J. Intell. Syst. **20** (2011), 147–166
DOI 10.1515/JISYS.2011.009

Use of Stability and Seasonality Analysis for Optimal Inventory Prediction Models

Peng Zhang, Manish Joshi and Pawan Lingras

Abstract. Inventory prediction and management is a key issue in a retail store. There are a number of inventory prediction techniques. However, it is difficult to identify a time series prediction model for inventory forecasting that provides uniformly good results for all the products in a store. This paper uses data from a small retail store to demonstrate the variability of results for different modeling techniques and different products. We demonstrate inadequacy of a generic inventory model. Stability and seasonality analysis of the time series is used to identify groups of products (local groups) exhibiting similar sales patterns. Different clustering techniques are applied to determine reasonable local groups. With the help of Mean absolute percentage error (MAPE), the effectiveness of dataset partitioning for better inventory management is demonstrated. Appropriate inventory management strategies are proposed based on the stability and seasonality analysis.

Keywords. Time series, clustering, prediction, inventory management.

2010 Mathematics Subject Classification. 62H30, 68T10.

1 Introduction

Profitability of a retailing outfit depends upon maintaining an optimal level of inventory. A store wants to minimize the amount of inventory on hand while ensuring that every customer order is fulfilled. In order to achieve this goal, accurate inventory forecasting plays a crucial role. Inventory forecasting predicts the future inventory demand based on historical and current demand [14]. The demand for a particular item or a group of related items can be considered as a time series [13]. Time series prediction, which predicts future values of a time series, plays a critical role in forecasting quantity demand in business operations [6, 13]. Regression analysis, neural networks, exponential smoothing and autoregressive integrated moving average (ARIMA) are some of the widely used time series prediction techniques in inventory management [1–4, 7, 9, 10, 16, 17]. Many researchers focus on finding a generic forecasting solution for all the products. However, products are distinguished by their seasonal sales patterns and volatilities in sales demand. One generic solution may not always be able to predict optimal quantity demand for each product. Two illustrative case studies are presented in the paper to support this claim.

We propose an approach that uses time series clustering and time series prediction techniques to forecast future demand of each product in an inventory management system. We use algorithms such as K-Means [11] and EM [5], to categorize products into several reasonable groups based on product sales patterns. Furthermore, an optimal forecasting solution for each product is identified by comparison and evaluation of inventory forecasting models. The models are compared using one of the statistical indicators, mean absolute percentage error (MAPE) [13].

The remaining paper is organized as follows. Section 2 presents related work while Section 3 describes the details of a dataset used for experiments. Two case studies illustrated in Section 4 that underline limitations of generic time series prediction modeling. Section 5 elaborates procedural steps to accomplish stability and seasonality analysis. Section 6 demonstrates computations of weekly, monthly stability analysis and quarterly seasonality analysis. The results are summarized and discussed in Section 7 followed by conclusions in Section 8.

2 Review of Inventory Prediction Models

Researchers and practitioners have proposed different models for inventory management. These models are based on experiments with a variety of time series prediction techniques. But use of clustering to group products exhibiting similar sales pattern followed by an appropriate time series prediction model is experimented by very few researchers.

Mehrez et al. [12] in their review analyzed results of 16 various forecasting models which were clustered along four defined dimensions. Mehrez et al. found that none of the forecasting methods have demonstrated the expected advantage. They recommended more specific knowledge as a prerequisite to assign a forecasting method to specific data. In our approach we propose to form cluster of products before applying any forecasting model.

In a related work proposed by Stefanovic et al. [15] clustering is used as a basis for supply chain inventory forecasting. In order to obtain out-of-stock forecasts at the store/product level clustering is used as a first phase followed by decision trees and neural network mining algorithms. Stores that have similar aggregate sales patterns across the chain are clustered in this case whereas we propose to concentrate on forming local groups of products based on their sales pattern.

Similar to our work, Kumar et al. [8] also show the importance of clustering products prior to prediction. While they develop a single model for the entire cluster, we actually identify the best techniques for each cluster. Our clustering is also more versatile. Kumar et al. use clustering to identify the seasonality of the products. We use seasonality as well as stability. The stability allows us to identify the best prediction period for each product. We also identify products that do not need any prediction, since they are only sold in single or pair of periods.

3 Experimental Data Set

The experimental data set is obtained from an independent small retail chain of specialty stores. As a retail chain, the data set captures information on customers, products, suppliers, and business operations from January 2005 to September 2007. According to the data set, more than 177,000 sales transactions are made in 33 months. Sales revenue, number of products, and number of customers are included in an overview of annual business operations as shown in Table 1.

Attribute	2005	2006	2007 (9 months)	Total
Sales revenue	$1,574,079	$3,885,394	$3,278,189	$8,737,662
Number of products	5782	7567	7587	10841
Number of customers	3824	9818	9371	20812

Table 1. An overview of annual business operations.

In total, there are 20,812 distinct customers, 10,841 different products and $8,737,000 sales revenue over two years and nine months. The sum of the products over three years is significantly less than total number of distinct products over the three years. This observation seems to suggest that there was a reasonable amount of overlap between the products between three years. However, the sum of the number of customers in each year is approximately the same as the total number of distinct customers over these three years. That means the number of customers that visited in each of the three years was relatively small.

4 Inadequacy of a Generic Model

A simple solution to inventory prediction is to use the same model for predicting demand for each product in the store. We will refer to such a model as a generic model. The experiments reported in this section demonstrate why a generic model could not generate an optimal solution for certain products. The first subsection explains how to identify a generic model for the dataset, while the second subsection presents two case studies that highlights limitations of a generic inventory model.

4.1 Deriving a Generic Model

Some of the popular time-series prediction techniques such as Simple Exponential Smoothing, Holt's Exponential Smoothing, Brown's Exponential Smoothing, Damp Trend Exponential Smoothing, and Autoregressive Integrated Moving Average (ARIMA) inventory forecasting models are applied on the dataset. Mean absolute percentage error (MAPE) is one of the frequently used statistical evaluation metrics. It is applied to evaluate various inventory forecasting models used in this study.

We define a time series prediction technique as the best-fit solution for a product if it has the lowest MAPE compared with other time series prediction techniques. A generic optimal solution is the time series prediction model that most frequently appeared to be the best-fit solution in the entire product set. We compared frequen-

Year	2005		2006	
Level	Monthly	Quarterly	Monthly	Quarterly
Simple Exponential Smoothing	688	765	958	991
Holt's Exponential Smoothing	948	**1703**	1200	**2009**
Brown's Exponential Smoothing	580	1077	1085	1586
Damp Trend Exponential Smoothing	**1771**	914	1922	1279
ARIMA	1750	1370	**2360**	1788
Generic optimal solutions	**Damp Trend Exponential Smoothing**	**Holt's Exponential Smoothing**	**ARIMA**	**Holt's Exponential Smoothing**

Table 2. Generic solutions in 2005 and 2006.

cies of the best-fit solutions' occurrences to determine generic optimal solutions. Two levels of inventory forecasting are performed to find generic optimal solutions at month and quarter level. Due to the large number of computations, inventory forecasting is not performed at week level.

Table 2 illustrates generic optimal solutions at month and quarter level in 2005 and 2006. It compares frequencies of the best-fit solutions' occurrences. For example, in 2005, the occurrence frequency of Damp Trend Exponential Smoothing (1771) is the highest. That is, it is the best-fit solution for 1771 products. Thus, Damp Trend Exponential Smoothing is the generic optimal solution for the entire product set at month level in 2005. Autoregressive Integrated Moving Average (ARIMA) is the generic optimal solution at month level in 2006 since it is the best-fit solution for the most number of products (2360). The generic optimal solution at quarter level in 2005 is Holt's Exponential Smoothing, and the same is repeated as the generic optimal solution in 2006.

4.2 Test Cases

Following two case studies reveal whether derived generic model is really able to outperform all remaining models for various groups of products.

Case-1: The generic optimal solution at quarter level is Holt's Exponential Smoothing in 2005, as shown in Table 2. However, Brown's Exponential Smoothing is the optimal solution for a typical group (named G-05Q20) consisting of 135 products. Brown's Exponential Smoothing is the best-fit solution for all products within this group. The comparison of MAPE values of time series prediction models for products of the group G-05Q20 is displayed in Table 3. MAPE values associated with Brown's Exponential Smoothing (7.50035) are lower than Holt's Exponential Smoothing values (7.541999), which is the generic optimal solution.

All products in this group G-05Q20 exhibit similar sales pattern and the corresponding details are presented in Section 7.

Simple Exponential Smoothing	Holt's Exponential Smoothing	Brown's Exponential Smoothing	Damp Trend Exponential Smoothing	Autoregressive Integrated Moving Average
7.716615	7.541999	**7.50035**	7.541548	7.6875

Table 3. MAPE comparison results for products in a group G-05Q20.

Case-2: In case-1 the usefulness of a generic model was tested. Furthermore, this case study indicates whether use of any inventory forecasting model is at all required for products with strong sales pattern. For example, products that are sold same number of times in each quarter forms a stable quarterly group. A simple solution is to keep these products at the stable quantity all quarters. Similarly, seasonal sales groups may also have very strong sales patterns. A closer look at our dataset uncovers the fact that some products were only sold in July in 2006 (group G-06M60). The MAPE comparison results for these products shows high MAPE values for all the prediction techniques (Table 4). It suggests that applying any prediction models to forecast products with strong sales pattern might produce incorrect predictions.

Simple Exponential Smoothing	Holt's Exponential Smoothing	Brown's Exponential Smoothing	Damp Trend Exponential Smoothing	Autoregressive Integrated Moving Average
9.248119	9.247283	9.243259	9.249918	**9.243056**

Table 4. MAPE comparison results for products in group G-06M60.

5 Identification of Sales Patterns

Products have their own sales patterns. Some products are sold throughout a whole year, while others are sold seasonally. Moreover, some products may share the same patterns. Seasonal selling distributions are clearly shown for some products. Categorizing products, based on their sales patterns, may ease inventory management.

Following description shows how to discover products that are mostly sold in certain month(s) or quarter(s). Products are considered as single period selling products if their selling ratios in one period are high. Here, the ratio is calculated as (1):

$$\frac{pk}{P - pk}, \tag{1}$$

where pk is the sales quantity in period k, P is the total sales quantity throughout the year. The ratio is considered high if its value is greater than 10. Table 5 illustrates the distribution of single quarter selling products. For instance, in 2006, 800 products were mostly sold in the first quarter, 336 products are mostly sold in

Quarters	Quarter 1	Quarter 2	Quarter 3	Quarter 4
Number of products (2005)	478	292	332	662
Number of products (2006)	800	336	306	579

Table 5. The distribution of single quarter selling products.

the second quarter, etc. Table 6 illustrates the distribution of single month selling products.

Inadequacy of generic inventory prediction model to fit best to entire dataset and the information of products' sales pattern provide a strong motivation for a better combined inventory management strategy that includes in-depth analysis of dataset prior to application of any time series prediction model. The study proposes to use time series clustering techniques, using algorithms such as K-Means and EM, to categorize products into several reasonable groups based on product sales patterns.

Months	Number of products (2005)	Number of products (2006)
Jan	132	491
Feb	156	93
Mar	106	99
Apr	79	110
May	96	76
Jun	74	105
Jul	104	85
Aug	92	91
Sep	91	81
Oct	108	92
Nov	115	127
Dec	300	188

Table 6. The distribution of single month selling products.

Our objective is to identify groups of products with similar sales pattern. The first subsection describes stability analysis that groups products whose sales quantities are approximately the same for a certain period (week, month or quarter) throughout a year. Products which are not stable for any period are unstable products. The next subsection describes seasonality analysis that categorizes the unstable products into groups based on their monthly and quarterly sales pattern.

5.1 Stability Analysis

Products are called stable if their sales quantities change negligibly over allotted periods. On the other hand, products are defined as unstable if their sales quantities change greatly with time. The products are categorized based on their stabilities of sales quantity in three levels: weekly, monthly and quarterly. This stability analysis is carried out from “the least” to “the greatest”, that is, the experiment starts with weekly sales quantities analysis, and then follows by monthly and quarterly sales quantities analysis. K-Means and EM algorithms are applied to the three level analyses. The possible categorizing results are weekly stable products, monthly stable products, quarterly stable products, and unstable products.

Each level of analysis categorizes products into two groups: a stable and an unstable group. Moreover, stable weekly products can be considered stable monthly products, and stable monthly products can be considered stable quarterly products, and so on. Thus, we perform monthly stability analysis on unstable weekly products and quarterly stability analysis on unstable monthly products. Finally, products in unstable quarterly groups are used in seasonality analysis.

In order to perform this study, we use two reasonable criteria: the mean of absolute z-score and the percentage of non-zero values. The mean of absolute z-score, denoted by A, is:

$$Z = \frac{x - \mu}{\sigma}, \tag{2}$$

$$A = \frac{\sum Z}{n}, \tag{3}$$

where x is the value of data object, which is the sales quantity of a given period, μ is the mean of the population, σ is the standard deviation of the population and n is the total number of data objects. The value of A indicates how stable the product is based on periodical sales quantities. The lower the value of A, the more stable the product.The percentage of non-zero values, denoted by P, is:

$$P = \frac{m}{n}, \tag{4}$$

where m is the number of data objects with any non-zero sales quantity. The value of P indicates the frequency of the product sold in corresponding periods. Therefore, the range of value P is from 0 to 1. Zero means that the product has no sales records in any period. One means that the product has been sold in every period. Products with higher values of P are considered to be sold more often, which makes them more stable.

Table 7 shows a sample of time series clustering data sets used for stability analysis. For example, P128 has a higher value of P compared to P114. That is, P128 has been sold more often than P114. In addition, the A value of P128, 0.870237, is lower than the A value of P114, which is 0.913061. Therefore, P128 is more stable than P114.

The stability analysis is performed in two tiers. Tier-1 clustering analysis figures out broad stable groups. Tier-2 clustering further analyze the most appropriate stable group to determine a list of stable products.

Product ID	Mean of absolute z-score (A)	Percentage of non-zero values (P)
P061998079829	0.905559	0.980769
P061998084458	0.940804	1
P068958011219	0.821932	0.884616
P107	0.84088	0.903846
P114	0.913061	0.903846
P128	0.870237	1
P408	0.889739	0.980769
P418_120	0.886479	0.961538
P418_60	0.836407	0.903846
P424	0.930193	0.884616
P427	0.849136	0.884616
P430	0.829077	0.942307
P503	0.894007	0.961538
P624917040852	0.898048	0.961538
P624917060027	0.936271	0.923077
P624917060157	0.825782	0.865385
P628747100045	0.810335	0.923077
P631257534507	0.876968	0.884616
P635824000013	0.801448	0.961538
P693749015017	0.886347	1
P727783005601	0.893142	0.884616

Table 7. A sample of time series clustering data sets.

5.2 Seasonality Analysis

Within the group of unstable products, 'seasonal patterns' is considered as another criterion to further categorize products. That is, unstable products are categorized into groups based on their monthly and quarterly sales patterns. K-Means and EM clustering algorithms are applied again to cluster products.

Products are considered seasonal products if their sales quantities periodically change with time. If products' selling ratios in one period are high, they are called single period selling products (1). In addition, products, which were mostly sold in two periods, are called two-period selling products.

We perform monthly and quarterly seasonality analysis, to discover products' sales patterns. Similar to stability analysis, two tiers of clustering analysis is performed for each level of seasonality analyses. That is, the Tier-1 clustering analysis indicates reasonable seasonal groups and the Tier-2 clustering analysis categorizes these seasonal groups into finer groups. Since the amount of computation in seasonality analysis is rather intensive, the K-Means algorithm is chosen to facilitate seasonality analysis.

6 Experimental Results

In this section we describe the actual computations used for identifying stable and seasonal products. The first sub section elaborates weekly and monthly stability analysis for year 2005 whereas the second subsection illustrates quarterly seasonality analysis for 2006.

6.1 Obtaining Stable Products

Stability analysis gives us a list of products exhibiting weekly, monthly and quarterly stable sales patterns. Each of these weekly, monthly and quarterly stability analyses is performed at two levels. The process of obtaining weekly stable products is as follows.

Weekly Stability Analysis

If products' sales quantities are approximately the same in all weeks, they are called stable weekly products. This subsection describes two tier clustering analysis to identify weekly stable products.

Tier-1 weekly stability analysis: The goal of the Tier-1 time series clustering is to define stable groups from the noisy data set. The stability analysis starts with the EM algorithm.

Table 8 shows the EM clustering results of the Tier-1 2005 weekly stability analysis. According to the results, 5737 products are clustered based on their weekly sales quantities and 4 groups (clusters) are generated as an optimal clustering solution. Among them, Cluster 3 has the highest value of P (0.3643), which means that products in this group were sold more often than products in other groups. Since the P values of Clusters 0, 1 and 2 are low, as 0.125, 0.0263 and 0.0665, respectively, products in these groups are not frequently sold products. Thus, Cluster 3 is the most stable group. That is, 1417 products, which are 25% of total products, are considered stable products in the Tier-1weekly stability analysis. The Tier-2 clustering analysis is performed on these products.

Product groups	Number of products (percentage)	Mean of absolute z-score (A)	Percentage of non-zero values (P)
Cluster 0	1203 (21%)	0.6153	0.125
Cluster 1	2067 (36%)	0.3105	0.0263
Cluster 2	1050 (18%)	0.4796	0.0665
Cluster 3	1417 (25%)	0.7785	0.3643

Table 8. Tier-1 weekly stability analysis with the EM algorithm in 2005.

Tier-2 weekly stability analysis: In the Tier-2 clustering analysis, time series clustering techniques, such as EM and K-Means, are applied. The normalized data set is clustered with the EM algorithm. However, the K-Means algorithm is used as an alternative solution to facilitate the Tier-2 clustering tasks when EM clustering results are not descriptive. Moreover, the number of output clusters is defined as 5 to distinguish products into finer groups, such as very stable, relatively stable, normal, relatively unstable and unstable groups.

Table 9 shows the Tier-2 clustering results carried on 1417 products (Cluster 3) filtered in the Tier-1 2005 weekly stability analysis. The results show that Cluster 11 is the most stable group since it has the highest P value (0.9128) and reasonable A value (0.7878). Therefore, 32 products are categorized as weekly stable products in 2005.

Monthly Stability Analysis

Products are considered stable monthly if their sales quantities are approximately the same in all months. In the weekly stability analysis, stable weekly products and unstable weekly products are distinguished from original data set. Unstable

Product groups	Number of products (percentage)	Mean of absolute z-score (A)	Percentage of non-zero values (P)
Cluster 0	203 (14%)	0.8553	0.3763
Cluster 1	53 (4%)	0.7295	0.3506
Cluster 2	129 (9%)	0.7542	0.5721
Cluster 3	35 (2%)	0.5943	0.4815
Cluster 4	86 (6%)	0.7501	0.1779
Cluster 5	55 (4%)	0.7815	0.297
Cluster 6	29 (2%)	0.847	0.6301
Cluster 7	53 (4%)	0.8544	0.4777
Cluster 8	101 (7%)	0.7243	0.1974
Cluster 9	99 (7%)	0.7491	0.225
Cluster 10	140 (10%)	0.8369	0.2787
Cluster 11	**32 (2%)**	**0.7878**	**0.9128**
Cluster 12	65 (5%)	0.813	0.7518
Cluster 13	80 (6%)	0.8019	0.5042
Cluster 14	53 (4%)	0.6704	0.2531
Cluster 15	83 (6%)	0.7795	0.205
Cluster 16	121 (9%)	0.8061	0.2291

Table 9. Tier-2 weekly stability analysis with the EM algorithm in 2005.

weekly products are further analyzed in the monthly stability analysis. Likewise weekly stability analysis, we perform two-tier clustering analysis with EM and K-Means algorithms for monthly stability analysis as follows.

Tier-1 monthly stability analysis: Table 10 shows the EM clustering results of the Tier-1 2005 monthly stability analysis. In total, 5705 products are clustered based on their monthly sales quantities. The optimal clustering solution categorizes products into 7 groups (clusters). Cluster 0 has the highest value of P (0.9301), which means products in this group were sold more often than the others. The A value (0.8201) in Cluster 0 is reasonable. Thus, products in Cluster 0 are the most stable products in terms of monthly sales quantities in 2005. Therefore, 644 products, which are 11% of total products, are to be analyzed in the Tier-2 monthly stability analysis.

Product groups	Number of products (percentage)	Mean of absolute z-score (A)	Percentage of non-zero values (P)
Cluster 0	644 (11%)	0.8201	0.9301
Cluster 1	845 (15%)	0.8406	0.2871
Cluster 2	1543 (27%)	0.5528	0.0833
Cluster 3	878 (15%)	0.7252	0.1765
Cluster 4	262 (5%)	0.6947	0.6682
Cluster 5	719 (13%)	0.9273	0.4137
Cluster 6	814 (14%)	0.8359	0.6142

Table 10. Tier-1 monthly stability analysis with the EM algorithm in 2005.

The Tier-2 EM clustering results of the Cluster 1 (644 products) in the Tier-1 2005 monthly stability analysis is illustrated in Table 11. Statistically, the results show that one big group of all the products is the optimal solution. It does not meet the group refinement goal of the Tier-2 stability analysis. In this case, we use the K-Means algorithm to categorize products into 5 groups. The K-Means clustering results are shown in Table 12. According to the results, Cluster 2 is the most stable group since it's A value (0.7664) is the lowest. Therefore, 87 products are categorized as stable monthly products in 2005.

Product groups	Number of products (percentage)	Mean of absolute z-score (A)	Percentage of non-zero values (P)
Cluster 0	644 (100%)	0.8183	0.9301

Table 11. Tier-2 monthly stability analysis with the EM algorithm in 2005.

Product groups	Number of products (percentage)	Mean of absolute z-score (A)	Percentage of non-zero values (P)
Cluster 0	150 (23%)	0.7802	1
Cluster 1	109 (17%)	0.8483	0.9167
Cluster 2	87 (14%)	0.7664	0.9167
Cluster 3	172 (27%)	0.821	0.8333
Cluster 4	126 (20%)	0.87	1

Table 12. Tier-2 monthly stability analysis with the K-Means algorithm in 2005.

Likewise weekly and monthly stability analysis, quarterly stability analysis is performed. In Section 7, we present weekly, monthly and quarterly stable products obtained from 2005 as well as from 2006 sales data.

6.2 Obtaining Seasonal Products

Normalized seasonal (monthly, quarterly) sales values of products for all months (or for all quarters) are obtained and such transformed data set is used in seasonality analysis. We present here quarterly seasonality analysis for year 2006 data.

Quarterly Seasonality Analysis

Some possible groups are 4 single sales-season groups, 2 double sales-season groups and 1 random selling group. Thus, we cluster products into 7 reasonable groups in the Tier-1 analysis.

Tier-1 quarterly seasonality analysis: Table 13 shows the Tier-1 2006 quarterly seasonality analysis results. According to the results, Cluster 2 has the most significant seasonal sales patterns. Products in this group were sold mostly in Quarter 4; they are single quarter selling products. The following groups are Clusters 4 and 0. They have clear sales patterns, but not as significant as Cluster 2. Cluster 6 is a tricky group. It can be considered single quarter selling group since products' sales quantities in the first quarter are relatively high. However, products' sales quantities in the first and second quarters are dominating in this group. That is, it is a two-quarters selling group in Quarters 1 and 2. In addition, products

Product groups	Number of products (percentage)	Quarter1	Quarter2	Quarter3	Quarter4
Cluster 0	434 (8%)	0.3431	0.735	2.4959	0.4259
Cluster 1	1001 (19%)	0.4543	0.7123	1.3548	1.4786
Cluster 2	387 (7%)	0.255	0.1847	0.1689	3.3914
Cluster 3	981 (19%)	1.6348	0.4959	0.9308	0.9385
Cluster 4	384 (7%)	0.3653	2.553	0.3459	0.7358
Cluster 5	1560 (30%)	1.0908	1.2294	0.7693	0.9106
Cluster 6	505 (10%)	2.7648	0.8854	0.1177	0.2321

Table 13. Tier-1 quarterly seasonality analysis in 2006.

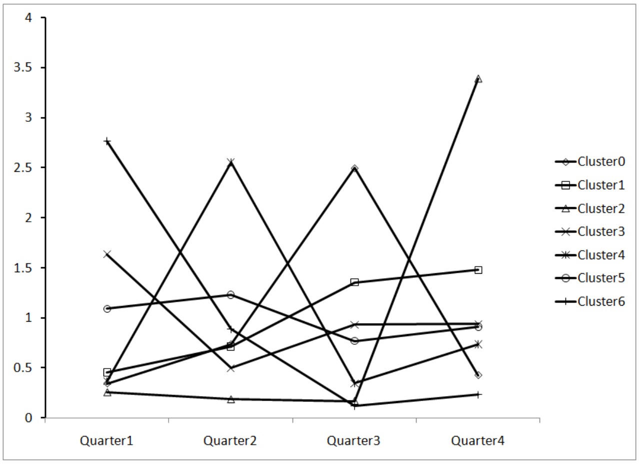

Figure 1. Tier-1 seasonality analysis – 2006 quarterly sales trend.

in this group were sold more in Quarter 1 than in Quarter 2. The Tier-2 quarterly seasonality analysis may identify products' sales patterns more clearly.

Figure 1 illustrates sales trends of each group in the Tier-1 2006 monthly seasonality analysis. It visually confirms our findings and supports the proposed inventory management strategies.

Tier-2 quarterly seasonality analysis: Several reasonable groups are provided by the Tier-1 quarterly seasonality analysis. The Tier-2 clustering analysis refines the clustering results. Here, we cluster products into 5 groups based on their quarterly sales patterns. Some of these Tier-2 clustering results are discussed below. According to the Tier-1 2006 quarterly seasonality analysis, Cluster 2 (387

Product groups	Number of products (percentage)	Quarter1	Quarter2	Quarter3	Quarter4
Cluster 0	200 (52%)	0.0083	0.0023	0.0137	3.9757
Cluster 1	24 (6%)	0.8134	0.0181	0.0235	3.145
Cluster 2	70 (18%)	0.2307	0.1275	0.7696	2.8722
Cluster 3	42 (11%)	1.1932	0.1293	0.0969	2.5806
Cluster 4	51 (13%)	0.2207	1.1021	0.0805	2.5966

Table 14. Tier-2 2006 quarterly seasonality analysis – Cluster 2.

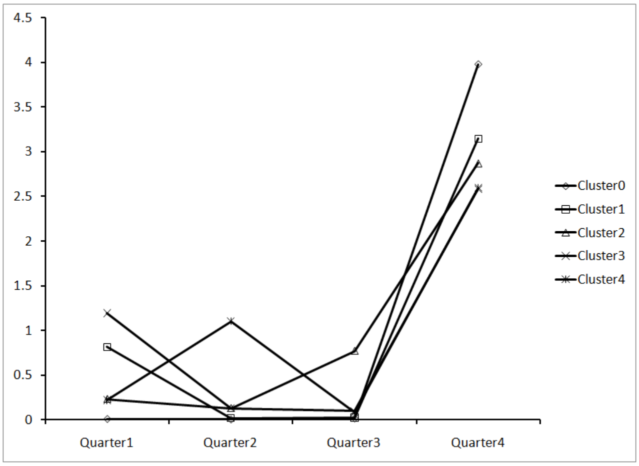

Figure 2. Tier-2 seasonality analysis – 2006 quarterly sales trend of Cluster 2.

products) is a single quarter selling group in Quarter 4. The Tier-2 quarterly seasonality analysis refines the grouping results as shown in Table 14. Obviously, Cluster 0 is a single quarter selling group with the most significant seasonal sales pattern. Products in this group were sold mostly in Quarter 4. Clusters 2 and 4 are two-quarters selling groups in Quarters 3–4 and Quarters 2–4, respectively. Coincidently, Clusters 1 and 3 are both two-quarters selling groups in Quarters 1–4. However, they are two seasonal groups because the weights of sales quantities between these two quarters are different. According to products' sales pattern in these seasonal groups, our seasonal inventory management strategies can be smoothly applied. Figure 2 illustrates the sales trends of these seasonal groups in Table 14. It graphically proves our findings reported above and supports the proposed inventory management strategies.

We illustrated quarterly seasonality analysis. Monthly seasonal analysis is obtained in similar fashion.

In the next section, we shall give detailed results of monthly and quarterly seasonality analysis for year 2005 and year 2006 data.

7 Summary of Results

This study analyzes a retail chain data set for stability and seasonality. The EM and K-Means clustering algorithms are used to facilitate product analyses. According to the analysis, 3 stable groups are identified at three levels: weekly (32 products), monthly (87 products) and quarterly (23 products) in 2005. Similarly, in 2006, stable weekly (21 products), monthly (26 products) and quarterly (32 products) groups are identified. There are 50 seasonal groups, including 1485 products, located in 2005 and 49 seasonal groups, including 2194 products, located in 2006. Figures 3 and 4 illustrate product distributions in 2005 and 2006, respectively. Product groups and numbers of products they contained are labeled in the Figures. It can be observed that sales patterns of more than 50% of products have been identified through product analyses.

A comparison between generic optimal solutions and local optimal solutions for specific groups reveals that local optimal solutions outperformed generic optimal solutions based on the lower MAPE criterion. The group G-05Q20 represented in case study-1 of Section 3 is a local group that exhibits seasonal sells pattern. All 135 products in this group are sold mostly in quarter 1. Hence the generic model could not obtain optimal solution. A different model suits best and generates the optimal solution for products of this group.

Simple inventory management strategies are proposed to control seasonal inventories, such as groups discussed in case study-1.

(i) Carry very few seasonal products in their off-sales periods. The inventory in off-season should be based on quantities from previous year sales during off-season.

(ii) Order a lot of seasonal products in their sales periods. The size of order should be based on quantities from previous year sales during the same season.

Case study-2 presented a group (G-06M60) that has a strong stable sales pattern. Several such groups are identified by analyzing the data set thoroughly. For stable products group, the use of time series predication models is not necessary. Sometime it may be disguised. Hence, better strategy is to order stable products regularly for the pertinent period. The size of order should be stable and based on quantities that came forward as a result of stability analysis.

8 Conclusions

Products in a retail store tend to exhibit different sales patterns. Some of the products sell more or less uniformly day after day. These products demonstrate the most stable sales pattern. Another group of products may sell more or less uniformly throughout the year. While their sales patterns may be similar from week to week, they may show significant daily variations. Similarly, some products may have a stable sales pattern from month to month, with fluctuating daily or weekly demand. Identification of these patterns can be referred to as stability analysis. In contrast to the products that sell more or less uniformly throughout the year, stores

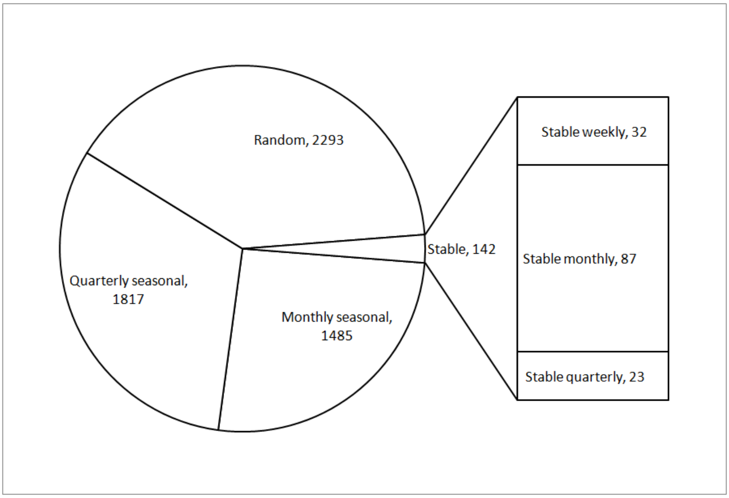

Figure 3. Product distribution in 2005.

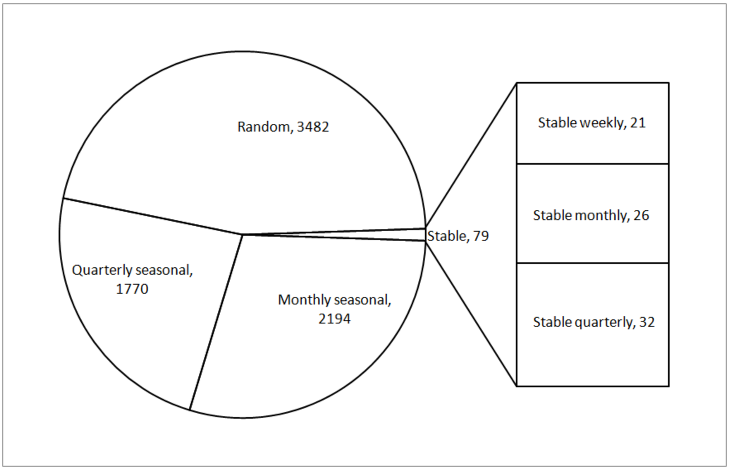

Figure 4. Product distribution in 2006.

also carry products that only sell during a given season such as spring, summer, autumn, and winter. Seasonality analysis allows us to identify such seasonal sales patterns.

Due to variations in such sales patterns, it is difficult to obtain a generic time series prediction model that fits best to entire data set. We propose time series clustering to analyze the data set to identify local groups of products that exhibit typical seasonal sales pattern or stability in sales pattern for a certain period. For such local groups depending upon seasonality or stability, we propose certain recommendations that result in better inventory forecasting.

We experimented with sales transactions of a retail chain store collected over a duration of 33 months. MAPE evaluations are studied between time series prediction models and the best-fit solution for each product is identified, the local optimal solution for each group and the generic optimal solution for the entire product set in each year at month and quarter level. The proposed inventory forecasting will provide simpler and more accurate management strategies to control seasonal inventories, especially for smaller retailers.

The clustering analysis reported in this study is based on crisp clustering that assigns sales patterns to one and only one group. Soft clustering alternatives such as fuzzy and rough clustering make it possible to assign sales patterns to more than one group. For example, we may be able to assign a product to both spring and summer clusters. Instead of committing them to either spring or summer cluster. Such an assignment could be used to create hybrid prediction models by combining prediction models from different groups. We plan to study the application of soft clustering techniques to seasonal and stability analysis of products for further improving the accuracy of inventory predictions.

Bibliography

[1] Y. Aviv, A Time-Series Framework for Supply-Chain Inventory Management, *Operations Research* **51** (2003), 210–227.

[2] B. Billah, M. L. King, R. D. Snyder and A. B. Koehler, Exponential smoothing model selection for forecasting, *International Journal of Forecasting* **22** (2006), 239–247.

[3] S. Chatterjee and A. S. Hadi, *Regression analysis of example*, 4 ed, Wiley-Interscience, 2006.

[4] J. Contreras, R. Espinola, F. J. Nogales and A. J. Conejo, ARIMA models to predict next-day electricity prices, *IEEE Transactions on Power Systems* **18** (2003), 1014–1020.

[5] A. P. Dempster, N. M. Laird and D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society, B* **39** (1977), 1–38.

[6] R. M. Hogarth and S. Makridakis, Forecasting and Planning: An Evaluation, *MANAGEMENT SCIENCE* **27** (1981), 115–138.

[7] G. A. Kruger, A Statistician Looks at Inventory Management, *QUALITY PROGRESS* **38** (2005), 36–44.

[8] M. Kumar and N. R. Patel, Using clustering to improve sales forecasts in retail merchandising, *Annals OR* **174** (2010), 33–46.

[9] J. Lawrence, *Introduction to neural networks: design, theory, and applications*, 3 ed, California Scientific Software Press, 1993.

[10] P. Lingras, M. Zhong and S. Sharma, *Evolutionary regression and neural imputations of missing values*, Soft Computing applications in industry, 226.

[11] J. MacQueen et al., Some methods for classification and analysis of multivariate observations, in: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 1, California, USA, p. 14, 1967.

[12] A. Mehrez and M. Y. Hu, A clustering analysis of forecasting methods in a multi-objective inventory system, *International Journal of Production Economics* **27** (1992), 1–8.

[13] E. A. Silver, D. F. Pyke and R. Peterson, *Inventory management and production planning and scheduling*, 3 ed, John Wiley and Sons, 1998.

[14] A. Sorjamaa, J. Hao, N. Reyhani, Y. Ji and A. Lendasse, Methodology for long-term prediction of time series, *Neurocomputing* **70** (2007), 2861–2869.

[15] N. Stefanovic, D. Stefanovic and B. Radenkovic, *Application of Data Mining for Supply Chain Inventory Forecasting*, Applications and Innovations in Intelligent Systems XV (Richard Ellis, Tony Allen and Miltos Petridis, eds.), Springer London, pp. 175–188.

[16] J. W. Taylor, Short-Term Electricity Demand Forecasting Using Double Seasonal Exponential Smoothing, *The Journal of the Operational Research Society* **54** (2003), 799–805.

[17] F.-M. Tseng, H.-C. Yu and G.-H. Tzeng, Combining neural network model with seasonal time series ARIMA model, *Technological Forecasting and Social Change* **69** (2003), 71–87.

Received May 1, 2011.

Author information

Peng Zhang, Department of Mathematics and Computing Science,
Saint Mary's University, Canada.
E-mail: alan.p.zhang@gmail.com

Manish Joshi, Department of Computer Science, North Maharashtra University, India.
E-mail: joshmanish@gmail.com

Pawan Lingras, Department of Mathematics and Computing Science,
Saint Mary's University, Canada.
E-mail: pawan@cs.smu.ca