Testing Asset Pricing Models With Individual Stocks

Charlie Clarke
University of Kentucky

Morteza Momeni
University of Kentucky

January 4, 2023

Motivation

- Portfolio sorts require a choice about which characteristics predict expected returns, but growing evidence suggests many characteristics perform poorly out-of-sample.
- If test asset is "data snooped," the true asset pricing model may not be able to price it.
- Newer tests of mean-variance efficiency (Barillas and Shanken (2017, 2018)) may compound the "data snooping" problem.
- We adapt methods from the fund performance literature to perform traditional time-series tests of factor models on individual stocks.

New Technology from the Fund Performance Literature

- We adapt a methodology from Kosowski, Timmerman, Wermers, and White (2006) and Fama and French (2010)
- Do fund managers have skill? Can we answer this without knowing the predictors of firm skill?
- Insight: We can bootstrap data under the null hypothesis (no alpha) and compare observed alpha to the zero-alpha null.
- We follow the insight into individual stocks. Create population data that has zero alpha. Bootstrap "observed alpha" confidence intervals. Does sample alpha lie within zero-alpha confidence intervals?

Testing Asset Pricing Models

Time-series test:

$$R_{i,t} = \alpha_i + \beta_{i,1} F_{1,t} + \epsilon_{i,t}$$

Null Hypothesis: Do all stocks have zero alpha? (No Mispricing)

$$H_0 : \alpha_1 = \alpha_2 = ... = \alpha_N = 0$$

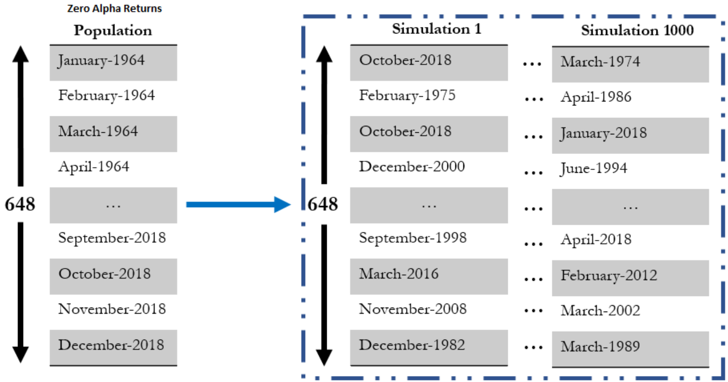
Estimate alphas and create Zero-alpha Returns

$$Z_{i,t} = R_{i,t} - \widehat{\alpha_i} = \widehat{\beta_{i,1}} F_{1,t} + \widehat{\epsilon_{i,t}}$$

Simulate the world under the null hypothesis of no alpha.

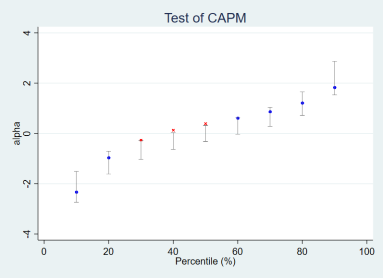
Is the "observed" data similar to the what we would expect, if there was no alpha?

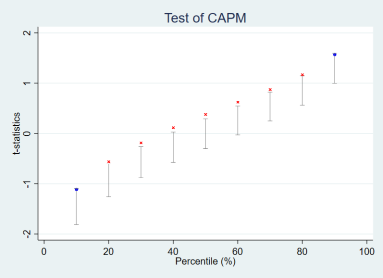
Bootstrap Resampling from Zero Alpha Returns to Simulate Data Under Null Hypothesis



Resampling Creates Confidence Intervals Under Null Hypothesis, Compare Observed Alphas

We condense N individual stock alphas (really t-statistics) to nine deciles (10%, 20%, . . . , 90%)





Do the sample observed alphas fall within the bootstrapped confidence intervals (adjusted for multiple hypotheses)?

Does It Work? Test Size and Power

- Create a Population where the CAPM is the Data Generating Process
- Test the CAPM and the FF6 Model using our procedure
- We should reject the CAPM ($< 5\%$) of the time (Test Size)
- We hope to reject the FF6 Model as much as possible (Test Power)
- Our Population: $R_{i,t} - \widehat{\alpha}_i = \widehat{\beta}_i F_t + \widehat{\epsilon_{i,t}}$

Test Size and Power

Panel A: Bonferonni Confidence Intervals

Model	Stat	Test	Rejection
			Pop: $\alpha_{CAPM} = 0$
CAPM	t	Size	1.5%
FF6	t	Power	97.5%
			Pop: $\alpha_{FF6} = 0$
FF6	t	Size	3.3%
CAPM	t	Power	83.2%

Bonferonni is conservative. We explore "optimized" confidence intervals (max power given size) in the paper and find very similar results.

What is HML when CAPM describes the population?

$$E[HML_t] = \alpha + \beta E[R_{m,t}]$$

.

HML is a set a of weights, $w_{i,t}$, on each stock each period.

$$HML_t = \sum w_{i,t} R_{i,t} = \sum w_{i,t}(\beta_i R_{m,t} + \epsilon_{i,t})$$

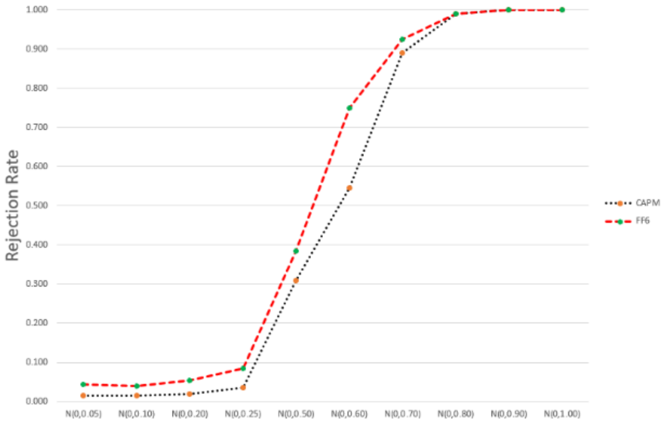
The last equal sign follows, because the CAPM holds.

$$E[HML_t] = \beta_{HML} E[R_{m,t}] + E[\sum w_{i,t}\epsilon_{i,t}]$$

Creating the HML's CAPM alpha when the CAPM is true requires choosing weights so that the error terms sum to the alpha.

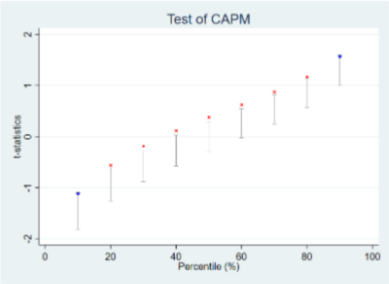
$$\alpha = E[\sum w_{i,t}\epsilon_{i,t}]$$

Power Given Distribution of Alpha

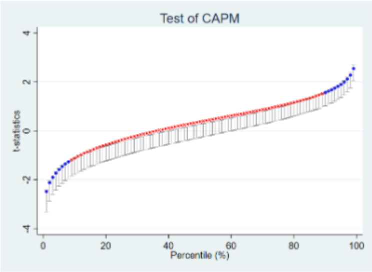


$$\alpha_i \sim \mathcal{N}(0, \sigma)$$

Test the CAPM on Individual Stocks



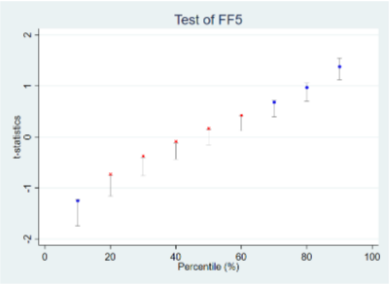
(a) 9 Percentiles



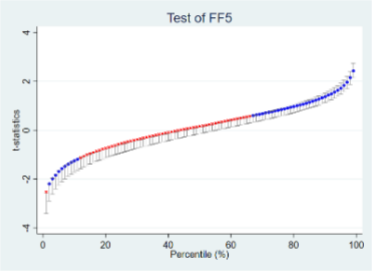
(b) 99 Percentiles

Figure 5: Testing the CAPM Using All Stocks

Test the FF5 on Individual Stocks

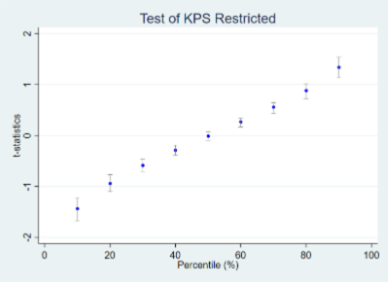


(c) 9 Percentiles

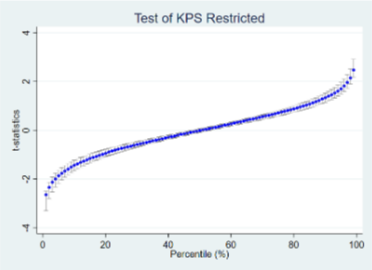


(d) 99 Percentiles

Test the KPS "Restricted" on Individual Stocks



(a) 9 Percentiles



(b) 99 Percentiles

Reject Most Models

Rejected:

CAPM, Fama and French (1993, 2015) 3, 5 and 6 factors.

Hou, Xue, Xhang (2015, 2020) Q-factor model (Market, Size, ROE, Inv), HXZ (2020) (+ Expected Growth in Profitability)

Pastor and Stambaugh (2003) (FF3 + Momentum + Liquidity Factor)

Pastor and Yuan (2017) Mispricing Factors (Market, Size, Performance, Management)

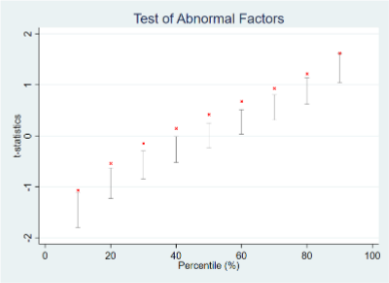
Not Rejected:

Kelly, Pruitt, Su (2019) - Restricted, Out-of-Sample

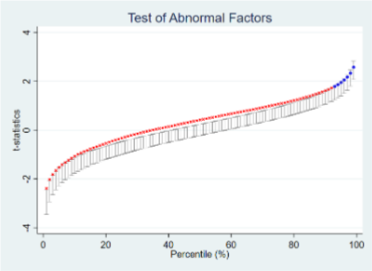
Barillas and Shanken Result

- When comparing two models, if the factors of each model are included as test assets, the other test assets are *irrelevant* to the model's success under traditional asset pricing metrics (for example, HJ distance). The only thing that matters is the squared Sharpe Ratio of the factors.
- Obvious Problem: "licence to fish" – the best model is now the ex post mean variance efficient portfolio.
- Unpromising solution: compare to sorted portfolios – which ones? these have the same "data snooping" issues.
- Our solution: compare to individual stocks. If a model as a high Sharpe ratio, but can't price individual stocks, it should be rejected.

“Abnormal Factors” Model



(a) 9 Percentiles



(b) 99 Percentiles

Figure 12: Testing Abnormal Factors Using All Stocks

Cherry picked - Market plus: CAR Earnings Announcements (4.81), Change in Analysts Earnings Forecast (4.71), 12-month Industry Lead-Lag Effect (3.45), 12-month Quarterly E/P (2.21), Change in Net Operating Assets (2.59), 4-quarter change in ROE (5.23), and Seasonality (4.19)

Mispricing Measure

An absolute measure of mispricing is the average of the absoluted difference at each of the nine deciles from the mean of the simulations:

$$|M| = \frac{1}{9} \sum_{i=10}^{90} |t(\alpha)_i^o - t(\alpha)_i^\mu|$$

And a squared measure of mispricing is the squared deviations from the average of the simulations:

$$M^2 = \frac{1}{9} \sum_{i=10}^{90} (t(\alpha)_i^o - t(\alpha)_i^\mu)^2$$

Model Mispricing

Model	$\lvert M \rvert$	**P Value**	M^2	**P Value**
KPS-R	0.01	0.89	0.000	0.90
FF3	0.14	0.01	0.022	0.01
FF5	0.17	0.00	0.029	0.00
Stambaugh & Yuan	0.20	0.00	0.043	0.00
FF4 + Liquidity	0.23	0.00	0.056	0.00
FF6	0.23	0.00	0.057	0.00
KPS-OOS	0.26	0.08	0.068	0.09
HXZ4	0.28	0.00	0.083	0.00
CAPM	0.35	0.00	0.122	0.00
KPS-U	0.35	0.00	0.134	0.00

Conclusion

- We adapt a methodology from the fund performance literature to testing asset pricing models on individual stocks.
- We show our procedure has appropriate size and yet maintains the power to reject models.
- We reject a number of leading models. IPCA models perform the best.