Multi-dimensional time series-based approach for banking regulatory stress testing purposes: Introduction to dualtime dynamics
Abstract

Under the regulatory paradigm of banking risk management, banks are required to perform stress testing of internally computed risk parameters to ensure they are holding adequate capital to offset the effects of downturn events. For this purpose, most of the contemporary stress-testing practices are limited to one dimension of the calculation, where endogenous risk parameters are predicted by modeling and scenario-based values of exogenous parameters (macroeconomic variables). This approach inherently fails to consider the simultaneous impact of other endogenous variables in predicting the risk factors. This real-life limitation is approached from a multi-dimensional time series standpoint. A multi-dimensional time series approach is adopted to combine the impacts of natural portfolio dynamics (endogenous characteristics) and macroeconomic performances (exogenous characteristics) to model and subsequently predict the portfolio performance.

As part of this approach, a vintage-level model is introduced, wherein customer vintage and age in the portfolio are considered to be additional endogenous characteristics contributing to portfolio performance. The approach has been tested on live data and it has been observed that the proposed model is more accurate in predicting the portfolio performance than other contemporary approaches such as one-dimensional models, generalized additive models (GAM), cross-sectional models, two-way proportional hazard models and age-period cohort (APC) models. This approach has also been adopted and tested on several historical downturn events and it has successfully and accurately predicted the occurring events.

Problem statement

The dynamics underlying retail banking portfolios are far from simple linear systems. For example, a model to predict for purposes of measuring capital might employ key risk identification parameters such as default rate (DR), Probability of Default (PD), Exposure at Default (EAD), Loss Given at Default (LGD) and Active Account Rate (AAR). The individual impact of these risk parameters cannot be pre-assumed but must be derived analytically.

Components of portfolio performance can include:

- Vintage life cycle: Maturation (age based)
- Seasonality (Exogenous: time-based)
- Management actions (Exogenous: time-based)
- Competitive and economic environment (Exogenous: time-based)
In contemporary practices (including global and local regulatory guidelines), single equation-based regression models and scenario-based assessment techniques are recommended to predict an endogenous variable (dependent variable) by modeling fluctuations in exogenous macroeconomic variables (independent variables). This technique fails to consider the impact of other independent endogenous variables in performance prediction. In banking portfolio performance prediction, both endogenous (natural portfolio dynamics) and exogenous (macroeconomic parameters) characteristics either jointly or independently impact the portfolio performance.

The question is, how does one combine the impacts of natural portfolio dynamics (endogenous characteristics) and macroeconomic performances (exogenous characteristics) in determining the predictive portfolio performance?

**Approach**

Since the prediction of portfolio performance is a time-driven event (trend is modeled by using historical information and potential occurrence of a scenario is used to predict futuristic portfolio performance), this paper attempts to solve the problem by using time series analysis.

Time series is an ordered sequence of values for a variable at equally spaced time intervals. Using the time series model in addressing the aforementioned problem can be twofold:

- Obtain an understanding of the underlying forces and structure that produced the observed data
- Fit a model and proceed to forecasting, monitoring or even feedback and feed-forward control

A conventional/basic time series model looks like this:

\[ y_t = x_t \beta + \epsilon_t, \quad t=1,2,\ldots,T \]

where \( \epsilon_t \) stands for the residual or error terms of a single equation-based regression model. In the modern view, the error terms can also be modeled, assuming that the residuals or errors in the model follow a first-order autoregressive process.

\[ \epsilon_t = \rho \epsilon_{(t-1)} + \delta_t, \quad \text{where} \ -1 < \rho < 1 \]

Time series patterns can be described in terms of two basic classes of components: trend and seasonality. The former represents a general systematic linear or non-linear component that changes over time and does not repeat. The latter may have a formally similar nature; however, it repeats itself at systematic intervals over time.

These two general classes of time series components are expected to coexist in real-life historical performance data for a retail bank. For example, customer probability of default in the credit card portfolio of a retail bank can rapidly grow during stressed periods (economic downturn), but they may still follow consistent seasonal patterns (e.g., as significantly low default tendency during festive events such as Halloween, Thanksgiving, Christmas, etc). In real life, modeling these two trends together is not easy, since a lot of performance-related data problems are multivariate and dynamic in nature [1]. For example, how is the performance of a mortgage portfolio related to the aggregate economic performance of the country? In this example, it is possible to write a single equation by considering customer default as the dependent variable and macroeconomic parameters as independent variables. But it is likely that in this example there is simultaneity, and that potentially there exists a second equation between the roles of independent and dependent variables.
In the above example, macroeconomic performance indicators are exogenous, whereas default tendency is an endogenous variable. One would expect that additional factors that may explain change in the composition and sensitivity of the portfolio are endogenously and dynamically related to the portfolio performance.

The conventional practice of single equation-based regression models (for predicting portfolio performance) generally ignores the fact that for endogenous dynamic relationships, there is either explicitly or implicitly more than one regression equation [1]. One may choose to continue estimating a single regression and hope that statistical interferences are not too flawed, or decide to estimate a multiple-equation model using a variety of techniques. The need for these dynamic multiple-equation models stems from two very common realities in the risk prediction models. First, variables simultaneously influence one another, so both are referred to as endogenous variables. Second, when considering the relationship among multiple dependent variables (a multiple-equation system may or may not have the same number of endogenous or dependent variables as equations), the unique or identified relationships for each equation of interest can be made only with reference to the system as a whole. Properly determining these relationships requires that information from all equations be used. For identification, there must be enough exogenous variables, specified in the correct way, in order to connect all the equations in a system and the estimate. Estimation requires that exogenous variables from the entire system be used to provide the most unbiased and efficient estimates of the relationships among the variables as possible.

If the customer default rate (PD or DR: dependent variable) is considered to be impacted by two endogenous variables such as customer delinquency status and loan utilization ratio, and exogenous variables such as bureau variables, the equation may stand as [2],

\[
DR(a,v,t)=\beta_m(v)f_m(a)e^{ag(v)fg(t)}
\]

where \(DR(a,v,t)\) is the dependent variable, influenced by two endogenous variables such as loan utilization ratio \((a)\) and customer delinquency status \((v)\) and exogenous bureau variables, which are a function of calendar date \((t)\).

In this equation, \((a)\) is the utilization ratio of the customer and \(f_m(a)\) is the function of the utilization ratio.

\(\beta_m(v)\) and \(\alpha_g(v)\) are the functions of delinquency status \(v\)

\(f_g(t)\) is the exogenous function of calendar date \((t)\), - which can be modeled by using bureau parameters as independent variables.

A multiple-equation time series model [1] can be developed by considering the simultaneous equations (SEQ) paradigm. Model-building with SEqs is based on taking the representation of a single theory or approach and rendering it into a set of equations. Using a single theory to specify the relationships among several variables leads to the identification of exogenous and endogenous variables. The exogenous variables are those that are determined to be outside the system or are considered fixed (at a point in time or in the past), i.e., bureau variables, macroeconomic parameters, etc.

Individually, each of these endogenous and exogenous variables holds a relationship with the default rate (either linear or nonlinear). i.e.,
\[ DR_1 (a,t) = \omega_m (a) e^{\alpha t} \] ..........................................(1)

\[ DR_2 (v,t) = \beta_m (v) e^{\gamma v} \] ..........................................(2)

\[ DR_3 (t) = \epsilon + \nabla_1 t_1 + \nabla_2 t_2 + \nabla_3 t_3 + \cdots + \nabla_n t_n \] ..............(3), where \( t_1, t_2, t_3 \ldots \) are macroeconomic parameters and \( \nabla_1, \nabla_2, \nabla_3 \ldots \) are their respective coefficients.

Equations 1, 2, and 3 can be combined by using exogenous variables, i.e., calendar time (t) of macroeconomic parameters, to derive a consolidated equation:

\[ DR(a,v,t) = \beta_m (v) f_m (a) e^{\alpha g (v) g (t)} \] .........(4)

**Dual-time Dynamics**

**Introduction**

The question of “how does one combine the impact of natural portfolio dynamics (endogenous characteristics) and macroeconomic performance (exogenous characteristics) in determining the predictive portfolio performance?” can be addressed through the vintage concept.

Dual-time dynamics (DtD) [2] is a method of analyzing simultaneous time series effects on risk parameters. DtD operates on vintage data to create scenario-based forecasting models for retail loan portfolios. Vintage performance is measured at regular intervals from the origination date.

DtD separates loan performance dynamics into three components:

- A maturation function of months-on-books (endogenous),
- An exogenous function of calendar date
- A quality function of vintage origination date (endogenous)

Of these three, the exogenous function captures the impact from the macroeconomic environment.

Dual-time dynamics measures factors driving portfolio performance from historical performance data. The lifecycle, environment, and vintage quality components measured by DtD provides a unique view into the factors driving portfolio performance and serve as individual controls on scenarios that will drive future performance [2]. Traditional portfolio models assume that a predetermined set of variables drive portfolio performance. These models are biased towards the selected model variables and the performance period of the data used to train the model. DtD makes no assumptions about which factors drive portfolio performance. Instead, it measures performance along the dimensions of age, time, and origination date. Dynamics such as lifecycles and seasonality tend to be stable over time, enabling users to focus on marketing and economic scenarios to drive forecasts.

Separating portfolio drivers into lifecycle, vintage quality, seasonality, policy changes, and economics provides unprecedented flexibility in using scenarios to drive portfolio forecasts. Banks can choose economic indicators for forecasting and specify origination plans. Scenario components are underlaid to produce forecasts. Banks can run forecasts against multiple economic scenarios to stress test portfolios [3] and DtD can quantify the amount that each scenario component contributes to the forecast.

DtD studies the rate of events occurring in aggregate rather than individual events such as default or early repayment that occur at the account level. The idea is that the rate of events (r), is a function of the age (a) of the account, the vintage origination date (v), and the calendar time (t).

DtD separates loan performance dynamics into three components:

**Computation technique**

With DtD, the dependent variable (y) (performance rate can be Probability of Default (PD), Loss Given Default (LGD), Exposure At Default (EAD), Expected Loss (EL), Active Account Rate (AAR), etc) is represented as a combination of three separate functions:
Vintage-level Performance Rate = (Maturation Function of Month-On-Book) x (Exogenous Function of Calendar Date) x (Quality Function of Vintage).

Model structure and assumptions
• Assume a mathematical form of the model to estimate i.e. \( y(a,v,t) = f_m(a), f_g(t), \beta_m(v), \alpha_g(v) \)
• \( f_m(a) \) is the maturation function of MOB \( \alpha \)
• \( f_g(t) \) is the exogenous function of calendar date \( t \)
• \( \beta_m(v) \) and \( \alpha_g(v) \) are the quality functions of vintage \( v \)
• A potential form of the relationship could stand as: \( y(a,v,t) = \beta_m(v) f_m(a) e^{\alpha_g(v)f_g(t)} \)

Modeling fitting – Non-parametric estimation of \( f_m(a), f_g(t), \beta_m(v), \alpha_g(v) \)
• As the functional form of \( f_m(a), f_g(t) \) is unknown, the values of \( f_m(a), f_g(t) \) are to be estimated
• Estimation is done by using iterative non-parametric technique with a proper convergence criterion
• Proper convergence criterion to be set with a presumed error bound on Mean square Error (MSE), whose range may vary based on quality of data

Model fitting – Parametric estimation of \( f_m(a), f_g(t) \)
• Establish a parametric relationship between age and \( f_m(a) \)
• Build relationship between \( f_g(t) \) and macroeconomic factors
• Forecast maturity for a given age using \( f_m(a) \) model (using classical approach e.g. exponential smoothing)

Model execution – Estimation of \( y(a+1,v,t+1) \)
• Using the relationship of \( f_g(t) \) and macroeconomic factors to estimate \( f_g(t+1) \) under different stress scenarios (Bank developed / regulator guided)
• Get \( f_m(a+1) \) from forecast of maturity
• Plug the value of \( f_g(t+1) \) and \( f_m(a+1) \) in DtD model to estimate \( y(a+1,v,t+1) \) i.e. stressed risk factor

Results and interpretation
Figure 1 shows the distribution of the observed default rate across time and vintage. This is also the combined effect of maturation (credit life cycle: top plot in Figure 2), exogenous factors (environment, i.e., macroeconomic parameters: the bottom plot in Figure 2), and the vintage quality (credit
quality: the right-hand side, mid-plot in the above representation) on the default rate (dependent variable), which is decomposed into three mutually independent dimensions. Broad steps followed are:

1. Separate estimation of the independent effect of maturation, exogeneity and vintage quality with a non-parametric approach (as explained under Computation technique)

2. Model these independent effects under combined impact on the default rate through a parametric estimation approach

3. Express the default rate with the parametric forms of maturation, exogeneity and vintage effect

Below are the ways through which the results are attained:

**Parametric estimation of maturation curve**

- The non-parametric estimates of the maturation function, \( f_m(a) \) and the exogenous function are to be obtained from an iterative process and then modeled parametrically. \( f_m(a) \) would be modeled with age, \( a \) and \( f_g(t) \) with the exogenous macroeconomic variables

- These parametric models of \( f_m(a) \) and \( f_g(t) \) would be used for forecasting and scenario generation

- Theoretically the best fitting model for \( f_m(a) \) comes out to be a polynomial of degree 6 (as per analyses base data). However, forecasting the high-degree polynomial may yield misleading results because of high variance in the data near the tail (Figure 3); thus it is not recommended

- The \( f_m(a) \) graph has been split into two parts and models are fitted for both the parts separately as shown in Figure 3
Figure 3

\[ y = 3E-06x^2 + 9E-05x + 0.0008 \]
\[ R^2 = 0.8923 \]

\[ y = -8E-04\ln(x) + 0.004 \]
\[ R^2 = 0.847 \]
Parametric estimation of exogenous curve

- To obtain the parametric relationship of \( f_g(t) \) with exogenous factors, different macroeconomic predictors are collected from external sources.
- Due to the time series effect, the lag correlation with lag of 0 to 6 months is calculated. The factors which are highly correlated with \( f_g(t) \) are taken with corresponding lag.
- Log transformation on actual value of macro factors is also used to obtain an efficient set of predictors.
- The exogenous function of calendar time is modeled with selected macroeconomic factors.

By replicating the impact on macroeconomic parameters during historical downturn events, futuristic scenarios are generated by using the scenarios’ default rate (dependent variable), predicted by using the modeled exogenous curve.

- Simple linear regression is used to explain \( f_g(t) \) with the macroeconomic parameters.
- Based on the sample data, R-square of the \( f_g(t) \) model came out to be 66.15% (Figure 4).
- The comparison of actual and predicted default rate presented in Figure 5 explains the strong predictive power of the model.

![Graph showing actual vs predicted default rates](image-url)
Conclusion

The dual-time dynamics technique adopted for predicting retail portfolio performance can not only consider multiple time series effects across portfolio dynamics and environmental fluctuations on portfolio risk parameters, it also overlays additional layers above standard one-equation macroeconomic regression models, thus reducing modeling error residuals. Furthermore, since this approach assumes that within a given vintage customers share the same maturation and exogenous curves, granular environmental impact can be assessed in detail at independent vintage levels.

The DtD methodology has been tested across the globe on several portfolios, specifically on the retail segments. Its forecast remained consistent and apt through the 2001 global recession, the 2003 Hong Kong SARS recession, the great U.S. recession in 2009 and the 2009, global financial crisis. Furthermore, it has been used to successfully back-test the Asian economic crisis of 1997.
About Genpact

Genpact (NYSE: G) stands for “generating business impact.” We design, transform, and run intelligent business operations including those that are complex and specific to a set of chosen industries. The result is advanced operating models that foster growth and manage cost, risk, and compliance across a range of functions such as finance and procurement, financial services account servicing, claims management, regulatory affairs, and industrial asset optimization. Our Smart Enterprise Processes (SEPSM) proprietary framework helps companies reimagine how they operate by integrating effective Systems of Engagement℠, core IT, and Data-to-Action Analytics℠. Our hundreds of long-term clients include more than one-fourth of the Fortune Global 500. We have grown to over 67,000 people in 25 countries with key management and a corporate office in New York City. Behind our passion for process and operational excellence is the Lean and Six Sigma heritage of a former General Electric division that has served GE businesses for more than 16 years.

For more information, contact, banking.solutions@genpact.com and visit, www.genpact.com/home/industries/banking-financial-services

Follow us on Twitter, Facebook and LinkedIn.

© 2015 Copyright Genpact. All Rights Reserved.

References