University of Warsaw
Faculty of Economic Sciences

Maastricht University
School of Business and Economics

APPLICATIONS OF BIG DATA METHODS IN FINANCE: INDEX TRACKING

Kamil Simka[1], Luca Margaritella[2]

[1] Faculty of Economic Sciences, University of Warsaw (kamilsimka@gmail.com)

[2] School of Business and Economics (SBE), Maastricht University (luca.margaritella@gmail.com)

RESEARCH OBJECTIVES

Although the *curse of dimensionality* does not relate to most financial settings, high-dimensional methods gained some relevance in the recent finance literature. **Index tracking** aims at finding an optimal sample of stocks able to mimic the behavior of an equity index. We solve the problem by imposing a ℓ_1 regularization in the optimization procedure and we consider the LARS of Efron et al. (2004) algorithm in order to solve the problem.

DATA

Most researchers focus their analyses on indexes of developed economies. But the biggest challenge for index tracking models is imposed by emerging markets. Such markets frequently lack in historical data and consist of numerous stocks. Here, market indexes of the Warsaw Stock Exchange are considered, namely the **WIG**, **WIG20**, and **mWIG40** between January 2009 and December 2016 with weekly frequency. Data are collected from bossa.pl.

CONSTRAINT OPTIMIZATION PROBLEM

To track the indexes weekly returns (R_I) a dynamic portfolio optimization algorithm is constructed by selecting a predefined number of the most significant stocks (K). To this aim, we firstly imply the shrinking ℓ_1 penalty formulation of the Penalized least squares equation (LASSO, Tibshirani (1996)) to obtain variable selection. Secondly, the underlying parameters ω's are set to satisfy a collection of linear constraints, thus yielding the following **Constrained LASSO** (CLASSO) parametrization:

$$\underset{\omega}{minimize} \qquad ||R_I - R_P\omega||_2^2/T + \lambda||\omega||_1^1$$

$$s.t. \qquad \sum_{i=1}^{\#J^*} \omega_i \leq 1; \qquad 0 \leq \omega \leq 1; \qquad \#J^* = K$$

As LASSO imposes an unwanted bias to coefficients, we carry out a **post OLS** estimation with variables selected by the CLASSO procedure (P-CLASSO):

$$\underset{\omega}{minimize} \qquad ||R_I - R_P\omega||_2^2/T$$

$$s.t. \qquad \sum_{i=1}^{\#J^*} \omega_i = 1; \qquad 0 \leq \omega \leq 1$$

EVALUATION OF TRACKING QUALITY

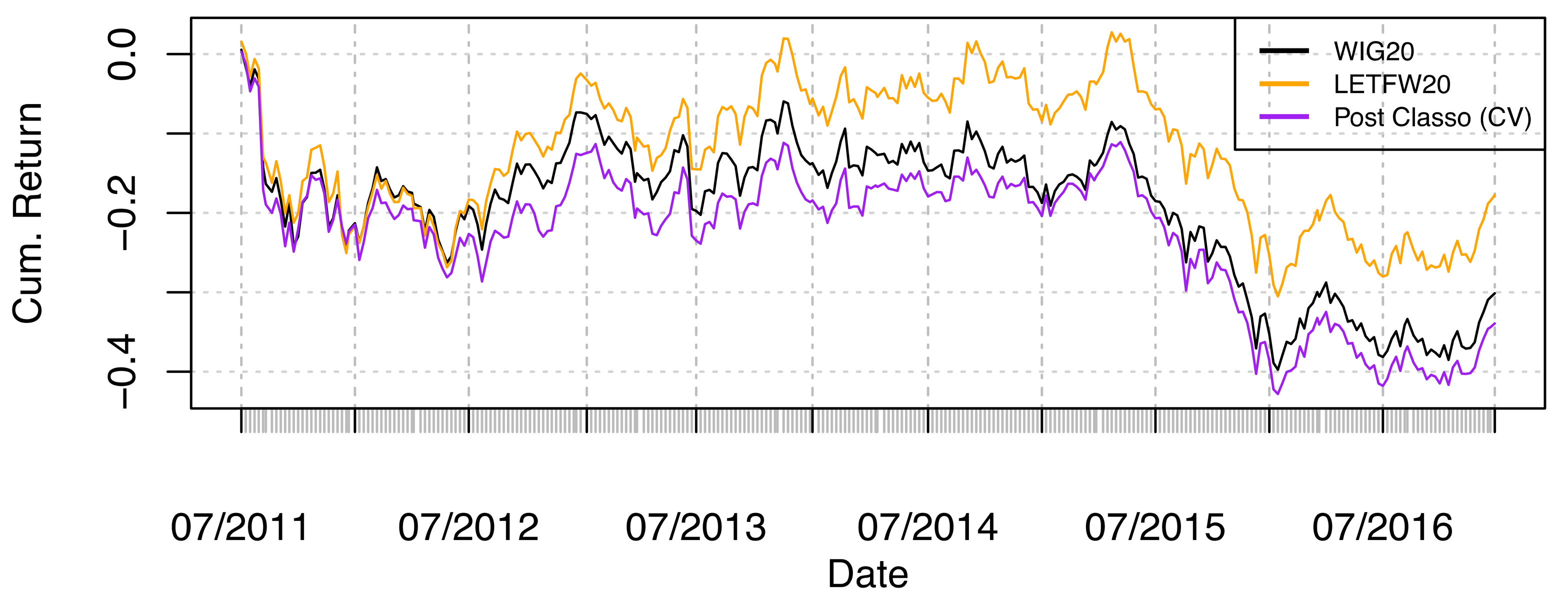
$$TE = \frac{1}{T}||R_{I,i} - R_{P,i} - (\bar{R}_{I,i} - \bar{R}_{P,i})||_2^1$$

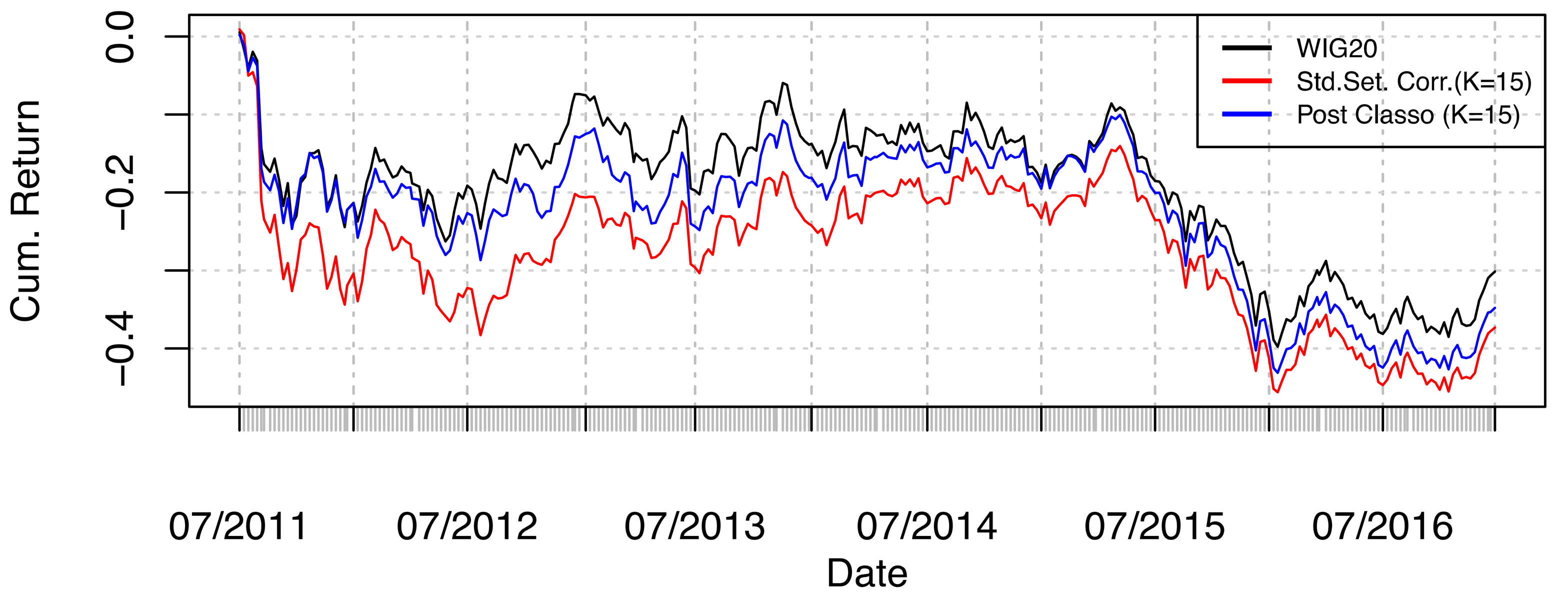
$$RMSE = \frac{1}{T}||R_{I,t} - R_{P,t}||_2^1 \qquad MAD = \frac{1}{T}||R_{I,t} - R_{P,t}||_1^1$$

		TE		RMSE		MAD	
Index	**K**	corr	p-classo	corr	p-classo	corr	p-classo
WIG	15	0.457	0.451	0.064	0.063	0.416	0.345
WIG20	15	0.385	0.275	0.053	0.038	0.306	0.228
mWIG40	15	0.666	0.628	0.093	0.087	0.624	0.637

Notes: "corr" stands for the portfolio strategy, which is based on selection of K most correlated stocks with the index. "p-classo" reports the estimates using the post constrained lasso as defined above. Portfolio re-balancing every half-year, 1% transaction costs.

FITTING PERFORMANCE





Notes: the *first graph:* cumulative returns of the WIG20 market index (black), the index funds Lyxor ETF WIG20 (yellow) and the post constrained lasso (purple) where the tuning parameter λ has been selected by means of cross-validation (CV). The *second graph:* comparison between the standard setting of correlation selection criteria (red) and the post constrained lasso (blue), both with a pre-specified number (K) of selected stock.

Index	TE	RMSE	MAD
WIG	0.260	0.037	0.232
WIG20	0.267	0.037	0.225
MWIG40	0.457	0.064	0.468

Fund Name	TE	RMSE	MAD
LYXOR WIG20	0.422	0.059	0.299

Notes: *the first table:* results of post constrained lasso with the tuning parameter selected my means of cross-validation. *the second table:* results of Lyxor ETF WIG20. *TE* stands for the tracking error, *RMSE* the root mean square error and *MAD* the mean absolute deviation. Portfolio rebalancing every half-year, 1% transaction costs.

CONCLUDING REMARKS AND FUTURE WORK

We make a comparison of portfolio management approaches which are based on correlation selection criteria and the constrained post lasso. The second approach is more effective in terms of efficiency of tracking the market indexes. Additionally, The method of P-CLASSO has been compared with the LYXOR WIG20 ETF, which is the only ETF that tracks the WIG20 market index. In the lasso optimization procedure we have used the ten-fold cross-validation method in order to select the tuning parameter. The tracking quality measures are substantially lower for the P-CLASSO model. Further investigations will be needed in order consider the sub-period re-balancing of portfolios, as well as different transaction costs schemes.

R PACKAGES USED

data.table; dplyr; glmnet; hydroGOF; lars; lubridate; PerformanceAnalytics; quadprog; quantmod; xts.

REFERENCES

Efron, Bradley et al. (2004). "Least angle regression". In: *The Annals of statistics* 32.2, pp. 407–499.

Tibshirani, Robert (1996). "Regression shrinkage and selection via the lasso". In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288.