# Multifactor dynamic credit risk model

Jaroslav Dufek[1], Martin Šmíd[2]

**Abstract.**

We propose a new dynamic model of the Merton type, based on the Vasicek model. We generalize Vasicek model in three ways: we add model for loss given default (LGD), we add dynamics to the model and we allow non-normal distributions of risk factors. Then we add a retrospective interaction of underlying factors and found a non-linear behaviour of these factors.

In particular, the evolution of factors underlying the DR and the LGD is assumed to be ruled by a non-linear vector AR process with lagged DR and LGD and their non-linear transformations.

We apply our new model on real US mortgage data and demonstrate its statistical significance.

**Keywords:** loss given default, default rate, credit risk.

**JEL classification:** G21, C58
**AMS classification:** 62P20

## 1 Introduction

Asking a question "Why do we do this?" the answer could be: because the risk which follows from the real estate market is bigger than what was expected. This was shown in a recent crisis a few years ago.

Our study is based on the famous Vasicek model. Modifications of the Vasicek model are in plentiful supply (for example [2], [5] or [1] etc.). In [1] is quite huge literature review for more details see there. But on the other hand just a few of these modifications are dynamic ones (for example [3]). So we will focus on a rising new dynamic model of the Merton type, based on the Vasicek model.

There is a vast amount of literature in this area of interest. By far the most famous and most frequently used model is the Vasicek model for default rate, see [6]. The Vasicek assuming a fixed LGD. There are many models for random LDG, see [5], [1], [2], [3] and references therein. The model in [3] is a dynamic one based on the Vasicek model.

Now we will mention several studies which react on recent crisis. In [7], [9] and [4] is shown that LGD non-linearly depends on house price index and its history, which is not surprising. In [8] is summary of current state of art at the mortgage risk modelling.

Our model is based on [3]. We use the same structure: models for the default rate and for the LGD are same. Our original contribution is the creation of sub-models for underlying factors. We want to model the situation when bank losses retrospectively affect the default rate. Then we obtain a non-linear model.

## 2 Definition of the model

We want to model a situation when we have one creditor (for example a bank) with a countable number $n$ of debtors (clients). The value of the $i$-th debtor's assets at time $t$ is $A_{i,t}$. We assume that each debtor pays a regular instalment $b$.

[1]Department of Probability and Mathematical Statistics, MFF UK, Prague, The Czech Republic, jaroslav.dufek.2@seznam.cz

[2]Institute of Information Theory and Automation, Academy of Sciences of The Czech Republic, martinsmid.eu@gmail.com

The default of the $i$-th debtor is a state when the value of assets decrease under a given threshold $B_i$. Then the definition of the probability of default at time $t$ is $P[A_{i,t} < B_i]$. We will use $DR_t$ – the default rate – very frequently. The default rate is a simple ratio

$$DR_t = \frac{number\ of\ defaults}{number\ of\ loans} \tag{1}$$

### 2.1 Model for DR – default rate

We assume that

$$\log A_{i,t} = \log A_{i,t-1} + \Delta Y_t + \Delta V_{i,t}, \qquad i \leq n \tag{2}$$

where $n$ is a number of debtors, $A_{i,t}$ is a value of assets of $i$-th debtor at time $t$, and $\Delta Y_t := Y_t - Y_{t-1}$, $Y_t$ is a common factor following the stochastic process.

We assume that the duration of the debt is just one period and that the value of assets in each period is

$$\log A_{i,t-1} = Y_{t-1} + V_{i,t-1} \qquad i \leq n, \tag{3}$$

where $V_{i,t}$ is r. v. specific to the $i$-th debtor. We assume that $\{V_{i,t}\}_{i\leq n,\ t\in\mathbb{N}}$ are mutually independent and independent with respect to $\Delta Y_t,\ t \in \mathbb{N}$.

From (2) and (3), and from the assumption of independence, we can obtain that conditional probability of the default of the $i$-th debtor at time $t$ for a given $\overline{Y}_t := (\Delta Y_1, \dots, \Delta Y_{t-1})$ is

$$P[A_{i,t} < B_i|\overline{Y_t}] = P[\Delta V_{i,t} + V_{i,t-1} < \log B_i - Y_t|\overline{Y_t}] = \Psi(\log B_i - Y_t), \tag{4}$$

where $\Psi$ is the distribution function of r. v. $V_{i,t}$ which is identically distributed with $\mathbb{E}V_{i,t} = 0$ and $\mathrm{var}V_{i,t} = \sigma^2,\ \sigma > 0$.

If we assume that the debts are identical for all periods - $\log B_{i,t} = b$, if we approximate $DR_t = \lim_n \frac{\text{number of defaults at time } t}{n}$ we may apply the Law of Large Numbers to the conditional probabilities from (3) (we can do this when $A_{1,t},\ A_{2,t}, \dots$ are conditionally independent with respect to $\overline{Y_t}$), then we obtain $DR_t = P[A_{i,t} < b|\overline{Y_t}] \doteq \Psi(b - Y_t)\ t \in \mathbb{N}$, further implying that (under assumption that function $\Psi$ is monotonic)

$$\Delta Y_t \doteq \Psi^{-1}(DR_{t-1}) - \Psi^{-1}(DR_t). \tag{5}$$

Let us note that $\Psi$ is a general distribution function but for our valuation we will assume that $\Psi(x) = \Phi(x)$, where $\Phi$ is a distribution function of a standard normal distribution.

### 2.2 Model for loss

Now we will introduce our model for loss of the bank. From formula (14) in [3] we have that

$$L_t = DR_t \cdot h(I_t), \tag{6}$$

where $L_t$ is the realised bank loss, $DR_t$ is the default rate and $I_t$ represents the price index of properties. From formula (17) in [3] we directly obtain

$$h(t) = \Phi(\frac{-t}{\sigma}) - \exp\{t + \frac{1}{2}\sigma^2\}\Phi(\frac{-t}{\sigma} - \sigma). \tag{7}$$

Justification (under the assumption that a property price follows geometric Brownian motion and $h(t) = 1 - RR(t)$, where $RR$ is a recovery rate) and valuation of the function $h$ is in Appendix in [3].

### 2.3 Analysis of function $h$

The function $h$ is one of the main pillars of our model; so the behaviour of this function is very important to us. The function $h$ is a convex-concave function, so its inflexion point is the most important to us because changes in house price index has the biggest impact to loss in neighbourhood of its inflexion point. After some algebra we obtain the first and second derivations:

$$h(t) = \Phi(\frac{-t}{\sigma}) - \exp\{t + \frac{1}{2}\sigma^2\}\Phi(\frac{-t}{\sigma} - \sigma), \tag{8}$$

$$h'(t) = -\exp t + \frac{1}{2}\sigma^2\Phi(\frac{-t}{\sigma} - \sigma), \tag{9}$$

$$h''(t) = \frac{-1}{\sigma\sqrt{2\pi}} \exp \frac{t^2}{2\sigma^2} - \exp t + \frac{1}{2}\sigma\Phi(\frac{-t}{\sigma} - \sigma). \tag{10}$$

But we can't obtain the inflexion point analytically because of the Equation (11), which is equivalent to $h''(t) = 0$.

$$\Phi(\frac{-t}{\sigma} - \sigma) = \frac{1}{\sigma}\varphi(\frac{-t}{\sigma} - \sigma). \tag{11}$$

We can obtain it only numerically. We know that the inflexion point is unique from the graphical solution.

### 2.4 Evolution of factors

When we consider about evolution of underlining factors we try to model situation that default rate depend on house price index, loss of the bank and theirs previous values and itself lagged values. Thus we assume that the number of people who are not able to pay their loan is growing significantly, the ratio of unpaid loans increases in all banks. Banks have their investments covered by real properties so they are losing part of their liquidity. If a bank wants to recover lost liquidity it must sell some of its real properties, if all banks chose this strategy, the value would decrease, equity would not be sufficient and the $LGD$ would increase.

We assume that common factors $Y_t$ and $I_t$ are driven by these equations:

$$\Delta Y_t = C_1 + a_1\Delta Y_{t-1} + b_1\Delta Y_{t-2} + \underbrace{c_1\Delta L_{t-3} + d_1\Delta L_{t-4}}_{\text{retrospective interaction}} + \\ + e_1\Delta I_{t-2} + \varepsilon_{1,t} \tag{12}$$

$$\Delta I_t = C_2 + a_2\Delta Y_{t-2} + b_2\Delta Y_{t-3} + c_2\Delta DR_{t-3} + d_2\Delta DR_{t-4} + \\ + e_2\Delta I_{t-1} + f_2 I_{t-2} + g_2\Delta I_{t-3} + \varepsilon_{2,t} \tag{13}$$

where $\varepsilon_t$ are iid independent, non-correlated and normally distributed.

## 3 Empirical results

We tested our proposed model on a real dataset which is described below.

### 3.1 Description of the data set

The dataset for our empirical work contains quarterly delinquency rates[1] on mortgage loans from the US economy, which are provided by the US Department of Housing and Urban Development and the Mortgage Bankers Association.[2] We used the Standard & Poor Price Index of properties. The data for the default rate starts in the first quarter of 1979 and ends in the first quarter of 2012. The data for house price index starts in the first quarter 1987 and ends in the third quarter of 2012. We will use our data only from the first quarter 1987 forward (due to missing values for the house price index prior to that date).

In Figure 1 we can see a peak in 2008 which corresponds to the recent crisis in 2009. That is quite interesting, because in Figure 2 we can see a peak in 2010 (values of foreclosures are in percent); so the peak in the house price index should indicate a peak in delinquency rates.

[1]The 90+ deliquency rate is the proportion of all receivables 90 or more days past in a given quarter

[2]The Mortgage Bankers Association is the largest US society representing the US real estate market, with over 2,400 members(banks, mortgage brokers, mortgage companies, life insurance companies, etc.)

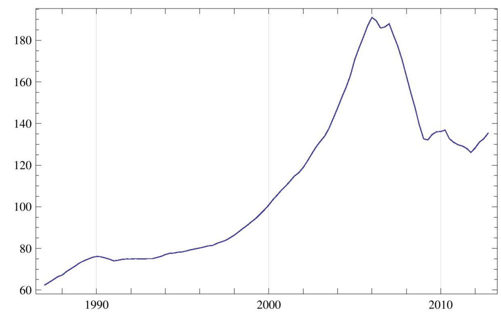


**Figure 1** The house price index - underlying factor $I_t$

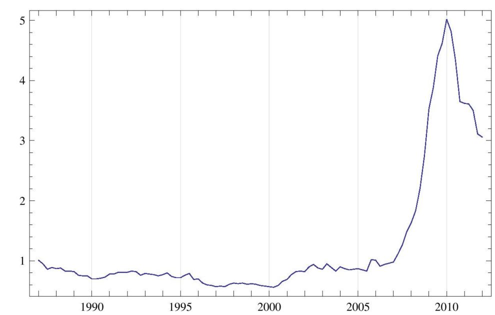


**Figure 2** The US 90+ Delinquency Rates - factor $DR_t$

### 3.2 Estimation

We took default rate $DR_t$ as the delinquency rate from the dataset; the factor $I_t$ was taken as the house price index from the dataset. Then we evaluated $\Delta Y_t$ according to Equation (5), where $\Psi(x) = \Phi(x)$ is a distribution function of a standard normal distribution. Then we evaluated the difference of $I_t$ and $L_t = DR_t \cdot h(\Delta I_t)$. Finally, we were fitted our model.

In Table 1 and Table 2 we can see estimation of coefficients from Equation 12 and Equation 13 (all coefficients are significant) with standard deviation and p-value obtained by t-test.

| | Coefficient | Standard dev. | p-value | |
|---|---|---|---|---|
| const | $-0.0030$ | 0.0026 | 0.2406 | |
| $\Delta L_{t-3}$ | 173.5630 | 32.5078 | 0.0000 | *** |
| $\Delta L_{t-4}$ | $-46.4339$ | 24.5574 | 0.0619 | * |
| $\Delta I_{t-2}$ | 0.0023 | 0.0009 | 0.0095 | *** |
| $\Delta Y_{t-1}$ | $-0.6236$ | 0.0923 | 0.0000 | *** |
| $\Delta Y_{t-2}$ | $-0.6092$ | 0.1011 | 0.0000 | *** |

**Table 1** Fitting of Equation 12 - dependent variable $\Delta Y_t$

## 4 Forecast

We have forecast the default rate $DR_t$, the loss given default $LGD_t = h(\Delta I_t)$ and the loss of the bank $L_t = h(\Delta I_t) \cdot DR_t$ for the following quarter, i.e., 2013Q3. The data for the default rate ends in 2012Q1 but the data for the house price index ends in 2012Q3. We constructed the forecast in two steps. In the first step we forecast the default rate up to 2012Q3 and in the second step we simultaneously forecast

| | Coefficient | Standard dev. | p-value | |
|---|---|---|---|---|
| const | 0.0989 | 0.1483 | 0.5066 | |
| $\Delta DR_{t-3}$ | 520.2640 | 183.2700 | 0.0056 | *** |
| $\Delta DR_{t-4}$ | −557.7900 | 192.9140 | 0.0048 | *** |
| $\Delta Y_{t-2}$ | 12.2806 | 5.5212 | 0.0286 | ** |
| $\Delta Y_{t-3}$ | 29.4992 | 8.7178 | 0.0011 | *** |
| $\Delta I_{t-1}$ | 1.0413 | 0.0971 | 0.0000 | *** |
| $\Delta I_{t-2}$ | −0.3774 | 0.1370 | 0.0071 | *** |
| $\Delta I_{t-3}$ | 0.2122 | 0.0989 | 0.0347 | ** |

**Table 2** Fitting of Equation 13 - dependent variable: $\Delta I_t$

the house price index and the default rate. The forecast is shown in Figure 3. The default rate is ratio form Equation 1, values of LGD and Loss are under assumption that exposure at default is unit.

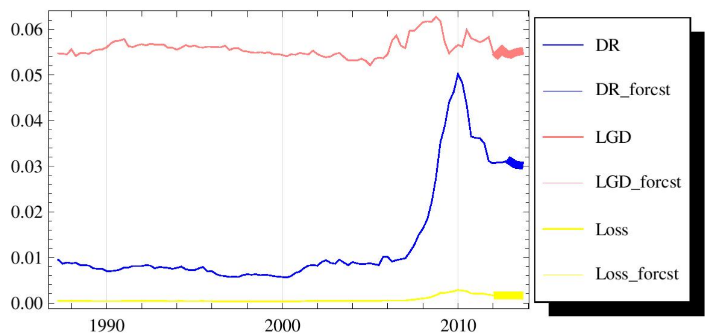


**Figure 3** Forecast of $DR_t, LGD_t$ and $L_t$ for 2013Q3

## 5 Conclusions

We generalised [3] model, changed a linear sub-model into non-linear one, showed the statistical significance of non-linear dynamics. We applied our model to the real data and construct the forecast. We think that non-linearity is the key property of our model.

There are several topic for future research the main one is study of properties of functional AR process, especially existence of stationary distribution.

## References

[1] Dullmann, K. and Trapp, M.: Systematic risk in recovery rates - an empirical analysis of U.S. corporate credit exposure. Working paper, Deutsche Bundesbank, Frankfurt, Germany.

[2] Frye, J. : Depressing recoveries. *Risk*,13(11): 106-111, 2000.

[3] Gapko, P. and Šmíd, M.: Dynamic Multi-Factor Credit Risk Model with Fat-Tailed Factors. Czech Journal of Economics and Finance, 62(2): 125–140, 2012.

[4] Park, Y. and Bang,D.: Bang. Loss given default of residential mortgages in a low ltv regime: Role of foreclosure auction process and housing market cycles. *Journal of Banking & Financeournal of Banking & Finance*, 39: 192–210, 2014.

[5] Pykhtin, MV : Unexpected Recovery Risk. *Risk*, 16(8):74–78, 2003.

[6] Vasicek, O.: The distribution of loan portfolio value. *RISK*, 15(12): $160 - 162$, 2002.

[7] Qi, M. and Yang, X.: Loss given default of high loan-to-value residential mortgages. *Journal of Banking & Financeournal of Banking & Finance*, 33(5): 788–799, 2009.

[8] Qi, M.: *Credit Securitizations and Derivatives: Challenges for the Global Markets* , pages 33–52, Mortgage Credit Risk, 2013.

[9] Zhang, Y., Chi, L., Liu, F. and Ji, L.: Local Housing Market Cycle and Loss Given Default: Evidence from Sub-Prime Residential Mortgages. *International Monetary Fund.* Working paper, 2010.