymbol{W} = \boldsymbol{I}_N$								
	RAP1	CAPM1	FF3	FFC4	AFP4	HXZ4	FF5	PS5	ALL11
R^2	81.6	12.1	44.6	84.4	55.8	82.1	74.0	84.4	88.5
RAP1	—	69.6	37.0	-2.8	25.8	-0.5	7.6	-2.8	-6.8
		0.00	0.00	0.22	0.02	0.23	0.18	0.19	0.65
CAPM1		—	-32.5	-72.3	-43.7	-70.1	-61.9	-72.4	-76.4
			0.07	0.01	0.08	0.02	0.02	0.00	0.01
FF3			—	-39.8	-11.2	-37.5	-29.4	-39.8	-43.9
				0.02	0.05	0.02	0.01	0.02	0.01
FFC4				—	28.6	2.3	10.4	-0.0	-4.1
					0.10	0.97	0.61	0.44	0.17
AFP4					—	-26.3	-18.2	-28.6	-32.6
						0.11	0.06	0.10	0.07
HXZ4						—	8.1	-2.3	-6.3
							0.57	0.96	0.47
FF5							—	-10.4	-14.5
								0.62	0.40
PS5								—	-4.0
									0.13
	GLS, $\boldsymbol{W} = \widehat{\boldsymbol{\Sigma}}_R^{-1}$								
Model	RAP1	CAPM1	FF3	FFC4	AFP4	HXZ4	FF5	PS5	ALL11
R^2	39.4	0.8	7.3	17.8	8.7	15.0	9.6	19.0	22.6
RAP1	—	38.7	32.1	21.6	30.7	24.4	29.8	20.5	16.8
		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CAPM1		—	-6.6	-17.1	-8.0	-14.3	-8.9	-18.2	-21.9
			0.13	0.01	0.13	0.01	0.14	0.01	0.01
FF3			—	-10.5	-1.4	-7.7	-2.3	-11.6	-15.3
				0.02	0.52	0.02	0.56	0.02	0.04
FFC4				—	9.1	2.8	8.2	-1.1	-4.8
					0.03	0.66	0.03	0.41	0.76
AFP4					—	-6.3	-0.9	-10.2	-13.9
						0.02	0.81	0.03	0.06
HXZ4						—	5.4	-3.9	-7.6
							0.01	0.53	0.48
FF5							—	-9.3	-13.0
								0.03	0.04
PS5								—	-3.7
									0.93

This table reports the OLS and GLS pricing performance comparison test results for the non-nested models, following Kan, Robotti, and Shanken (2013). The first number of each comparison reports the difference of row model R^2 minus the column model R^2, and the second number reports the p-value of the test that the difference is zero. The sample period is 1967.06.30—2014.12.31 with 511 effective months after excluding the first 5-year starting sample.

Table 6: **Incremental Pricing Power, Covariance Risk (Scaled Simple Beta)**

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	2-Pass OLS, Covariance Risk, $\boldsymbol{W} = \boldsymbol{I}_N$														
	λ_0	λ_{MKT}	λ_{SMB}	λ_{HML}	λ_{RMW}	λ_{CMA}	λ_{MOM}	λ_{qME}	λ_{qIA}	λ_{qROE}	λ_{QMJ}	λ_{LIQV}	λ_{m^n}	R^2	$R^2_{w/o}$
RAP1	5.96												4.82	81.6	—
	[1.46]												[2.49]		
CAPM1+1	-1.07	0.29											5.85	88.0	12.1
	[-0.21]	[1.24]											[3.03]		
FF3+1	1.86	0.11	0.18	0.04									5.24	91.2	44.6
	[0.48]	[0.64]	[0.81]	[0.14]									[2.55]		
FFC4+1	1.16	-0.05	0.13	-0.44			-0.34						8.28	92.3	84.4
	[0.28]	[-0.23]	[0.53]	[-1.16]			[-1.37]						[3.51]		
AFP4+1	2.38	0.05	0.13	0.01							-0.15		5.35	91.2	55.8
	[0.51]	[0.13]	[0.36]	[0.02]							[-0.18]		[2.34]		
HXZ4+1	3.06	-0.06						-0.05	-0.56	-0.51			6.84	91.7	82.1
	[0.67]	[-0.19]						[-0.10]	[-0.43]	[-0.65]			[1.72]		
PS4+1	1.74	0.12	0.18	0.06								-0.21	5.24	91.2	
	[0.47]	[0.72]	[0.79]	[0.22]								[-0.36]	[2.61]		
PS5+1	0.72	-0.08	0.11	-0.50			-0.41					-0.51	8.88	92.7	84.4
	[0.18]	[-0.33]	[0.42]	[-1.27]			[-1.72]					[-0.70]	[3.76]		
FF5+1	3.56	-0.13	0.17	0.58	-0.26	-1.43							5.95	91.5	74.0
	[0.66]	[-0.29]	[0.50]	[0.67]	[-0.32]	[-0.63]							[2.39]		
ALL11+1	2.38	-0.15	4.03	0.08	-0.62	-0.63	-0.37	-4.08	-0.75	-0.30	0.65	-0.52	10.86	93.9	88.5
	[0.33]	[-0.21]	[1.12]	[0.08]	[-0.30]	[-0.14]	[-0.63]	[-1.11]	[-0.20]	[-0.09]	[0.25]	[-0.47]	[2.61]		

Continued on next page

Table 6 **Incremental Pricing Power, Covariance Risk (Scaled Simple Beta), Continued**

Continued from previous page

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
	2-Pass GLS, Covariance Risk, $W = \widehat{\Sigma}_R^{-1}$														
	λ_0	λ_{MKT}	λ_{SMB}	λ_{HML}	λ_{RMW}	λ_{CMA}	λ_{MOM}	λ_{qME}	λ_{qIA}	λ_{qROE}	λ_{QMJ}	λ_{LIQV}	λ_{m^n}	R^2	$R^2_{w/o}$
RAP1	3.53												6.05	39.4	—
	[1.31]												[3.18]		
CAPM1+1	2.26	0.11											6.20	39.8	0.8
	[0.61]	[0.66]											[3.20]		
FF3+1	2.95	0.03	0.10	-0.13									6.34	40.4	7.3
	[0.75]	[0.15]	[0.44]	[-0.46]									[3.11]		
FFC4+1	2.44	-0.11	0.05	-0.53			-0.39						8.65	43.5	17.8
	[0.61]	[-0.44]	[0.19]	[-1.35]			[-1.59]						[3.77]		
AFP4+1	3.59	-0.07	0.03	-0.20							-0.25		6.50	40.6	8.7
	[0.89]	[-0.26]	[0.08]	[-0.65]							[-0.43]		[2.93]		
HXZ4+1	3.52	-0.18						-0.21	-1.14	-0.66			8.21	43.2	15.0
	[0.80]	[-0.60]						[-0.54]	[-1.07]	[-1.07]			[2.52]		
PS4+1	3.62	-0.02	0.12	-0.13								-0.53	6.33	42.2	
	[0.98]	[-0.09]	[0.52]	[-0.45]								[-1.09]	[3.14]		
FF5+1	3.67	-0.17	0.09	0.44	-0.24	-1.51							7.08	42.2	9.6
	[0.86]	[-0.51]	[0.30]	[0.61]	[-0.45]	[-0.82]							[2.86]		
ALL11+1	3.02	-0.16	3.63	0.03	-0.32	-2.56	-0.18	-3.69	1.14	-1.33	1.05	-0.65	10.86	51.8	22.6
	[0.53]	[-0.26]	[0.95]	[0.03]	[-0.15]	[-0.70]	[-0.37]	[-0.97]	[0.31]	[-0.49]	[0.45]	[-0.86]	[2.83]		

This table reports results for the 2-step GLS and OLS cross-sectional regression. The second-step uses covariance between returns and factors (scaled simple regression beta) as the explanatory variables following Kan, Robotti, and Shanken (2013) assuming miss-specification. The LHS variables are the excess returns of the 42 portfolios (25SZBM and 25 SZMOM less the 8 corner portfolios used to construct the efficient factor). The first row of each model reports the estimated covariance risk-premium λ's annualized, the second row reports the Gospodinov, Kan, and Robotti (2014) miss-specification robust t-statistics. Column (15) report the 2-step cross-sectional R^2; column (16) reports the original R^2 of the common multi-factor models without adding the efficient factor m^n (same as in Table 3 and 4). The sample period is 1967.06.30—2014.12.31 with 511 effective months after excluding the first 5-year starting sample.

Table 7: **Nesting Model Comparison: R^2 Difference and p-Value**

Panel A: $\boldsymbol{X}+\boldsymbol{m} \approx \boldsymbol{m}$? OLS $\boldsymbol{W}=\boldsymbol{I}_N$									
Statistics	CAPM1	FF3	FFC4	AFP4	HXZ4	PS4	FF5	PS5	ALL11
$\Delta = R^2_{X+m} - R^2_m$	6.4	9.5	10.7	9.6	10.1	9.6	9.9	11.1	12.3
χ^2-test $p[\Delta = 0 \vert R^2 = 1]$	0.24	0.47	0.53	0.53	0.50	0.50	0.56	0.56	0.70
Wald-test $p[\Delta = 0 \vert R^2 = 1]$	0.24	0.57	0.41	0.71	0.67	0.62	0.62	0.23	0.63
χ^2-test $p[\Delta = 0 \vert R^2 < 1]$	0.26	0.47	0.53	0.52	0.52	0.50	0.58	0.58	0.74
Wald-test $p[\Delta = 0 \vert R^2 < 1]$	0.26	0.58	0.45	0.72	0.70	0.67	0.83	0.35	0.76

Panel B: $\boldsymbol{X}+\boldsymbol{m} \approx \boldsymbol{m}$? GLS $\boldsymbol{W}=\widehat{\boldsymbol{\Sigma}}_R^{-1}$									
Statistics	CAPM1	FF3	FFC4	AFP4	HXZ4	PS4	FF5	PS5	ALL11
$\Delta = R^2_{X+m} - R^2_m$	0.4	0.9	4.1	1.2	3.7	2.8	2.7	6.8	12.4
χ^2-test $p[\Delta = 0 \vert R^2 = 1]$	0.40	0.83	0.53	0.90	0.51	0.61	0.80	0.39	0.73
Wald-test $p[\Delta = 0 \vert R^2 = 1]$	0.40	0.73	0.54	0.82	0.30	0.50	0.69	0.28	0.69
χ^2-test $p[\Delta = 0 \vert R^2 < 1]$	0.56	0.84	0.60	0.90	0.68	0.68	0.86	0.48	0.87
Wald-test $p[\Delta = 0 \vert R^2 < 1]$	0.56	0.74	0.64	0.82	0.68	0.52	0.89	0.43	0.89

Panel C: $\boldsymbol{X}+\boldsymbol{m} \approx \boldsymbol{X}$? OLS $\boldsymbol{W}=\boldsymbol{I}_N$									
Statistics	CAPM1	FF3	FFC4	AFP4	HXZ4	PS4	FF5	PS5	ALL11
$\Delta = R^2_{X+m} - R^2_X$	75.9	46.5	7.9	35.4	9.6	46.3	17.5	8.3	5.5
χ^2-test $p[\Delta = 0 \vert R^2 = 1]$	0.00	0.01	0.00	0.02	0.07	0.01	0.00	0.00	0.00
Wald-test $p[\Delta = 0 \vert R^2 = 1]$	0.00	0.01	0.00	0.02	0.07	0.01	0.00	0.00	0.00
χ^2-test $p[\Delta = 0 \vert R^2 < 1]$	0.00	0.01	0.00	0.02	0.09	0.01	0.01	0.00	0.00
Wald-test $p[\Delta = 0 \vert R^2 < 1]$	0.00	0.01	0.00	0.02	0.09	0.01	0.01	0.00	0.00

Panel D: $\boldsymbol{X}+\boldsymbol{m} \approx \boldsymbol{X}$? GLS $\boldsymbol{W}=\widehat{\boldsymbol{\Sigma}}_R^{-1}$									
Statistics	CAPM1	FF3	FFC4	AFP4	HXZ4	PS4	FF5	PS5	ALL11
$\Delta = R^2_{X+m} - R^2_X$	39.1	33.0	25.7	31.9	28.1	33.0	32.5	27.3	29.2
χ^2-test $p[\Delta = 0 \vert R^2 = 1]$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Wald-test $p[\Delta = 0 \vert R^2 = 1]$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
χ^2-test $p[\Delta = 0 \vert R^2 < 1]$	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
Wald-test $p[\Delta = 0 \vert R^2 < 1]$	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00

This table reports pricing performance comparison based on R^2 following Kan, Robotti, and Shanken (2013) with GLS (panels B and D) weighting matrix assumes miss-specification. The LHS variables are the excess returns of the 42 portfolios (25SZBM and 25 SZMOM less the 8 corner portfolios used to construct the efficient factor). Panels A and B tests whether common multi-factor model "X" adds additional pricing power to m^n. Panels C and D tests whether m^n adds additional pricing power to the model "X". The first row of each panel reports the R^2 difference between the larger and smaller models; the second and third rows report the p-value of the test that $\Delta = 0$ under correct specification, based on χ^2-test and Wald-test. the fourth and fifth rows report the p-values when miss-specification is assumed. The sample period is 1967.06.30—2014.12.31 with 511 effective months after excluding the first 5-year starting sample.

3.2.3 Spurious Factor?

Here I present evidence that the robust pricing performance of the efficient 1-factor model is not spurious.

Pointed out by recent study Bryzgalova (2014), a factor that is weakly correlated with the stock returns, can sometimes disguise as a strong factor that generate high statistical significance, high measure of fit via cross-sectional regression. Including such factor in the cross-sectional regression can crowd-out useful factors and make the estimation results unreliable. The study proposes to penalize the factors that have a small absolute value of beta in the cross-section.

There is also another potential problem for identification in the second-stage regression. Weak identification can also arise if a factor generates large but almost constant value of beta in the cross-section. Since there is close to no variation in the betas which enters the second-stage cross-sectional regression as an explanatory variable, the matrix of explanatory variables will be close to rank deficient when an intercept is also included (the intercept is required if zero-beta rates are estimated). The resulting slope, i.e factor premium, is also weakly identified[5]. To address this additional potential sources for weak identification concern, I present in Table 8 the cross-sectional regression reporting both the cross-sectional average of absolute value of beta as well as the cross-sectional standard deviation of betas for each factor. As reported, asset's beta w.r.t the efficient factor m^n is high in both magnitude and cross-sectional dispersion. When the efficient factor is incrementally added to each multi-factor models, the high magnitude and dispersion of betas persists.

3.3 Pricing Other Anomalies

Here I present the evidence that the pricing performance of the efficient factor m^n significantly extends to other assets: (1) The efficient factor can price a variety of different test assets; (2) The efficient factor formed on merely 3 anomalies can drive out 90% of other acclaimed anomalies.

3.3.1 Different Test Assets

Aside from testing on the 42 double-sorted portfolios, I also test the 30 univariate decile portfolios sorted on the Size, B/M and Momentum. Moreover, the pricing power of the efficient portfolio extracted from the 3 anomalies is able to price assets sorted on other predictive characteristics as well. I illustrate the performances with 3 additional sets: 25 Size-Investment portfolios, 25 Size-Operating Profitability portfolios and 25 Size-Net Issuance portfolios.

[5]When factors are excess returns, the leverage of the factors (something that is irrelevant to pricing) will automatically affect the magnitude of beta and the standard deviation. If one leverage up the excess return factor by 2, the volatility of the factor is scaled by 2, and beta will be scaled by 1/2. However, this issue is not a concern since the magnitude of the return of the efficient factor m^n is comparable to the other factors as shown in Table 2

Table 8: **Magnitudes and Variations of Beta's**

		(1) MKT	(2) SMB	(3) HML	(4) RMW	(5) CMA	(6) MOM	(7) qME	(8) qIA	(9) qROE	(10) QMJ	(11) LIQ	(12) m^n
RAP1	$\overline{\lvert\beta\rvert}$												$\lvert 0.80\rvert$
	$\widehat{\sigma}(\beta)$												(0.60)
CAPM1+1	$\overline{\lvert\beta\rvert}$	$\lvert 1.06\rvert$											$\lvert 0.56\rvert$
	$\widehat{\sigma}(\beta)$	(0.15)											(0.64)
FF3+1	$\overline{\lvert\beta\rvert}$	$\lvert 1.03\rvert$	$\lvert 0.57\rvert$	$\lvert 0.37\rvert$									$\lvert 0.43\rvert$
	$\widehat{\sigma}(\beta)$	(0.12)	(0.40)	(0.32)									(0.69)
FFC4+1	$\overline{\lvert\beta\rvert}$	$\lvert 1.01\rvert$	$\lvert 0.56\rvert$	$\lvert 0.28\rvert$			$\lvert 0.17\rvert$						$\lvert 0.18\rvert$
	$\widehat{\sigma}(\beta)$	(0.09)	(0.40)	(0.29)			(0.25)						(0.20)
AFP4+1	$\overline{\lvert\beta\rvert}$	$\lvert 1.06\rvert$	$\lvert 0.58\rvert$	$\lvert 0.39\rvert$							$\lvert 0.18\rvert$		$\lvert 0.43\rvert$
	$\widehat{\sigma}(\beta)$	(0.10)	(0.39)	(0.34)							(0.20)		(0.65)
HXZ4+1	$\overline{\lvert\beta\rvert}$	$\lvert 1.01\rvert$						$\lvert 0.52\rvert$	$\lvert 0.35\rvert$	$\lvert 0.17\rvert$			$\lvert 0.33\rvert$
	$\widehat{\sigma}(\beta)$	(0.10)						(0.37)	(0.33)	(0.24)			(0.45)
PS4+1	$\overline{\lvert\beta\rvert}$	$\lvert 1.03\rvert$	$\lvert 0.57\rvert$	$\lvert 0.37\rvert$								$\lvert 0.03\rvert$	$\lvert 0.43\rvert$
	$\widehat{\sigma}(\beta)$	(0.12)	(0.40)	(0.32)								(0.03)	(0.69)
FF5+1	$\overline{\lvert\beta\rvert}$	$\lvert 1.04\rvert$	$\lvert 0.58\rvert$	$\lvert 0.35\rvert$	$\lvert 0.18\rvert$	$\lvert 0.10\rvert$							$\lvert 0.43\rvert$
	$\widehat{\sigma}(\beta)$	(0.11)	(0.39)	(0.32)	(0.17)	(0.13)							(0.66)
PS5+1	$\overline{\lvert\beta\rvert}$	$\lvert 1.01\rvert$	$\lvert 0.56\rvert$	$\lvert 0.28\rvert$			$\lvert 0.17\rvert$					$\lvert 0.03\rvert$	$\lvert 0.18\rvert$
	$\widehat{\sigma}(\beta)$	(0.09)	(0.40)	(0.29)			(0.25)					(0.03)	(0.20)
ALL11+1	$\overline{\lvert\beta\rvert}$	$\lvert 1.02\rvert$	$\lvert 0.56\rvert$	$\lvert 0.27\rvert$	$\lvert 0.11\rvert$	$\lvert 0.11\rvert$	$\lvert 0.18\rvert$	$\lvert 0.14\rvert$	$\lvert 0.11\rvert$	$\lvert 0.13\rvert$	$\lvert 0.14\rvert$	$\lvert 0.03\rvert$	$\lvert 0.13\rvert$
	$\widehat{\sigma}(\beta)$	(0.06)	(0.43)	(0.29)	(0.13)	(0.14)	(0.25)	(0.16)	(0.14)	(0.12)	(0.18)	(0.03)	(0.17)

This table reports the magnitudes (as cross-sectional average of the absolute value) and variations (as cross-sectional sample standard deviations) of the β estimated in the first step of the 2-step cross-sectional regression. The first row encased by $|\cdot|$ reports the magnitudes and the second row in parenthesis reports the variations. The LHS variables are the excess returns of the 42 portfolios (25SZBM and 25 SZMOM less the 8 corner portfolios used to construct the efficient factor). The sample period is 1967.06.30—2014.12.31 with 511 effective months after excluding the first 5-year starting sample.

Table 9 summarizes the robust pricing performance of the efficient factor on these alternative test assets. It is evident that on all of the assets the risk-premium are estimated at level comparable to the average excess return of m^n, 7.34%. The robust t-statistics is highly significant, compared with the typical statistical significance of multi-factor models. The 1-factor model is able to price all these alternative test assets at high R^2 comparable to the levels achieved by the combined 11-factor model. Model specification tests strongly favor the 1-factor model. At 95% confidence:

(1) The hypothesis for correct specification ($H_0 : R^2 = 1$) is far from being rejected by the 1-factor model for each test assets, while the combined 11-factor model has smaller p-values and being borderline rejected for the 25 Size-Net Issuance portfolios.

(2) The hypothesis for complete miss-specification ($H_0 : R^2 = 0$) is consistently rejected for all test assets with p-value less that 0.01 when the 1-factor model is in

Table 9: **Different Test Assets: "RAP1" vs. "ALL11"**

Test Assets	(1) 21SZ-BM 21SZ-MOM	(2) 10SZ 10BM 10MOM	(3) 25SZ-INV	(4) 25SZ-OP	(5) 25SZ-NI
			OLS, $\boldsymbol{W} = \boldsymbol{I}_N$		
γ_{m^n}	4.21	4.67	6.58	5.66	5.71
t_{m^n}	[3.18]	[2.66]	[4.39]	[3.14]	[3.57]
R^2	81.6% vs. 88.5%	83.3% vs. 91.4%	81.1% vs. 82.0%	78.1% vs. 93.4%	62.8% vs. 76.1%
$p[R^2 = 1]$	0.50 vs. 0.17	0.63 vs. 0.20	0.66 vs. 0.17	0.64 vs. 0.50	0.18 vs. 0.05
$p[R^2 = 0]$	0.00 vs. 0.01	0.01 vs. 0.01	0.00 vs. 0.07	0.00 vs. 0.08	0.00 vs. 0.07
$p[\alpha(i) = 0]$	0.07 vs. 0.01	0.14 vs. 0.04	0.10 vs. 0.01	0.18 vs. 0.09	0.05 vs. 0.02
			WLS, $\boldsymbol{W} = diag(\widehat{\boldsymbol{\Sigma}}_u)$		
γ_{m^n}	4.31	4.35	6.63	5.80	5.32
t_{m^n}	[3.45]	[2.72]	[4.03]	[2.93]	[3.36]
R^2	74.7% vs. 86.4%	74.0% vs. 90.8%	79.3% vs. 82.6%	75.3% vs. 93.8%	54.5% vs. 77.8%
$p[R^2 = 1]$	0.42 vs. 0.12	0.59 vs. 0.36	0.70 vs. 0.08	0.59 vs. 0.51	0.19 vs. 0.09
$p[R^2 = 0]$	0.00 vs. 0.01	0.01 vs. 0.06	0.00 vs. 0.06	0.00 vs. 0.09	0.00 vs. 0.04
$p[\alpha(i) = 0]$	0.07 vs. 0.01	0.14 vs. 0.04	0.10 vs. 0.01	0.18 vs. 0.09	0.05 vs. 0.02

Top panel reports the 2-step OLS cross-sectional regression, and bottom panel for weighed least squares with weights being the diagonal of the residual variance matrix. Results are reported across different test assets. The LHS variables are the excess returns of the corresponding test assets. $\gamma(m^n)$ reports the estimated risk-premium annualized for the efficient factor, $t(m^n)$ is the Jagannathan and Wang (1998) heteroskedasticity and error-in-variable corrected t-statistics in square brackets; The rest is same as in Table 3. The sample period is 1967.06.30—2014.12.31 with 511 effective months after excluding the first 5-year starting sample.

use, while for the combined 11-factor model, it fails to reject miss-specification 6 out of 10 times.

(3) GRS test for the joint hypothesis that all asset pricing errors are zero cannot be rejected for the 1-factor model (except the borderline case for 25 Size-Net Issuance portfolios with a p-value of 0.05), yet for the 11-factor model it is rejected 8 out of 10 times.

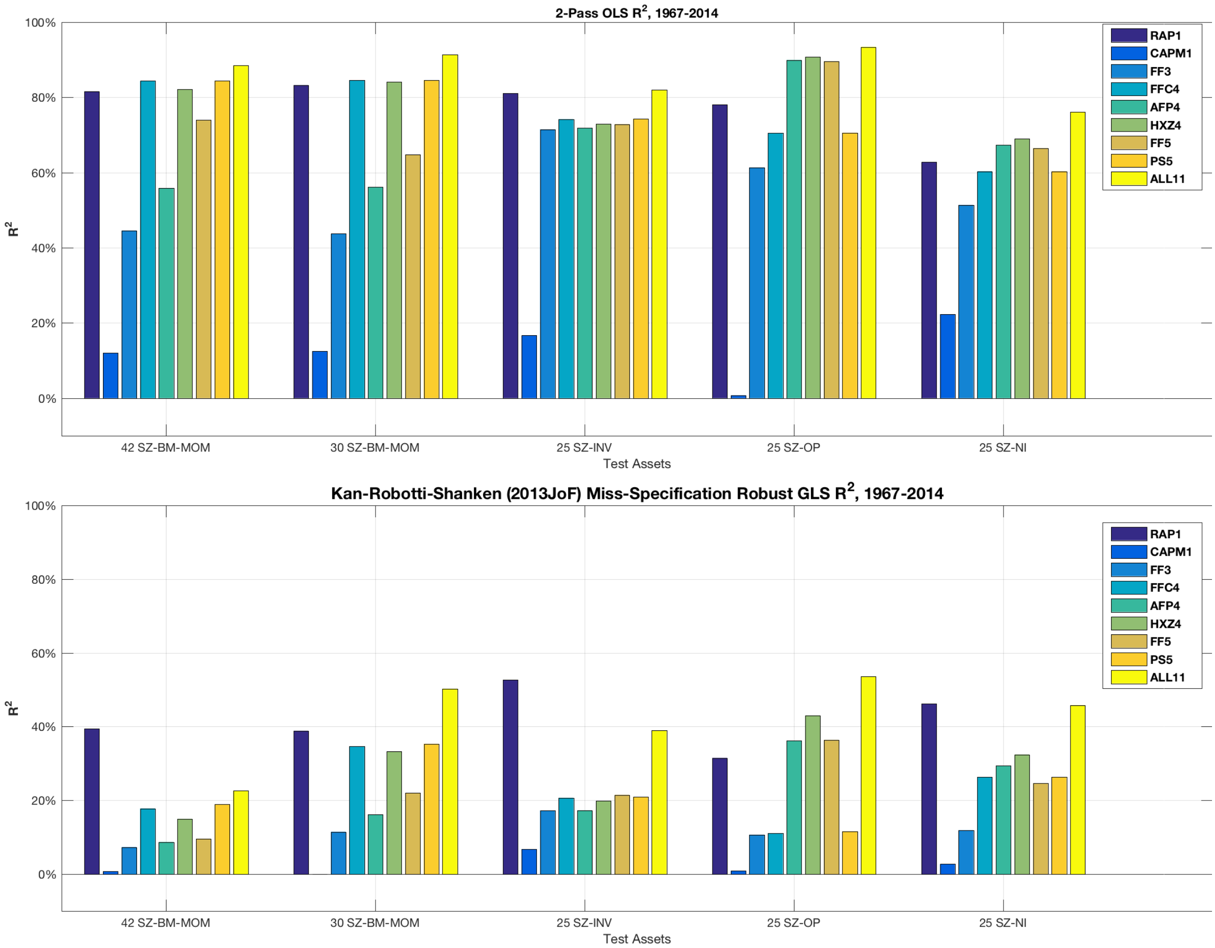
Additionally, for a quick overall view of the pricing power of the 1-factor model, I calculate the model predicted excess returns for the pooled 147 portfolios[6] (42+30+25+25+25). Figure 4 plots the predicted return versus the realized excess return comparing the 1-factor model with all 9 different multi-factor specifications including the combined 11-factor model.

3.3.2 Pricing Other "Unique" Factors

I present the evidence that the efficient factor is able to drive out a large number of anomalies within the limits of market friction.

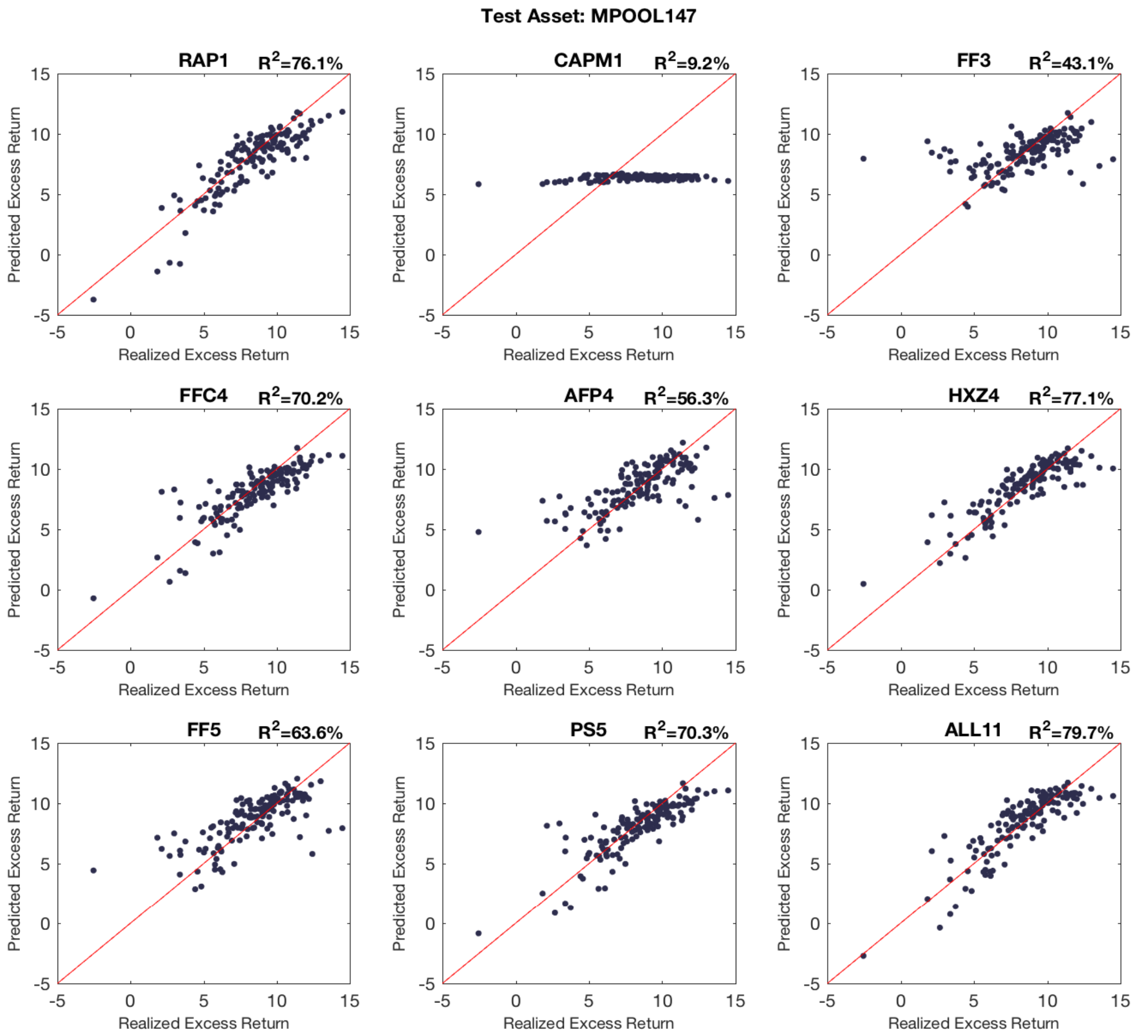
[6] Since the cross-sectional dimension is close to the time-series length, I only report the cross-sectional predicted returns

Figure 3: **R^2 across Test Assets**



The Efficient 1-factor extracted from the 8 portfolios utilizing the Size B/M and Momentum anomalies is able to drive out 39 out of 42 "Unique" anomalies, with 3 anomalies likely to diminish after implementation cost. Table 10 reports the time-series alpha's of the 42 anomalies estimated according to various factor models. The test assets include: 6 pricing factors from Kenneth French (SMB, HML, RMW, CMA, MOM, LTR), 3 q-factors from Lu Zhang (qME, qIA, qROE), 3 pricing factors from Andrea Frazzini (QMJ, BAB, DEV), 1 liquidity factor from Lubos Pastor and 32 trading strategies from Robert Novy-Marx, with 7 high turnover strategies. The point estimates of alpha's are reported for every anomaly/model pair, and significance at 95% confidence level is indicated by shaded point estimate. Different trading strategies may have very different time-series distributional characteristics, and may lead to different accuracy in the estimation of alpha. To put the significance into context, I simulated a probability that a random 1-factor model number could eliminate the significance of alpha's for a given anomaly, reported in the last column. A high probability that a random number could eliminate the significance of alpha indicates that the anomaly itself maybe extremely unstable, while a low probability indicates that the anomaly is

Figure 4: **Pricing Performance: Realized vs Predicted Return**



highly significant.

Out of all 42 anomalies, 4 has a shaded region, while one with "ValMom" has reversed sign of alpha in many factor models, moreover the probability that a random number could "explain" the anomaly is 42%, this indicates that the anomaly itself is very unstable, or that trading on such strategy may in fact lose money relative to the multi-factors. The 3 stable anomalies remaining are liquidity (LIQ) from Pástor and Stambaugh (2003), Short-Run Reversal and Seasonality. Notice that these 3 anomalies all have their spreads reduced by the 1-factor model: the 1-factor model is the only model that reduces the spread of liquidity factor (from 5.19 to 5.03), although the difference is small; while the spreads of the 2 high turnover strategies are reduced toward zero (from -16.16 to -13.38 and from -8.30 to -7.55). As shown by Novy-Marx and Velikov (2016), page 118, the Short-Run Reversal and Seasonality has monthly turnover over 90%, and do not even have positive excess return after cost. On the other hand, for the liquidity factor, it is unclear whether the implementation cost will exceed the liquidity spread, however, Pástor and Stambaugh (2003) Table 9 does show that firms with high liquidity beta's are the ones with the worst liquidity measure. Trading on such firms could incur significant cost that will reduce the spread.

Table 10: **Pricing 42 "Unique" Anomalies, Alpha's**

(1)	(2) Avg.	(3) CAPM1	(4) FF3	(5) FFC4	(6) HSZ4	(7) AFP4	(8) FF5	(9) PS5	(10) ALL11	(11) RAP1	(12) $p\,[Rand]$
m^n	7.34	7.23	6.53	5.40	4.95	5.39	5.67	5.37	4.37		0.00
	Panel A: from Kenneth French										
	Reference: Fama and French (1993, 2014); Carhart (1997)										
SMB	2.78	1.64			0.52					0.77	0.82
HML	4.39	5.60			0.32					-2.05	0.28
RMW	3.12	3.74	4.01	3.60	0.38	-2.71		3.50		0.21	0.36
CMA	4.34	5.40	2.84	2.36	-0.03	3.07		2.46		0.03	0.01
MOM	8.09	8.84	10.69		1.42	7.03	8.19			-4.93	0.04
LTR	3.46	3.56	0.76	0.56	0.95	2.11	0.32	0.86	0.58	-1.37	0.41
	Panel B: from Lu Zhang										
	Reference: Hou, Xue, and Zhang (2014)										
qME	3.63	2.55	0.69	0.39		0.48	0.57	0.47		0.76	0.52
qIA	5.35	6.30	4.11	3.58		3.91	1.56	3.67		0.73	0.00
qROE	6.57	7.23	8.67	5.79		2.40	5.20	5.96		1.60	0.00
	Panel C: from Andrea Frazzini										
	Reference: Asness et al. (2013, 2014); Asness and Frazzini (2013)										
QMJ	4.58	6.31	7.71	6.79	3.48		4.14	6.76		2.42	0.04
BAB	10.78	11.15	8.49	6.36	3.67	4.28	5.57	6.02	2.02	-0.61	0.00
DEV	4.46	5.08	-0.44	4.19	4.47	2.55	1.02	4.02	4.99	6.08	0.49
	Panel D: from Lubos Pastor										
	Reference: Pástor and Stambaugh (2003)										
LIQ	5.19	5.45	5.40	5.66	6.47	5.34	5.41			5.03	0.14
	Panel E: from Robert Novy-Marx, Low or Medium Turnover Strategies										
	Reference: Novy-Marx and Velikov (2016)										
Gross Prof.	4.22	4.31	6.43	5.74	1.64	-1.49	1.87	5.58	-0.43	0.46	0.43
ValProf	9.29	10.13	5.21	5.87	5.01	0.04	3.01	4.78	0.25	1.31	0.00
Accruals	2.15	2.99	2.47	2.21	3.94	5.10	2.99	1.58	3.02	1.85	0.91
Net Issu. Ann.	8.50	9.75	8.36	7.38	4.67	2.87	4.14	7.10	0.76	2.25	0.00
Asset Growth	4.03	5.10	0.78	0.23	-2.60	1.65	-2.29	0.86	-2.27	-1.03	0.64
Investment	5.33	6.04	3.90	3.24	1.95	5.25	2.42	3.79	1.38	2.31	0.07
Piotroski's F	1.30	3.10	3.97	2.41	-2.83	-4.40	0.06	2.34	-3.95	-0.92	0.99

Continued on next page

Table 10 **Pricing 42 "Unique" Anomalies, Alpha's**

Continued from previous page

(1)	(2) Avg.	(3) CAPM1	(4) FF3	(5) FFC4	(6) HSZ4	(7) AFP4	(8) FF5	(9) PS5	(10) ALL11	(11) RAP1	(12) $p\,[Rand]$
Asset Turn.	4.84	4.16	4.45	4.10	-0.64	-4.99	-1.06	3.71	-3.41	-1.43	0.47
Gross Margins	0.34	1.15	4.53	4.99	4.16	1.17	3.96	5.14	2.09	1.88	1.00
Ohlson's O	-0.37	2.09	5.87	3.14	-0.61	-1.45	1.87	3.20	-2.92	-4.20	1.00
Net Issu. Mon.	5.31	6.66	5.37	5.03	1.87	-0.94	0.87	4.21	-2.76	0.93	0.15
Ret-on-Bk-Eq.	3.96	6.24	9.09	5.08	-4.60	-3.11	1.72	5.82	-5.08	-2.73	0.93
Failure Prob.	2.94	7.74	13.07	3.60	-2.69	0.54	5.68	3.62	-5.64	-9.61	0.99
ValMomProf	12.26	13.63	12.42	4.04	5.65	7.35	9.39	3.72	2.06	-1.46	0.00
ValMom	6.24	7.68	3.34	-5.99	-2.92	2.46	2.34	-6.02	-4.36	-9.85	0.42
Idio. Vol.	2.66	7.92	7.38	3.66	-2.85	-5.00	0.46	3.70	-7.71	-9.47	0.99
PEAD (SUE)	3.09	3.36	5.67	1.22	-2.91	2.64	3.48	1.70	-2.29	-2.16	0.75
PEAD (CAR3)	4.08	4.56	5.74	3.39	3.41	4.52	5.35	3.27	2.11	1.84	0.28
Long-Run Rev.	1.22	1.06	-4.57	-3.34	-1.00	0.14	-3.54	-3.32	-1.73	-3.50	1.00
Ret-on-Mkt-Eq.	6.84	9.39	7.07	2.88	-3.35	-1.93	2.40	2.62	-4.07	-2.49	0.46
Ret-on-Assets	2.71	4.91	8.10	4.49	-3.45	-2.88	2.29	5.14	-4.34	-2.81	0.96
Beta Arb.	3.46	2.98	-0.06	-0.77	-2.64	-5.75	-3.21	-0.83	-7.61	-6.11	0.90
	Panel F: from Robert Novy-Marx, High Turnover Strategies Reference: Novy-Marx and Velikov (2016)										
Ind. Mom.	-3.83	-2.39	-1.68	-4.85	-4.12	-1.60	-1.83	-4.49	-4.96	-5.16	0.94
Ind. Rel. Rev.	-10.47	-12.51	-13.36	-9.74	-9.64	-11.51	-11.67	-10.19	-8.86	-7.17	0.00
Hi-Freq Comb.	1.49	0.93	0.36	0.00	0.30	1.29	1.12	-0.09	-0.21	0.15	0.99
Short-Run Rev.	-16.16	-18.46	-19.27	-15.14	-15.75	-17.80	-18.17	-15.70	-14.91	-13.38	0.00
Seasonality	-8.30	-9.06	-7.58	-8.49	-7.14	-8.22	-7.21	-8.55	-6.92	-7.55	0.01
Ind.Rel.Rev.(LV)	1.41	0.20	-0.88	0.44	-0.15	0.57	-0.17	0.24	0.50	-1.27	0.99
Hi-Freq.(HS)	2.52	1.64	1.95	0.98	2.68	3.45	3.35	1.11	2.98	2.02	0.93

This table reports the time series alpha's of the various returns against the multi-factor models and the 1-factor model. Significance at 95% confidence level is marked by shaded numbers. Standard errors are calculated according to Newey and West (1987) with Newey and West (1994) automatic lag selection. The LHS variables are the return series obtained from different sources according to the corresponding publications. The last column reports the probability that a random 1-factor model could eliminate the alpha of the corresponding anomaly. The random factor is drawn from normal distribution with mean and variance equal to the efficient factor's for 1000 iterations. Data series are either directly provided by the corresponding author or downloaded from their websites. The detailed description of the trading strategies is in Novy-Marx and Velikov (2016) Appendix B. The sample period is 1967.06.30—2014.12.31 with 511 effective months after excluding the first 5-year starting sample.

4 Pricing at the Firm-Level

This section I present the evidence that the efficient factor is priced significantly in the cross-section of stock returns at the firm-level. I first form a forward-bias-free decile portfolio on the expected efficient factor beta to demonstrate its significant spread w.r.t existing multi-factor models. Then I show that one could engineer an anomaly using an optimal mixture of exotic characteristics to form an anomaly portfolio that would pass existing tools as a "unique" anomaly, yet is completely manifested by the efficient factor.

4.1 Portfolios Formed on Expected Beta

In this section I group portfolios according to assets's expected beta to the efficient factor to demonstrate a significant 12% spread that cannot be explained by many multi-factor models, except the Carhart (1997) 4-factor model and Hou, Xue, and Zhang (2014) 4-factor model, and the Efficient 1-factor model.

If the efficient factor is a valid risk factor for cross-sectional asset returns, and that the risk-premium is significantly positive in the cross-sectional regression, then portfolios with securities having high conditional sensitivity (expected beta) to the efficient factor should earn a higher expected return. To demonstrate this, I adopt the expected beta approach similar to Pástor and Stambaugh (2003) by specifying a set of instrumental variables as the conditional information set. Also, in the theoretical framework, I give exact form of the expected beta as a function of the characteristics instruments. The expected beta for an asset i formed on time t information is determined by the set of time t normalized instruments $\boldsymbol{z}_t(i)$:

$$\beta_t(i) = \phi_{1,t} + \boldsymbol{z}_t(i)'\boldsymbol{\phi}_{2,t} \tag{6}$$

The characteristic instruments specified need to be "comprehensive" enough to reflect the true conditional information set, but at the same time "succinct" enough to allow reliable estimation. To do so, I include the 3 basis characteristics as well as an additional total volatility characteristics. As shown by MacKinlay and Pástor (2000) and further by Chen and Petkova (2012), the residual variance could proxy the amount of miss-pricing a security has with respect to a particular asset pricing model. By the same token, in the absence of factors, the total variance could proxy for the amount of total miss-pricing that is available for a factor model to capture. Thus I set:

$$\boldsymbol{Z}_t(i) = \{SZ_t(i), BM_t(i), MOM_t(i), TV_t(i)\}'$$

they are the firm size, B/M ratio, momentum (past 12 to 2 month return), total variance (as measured by realized variance estimated on past month daily returns), respectively.

For each original instrument $Z_t^k(i)$ in $\boldsymbol{Z}_t(i)$, the normalized instrument $z_t^k(\gamma)$ is transformed to have cross-sectional uniform distribution on $[0,1]$ while preserving the cross-sectional ranking[7].

$$
\begin{aligned}
z_t^k(i) =& F_t^k\left[Z_t^k(i)\right] \\
F_t^k(a) =& \frac{\sum_i \mathbb{1}_{[Z_t^k(i)<a]}}{\sum_i 1}
\end{aligned}
$$

meaning that at time t if asset i has the highest value of size characteristics $SZ_t(i)$ among all assets, then $sz_t(i) = 1$; while if asset j has the median value of characteristics $BM_t(j)$ among all assets, then $bm_t(j) = 0.5$. Moreover, as illustrated in Section 2.1.4, there exists non-monotonic or even reversal of benefits from a given characteristics. To accommodate the non-monotonicity, I include 3 "corner" indicators on the momentum characteristics in order to capture any "reversal of benefits":

$$
\begin{aligned}
mom^5 sz_t^1(i) &= \mathbb{1}_{[sz_t(i)<0.2, mom_t(i)>0.8]} \\
mom^5 bm_t^5(i) &= \mathbb{1}_{[bm_t(i)>0.8, mom_t(i)>0.8]} \\
mom^5 tv_t^1(i) &= \mathbb{1}_{[tv_t(i)<0.2, mom_t(i)>0.8]}
\end{aligned}
$$

resulting in the following specification of $\boldsymbol{z}_t(i)$

$$
\boldsymbol{z}_t(i) = \left\{sz_t(i), bm_t(i), mom_t(i), tv_t(i), mom^5 sz_t^1(i), mom^5 bm_t^5(i), mom^5 tv_t^1(i)\right\}' \tag{7}
$$

The coefficients $\boldsymbol{\phi}$'s are estimated in a pooled regression with expanding windows, using all available observations up to time t:

$$
\widetilde{r}_{s+1}(i) = \phi_{0,t} + \left(\phi_{1,t} + \boldsymbol{z}_s(i)'\boldsymbol{\phi}_{2,t}\right) m_{s+1}^n + u_{s+1}(i) \qquad s \leq t-1 \tag{8}
$$

Thus we can calculate the expected relative beta given the estimated $\widehat{\phi}_{1,t}, \widehat{\boldsymbol{\phi}}_{2,t}$ via Equation (6). Table 11 reports the summary of the estimated coefficients $\boldsymbol{\phi}_t$, together with a full-sample estimated $\boldsymbol{\phi}_T$ for comparison. As shown in the table, while momentum is the strongest positive predictor of expected beta, size and volatility are strong negative predictor of expected beta. If monotonic specification captures all conditional information, then the non-monotonicity terms should be positive or insignificant, however, it is evident that the combination of high momentum and low volatility stocks in fact strongly and negatively predict expected beta, contrasting the direction of the char-

[7] This is equivalent to assigning quantile portfolios weights to each stock based on the characteristics ranking. It is also motivated by the recursive asset pricing model in Section 5.

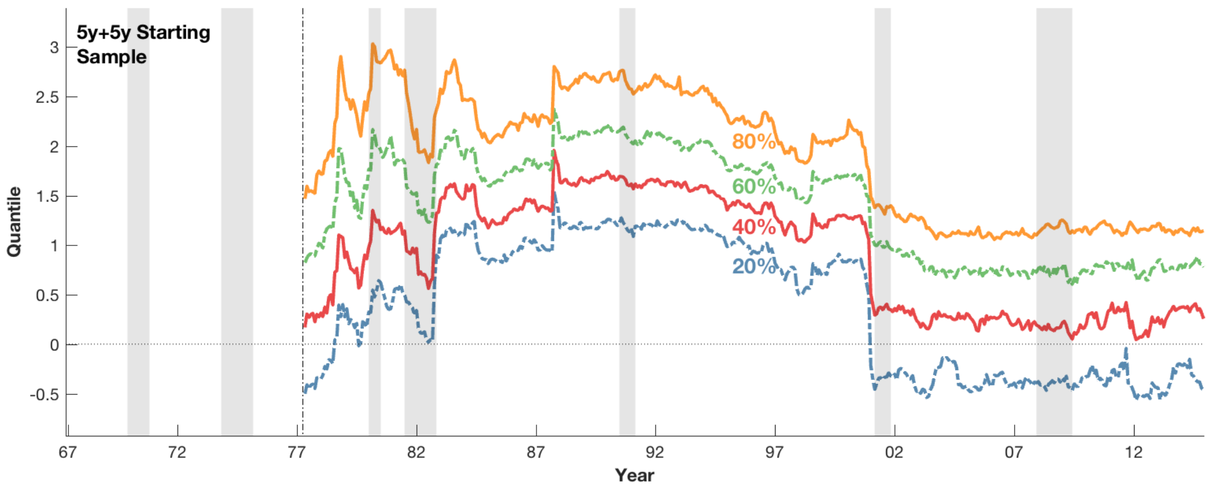
acteristics' monotonic predictability. This non-monotonicity is again similar to the "reversal of benefits" observed earlier. Figure 5 plots the quantiles of the estimated efficient factor beta over time.

Table 11: **Estimated ϕ**

(1)	(2) ϕ_0	(3) ϕ_1	(4) $sz_t(i)$	(5) $bm_t(i)$	(6) $mom_t(i)$	(7) $tv_t(i)$	(8) $mom^5 sz_t^1(i)$	(9) $mom^5 bm_t^5(i)$	(10) $mom^5 tv_t^1(i)$
ϕ_T	0.47	-0.08	-0.55	-0.01	2.47	-0.89	-0.04	-0.08	-0.65
$\sigma(\phi_T)$	0.03	0.09	0.07	0.07	0.09	0.08	0.16	0.10	0.07
$t(\phi_T)$	16.94	-0.85	-7.32	-0.20	27.48	-10.65	-0.24	-0.73	-9.73
$\overline{\phi_t}$	-0.01	0.45	-0.93	0.03	2.74	-0.35	-0.28	-0.08	-0.90

This table summarize the time series of the estimated coefficients ϕ_t in the calculation of expected beta. Sample Period: 1967.06.30—2014.12.31;

Figure 5: **Quintiles of the Expected Efficient Factor Beta**



Securities are grouped into quintile portfolios at the end of every month t, using expected beta $\widetilde{\beta}_t$ which is estimated using information up to time t. The securities are then held for 1 month earning portfolio return for the next month $t+1$. The securities are re-grouped into quintile portfolios every month. Figure 6 demonstrate the quintile portfolios relative performance over time. Table 12 summarizes the portfolio's details, as well as reports their alpha's with respect to each asset pricing model.

Patton and Timmermann (2010) monotonicity test

[Insert Figure Average Return About Here]

4.2 Engineering An "Anomaly"

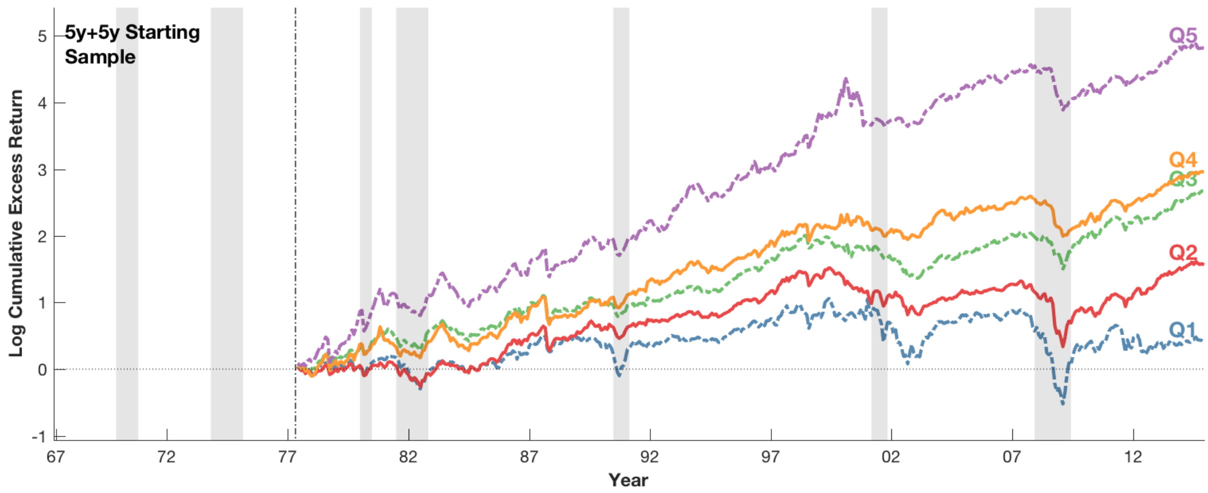
Using the expect beta framework, one can essentially engineer an anomaly based on the full sample estimated ϕ's. If one wish to find a certain function of observable firm

Table 12: **Portfolios Grouped on Expected Efficient Factor Beta**

Quintile	Low	2	3	4	High	High-Low
			Value Weighted, Excess Return			
Avg. Ret.	3.71	5.59	8.36	9.45	15.60	11.89
	[0.99]	[2.05]	[3.29]	[3.34]	[4.17]	[3.12]
CAPM1	-5.19	-1.79	1.36	1.80	6.51	11.70
	[-2.33]	[-1.65]	[1.34]	[1.59]	[3.19]	[3.26]
FF3	-6.79	-3.21	0.73	2.01	7.34	14.13
	[-3.19]	[-3.25]	[0.69]	[1.93]	[4.14]	[4.33]
FFC4	0.43	-1.17	-0.07	-0.78	2.40	1.97
	[0.27]	[-1.27]	[-0.07]	[-0.80]	[1.63]	[0.89]
HSZ4	1.30	-2.35	-1.09	-1.52	3.99	2.68
	[0.43]	[-1.58]	[-0.98]	[-1.36]	[1.81]	[0.58]
PS4	-7.29	-3.40	0.88	1.94	7.61	14.90
	[-3.44]	[-3.30]	[0.80]	[1.79]	[4.27]	[4.57]
FF5	-3.22	-3.50	-0.59	-0.06	7.38	10.60
	[-1.24]	[-2.95]	[-0.62]	[-0.05]	[3.46]	[2.53]
ALL11	1.84	-2.30	-0.42	-1.25	5.89	4.05
	[1.27]	[-2.22]	[-0.38]	[-1.22]	[3.94]	[2.10]
RAP1	8.65	2.94	4.40	2.82	7.32	-1.33
	[1.28]	[0.73]	[1.38]	[0.82]	[1.67]	[-0.23]
			Equal Weighted, Excess Return			
Avg. Ret.	7.08	10.48	10.03	12.36	14.81	7.73
	[1.65]	[3.54]	[4.00]	[4.71]	[4.49]	[2.57]
CAPM1	-2.44	3.09	3.41	5.47	6.86	9.29
	[-0.88]	[1.76]	[2.32]	[3.53]	[3.41]	[3.52]
FF3	-5.04	0.42	1.13	3.63	5.77	10.81
	[-2.38]	[0.38]	[1.31]	[4.11]	[5.10]	[4.42]
FFC4	2.39	3.13	1.88	2.76	3.24	0.85
	[1.21]	[2.96]	[2.12]	[3.07]	[3.19]	[0.49]
HSZ4	5.15	3.50	1.03	2.76	4.88	-0.27
	[1.55]	[2.13]	[0.84]	[2.75]	[3.52]	[-0.07]
PS4	-5.31	0.39	1.09	3.61	5.88	11.19
	[-2.57]	[0.34]	[1.26]	[3.97]	[5.07]	[4.62]
FF5	-0.42	1.21	0.08	2.57	5.50	5.92
	[-0.14]	[0.90]	[0.09]	[2.93]	[4.49]	[1.76]
ALL11	8.10	4.06	0.99	3.02	6.13	-1.97
	[3.79]	[3.83]	[1.16]	[3.41]	[5.38]	[-1.00]
RAP1	13.67	8.19	4.25	5.25	6.13	-7.54
	[1.61]	[1.55]	[1.02]	[1.29]	[1.34]	[-1.39]
			Portfolio Details, Averaged			
# Firms.	225	226	226	226	225	
Size (bl.$)	1.04	1.71	2.24	2.07	0.81	
Expected β, vw	0.07	0.65	1.16	1.66	2.23	
Post-Form. β, vw	-0.08	0.67	0.90	1.18	1.58	

This table reports the annualized alpha's of the quintile portfolios grouped on expected efficient factor beta. The top panel presents the value-weighted portfolios and the bottom panel for the equal-weighted portfolios. The first row reports the return and the second row reports the Newey and West (1987) robust t-statistics with Newey and West (1994) automatic lag selection. Sample period: 1967.06.30—2014.12.31, with 10 years of starting sample.

Figure 6: **Value-Weighted Quintile Portfolios on Expected Efficient Factor Beta**



characteristics such that it predicts future return but robust to all existing factors in a multi-factor test. One could simply do the following steps:

(1) Choose a set of observable firm characteristics $\boldsymbol{z}_t^e(i)$. For novelty, one could set such observable characteristics to some un-explored features.
(2) Run a full-sample regression using Equation (8), with the chosen $\boldsymbol{z}_t^e(i)$. One thus obtains a full-sample optimal $\boldsymbol{\phi}_T^e$.
(3) Compute the engineered best mix of characteristics using the full-sample formula $a_t^e(i) = \boldsymbol{z}_t^e(i)'\boldsymbol{\phi}_{2,T}^e$
(4) Group portfolios according to the engineered best mix $a_t^e(i)$. By design, it maximally proxies for the expected relative beta ex-post, and should significantly predict return. Moreover, this characteristics is by design, robust to all other risk-factors. Neither multi-factor tests, nor double sorts will be able to explain such return predictability, therefore, a "unique anomaly" has been discovered.
(5) Seek economic or behavioral motivations for why such characteristics will predict returns, even though it will be completely explained by the efficient excess return portfolio.

Table 13 shows the the engineered anomaly with the same set of characteristics as in the previous section.

Double Sort controlling an array of other characteristics

[Insert Table **??** About Here]

Table 13: **Engineered "Anomaly": Alpha's**

	Value Weighted, Excess Return					
Quintiles	Low	2	3	4	High	High-Low
Avg. Ret.	-0.67	1.84	5.60	5.22	14.48	15.16
	[-0.18]	[0.53]	[1.99]	[1.87]	[4.86]	[4.53]
CAPM1	-8.87	-4.82	0.02	-0.49	7.62	16.50
	[-3.78]	[-1.98]	[0.01]	[-0.27]	[5.38]	[5.54]
FF3	-10.53	-6.22	-0.36	-0.69	7.91	18.44
	[-4.64]	[-2.45]	[-0.23]	[-0.36]	[6.05]	[6.38]
FFC4	-1.99	-2.61	-0.17	-2.44	2.92	4.91
	[-1.11]	[-1.10]	[-0.10]	[-1.32]	[2.54]	[2.37]
HSZ4	-2.60	-4.02	-0.20	-2.64	4.23	6.83
	[-0.83]	[-1.53]	[-0.12]	[-1.38]	[2.57]	[1.63]
PS4	-10.89	-6.13	-0.21	-0.69	8.33	19.22
	[-4.96]	[-2.42]	[-0.13]	[-0.37]	[6.10]	[6.72]
FF5	-8.08	-6.11	-0.60	-1.56	7.31	15.40
	[-2.94]	[-2.46]	[-0.40]	[-0.79]	[4.52]	[4.04]
ALL11	-0.46	-3.91	0.84	-1.99	4.21	4.67
	[-0.28]	[-1.83]	[0.41]	[-1.22]	[3.22]	[2.23]
RAP1	5.00	1.62	2.52	1.33	5.90	0.89
	[0.72]	[0.39]	[0.81]	[0.41]	[1.52]	[0.15]
	Equal Weighted, Excess Return					
Quintiles	Low	2	3	4	High	High-Low
Avg. Ret.	6.14	7.84	7.65	8.60	14.09	7.96
	[1.41]	[2.20]	[2.58]	[3.04]	[4.98]	[2.87]
CAPM1	-2.37	1.17	2.01	3.05	7.66	10.04
	[-0.81]	[0.44]	[1.09]	[1.46]	[4.26]	[4.20]
FF3	-5.46	-1.40	-0.24	1.39	6.33	11.79
	[-2.59]	[-0.62]	[-0.16]	[0.79]	[6.04]	[5.27]
FFC4	2.03	1.65	1.27	0.37	3.70	1.67
	[0.92]	[0.77]	[0.80]	[0.21]	[3.86]	[0.95]
HSZ4	4.32	1.50	1.12	0.67	4.98	0.66
	[1.33]	[0.65]	[0.66]	[0.37]	[4.11]	[0.20]
PS4	-5.63	-1.34	-0.14	1.49	6.53	12.17
	[-2.71]	[-0.60]	[-0.09]	[0.86]	[5.96]	[5.37]
FF5	-1.98	-1.04	-0.32	0.80	5.98	7.96
	[-0.69]	[-0.46]	[-0.22]	[0.45]	[5.62]	[2.57]
ALL11	8.03	1.63	1.97	1.29	5.12	-2.91
	[3.39]	[0.91]	[1.09]	[0.91]	[4.90]	[-1.44]
RAP1	11.95	7.86	3.25	3.49	5.07	-6.88
	[1.41]	[1.51]	[0.78]	[0.89]	[1.17]	[-1.30]

This table reports the annualized alpha's of the quintile portfolios grouped on engineered anomaly characteristics $a_t(i)$. The top panel presents the value-weighted portfolios and the bottom panel for the equal-weighted portfolios. The first row reports the return and the second row reports the Newey and West (1987) robust t-statistics with Newey and West (1994) automatic lag selection. Sample period: 1967.06.30—2014.12.31, with 10 years of starting sample.

5 A Recursive Asset Pricing Model

To formally setup the theoretical framework, I adopt an approach similar to that of Gagliardini, Ossola, and Scaillet (2016) in order to specify both time and cross-sectional dependence.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Let τ be a measure-preserving, ergodic time-shift transformation mapping Ω onto itself. If $\omega \in \Omega$ is the state of the world at time 0, then $\tau^t(\omega)$ is the state at time t. In order to define the information sets, let $\mathcal{F}_0 \subset \mathcal{F}$ be a sub sigma-field. Define $\mathcal{F}_t = \{\tau^{-t}(A), A \in \mathcal{F}_0\}$, $t = 1, 2, ...$, through the inverse mapping τ^{-t} and assume that $\mathcal{F}_1$ contains $\mathcal{F}_0$. Then the filtration $\mathcal{F}_t$, $t = 1, 2, ...$, characterizes the flow of information available to investors. The conditional expectations are defined as $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot|\mathcal{F}_t]$ for $t = 1, 2, ...$

The economy has a continuum of assets indexed by $i \in \Gamma \equiv [0, 1]$. The index set is endowed with the Borel sigma-field $\mathcal{B}$ and a probability distribution G that is absolutely continuous w.r.t the Lebesgue measure[8].

On this probability space, let κ, $\boldsymbol{\delta}$, together with $\boldsymbol{z}(i)$ and $e(i)$ for any $i \in \Gamma$ be measurable functions w.r.t $\mathcal{F}_0$. Further define the random vector $\boldsymbol{\delta}_t(\omega) = \boldsymbol{\delta}(\tau^{t-1}\omega)$ admitting values in $\mathbb{R}^K$, random variable $\kappa_t(w) = \kappa(\tau^{t-1}(\omega))$ admitting values in $\mathbb{R}$; and for any $i \in \Gamma$, the collection of random vectors $\boldsymbol{z}_t(i, \omega) = \boldsymbol{z}(i, \tau^{t-1}(\omega)) = \left\{z_t^1(i, \omega), z_t^2(i, \omega), ..., z_t^K(i, \omega)\right\}'$ admitting values in $\mathbb{R}^K$, the collection of random variables $e_t(i, w) = e(i, \tau^{t-1}(\omega))$, both admitting values in $\mathbb{R}$.

5.1 Predictable Return

The following assumptions are made regarding the return $r_{t+1}(i)$ and prediction dynamics from relevant features $\boldsymbol{z}_t(i)$:

Assumption 1. ***Linear Prediction:***
For all asset $i \in \Gamma$, the net return $r_{t+1}(i)$ at dates $t = 1, 2, ...$, satisfy the conditional linear prediction model

$$r_{t+1}(i) = \kappa_{t+1} + \boldsymbol{z}_t(i)'\boldsymbol{\delta}_{t+1} + e_{t+1}(i) \tag{9}$$

where $\boldsymbol{z}_t(i)$ are the collection of all features known at time t that satisfy:

$$\begin{array}{ll} \textit{Non-degenerated} & 0 < \int_\Gamma \left[z_t^k(i)\right]^2 dG(i) < \infty \\ \textit{Normalized} & \int_\Gamma z_t^k(i) dG(i) = 0 \\ \textit{Relevant} & \mathbb{P}\left[\delta_{t+1}^k = 0 \middle| \mathcal{F}_t\right] < 1 \end{array}$$

for all $k = 1, 2, ..., K$

The last condition of Assumption 1 states that if there are any other asset feature $z_t^{K+1}(i)$ that are non-degenerated and normalized with a δ_{t+1}^{K+1} that satisfies Equation

[8] A countable collection of assets from this economy is a sequence (i_j) in $\Gamma^\infty = [0, 1]^\infty$, endowed with the product sigma-field $\mathcal{B}^\infty$ and the product measure $\mu = G^\infty$

9, the feature must be useless because $\mathbb{P}\left[\delta^{K+1}_{t+1} = 0 \middle| \mathcal{F}_t\right] = 1$

Assumption 2. ***Identification:***
For all $i \in \Gamma$:

(1) For any weighting scheme $-\infty < w_t(i) < \infty$ known at time t, non-degenerated, i.e. $0 < \int_\Gamma [w_t(i)]^2\, dG(i) < \infty$, the individual errors $e_{t+1}(i)$ satisfy

$$\mathbb{P}\left[\int_\Gamma w_t(i) e_{t+1}(i) dG(i) = 0 \middle| \mathcal{F}_t\right] = 1 \tag{10}$$

(2) The cross-sectional dispersion matrix $\boldsymbol{S}_t := \int_\Gamma \boldsymbol{z}_t(i)\boldsymbol{z}_t(i)' dG(i)$ is positive definite, almost surely

Assumption 2 part (1) ensure that the error terms are diversified away, as long as the weighting scheme assigns finite, non-zero weights to infinitely many assets. However, it does not mean that the error terms are zero for each asset, because a weighting scheme that picks finitely many assets will inevitably be degenerated as a finite collection of asset has measure 0 according to $G(\cdot)$. Assumption 2 part (2) ensures that the features are themselves unique, that one cannot be completely linearly dependent on others. The Assumption 2 is not restrictive as it is the *necessary and sufficient* condition for identification of the model in Assumption 1 by the following proposition:

Proposition 1. ***Identification***
κ_{t+1} and $\boldsymbol{\delta}_{t+1}$ in equation (9) are identified with probability 1 with respect to $\mathcal{F}_{t+1}$ if and only if Assumption 2 is satisfied.

For readability, for the remainder of this section, I suppress all notations regarding "almost sure" except when necessary, i.e. the remaining assumptions and derivations holds when the above defined random-functions are restricted to $\bar{\Omega} \subset \Omega$ where $\mathbb{P}(\bar{\Omega}) = 1$.

***Remark* 1.** *κ_{t+1} is the market return*
The time-specific intercept κ_{t+1} has the special interpretation of the aggregate market return. To see this, given Assumption 1 and Assumption 2, integrate over the asset index set Γ immediately arrive at:

$$\int_\Gamma r_{t+1}(i) dG(i) = \kappa_{t+1} \quad \text{for any } t = 1, 2, ... \tag{11}$$

Now, consider the following investment strategy to exploit predictive feature k. At the end of time t, with asset features $\boldsymbol{z}_t$ known, first form a long-leg with assets high in feature k:

$$l^k_{t+1} = \int_\Gamma r_{t+1}(i) \left|z^k_t(i)\right| \mathbb{1}_{[z^k_t(i)>0]} dG(i)$$

where each asset i is weighted by $|z_t^k(i)|\mathbb{1}_{[z_t^k(i)>0]}$. Similarly a short-leg with assets low in feature k:

$$s_{t+1}^k = \int_\Gamma r_{t+1}(i)\left|z_t^k(i)\right| \mathbb{1}_{[z_t^k(i)<0]} dG(i)$$

The long-short portfolio is then

$$x_{t+1}^k = l_{t+1}^k - s_{t+1}^k = \int_\Gamma r_{t+1}(i) z_t^k(i) dG(i) \tag{12}$$

Notice that this portfolio is zero cost since the total value invested $\int z_t^k(i)dG(i) = 0$ from the normalization in Assumption 1. Thus K long-short zero cost portfolios could be formed for all K features:

$$\boldsymbol{x}_{t+1} = \left\{x_{t+1}^1, x_{t+1}^2, ...\right\}' = \int_\Gamma \boldsymbol{z}_t(i) r_{t+1}(i) dG(i) \tag{13}$$

Directly substitute equation (9) into equation (13) and use the cross-sectional averages from Assumption 1 and Assumption 2, we have the following diversification result:

Remark 2. *$\boldsymbol{x}_{t+1}$ are multi-factor portfolios*
The randomness of the collection of long-short portfolios form by equation (13) for each of the features depends only on the vector $\boldsymbol{\delta}_{t+1}$, thus are well-diversified such that:

$$\boldsymbol{x}_{t+1} = \boldsymbol{S}_t \boldsymbol{\delta}_{t+1} \tag{14}$$

Note that $\boldsymbol{S}_t$ is known at time t, while the individual error terms $e_{t+1}(i)$, and time-specific intercept, or market return κ_{t+1} that belong to time $t+1$ information set disappears.

This is a generalization of the long-short factors various papers have proposed. For example, an equal-weighted Size factor in Fama and French (1993) are such grouping methods that set $z_t^k(i) = \mathbb{1}_{[Size_t(i)>q_t]} - \mathbb{1}_{[Size_t(i)<q_t]}$, thereby going long in larger firms and short smaller firms.

Remark 3. *Consistent predictability do not imply consistent portfolio performance*
The above Remark 2 illustrates that the long-short anomaly portfolios grouped on features depend on both the predictability coefficient $\boldsymbol{\delta}_{t+1}$ and the cross-sectional dispersion of all relevant features $\boldsymbol{S}_t$. Even when a certain feature consistently predict asset returns, $\boldsymbol{x}_{t+1}$ may not be stable over time at all, since the dispersion matrix $\boldsymbol{S}_t$ fluctuates and uses *all* relevant features.

5.2 One-Factor Relative Pricing

The next step is to construct a global efficient excess return portfolio. Let $\boldsymbol{\Sigma}_t = Var_t(\boldsymbol{x}_{t+1})$ and $\boldsymbol{\mu}_t = \mathbb{E}_t[\boldsymbol{x}_{t+1}]$ be the conditional variance and mean of the long-short portfolios $\boldsymbol{x}_{t+1}$, and consider the following 2 problems an investor faces while allocating capital among the long-short portfolios $\boldsymbol{x}_{t+1}$

Problem 1. *Minimize Variance Given Return*

$$\underset{\boldsymbol{w}}{\arg\min} \quad Var_t[\boldsymbol{w}'\boldsymbol{x}_{t+1}] \quad \text{s.t. } \mathbb{E}_t[\boldsymbol{w}'\boldsymbol{x}_{t+1}] = y_1 \tag{15}$$

Problem 2. *Minimize Tracking Error towards a Fixed Return*
For $y_2 \in \mathbb{R}$

$$\underset{\boldsymbol{w}}{\arg\min} \quad \mathbb{E}_t\left[(y_2 - \boldsymbol{w}'\boldsymbol{x}_{t+1})^2\right] \tag{16}$$

Assumption 3. ***No Redundancy:***
For all $i \in \Gamma$, the K predictive features $\boldsymbol{z}_t(i)$ are not redundant, i.e.

$$\boldsymbol{D}_t = \mathbb{E}_t\left[\boldsymbol{\delta}_{t+1}\boldsymbol{\delta}'_{t+1}\right] - \mathbb{E}_t[\boldsymbol{\delta}_{t+1}]\,\mathbb{E}_t\left[\boldsymbol{\delta}'_{t+1}\right] \tag{17}$$

is positive definite with probability 1, for $t = 1, 2, ...$

It is easy to show that the 2 problems are equivalent and the following special portfolio:

$$m_{t+1} = \boldsymbol{x}'_{t+1}\boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mu}_t = \boldsymbol{\delta}'_{t+1}\boldsymbol{D}_t^{-1}\mathbb{E}_t[\boldsymbol{\delta}_{t+1}] \tag{18}$$

with weights

$$\boldsymbol{w}_t^* = \boldsymbol{\Sigma}_t^{-1}\boldsymbol{\mu}_t \tag{19}$$

is the unique (up to a scaler) solution to both problems:

Proposition 2. m_{t+1} ***is an efficient portfolio among*** $\boldsymbol{x}_{t+1}$ *The set of solutions to Problems 1 and 2 are the same and if Assumptions 1, 2 and 3 are satisfied, the solutions are given by*

$$\{\boldsymbol{w} | \boldsymbol{w} = c\boldsymbol{w}_t^*, c \in \mathbb{R}\} \tag{20}$$

Not only m_{t+1} is efficient among the particular portfolios $\boldsymbol{x}_{t+1}$, the efficiency in fact extends to all universe of excess returns via the following proposition. One could obtain global efficient excess return portfolio by only using the several "anomalous" portfolios, thus significantly reduce the dimensionality of portfolio optimization.

Assumption 4. ***Strict Exogenous Errors***
For all $i \in \Gamma$, the individual residual errors $e_{t+1}(i)$ of Equation (9) satisfy:

$$\mathbb{E}_t \left[e_{t+1}(i) | \boldsymbol{\delta}_{t+1} \right] = 0 \tag{21}$$

for $t = 1, 2, ...$

Proposition 3. m_{t+1} ***is efficient for all excess returns***
Given Assumptions 1, 2, 3 and 4, the zero-cost portfolio defined by Equation (18), with weights defined by Equation (19) is a conditional efficient portfolio for all excess returns. Specifically, any excess return portfolio r^e_{t+1}, formed at time t, satisfy the orthogonal decomposition

$$r^e_{t+1} = \boldsymbol{\omega}'_t \boldsymbol{x}_{t+1} + v_{t+1} \tag{22}$$

for some $\boldsymbol{\omega}_t$ and v_{t+1}, where $\mathbb{E}_t \left[v_{t+1} | \boldsymbol{x}_{t+1} \right] = 0$

With the ex-ante efficient excess return portfolio obtained, we are ready to present the main theoretical result.

Proposition 4. ***One-Factor Relative Pricing (extension of Roll (1977))***
Given Assumptions 1, 2, 3 and 4, every asset $i \in \Gamma$ satisfies the one-factor relative asset pricing equation

$$\begin{aligned} \widetilde{r}_{t+1}(i) \equiv & r_{t+1}(i) - \kappa_{t+1} \\ = & \widetilde{\beta}_t(i) m_{t+1} + u_{t+1}(i) \end{aligned} \tag{23}$$

where Relative Beta is defined as

$$\widetilde{\beta}_t(i) \equiv \frac{Cov_t \left[\widetilde{r}_{t+1}(i), m_{t+1} \right]}{Var_t \left[m_{t+1} \right]} = \frac{\boldsymbol{z}_t(i)' \mathbb{E}_t \left[\boldsymbol{\delta}_{t+1} \right]}{\mathbb{E}_t \left[m_{t+1} \right]} \tag{24}$$

and has zero cross-sectional average:

$$\int_\Gamma \widetilde{\beta}_t(i) dG(i) = 0 \tag{25}$$

Moreover the idiosyncratic errors $u_{t+1}(i)$ has zero cross-sectional average:

$$\int_\Gamma u_{t+1}(i) dG(i) = 0 \tag{26}$$

and time-series strict exogeneity:

$$\mathbb{E}_t \left[u_{t+1}(i) \right] = \mathbb{E}_t \left[u_{t+1}(i) | m_{t+1} \right] = 0 \tag{27}$$

Proposition 4 shows that in the presence of cross-sectional predictability, the assets' relative return has a 1-factor pricing structure. This is consistent with Roll (1977)'s result that a mean-variance efficient return automatically prices assets. Here I present the exact condition to construct the mean-variance efficient portfolio from a set of factor portfolios, instead of having to construct from the entire asset universe. A problem that is often impossible due to the curse of dimensionality.

The proposition also implies that one could test the cross-sectional asset pricing models in relative return form, rather than excess return form. It says that the market return, in the perspective of cross-sectional asset pricing within the equity market, is nothing but a intercept that shifts the level of returns. Although, the model does not say whether all stocks have a market beta of 1, since market beta could be one of the time t available characteristics that predict returns cross-sectionally. However, if one is able to obtain the complete set of cross-sectional predictive characteristics, i.e. market beta's predictive power diminish in the presence of all predictable characteristics, market becomes a mere level shift for all asset returns. Subtracting market return would not cause any effect to the cross-sectional asset pricing, but will provide a convenience that the zero-beta rate is gone.

5.3 A Recursive View of Asset Prices

Proposition 4 is simple and intuitive, that for the entire universe of available assets, one only need to find the efficient excess return portfolio, it would price all assets available. Certainly, for any specific class of asset, the global efficient excess return portfolio will price the assets for that class. However, the recursive nature of the relative asset pricing model offers an insight: To price a class of asset, one does not need the globally efficient excess return, an efficient excess return derived within the class suffices. Intuitively, one can just pretend that assets out of the class do not exist, the pricing derivation will survive as long as there are still predictabilities within the class. To proceed, I assume:

Assumption 5. ***Non-Degenerate Class*** C
For a class of assets $i \in C \subsetneq \Gamma$*, the cross-sectional weight* $G^c > 0$ *and the symmetric matrix* $\boldsymbol{S}_t^c$ *has rank* $L \geq 1$ *P-a.s, for* $t = 1, 2, ...$ *where*

$$\begin{array}{lrl} \textit{Group Cross-Sectional Weight:} & G^c = & \int_C dG(i) \\ \textit{Group Average Return:} & \kappa_{t+1}^c \equiv & \int_C r_{t+1}(i) dG(i)/G^c \\ \textit{Group Centered Features:} & \overline{\boldsymbol{z}}_t^c(i) \equiv & \boldsymbol{z}_t(i) - \int_C \boldsymbol{z}_t(i) dG(i)/G^c \\ \textit{Group Dispersion of Features:} & \overline{\boldsymbol{S}}_t^c \equiv & \int_C \overline{\boldsymbol{z}}_t^c(i) \overline{\boldsymbol{z}}_t^c(i)' dG(i)/G^c \end{array} \tag{28}$$

Using the above definition, Equation (9) can thus be re-written for all assets in the class $i \in C$:

$$r_{t+1}(i) = \kappa_{t+1}^{c} + \overline{\boldsymbol{z}}_t^c(i)'\boldsymbol{\delta}_{t+1} + e_{t+1}(i) \tag{29}$$

Notice that Equation (29) satisfies Assumptions 1 and Assumption 2 part (1). To derive result similar to Proposition 4, we need to ensure no redundancy in the features pertaining to this asset class. However, it is possible that due to the restricted class C, some variations among the features in the vector $\boldsymbol{z}_t(i)$ had collapsed, making the symmetric matrix $\overline{\boldsymbol{S}}_t^c$ no-longer positive definite. For example, one can construct C such that one component of the features vector $\boldsymbol{z}_t(i)$ is constant (e.g. assets of a specific size, if size is a predictive feature), causing the re-defined features $\boldsymbol{z}_t^c(i)$ having a zero component. Therefore the goal is to select a cross-sectionally linearly independent subset $\boldsymbol{z}_t^c(i)$ among the features, as well as their corresponding predictabilities $\boldsymbol{\delta}_{t+1}^c$ that could equivalently represent all asset returns in this class in the following way:

$$r_{t+1}(i) = \kappa_{t+1}^{c} + \boldsymbol{z}_t^c(i)'\boldsymbol{\delta}_{t+1}^c + e_{t+1}(i) \text{ for all } i \in \Gamma \tag{30}$$

The following proposition selects such subset of $\boldsymbol{z}_t^c(i)$ and $\boldsymbol{\delta}_{t+1}^c$

Proposition 5. ***Selecting Relevant Features***
If the collection of variables

$$\{r_{t+1}(i), \kappa_{t+1}, \boldsymbol{z}_t(i), \boldsymbol{S}_t, \boldsymbol{\delta}_{t+1}, e_{t+1}(i)\}$$

satisfy Assumptions 1, 2 3 and 4 for all $i \in \Gamma$. Moreover, if for a class of assets $C \subsetneq \Gamma$, Assumption 5 is also satisfied, then there exist a $L \times K$ matrix $\boldsymbol{B}_t^c$ with rank L known at time t, such that:
The collection of variables

$$\left\{r_{t+1}(i), \kappa_{t+1}^c, \boldsymbol{z}_t^c(i), \boldsymbol{S}_t^c, \boldsymbol{\delta}_{t+1}^c, e_{t+1}(i)\right\} \tag{31}$$

satisfy Assumptions 1, 2 3 and 4 for all $i \in C$ where κ_{t+1}^c is defined in Equation (28), and

$$\boldsymbol{z}_t^c(i) = \boldsymbol{B}_t^c \overline{\boldsymbol{z}}_t^c(i) \tag{32}$$

$$\boldsymbol{\delta}_{t+1}^c = \boldsymbol{B}_t^c \boldsymbol{\delta}_{t+1} \tag{33}$$

$$\boldsymbol{S}_t^c = \int_C \boldsymbol{z}_t^c(i)\boldsymbol{z}_t^c(i)'dG(i) \tag{34}$$

Then we can form the long-short portfolios, and excess return portfolio efficient among all assets in class C, using only assets of this class C:

$$\boldsymbol{x}_{t+1}^{c} = \int_{C} \boldsymbol{z}_{t}^{c}(i) r_{t+1}(i) dG(i) = \boldsymbol{S}_{t}^{c} \boldsymbol{\delta}_{t+1}^{c} \tag{35}$$

$$m_{t+1}^{c} = \left(\boldsymbol{x}_{t+1}^{c}\right)' Var_{t}\left[\boldsymbol{x}_{t+1}^{c}\right]^{-1} \mathbb{E}_{t}\left[\boldsymbol{x}_{t+1}^{c}\right] = \boldsymbol{\delta}_{t+1}^{c}\left(\boldsymbol{D}_{t}^{c}\right)^{-1} \mathbb{E}_{t}\left[(\boldsymbol{\delta}_{t+1}^{c})'\right] \tag{36}$$

Proposition 6. ***Recursive Relative Pricing:***
Given Assumptions 1, 2, 3, and 4 for an asset class, $C \subsetneq \Gamma$, satisfies Assumption 5, then for every asset $i \in C$, the recursive one-factor relative asset pricing relation holds:

$$\begin{aligned} \widetilde{r}_{t+1}^{c}(i) \equiv & r_{t+1}(i) - \kappa_{t+1}^{c} \\ = & \widetilde{\beta}_{t}^{c}(i) m_{t+1}^{c} + u_{t+1}^{c}(i) \end{aligned} \tag{37}$$

where Recursive Relative Beta is defined as

$$\widetilde{\beta}_{t}^{c}(i) \equiv \frac{Cov_{t}\left[\widetilde{r}_{t+1}^{c}(i), m_{t+1}^{c}\right]}{Var_{t}\left[m_{t+1}^{c}\right]} = \frac{\boldsymbol{z}_{t}^{c}(i)' \mathbb{E}_{t}\left[\boldsymbol{\delta}_{t+1}^{c}\right]}{\mathbb{E}_{t}\left[m_{t+1}^{c}\right]} \tag{38}$$

and has zero cross-sectional average:

$$\int_{C} \widetilde{\beta}_{t}^{c}(i) dG(i) = 0 \tag{39}$$

Moreover the idiosyncratic errors $u_{t+1}^{c}(i)$ has zero cross-sectional average:

$$\int_{C} u_{t+1}^{c}(i) dG(i) = 0 \tag{40}$$

and time-series strict exogeneity:

$$\mathbb{E}_{t}\left[u_{t+1}^{c}(i)\right] = \mathbb{E}_{t}\left[u_{t+1}^{c}(i) | m_{t+1}^{c}\right] = 0 \tag{41}$$

5.4 The Implied Stochastic Discount Factor

Assumption 6. ***Existence of Stochastic Discount Factor***
There exists an investable unit cost asset at time t with payoff d_{t+1}^{*} and net return $\kappa_{t+1}^{*} \equiv d_{t+1}^{*} - 1$, such that for any asset $i \in \Gamma$ satisfies

$$\mathbb{E}_{t}\left[r_{t+1}(i) d_{t+1}^{*}\right] = \mathbb{E}_{t}\left[\left(d_{t+1}^{*}\right)^{2}\right] - \mathbb{E}_{t}\left[d_{t+1}^{*}\right] \tag{42}$$

or equivalently

$$\mathbb{E}_t\left\{\left[r_{t+1}(i)+1\right]\frac{d^*_{t+1}}{\mathbb{E}_t\left[\left(d^*_{t+1}\right)^2\right]}\right\}=1 \tag{43}$$

Assumption 6 is directly implied if the underlying economy satisfies *Law of One Price*, or linear pricing. See Theorem 2.1 of Hansen and Richard (1987) for derivation. Moreover, when the stochastic discount factor payoff (gross return) d^*_{t+1} is positive almost surely, then there is no arbitrage in the economy.

The following proposition helps relate the efficient portfolio m_{t+1} with existing literature. The efficient excess return is closely related to the orthogonal decomposition of Hansen and Richard (1987):

Proposition 7. ***Hansen and Richard (1987) Decomposition***
Given Assumptions 1, 2, 3, 4 and 6, every asset $i \in \Gamma$ satisfy the decomposition:

$$r_{t+1}(i)=\kappa^*_{t+1}+\left[\beta_t(i)-\beta^*_t\right]m_{t+1}+u^*_{t+1}(i) \tag{44}$$

where

$$\beta_t(i)\equiv\frac{Cov_t\left[r_{t+1}(i),m_{t+1}\right]}{Var_t\left[m_{t+1}\right]} \tag{45}$$

$$\beta^*_t\equiv\frac{Cov_t\left[\kappa^*_{t+1},m_{t+1}\right]}{Var_t\left[m_{t+1}\right]} \tag{46}$$

Moreover the three components satisfies:

$$\mathbb{E}_t\left[u^*_{t+1}(i)\right]=\mathbb{E}_t\left[u^*_{t+1}(i)m_{t+1}\right]=\mathbb{E}_t\left[u^*_{t+1}(i)\kappa^*_{t+1}\right]=\mathbb{E}_t\left[m_{t+1}\left(\kappa^*_{t+1}+1\right)\right]=0 \tag{47}$$

Assumption 7. ***Existence of Risk-Free Rate***
There exists an investable asset with return r^f_t known at time t, i.e. $\exists\, i^f \in \Gamma$ such that

$$r_{t+1}(i^f)=r^f_t$$

Proposition 8. ***Relations between r^f_t, m_{t+1} and κ^*_{t+1}***
Given Assumptions 1,2,3,4,6,7 the following equation holds for any $t=1,2,\ldots$

$$r^f_t=\kappa^*_{t+1}-\beta^*_t m_{t+1} \tag{48}$$

Proposition 9. ***One Factor Absolute Pricing***

Given Assumptions 1,2,3,4,6,7, any asset $i \in \Gamma$ satisfy

$$r_{t+1}(i) = r_t^f + \beta_t(i) m_{t+1} + u_{t+1}^*(i) \tag{49}$$

where

$$\mathbb{E}_t\left[u_{t+1}^*(i)\right] = \mathbb{E}_t\left[m_{t+1} u_{t+1}^*(i)\right] = 0$$

Proposition 10. ***Implied Stochastic Discount Factor***
Given Assumptions 1,2,3,4,6,7 the Stochastic Discount Factor implied by asset prices can be constructed in the following unit cost tradable portfolio

$$\kappa_{t+1}^* = r_t^f - \frac{1 + r_t^f}{1 + \mathbb{E}_t\left[m_{t+1}\right]} m_{t+1} \tag{50}$$

or in gross-return form

$$d_{t+1}^* = 1 + r_t^f - \frac{1 + r_t^f}{\mathbb{E}_t\left[1 + m_{t+1}\right]} m_{t+1} \tag{51}$$

with weight 1 *on the risk-free asset and* $-\frac{1 + r_t^f}{1 + \mathbb{E}_t\left[m_{t+1}\right]}$ *on the zero-cost efficient portfolio* m_{t+1}

5.5 Stochastic Discount Factor and the Conditional Risk-Premium

Given Proposition 10, the conditional expectation of the Stochastic Discount Factor return is

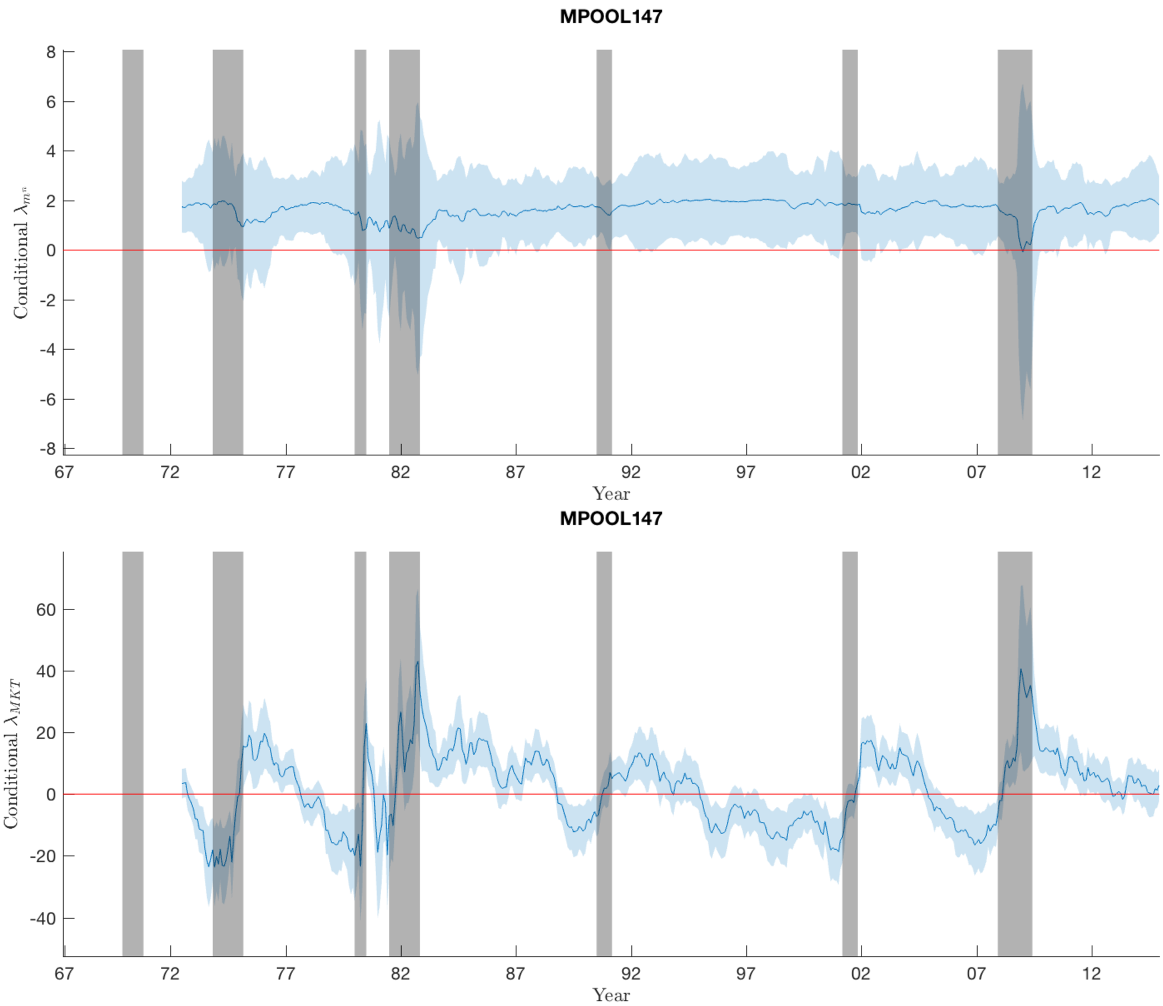
$$\mathbb{E}_t\left[d_{t+1}^*\right] = \frac{1 + r_t^f}{1 + \mathbb{E}_t\left[m_{t+1}\right]} \tag{52}$$

Economic theory suggests that this conditional quantity should be high during bad times and low during good times (counter-cyclical), since an economic agent values future payoff higher during bad times. Given the inverse relationships between $\mathbb{E}_t\left[m_{t+1}\right]$ and $\mathbb{E}_t\left[d_{t+1}^*\right]$, it means that the conditional expectation of the efficient excess return must be strongly *pro-cyclical*. I estimate this conditional quantity to show that it is indeed strongly pro-cyclical. Moreover for comparison I show that the market return cannot be the efficient return leading to the Stochastic Discount Factor, as its conditional expectation is highly *counter-cyclical*.

Since we only have 1 universe, thus 1-observation for each conditional information set, estimating the conditional expectation is difficult. It cannot be obtained via making usual smoothness assumptions, i.e. assuming that $\mathbb{E}_t\left[m_{t+1}\right]$ do not change and equals to the local average. Doing so is equivalent to making the strong assumption that the conditional information set do not change. Instead, I propose another way around, by going into the cross-sectional dimension. A priced risk-factor should have its conditional expectation equals to its conditional risk-premium in the cross-section. Even though we have only 1 observation in the time-series dimension, we have an abundance of observations in the cross-section given any point in time, allowing us to obtain a reliable estimate of the conditional risk-premium. Therefore, the quantity $\mathbb{E}_t[m_{t+1}]$ has the special interpretation that it is the conditional risk-premium of the efficient factor. Given the large cross-section, we are able to estimate it using the most recent econometric tools.

I employ the recent development of Gagliardini, Ossola, and Scaillet (2016) to achieve this goal. The appendix provides details on the variable specification including test assets, common and firm-specific instruments. Figure 7 plots the estimated

Figure 7: **Conditional Risk Premium of m^n and Mkt-Rf**



conditional risk-premium of the efficient factor. The conditional λ_{t,m^n} is positive over time, consistent with the previously documented evidence that its un-conditional risk-premium λ_{m^n} is significantly positive. Moreover, it exhibits strongly **pro-cyclical** pattern, dropping significantly during NBER recessions, while remaining at high level during economic expansion.

In stark contrast, the conditional risk-premium of the market is not significantly above or below zero, and is highly **counter-cyclical**, with conditional risk-premium being high during recession while low during expansion. This behavior of the market risk-premium indicates that market return is not the mean-variance efficient return.

A Empirical Details

A.1 Details on Data

Daily and monthly stock return files are obtained from CRSP(Center for Research in Security Prices), using all common stocks (with share code of 10 or 11), listed in NYSE, AMEX or NASDAQ (exchange code 1, 2 or 3), from June 30, 1967 to December 31, 2014, resulting in 23,142 unique securities, and on average about 5,000 securities at a given point in time. To help avoid survivorship bias, I further adjust the individual stock returns for delisting, using the delisting return provided by CRSP as the return after the last trading month. Firm book information is obtained from COMPUSTAT database and merged with CRSP database via the CRSP PERMNO number. The firm-level characteristics (Size, B/M ratio, Momentum, Long-term Reversal, Short-term Reversal, Operating Profitability and Investment) are calculated according to Fama and French (2014). Daily and monthly return series of the 8 factors (Mkt-Rf, Small-Minus-Big, High-Minus-Low, Robust-Minus-Weak, Conservative-Minus-Aggressive, Momentum, Long-term Reversal, Short-term Reversal), 5 sets of 5×5 portfolios (25 Size—Book-to-Market, 25 Size—Momentum, 25 Book-to-Market—Investment, 25 Size—Investment, 25 Size—Operating Profitability), and monthly return series of the 3 sets of 5×5 portfolios (25 Size—Net Income, 25 Size—Total Variance, 25 Size—Residual Variance) are taken from the website of Kenneth French[9]. Default spread (return difference between Aaa and Baa bonds) and term spread (return difference between 10-year U.S. government bond and 3-month Treasury Bill) series are obtained from WRDS (Wharton Research Data Services). Monthly series of liquidity factor and decile portfolios are obtained from Stambaugh's website[10]. Monthly series of Quality-Minus-Junk, Betting-Against-Beta and the Devil factors and respectively their decile portfolios are obtained from Andrea Frazzini's website[11]. Monthly series of q-factors are provided by Lu Zhang. Monthly series of Novy-Marx and Velikov (2016) anomalies are obtained from Robert Novy-Marx's website[12].

A.2 Details on the Efficient Portfolio

A.2.1 Risk-Parity Efficient Allocation

In order to compare directly with the market, I compute a "risk-parity" scaled weight $\widehat{\boldsymbol{w}}_t^p$, so that the resulting portfolio have the same *expected* volatility as the market excess

[9] URL:http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
[10] URL:http://finance.wharton.upenn.edu/~stambaugh/liq_data_1962_2014.txt
[11] URL:http://www.econ.yale.edu/~af227/data_library.htm
[12] URL:http://rnm.simon.rochester.edu/data_lib/index.html

return:

$$\widehat{\boldsymbol{w}}_t^p = \widehat{\boldsymbol{w}}_t^* \cdot \left[\frac{\left(\widehat{\sigma}_t^{mkt}\right)^2}{\left(\widehat{\boldsymbol{w}}_t^*\right)' \widehat{\boldsymbol{\Sigma}}_t^{HF} \widehat{\boldsymbol{w}}_t^*} \frac{det\left(\widehat{\boldsymbol{\Sigma}}_t^{HF}\right)}{det\left(\widehat{\boldsymbol{\Sigma}}_t^{LF}\right)} \right]^{1/2} \tag{53}$$

where $\widehat{\sigma}_t^{mkt}$ is the realized volatility of the market excess return in the same expanding window.

A.2.2 Time-Varying Allocations for 8 Assets

Figure 8: Efficient Portfolio Allocation for 8 Assets

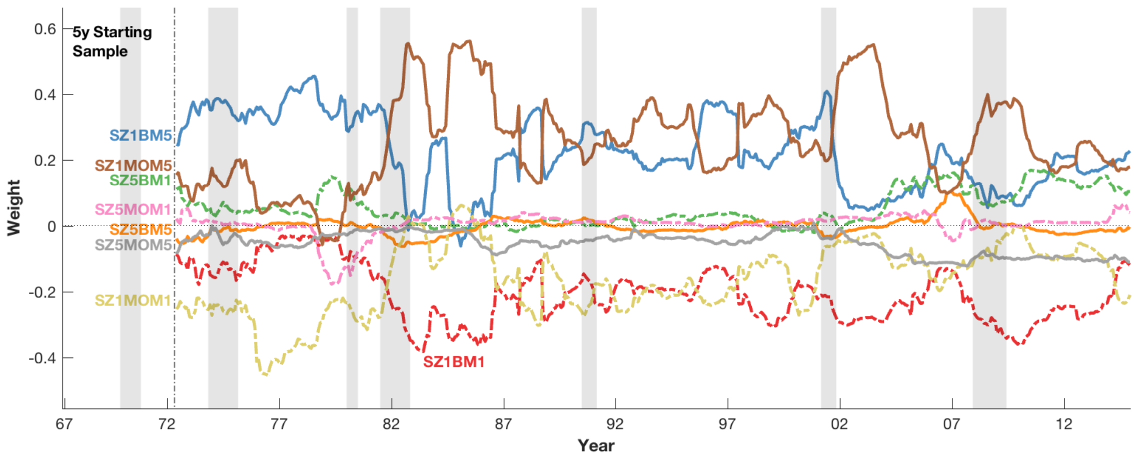


Figure 8 plots the time-series dynamics of the efficient allocation over time. Among smaller stocks, the efficient allocation go short in low value and momentum stocks (SZ1BM1, SZ1MOM1) and go long in high value and momentum stocks (SZ1BM5, SZ1MOM5) while within each pair are highly negatively correlated over time. The reverse is observed for larger stocks, the efficient allocation go long in low value and momentum stocks (SZ5BM1,SZ5MOM1) and short in high value momentum stocks (SZ5BM5, SZ5MOM5);

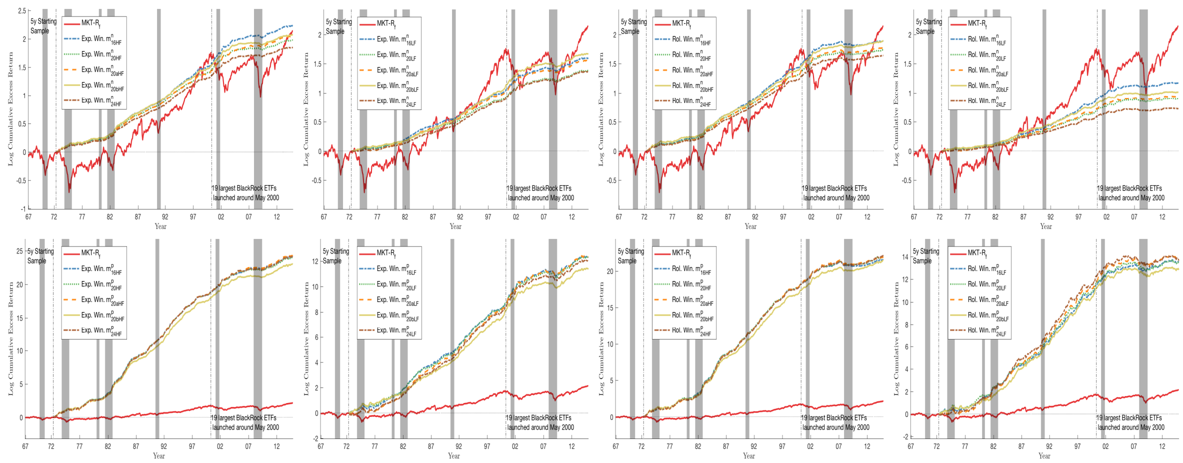
It is evident from the time-series plot that the efficient weights between the pairs (SZ1BM1, SZ1MOM1) and (SZ1BM5, SZ1MOM5) are highly negatively correlated. It is also evident (but harder to see because of the scale) that the pairs (SZ5BM1,SZ5MOM1, green and pink), and (SZ5BM5,SZ5MOM5, orange and gray) are highly negatively correlated.

A.2.3 Sanity Check

Table 1 and Figure 8 suggest that the efficient allocation is concentrated on strong long short positions on the two pairs of asset: (SZ1BM1, SZ5BM1) and (SZ1MOM1, SZ5MOM1). Indeed, the two pairs contribute large part of efficiency gain in terms of Sharpe Ratio for the efficiency portfolio.

A.2.4 Alternative Efficient Portfolios

Figure 9: The Efficient Portfolios: Alternative Construction



A.3 Robustness Checks

A.4 Common Questions

Is it a tautology?

"The Fama-French model is not a tautology, despite the fact that factors and test portfolios are based on the same set of characteristics."

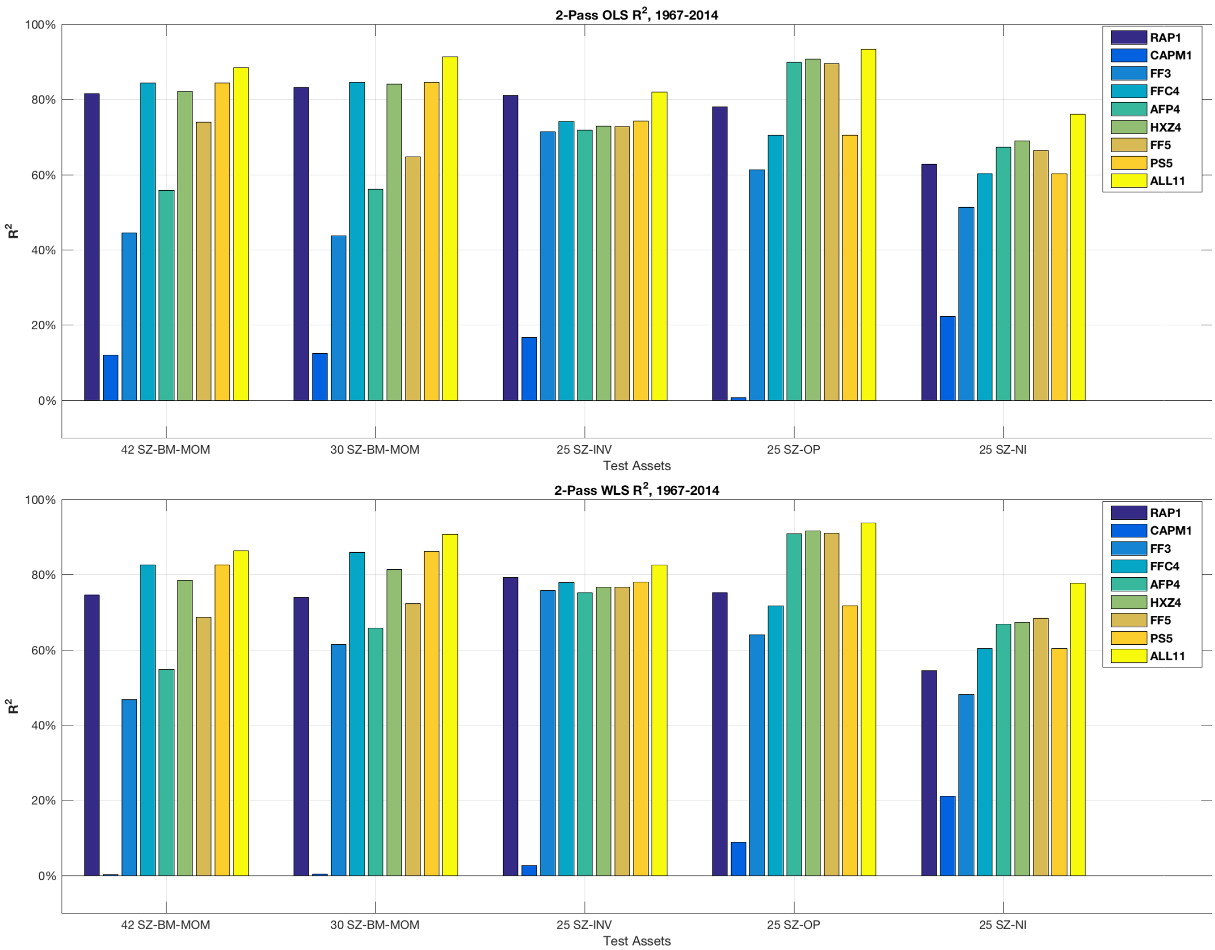
— John Cochrane *Asset Pricing* 2005, p. 442, para. 3

FF3 model(s) take(s) the portfolios that cannot be priced well by CAPM as base assets, then forms 2 new factor(s) from the base assets to price base anomalies and many other new anomalies.

ALL12 - FF3 model(s) take(s) the portfolios that cannot be priced well by FF3 as base assets, then forms 8 new factor(s) from the base assets to price base anomalies and many other new anomalies.

RAP1 model(s) take(s) the portfolios that cannot be priced well by FF3 as base assets, then forms 1 new factor(s) from the base assets to price base anomalies and many other new anomalies.

Figure 10: Non-Miss-Specification-Robust R^2 Across Models and Test Assets



A.5 Conditional Risk-Premium Estimation

B Proofs

Proof of Proposition 1

(a) Assumption 2 $\Rightarrow$ "identification"

The Equation (9) can be written as

$$r_{t+1}(i) = \{1, \boldsymbol{z}_t(i)'\} \times \left\{\kappa_{t+1}, \boldsymbol{\delta}'_{t+1}\right\}' + e_{t+1}(i)$$

Then using Assumption 2, with probability 1, we have

$$\begin{aligned}
\int_\Gamma \{1, \boldsymbol{z}_t(i)'\}' r_{t+1}(i) dG(i) &= \int_\Gamma \{1, \boldsymbol{z}_t(i)'\}' \times \{1, \boldsymbol{z}_t(i)'\} \times \left\{\kappa_{t+1}, \boldsymbol{\delta}'_{t+1}\right\}' dG(i) \\
&= \left\{ \begin{array}{cc} 1 & 0 \\ 0 & \boldsymbol{S}_t \end{array} \right\} \times \left\{\kappa_{t+1}, \boldsymbol{\delta}'_{t+1}\right\}' + \int_\Gamma \{1, \boldsymbol{z}_t(i)'\}' e_{t+1}(i) dG(i) \\
&= \left\{ \begin{array}{cc} 1 & 0 \\ 0 & \boldsymbol{S}_t \end{array} \right\} \times \left\{\kappa_{t+1}, \boldsymbol{\delta}'_{t+1}\right\}'
\end{aligned}$$

Using invertibility of $\boldsymbol{S}_t$, with probability 1, we have

$$\left\{\kappa_{t+1}, \boldsymbol{\delta}'_{t+1}\right\}' = \left\{ \begin{array}{cc} 1 & 0 \\ 0 & \boldsymbol{S}_t^{-1} \end{array} \right\} \times \int_\Gamma \{1, \boldsymbol{z}_t(i)'\}' r_{t+1}(i) dG(i)$$

a unique solution to Equation (9)

(b) Not "Assumption 2 part (1)" $\Rightarrow$ No "identification"

First suppose there exists a weighting scheme $w_t(i) < \infty$ known at time t, such that

$$x_t = \int_\Gamma \left[w_t(i)\right]^2 dG(i) \in (0, \infty)$$

and, there exists a $\Omega^0 \subset \Omega$, with $\mathbb{P}(\Omega^0) > 0$, such that the realizations of the random variables restricted to Ω^0 makes

$$\int_\Gamma w_t(i) e_{t+1}(i) dG(i) = c_{t+1} \neq 0$$

Then define

$$\begin{aligned}
\overline{w}_t &= \int_\Gamma w_t(i) dG(i) \\
\widetilde{w}_t(i) &= w_t(i) - \overline{w}_t \\
y_{t+1} &= \int_\Gamma \widetilde{w}_t(i) e_{t+1}(i) dG(i) \\
\kappa^0_{t+1} &= \overline{w}_t \int_\Gamma e_{t+1}(i) dG(i) \\
\widetilde{e}_{t+1}(i) &= e_{t+1}(i) - \kappa^0_{t+1} - \widetilde{w}_t(i) \frac{y_{t+1}}{x_t} \\
\widetilde{\kappa}_{t+1} &= \kappa_{t+1} + \kappa^0_{t+1} \\
\widetilde{\boldsymbol{\delta}}_{t+1} &= \left\{ \boldsymbol{\delta}'_{t+1}, \frac{y_{t+1}}{x_t} \right\}' \\
\widetilde{\boldsymbol{z}}_{t+1}(i) &= \{\boldsymbol{z}_{t+1}(i)', \widetilde{w}_t(i)\}'
\end{aligned}$$

Notice for $\omega \in \Omega^0$, by construction, $c_{t+1} = \kappa^0_{t+1} + y_{t+1} \neq 0$, thus at least one of κ^0_{t+1}, y_{t+1} not zero for $\omega \in \Omega^0$.

Moreover, for all $\omega \in \Omega$, since $x_t \geq \overline{w}_t^2$ (second moment versus square of first moment), consider four possible cases:

Case 1: $\omega \in \Omega^1 := \left\{\omega \middle| \overline{w}_t^2 = x_t, \omega \in \Omega^0\right\}$,
This means that the weights $w_t(i)$ is almost constant in the cross-section, in this case, $y_{t+1} = 0$ then $\kappa^0_{t+1} = c_{t+1} \neq 0$. The collection of variables

$$\{\widetilde{\kappa}_{t+1}, \boldsymbol{\delta}_{t+1}, \widetilde{e}_{t+1}(i), \boldsymbol{z}_{t+1}(i)\}$$

satisfy all the conditions of the Assumption 1, they generate exactly the same distribution for $r_{t+1}(i)$ and $\boldsymbol{z}_t(i)$, while $\widetilde{\kappa}_{t+1} \neq \kappa_{t+1}$, so not identified.

Case 2: $\omega \in \Omega^2 := \left\{\omega \middle| \overline{w}_t^2 < x_t, y_{t+1} = 0, \omega \in \Omega^0\right\}$
$y_{t+1} = 0$ implies that $\kappa^0_{t+1} = c_{t+1} \neq 0$, thus the collection of variables

$$\{\widetilde{\kappa}_{t+1}, \boldsymbol{\delta}_{t+1}, \widetilde{e}_{t+1}(i), \boldsymbol{z}_{t+1}(i)\}$$

satisfy all the conditions of the Assumption 1, they generate exactly the same distribution for $r_{t+1}(i)$ and $\boldsymbol{z}_t(i)$ as the original variables, while $\widetilde{\kappa}_{t+1} \neq \kappa_{t+1}$, so not identified.

Case 3: $\omega \in \Omega^3 := \left\{\omega \middle| \overline{w}_t^2 < x_t, \kappa^0_{t+1} = 0, \omega \in \Omega^0\right\}$
This means that $\widetilde{w}_t(i)$ satisfies $\int_\Gamma [w_t(i)]^2 \, dG(i) = x_t - \overline{w}_t^2 \in (0, \infty)$.
Also $\kappa^0_{t+1} = 0$ implies that $y_{t+1} \neq 0$, then the collection of variables

$$\left\{\widetilde{\kappa}_{t+1}, \widetilde{\boldsymbol{\delta}}_{t+1}, \widetilde{e}_{t+1}(i), \widetilde{\boldsymbol{z}}_{t+1}(i)\right\}$$

satisfy all conditions of the Assumption 1, and the new additional feature $\widetilde{w}_t(i)$ is relevant with its corresponding component in vector $\widetilde{\boldsymbol{\delta}}_{t+1}$, i.e. $\dfrac{y_{t+1}}{x_t}$ non-zero. They generate exactly the same distribution for $r_{t+1}(i)$ and $\widetilde{\boldsymbol{z}}_t(i)$ as the original variables while $\widetilde{\boldsymbol{\delta}}_{t+1} \neq \boldsymbol{\delta}_{t+1}$, so not identified.

Case 4: $\omega \in \Omega^4 := \left\{\omega \middle| \overline{w}_t^2 < x_t, \kappa^0_{t+1} \neq 0, y_{t+1} \neq 0, \omega \in \Omega^0\right\}$
The collection of variables

$$\left\{\widetilde{\kappa}_{t+1}, \widetilde{\boldsymbol{\delta}}_{t+1}, \widetilde{e}_{t+1}(i), \widetilde{\boldsymbol{z}}_{t+1}(i)\right\}$$

satisfy all conditions of the Assumption 1, and the new additional feature $\widetilde{w}_t(i)$ is relevant with its corresponding component in vector $\widetilde{\boldsymbol{\delta}}_{t+1}$, i.e. $\dfrac{y_{t+1}}{x_t}$ non-zero. They generate exactly the same distribution for $r_{t+1}(i)$ and $\widetilde{\boldsymbol{z}}_t(i)$ as the original variables while $\widetilde{\boldsymbol{\delta}}_{t+1} \neq \boldsymbol{\delta}_{t+1}$ and $\widetilde{\kappa}_{t+1} \neq \kappa_{t+1}$, so not identified.

Since $\Omega^0 = \Omega^1 \cup \Omega^2 \cup \Omega^3 \cup \Omega^4$, and $\mathbb{P}(\Omega^0) > 0$, for all $\omega \in \Omega^0$, it is alway possible to redefine variables that are different from original parameters such that the resulting distributions are observationally equivalent, thus Equation (9) is not identified.

(c) Not "Assumption 2 part (2)" $\Rightarrow$ No "identification"
Suppose there exist $\Omega^0 \subset \Omega$, with $\mathbb{P}(\Omega^0) > 0$, for $\omega \in \Omega^0$, $\boldsymbol{S}_t$ is singular.
Let $L = rank(\boldsymbol{S}_t)$, we have $1 \leq L < K$. Since $\boldsymbol{S}_t$ is symmetric positive semi-definite real matrix, there exists an orthonormal matrix $\boldsymbol{Q}_t$ and a diagonal matrix $\boldsymbol{\Lambda}_t$, such that $\boldsymbol{S}_t$ have the following singular value decomposition

$$\boldsymbol{S}_t = \boldsymbol{Q}_t \boldsymbol{\Lambda}_t \boldsymbol{Q}_t'$$

By construction, $\boldsymbol{\Lambda}_t$ has rank L and thus L non-zero entries on the diagonal (and in the whole matrix), and $N = K - L$ zero entries on the diagonal. Let their locations be: Non-zero entries: $\{l_1, l_2, ..., l_L\}$, zero-entries $\{i_1, i_2, ..., i_N\}$.
Define the selection matrices:

$$\boldsymbol{J}_L = \begin{pmatrix} \cdots & \overset{l_1\text{-th col.}}{1} & 0 & 0 & 0 & 0 & 0 & \cdots \\ \cdots & 0 & \overset{l_2\text{-th col.}}{1} & 0 & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & 0 & 0 & 0 & 0 & \overset{l_L\text{-th col.}}{1} & 0 & \cdots \end{pmatrix}_{L \times K}$$

$$\boldsymbol{J}_N = \begin{pmatrix} \cdots & \overset{i_1\text{-th col.}}{1} & 0 & 0 & 0 & 0 & 0 & \cdots \\ \cdots & 0 & \overset{i_2\text{-th col.}}{1} & 0 & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & 0 & 0 & 0 & 0 & \overset{i_N\text{-th col.}}{1} & 0 & \cdots \end{pmatrix}_{L \times N}$$

By construction, $\boldsymbol{J}_L \boldsymbol{\Lambda}_t \boldsymbol{J}_L'$ selects non-zero entries of $\boldsymbol{\Lambda}_t$ and produce a $L \times L$ full rank diagonal $\boldsymbol{S}_L$ matrix, and $\boldsymbol{J}_N \boldsymbol{\Lambda}_t \boldsymbol{J}_N'$ produce a $N \times N$ null matrix $\boldsymbol{S}_N$ with all entries equal to zero. Moreover the identity matrix $\boldsymbol{I}_K$ can be decomposed as

$$\boldsymbol{I}_K = \boldsymbol{J}_L' \boldsymbol{J}_L + \boldsymbol{J}_N' \boldsymbol{J}_N$$

The $K \times K$ matrix $\boldsymbol{Q}_t \boldsymbol{J}_N' \boldsymbol{J}_N \boldsymbol{Q}_t'$ has rank $N > 0$, therefore there exists a $K \times 1$ vector $\boldsymbol{\eta}_t$, such that the $K \times 1$ vector $\boldsymbol{\theta}_t := \boldsymbol{Q}_t \boldsymbol{J}_N' \boldsymbol{J}_N \boldsymbol{Q}_t' \boldsymbol{\eta}_t \neq \boldsymbol{0}_{K \times 1}$

Now consider the following scaler:

$$a_t(i) = \boldsymbol{z}_t(i)'\boldsymbol{\theta}_t = \boldsymbol{z}_t(i)'\boldsymbol{Q}_t\boldsymbol{J}_N'\boldsymbol{J}_N\boldsymbol{Q}_t'\boldsymbol{\eta}_t$$

This scaler is zero, to see this

$$\begin{aligned}
\int_\Gamma a_t(i)^2 dG(i) = & \int_\Gamma \boldsymbol{Q}_t\boldsymbol{J}_N'\boldsymbol{J}_N\boldsymbol{Q}_t'\boldsymbol{z}_t(i)\boldsymbol{z}_t(i)'\boldsymbol{Q}_t\boldsymbol{J}_N'\boldsymbol{J}_N\boldsymbol{Q}_t' dG(i) \\
= & \boldsymbol{Q}_t\boldsymbol{J}_N'\boldsymbol{J}_N\boldsymbol{Q}_t' \left[\int_\Gamma \boldsymbol{z}_t(i)\boldsymbol{z}_t(i)' dG(i)\right] \boldsymbol{Q}_t\boldsymbol{J}_N'\boldsymbol{J}_N\boldsymbol{Q}_t' \\
= & \boldsymbol{Q}_t\boldsymbol{J}_N'\boldsymbol{J}_N\boldsymbol{Q}_t'\boldsymbol{S}_t\boldsymbol{Q}_t\boldsymbol{J}_N'\boldsymbol{J}_N\boldsymbol{Q}_t' \\
= & \boldsymbol{Q}_t\boldsymbol{J}_N'\boldsymbol{J}_N\boldsymbol{Q}_t'\boldsymbol{Q}_t\boldsymbol{\Lambda}_t\boldsymbol{Q}_t'\boldsymbol{Q}_t\boldsymbol{J}_N'\boldsymbol{J}_N\boldsymbol{Q}_t' \\
= & \boldsymbol{Q}_t\boldsymbol{J}_N'\boldsymbol{J}_N\boldsymbol{\Lambda}_t\boldsymbol{J}_N'\boldsymbol{J}_N\boldsymbol{Q}_t' \\
= & 0
\end{aligned}$$

The last identity comes from the fact that $\boldsymbol{J}_N\boldsymbol{\Lambda}_t\boldsymbol{J}_N' = \boldsymbol{0}_{N\times N}$. Thus, it must be that $a_t(i) = \boldsymbol{z}_t(i)'\boldsymbol{\theta}_t = 0$, even though $\boldsymbol{\theta}_t \neq \boldsymbol{0}_{K\times 1}$ Thus we have showed that the re-defined vector $\widetilde{\boldsymbol{\delta}}_{t+1} = \boldsymbol{\delta}_{t+1} + \boldsymbol{\theta}_t \neq \boldsymbol{\delta}_{t+1}$ satisfies Equation (9) and generate exactly the same distribution as the original parameters. Thus Equation (9) is not identified with probability $\mathbb{P}(\Omega^0) > 0$

□

Proof of Proposition 2
First, I show the equivalence of Problems 1 and 2, then derive the solution.

The objective function of Problem 1 is

$$\mathcal{L}_1(y_1, \boldsymbol{w}) = Var_t\left[\boldsymbol{w}'\boldsymbol{x}_{t+1}\right] \quad \text{s.t. } \mathbb{E}_t\left[\boldsymbol{w}'\boldsymbol{x}_{t+1}\right] = y_1$$

The objective function of Problem 2 is

$$\begin{aligned}
\mathcal{L}_2(y_2, \boldsymbol{w}) = & y_2^2 - 2y_2\mathbb{E}_t\left[\boldsymbol{w}'\boldsymbol{x}_{t+1}\right] + \mathbb{E}_t\left[\left(\boldsymbol{w}'\boldsymbol{x}_{t+1}\right)^2\right] \\
= & \left\{y_2 - \mathbb{E}_t\left[\boldsymbol{w}'\boldsymbol{x}_{t+1}\right]\right\}^2 + Var_t\left[\boldsymbol{w}'\boldsymbol{x}_{t+1}\right]
\end{aligned}$$

Thus

$$\mathcal{L}_2(y_1, \boldsymbol{w}) = \mathcal{L}_1(y_1, \boldsymbol{w})$$

Next, I derive the solution.

$$\mathcal{L}_1(y, \boldsymbol{w}) = \boldsymbol{w}'\boldsymbol{\Sigma}_t\boldsymbol{w} \quad \text{s.t.} \boldsymbol{w}'\boldsymbol{\mu}_t = y$$

Using Equation 14, we have $\mathbf{\Sigma}_t = Var_t(\boldsymbol{p}_{t+1}) = \boldsymbol{S}_t\boldsymbol{D}_t\boldsymbol{S}_t$, and by Assumption 3, we have that

$$\frac{\partial^2}{(\partial \boldsymbol{w})(\partial \boldsymbol{w}')}\boldsymbol{w}'\mathbf{\Sigma}_t\boldsymbol{w} = \mathbf{\Sigma}_t = \boldsymbol{S}_t\boldsymbol{D}_t\boldsymbol{S}_t$$

is positive definite. Therefore the objective function in $\mathcal{L}_1(y)$ is globally convex, continuous and has a global minimum.

Now consider the Lagrangian

$$\mathcal{L}(\boldsymbol{w}, \lambda) = \boldsymbol{w}'\mathbf{\Sigma}_t\boldsymbol{w} + \lambda(\boldsymbol{w}'\boldsymbol{\mu}_t - y)$$

The first order conditions are:

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}} &: \quad 2\mathbf{\Sigma}_t\boldsymbol{w} + \lambda\boldsymbol{\mu}_t = 0\\ \frac{\partial \mathcal{L}}{\partial \lambda} &: \quad \boldsymbol{w}'\boldsymbol{\mu}_t - y = 0\end{aligned}$$

Solving for λ first and then for $\boldsymbol{w}$ to arrive at the solution

$$\begin{aligned}\lambda =& \frac{-2y}{\boldsymbol{\mu}_t'\mathbf{\Sigma}_t^{-1}\boldsymbol{\mu}_t}\\ \boldsymbol{w} =& \frac{y}{\boldsymbol{\mu}_t'\mathbf{\Sigma}_t^{-1}\boldsymbol{\mu}_t}\mathbf{\Sigma}_t^{-1}\boldsymbol{\mu}_t = c\boldsymbol{w}_t^*\end{aligned}$$

where the scaling constant $c = \dfrac{y}{\boldsymbol{\mu}_t'\mathbf{\Sigma}_t^{-1}\boldsymbol{\mu}_t}$ □

Proof of Proposition 3

Consider any excess return r^e_{t+1} formed at time t with weights $w_t(i)$ for asset i

$$r^e_{t+1} = \int_\Gamma w_t(i)r_{t+1}(i)dG(i)$$

For it to be excess return it must have zero cost, thus

$$\int_\Gamma w_t(i)dG(i) = 0$$

Using Equation (9), we have

$$\begin{aligned}r^e_{t+1} =& \int_\Gamma w_t(i)\kappa_{t+1}dG(i) + \int_\Gamma w_t(i)\boldsymbol{z}_t(i)'\boldsymbol{\delta}_{t+1}dG(i) + \int_\Gamma w_t(i)e_{t+1}(i)dG(i)\\ =& 0 + \left[\int_\Gamma w_t(i)\boldsymbol{z}_t(i)'dG(i)\right]\boldsymbol{\delta}_{t+1} + \int_\Gamma w_t(i)e_{t+1}(i)dG(i)\end{aligned}$$

Using Equation 14, we have

$$r^e_{t+1} = \boldsymbol{\omega}_t' \boldsymbol{x}_{t+1} + v_{t+1}$$

where

$$\begin{aligned}\boldsymbol{\omega}_t' &= \left[\int_\Gamma w_t(i) \boldsymbol{z}_t(i)' dG(i)\right] \boldsymbol{S}_t^{-1} \\ v_{t+1} &= \int_\Gamma w_t(i) e_{t+1}(i) dG(i)\end{aligned}$$

and

$$\mathbb{E}_t \left[v_{t+1} | \boldsymbol{x}_{t+1}\right] = \int_\Gamma w_t(i) \mathbb{E}_t \left[e_{t+1}(i) | \boldsymbol{S}_t \boldsymbol{\delta}_{t+1}\right] dG(i) = 0$$

by the conditional exogeneity in Assumption 4.

Further since m_{t+1} is efficient for all portfolios formed among $\boldsymbol{x}_{t+1}$, m_{t+1} is therefore efficient for all such r^e_{t+1} □

Proof of Proposition 4

First I show the relationships among expectations:

$$\mathbb{E}_t \left[r_{t+1}(i) - \kappa_{t+1}\right] = \widetilde{\beta}_t(i) \mathbb{E}_t \left[m_{t+1}\right] \quad \forall i \in [0, 1]$$

By direct algebraic manipulation, using Equation 14 and Equation (9)

$$\begin{aligned} m_{t+1} =& \boldsymbol{x}_{t+1}' \boldsymbol{\Sigma}_t^{-1} \boldsymbol{\mu}_t \\ =& \boldsymbol{\delta}_{t+1}' \boldsymbol{S}_t \left(\boldsymbol{S}_t \boldsymbol{D}_t \boldsymbol{S}_t'\right)^{-1} \boldsymbol{S}_t \mathbb{E}_t \left[\boldsymbol{\delta}_{t+1}\right] \\ =& \boldsymbol{\delta}_{t+1}{\boldsymbol{'}} \boldsymbol{D}_t^{-1} \mathbb{E}_t \left[\boldsymbol{\delta}_{t+1}\right]\end{aligned}$$

$$\begin{aligned} Cov_t \left[r_{t+1}(i), m_{t+1}\right] =& Cov_t \left[\kappa_{t+1}, m_{t+1}\right] \\ &+ \boldsymbol{z}_t(i)' \mathbb{E}_t \left[\boldsymbol{\delta}_{t+1} \boldsymbol{\delta}_{t+1}'\right] \boldsymbol{D}_t^{-1} \mathbb{E}_t \left[\boldsymbol{\delta}_{t+1}\right] \\ &- \boldsymbol{z}_t(i)' \mathbb{E}_t \left[\boldsymbol{\delta}_{t+1}\right] \mathbb{E}_t \left[\boldsymbol{\delta}_{t+1}'\right] \boldsymbol{D}_t^{-1} \mathbb{E}_t \left[\boldsymbol{\delta}_{t+1}\right] \\ &+ Cov_t \left[e_{t+1}(i), m_{t+1}\right] \\ =& Cov_t \left[\kappa_{t+1}, m_{t+1}\right] + \boldsymbol{z_t}(i)' \mathbb{E}_t \left[\boldsymbol{\delta}_{t+1}\right]\end{aligned}$$

$$\begin{aligned} Var_t \left[m_{t+1}\right] =& \mathbb{E}_t \left[m_{t+1}' m_{t+1}\right] - \mathbb{E}_t \left[m_{t+1}'\right] \mathbb{E}_t \left[m_{t+1}\right] \\ =& \mathbb{E}_t \left[\boldsymbol{\delta}_{t+1}'\right] \boldsymbol{D}_t^{-1} \mathbb{E}_t \left[\boldsymbol{\delta}_{t+1}\right] \\ =& \mathbb{E}_t \left[m_{t+1}\right]\end{aligned}$$

Therefore, we have

$$
\begin{aligned}
\mathbb{E}_t\left[r_{t+1}(i)\right] &= \mathbb{E}_t\left[\kappa_{t+1}\right] + \boldsymbol{z_t}(i)'\mathbb{E}_t\left[\boldsymbol{\delta}_{t+1}\right] \\
&= \mathbb{E}_t\left[\kappa_{t+1}\right] + \frac{Cov_t\left[r_{t+1}(i), m_{t+1}\right]}{Var_t\left[m_{t+1}\right]}\mathbb{E}_t\left[m_{t+1}\right] - \frac{Cov_t\left[\kappa_{t+1}, m_{t+1}\right]}{Var_t\left[m_{t+1}\right]}\mathbb{E}_t\left[m_{t+1}\right] \\
&= \mathbb{E}_t\left[\kappa_{t+1}\right] + \widetilde{\beta}_t(i)\mathbb{E}_t\left[m_{t+1}\right]
\end{aligned}
$$

where

$$
\widetilde{\beta}_t(i) \equiv \frac{Cov_t\left[\widetilde{r}_{t+1}(i), m_{t+1}\right]}{Var_t\left[m_{t+1}\right]}
$$

and

$$
\int_\Gamma \widetilde{\beta}_t(i)dG(i) = 0
$$

follows immediately.

Now define $u_{t+1}(i) \equiv r_{t+1}(i) - \kappa_{t+1} - \widetilde{\beta}_t(i)m_{t+1}$, we have $\mathbb{E}_t\left[u_{t+1}(i)\right] = 0$ by definition.

Moreover,

$$
\mathbb{E}_t\left[u_{t+1}|m_{t+1}\right] = \mathbb{E}_t\left[\widetilde{r}_{t+1}(i)|m_{t+1}\right] - \mathbb{E}_t\left\{\widetilde{\beta}_t(i)m_{t+1}|m_{t+1}\right\}
$$

Using Assumption 1, the law of iterated expectation, and the fact that $m_{t+1} = \boldsymbol{\delta}_{t+1}'\boldsymbol{D}_t^{-1}\mathbb{E}_t\left[\boldsymbol{\delta}_{t+1}\right]$ we have

$$
\begin{aligned}
\mathbb{E}_t\left[\widetilde{r}_{t+1}(i) \mid m_{t+1}\right] =& \mathbb{E}_t\left[\boldsymbol{z}_t(i)'\boldsymbol{\delta}_{t+1} + e_{t+1}(i) \mid m_{t+1}\right] \\
=& \mathbb{E}_t\left\{\mathbb{E}_t\left[\boldsymbol{z}_t(i)'\boldsymbol{\delta}_{t+1} + e_{t+1}(i) \mid \boldsymbol{\delta}_{t+1}\right] \mid m_{t+1}\right\} \\
=& \boldsymbol{z}_t(i)'\mathbb{E}_t\left[\boldsymbol{\delta}_{t+1} \mid m_{t+1}\right]
\end{aligned}
$$

and by definition

$$
\begin{aligned}
\mathbb{E}_t\left\{\widetilde{\beta}_t(i)m_{t+1} \mid m_{t+1}\right\} =& \mathbb{E}_t\left\{\frac{Cov_t\left[\widetilde{r}_{t+1}(i), m_{t+1}\right]}{Var_t\left[m_{t+1}\right]}m_{t+1} \mid m_{t+1}\right\} \\
=& \mathbb{E}_t\left\{\frac{\boldsymbol{z}_t(i)'\mathbb{E}_t\left[\boldsymbol{\delta}_{t+1}\right]}{\mathbb{E}_t\left[\boldsymbol{\delta}_{t+1}'\right]\boldsymbol{D}_t^{-1}\mathbb{E}_t\left[\boldsymbol{\delta}_{t+1}\right]}\boldsymbol{\delta}_{t+1}'\boldsymbol{D}_t^{-1}\mathbb{E}_t\left[\boldsymbol{\delta}_{t+1}\right] \mid m_{t+1}\right\} \\
=& \mathbb{E}_t\left\{\mathbb{E}_t\left\{\frac{\boldsymbol{z}_t(i)'\mathbb{E}_t\left[\boldsymbol{\delta}_{t+1}\right]}{\mathbb{E}_t\left[\boldsymbol{\delta}_{t+1}'\right]\boldsymbol{D}_t^{-1}\mathbb{E}_t\left[\boldsymbol{\delta}_{t+1}\right]}\boldsymbol{\delta}_{t+1}'\boldsymbol{D}_t^{-1}\mathbb{E}_t\left[\boldsymbol{\delta}_{t+1}\right] \mid \boldsymbol{\delta}_{t+1}\right\} \mid m_{t+1}\right\} \\
=& \boldsymbol{z}_t(i)'\mathbb{E}_t\left[\boldsymbol{\delta}_{t+1} \mid m_{t+1}\right]
\end{aligned}
$$

Therefore,

$$\mathbb{E}_t\left[u_{t+1}|m_{t+1}\right] = 0$$

Next, integrate the equation

$$r_{t+1}(i) = \kappa_{t+1} + \widetilde{\beta}_t(i)m_{t+1} + u_{t+1}(i)$$

over asset index set Γ we have that

$$\int_\Gamma u_{t+1}(i)dG(i) = 0$$

$\square$

Proof of Proposition 5

Using Assumption 5, the matrix $\overline{\boldsymbol{S}}_t^c$ has rank $L \geq 1$ for $\omega \in \Omega^0 \subset \Omega$, and $\mathbb{P}(\Omega^0) = 1$.

Let $N = K - L > 0$, similar to the proof of Proposition 1, we can use singular value decomposition of $\overline{\boldsymbol{S}}_t^c = \boldsymbol{Q}_t\boldsymbol{\Lambda}_t\boldsymbol{Q}_t'$, with $\boldsymbol{Q}_t$ orthonormal matrix and $\boldsymbol{\Lambda}_t$ diagonal, and define the selection matrices exactly the same as before, $\boldsymbol{J}_L$ and $\boldsymbol{J}_N$, such that $\boldsymbol{J}_N\boldsymbol{\Lambda}_t\boldsymbol{J}_N'$ produce a $N \times N$ null matrix with all entries being zero, and $\boldsymbol{J}_L\boldsymbol{\Lambda}_t\boldsymbol{J}_L'$ produce a $L \times L$ full rank matrix with diagonals being the L non-zero elements of $\boldsymbol{\Lambda}_t$. Moreover, the following decomposition of identity matrix holds:

$$\boldsymbol{I}_K = \boldsymbol{J}_N'\boldsymbol{J}_N + \boldsymbol{J}_L'\boldsymbol{J}_L$$

Now define the $L \times K$ matrix

$$\boldsymbol{B}_t^c := \boldsymbol{J}_L\boldsymbol{Q}_t'$$

it has rank L since $rank(J_L) = L$ and $\boldsymbol{Q}_t$ is full rank.

Let

$$\begin{aligned}
\boldsymbol{z}_t^c(i) =& \boldsymbol{B}_t^c\overline{\boldsymbol{z}}_t^c(i)\\
\boldsymbol{\delta}_{t+1}^c =& \boldsymbol{B}_t^c\boldsymbol{\delta}_{t+1}
\end{aligned}$$

and verify all parts of Assumptions 1,2,3,4 are satisfied for all $i \in C$.

(a) "Assumption 1, Equation (9)": Notice that since $\boldsymbol{Q}_t$ is orthonormal, its inverse is its transpose, the following identity holds

$$
\begin{aligned}
\overline{\boldsymbol{z}}_t^c(i)'\boldsymbol{\delta}_{t+1} =& \overline{\boldsymbol{z}}_t^c(i)'\boldsymbol{Q}_t\boldsymbol{I}_K\boldsymbol{Q}_t'\boldsymbol{\delta}_{t+1} \\
=& \overline{\boldsymbol{z}}_t^c(i)'\boldsymbol{Q}_t\left\{\boldsymbol{J}_N'\boldsymbol{J}_N + \boldsymbol{J}_L'\boldsymbol{J}_L\right\}\boldsymbol{Q}_t'\boldsymbol{\delta}_{t+1} \\
=& \overline{\boldsymbol{z}}_t^c(i)'\boldsymbol{Q}_t\boldsymbol{J}_N'\boldsymbol{J}_N\boldsymbol{Q}_t'\boldsymbol{\delta}_{t+1} \\
&+ \overline{\boldsymbol{z}}_t^c(i)'\boldsymbol{Q}_t\boldsymbol{J}_L'\boldsymbol{J}_L\boldsymbol{Q}_t'\boldsymbol{\delta}_{t+1} \\
=& \overline{\boldsymbol{z}}_t^c(i)'\boldsymbol{Q}_t\boldsymbol{J}_N'\boldsymbol{J}_N\boldsymbol{Q}_t'\boldsymbol{\delta}_{t+1} \\
&+ \boldsymbol{z}_t^c(i)'\boldsymbol{\delta}_{t+1}^c
\end{aligned}
$$

What is the scaler $a_t(i) := \overline{\boldsymbol{z}}_t^c(i)'\boldsymbol{Q}_t\boldsymbol{J}_N'\boldsymbol{J}_N\boldsymbol{Q}_t'\boldsymbol{\delta}_{t+1}$?
Lets integrate it's square over the set C to uncover its value

$$
\begin{aligned}
\int_C [a_t(i)]^2\, dG(i) =& \int_C \boldsymbol{\delta}_{t+1}'\boldsymbol{Q}_t\boldsymbol{J}_N'\boldsymbol{J}_N\boldsymbol{Q}_t'\overline{\boldsymbol{z}}_t^c(i)\overline{\boldsymbol{z}}_t^c(i)'\boldsymbol{Q}_t\boldsymbol{J}_N'\boldsymbol{J}_N\boldsymbol{Q}_t'\boldsymbol{\delta}_{t+1}dG(i) \\
=& \boldsymbol{\delta}_{t+1}'\boldsymbol{Q}_t\boldsymbol{J}_N'\boldsymbol{J}_N\boldsymbol{Q}_t'\left\{\int_C \overline{\boldsymbol{z}}_t^c(i)\overline{\boldsymbol{z}}_t^c(i)'dG(i)\right\}\boldsymbol{Q}_t\boldsymbol{J}_N'\boldsymbol{J}_N\boldsymbol{Q}_t'\boldsymbol{\delta}_{t+1} \\
=& \boldsymbol{\delta}_{t+1}'\boldsymbol{Q}_t\boldsymbol{J}_N'\boldsymbol{J}_N\boldsymbol{Q}_t'\left\{\boldsymbol{S}_t\right\}\boldsymbol{Q}_t\boldsymbol{J}_N'\boldsymbol{J}_N\boldsymbol{Q}_t'\boldsymbol{\delta}_{t+1} \\
=& \boldsymbol{\delta}_{t+1}'\boldsymbol{Q}_t\boldsymbol{J}_N'\boldsymbol{J}_N\boldsymbol{Q}_t'\boldsymbol{Q}_t\boldsymbol{\Lambda}_t\boldsymbol{Q}_t'\boldsymbol{Q}_t\boldsymbol{J}_N'\boldsymbol{J}_N\boldsymbol{Q}_t'\boldsymbol{\delta}_{t+1} \\
=& \boldsymbol{\delta}_{t+1}'\boldsymbol{Q}_t\boldsymbol{J}_N'\boldsymbol{J}_N\boldsymbol{\Lambda}_t\boldsymbol{J}_N'\boldsymbol{J}_N\boldsymbol{Q}_t'\boldsymbol{\delta}_{t+1} \\
=& \boldsymbol{\delta}_{t+1}'\boldsymbol{Q}_t\boldsymbol{J}_N'\boldsymbol{0}_{K\times K}\boldsymbol{J}_N\boldsymbol{Q}_t'\boldsymbol{\delta}_{t+1} \\
=& 0
\end{aligned}
$$

Therefore, we have established Equation (30).

(b) "Assumption 1, non-degeneracy part" + "Assumption 2, part (2)":

$$
\boldsymbol{S}_t^c = \int_C \boldsymbol{z}_t^c(i)\boldsymbol{z}_t^c(i)'dG(i) = \boldsymbol{B}_t^c\overline{\boldsymbol{S}}_t^c(\boldsymbol{B}_t^c)'
$$

has rank L, full rank, therefore, it is symmetric positive definite. Assumption 2 part (2). Also, all symmetric positive definite matrices has positive diagonal elements, thus

$$
\int_C \left[z_t^{c,l}(i)\right]^2 dG(i) > 0
$$

for all $l \leq L$, and Assumption 1 non-degeneracy are satisfied.

(c) "Assumption 1 normalization"

Integrate

$$\begin{aligned}\int_C \boldsymbol{z}_t^c(i)dG(i) = &\int_C \boldsymbol{B}_t^c \overline{\boldsymbol{z}}_t^c(i)dG(i)\\ =&\boldsymbol{B}_t^c \int_C \overline{\boldsymbol{z}}_t^c(i)dG(i)\\ =&\,0\end{aligned}$$

thus Assumption 1 normalization is satisfied.

(d) "Assumption 1 relevance"

Next, suppose for some l element of vector $\boldsymbol{\delta}_{t+1}^c$, $\mathbb{P}\left[\delta_{t+1}^{c,l}=0\middle|\mathcal{F}_t\right]=1$, that means there exist a $L\times 1$ vector $\boldsymbol{a}_t=(...,0,1,0,...)'$ with l-th element equals to 1 while every other element equals to zero, such that $\mathbb{P}\left[\boldsymbol{a}_t'\boldsymbol{\delta}_{t+1}^c=0\middle|\mathcal{F}_t\right]=1$. Now compute its conditional second moment, with probability 1,

$$0=\mathbb{E}_t\left[\boldsymbol{a}_t'\boldsymbol{\delta}_{t+1}^c(\boldsymbol{\delta}_{t+1}^c)'\boldsymbol{a}_t\right]=\boldsymbol{a}_t'\boldsymbol{B}_t^c\mathbb{E}_t\left[\boldsymbol{\delta}_{t+1}\boldsymbol{\delta}_{t+1}'\right](\boldsymbol{B}_t^c)'\boldsymbol{a}_t$$

subtract both sides by a number squared $\boldsymbol{a}_t'\boldsymbol{B}_t^c\mathbb{E}_t\left[\boldsymbol{\delta}_{t+1}\right]\mathbb{E}_t\left[\boldsymbol{\delta}_{t+1}'\right](\boldsymbol{B}_t^c)'\boldsymbol{a}_t$

$$\begin{aligned}0>&-\boldsymbol{a}_t'\boldsymbol{B}_t^c\mathbb{E}_t\left[\boldsymbol{\delta}_{t+1}\right]\mathbb{E}_t\left[\boldsymbol{\delta}_{t+1}'\right](\boldsymbol{B}^c)'\boldsymbol{a}_t\\ =&\boldsymbol{a}_t'\boldsymbol{B}_t^c\mathbb{E}_t\left[\boldsymbol{\delta}_{t+1}\boldsymbol{\delta}_{t+1}'\right](\boldsymbol{B}_t^c)'\boldsymbol{a}_t\\ &-\boldsymbol{a}_t'\boldsymbol{B}_t^c\mathbb{E}_t\left[\boldsymbol{\delta}_{t+1}\right]\mathbb{E}_t\left[\boldsymbol{\delta}_{t+1}'\right](\boldsymbol{B}_t^c)'\boldsymbol{a}_t\\ =&\boldsymbol{a}_t'\boldsymbol{B}_t^c\boldsymbol{D}_t(\boldsymbol{B}_t^c)'\boldsymbol{a}_t\end{aligned}$$

Consider the $K\times 1$ vector $\boldsymbol{b}_t=(\boldsymbol{B}_t^c)'\boldsymbol{a}_t$, its inner product with itself is

$$\begin{aligned}\boldsymbol{a}_t'\boldsymbol{B}_t^c(\boldsymbol{B}_t^c)'\boldsymbol{a}_t=&\boldsymbol{a}_t'\boldsymbol{J}_L\boldsymbol{Q}_t'\boldsymbol{Q}_t\boldsymbol{J}_L'\boldsymbol{a}_t\\ =&\boldsymbol{a}_t'\boldsymbol{J}_L\boldsymbol{J}_L'\boldsymbol{a}_t\\ =&\boldsymbol{a}_t'\boldsymbol{I}_L\boldsymbol{a}_t\\ =&\boldsymbol{a}_t'\boldsymbol{a}_t=1\end{aligned}$$

Thus with probability 1, we have $\boldsymbol{b}_t\neq\boldsymbol{0}_{K\times 1}$, yet $\boldsymbol{b}_t'\boldsymbol{D}_t\boldsymbol{b}_t<0$, contradict with the almost sure positive definite assumption about $\boldsymbol{D}_t$ from Assumption 3. Therefore, no element of vector $\boldsymbol{\delta}_{t+1}^c$ is zero with probability 1. Hence, Assumption 1 relevance part is satisfied.

(e) "Assumption 2 part (1)"

This is automatically satisfied since $e_{t+1}(i)$ is the same as original, and $C\subset\Gamma$

(f) "Assumption 3

Define

$$
\begin{aligned}
\boldsymbol{D}_t^c := & \mathbb{E}_t\left[\boldsymbol{\delta}_{t+1}^c\left(\boldsymbol{\delta}_{t+1}^c\right)'\right] - \mathbb{E}_t\left[\boldsymbol{\delta}_{t+1}^c\right]\mathbb{E}_t\left[\left(\boldsymbol{\delta}_{t+1}^c\right)'\right] \\
= & \boldsymbol{B}_t^c \boldsymbol{D}_t \left(\boldsymbol{B}_t^c\right)'
\end{aligned}
$$

Let any $\boldsymbol{a} \neq \mathbf{0}_{L\times 1}$, we have previously shown that $(\boldsymbol{B}_t^c)'\boldsymbol{a} \neq \mathbf{0}_{K\times 1}$ Since, $\boldsymbol{D}_t$ is symmetric positive definite,

$$
\boldsymbol{a}'\boldsymbol{D}_t^c\boldsymbol{a} = \boldsymbol{a}'\boldsymbol{B}_t^c\boldsymbol{D}_t(\boldsymbol{B}_t^c)'\boldsymbol{a} > 0
$$

Thus, $\boldsymbol{D}_t^c$ is also symmetric positive definite.

(g) "Assumption 4"

Notice for all $i \in C \subset \Gamma$

$$
\mathbb{E}_t\left[e_{t+1}(i)\middle|\boldsymbol{\delta}_{t+1}^c\right] = \mathbb{E}_t\left[e_{t+1}(i)\middle|\boldsymbol{B}_t^c\boldsymbol{\delta}_{t+1}\right] = \mathbb{E}_t\left[e_{t+1}(i)\middle|\boldsymbol{\delta}_{t+1}\right] = 0
$$

□

Proof of Proposition 6

Apply directly Proposition 5, and Theorem 4 to the collection of variables:

$$
\left\{r_{t+1}(i), \kappa_{t+1}^c, \boldsymbol{z}_t^c(i), \boldsymbol{S}_t^c, \boldsymbol{\delta}_{t+1}^c, e_{t+1}(i)\right\}
$$

and set C □

Proof of Proposition 7

Consider the decomposition of returns in Equation (9):

$$
r_{t+1}(i) = \kappa_{t+1}^* + \left(\kappa_{t+1} - \kappa_{t+1}^*\right) + \boldsymbol{z}_t(i)'\boldsymbol{\delta}_{t+1} + e_{t+1}(i)
$$

The portfolio $\kappa_{t+1} - \kappa_{t+1}^*$ is an excess return, therefore by Equation 14 and 3, it can be decomposed into

$$
\begin{aligned}
\kappa_{t+1} - \kappa_{t+1}^* = & (\boldsymbol{\omega}_t^*)'\boldsymbol{p}_{t+1} + v_{t+1}^* \\
= & (\boldsymbol{\omega}_t^*)'\boldsymbol{S}_t\boldsymbol{\delta}_{t+1} + v_{t+1}^*
\end{aligned}
$$

where $\mathbb{E}_t\left[v^*_{t+1}|\boldsymbol{\delta}_{t+1}\right]=0$ We have

$$
\begin{aligned}
Cov_t\left[r_{t+1}(i),m_{t+1}\right]=&Cov_t\left[\kappa^*_{t+1},m_{t+1}\right]\\
&+Cov_t\left[\kappa_{t+1}-\kappa^*_{t+1},m_{t+1}\right]\\
&+\boldsymbol{z_t}(i)'\mathbb{E}_t\left[\boldsymbol{\delta}_{t+1}\boldsymbol{\delta}'_{t+1}\right]\boldsymbol{D}_t^{-1}\mathbb{E}_t\left[\boldsymbol{\delta}_{t+1}\right]\\
&-\boldsymbol{z}_t(i)'\mathbb{E}_t\left[\boldsymbol{\delta}_{t+1}\right]\mathbb{E}_t\left[\boldsymbol{\delta}'_{t+1}\right]\boldsymbol{D}_t^{-1}\mathbb{E}_t\left[\boldsymbol{\delta}_{t+1}\right]\\
&+Cov_t\left[e_{t+1}(i),m_{t+1}\right]\\
=&Cov_t\left[\kappa^*_{t+1},m_{t+1}\right]+\mathbb{E}_t\left[\kappa_{t+1}-\kappa^*_{t+1}\right]+\boldsymbol{z_t}(i)'\mathbb{E}_t\left[\boldsymbol{\delta}_{t+1}\right]
\end{aligned}
$$

Thus

$$
\begin{aligned}
\mathbb{E}_t\left[r_{t+1}(i)\right]=&\mathbb{E}_t\left[\kappa^*_{t+1}\right]+\mathbb{E}_t\left[\kappa_{t+1}-\kappa^*_{t+1}\right]+\boldsymbol{z_t}(i)'\mathbb{E}_t\left[\boldsymbol{\delta}_{t+1}\right]+\mathbb{E}_t\left[e_{t+1}(i)\right]\\
=&\mathbb{E}_t\left[\kappa^*_{t+1}\right]+\frac{Cov_t\left[r_{t+1}(i),m_{t+1}\right]-Cov_t\left[\kappa^*_{t+1},m_{t+1}\right]}{Var_t\left(m_{t+1}\right)}\mathbb{E}_t\left[m_{t+1}\right]\\
=&\mathbb{E}_t\left[\kappa^*_{t+1}\right]+\left[\beta_t(i)-\beta^*_{t+1}\right]\mathbb{E}_t\left[m_{t+1}\right]
\end{aligned}
$$

Define $u^*_{t+1}(i)\equiv r_{t+1}(i)-\kappa^*_{t+1}-\left[\beta_t(i)-\beta^*_{t+1}\right]m_{t+1}$, we have $\mathbb{E}_t\left[u^*_{t+1}(i)\right]=0$. Also

$$
\begin{aligned}
\beta_t(i)Var_t\left[m_{t+1}\right]=&Cov_t\left[r_{t+1}(i),m_{t+1}\right]\\
=&Cov_t\left[\kappa^*_{t+1}-\beta^*_{t+1}m_{t+1},m_{t+1}\right]\\
&+Cov_t\left[\beta_t(i)m_{t+1},m_{t+1}\right]\\
&+Cov_t\left[u^*_{t+1}(i),m_{t+1}\right]\\
=&0+\beta_t(i)Var_t\left[m_{t+1}\right]+Cov_t\left[u^*_{t+1}(i),m_{t+1}\right]
\end{aligned}
$$

Therefore $\mathbb{E}_t\left[u^*_{t+1}(i)m_{t+1}\right]=0$
Moreover, since κ^*_{t+1} is the Stochastic Discount Factor net return, and $d^*_{t+1}=\kappa^*_{t+1}+1$ is the Stochastic Discount Factor gross return, using Assumption 4, and the fact that m_{t+1} is an excess return. We have $\mathbb{E}_t\left[m_{t+1}d^*_{t+1}\right]=0$, and

$$
\begin{aligned}
\mathbb{E}_t\left[r_{t+1}(i)d^*_{t+1}\right]=&\mathbb{E}_t\left[(d^*_{t+1})^2\right]-\mathbb{E}_t\left[d^*_{t+1}\right]\\
=&\mathbb{E}_t\left[\kappa^*_{t+1}d^*_{t+1}\right]+\mathbb{E}_t\left[\left[\beta_t(i)-\beta^*_t\right]m_{t+1}d^*_{t+1}\right]+\mathbb{E}_t\left[d^*_{t+1}u^*_{t+1}(i)\right]
\end{aligned}
$$

Therefore $\mathbb{E}_t\left[d^*_{t+1}u^*_{t+1}(i)\right]=\mathbb{E}_t\left[\kappa^*_{t+1}u^*_{t+1}(i)\right]=0$ □

Proof of Proposition 8

Using Proposition 5

$$r_t^f = r_{t+1}(i^f) = \kappa_{t+1}^* + \left[\beta_t(i^f) - \beta_t^*\right] m_{t+1} + u_{t+1}^*(i^f)$$

Since

$$\mathbb{E}_t\left[u_{t+1}^*(i)\right] = \mathbb{E}_t\left[u_{t+1}^*(i)m_{t+1}\right] = \mathbb{E}_t\left[u_{t+1}^*(i)\kappa_{t+1}^*\right] = 0$$

It must be that

$$\begin{aligned}
0 = Var_t\left[r_t^f\right] =& Var_t\left[\kappa_{t+1}^* + \left[\beta_t(i^f) - \beta_t^*\right] m_{t+1}\right] \\
&+ 2Cov_t\left[\kappa_{t+1}^* + \left[\beta_t(i^f) - \beta_t^*\right] m_{t+1}, u_{t+1}^*(i^f)\right] \\
&+ Var_t\left[u_{t+1}^*\right] \\
=& Var_t\left[\kappa_{t+1}^* + \left[\beta_t(i^f) - \beta_t^*\right] m_{t+1}\right] + 0 + Var_t\left[u_{t+1}^*\right]
\end{aligned}$$

Therefore $u_{t+1}^* = 0$ and $\kappa_{t+1}^* + \left[\beta_t(i^f) - \beta_t^*\right] m_{t+1} = r_t^f$

Moreover, by definition

$$\beta_t(i^f) = Cov_t\left[r_t^f, m_{t+1}\right] / Var_t\left[m_{t+1}\right] = 0$$

Therefore we have

$$r_t^f = \kappa_{t+1}^* - \beta_t^* m_{t+1}$$

□

Proof of Proposition 10

Using Propositions 8 and 9

$$\begin{aligned}
1 + r_t^f =& d_{t+1}^* - \beta_t^* m_{t+1} \\
=& d_{t+1}^* - \frac{Cov_t\left[d_{t+1}^* - 1, m_{t+1}\right]}{Var_t\left[m_{t+1}\right]} m_{t+1} \\
=& d_{t+1}^* - \frac{\mathbb{E}_t\left[d_{t+1}^* m_{t+1}\right] - \mathbb{E}_t\left[d_{t+1}^*\right]\mathbb{E}_t\left[m_{t+1}\right]}{Var_t\left[m_{t+1}\right]} m_{t+1} \\
=& d_{t+1}^* + \mathbb{E}_t\left[d_{t+1}^*\right] m_{t+1}
\end{aligned}$$

Take expectation w.r.t time t information set,

$$1 + r_t^f = \mathbb{E}_t\left[d_{t+1}^*\right] + \mathbb{E}_t\left[d_{t+1}^*\right]\mathbb{E}_t\left[m_{t+1}\right]$$

solve for $\mathbb{E}_t\left[d^*_{t+1}\right]$, which is equal to $-\beta^*_t$ to arrive at

$$r^f_t = \kappa^*_{t+1} + \frac{1 + r^f_{t+1}}{1 + \mathbb{E}_t\left[m_{t+1}\right]} m_{t+1}$$

Thus we have

$$\kappa^*_{t+1} = r^f_t - \frac{1 + r^f_t}{1 + \mathbb{E}_t\left[m_{t+1}\right]} m_{t+1}$$

Notice that m_{t+1} is the zero-cost efficient portfolio. $\square$

References

Ahn, D.-H., J. Conrad, and R. F. Dittmar (2009). Basis assets. *Review of Financial Studies 22*(12), 5133–5174.

Andersen, T. G. and T. Bollerslev (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review 39*, 885–905.

Asness, C. and A. Frazzini (2013). The devil in hml's details. *Journal of Portfolio Management 39*(4), 49.

Asness, C. S., A. Frazzini, R. Israel, T. J. Moskowitz, and L. H. Pedersen (2015). Size matters, if you control your junk. *Fama-Miller Working Paper*.

Asness, C. S., A. Frazzini, and L. H. Pedersen (2014). Quality minus junk. *Available at SSRN 2312432*.

Asness, C. S., T. J. Moskowitz, and L. H. Pedersen (2013). Value and momentum everywhere. *The Journal of Finance 68*(3), 929–985.

Brandt, M. W., P. Santa-Clara, and R. Valkanov (2009). Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns. *Review of Financial Studies 22*(9), 3411–3447.

Bryzgalova, S. (2014). Spurious factors in linear asset pricing models. Technical report, Working Paper.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of finance 52*(1), 57–82.

Chen, Z. and R. Petkova (2012). Does idiosyncratic volatility proxy for risk exposure? *Review of Financial Studies 25*(9), 2745–2787.

Cochrane, J. H. (2009). *Asset Pricing:(Revised Edition)*. Princeton university press.

Epps, T. W. (1979). Comovements in stock prices in the very short run. *Journal of the American Statistical Association 74*(366a), 291–298.

Fama, E. and K. French (2014). A five-factor asset pricing model, fama-miller working paper.

Fama, E. F. and K. R. French (1993). Common risk factors in the returns on stocks and bonds. *Journal of financial economics 33*(1), 3–56.

Fama, E. F. and J. D. MacBeth (1973). Risk, return, and equilibrium: Empirical tests. *The Journal of Political Economy*, 607–636.

Frazzini, A. and L. H. Pedersen (2014). Betting against beta. *Journal of Financial Economics 111*(1), 1–25.

Gagliardini, P., E. Ossola, and O. Scaillet (2016). Time-varying risk premium in large cross-sectional equity data sets. *Econometrica 84*(3), 985–1046.

Gandhi, P. and H. Lustig (2015). Size anomalies in us bank stock returns. *The Journal of Finance 70*(2), 733–768.

Gibbons, M. R., S. A. Ross, and J. Shanken (1989). A test of the efficiency of a given portfolio. *Econometrica: Journal of the Econometric Society*, 1121–1152.

Gospodinov, N., R. Kan, and C. Robotti (2014). Misspecification-robust inference in linear asset-pricing models with irrelevant risk factors. *Review of Financial Studies 27*(7), 2139–2170.

Hansen, L. P. and R. Jagannathan (1997). Assessing specification errors in stochastic discount factor models. *The Journal of Finance 52*(2), 557–590.

Hansen, L. P. and S. F. Richard (1987). The role of conditioning information in deducing testable restrictions implied by dynamic asset pricing models. *Econometrica: Journal of the Econometric Society*, 587–613.

Harvey, C. R., Y. Liu, and H. Zhu (2015). ... and the cross-section of expected returns. *Review of Financial Studies*, hhv059.

Hou, K., C. Xue, and L. Zhang (2014). Digesting anomalies: An investment approach. *Review of Financial Studies*, hhu068.

Jagannathan, R. and Z. Wang (1998). An asymptotic theory for estimating beta-pricing models using cross-sectional regression. *The Journal of Finance 53*(4), 1285–1309.

Kan, R. and C. Robotti (2009). Model comparison using the hansen-jagannathan distance. *Review of Financial Studies 22*(9), 3449–3490.

Kan, R., C. Robotti, and J. Shanken (2013). Pricing model performance and the two-pass cross-sectional regression methodology. *The Journal of Finance 68*(6), 2617–2649.

Lewellen, J., S. Nagel, and J. Shanken (2010). A skeptical appraisal of asset pricing tests. *Journal of Financial Economics 96*(2), 175–194.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The review of economics and statistics*, 13–37.

MacKinlay, A. C. and L. Pástor (2000). Asset pricing models: Implications for expected returns and portfolio selection. *Review of Financial studies 13*(4), 883–916.

Malamud, S. and G. Vilkov (2015). Non-myopic betas. *Available at SSRN*.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–08.

Newey, W. K. and K. D. West (1994). Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies 61*(4), 631–653.

Novy-Marx, R. (2014). Predicting anomaly performance with politics, the weather, global warming, sunspots, and the stars. *Journal of Financial Economics 112*(2), 137–146.

Novy-Marx, R. (2015). Fundamentally, momentum is fundamental momentum. Technical report, National Bureau of Economic Research.

Novy-Marx, R. and M. Velikov (2016). A taxonomy of anomalies and their trading costs. *Review of Financial Studies 29*(1), 104–147.

Pástor, L. and R. F. Stambaugh (2003). Liquidity risk and expected stock returns. *Journal of Political Economy 111*(3), 642–685.

Patton, A. J. and A. Timmermann (2010). Monotonicity in asset returns: New tests with applications to the term structure, the capm, and portfolio sorts. *Journal of Financial Economics 98*(3), 605–625.

Roll, R. (1977). A critique of the asset pricing theory's tests part i: On past and potential testability of the theory. *Journal of financial economics 4*(2), 129–176.

Scholes, M. and J. Williams (1977). Estimating betas from nonsynchronous data. *Journal of financial economics 5*(3), 309–327.

Shanken, J. (1992). On the estimation of beta-pricing models. *Review of Financial studies 5*(1), 1–33.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The journal of finance 19*(3), 425–442.