Expert Systems with Applications 42 (2015) 554–561

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Uniform accrual generating process grouping with self-organizing maps

Henrik Höglund *

Hanken School of Economics, Biblioteksgatan 16, 65101 Vasa, Finland

ARTICLE INFO

Article history:
Available online 19 August 2014

Keywords:
Earnings management
Discretionary accruals
Self-organizing maps
Accrual generating process

ABSTRACT

Most earnings management and earnings quality studies rely on various types of discretionary accrual estimation models. Common assumptions when using these models is that the accrual generating process (AGP) is stable over time or that firms within the same industry have similar AGPs. These assumptions have, however, been challenged in a number of studies. Instead, it has been suggested that AGP is depicted by various accrual determinants and that firms should be grouped according to similarities in the AGP. The purpose of this study is to develop and assess the performance of a self-organizing map (SOM) local regression-based discretionary accrual estimation model. Overall, the results show that the SOM local regression model outperforms previously suggested discretionary accrual estimation models. For example, the detection rate of simulated earnings management for the SOM local regression model is almost twice the detection rate of the commonly used cross-sectional Jones model. In addition to outperforming previously suggested models, the SOM local regression model also gives a visual representation of the AGP of a specific firm in relation to other firms.

1. Introduction

Most earnings management and earnings quality studies rely on various types of discretionary accrual estimation models. A crucial requirement for these models is that they capture the magnitude of earnings management without error. Some studies have cast doubt on the earnings management detection ability of commonly used discretionary accrual estimation models. For example, Thomas and Zhang (2000) showed that the simple assumption that the expected total accruals for all firms equal −5% of lagged total assets outperforms most models. Other studies, such as Young (1999) and Bartov, Gul, and Tsui (2000), have also reported similar findings. The discretionary accruals estimation models are based on various assumptions regarding the accrual generating process (AGP). Typically, it is assumed that the AGP is stable over time or that firms within the same industry have similar AGPs. The results from Dopuch, Mashruwala, Seethamraju, and Zach (2012), however, challenge these assumptions. Instead, they argue that the AGP is best described with various accrual determinants. Thus, they conclude that researchers should focus on estimating the discretionary accrual models in groups that have as similar AGPs as possible.

One alternative approach to improve the discretionary accrual estimation models is to model the AGP by augmenting the models with accrual determinant variables. This approach is, however, problematic as there is a strong collinearity between the accrual determinants as shown by Dopuch et al. (2012). Furthermore, a number of studies have shown that the AGP is non-linear (e.g. Dechow, Sloan, & Sweeney, 1995; Jeter & Shivakumar, 1999; Kothari, Leone, & Wasley, 2005) which makes it difficult to model using linear regression. Relatively few studies have addressed the collinearity and non-linearity issues with discretionary accrual estimation models. Höglund (2012) showed that multilayer perceptrons (MLPs) and general regression neural networks (GRNNs) outperforms OLS regression when used with discretionary accrual estimation models. In another study Höglund (2013) showed that fuzzy linear regression (FLR) outperforms OLS regression when only short time-series of data are available. Alternatively, to address the issues with collinearity and non-linearity between the variables depicting the AGP a self-organizing map (SOM) local regression approach could be considered. That is, the SOM would be used for grouping the firms based on the AGP and local regression models would be estimated using firms with similar AGPs. The SOM has several appealing features, such as being able to model non-linear relationships and not assuming a priori knowledge of the data distribution. The purpose of this study is to develop and assess the performance of a SOM local regression-based discretionary accrual estimation model. In previous studies where SOM local regression models have been used a common approach has been to build the SOM using the regression model variables. In this study the SOM is built using various accrual determinants suggested by Dechow, Kothari, and Watts (1998) and Dopuch et al. (2012)

* Tel.: +358 (0)40 3521768.
E-mail address: henrik.hoglund@hanken.fi

http://dx.doi.org/10.1016/j.eswa.2014.08.010
0957-4174/

whereas the local regression models are based on a current accruals version of the discretionary accruals estimation model suggested by Jones (1991).

The remainder of this study is organized as follows. The basic operating principle of the linear regression-based accrual models and the accrual generating process are covered in Section 2. In Section 3 an overview of self-organizing maps and SOM local regression models are given. The research design is presented in Section 4 and the results from the empirical study are presented in Section 5. Section 6 concludes the study.

2. Discretionary accrual estimation models

Discretionary accrual estimation models are an essential tool for measuring earnings management and earnings quality. A large number of different models have been suggested but most discretionary accrual estimation models are derived from the linear regression model suggested by Jones (1991). In the Jones model the reciprocal of lagged total assets, the change in sales and property, plant and equipment are regressed on total accruals. Jones used a firm-specific time-series prior to the event year to estimate the model regression coefficients. Another, more common, approach for estimating the regression coefficients is the cross-sectional approach where firms from the same year and industry are used (DeFond & Jiambalvo, 1994).

Total accruals are calculated either using the balance sheet approach or the cash flows approach. With the balance sheet approach total accruals equals the change in working capital accruals minus depreciation and amortization. Using the cash flows approach, total accruals are calculated by subtracting cash flows from operations from earnings before extraordinary items. The purpose of the discretionary accruals estimation models is to decompose the total accruals into discretionary accruals and non-discretionary accruals. The discretionary accruals are considered a proxy for earnings management whereas the non-discretionary accruals constitute expected or normal accruals. To calculate the non-discretionary accruals, the regression coefficients are first estimated using either a pre-event year time-series or cross-sectional industry matched data. The coefficients are then used together with data for the firms for which earnings management is examined. Once the non-discretionary accruals have been calculated, the discretionary accruals are obtained by subtracting the non-discretionary accruals from the total accruals.

2.1. Accrual generating process

Understanding the underlying AGP is essential when estimating the discretionary or abnormal accruals. That is, one needs to understand the innate determinants that drive accruals. Typically, it is assumed in discretionary accrual estimation models that normal or expected accruals are driven by a change in sales and by the level of property, plant and equipment (Jones, 1991). Also other determinants for accruals, such as return on assets (Kothari et al., 2005) and cash flows from operations (Kasznik, 1999; Rees, Gil, & Gore, 1996), have been suggested in previous studies. Furthermore, common assumptions are that the AGP is stable over time (Jones, 1991) and that the AGP is uniform within industries (DeFond & Jiambalvo, 1994).

Total accruals comprise both current and non-current accruals. Thomas and Zhang (2000) showed that most of the variance in total accruals comes from the variance in current accruals and that the non-current accruals are mainly driven by the level of property, plant and equipment. This implies that the variance in current accruals is mainly driven by the change in sales. Dopuch et al. (2012) argue, however, that the relationship between the change in sales and accruals is more complex than typically modeled. They suggest that it is the change in sales together with firm-specific accrual determinants that best describe the AGP. The accrual determinants examined by Dopuch et al. are receivable turnover, payable turnover, inventory turnover and profit margin. They find that all four determinants are significantly associated with the change in sales and that the correlations have the expected signs. Dechow et al. (1998) also report similar results when examining the relationship between the change in sales and accrual determinants. Furthermore, Dopuch et al. show results indicating that the degree of homogeneity in the accrual determinants is positively related to the performance of the discretionary accrual estimation models. In other words, the more uniform the AGP among the data set firms, the better the performance of the models. Finally, Dopuch et al. also demonstrates that the common assumption of a uniform AGP within an industry does not hold.

3. Self-organizing maps

The SOM is a type of neural network based on competitive, unsupervised learning. The operating principles of the SOM were first introduced by Kohonen (1982). The purpose of the SOM is to represent multidimensional data in lower, typically one or two, dimensions. One feature of the SOM is that it is topology preserving. That is, observations close to each other in the input data are mapped to adjacent nodes in the SOM. This feature makes the SOM a useful tool for data visualization and cluster analysis. The SOM has also generalization capability as it can characterize data that has not been used in the training process.

A typical SOM is made up of a two-dimensional lattice of nodes where each node is connected to each variable in the input layer (see Fig. 1). The SOM comprises an n dimensional input vector $(V_1, V_2, \ldots, V_n)$ and each node in the SOM is connected to the input vector via an n dimensional weight vector $(w_1, w_2, \ldots, w_n)$. Before initiating the training process a number of parameters need to be set. The first two parameters are the shape and size of the SOM. Concerning the shape of the SOM, Kohonen (2001) recommends the use of a hexagonal lattice rather than a rectangular lattice and a rectangular map rather than a square map. The number of nodes in the map depends on the number of observations in the estimation data set and the intended use of the map. Other parameters that also need to be defined are the initial learning rate and neighborhood radius. The initial value of the learning rate should be close to one and the initial neighborhood radius can be more than half of the SOM diameter (Kohonen, 2001).

The first step in the training process is to initialize the weight vectors. The weights are usually initialized using small random values (Van Laerhoven, 2001). Once the weights have been

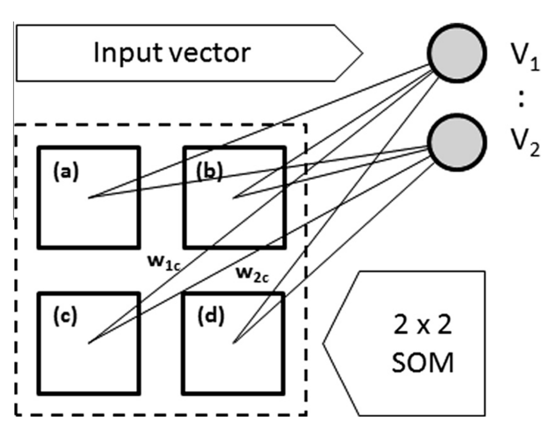

Fig. 1. Two-dimensional SOM and input layer with two variables.