Frequent Problems of Model Specification and Forecasting of Time Series in Goods Management Systems

Gerhard Arminger [1]

Carsten Schneider [1]

Technical Report – SFB 475

1. Bergische Universität – Gesamthochschule Wuppertal, FB 6, Gaußstr. 20, D-42097 Wuppertal.
The authors are grateful to Norman Götz for reading various drafts of the paper.

Abstract

The forecasting of time series in goods management systems causes various problems that we identify and indicate possible solutions. The implementation of auxiliary information like promotional activities or calendar effects in forecasts using ARMA models and exponential smoothing methods may be difficult, especially if these effects did not yet occur in the past. The effects usually cause transient shocks that have to be corrected when using ARMAX models and that yield high forecast variances. These high variances have to be corrected when estimating the parameters. The calendar effects do not for all items have a fixed seasonal figure but may have a variable seasonal behaviour, e.g. depending on the weather. We have to identify and perhaps to eliminate outliers that would yield not optimal forecasts. In addition to these problems dealing with univariate time series, we introduce some similarity-measures for time series which are prerequisite for the multivariate analysis of time series.

Keywords: forecasting of time series, auxiliary information, seasonal behaviour, similarity of time series, Goods Management Systems

1. Introduction

Classical time series analysis, e.g. exponential smoothing or ARIMA-modelling, is not the optimal way for forecasting time series in goods management systems for the following reasons:

- Usually there exists auxiliary information about the items for which the sales are to be predicted. This information should be integrated into the forecast because all available information should be used to improve the forecasts
- Due to time lags induced by delivery schedules we need forecasts for the lags t+1, t+2 and t+3. Forecasting 2 or 3 steps ahead can cause some problems and is often more difficult than just 1-step ahead forecasts.
- Usually the forecast quality is measured by the MSE (Mean Squared Error) or MAD (Mean Absolute Deviation), and for example the estimation of the parameters of an ARIMA model generates a minimum in the MSE criterion. However, in goods management systems we are interested in costs of forecasting sales that yield either stockouts or stocks that are too high. These costs may be not symmetric and therefore we need to measure the quality of forecasts by an asymmetric loss function. Possible asymmetric loss functions are derived in a companion paper by Arminger and Götz (1999).

In this paper we will focus on frequent problems in modelling and forecasting sales in goods management systems and indicate possible solutions applying special models of exponential smoothing, ARMAX-processes and dynamic linear systems. These solutions are all tentative. The solutions and possible improvement will have to be checked using the criteria provided by Arminger and Götz (1999).

2. Available Information

Available for analysis and forecasting are time series of sales in many items, usually a few thousand different time series. Each of these time series is rather short (for instance 52 weeks, 104 weeks, 365 days, 730 days). The different time series may be related, but we usually do not know whether a time series belongs to a set of time series with common properties or not and how the relationship within a set may be characterised. Therefore, clustering and grouping may have to be applied before methods for analysing vectors of time series may be used. The different time series may be related, but the type of relationship is usually unknown.

In all cases we need some information about the past to specify any model. Usually the first 20 observations are used for this first model-specification (this is also called "the past" of a time series), we start forecasting with t=21 and modify the model after each new available observation. If graphs are given as example, they start with the first available observation.

In addition some auxiliary information may be available:

- Item specific information
    - Promotional Activities
    - Prices, Price changes
    - Price of competing products
- Calendar information, which is usually relevant for many items, but it is unknown for which. This information may include
    - Holidays (Easter, Christmas)
    - Vacation terms (Easter-, summer vacation)
    - Special occasions (Mother´s day for drug markets; Father´s day for hardware markets)

3. Frequent Problems

In this chapter we give an overview of the most frequent problems in predicting the time series of sales of many items. In addition we try to indicate possible solutions for these problems without having checked their usefulness or applicability in practice.

3.1 Item Specific Information

3.1.1 Promotional activities without precedence in the past

Promotions are all kinds of additional activity, for example a special presentation of single goods, a second sales location for some days, or simply “special offers”, which usually includes changes in price for some time. The last case is therefore a combination of promotions and price changes, regarded in chapter 3.1.5.

If Promotions have not yet appeared, it is difficult to predict the future sales.

A typical graph using weekly data is shown in figure 1.

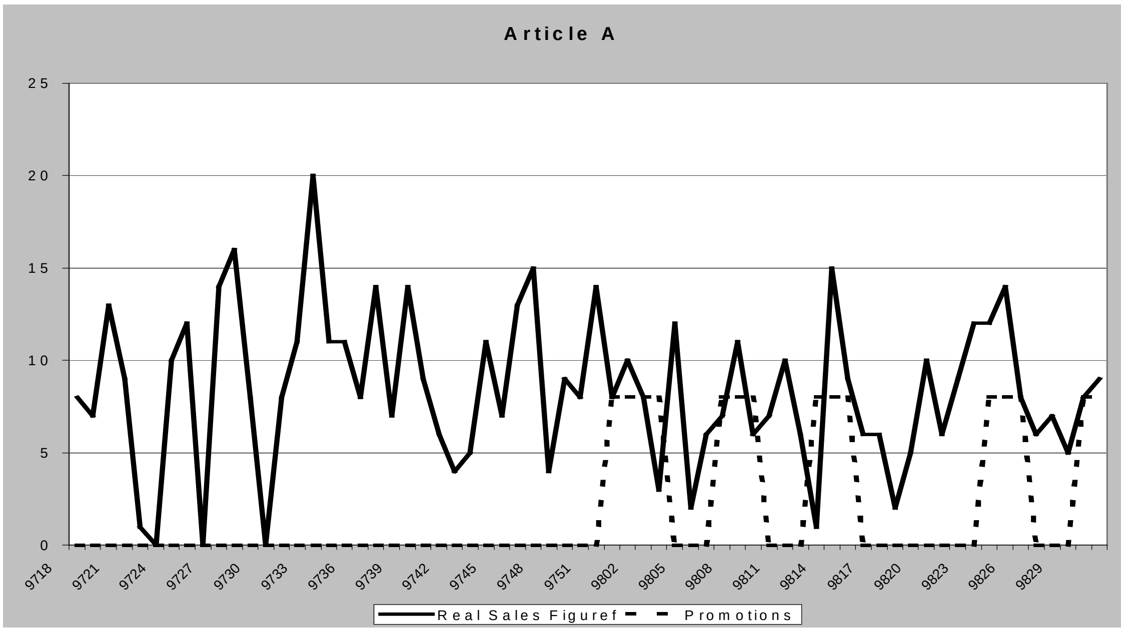

Figure 1: Sales Figure without Promotion Activity in the Past

The graph shows that no promotional activity occurred till calendar week 98/02. Afterwards, the promotional activities do not seem to indicate any additional sales effect.

Possible solutions for this problem may be:

- Ahead Monitoring. In this case, the user receives some periods before the first promotional activity occurs, a signal that warns of possible changes in the sales because of the future promotional activity.
- Introduction of a reference time series of a specific product. This implies the comparability of these two time series and the similar behaviour in the case of promotional activities. In practice this is an identification problem which needs a fair amount of additional information about the reference product to verify the comparability before making any forecast on the base of a reference time series. Usually, such a reference item can only be provided by the retail chain management.
- Introduction of a reference time series constructed from a whole set of items to which the item that is analysed belongs. Such a set may be defined as a group or category of similar items such as perfumes or baby food or tools, but they are always thought to exhibit a certain homogeneity of the items in the set. Therefore, there is a need to develop methods that allow to find homogeneous sets of items using the time series information.

3.1.2 Promotional activities with precedence in the past

A typical graph using weekly data is shown in figure 2.

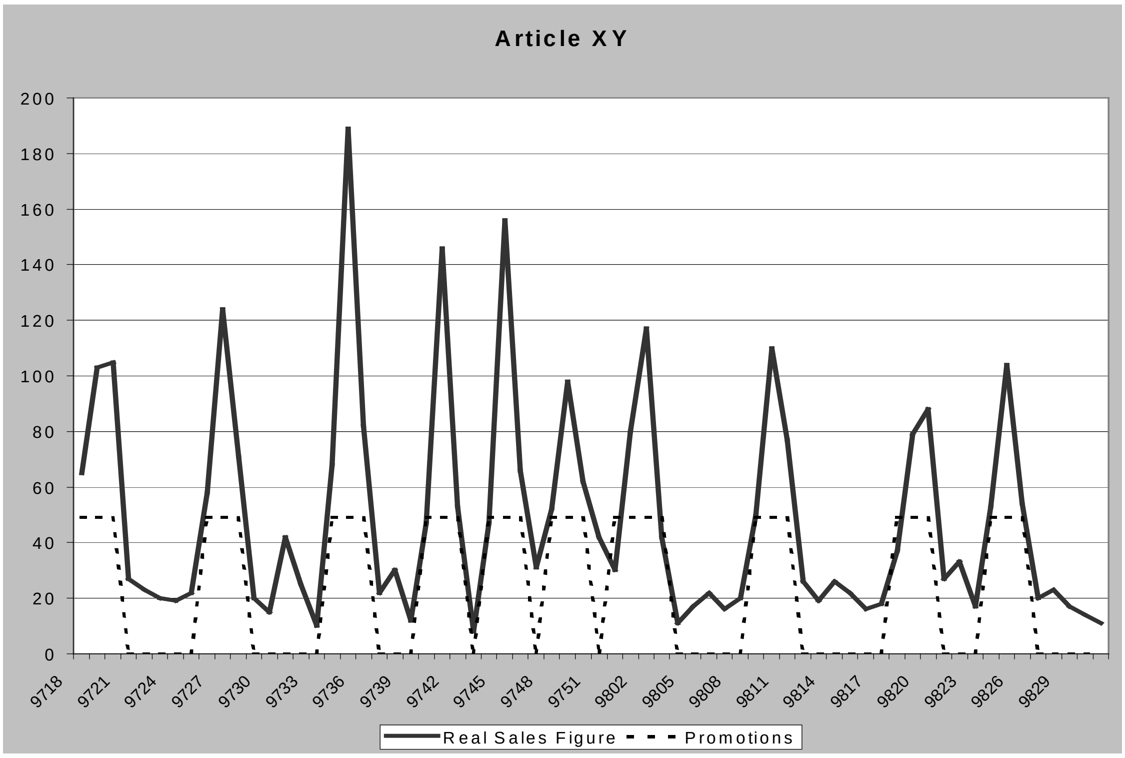

Figure 2: sales figure with promotional activities in the past

The graph shows that promotional activities have occurred in the past and that there exists a strong effect of the promotional campaign on the sales.

The effect of promotional activity may usually be considered as a transient shock without a persisting effect. A simple forecasting method would be a standard ARX(1)-model of the following form:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 a_t + \varepsilon_t \tag{1}$$

with a_t as indicator-function for promotions. This model yields too high a forecast for y_{t+1} if $a_t = 1$ because of the correlation between y_t and y_{t+1} induced by β_1.

A possible solution for this problem may be to correct the influence of a_t in an ARX(1) model by subtracting the effect of a_t on the time series yielding the model

$$y_t = \beta_0 + \beta_1 (y_{t-1} - \beta_2 a_{t-1}) + \beta_2 a_t + \varepsilon_t . \tag{2}$$

The problem of this solution is that it leads to a regression model of the form

$$y_t = \beta_0 + \beta_1 y_{t-1} - \beta_1 \beta_2 a_{t-1} + \beta_2 a_t + \varepsilon_t \tag{3}$$

with a non-linear restriction on the parameter for a_{t-1}. Another problem is the extension to more complicated ARMAX models.

3.1.3 High forecast variance induced by promotional activities

Transient shocks such as promotional activities generate a very high variance of the forecast, especially if only few activities have occurred before. If this additional variance is not removed from the estimated variance of the forecast the consequence will be a fair amount of overstocking. A possible solution is to use only the residuals of past forecasts as an estimate of the forecast variance without the additional variance induced by the estimation of the activity parameter β_2 .

3.1.4 Short time series with promotional activities

If model building with an ARMAX model is not possible exponential smoothing instead of ARMAX-models could be used. In this case, the problem arises, how additional information like promotions can be integrated in exponential smoothing forecasts.

3.1.5 Prices and price changes

Prices should generate a permanent change and therefore may be modelled as

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 p_t + \varepsilon_t \tag{4}$$

with p_t as metric variable for the price.

If price changes have not yet occurred in the past the corresponding design-matrix will be singular. A possible solution is the use of the g_2-inverse matrix for computing the parameter estimator.

A typical graph using weekly data with regards to prices is shown in figure 3.

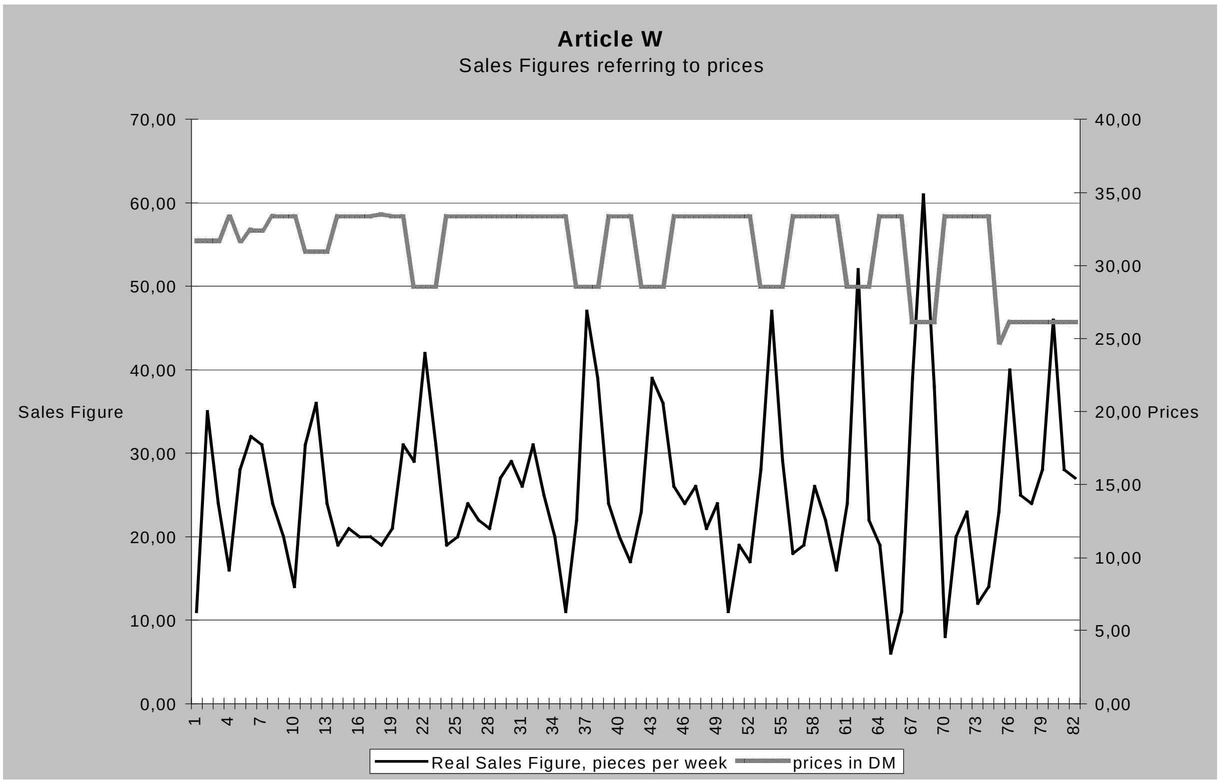

Figure 3: Sales figures referring to prices

The graph shows the influence of price changes to the time series. Decreasing the price leads to higher sales and reverse.

3.2 Calendar effects

3.2.1 Christmas effect

The Christmas effect usually starts to influence the time series some weeks before Christmas and increases up to the Christmas-week where often a peak occurs. For some items only a single peak in the Christmas-week occurs. Both kinds of effect are usually followed by one or a few weeks of very low sales figures. These transient shocks cause different problems:

- The first occurrence cannot be predicted. Even if it is possible to predict "some Christmas effect", the prediction of the sales is not trivial. This effect is similar to the first occurrence of promotional activities in 3.1.1.
- The often used ARX-models generate wrong forecasts and wrong forecast variances just as they do in 3.1.1 – 3.1.3. Possibly model (2) can be used to correct the transient shock caused by the Christmas effect, but the problems start to increase when an ARX(1)-model is not complicated enough to capture the movement of sales during this season.

The following figure 4 shows the cumulated sales of all items in an outlet of a retail chain with a weekly average sale of more than 30 pieces.

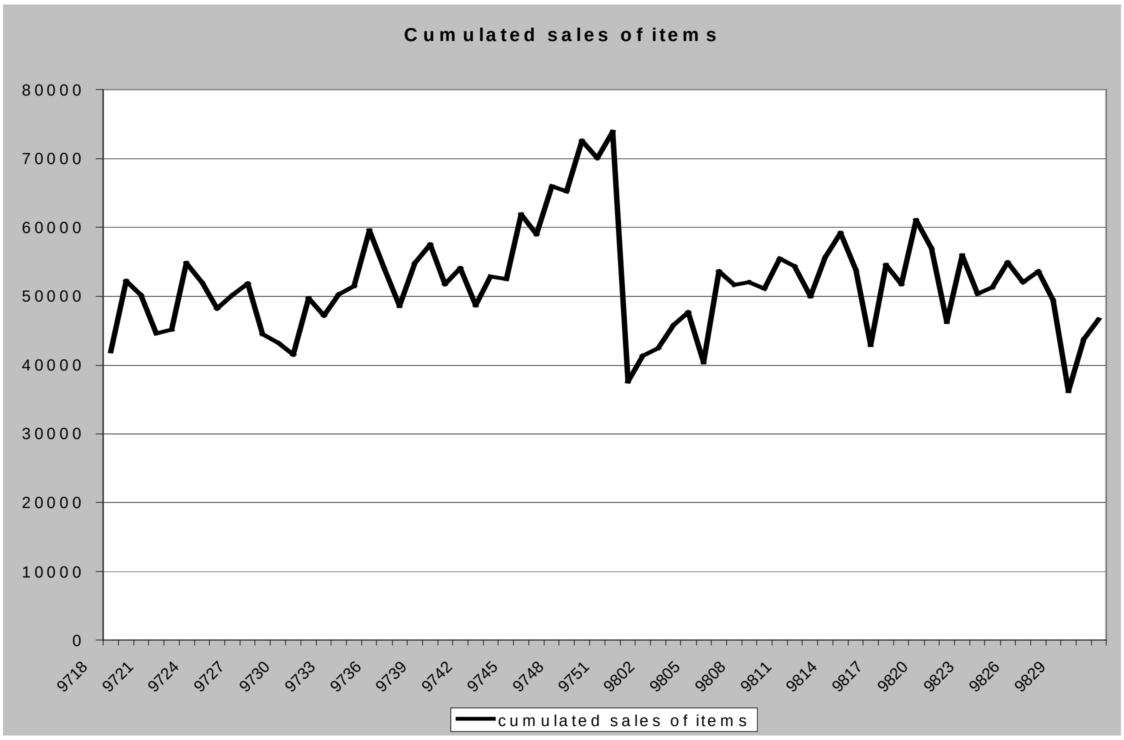

Figure 4: Cumulated sales of items

The graph shows the Christmas-effect with increasing sales a few weeks before Christmas (week 52/1997, Dec. 20-26) and very low sales in the weeks after Christmas took place.

Solutions may be similar to the solutions as in 3.1.1 – 3.1.3. Special attention has to be given to the weeks after Christmas. Here, a model of the kind

$$y_t = \beta_0 + \beta_1 (y_{t-1} - \beta_2 c_{t-1} - \beta_3 d_{t-1}) + \beta_2 c_t + \beta_3 d_t + \varepsilon_t \tag{5}$$

with c_t as dummy variable for Christmas and d_t as dummy variable for After-Christmas can be useful.

3.2.2 Easter- and Vacation effect

Easter effects can be modelled together with the vacation effects because it is not possible to model the Easter effect alone due to the Easter-vacations. Usually we have increasing or decreasing sales some time before the actual event occurs, for example a few weeks before the beginning of the vacations. During the vacations, the sales have usually a constant level which is either higher or lower than in other times. In some cases the vacation effect reduces to a single peak at the beginning or the end of the vacation period.

Possible solutions:

- Single peaks can be regarded as transient shocks as in the case of promotions. The corresponding model is of the kind (2), but usually the vacation effect occurs only once or twice a year, while promotions take place more often. Model (2) implies information about the behaviour of a time series in the concrete situation which could already been observed. This information is often not available because in short time series of length 52. The effect of Easter and summer vacations is not necessarily the same, so it is impossible to predict the effect of summer vacation if the time series is not longer than one year.
- If information about the seasonal behaviour is available, sales figures can be modelled as in (2) with one dummy variable for every kind of vacations to include differences between the summer vacation effect and the Easter effect.

3.2.3 Special Holidays

Some special days have very interesting effect in few products, for example at the 1st of November occurs a peak in sales figures of candles and flowers, Mother´s day has an effect on make-up and perfume and on Father´s day sales of machine tools and after shaves increase. Usually these sales figures start to increase one or two periods before the actual event, so prediction has to take not only future holidays themselves but also the weeks before into consideration.

3.2.4 Seasonal effects

Seasonal effect may cause various problems:

- Christmas may occur 52 or 53 weeks earlier if the forecast periods are weeks. For example 1997 had 52 weeks, 1998 had 53 weeks and for the prediction of the Christmas effect 1999 53 weeks will have to be taken into account. Hence, the seasonal lag has to be adapted over the years.
- One effect can be observed in every holidays, but especially at Christmas or Easter time: People start to buy at some time before the actual event. This is similar to investment buying for a central warehouse. This implies increasing sales figures before any kind of holidays and possibly lower sales figures after the holiday because personal stocks have to be used up.
- Some sales depend on weather, for example gardening tools with a peak as soon as spring-weather starts or winter tires when the first snow is falling. This can not be predicted by seasonal time series analysis but can only be corrected beforehand by an early increasing of stock capacities.

A typical graph using weekly data is shown in figure 5:

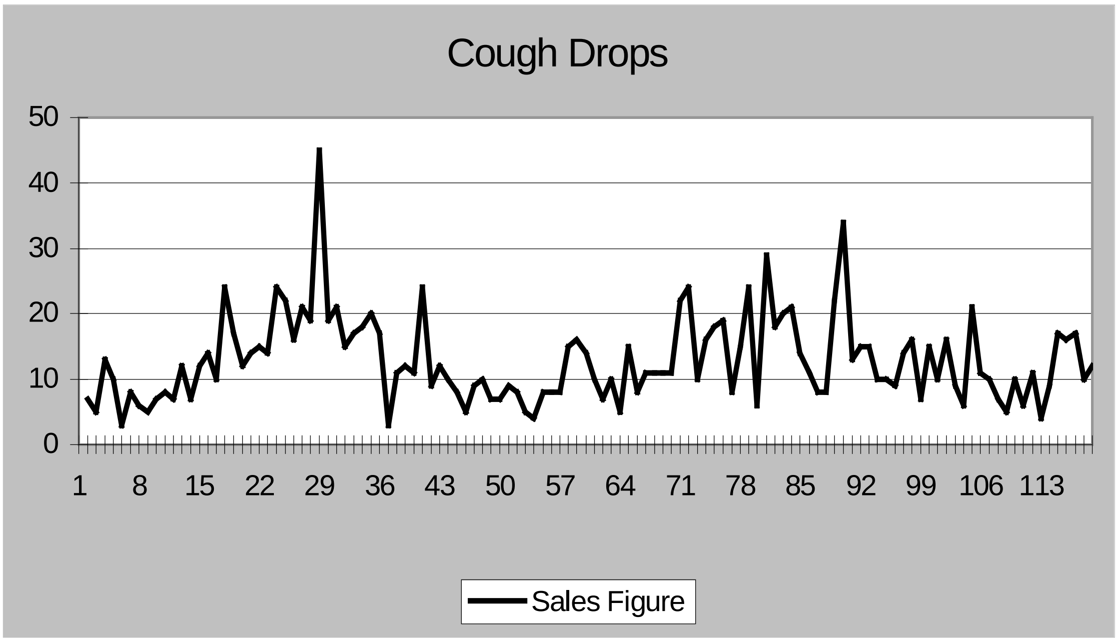

Figure 5: Seasonal Behaviour of time series of Cough Drops

The graph shows the seasonal behaviour of the sales of cough drop. The length of the season is not exactly 52 week because the weather has an effect on the sales.

Possible solutions:

- If the effect of "investment buying" has already occurred, it can be predicted by using an asymmetric window for seasonal smoothing, for instance

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 (1/3(y_{t-50} + y_{t-51} + y_{t-52})) + \varepsilon_t . \quad (6)$$

- Regular seasons can be modelled in the case of a fixed number of periods between the observations (either 52 or 53), for instance

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-52} + \varepsilon_t . \quad (7)$$

In the case of weather effects, stocks should be built and one could adapt the formula by:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-50} + \varepsilon_t \,. \qquad (8)$$

This implies, that weather effects do not occur earlier than two periods before they appeared last year. From this moment on, stocks grow and in the moment of concrete weather effects, demand can be gratified.

3.3 Outlier Identification

A special problem is the identification of outliers to avoid high forecast variances due to the influence of the outliers in the past. Possible solution to detect outliers online may be found in multiprocess dynamic linear models for online monitoring (cf. Brandt-Lassen, Dahl, 1995) or in phase space models currently researched by Bauer, Gather, Imhoff (1999).

4 Using the information from multiple time series

Using the information of one time series for forecasting another one is a very promising idea, but in the case of thousands of time series very difficult to do. The reason is that usually the user of a forecasting system does not know which time series he should use as predictor series. Therefore one needs to identify automatically similar time series or time series that can be used for increasing the forecast quality and these identification procedures have to be automatized.

4.1 Similarity of time series

The first question here is, how similarity between time series can be defined and measured. Possible options for similarity are

a) If both time series 1 and time series 2 can be used for forecast (Forecast similarity).
b) If time series 1 and time series 2 "look" similar after proper standardisation (Cluster similarity). The problem here is that cluster similarity does not take the serial structure into account.
c) If time series 1 and time series 2 can be predicted with the same models (Model similarity).

These types of similarity need to be measured precisely, which is until now hardly possible. For instance, cluster similarity could be measured by defining a maximum sum of squared differences depending on the length of the time series after a certain standardisation. But this would not necessarily solve the problem of not taking into account the serial structure.

4.2 Identification of time series that may be useful for forecasting others

There are several possibilities for forecasting one time series with help of a predictor series if the prediction is already identified. These are

- Positively correlated time series at time t (orange juice and Campari, Cola and Bacardi etc.). This may be especially useful if a promotional campaign for one item has a spill over effect on another item.
- Negatively correlated time series of substitution goods (different types of taste in baby food, cough drops etc.).
- Using one item as indicator for trends or seasonal variations (lead items), for instance the average sales of an item category from last year.

4.3 Model specification

If correlated or homogeneous time series have been found, how can predictive models be specified? Possible options are:

a) VAR-models
b) "lead time" models of the following kind:

$$y_t^{(1)} = \beta_0 + \beta_1 y_{t-1}^{(1)} + \beta_2 y_{t-1}^{(2)} + \varepsilon_t$$

with $y_{t-1}^{(2)}$ as realised time series of the lead item.

5 Automatic model selection

Due to the big number of time series which have to be predicted, it is absolutely necessary to generate forecasts automatically. This causes a lot of new problems because it seems to be possible to find solution for most problems earlier mentioned, but how can these solutions be automatized? Possible key problems in this context are:

- fractional methods
- forecast quality using asymmetric loss functions
- a mix of forecast strategies
- creating similarity indices for time series

References

Arminger, G., Götz, N. (1999): *Asymmetric Loss Functions for Evaluating the Quality of Forecasts in Time Series for Goods Management Systems*, SFB 475, TR XX/99

Bauer, M., Gather, U., Imhoff, M. (1999): *Statistical Pattern Detection in Univariate Time Series of Intensive care On-Line Monitoring data*, SFB 475, TR 1/99

Brandt-Lassen, M., Dahl, U.L. (1995): *A Multiprocess Dynamic Linear Model in Monitoring Postoperative Breast Cancer Patients*, Master of Science Thesis, Aalborg University, Dänemark