

Electronics and Computer Science
Faculty of Physical and Applied Sciences
University of Southampton

Alejandro Saucedo

29th April, 2014

SPARSE AND GROUP REGRESSION MODELS
IN PORTFOLIO OPTIMIZATION

Project Supervisor: Prof. Mahesan Niranjan

Second Examiner: Prof. Vladimiro Sassone

A final report submitted for the award of
Software Engineering BSc

Abstract

Current approaches to portfolio optimization only consider financial instruments as individual entities, not taking any advantage of the huge amounts of categorizing information available on the underlying financial instruments (i.e. instrument type, industry, sector, volatility level, etc). In this paper we take a novel approach to portfolio optimization, where our main objective is to show that it is possible to exploit categorizing information on the financial instruments in the portfolio and benefit from group correlations present in the data to obtain richer results.

This paper focuses in a very popular branch of portfolio optimization – namely, market index tracking, where the aim is, chosen a Market Index (i.e. a set of high-performing stocks), we want to find a subset that follows the behaviour of its respective Market Index as close as possible. Our approach to solving this problem consists of feature-level regression models with a cardinality (L0-norm) constraint¹, and sparse-inducing group-level regression models². These two approaches will be introduced, analysed and compared in order to provide an insight on the effect group characteristics have when implemented in financial datasets.

Given that the Sparse Group Model analysed in this paper is limited to a single category (i.e. only one category of groups can be taken into consideration) a new regression model was proposed based on our results. This model suggests to take into consideration multiple categories of groups (i.e. type of financial instrument, sector, volatility, etc.) in order to provide more diverse portfolios, and more accurate results.

Finally, the code that was written for the implementation of these models can be found online in a well documented GitHub repository, which has been registered on an Open Source licence and is available at github.com/axsauze/sparse.

¹ Simple Regression Models: Sum of Absolute Values, Sum of Squares, Ridge Regression, Conditional-Value-at-Risk and Lasso.

² Group Lasso and Sparse Group Lasso

Contents

Abstract	2
Contents	3
1 Introduction	1
2 Background and Report of Literature Search	2
2.1 Core Definitions	2
2.1.1 General.....	2
2.1.2 Group Notation.....	3
2.1.3 Index Tracking.....	3
2.2 Feature-Level Regression Models.....	4
2.2.1 CVaR Minimization.....	4
2.2.2 Absolute Error Model (Abs).....	5
2.2.3 Least Squares Minimization (Squares).....	5
2.2.4 L0+L2-norm Model (Ridge Regression).....	5
2.2.5 Lasso Regression.....	6
2.3 Model Selection Approaches.....	6
2.3.1 Feature Selection.....	6
2.3.2 Group Selection.....	6
2.4 Group Regression Approaches.....	7
2.4.1 Group Lasso Regression.....	7
2.4.2 Sparse group lasso.....	7
2.5 Zero-Constrained Group Lasso Application	9
2.5.1 Mathematical Definitions	9
2.5.2 Constraints.....	9
2.5.3 Model Definition	10
3 Implementation	12
3.1 Working Environment.....	12
3.1.1 CVX	12
3.1.2 SPAMS	13
3.2 Implementation Design.....	13
3.2.1 Financial Implementation	13
3.3 Feature-Level Regression Models.....	14
3.3.1 L0-norm constrained models	14
3.3.2 The Lasso sparse inducing model.....	16
3.4 Group-Level Regression Models.....	17
3.4.1 Group Selection.....	18
3.5 Sparse Group Models.....	18
3.5.1 Group Lasso Regression Model.....	19
3.5.2 Sparse Group Lasso	20
4 Analysis	26

4.1	Quantitative Model Comparison	26
4.1.1	Accuracy	26
4.1.2	Sparsity	27
4.1.3	Speed	30
5	Expansion	31
5.1	Multiple Sparse Group Regression (Proposed Concept)	31
5.2	Notation	32
5.3	MSGGR Model definition	34
6	Conclusion	35
6.1	Feature Selection	35
6.2	Group Selection	35
6.3	Lasso Model	36
6.4	Sparse Group Lasso	36
6.5	Effects of Group and Sparse models in Financial Datasets	36
7	References	37