Journal of Economic Dynamics & Control 32 (2008) 3939–3959

Contents lists available at ScienceDirect

Journal of Economic Dynamics & Control

journal homepage: www.elsevier.com/locate/jedc

Emergent and spontaneous computation of factor relationships from a large factor set

Zitian Wang [a,*], Lili Wang [b], Shaohua Tan [a]

[a] State Key Laboratory of Machine Perception, Center for Information Science, Peking University, Beijing 100871, PR China
[b] Department of Accounting and Finance, School of Economics, Finance and Management, University of Bristol, Bristol BS8 1TN, UK

A R T I C L E I N F O

Article history:
Received 31 March 2006
Accepted 21 April 2008
Available online 14 May 2008

JEL Classification:
C49
C51
G12

Keywords:
Asset pricing
Financial factor analysis
Bayesian Network
Bayesian inference

A B S T R A C T

We propose a systematic factor analysis approach using the Bayesian Network (BN) framework by taking great advantage of the information conveyed in the large amount of financial data, as well as experts' personal insight and additional evidence. First, we build the BN for a large set of macroeconomic and firm-specific financial factors, describing factor interrelationships in the market. As a subgraph of the learned BN structure, a compact set of influential factors for return are extracted which jointly have high influence on the rate of return and low mutual redundancy among themselves. Then in the individual firm analysis, based on the general model of the market, additional information like human expert judgment can be incorporated in the model to conduct the specific and concrete individual firm study. Empirical results show the efficiency and flexibility of our method.

1. Introduction

Discovering factor relationships is one of the fundamental issues in financial analysis. Researchers are especially interested in identifying factors that affect the rate of return on assets. There are quite a few influential pricing kernels developed in the past such as CAPM of Sharpe (1964), Lintner (1965), and Black (1972), three-factor model of Fama and French (1996), dividend yields model of Naranjo et al. (1998), consumption-based models of Lucas (1978) and Breeden (1979), Jacobs and Wang (2004) and Lettau and Ludvigson (2001, 2005), etc. With theoretical and empirical support,

* Corresponding author. Tel.: +86 10 62755745.
E-mail address: wzt@cis.pku.edu.cn (Z. Wang).

0165-1889/$ - see front matter
doi:10.1016/j.jedc.2008.04.005

these various pricing kernels describe the risk–return relationship from different aspects. Despite the intellectual implications of factors in these models, by confining the study to a small set of factors, the key relationships which may be buried in an extended set of factors beyond the current choice tend to be overlooked. It has been noted by Ludvigson and Ng (2007) that the model may be highly misleading if investors have additional information not fully conveyed in the chosen conditioning variables.

There are various financial and macroeconomic factors containing valuable information from different economic levels. There has been growing interest in the study of large dimensionality problems. The dilemma that researchers constantly face is that they are forced to select, among large number of economic factors, a small candidate set of factors which may subjectively seem to have strong influence on the asset return and then carry on a statistical trial to determine their relevance. It has been an open question to develop a systematic approach that spontaneously discovers the buried factor relationships directly among a large set of factors.

While modeling the risk–return relationship, traditional methods just concentrate on the risk's role on return while usually disregarding the interrelationships of the risk factors. Traditional independence assumption among the dependent variables is usually too strong in reality especially when the studied factor set is large, e.g., with hundreds of various factors included. Moreover, in the study of financial or economic prospect, there are few mechanisms that incorporate financial analysts' subjective judgment which can be very valuable for individual firm everyday analysis. Most existing methods just take the quantitative information into model and neglect the qualitative aspect, which make the model not flexible enough.

In this paper, we propose a systematic factor analysis approach using the Bayesian Network (BN) framework. First, we build the general factor-relationships of the market based on the data of a large set of macroeconomic and firm-specific financial factors. The hierarchical *macroeconomic-firm_specific-return* relationships are represented in the BN graphical structure with the probabilistic distribution of each factor specified. Because the identification of strong influential factors of return is of primary interest, we study the subgraph related to the asset return in detail. With the mutual information based minimal-redundancy-maximal-relevance (mRMR) criterion (Peng et al., 2005), factors with largest joint relevance to return are selected, and meanwhile the information redundancy of the selected factors are minimized.

Then in the individual firm analysis, based on the general model of the market, additional information like human expert judgment can be incorporated in the model to conduct the specific and concrete individual firm study. The BN inference mechanism will make the analysis work well even with only imprecise and partial information about the economy and the firm.

It is believed that there are three key novelties in our approach as compared to earlier work. First, our approach takes great advantage of information conveyed in the large amount of financial data, as well as experts' personal insight and additional evidence. By adopting the BN structure learning algorithm, the factor interrelationships are modeled in the graphical structure, representing the information from the data. Shenoy and Shenoy (2000) used BNs for modeling portfolio returns. However, the shortcoming of their work lies in that the BN structure in their setting was totally specified by a subjective choice. For small scaled models with only a few factors included (like 9 in Shenoy and Shenoy, 2000), financial experts are competent for specifying the factor interrelationships. But as the dimension becomes large (like hundreds of factors in our setting, or more in other complicated problems), the relationships become unclear even for experts. It will be very difficult to construct the graphical structure only from human knowledge. By learning the structure from data, useful information can be extracted with legible representation of factor interactions. From the other aspect, after the general model is given, in the individual firm analysis with various concrete everyday situations, the BN model can be updated from new information like new observations of the firm data or analysts' subjective judgment. In everyday analysis where imprecise and incomplete evidence is very common, this mechanism makes the model flexible and practical for individual firm monitoring and analysis.

Second, our approach extracts a compact set of influential factors of return from compound large factor sets. The selected factors jointly have high influence on the rate of return and low