INTEGRATING SEASONAL OSCILLATIONS INTO BASEL II BEHAVIOURAL SCORING MODELS

Goran KLEPAC, PhD
Raiffeisen bank Austria Ltd, Zagreb

Article*
UDC 657.92(035)
JEL G21

Abstract

The article introduces a new methodology of temporal influence measurement (seasonal oscillations, temporal patterns) for behavioural scoring development purposes. The paper shows how significant temporal variables can be recognised and then integrated into the behavioural scoring models in order to improve model performance.

Behavioural scoring models are integral parts of the Basel II standard on Internal Ratings-Based Approaches (IRB). The IRB approach much more precisely reflects individual risk bank profile.

A solution of the problem of how to analyze and integrate macroeconomic and microeconomic factors represented in time series into behavioural scorecard models will be shown in the paper by using the REF II model.

Keywords: credit scoring, REF II, time series analyze, data mining, temporal influence, seasonal oscillation, Basel II

1 Introduction

The basic scoring models within the Basel II standard, according to the internal rating system, are application scorecards. Their basic task is the initial recognition of client risk during the process of considering what the credit risk attending a loan made to this client might be.

* *Received:* June 1, 2007
Accepted: September 4, 2007

Behaviour scorecards are used for predicting potentially risky behaviour (failure to pay liabilities due) on the part of users of risky banking products and services as defined by Basel II standards. For the purpose of the construction of such models information about the behaviour of these clients in the past is used, and an attempt is made to predict the likelihood of risky behaviour in the future, in the continuation of use of these products and services, on the basis of the patterns revealed in the data samples (Thomas, 2002).

With the employment of such models, financial institutions can more easily manage their portfolios. As a portfolio matures over time, the properties of it will change either because of the entry of new clients into the portfolio, or because of the effects of micro-economic and macro-economic factors. Hence, it is necessary periodically to measure these effects on the risks inherent in the portfolio, as well as the strength of these effects.

For this purpose we use behaviour scoring models .

In the development of behaviour scoring models, an integral part of the Basel II standards, on the whole, for predictive purposes, we use derived attributes that demonstrate certain forms of behaviour over some time period, presented in the forms of value variables. This kind of methodology is suitable for the design of scoring models with the use of the usual methods such as binominal logistic regression, neural networks, decision trees and so on.

The traditional approach to the development of behaviour scorecards does not on the whole deal with issues related to the effects of seasonal oscillations on certain forms of risky behaviour (delays in repayments). In the literature, authors are on the whole focused on the creation of a data sample for the construction of a model in which it neglects any seasonal influences on risky behaviour (Siddiqi, 2005; Thomas, 2003). If we want a stable sample, this kind of approach is justified, but it does raise additional issues.

If seasonal oscillations have a strong effect on the portfolio of a client for risky services, and if we nevertheless ignore these seasonal influences, we may well develop a non-transparent scoring model for an outcome period that is shorter or longer than a year. The question arises of why we should not measure these influences and, if they should turn out to be significant, incorporate them into the model in the form of predictive variables, for in this case they will be bound to contribute to the reliability and greater stability of the model.

The key target variable, the state of late payments (defaults) can be observed as a time series, and it is possible to measure seasonal influences, and integrate them into the scoring model in proportion to the degree of impact revealed.[1]

Furthermore, macroeconomic phenomena such as exchange rate fluctuations, increased rates of unemployment and falls of solvency can also have an important effect on variations of the degree of credit risk.

In the case of the estimated influence of macroeconomic factors presented in the form of temporal series on the risky behaviour of clients, by integrating them as predictive variables, we can also increase the degree of transparency and reliability of the scoring mo-

[1] According to Basel II standards default is lateness in meeting monetary liabilities of 90 or more days.

dels that take into account debtor behaviour (for example, softening of the kuna against the euro will result in an increased risk of repayment of loans with a foreign currency clause expressed in euros).

One of the possible reasons for the traditional habit of avoiding modelling seasonal influences and macroeconomic indicators presented in the form of time series lies in the fact that there are problems about how to integrate time series into classic scoring models. These problems do not come out only in scoring, but are also characteristic and frequently mentioned in data mining (Berry, 1997; Han, 2001; Pyle, 2001; Williams, 2002).

In order successfully to settle the problem of integrating time series into classical scoring models and to increase the degree or reliability of the said models, the REF II model of temporal series transformation will be employed (Klepac, 2005; 2006).[2]

The introduction of seasonal variables into scoring models that take debtor behaviour into account requires a new approach to the creation of the sample for the implementation of the analysis, because in the outcome period it is necessary to devote additional attention to the period in which the debtor gets into default. This means it is necessary to record not only whether a state of default has arisen in the outcome period, but also when.

2 Transformation of time series in the REF II model

A precondition for the integration of seasonal variables into scoring models is that they be transformed into an REF II model, which takes place in several steps (Klepac, 2005; 2006).

We can state a time series to be a series of values S (S1…Sn), where S represents the time series (S1…Sn) elements of the series S. Let us look how this appears in eight basic steps:

- Temporal interpolation (1)

The formation of an independent temporal sequence Vz (vi1, ….,vin) at the interval <1..n> (days, weeks, months, quarters, years) with values 0. On the basis of a sequence formed in this way it is necessary to interpolate the missing values into S (S1,…,Sn) with 0 on the basis of the Vi sequence formed. The result of this treatment is the sequence S(s1,..,Sn) with interpolated missing values from the sequence Vi(Vi1,..,vin).

- Temporal granulation (2)

In this step we define the degree of the condensation of the temporal series S(s1,..sn) that is located in the elementary temporal unit (the day, week, month…). We condense elements of the existing temporal series with the use of statistical functions such as sum, average values and mode at the level of the granulated segment. In this way we can reduce the time series to a greater degree of granulation (days into weeks, weeks into months), and obtain a time series S(s1,..,sn) with a great degree of granulation.

Depending on the objectives, we can return to this step during the analysis process, which assumes the obligatory repeat implementation of the processes described in the following steps.

[2] REF is an acronym derived from Rise, Equal Fall, while II marks the second generation model.

- Standardising (3)

The standardising procedure implies the transformation of the time series S(s1,..,sn) into the time series T(t1,..,tn), in which every element of the sequence is subject to the min-max standardisation procedure at the interval of <0,1> as follows

– Time series T consists of the elements (t1,..,tn), in which ti is counted as ti=((si-min(S))/max(S)-min(S)), where min(S) and max(S) are the minimum and maximum values of time series S.

– The time shift between the elementary samples (the criterion of temporal complexity) of the segment of the X axis is defined as s d((ti,ti+1)=a

- Transformation into REF notation according to the formula Tr=ti+1- ti Tr > 0 =>R; Tr< 0 =>F; Tr=0 =>E , where Yi are elements of the sequence Ns (4)
- Computation of slope on the basis of the angle (5)

Coefficient of angle movement=>

Tr > 0 (R) Coefficient = t i + 1 - t i

Tr < 0 (F) Coefficient = t i - t i + 1

Tr = 0 (E) Coefficient = 0

- Computation of area under the curve (6)

Numerical integration using the rectangle method

$$p = ((t\ i*a)+(t\ i+1*a))/2$$

- Creation of time indexes (7)

The construction of a hierarchical tree of indexes depending on the character of the analysis, where an element of the structured index can be an attribute such as client code.

- Creation of classes (8)

The creation of derived values of attributes on the basis of the area below the curve and movement of angle. It is possible to create classes with the use of crisp or fuzzy logic.

These eight fundamental steps are the basic for the algorithmised procedure on which the REF II model is based; the ultimate result of this is the formation of a transformation matrix. The transformation matrix is the basis for the implementation of subsequent analytical procedures in the analysis of the time series.

The REF II model is at base a concept composed of three subunits (REF, area beneath the curve, and the coefficient of angle movement), the primary objective of which is the transformation of a time series into a sequence of indicators that unambiguously define the time series. REF, area under the curve, and coefficient of angle movement are indicators that do unambiguously describe the segment of the time series, and a number of such sections will together make up the transformed time series. Such a sequence of indicators is sequenced according to the order appearing in the unit of time and united into a joint conceptual structure of a transformed time series that we can call the transformation matrix.

A depiction of this structure is given in Table 1:

Table 1 Transformation matrix

Time segment index	I1	I2	…	In
REF notation	REF 1	REF 2	…	REF n
Coefficient of angle movement	KKO 1	KKO 2	…	KKO n
Area of time segment	P1	P2	…	Pn

Source: authors calculation

Time segment indicators are calculated on the basis of the coordinates of the two adjacent values in the time series. Thus for example a section with the index I1 is formed on the basis of the value of the coordinates of time series t0 and t1. The time segment index is used for an unambiguous identification of the time segment so that it can be analysed, i.e., so that the basic sample can be created.

Indexes can be complexly structured and can contain hierarchical elements, belonging elements, and conjunctive elements to other data sources.

The elements covered in the preceding table represent the basis of the REF II model, and with them it is possible to describe a curve unambiguously and to carry out all the analysis for the sake of which the model was developed. As well as those described it is possible to comprehend derived indicators, but that depends on the character of the analysis. After the transformation of the time series via REF II, we can obtain the transformed time series as shown in Table 1.

3 Measurement of microeconomic and macroeconomic seasonal impacts on meeting liabilities

Timely meeting of due liabilities of risky banking services may depend on the profile of the client and the part of the year in which the user has to meet its liabilities. There must be some profiles of clients whose behaviour is determined by, for example, the season of the Christmas and New Year festivities, or by the season of the summer holidays.

These impacts on the risk of meeting liabilities that have fallen due is more or less expressed depending on the market segment that constitutes our client portfolio. Using the REF II model we can measure how much promptness in repayment of loans is affected by seasonal factors in the segment for which the behavioural scoring model was developed. We can classify this kind of impact among microeconomic temporal factors for they are characteristic of a given segment at the level of the financial institution observed.

Exchange rate oscillations can also have an important effect on the promptness with which due liabilities are met with respect to risky banking services bound to a foreign currency. Exchange rates can be seen in the form of time series, and via the REF II model their impacts on the risk of liabilities not being met can be measured. We can classify this kind of impact among macroeconomic temporal factors. As well as exchange rate oscillations, we can measure trends in the rate of unemployment, or perhaps the liquidity of the sector in which the risky services work (on the basis of segmentation) and thus it is possible to measure the promptness with which due liabilities are met. The Basel II methodology allows the development of scoring at the level of certain asset class segments, where among other things sectors of the economy can be criteria for the implementation of segmentation.

These indicators can be combined with the traditional variables that are used during the development of scoring models that take into account debtor behaviour such as: number of defaults in a given period, total number of days late in a given period, number of days elapsed since emerging from the last default state and so on.

During measurement of the relevance of attributes and univariate analysis we usually rely on weight of evidence (WE) and information value (IV) (Siddiqi, 2005; Thomas, 2002).

Although traditional techniques for developing scoring imply the application of these variables, it is important to mention that after their computation for the understanding and profiling of the conduct of a risky client it is essential to find out the logical connection among the analysed factors and the default indicator. The possibility for the logical interpretation of the risk factors is a crucial component while choosing the variables for the creation of the model. On the whole the aim is to produce a logical explanation of the dependence of each of the variables that form part of the model.

The introduction of seasonal variables into scoring models can be very difficult.

Table 2 Univariate analysis of a time series with the use of the REF II model

Months	Default		Analysis of relevance	
	Rising trend (R)	Falling trend (F)		
	(%)	(%)	WE	IV
January	8.0	4.0	-0.693150	0.027726
February	10.0	9.0	-0.105360	0.001054
March	13.0	6.5	-0.690400	0.044752
April	6.0	7.0	0.154151	0.001542
May	6.7	9.0	0.295117	0.006788
June	9.0	10.0	0.105361	0.001054
July	9.3	36.0	1.353505	0.361386
August	5.0	6.5	0.265116	0.004024
September	3.0	8.0	0.980829	0.049041
October	20.0	2.0	-2.302590	0.414465
November	5.0	1.0	-1.609440	0.064378
December	5.0	1.0	-1.609440	0.064378
				1.040586

Source: authors calculation

In order to avoid this problem it is necessary to transform a time series (default in terms of months) into the REF II model. A time series of default states is created as an aggregation function of the number of default states and cases where no default at the monthly level is recorded. In order to present the methodology, we transform monthly trends of defaults for the previous seven years into REF II notation, which enables the implementation of an analysis of the relevance of attributes to the traditional manner within the scoring model with the use of WE and IV measures. Table 2 illustrates the procedure of univariate analysis of a time series with the use of the REF II model according to formulae / procedures (1), (2), (3), (4) and (5).

From this table it can be seen (IV = 1.04) how variables of the trend in default state at the monthly level are significant for the creation of debtor behaviour scoring models. It is also seen that during April, May, June and July (WE criteria), we have an expected smaller number of irregular repayments of loans, while at the end and the beginning of the year we can expect a greater probability of irregular loan repayments.

From the microlevel, we can hypothetically explain these end-of-year trends by the New Year holidays and the profile of the clients for loan services within the observed profile. The question arises of how to explain the regular repayments during the summer period. We can assume that this analysis was made on the basis of loan contracts containing a foreign currency clause expressed in euros. If we wanted to carry out a more profound analysis we might analyse the trend in the euro as against the kuna, that is, in other words analyse the seasonal oscillations in this trend.

Table 3 Analysis of seasonal oscillations of euro exchange rates against the kuna with the use of the REF II model

Months	REF Euro (F)	REF Euro (R)
January[a]	11.1	77.8
February	37.5	62.5
March	62.5	37.5
April	87.5	12.5
May	100	0
June	87.5	12.5
July	87.5	12.5
August	37.5	62.5
September	37.5	62.5
October	37.5	62.5
November	37.5	62.5
December	37.5	62.5

[a] *The 11.1% in January 1999 is an indeterminate trend because the values for the rate in December 1998 are not available.*

Source: authors calculation

The basic transformation into the REF II model is carried out on a time series of euro and kuna exchange rates in the period from January 1999 to January 2007. The time series of exchange rates at monthly levels is obtained as an average of daily mean values of the CNB exchange rate per month in this period. The time series so formed is transformed into REF II notation according to formulae / procedures (1), (2), (3), (4), (5).

For an analysis of seasonal oscillations of the exchange rate a table of the following structure is required.

From this table it can be seen that we can determine with a great degree of certainty that during May the euro will soften against the kuna, and that this trend will with a smaller degree of probability continue through June and July, also being present in April. This analysis shows us the zones of seasonal oscillations. Depending on the acceptable degree of reliability we can accept the hypothesis concerning seasonal oscillations. If this degree of reliability is greater than 85%, then we can accept the seasonal oscillation of the exchange rate hypothesis in the April-July period. Since the REF II model can transform a time series with much greater precision we can make this analysis more accurate by taking into account the area under the curve segment and the precise deviations of the trends. Since our objective here is to present the methodology for integration time series into scoring models, we shall direct our attention primarily to this simple transformation. Of course, we can use the methodology in more complex transformations as well. On the basis of the results obtained we can erect a hypothesis concerning the impact of an oscillation in the exchange rate on the promptness of repayments of loans with a euro-denominated exchange rate clause.

Apart from the impact of exchange rate oscillations, we can also measure the impact of the rate of unemployment of a certain sector in which loan customers prevail, and in the same way we can include other macroeconomic factors into the model, such as solvency, retail price trends and so on.

4 Predictions on the basis of seasonal factors

It is possible to apply the method of grouping and creating dummy variables during the development of scoring models that do not take seasonal influences into account in the model in which seasonal influences are included.

Dummy variables can be created on the basis of the seasonal variables, in such a way that the dummy variable that relates to, for example, the January-March period takes the value of 1, if it belongs to the January-March period, otherwise it will acquire the value of 0.

The created variables enter the model as predictors after the implementation of correlation analysis and then take part in the creation of the scoring models.

Although situations that correspond with the seasonality of the time series are taken as examples, for a prediction of default state it is possible on the basis of temporal attributes to build in a kind of what-if mechanism. If for example we analyse the rate of solvency within a certain sector that is dominant in the observed portfolio as employer, then this variable can be shown in temporal form.

With a univariate analysis of a time series so created, according to the previously shown example, we can analysis the size of the effect of a fall on the rate of solvency within a given sector at the macroeconomic level on the promptness of repayment of loans within a given portfolio.

This information can help us in future predictions (which is actually the task of scoring models that taken borrower behaviour into account) to determine with greater precision the degree of risk in our portfolio in the event of some future change in the rate of solvency in the given sector. In this manner, debtor behaviour-based scoring models can, in addition to their fundamental tasks, serve for the creation of scenarios if in this way we can integrate into them a sufficient number of macroeconomic variables. And not all the macroeconomic variables necessarily have to take on a temporal character. This approach that depends on the creation of a scenario enables us to recognise behaviour models and the vulnerabilities of segments of our portfolio, which can be a signpost for us while making strategic decisions. This approach is also in line with the stress testing method of portfolio management.

5 Assessment of the reliability of behaviour scoring methods with integrated seasonal predictions

In order to be able to test out the hypothesis that better scoring results can be obtained when seasonal predictions are integrated a testing model without the prediction of seasonal variables and a model in which seasonal variables are included were developed. For this purpose data runs were used (Pyle, 2001; Thomas, 2002; Modelandmine, 2007 and Kdnuggets, 2007).

Each data sample is divided into a training sample and a testing sample, in the ratio of 80:20. The trend in default states is interpreted as a time series and transformed into an REF II model. Univariate analysis showed the seasonal character of this variable with an important significance pursuant to the IV value. After the creation of a dummy variable pursuant to the significant characteristics, correlation analysis was carried out.

In the training sample, with the help of binominal logistical regression two scoring models were made, with the proviso that for the development of the first model the temporal model was not and for the development of the second model the temporal variable was used.

The reliability of each model individually was tested on the testing sample, and the results of the analysis of the model from the first data run are given in the following table.

Table 4 Testing the results of modelling with or without predictions of seasonal variables (%)

Kind of test	Credit scoring model		Difference in the results
	Without predicition of seasonal variables	With prediction of seasonal variables	
Kolmogorov-Smirnov test	61.22	62.83	+1.61
Gini index	71.4	73.51	+2.08

Source: authors calculation

From this table it can be seen that the model in which seasonal variables are predicted gives more reliable results than the model in which they are not included.

Three more data samples from the sources mentioned were also tested out, and the results obtained confirm the hypothesis erected.

It is necessary to take into account the effect of seasonality apparent in an outcome period longer or shorter than a year because within the outcome period of precisely within a single year we will eliminate the effect of seasonality.

6 Conclusion

Although during the construction of behavioural scoring models there tends to be the effect of neglecting seasonal impacts via the maximum enlargement of the outcome periods, this approach is inappropriate when we want to predict the debt-servicing promptness of customers for risky products for periods that are shorter than a year (three months, for example) or that do not overlap with the calendar beginnings and endings of the year. In this case for more exact prediction, seasonal factors and influences are significant for us.

Furthermore, seasonal factors can improve the results of scoring models, as is visible from the presented reliability assessment of models in which seasonal impacts are and are not predicted. Models in which seasonal influences are predicted give better results, particularly within shorter observed outcome periods.

The proposed model heightens rather than flattens seasonal factors, which is particularly important in the case of outcome periods of less than a year, and in this manner it allows more precise predictions.

The methodology proposed, furthermore, gives a solution of how to integrate into scoring models time series as predictors, introducing into them both macroeconomic and microeconomic factors. This is particularly significant since it facilitates, through the scoring models, a better understanding of portfolio structure, the patterns of behaviour of members of the portfolio, and gives answers about how certain factors (including macroeconomic) have a direct effect on the concrete portfolio.

This methodology can be used during the development of behavioural scoring models for risk evaluation with respect to potential frauds in financial operations, and in all of the behavioural scoring models from the domain of the financial sector related to the Basel II standards.

The methodology does not necessarily need to be applied to risk assessment in the strict sense, but can be used during the creation of a strategy for increasing market share. Thus for example, while a marketing campaign is being devised, during which the wish is to attract new clients to a financial institution, risk profile criteria have to be considered while the target group is being defined. In this manner we have to aim for the optimisation of the risk in the portfolio over a longer period if this is, for example, the strategic decision we have made.

LITERATURE

Basel Comitee on Banking Supervision, 2001. *Overview of the new capital accordp*. Basel: Bank for International Settelments.

Berry, M. J. A. and Linoff, G., 1997. *Data mining techniques for marketing sales and customer support*. San Francisco: John Wiley & Sons Inc.

Han, J. and Kamber, M., 2001. *Data mining-concepts and techniques*. London: Morgan Kaufmann publishers.

Kdnuggest, 2007. http://www.kdnuggets.com/datasets/index.html.

Klepac, G. and Panian, Ž., 2006. *Poslovna inteligencija.* Zagreb: Masmedia.

Klepac, G., 2005. *Otkrivanje zakonitosti temeljem jedinstvenoga modela transformacije vremenske serije*. Doktorska disertacija. Varaždin: Fakultet organizacije i informatike.

Lee, Y. [et al.], 1997. *Solving Data Mining Problems Through Pattern Recognition*. New York: Prentice Hall.

Modelandmine, 2007. http://www.modelandmine.com/dataprep.htm.

Naeem, S., 2005. *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring.* San Francisco: John Wiley&Sons Inc.

Ozsu, T. and Chen, L., 2003. "Multi-Scale Histograms for Answering Queries over Time Series Data" in: *Proceedings of the 20th International Conference on Data Engineering, March 30* - April 2. Boston, MA.

Pyle, D., 2001. *Data preparation for data mining*. London: Morgan Kaufmann publishers.

Thomas, L., Edelman, D. and Crook, J., 2002. *Credit Scoring and Its Applications*. Philadelphia: SIAM.

Wang, W. and Yang, J. Yu P., 2001. "Mining long sequential patterns in a noisy environment" [online]. *IBM research report.* Available from: [http://www-sal.cs.uiuc.edu/~hanj/pdf/ww1sigmod02_1.pdf].

Williams, J. G., Weiqiang, L. and Mehmet, A. O., 2002. "An Overview of Temporal Data Mining". Proceedings of the 1st Australian Data Mining Workshop, Canberra, 2002.