

# MPRA

Munich Personal RePEc Archive

## **On Models of Stochastic Recovery for Base Correlation**

Li, Hui

June 2009

Online at <https://mpra.ub.uni-muenchen.de/15750/>  
MPRA Paper No. 15750, posted 18 Jun 2009 04:45 UTC

# On Models of Stochastic Recovery for Base Correlation

Hui Li<sup>1</sup>

June 15, 2009

## Abstract

This paper discusses various ways to add correlated stochastic recovery to the base correlation framework for pricing CDOs. Several recent models are extended to more general framework. The pros and cons of these models for calibration to single name CDS and index CDO tranches are discussed. It is shown that negative forward recovery rate under fixed systematic factor appears in these models. This suggests that current static copula models of correlated default and recovery processes are inherently inconsistent.

Keywords: CDO, Gaussian Copula, Base Correlation, Stochastic Recovery, Correlated Loss Given Default

JEL classification: G13

---

<sup>1</sup> Email: [xhuili@yahoo.com](mailto:xhuili@yahoo.com). The opinions expressed in this article are the author's own and do not reflect in any way the views of his employer.

# 1 Introduction

Recent credit market turmoil has seen the spreads on the senior tranches of the CDX investment grade index widening so much that the standard Gaussian Copula model can not calibrate to the market any more. Especially for the 15% - 30% senior tranche and the 30% -100% super senior tranche, even 100% correlation would not be able to give the market spreads under the fixed 40% recovery rate assumption. Besides, market has also started to trade the super duper 60% - 100% tranche, which is in direct contradiction with the 40% fixed recovery rate assumption.

A quick fix that would still be consistent with index and single name CDS market would be to relax the deterministic recovery rate assumption and to introduce stochastic recovery rate into the base correlation framework while keeping the expected recovery rate at the same 40%. Historically, Andersen and Sidenius (2004) were the first to explore stochastic recovery in the Gaussian Copula framework. They modelled the recovery rate as a function of the systematic factors driving the default process and additional idiosyncratic factors that are different from the ones driving the default. The rationale is that empirical facts suggest a negative correlation between default probability and recovery rate. When default rate is high, recovery is usually low. There are some technical issues with the original specification of Andersen and Sidenius (2004). But the negative correlation is generally agreed to be an important feature of any stochastic recovery model. Recently, a number of specifications have been proposed along this line of thinking, see for example Krekel (2008), Amraoui and Hitier (2008) and Ech-Chatbi (2008). The Krekel model assumes that the stochastic recovery is driven by the same factors that trigger the default. The Amraoui-Hitier model assumes that recovery is a deterministic function of the systematic factor in the default triggering variable without reference to any idiosyncratic factors. The Ech-Chatbi model uses a multiple default process to model the recovery rate, but the correlation is still driven by the same default triggering variable. So it can also be viewed as a special case in the framework of the Krekel model.

However, the stochastic recovery specification may not be internally consistent. Specifically, negative forward recovery rate may appear when the systematic factor takes a large negative value. The present paper will try to prove that this is generally true for models within the current framework.

Correlated stochastic loss given default has also been considered in the Basel II framework to model the unexpected tail loss on a credit portfolio, see for example Frye (2000), Pykhtin (2003), Tasche (2004) and Witzany (2009). Although the purpose of the models is different, the specifications share some common features that worth looking at. The Tasche model is essentially the same as the Krekel model with the same default triggering variable driving the recovery, but it is set up in a more general, continuous manner. The Frye model has been discussed by Andersen and Sidenius, and was extended to a more sensible specification. The Pykhtin model can be viewed as a generalization of the Frye model but with more correlation parameters. The Witzany paper summarizes all these models and makes some suggestions to improve the Basel regulatory capital formula to cover the tail risk due to the correlated loss given default.

The rest of the paper is organized as follows. In section 2, we setup a general framework to add correlation to any exogenous stochastic recovery specification following Tasche (2004). In section 3, we generalize the Krekel model to include both discrete and continuous recovery distribution, which has been discussed by Tasche in a different form. In section 4, we discuss a general form of stochastic recovery driven only by the systematic factor and reveal the relationship between the Amraoui-Hitier model and the Krekel model. In section 5, we discuss the general form of the Frye model and the Andersen-Sidenius model. In section 6, we discuss the general form of the Pykhtin model. In section 7, the negative forward recovery rate under fixed systematic factor is discussed and a proof is given for the generalized Krekel model. Section 8 concludes the paper.

## 2 General framework

First we define the default indicator  $I = 1_{\tau < t}$  as a random variable taking values 0 if an obligor does not default before time  $t$  or 1 otherwise. Then the cumulative distribution function for  $I$  is

$$F_I(i) = P(I \leq i) = 1 - p + p \cdot H(i-1) \quad \text{for } i \in [0,1] \quad (1)$$

where  $H$  is the Heaviside step function and  $p$  is the probability of default before time  $t$ .

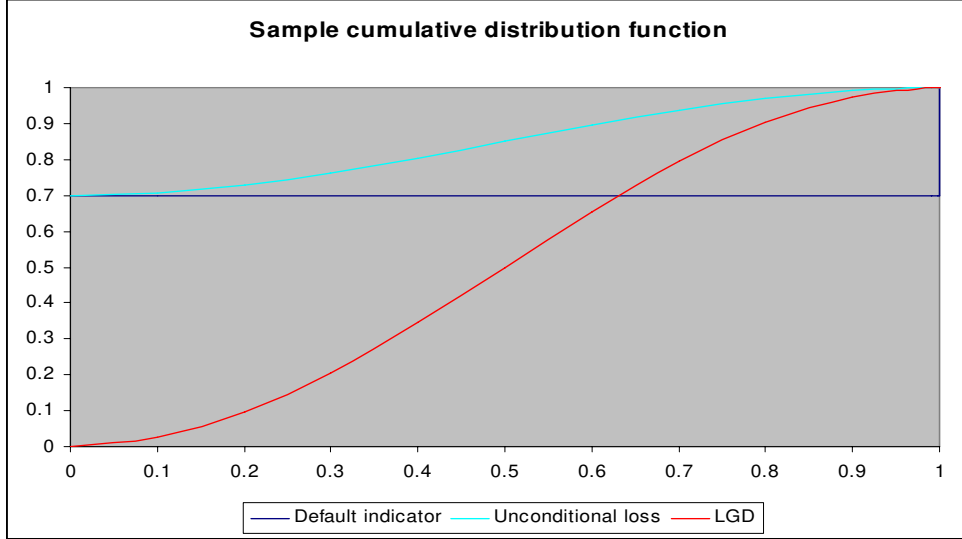
Let  $L$  be the unconditional stochastic loss as a percentage of the total exposure to an obligor. Then  $L$  will be zero with probability  $1 - p$  when the obligor is not in default.  $L$  will take non-negative values with probability  $p$  when the obligor defaults. Formally, the cumulative distribution function  $F_L$  of  $L$  has the following general form (see Tasche, 2004)

$$F_L(l) = P(L \leq l) = 1 - p + p \cdot F_D(l) \quad \text{for } l \in [0,1] \quad (2)$$

where  $F_D(l) = P(L \leq l | \text{default})$  is the cumulative distribution of loss given default. We will not make the assumption of hard default where obligor default is equivalent to loss greater than zero. So  $F_D(0) > 0$  is possible in the current framework. The default indicator  $I$  dominates the unconditional loss  $L$  in the sense that

$$F_I(x) \leq F_L(x) \quad \text{for } x \in [0,1] \quad (3)$$

This is the same as the statement that loss is conditional on default. The following chart shows sample cumulative distribution functions for default indicator, unconditional loss and loss given default (LGD) with default probability of 30%.



Note that the marginal cumulative distribution function of recovery upon default is

$$\begin{aligned}
 F_R(r) &= P(R \leq r | \text{default}) \\
 &= P(L \geq 1 - r | \text{default}) \\
 &= 1 - F_D(1 - r) + P(L = 1 - r | \text{default})
 \end{aligned} \tag{4}$$

It is obvious that the expected loss given default equals one minus the expected recovery. It is also easy to show that the variance of loss given default is the same as the variance of recovery.

$$\text{Var}(L) = E([L - E(L)]^2) = E([1 - R - E(1 - R)]^2) = \text{Var}(R) \tag{5}$$

Now we will try to add correlation between default probability and loss given default. Suppose obligor asset depends on a random variable  $V$ , which may have systematic factors and idiosyncratic factors in it. This kind of structural model is normally used in the Copula model for CDOs. Let  $\Phi(v) = P(V \leq v)$  be the cumulative distribution function of  $V$ . Assuming  $\Phi(v)$  is strictly increasing and has an inverse, which is normally the case for the distributions used in copula models, such as normal and Student  $t$  distributions. However,  $F_L(l)$  is not generally invertible as a function  $F_L : [0,1] \rightarrow [0,1]$ , since  $F_L(0) \geq 1 - p > 0$  and  $F_L$  is a step function for discrete distribution of loss. Since  $F_L$  is right continuous, one may still define, for  $y \in [0,1]$ ,

$$F_L^{-1}(y) = \inf_{l \in [0,1]} \{F_L(l) \geq y\} \tag{6}$$

It is easy to prove that the random variable  $\Phi(V)$  has the uniform cumulative distribution function

$$P(\Phi(V) \leq y) = P(V \leq \Phi^{-1}(y)) = \Phi(\Phi^{-1}(y)) = y \tag{7}$$

Now we can introduce the dependence of unconditional loss  $L$  on  $V$  as

$$L = F_L^{-1}(\Phi(V)) \quad (8)$$

This was discussed by Tasche (2004) to add correlated loss given default effect to the Basel II risk weighted capital charge formula. Here  $V$  can be interpreted as the negative change in obligor's asset value for certain time period. So this model makes the economic sense that loss is negatively correlated with asset value and thus positively correlated with default rate.

The marginal distribution of  $L$  will not change

$$P(L \leq l) = P(F_L^{-1}(\Phi(V)) \leq l) = P(\Phi(V) \leq F_L(l)) = F_L(l) \quad (9)$$

This way we can introduce correlation between default probability and loss given default, and correlation of loss given default between different obligors through the systematic factors in  $V$ . The default indicator can be also dependent on  $V$  as follows

$$I = F_I^{-1}(\Phi(V)) \quad (10)$$

In general, the random variable driving the loss may be different from the random variable that drives the credit default, but can be correlated through the same systematic factors or even through common idiosyncratic factors.

The same scheme can be applied to the conditional distribution function  $F_D$  for the loss given default, instead of the unconditional loss distribution function  $F_L$ . The loss given default  $L_D$  can be written as

$$L_D = F_D^{-1}(\Phi(V)) \quad (11)$$

to make it dependent on  $V$ . However, since default probability also depends on the systematic factors of  $V$  in this correlation framework, the true marginal distribution function of loss given default will be different from the function  $F_D$ . We will discuss this phenomenon in more details later. The choice of using  $F_L$  or  $F_D$  will be based on the requirement that loss is only meaningful upon default and potential loss without default is excluded from the model framework.

### 3 The Tasche/Krekel Model

The Krekel model provides a framework to add correlation to an arbitrary discrete marginal distribution of the random recovery. It can be easily generalized to any continuous marginal distribution as well. The idea is that the random variable driving loss given default is the same as the random variable driving the default event (Tasche, 2004). This will avoid certain technical issues with other models, but it is very restrictive.

Assume  $V_i = \sqrt{\rho}Z + \sqrt{1-\rho}\varepsilon_i$  drives the default of obligor  $i$  of a credit portfolio, where  $Z$  and  $\varepsilon_i$  are independent normal random variables  $\sim N(0,1)$  and  $Z$  is the systematic factor. The default event can be characterized as  $V_i \leq v_i = N^{-1}(p_i)$ , where  $p_i$  is the default probability of the obligor  $i$  and  $N(x)$  is the standard cumulative normal distribution function.

The dependence of loss  $L_i$  on  $V_i$  can be specified as  $L_i = F_L^{-1}(N(-V_i))$ , since here  $V_i$  is a proxy of the positive change in asset value. If  $-V_i$  does not have the same distribution as  $V_i$  like the normal distribution, we should use the distribution function for  $-V_i$ . Given  $z$ , the probability of default will be

$$\begin{aligned}
P_i(z) &= P(I_i > 0 | Z = z) \\
&= 1 - P(I_i = F_L^{-1}(N(-V_i)) \leq 0 | Z = z) \\
&= 1 - P(V_i \geq -N^{-1}(1 - p_i) | Z = z) \\
&= N\left(\frac{N^{-1}(p_i) - \sqrt{\rho}z}{\sqrt{1-\rho}}\right)
\end{aligned} \tag{12}$$

The cumulative loss distribution is

$$\begin{aligned}
P(L_i = F_L^{-1}(N(-V_i)) \leq l | Z = z) \\
&= N\left(\frac{N^{-1}(F_L(l)) + \sqrt{\rho}z}{\sqrt{1-\rho}}\right) \\
&= N\left(\frac{-N^{-1}(p_i(1 - F_D(l))) + \sqrt{\rho}z}{\sqrt{1-\rho}}\right) \\
&= 1 - P_i(z) + P_i(z) \cdot P(L_i \leq l | \text{default}, Z = z)
\end{aligned} \tag{13}$$

The last line in the above equation is just the definition of loss given default conditional on  $z$ . So

$$\begin{aligned}
P(L_i \leq l | \text{default}, Z = z) \\
&= 1 - P_i^{-1}(z) \cdot \left(1 - N\left(\frac{-N^{-1}(p_i(1 - F_D(l))) + \sqrt{\rho}z}{\sqrt{1-\rho}}\right)\right) \\
&= 1 - P_i^{-1}(z) \cdot N\left(\frac{N^{-1}(p_i(1 - F_D(l))) - \sqrt{\rho}z}{\sqrt{1-\rho}}\right)
\end{aligned} \tag{14}$$

The cumulative recovery distribution is

$$\begin{aligned}
P(R_i \leq r | \text{default}, Z = z) &= P(L_i \geq 1 - r | \text{default}, Z = z) \\
&= 1 - P(L_i \leq 1 - r | \text{default}, Z = z) + P(L_i = 1 - r | \text{default}, Z = z) \\
&= P_i^{-1}(z) \cdot N\left(\frac{N^{-1}(p_i(1 - F_D(1 - r))) - \sqrt{\rho}z}{\sqrt{1 - \rho}}\right) + P(L_i = 1 - r | \text{default}, Z = z) \\
&= P_i^{-1}(z) \cdot N\left(\frac{N^{-1}(p_i \cdot (F_R(r) - P(R_i = r | \text{default}))) - \sqrt{\rho}z}{\sqrt{1 - \rho}}\right) + P(R_i = r | \text{default}, Z = z)
\end{aligned} \tag{15}$$

If the marginal loss distribution is continuous at  $1 - r$ , then

$$P(L_i = 1 - r | \text{default}, Z = z) = 0 \tag{16}$$

and, comparing with equation (4),

$$F_R(r) = 1 - F_D(1 - r) \tag{17}$$

and

$$P(R_i \leq r | \text{default}, Z = z) = P_i^{-1}(z) \cdot N\left(\frac{N^{-1}(p_i F_R(r)) - \sqrt{\rho}z}{\sqrt{1 - \rho}}\right) \tag{18}$$

This is the continuous generalization of Krekel's stochastic recovery model.

If the marginal recovery distribution is not continuous at  $r$ , it will be right continuous at  $r$ . So we should still have

$$P(R_i \leq r | \text{default}, Z = z) = P_i^{-1}(z) \cdot N\left(\frac{N^{-1}(p_i(F_R(r))) - \sqrt{\rho}z}{\sqrt{1 - \rho}}\right) \tag{19}$$

and

$$\begin{aligned}
&P(R_i = r | \text{default}, Z = z) \\
&= P_i^{-1}(z) \cdot \left[ N\left(\frac{N^{-1}(p_i F_R(r)) - \sqrt{\rho}z}{\sqrt{1 - \rho}}\right) \right. \\
&\quad \left. - N\left(\frac{N^{-1}(p_i \cdot (F_R(r) - P(R_i = r | \text{default}))) - \sqrt{\rho}z}{\sqrt{1 - \rho}}\right) \right]
\end{aligned} \tag{20}$$

which is exactly Krekel's result. Therefore the continuous case formula (18) is also valid in the discrete case.



A key feature of this model is that the marginal distribution is preserved by construction. Therefore the calibration to single name CDS is done with the marginal distribution. This is generally not true with other models, as will be shown in later sections.

It should be noted that the Ech-Chatbi's model (Ech-Chatbi, 2008) can be viewed as a special case of the Krekel model, except that the distribution is motivated by a multiple loss process model instead of an arbitrary specification. The Ech-Chatbi's model also does not calibrate to single name CDS expected loss at all time points. It only calibrates to the five year expected loss to determine its parameter. Then the model expected recovery rate will be inversely related to the default rate and decreases with time. This is the reason why the Ech-Chatbi's model has wider calibration range for base correlation. One problem with Ech-Chatbi's model is that, if the recovery distribution when  $z$  is fixed is generated by a similar Poisson multiple loss process, it may not produce the original marginal distribution used in calibration to CDS. Instead, the distribution for fixed  $z$  should follow the same scheme as that of the Krekel model discussed above to make sure that marginal distribution is conserved.

## 4 The Amraoui-Hitier model

The Amraoui-Hitier model assumes the stochastic recovery rate depends only on the systematic factor  $Z$ . The general formulation is

$$L = F_D^{-1}(N(-Z)) \quad (21)$$

This is conditional on default. The unconditional loss function should not be used here since then the default function may not dominate the loss function, or loss could be positive even when default has not happened.

The cumulative distribution of loss for a fixed  $z$  is

$$P(L_i \leq l | Z = z) = 1 - P_i(z) + P_i(z) \cdot H(l - F_D^{-1}(N(-z))) \quad (22)$$

where  $H$  is the Heaviside function. Integrating over  $z$ , we have the marginal distribution for the unconditional loss as

$$P(L_i \leq l) = 1 - p_i + p_i \cdot N_2(N^{-1}(p_i), -N^{-1}(F_D(l)); \sqrt{\rho}) \quad (23)$$

where  $N_2(x, y; \rho)$  is the bivariate normal distribution function. Then the marginal distribution of loss given default is

$$P(L_i \leq l | default) = 1 - p_i^{-1} \cdot N_2(N^{-1}(p_i), -N^{-1}(F_D(l)); \sqrt{\rho}) \quad (24)$$

And the marginal distribution for recovery is

$$\begin{aligned} P(R_i \leq r | default) &= p_i^{-1} \cdot N_2(N^{-1}(p_i), -N^{-1}(F_D(1-r)); \sqrt{\rho}) \\ &= p_i^{-1} \cdot N_2(N^{-1}(p_i), N^{-1}(F_R(r)); \sqrt{\rho}) \end{aligned} \quad (25)$$

Note that the marginal distribution of recovery rate will be different from  $F_R$ .

For a fixed  $z$ , the recovery rate will also be fixed. One interpretation is to view this fixed recovery rate as the expected recovery rate under fixed  $z$  of a Krekel model. Then the Krekel model underlies the original Amraoui-Hitier model has the marginal distribution as follows

$$R = \begin{cases} \tilde{R} & \text{with probability } \frac{1-R_0}{1-\tilde{R}} \\ 1 & \text{with probability } \frac{R_0-\tilde{R}}{1-\tilde{R}} \end{cases} \quad (26)$$

where  $R_0$  is the standard 40% recovery rate used in single name calibration and  $\tilde{R}$  is the recovery mark down in the Amraoui-Hitier model. This distribution has the expected recovery rate of  $R_0 = 40\%$ . It is easy to write down the loss distribution function

$$F_L(l) = 1 - p_i + p_i \cdot \left( \frac{R_0 - \tilde{R}}{1 - \tilde{R}} + \frac{1 - R_0}{1 - \tilde{R}} H(l - 1 + \tilde{R}) \right) \quad (27)$$

and

$$F_D(l) = \frac{R_0 - \tilde{R}}{1 - \tilde{R}} + \frac{1 - R_0}{1 - \tilde{R}} \cdot H(l - 1 + \tilde{R}) \quad (28)$$

where  $H$  is the Heaviside function. Loss only happens at  $1 - \tilde{R}$  with probability  $p_i \cdot \frac{1 - R_0}{1 - \tilde{R}}$ . So the expect loss with fixed  $z$  is

$$E(L_i|Z = z) = (1 - \tilde{R}) \cdot N \left( \frac{N^{-1}(p_i \frac{1 - R_0}{1 - \tilde{R}}) - \sqrt{\rho}z}{\sqrt{1 - \rho}} \right) \quad (29)$$

or equivalently

$$1 - E(R_i|default, Z = z) = (1 - \tilde{R}) \cdot P_i^{-1}(z) \cdot N \left( \frac{N^{-1}(p_i \frac{1 - R_0}{1 - \tilde{R}}) - \sqrt{\rho}z}{\sqrt{1 - \rho}} \right) \quad (30)$$

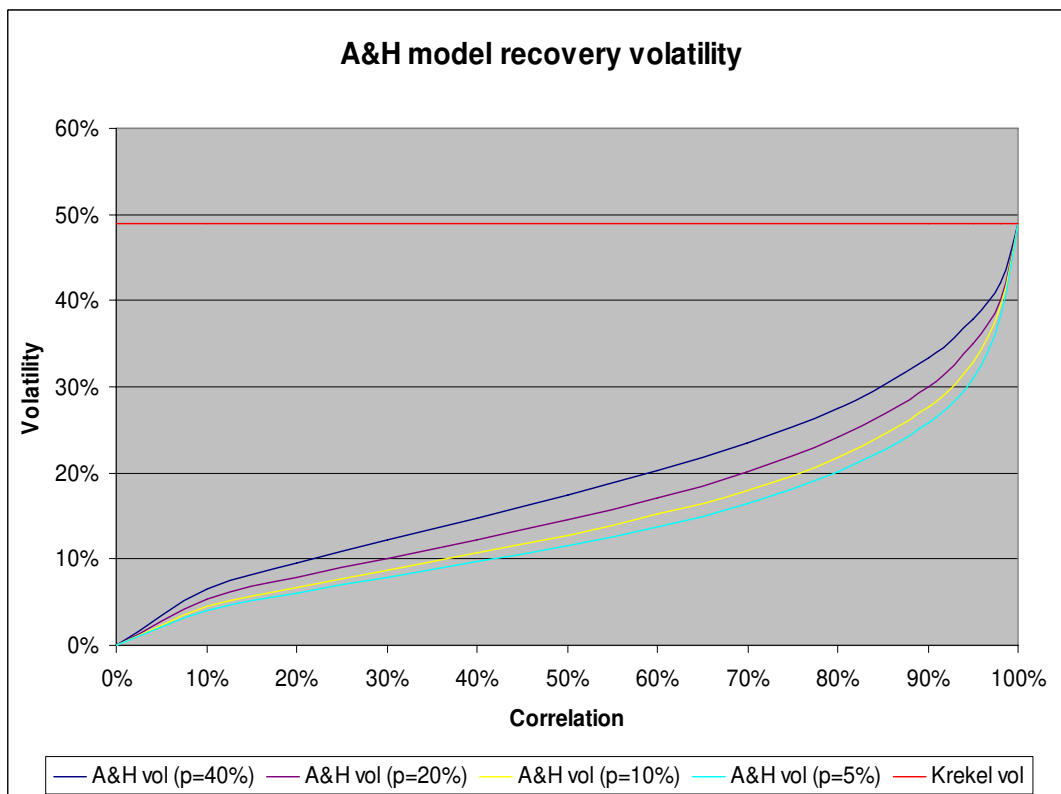
which is the same as the recovery rate specified in the original Amraoui-Hitier model.

For fixed  $z$ , the distribution function of loss for the Amraoui-Hitier model is

$$P(L_i \leq l | Z = z) = 1 - P_i(z) + P_i(z) \cdot H \left( l - (1 - \tilde{R}) \cdot P_i^{-1}(z) \cdot N \left( \frac{N^{-1}(p_i \frac{1 - R_0}{1 - \tilde{R}}) - \sqrt{\rho} z}{\sqrt{1 - \rho}} \right) \right) \quad (31)$$

The marginal distribution of loss is obtained by integrating out  $z$  dependence, which can only be done numerically. Note that the marginal loss distribution is no longer the same as the discrete distribution. However, as long as the underlying Krekel model is calibrated to  $R_0$ , the Amraoui-Hitier model will also be calibrated to the same expected recovery.

When  $\tilde{R} = 0$ , the marginal distribution of the underlying Krekel model has the highest variance if the expected recovery has to be fixed at  $R_0$ . This will give the maximum calibration range for the Amraoui-Hitier model. It is known that the variance of the Amraoui-Hitier recovery model depends on the correlation parameter  $\rho$ . Only when correlation is high will the variance be close to the Krekel model underlying it. That the variance of recovery rate increases with the correlation of corporate defaults needs further empirical evidence. The following chart shows the Amraoui-Hitier (A&H) model recovery volatility vs default correlation with  $\tilde{R} = 0$ . The line on top is the underlying Krekel model with maximum recovery volatility, which is fixed at 49% for 40% expected recovery. When  $\rho$  changes from 0% to 100%, the recovery distribution changes from the fixed rate of 40% to the Krekel model with maximum variance.



One way to explain the variance difference between the Amraoui-Hitier model and its underlying Krekel model is to view the variance of the underlying Krekel model as the sum of expected conditional variance and the variance of the conditional expectation conditional on  $Z$  :

$$\text{Var}(R) = E(\text{Var}(R|Z)) + \text{Var}(E(R|Z)) \quad (32)$$

The second term is just the variance of the Amraoui-Hitier model, while the first term is varying from zero to maximum when the recovery dependence on  $Z$  changes from perfect correlation to independence.

## 5. The Frye and Andersen-Sidenius model

Andersen and Sidenius (2004) proposed a different stochastic recovery model which depends on stochastic factors other than those driving the default. Frye (2000) has proposed this kind of model before for Basel capital calculation but his specification of a random recovery following normal distribution has the technical difficulty of loss over 100%.

Assume  $V_i = \sqrt{\rho_1}Z + \sqrt{1-\rho_1}\varepsilon_i$  drives the default of an obligor and  $W_i = \sqrt{\rho_2}Z + \sqrt{1-\rho_2}\xi_i$  drives the loss, where  $Z$ ,  $\varepsilon_i$ ,  $\xi_i$  are independent normal random variables. The default probability still follows

$$P_i(z) = P(V_i \leq v_i | Z = z) = N\left(\frac{v_i - \sqrt{\rho_1}z}{\sqrt{1-\rho_1}}\right) \quad (33)$$

where  $v_i = N^{-1}(p_i)$  is the default threshold. Dependence of loss is specified as  $L_i = F_D^{-1}(N(-W_i))$ , where we have used loss given default function instead of unconditional loss to avoid potential loss without default. The cumulative loss distribution becomes

$$\begin{aligned} P(L_i = F_D^{-1}(N(-W_i)) \leq l | \text{default}, Z = z) \\ = N\left(\frac{N^{-1}(F_D(l)) + \sqrt{\rho_2}z}{\sqrt{1-\rho_2}}\right) \end{aligned} \quad (34)$$

and

$$P(L_i \leq l | Z = z) = 1 - P_i(z) + P_i(z) \cdot P(L_i \leq l | \text{default}, Z = z) \quad (35)$$

It is easy to show that the recovery upon default distribution is as follows

$$P(R_i \leq r | \text{default}, Z = z) = N\left(\frac{N^{-1}(F_R(r)) - \sqrt{\rho_2}z}{\sqrt{1-\rho_2}}\right) \quad (36)$$

However, after integration over  $z$ , the marginal loss given default distribution will not be  $F_D$ , but is instead

$$\begin{aligned}
F_D^M(l) &= p_i^{-1} \cdot \int_{-\infty}^{+\infty} P_i(z) \cdot P(L \leq l | \text{default}, Z = z) \cdot f_N(z) dz \\
&= p_i^{-1} \cdot N_2(N^{-1}(F_D(l)), N^{-1}(p_i); -\sqrt{\rho_1 \rho_2}) \\
&= 1 - p_i^{-1} \cdot N_2(-N^{-1}(F_D(l)), N^{-1}(p_i); \sqrt{\rho_1 \rho_2})
\end{aligned} \tag{37}$$

where  $N_2(x, y; \rho)$  is the bivariate normal distribution function and  $f_N$  is the density function of the standard normal distribution. See the Appendix for the calculation of the Gaussian integral. The marginal recovery rate distribution is

$$F_R^M(r) = p_i^{-1} \cdot N_2(N^{-1}(F_R(r)), N^{-1}(p_i); \sqrt{\rho_1 \rho_2}) \tag{38}$$

where  $F_R(r) = 1 - F_D(1 - r)$ . Only when the correlation term  $\sqrt{\rho_1 \rho_2}$  is zero will  $F_D^M$  be the same as  $F_D$  and  $F_R^M$  be the same as  $F_R$ . In general, we have  $F_R^M \geq F_R$  so that the marginal recovery distribution  $F_R^M$  is stochastically dominated by the distribution  $F_R$ . This means the marginal expected recovery will be smaller than that implied by the distribution  $F_R$ . Besides, the marginal recovery distribution function decreases when default probability increases so that the marginal expected recovery increases with the default probability.

The marginal recovery rate distribution will now depend on correlation  $\sqrt{\rho_1 \rho_2}$  and default probability  $p_i$ , which makes single name calibration and index tranche calibration more complicated. The added idiosyncratic factor makes the model more flexible, but it does not necessarily increase the correlation calibration range because the correlation between default and loss is less tight. In a sense, the original Amraoui and Hitier model does have the optimal calibration range if the marginal expected recovery is fixed for all time.

In the case  $\rho_2 = 1$ , the model is reduced to that discussed in the previous section, with the systematic factor being the only driver for loss.

## 6 The Pykhtin model

The Pykhtin model extends the above models by assuming a more general correlation form. The loss driver now not only depends on the systematic factors, but also depends on the idiosyncratic factors that drive the default. The loss driver has the form of

$Y_i = \sqrt{\rho_2} Z + \sqrt{1 - \rho_2} (\sqrt{\rho_3} \varepsilon_i + \sqrt{1 - \rho_3} \xi_i)$ , see Witzany (2009). Originally Pykhtin used a lognormal distribution for recovery, which may lead to recovery rate higher than 100%. Besides, the original model assumes potential loss without default, which is excluded here by assuming  $Y_i$  drives only loss given default.

Default is still driven by  $V_i = \sqrt{\rho_1}Z + \sqrt{1-\rho_1}\varepsilon_i$ , such that

$$\begin{aligned}
P_i(z) &= P(V_i \leq v_i | Z = z) \\
&= P\left(\varepsilon_i \leq \frac{N^{-1}(p_i) - \sqrt{\rho_1}z}{\sqrt{1-\rho_1}} | Z = z\right) \\
&= N\left(\frac{N^{-1}(p_i) - \sqrt{\rho_1}z}{\sqrt{1-\rho_1}}\right)
\end{aligned} \tag{39}$$

The cumulative loss given default distribution is

$$\begin{aligned}
P(L_i = F_D^{-1}(N(-Y_i)) \leq l | \text{default}, Z = z) \\
&= P\left(\sqrt{\rho_3}\varepsilon_i + \sqrt{1-\rho_3}\xi_i \geq -\frac{N^{-1}(F_D(l)) + \sqrt{\rho_2}z}{\sqrt{1-\rho_2}} | \text{default}, Z = z\right) \\
&= 1 - P_i^{-1}(z) \cdot N_2\left(\frac{-N^{-1}(F_D(l)) - \sqrt{\rho_2}z}{\sqrt{1-\rho_2}}, \frac{N^{-1}(p_i) - \sqrt{\rho_1}z}{\sqrt{1-\rho_1}}; \sqrt{\rho_3}\right)
\end{aligned} \tag{40}$$

where  $N_2(x, y; \rho)$  is the bivariate normal cumulative distribution function with correlation  $\rho$ .

So

$$\begin{aligned}
P(L_i \leq l | Z = z) &= 1 - P_i(z) + P_i(z) - \\
&\quad - N_2\left(\frac{-N^{-1}(F_D(l)) - \sqrt{\rho_2}z}{\sqrt{1-\rho_2}}, \frac{N^{-1}(p_i) - \sqrt{\rho_1}z}{\sqrt{1-\rho_1}}; \sqrt{\rho_3}\right)
\end{aligned} \tag{41}$$

and

$$P(R_i \leq r | \text{default}, Z = z) = P_i^{-1}(z) \cdot N_2\left(\frac{N^{-1}(F_R(r)) - \sqrt{\rho_2}z}{\sqrt{1-\rho_2}}, \frac{N^{-1}(p_i) - \sqrt{\rho_1}z}{\sqrt{1-\rho_1}}; \sqrt{\rho_3}\right) \tag{42}$$

Then the marginal distribution of loss and recovery can be obtained by integration over  $z$  using the formula in the Appendix

$$F_D^M(l) = 1 - p_i^{-1} \cdot N_2(-N^{-1}(F_D(l)), N^{-1}(p_i); \sqrt{\rho_1\rho_2} + \sqrt{(1-\rho_1)(1-\rho_2)\rho_3}) \tag{43}$$

and

$$F_R^M(r) = p_i^{-1} \cdot N_2(N^{-1}(F_R(r)), N^{-1}(p_i); \sqrt{\rho_1\rho_2} + \sqrt{(1-\rho_1)(1-\rho_2)\rho_3}) \tag{44}$$

These have the same form as equations (37), (38) in the previous section except for different correlation. Note that again the marginal distribution functions  $F_D^M$  and  $F_R^M$  will be different from the distribution functions  $F_D$  and  $F_R$ . The calibration to single name CDS and index tranches will be tricky since both may depend on correlations  $\rho_1$ ,  $\rho_2$  and  $\rho_3$ . Besides, with fixed  $z$ , default and loss given default are still correlated through  $\varepsilon_i$ . Without conditional independence, the calculation will be more involved and less efficient.

In the case  $\rho_3 = 0$ , the Pykhtin model reduces to the Frye model discussed in the previous section. When  $\rho_3 = 1$ , if  $\rho_1 \neq \rho_2$ , then there will be potential loss without default and the model is not consistent with our requirement ; if  $\rho_1 = \rho_2$ , then the model reduces to the Tasche/Krekel model discussed in Section 3.

## 7 Negative forward expected recovery for fixed $z$

It is known that negative forward recovery rate conditional on  $Z = z$  could appear in the Krekel model and the Amraoui-Hitier model when  $z$  is sufficiently negative<sup>2</sup>.

As an example, we will have a look at the Krekel model. The expected loss for fixed  $z$  is

$$\begin{aligned}
E(L|Z = z) &= \int_0^1 l \cdot dP(L \leq l|Z = z) \\
&= 1 - \int_0^1 P(L \leq l|Z = z) \cdot dl \\
&= P_i(z) - P_i(z) \cdot \int_0^1 P(L \leq l|default, Z = z) \cdot dl \\
&= \int_0^1 N\left(\frac{N^{-1}(p_i(1 - F_D(l))) - \sqrt{\rho}z}{\sqrt{1 - \rho}}\right) \cdot dl
\end{aligned} \tag{45}$$

where equation (14) is used in the last step. The default probability is

$$P_i(z) = N\left(\frac{N^{-1}(p_i) - \sqrt{\rho}z}{\sqrt{1 - \rho}}\right) \tag{46}$$

The time dependence of the expected loss and the default probability is all through  $p_i$  for a static model. The instantaneous forward expected loss is defined as

---

<sup>2</sup> Thanks to Paul Bradshaw for pointing this out first.

$$\begin{aligned}
E_i(L|Z = z) &= \frac{dE(L|Z = z)/dt}{dP_i(z)/dt} \\
&= \frac{d_{p_i} E(L|Z = z)}{d_{p_i} P_i(z)} \\
&= \int_0^1 \frac{d_{p_i} N\left(\frac{N^{-1}(p_i(1 - F_D(l))) - \sqrt{\rho}z}{\sqrt{1 - \rho}}\right)}{d_{p_i} N\left(\frac{N^{-1}(p_i) - \sqrt{\rho}z}{\sqrt{1 - \rho}}\right)} \cdot dl \\
&= \int_0^1 \frac{f_N\left(\frac{N^{-1}(p_i(1 - F_D(l))) - \sqrt{\rho}z}{\sqrt{1 - \rho}}\right) / f_N(N^{-1}(p_i(1 - F_D(l))))}{f_N\left(\frac{N^{-1}(p_i) - \sqrt{\rho}z}{\sqrt{1 - \rho}}\right) / f_N(N^{-1}(p_i))} \cdot (1 - F_D(l)) \cdot dl \\
&= \int_0^1 \exp\left(\frac{N^{-1}(p_i)^2 - N^{-1}(p_i(1 - F_D(l)))^2 + 2\sqrt{\rho}z(N^{-1}(p_i(1 - F_D(l))) - N^{-1}(p_i))}{2(1 - \rho)}\right. \\
&\quad \left. + \frac{N^{-1}(p_i)^2 - N^{-1}(p_i(1 - F_D(l)))^2}{2}\right) \cdot (1 - F_D(l)) \cdot dl
\end{aligned} \tag{47}$$

where  $f_N$  is the density function of the standard normal distribution. It is obvious that, when  $z \rightarrow -\infty$ , the exponential part of the integration will go to  $+\infty$ , which makes the instantaneous forward expected loss to be unlimited. Thus, the instantaneous forward expected recovery rate will also be unlimited negative number. Note that, without the exponential term, the formula reduces to the marginal expected loss given default,

$$E(L|default) = \int_0^1 (1 - F_D(l)) \cdot dl \tag{48}$$

The Amraoui-Hitier model has the same expected loss as the underlying Krekel model. So it has the same problem of negative forward recovery conditional on  $z$ .

For the Frye and Andersen-Sidenius model, the loss given default distribution conditional on  $z$  does not depend on  $p_i$ , which seems to be able to avoid the negative forward recovery problem. However, if the marginal distribution were to be calibrated to the single name market with fixed recovery, there would be implicit  $p_i$  dependency in the distributions  $F_D$  and  $F_R$ . Since the marginal distribution  $F_R^M$  has higher expected value with increasing default probability  $p_i$ , the distribution  $F_R$  used for recovery correlation should increase with  $p_i$  to generate a lower expected value to compensate the marginally higher expected recovery. This way, the expected loss under fixed  $z$  will increase due both to increased default probability and increased expected loss given default. With a



highly negative systematic factor, the increase in default probability could be much smaller comparing with the increase of loss given default and results in negative forward recovery. The same thing can happen with the Pykhtin model.

## 8 Conclusion

We have defined a general framework to induce correlation between default and stochastic recovery and also extended several models recently proposed to more general forms. It is known that the current static Gaussian Copula framework will introduce arbitrage and inconsistency. We have shown that all correlated stochastic recovery models in the current static framework may have negative instantaneous forward recovery rate when the systematic factor becomes very negative. Although the problems are more of theoretical nature, it is worth building more consistent models, including, for example, dynamic models. For a recent discussion of a dynamic model framework for CDOs, see for example Li (2009). One specific assumption that needs careful look is the fixed recovery rate at all future times. Some people have suggested to relax it, but others have argued the tranche market should be consistent with the index and single name market for hedging. In a sense, this will depend on how the market on recovery risk evolves.

Although our discussion has been focused on Gaussian copula, it is obvious that the framework is in a more general setting. We noticed that a recent work has extended stochastic recovery in a nested Archimedean copula framework (Höcht and Zagst, 2009). We emphasize that the inconsistency is still inherent in all static models although different copula could introduce more tail risk. We also want to emphasize the technical fact that the marginal recovery distribution may be different from the distribution used to add correlation to stochastic recovery. The marginal distribution might depend on correlation parameters, which makes calibration more complicated.

## Appendix A useful Gaussian integral

$$\int_{-\infty}^{+\infty} N_2(az + b, cz + d; \rho) \cdot f_N(z) dz = N_2\left(\frac{b}{\sqrt{1+a^2}}, \frac{d}{\sqrt{1+c^2}}; \frac{ac + \rho}{\sqrt{(1+a^2)(1+c^2)}}\right)$$

**Proof:** Let  $z_1 = -az + \sqrt{\rho}\varepsilon + \sqrt{1-\rho}\varepsilon_1$  and  $z_2 = -cz + \sqrt{\rho}\varepsilon + \sqrt{1-\rho}\varepsilon_2$  where  $z, \varepsilon, \varepsilon_1, \varepsilon_2$  are independent standard normal random variables. Then

$$\begin{aligned} P(z_1 \leq b, z_2 \leq d) &= E(P(z_1 \leq b, z_2 \leq d | z)) \\ &= E(P(\sqrt{\rho}\varepsilon + \sqrt{1-\rho}\varepsilon_1 \leq az + b, \sqrt{\rho}\varepsilon + \sqrt{1-\rho}\varepsilon_2 \leq cz + d | z)) \\ &= \int_{-\infty}^{+\infty} N_2(az + b, cz + d; \rho) \cdot f_N(z) dz \end{aligned}$$

Meanwhile, the correlation between  $z_1$  and  $z_2$  is  $\frac{ac + \rho}{\sqrt{(1+a^2)(1+c^2)}}$ . So we have

$$P(z_1 \leq b, z_2 \leq d) = N_2 \left( \frac{b}{\sqrt{1+a^2}}, \frac{d}{\sqrt{1+c^2}}; \frac{ac + \rho}{\sqrt{(1+a^2)(1+c^2)}} \right).$$

In the special case  $\rho = 0$ , the integral reduces to

$$\int_{-\infty}^{+\infty} N(az + b) \cdot N(cz + d) \cdot f_N(z) dz = N_2 \left( \frac{b}{\sqrt{1+a^2}}, \frac{d}{\sqrt{1+c^2}}; \frac{ac}{\sqrt{(1+a^2)(1+c^2)}} \right).$$

## References

- S. Amraoui, S. Hitier (2008): Optimal Stochastic Recovery for Base Correlation. defaultrisk.com.
- L. Andersen, J. Sidenius (2004): Extensions to the Gaussian Copula: Random Recovery and Random Factor Loadings. *Journal of Credit Risk* **1**(1), pp. 29-70.
- C. Ech-Chatbi (2008): CDS and CDO Pricing with Stochastic Recovery. SSRN working paper.
- J. Frye (2000): Collateral Damage. *RISK* **13**(4), pp. 91-94.
- S. Höcht, R. Zagst (2009): Pricing Distressed CDOs with Stochastic Recovery. defaultrisk.com.
- M. Krekel (2008): Pricing Distressed CDOs with Base Correlation and Stochastic Recovery. defaultrisk.com.
- Y. Li (2009): A Dynamic Correlation Modelling Framework with Consistent Stochastic Recovery. defaultrisk.com.
- M. Pykhtin (2003): Unexpected Recovery Risk. *RISK* **16**(8), pp. 74-78.
- D. Tasche (2004): The Single Risk Factor Approach to Capital Charges in case of Correlated Loss Given Default Rates. Working paper, Deutsche Bundesbank.
- J. Witzany (2009): Loss, Default, and Loss Given Default Modelling. Working paper, IES.