Journal of Multivariate Analysis 154 (2017) 234–248

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Bayesian regularized quantile structural equation models

Xiang-Nan Feng [a], Yifan Wang [a], Bin Lu [b], Xin-Yuan Song [c,*]

[a] Department of Statistics, the Chinese University of Hong Kong, Shatin, New Territories, Hong Kong
[b] Institute of Economics and Finance, Nanjing Audit University, Nanjing, Jiangsu, China
[c] Shenzhen Research Institute & Department of Statistics, the Chinese University of Hong Kong, Hong Kong

ARTICLE INFO

Article history:
Received 1 March 2016
Available online 17 November 2016

AMS subject classifications:
62H25
62F15
62J07

Keywords:
Bayesian Lasso
Markov chain Monte Carlo methods
Quantile regression
Structural equation model

ABSTRACT

Latent variables that should be examined using multiple observed variables are common in substantive research. The structural equation model (SEM) is widely recognized as the most important statistical tool for assessing interrelationships among latent variables. As a recent advancement, Bayesian quantile SEM provides a comprehensive assessment of the conditional quantile of the response latent variables given the explanatory covariates and latent variables. In this study, we develop Bayesian least absolute shrinkage and selection operator (Lasso) and Bayesian adaptive Lasso procedures to conduct simultaneous estimation and variable selection in the context of quantile SEM. We propose the use of the Markov chain Monte Carlo method to conduct Bayesian inference. Various features, including the finite sample performance of the proposed procedures, are validated through simulation studies. The proposed method is applied to investigate the determinants of the capital structure of Chinese-listed companies.

1. Introduction

The structural equation model (SEM) is the most important tool for analyzing the interrelationships among latent variables that are measured using multiple correlated observable indicators [5,46]. SEM has two components, namely, the measurement equation that characterizes latent variables via a confirmatory factor analysis model and the structural equation that assesses the effects of explanatory latent variables on outcome latent variables via a regression model. In conventional SEMs, both components are mean regression-type models, and therefore, are vulnerable to outliers or non-normality of the response distributions. Moreover, the effects of explanatory observable and latent variables on the upper or lower tails of the response distribution are of particular interest in certain circumstances. Such tail-related behavior and relationships cannot be adequately presented using conventional SEMs. A promising direction for addressing the aforementioned problems is the use of quantile SEM that introduces quantile regression techniques into the SEM framework.

In recent years, quantile regression has become increasingly popular in the fields of finance, econometrics, medical sciences, and ecology, and it has been widely extended to a variety of complex models [3,10,22,30,50] and multivariate quantiles [20,21]. In contrast to mean regression, which solely summarizes the average of the conditional distribution, quantile regression utilizes different percentage points of the distribution and computes several regression lines. Therefore, the distinct effects of explanatory variables on the response at various quantiles can be fully elucidated. In addition, quantile regression exhibits high robustness [27]. As a special case, median regression that estimates the 50% conditional quantile

* Corresponding author.
E-mail addresses: fengxiangnan123@gmail.com (X.-N. Feng), yifan@link.cuhk.edu.hk (Y. Wang), b_lu@nau.edu.cn (B. Lu), xysong@sta.cuhk.edu.hk (X.-Y. Song).

http://dx.doi.org/10.1016/j.jmva.2016.11.002

is a popular robust regression method. Despite the appealing features of quantile regression, it is seldom developed in the context of SEM. One exception is the study of Wang et al. [54], which considered a quantile SEM in a Bayesian framework. However, this work focused solely on estimation, and the important issue of model selection was not addressed.

The selection of a suitable model is a crucial part of building an effective quantile regression model, given that the effects and significance of the predictors may not be the same for different quantile levels. Therefore, a small subset of important explanatory variables should be identified to obtain a parsimonious model, and a reasonable interpretation for each quantile level is necessary. Among the studies on variable selection in Bayesian quantile regression, stochastic search variable selection (SSVS) methods [1,25] and regularization methods with ℓ_1 shrinkage, including the least absolute shrinkage and selection operator (Lasso) and adaptive Lasso, play a dominant role. SSVS introduces a probability distribution on the set of all possible models in the model space by adding an extra indicator parameter for each regression coefficient, so that models with regression coefficients that are substantially deviant from zero are allocated with high posterior probabilities. Although SSVS is powerful, it is computationally demanding when model space is large. Moreover, computational instabilities can be elicited by the varying dimensional parameter space in implementing the SSVS algorithm, particularly when vague or non-informative priors are applied [18,28]. Therefore, we focus on the relatively stable and computationally efficient regularized methods in this study. Lasso was initially elaborated in the seminal work of Tibshirani [52] to simultaneously estimate and select important predictors in a linear model. Koenker [29] first used regularization in quantile regression, which imposed ℓ_1 penalty on random effects in a mixed-effect quantile regression model. Thereafter, Li and Zhu [38] developed the Lasso estimator for general linear quantile regression models. However, one problem with the Lasso estimator is its inconsistency in certain conditions, and thus, it may suffer from appreciable bias [13,55,58]. To address this issue, Zou [58] proposed adaptive Lasso by adapting penalties for different predictors. Wu and Liu [56] then investigated the adaptive Lasso procedure in quantile regression. Recently, the Bayesian versions of Lasso (BLasso) and adaptive Lasso (BaLasso) procedures have been proposed to fulfill the requirements of Bayesian methods to analyze complex models [36,42]. Motivated by the convenience of Bayesian methods, Li et al. [37] and Alhamzawi et al. [2] applied BLasso and BaLasso, respectively, to the Bayesian quantile regression model and obtained good results. Nevertheless, model selection has never been considered in quantile regression models with latent variables. In this study, we introduce BLasso and BaLasso into quantile SEM, which is the first attempt for this model.

This study identifies the potential determinants of the capital structure of Chinese listed companies. As the leading developing economy of the world, China attracts increasing attention from international capital markets. Chinese companies are expected to have features that are distinct from those of companies in Western countries. Thus, the capital structure, particularly the debt structure and its determinants of Chinese listed companies are of significant interest. We consider a data set collected from the annual reports of over 800 companies listed in the Shanghai and Shenzhen Stock Exchanges. The outcome of interest is debt. The potential determinants include non-debt tax shields, growth, size, profitability, liquidity, ownership, tangible assets, operational risk, and signal. However, the dependent variable debt is a latent trait that should be simultaneously assessed by the total debt ratio (TDR) and the short-term debt ratio (SDR). Similarly, several potential determinants, such as non-debt tax shields, growth, size, profitability, liquidity, and ownership, are also latent traits. Each of these traits is assessed using two or more correlated indicators from different perspectives (details are provided in Table 5). Thus, we adopt the SEM approach to manage latent variables and examine their interrelationships. Furthermore, traditional analyses [8,9,26] of the determinants of the capital structure of Chinese firms have focused only on central tendency. However, the effects of the determinants may vary across regression lines that relate to different conditional quantiles of the response. The determinants that are significant at the upper or lower tails of the conditional response distribution are also important factors for explaining the capital structure of Chinese listed companies. To address the aforementioned problems, we propose the regularized quantile SEM for the rigorous treatment of the latent dependent and explanatory variables, as well as for the simultaneous estimation and selection of important determinants across different conditional quantiles of the response latent variable. The analysis presented in Section 5 shows that the determinants exhibit different patterns across quantiles, thereby providing useful insights into determining which determinants are truly important and in what manner are they important. To our knowledge, real data analysis is the first application of the quantile regression-type model to Chinese firms to elicit a comprehensive view of the effects of potential determinants on the capital structure of firms.

The remainder of this paper is organized as follows. The quantile SEM is described in Section 2. In Section 3, we provide a brief review of regularization and quantile regression in the Bayesian framework and introduce the Bayesian modeling for the regularized quantile SEM. In Section 4, we examine the performance of the proposed methods through extensive simulation studies. In Section 5, real data analysis regarding the determinants of the capital structure of Chinese listed companies is presented. The paper is concluded with a brief discussion in Section 6. In addition, the technical details are provided in the Appendix.

2. Quantile structural equation model

The quantile structural equation model consists of two components, namely, the measurement equation and the structural equation. The measurement equation is a median regression-type confirmatory factor analysis model that groups the observable variables $\mathbf{y}_i$ $(p \times 1)$ into a potentially smaller number of latent factors $\boldsymbol{\omega}_i$ $(q \times 1)$ as follows:

$$\mathbf{y}_i = \mathbf{A}\mathbf{c}_i + \boldsymbol{\Lambda}\boldsymbol{\omega}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, n, \tag{1}$$