

PhD THESIS

PhD Candidate: *Giuseppe Brandi*

Summary of:

***Decompose et Impera:
Tensor methods in high-dimensional
data***

Keywords:

Tensor methods, Factor analysis, Multilinear regression

PhD in Economics

XXX Cycle

LUISS Guido Carli

Supervisor: Prof. Giuseppe Ragusa

Introduction

This thesis is written with the scope of exploring multiway data. Multiway data, also referred to as tensor data, is a collection of datapoints in multidimensional matrices. At a first glance one may think that these objects are only a convenient representation of a datasets. They are not *just* a collection of data, they have their own structure. For this reason, multiway data need specific models to be correctly analysed. In this spirit, I developed my personal idea on data analysis which can be represented by following statement:

“It is not the data that should fit models, but models that should fit the data”

However, this should not be taken literary I do think that models are important: giving a structure to our techniques is necessary. Nevertheless, I do think that data should be the main driver. This means that instead of trimming data at our necessity to fit existing models, researchers should develop new models to reflect the complexity of the data.

The purpose of this work is to provide an overview of tensor methods applied to Economics and Finance. Yet, the most important aspect of this thesis are ideas and applications rather than the mathematical content. New models are proposed and fitted to data in order to test their performance and get insights from the datasets analysed.

The description of the tensor methods provided in this thesis is not intended to be complete but rather restricted to the model applicable to the analysed data.

Content of the Thesis

As already stated, the main objective of this thesis is to overview existing tensor methods and develop new ones. In particular, the aim is to contribute in two ways. The first, is to link the world of tensor and related models to Economics and Finance, topic not yet well established in the two research communities. The second, more related to the development of new techniques, is to introduce two new tensor methods, namely Tensor Autoregression and the Slice-diagonal tensor Decomposition (SDT). Even if these methods are applied to economic data, their applicability remains general. This thesis is intended to be a starting point for researchers who want to

have an overview of some Tensor methods applied to Economics and Finance. In the remaining part of this introduction, I will explain the history and future directions of the tensor data analysis. Chapter 2 is devoted to the Tensor Autoregressive model while Chapter 3 to the analysis of correlation and covariance matrices over time with the help of the Slice-Diagonal Tensor decomposition. A brief summary of the two chapter is provided below:

Chapter 2: Tensor Autoregression in Economics and Finance

Multidimensional data (tensor data) is a relevant topic in statistical and machine learning research. Given their complexity, such data objects are usually reshaped into matrices and then analysed. However, this methodology presents other drawbacks. First of all, it destroys the intrinsic interconnections among datapoints in the multidimensional space. Secondly, the number of parameters to be estimated in a model increases exponentially. To alleviate these issues, we treat the data as it is and we develop a model able to deal with the multidimensionality of the dataset. In particular, a parsimonious tensor regression (in which both the regressor and the response are tensors) is build such that it retains the intrinsic multidimensional structure of the dataset. Tucker decomposition is employed to achieve parsimony and an ALS algorithm is developed to estimate the model parameters. A simulation exercise is produced to validate the model. In order to compare it with the existing model in a forecasting setting, an empirical application to Foursquares spatio-temporal dataset and macroeconomic time series is performed.

Chapter 3: Latent stock correlation projection via tensor decomposition

Correlation matrices are ubiquitous in financial statistics both in research papers and in the industry. The correlation is the most used method of association between stock returns. Portfolio allocation, risk management and network analysis are based on correlation matrices. Applied researchers debate about the correct correlation matrix to use. In this regard, one of the issues is related to the choice of the sample period used in the reference models. This is because the true correlation is not available. Several approaches exist in order to handle this problem; yet, all of them are heavily dependent on the time horizon used. In this paper, we introduce a method which

aims to alleviate the problem of true correlation availability by decomposing the time series of correlations into a static and a dynamic component. The static component encapsulates the latent, long-run behaviour of the correlations while the dynamic part represents the time dependent structure in the latent space. Hence, the latent correlation matrix projected by the tensor factorization can be plugged in the models as an alternative to the standard correlation matrix. We will show that the hidden correlation is empirically almost time invariant.

Conclusions

In this thesis, I explored some of the tensor methods available in the literature. I studied their performance and analysed the insights on the data they provide. I also proposed two novel tensor methods, namely the Tensor (Auto)regression, in which both the dependent variable and the independent variable are tensors. When compared to other models, this new methodology has proven to have satisfactory performance both in sample and in forecasting. In the second paper, I studied the time series of covariance and correlation matrices. In performing this analysis, I also introduce a decomposition approach I named Slice-diagonal tensor decomposition. This approach is at the crossroad between Tucker and PARAFAC decompositions, being a special case of the first and a generalization of the second. It is more parsimonious than the Tucker model but more flexible than the PARAFAC. In the paper, it proved to perform well in terms of fit. In the same paper, I studied the dynamic and static factors induced by the tensor factorization finding similarities with real world volatility and correlation indexes.