Non-linear Multi-factor Model

J. Dufek
Charles University in Prague, Faculty of Mathematics and Physics,
Prague, Czech Republic.

Abstract. We propose a new multi-factor dynamic credit risk model of the default rate (DR) and the loss given default (LGD). 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 data and show its significance.

Motivation

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 KMV Vasicek model. Modifications of the KMV Vasicek model are in plentiful supply [for example Frye, 2000; Pykhtin, 2003; Dullmann, 2004, etc.]. In Dullmann [2004] 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 Gapko, 2012]. So we will focus on a rising new dynamic model of the Merton type, based on the Vasicek model. This paper is organised as follows: the first part will be a short literature review followed by the definition of the model; in the last part of the paper we will focus on empirical results of our model.

Literature review

There is a vast amount of literature in this area of interest. By far the most famous and most frequently used model is the KMV Vasicek model for default rate Vasicek [see 2002]. And then there is the KMV assuming a fixed LGD. There are many models for random LDG [see Pykhtin, 2003; Dullmann, 2004; Frye, 2000; Gapko, 2012, and references therein]. The model in Gapko [2012] is a dynamic one based on the Vasicek model.

Our model is based on Gapko [2012]. 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.

Definition of the model

Our situation

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 $b$.

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 $DR_t = \frac{\text{number of defaults}}{\text{number of loans}}$.

Model for DR — default rate

We assume that

$$\log A_{i,t} = \log A_{i,t-1} + \Delta Y_t + \Delta V_{i,t}, \quad i \leq n$$ (1)
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

$$\log A_{i,t-1} = Y_{t-1} + V_{i,t-1}$$

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 (1) and (2), 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 $Y_t := (\Delta Y_1, \ldots, \Delta Y_{t-1})$ is

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

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

If we assume that the debts are identical for all the times: $\log B_{i,t} = b$, 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 $A_{1,t}, A_{2,t}, \ldots$ are conditionally independent with respect to $Y_t$), then we obtain $DR_t = P[A_{i,t} < b | Y_t] = \Psi(b - Y_t)$ $t \in \mathbb{N}$, further implying that (under assumption that function $\Psi$ is monotonic)

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

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.

Model for loss

Now we will introduce our model for loss of the bank. We assume that

$$L_t = DR_t \cdot h(I_t),$$

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 Gapko [2012] we directly obtain

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

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 Gapko [2012].

Evolution of factors

So when 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:

$$\Delta Y_t = C_1 + a_1 \Delta Y_{t-1} + b_1 \Delta Y_{t-2} + c_1 \Delta I_{t-3} + d_1 \Delta I_{t-4} + \epsilon_1 \Delta I_{t-2} + \epsilon_{1,t}$$

(7)
Dupek: Non-linear Multi-Factor Model

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

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

In Gapko [2012] \( I_t \) represents a latent common factor, but in our contribution \( I_t \) represents the house price index in our point of view seems straightforward that this factor should be the house price index.

Explanation for Equations (7) and (8) is that we think that default rate depend on house price index, loss of the bank and theirs previous values and itselfs lagged values. We decided that maximum lag should be one year and we have a quarterly data. So we started with a model where lags for all variables were 1, 2, 3 and 4. Then we dismissed (by one) all irrelevant variables (relevance threshold was 5%).

Empirical results

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

Description of the data set

The dataset for our empirical work contains quarterly delinquency rates on mortgage loans from the US economy, which are provided by the US Department of Housing and Urban Development and the Mortgage Bankers Association.\(^1\) 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 and the 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; so the peak in the house price index should indicate a peak in delinquency rates.

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 (4), where \( \Psi(x) = \Phi(x) \) is a distribution function of a standard normal distribution. Next we

\(^1\)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.)
computed the difference of $I_t$ and $L_t = DR_t \cdot h(\Delta I_t)$, and we were then able to estimate our model. We used expert judgement for estimate of parameter $\sigma = 1$. We now that’s one of the weaknesses of our model, which we want to focus on in a future research.

Table 1. Dependent variable: $\Delta Y_t$.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Standard dev.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>const</td>
<td>$-0.00301987$</td>
<td>$0.00255641$</td>
</tr>
<tr>
<td>$\Delta L_{t-3}$</td>
<td>$173.563$</td>
<td>$32.5078$</td>
</tr>
<tr>
<td>$\Delta L_{t-4}$</td>
<td>$-46.4339$</td>
<td>$24.5574$</td>
</tr>
<tr>
<td>$\Delta I_{t-2}$</td>
<td>$0.00229046$</td>
<td>$0.000864648$</td>
</tr>
<tr>
<td>$\Delta Y_{t-1}$</td>
<td>$-0.623598$</td>
<td>$0.0922916$</td>
</tr>
<tr>
<td>$\Delta Y_{t-2}$</td>
<td>$-0.609174$</td>
<td>$0.101066$</td>
</tr>
</tbody>
</table>

Table 2. Dependent variable: $\Delta I_t$.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Standard dev.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>const</td>
<td>$0.098875$</td>
<td>$0.148274$</td>
</tr>
<tr>
<td>$\Delta DR_{t-3}$</td>
<td>$520.264$</td>
<td>$183.270$</td>
</tr>
<tr>
<td>$\Delta DR_{t-4}$</td>
<td>$-557.790$</td>
<td>$192.914$</td>
</tr>
<tr>
<td>$\Delta Y_{t-2}$</td>
<td>$12.2806$</td>
<td>$5.52117$</td>
</tr>
<tr>
<td>$\Delta Y_{t-3}$</td>
<td>$29.4992$</td>
<td>$8.71783$</td>
</tr>
<tr>
<td>$\Delta I_{t-1}$</td>
<td>$1.04128$</td>
<td>$0.0971320$</td>
</tr>
<tr>
<td>$\Delta I_{t-2}$</td>
<td>$-0.377446$</td>
<td>$0.137017$</td>
</tr>
<tr>
<td>$\Delta I_{t-3}$</td>
<td>$0.212169$</td>
<td>$0.0989791$</td>
</tr>
</tbody>
</table>

Forecast

We have forecasted 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 forecast in two steps. In the first step we forecast the default rate up to 2012Q3 and in the second step we simultaneously forecast the house price index and the default rate. The forecast is shown in Figure 3.
Conclusions

We showed non-linear behaviour in our dataset, proposed a new model for the bank loss, and applied our model to the real data.

There are several topics for future research. The first is to assume that our data contains an error, for example: $I_t = HPI_t + \varepsilon_t$. The second is to develop a multi-periodic model, because our model is just a single-period model but mortgages usually last a long time period. The third one is to show a situation in which the value of the collateral decreases to less than the value of debt, because in the US the debtor can then give the real property back to the bank and take a new mortgage. And last but a very important topic is to study the theoretical properties of non-linear autoregressive processes, because we think that non-linearity is the key property of our model.

References