Bai and Pukthuanthong (2020) - "Machine Learning classification methods and portfolio allocation: an examination of market efficiency"

Discussion by Ahmed Guecioueur

INSEAD

13 February 2021
Miami Herbert Winter Conference on ML & Business

Summary

Typical cross-sectional return predictability setup I

General model of firms' next-period excess returns

$$r_{i,t+1} = E_t(r_{i,t+1}) + \epsilon_{i,t+1} = h(x_{it}) + \epsilon_{i,t+1}, \tag{1}$$

- stocks are indexed as $i = 1, \ldots, N_t$
- months are indexed by $t = 1, \ldots, T$
- x_{it} is an M-dimensional vector of publicly-available predictors, such as firm-level characteristics or market-level variables
- Examples: Gu, Kelly, and Xiu (2020), Freyberger, Neuhierl, and Weber (2020), Han et al. (2020), and Evgeniou, Guecioueur, and Prieto (2020)

Typical cross-sectional return predictability setup II

Predicting $r_{i,t+1}$ is a cross-sectional regression problem:

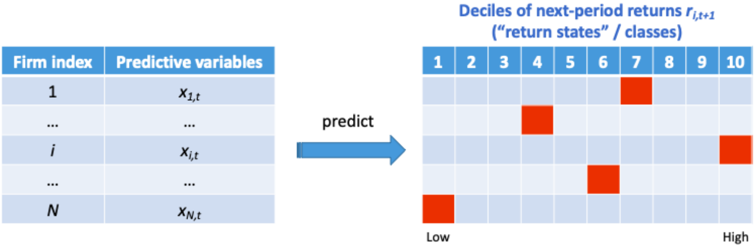
Firm index	Predictive variables
1	$x_{1,t}$
...	...
i	$x_{i,t}$
...	...
N	$x_{N,t}$

predict

Next-period returns
$r_{1,t+1}$
...
$r_{i,t+1}$
...
$r_{N,t+t}$

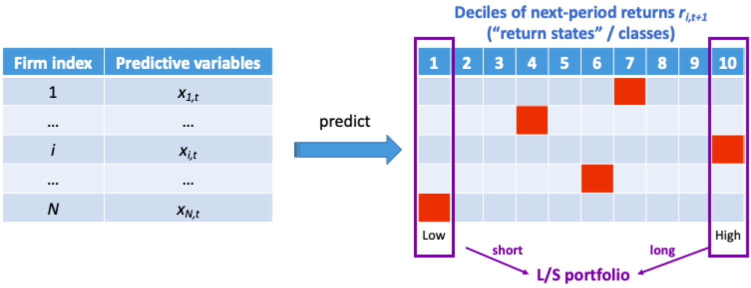
Classification setup of Bai and Pukthuanthong (2020)

Bai and Pukthuanthong (2020) classify firms into deciles of next-period returns (instead of predicting $r_{i,t+1}$ directly):



From “return states” to portfolios I

Predicted return deciles map nicely into forming long/short portfolios to test the economic significance of OOS predictability:



From "return states" to portfolios II

L/S portfolios with best OOS performance

- Equal-weighted: monthly SR of 0.87 (vs. market 0.13)
- Value-weighted: monthly SR of 0.42 (vs. market 0.12)
- Compare favourably to the cross-sectional regression-based approach of Gu, Kelly, and Xiu (2020)
- Not due to leverage or excessive concentration in microcaps (whether defined at bottom 5% of 10% of caps)

Other results

Compare classifier accuracies against benchmarks

- Benchmarks are outperformed, with statistical significance
- Benchmarks interpreted with respect to market efficiency

Predictability

- (Average) classification accuracy of future deciles is best for the highest and lowest realized future deciles

Feature importance (based on Total SSE reductions)

- Firm-level idiosyncratic volatility is the most important feature for both neural nets and tree-based models
- Bid-ask spread and return volatility are also especially important for tree-based models

Comments

Comment: advantages of a classification approach I

Stated advantages

1. Can easily calculate predictive accuracy
2. Can compare to benchmarks to test hypotheses
3. Can assess which stocks are more predictable than others
4. Can quantify probabilities of predictions

Other approaches

- (1)-(3) above also apply to regression approaches
- (4) applies to many probabilistic ML models

Unique advantages for classification?

- Would help to sharpen the importance

Comment: advantages of a classification approach II

Some leads

- Can be used for applications that do not require conditional expectations of individual firm-level returns, such as:
  - L/S portfolio formation (already in the paper)
  - risk management
- May be useful to characterise feature values across different quantiles – see Barnes and Hughes (2002)
- There is a theoretical link between classification and quantile regression – see Langford, Oliveira, and Zadrozny (2006)

Comment: benchmarks & Monte Carlo I

Tests of market efficiency

- 2 of the “no information” benchmarks use information about the in-sample/historical return states
  - Can this be framed as testing whether firm-level information beyond returns improves predictability? i.e. version of the weak-form vs. semistrong-form debate
- 2 of the “no information” benchmarks use information about the out-of-sample/future return states
  - Justification for using OOS information?
- “Martingale” benchmark: prediction for future return state is the current return state
  - Differs from the classic martingale hypothesis that *prices* follow a martingale and *returns* are unpredictable

Comment: benchmarks & Monte Carlo II

Usage of Monte Carlo

- 3 of the “no information” benchmarks are non-deterministic (i.e. random predictions)
- Monte Carlo samples can be used to produce an average performance measure for these – is this the usage?
- Is Monte Carlo also applied to (any non-deterministic) non-benchmark classifiers too?

Comment: statistical testing & Monte Carlo I

Would benefit from some more detail

- Why & how Monte Carlo sampling is performed
- How the binomial test is applied: independent trials?

Some ways classifiers can be statistically compared

- Within a dataset: trials = samples
- Across multiple datasets: trials = datasets

Comment: statistical testing & Monte Carlo II

Example of within-dataset binomial test, from Salzberg (1997)

- Two classifiers, A & B
- Let n be the number of samples within the dataset on which A & B predict different classes (i.e. drop ties)
- Let s be the number of successes for A (i.e. A is correct and B is wrong). So $n - s$ is the number of failures for A (i.e. A is wrong and B is correct)
- Suppose $s = 35$ (A correct, B wrong) out of $n = 50$ examples where A & B predicted different classes. Then the probability of this result under the null hypothesis of $P(success) = 0.5$ is

$$\sum_{s=35}^{n=50} \frac{n!}{s!(n-s)!}(0.5)^n = 0.0032$$

Comment: statistical testing & Monte Carlo III

Across-dataset testing

- Examples: Japkowicz and Shah (2011, Ch. 6.6 & 6.7)
- Demšar (2006) recommends the Wilcoxon signed-rank test

Independence & across-dataset comparisons

- Binomial test (like others) assumes sample independence
- Dietterich (1998) and Nadeau and Bengio (2003) show that the t-test (similar assumption) is not well calibrated in situations where independence of samples does not hold
- Informally, overlapping datasets can lead to bias
- Monte Carlo samples to produce multiple datasets for across-dataset comparison? Independence may not hold

Comment: realized volatility I

Good OOS performance among the extreme deciles of next-period returns, especially transitions between deciles $1 \leftrightarrow 10$:

	New 1	New 2	New 3	New 4	New 5	New 6	New 7	New 8	New 9	New 10
Old 1	0.5054	0.0228	0.0045	0.0020	0.0142	0.0446	0.0504	0.0476	0.0928	0.4449
Old 2	0.4980	0.0735	0.0122	0.0064	0.0456	0.1478	0.1309	0.1040	0.1516	0.2424
Old 3	0.4403	0.0902	0.0158	0.0125	0.0787	0.2515	0.1773	0.1052	0.1266	0.1753
Old 4	0.4098	0.0874	0.0195	0.0146	0.0995	0.3184	0.2009	0.1023	0.0993	0.1475
Old 5	0.4226	0.0832	0.0177	0.0152	0.1034	0.3350	0.2087	0.0923	0.0874	0.1404
Old 6	0.4244	0.0822	0.0200	0.0153	0.1101	0.3358	0.2028	0.0963	0.0848	0.1390
Old 7	0.4094	0.0942	0.0245	0.0159	0.1028	0.3333	0.2008	0.1025	0.0849	0.1303
Old 8	0.4376	0.1093	0.0295	0.0189	0.1016	0.3133	0.1748	0.0991	0.0876	0.1293
Old 9	0.5011	0.1472	0.0350	0.0221	0.0834	0.2463	0.1214	0.1033	0.0948	0.1270
Old 10	0.7522	0.1173	0.0257	0.0166	0.0446	0.1067	0.0586	0.0613	0.0667	0.0992

Comment: realized volatility II

It seems that these higher (absolute) return states are persistent from one month to the next:

Panel A: True Return State Transition Probability Matrix 196301:201912										
	New 1	New 2	New 3	New 4	New 5	New 6	New 7	New 8	New 9	New 10
Old 1	0.1741	0.1063	0.0816	0.0686	0.0665	0.0660	0.0719	0.0817	0.1052	0.1782
Old 2	0.1137	0.1073	0.0963	0.0891	0.0879	0.0875	0.0918	0.0993	0.1090	0.1180
Old 3	0.0859	0.0987	0.0997	0.1011	0.1007	0.1033	0.1050	0.1054	0.1059	0.0944
Old 4	0.0713	0.0899	0.1007	0.1073	0.1127	0.1134	0.1122	0.1092	0.1014	0.0817
Old 5	0.0696	0.0860	0.0992	0.1094	0.1128	0.1203	0.1167	0.1098	0.0981	0.0779
Old 6	0.0690	0.0868	0.1002	0.1084	0.1138	0.1177	0.1186	0.1116	0.0970	0.0768
Old 7	0.0675	0.0897	0.1025	0.1083	0.1134	0.1163	0.1164	0.1121	0.0980	0.0758
Old 8	0.0753	0.0973	0.1054	0.1067	0.1102	0.1123	0.1092	0.1058	0.0984	0.0794
Old 9	0.0958	0.1103	0.1061	0.1023	0.0976	0.0974	0.0971	0.1009	0.0999	0.0927
Old 10	0.1742	0.1236	0.0966	0.0825	0.0752	0.0736	0.0743	0.0802	0.0912	0.1284

Are high-volatility stocks the most predictable?

- The best-predictable high (absolute) return states appear to have high persistence (as in volatility clustering)
- Idiosyncratic volatility was an important firm-level feature

Comment: deciles (and other quantiles)

Empirical validation of deciles

- Frequency: how close is the partition to 1/10 firms per decile?
- Consistency: are mean returns in deciles non-overlapping?
- May help link classification framing to quantile regression

Performance across quantiles

- Currently, the top & bottom deciles are used to form L/S portfolios, as OOS performance is concentrated there
- Would using quintiles/terciles improve classification performance in intermediate quantiles?

Conclusion

Contributions

- Bai and Pukthuanthong (2020) take a fresh look at an asset pricing problem through the eyes of a ML classification setup
  - Predict the firm-distribution of conditional returns
  - This directly maps to long/short portfolio formation
- OOS performance: high L/S Sharpe Ratios & outperforms benchmarks in classification accuracy
- Analyze predictor importance and return state predictability

Main suggestions

- Other advantages of the classification framing and the distribution prediction
- More detail on statistical testing and benchmarks/market efficiency hypotheses
- Links to risk management and realized volatility?

References

References I

Bai, Yang, and Kuntara Pukthuanthong. 2020. "Machine Learning classification methods and portfolio allocation: an examination of market efficiency". *Available at SSRN* 3665051.

Barnes, Michelle L, and Anthony W Hughes. 2002. *A quantile regression analysis of the cross section of stock market returns*. Tech. rep. Federal Reserve Bank of Boston.

Chang, Woo-Jin, et al. 2021. "The higher moments of future earnings". *The Accounting Review* 96 (1): 91–116.

Demšar, Janez. 2006. "Statistical comparisons of classifiers over multiple data sets". *Journal of Machine Learning Research* 7:1–30.

Dietterich, Thomas G. 1998. "Approximate statistical tests for comparing supervised classification learning algorithms". *Neural Computation* 10 (7): 1895–1923.

References II

Evgeniou, Theodoros, Ahmed Guecioueur, and Rodolfo Prieto. 2020. “Uncovering sparsity and heterogeneity in firm-level return predictability using machine learning”. *Available at SSRN* 3604921.

Freyberger, Joachim, Andreas Neuhierl, and Michael Weber. 2020. “Dissecting characteristics nonparametrically”. *Review of Financial Studies* 33 (5): 2326–2377.

Gu, Shihao, Bryan Kelly, and Dacheng Xiu. 2020. “Empirical asset pricing via machine learning”. *Review of Financial Studies* 33 (5): 2223–2273.

Han, Yufeng, et al. 2020. “Firm characteristics and expected stock returns”. *Available at SSRN* 3185335.

Japkowicz, Nathalie, and Mohak Shah. 2011. *Evaluating learning algorithms: a classification perspective*. Cambridge University Press.

Langford, John, Roberto Oliveira, and Bianca Zadrozny. 2006. “Predicting conditional quantiles via reduction to classification”. In *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, 257–264.

References III

Nadeau, Claude, and Yoshua Bengio. 2003. "Inference for the generalization error". *Machine Learning* 52 (3): 239–281.

Salzberg, Steven L. 1997. "On comparing classifiers: Pitfalls to avoid and a recommended approach". *Data Mining and Knowledge Discovery* 1 (3): 317–328.