LEHMAN BROTHERS

Fixed Income
Quantitative Credit Research

Quantitative Credit Research

Quarterly

Volume 2005-Q1

CONTACTS

Quantitative Credit Research (U.S.)

Marco Naldi ..... 212-526-1728 ..... mnaldi@lehman.com
Prasun Baheti ..... 212-526-9251 ..... prbaheti@lehman.com
Juha Korpela ..... 212-526-7911 ..... jkorpela@lehman.com
Jin Liu ..... 212-526-3716 ..... jliu4@lehman.com
Roy Mashal ..... 972-(0)3-623-8675 ..... rmashal@lehman.com
Graham Rennison ..... 212-526-7907 ..... grrennis@lehman.com

Quantitative Credit Research (Europe)

Dominic O'Kane ..... 44-(0)20-7102-2628 ..... dokane@lehman.com
Andrei Greenberg ..... 44-(0)20-7102-1285 ..... andgreen@lehman.com
Lutz Schloegl ..... 44-(0)20-7102-2113 ..... luschloe@lehman.com
Saurav Sen ..... 44-(0)20-7102-2940 ..... sasen@lehman.com
Matthew Livesey ..... 44-(0)20-7102-5942 ..... mlivesey@lehman.com

Structured Credit Strategies

Sunita L. Ganapati ..... 415-274-5485 ..... sganapat@lehman.com
Vikas Chelluka ..... 212 526 3364 ..... vchelluk@lehman.com
Lorraine Fan ..... 212 526 1929 ..... lfan@lehman.com
Claude A. Laberge ..... 212-526-5450 ..... claberge@lehman.com
Takahiro Tazaki ..... 81-3-5571-7188 ..... ttazaki@lehman.com
Gaurav Tejwani ..... 212-526-4484 ..... gtejwani@lehman.com

Quantitative Strategies (Europe)

Vasant Naik ..... 44-(0)20-7102-2813 ..... vnaik@lehman.com
Srivaths Balakrishnan ..... 44-(0)20-7102-2180 ..... sbalakri@lehman.com
Albert Desclee ..... 44-(0)20-7102-2474 ..... adesclee@lehman.com
Jeremy Rosten ..... 44-(0)20-7102-1020 ..... jrosten@lehman.com
Minh Trinh ..... 44-(0)20-7102-1484 ..... mtrinh@lehman.com
Mukundan Devarajan ..... 44-(0)20-7102-9033 ..... mudevara@lehman.com

Additional Contacts

Prafulla Nabar ..... Global Head of Quantitative Research ..... 212-526-6108 ..... pnabar@lehman.com
Lev Dynkin ..... Head of Quantitative Portfolio Strategy ..... 212-526-6302 ..... ldynkin@lehman.com
Hua He ..... Head of Fixed Income Research, Asia ..... 81-3-5571-7215 ..... hhe@lehman.com

Relative Value in CDO Tranches: A View through ASTERION

Minh Trinh
Quantitative Strategies
+44 (0) 207 102 1484
mtrinh@lehman.com

Robin Thompson
Asset & Liability Management
+44 (0) 207 102 1647
rthompson@lehman.com

Mukundan Devarajan
Quantitative Strategies
+44 (0) 207 102 9033
mudevara@lehman.com

With the emergence of liquid credit tranche products based on CDX or iTraxx, investors are now able to implement their views on the market and on particular risk factors such as correlation, market risk and idiosyncratic risk. They can use implied correlation measures to assess the relative value of different tranches, but are limited by the complexity of the behaviour of these implied correlations. There is thus a need for a simple relative value model for CDOs that can intuitively guide investors among these factors. In this article, we discuss the relative value between CDO tranches using the "Asterion"[1] model: we show how the model can help investors understand the correlation smile, assess the relative cheapness and richness of tranches, and design correlation trades.

1. INTRODUCTION

With the emergence of liquid credit tranche products based on CDX or iTraxx, investors are now able to express their views on the market and on risk factors such as correlation, market risk and idiosyncratic risk. Examples of correlation trades include the following:

- Leveraged investors sell protection on the equity tranche to have a leveraged exposure to credit, while other investors such as insurance companies sell protection on mezzanine tranches of synthetic CDOs and leave dealers with exposure to correlation risk and axed to buy equity tranche protection.
- Leveraged investors sell protection on the equity tranche and delta-hedge with the index to benefit from the convexity profile of the equity tranche payoff. As spread volatility increases, the investor gains. The trade is also usually carry positive but is exposed to default risk.
- Leveraged investors invest in steepener trades such as selling protection on the 7-year 3-6% mezzanine tranche and buying protection on the 5-year 3-6% mezzanine tranche on a PV01 neutral basis, anticipating some synthetic CDO issuance in 7-year maturity and thus some new protection sellers in that maturity.
- Bank loan book managers hedge their portfolio by buying protection on the 5-year 3-6% mezzanine tranche. With the issuance of CDO squared (CDO of CDO mezzanine tranches) and thus rising demand from mezzanine protection sellers, mezzanine tranches have become expensive relative to other tranches.

The standard valuation model has been the LHP model which approximates the CDO collateral by a large homogenous portfolio of credits with the same default probability and default correlations (cf. Vasicek (1987)). In this article, we present a simple extension of the LHP model – named *Asterion* – and use it to develop a framework to assess the relative value between tranches in a capital structure.

Asterion allows us to model in a simple way the presence of a "correlation smile" observed in the market for tranche products. The "correlation smile" in CDO tranches is akin to the volatility smile observed in options on equities, interest rates and currencies. Just as the volatility implied by out-of-the money options on equities (using the Black-Scholes model) is typically larger than the at-the-money options, the value of default correlation implied by the equity and senior tranches (using the LHP model) is larger than that implied from the

[1] *Asterion is a star that belongs to the Canes Venatici ("hunting dogs") constellation which represents a smile in the sky.*

mezzanine tranche. The reason why *Asterion* can capture the observed correlation smile is that the distribution of portfolio losses in this model has fatter tails compared with the LHP. This in turn results from allowing the possibility of idiosyncratic as well as systematic jumps to default. Fatter tails in the loss distribution lead to an increased value of the equity tranche as their losses are capped at the detachment point. On the other hand, the fat tails in the loss distribution reduce the value of the senior tranche due to the increased probability of losses. The LHP model can rationalize these valuation effects only by an increase in default correlation – which allows *Asterion* to generate a correlation smile.

Asterion is aimed to be a convenient relative value tool to assess investment opportunities in the CDO tranche market. For this reason, it is parsimonious and has few parameters. It does not have the flexibility to price perfectly all tranches at all maturities[2] and it is not meant to. It is a simplified representation of reality that can be used to understand the general shape of the implied correlations[3] structure.

In section 2, we present the *Asterion* model. In section 3, we discuss the impact of the key parameters of the model on the correlation smile that it generates. In section 4, we explain the use of *Asterion* as a relative value tool and present conclusions in section 5. In the Appendix, we present the derivation of the pricing equations of the model.

2. THE ASTERION MODEL

2.1. Basic model set-up

Asterion builds on the LHP model[4] which assumes a large homogenous portfolio in which individual default is triggered by the asset value falling below a default barrier. In LHP, the asset value is driven by two Gaussian factors: a market factor and an idiosyncratic factor. These two factors are weighted by the asset correlation. We extend this model by allowing the possibility of idiosyncratic and systematic jumps-to-default. This is done as follows:

Consider a given time horizon T and a credit portfolio with m issuers. The default time τ_i of the i^{th} issuer is less than T if one of three events occurs:

1. if a continuous random variable representing the asset value of the firm $A(i) = \beta(i)Z + \sqrt{1-\beta(i)^2}Z(i)$ falls below a barrier $C(i)$ where Z and $Z(i), i = 1,\ldots,m$ are i.i.d. standard normal variables; or

2. if there is an idiosyncratic jump to default (modelled as the first jump time of a Poisson process $J(i,t)$ with a constant arrival rate $\lambda(i)$ being less than T); or

3. if there is a systematic jump to default (modelled as the first jump time of a Poisson process $N(t)$ with constant arrival rate λ being less than T). The difference between $J(i,t)$ and $N(t)$ is simply that when $N(t)$ jumps, all firms default simultaneously while the jump of $J(i,t)$ implies the default of the i^{th} firm alone.

2 *If it fitted the market data perfectly there would be no relative value since all tranches would be fairly priced already!*
3 *This is probably why the famous Black-Scholes equation for pricing options still survives at this time and is still widely used in spite of its limited features.*
4 *See O'Kane* et al. *(2003),* Lehman Brothers Guide to Exotic Credit Derivatives *for a description of LHP. The model is due to Vasicek (1987).*

All the probability distributions used in this paper are the ones under the risk-neutral probability measure because we use these distributions in the pricing of tranche products.

If there was no jump to default, the above is simply the LHP model if we assume that the credit portfolio is homogeneous (i.e. if there are no differences across issuers). The correlation in defaults in our model arises from the common factor Z governing the distribution of $A(i)$ as well as the possibility of a common jump to default.

Just like the LHP model, *Asterion* also assumes a homogeneous portfolio. Thus, all β are assumed to be the same, so that β^2 is the correlation between any two assets. Also, the default barrier, and the default probabilities and recovery rates are identical across firms. We denote by N the notional amount of the portfolio and by R , the recovery rate.

In the model described above, we can calculate explicitly the probability of default by the i^{th} issuer before T (denoted $p(i,T)$). This is simply given by one minus the probability that the default time τ_i of the issuer is greater than T which in turn is given by:

$$P(\tau_i \geq T) = e^{-(\lambda(i)+\lambda)T}\left[1-\Phi\left(C(i)\right)\right]$$

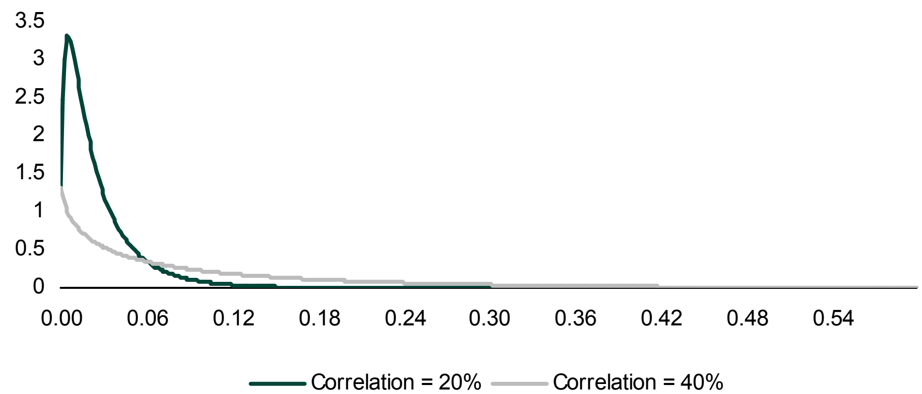
Here $e^{-(\lambda(i)+\lambda)T}$ is the probability of no idiosyncratic or systematic jump to default over the horizon T and $\Phi\left(C(i)\right)$ is the probability of $A(i)$ falling below the barrier $C(i)$ conditional on no idiosyncratic or market jump to default. Using the assumed structure on correlation of default times and the assumption of portfolio homogeneity, it is also straightforward to calculate the probability distributions of portfolio losses which then imply the pricing equations of various tranche products. Detailed derivations are provided in the Appendix.

Figure 1 provides an illustration of the portfolio loss distribution implied by the *Asterion* model. We assume the following values for the jump intensities, the correlation and the single-name default probability: $\lambda = 7bp, \lambda(i) = 1.5bp, \beta = \sqrt{20\%}, p = 45bp$.

The loss distribution has the usual shape: it is very skewed towards the low losses and exhibits fat tails because of the jumps, in particular because of the market jump for which there is a single probability mass at the 1-R level.

With a higher level of correlation, the distribution has a fatter tail, with a higher probability of small and large losses as shown in Figure 1. This is consistent with all the names surviving together or defaulting together when the correlation is high.

Figure 1. Probability density function of losses for different correlations



Source: Lehman Brothers.

With a higher level of market jump risk, the distribution has fatter tails, with a higher probability of small and large losses. The distribution becomes more binary: all names tend to survive or default together. With a higher level of idiosyncratic jump risk, a small loss is less likely and it thus affects the equity tranche (0-3%) more severely. Losses also tend to be concentrated in the mezzanine tranches and do not tend to affect the senior tranche too much. The default barrier is moving lower and the diffusive default probability is lower. As explained in section 3, these features of loss distribution in the model allow it to generate a reasonable implied correlation smile.

3. COMPOUND CORRELATION SMILE IN THE ASTERION FRAMEWORK

For a plain-vanilla option, the option implied volatility is the volatility we have to use in the Black-Scholes formula to be able to value the option. For different strikes there is typically a volatility smile. For a CDO, the analogy carries on with the correlation. The implied compound correlation of a tranche is the correlation we have to use in the LHP formula to be able to value the tranche (same present value).

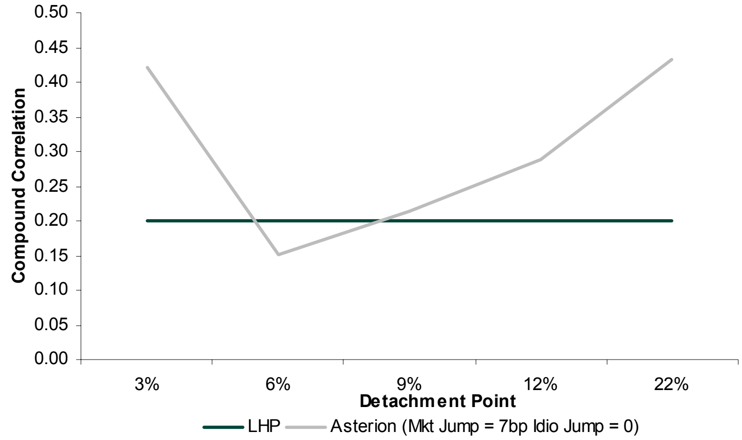
The presence of jumps allows for fatter tails in the loss distribution than the Gaussian set up of LHP. There is a higher probability of large losses which implies that losses have been redistributed from junior tranches to senior tranches relative to the Gaussian case. This increases the expected losses of senior tranches and decreases the expected losses of the equity tranche – thereby increasing the value of the equity tranche and decreasing the value of the senior tranche.

Given the fact that the equity tranche is long (compound) correlation and that the senior tranche is short (compound) correlation[5], the changes in values mentioned above can be captured by LHP only if one increases the compound correlation of the equity and senior tranches vis-à-vis the mezzanine tranche – thereby generating a correlation smile. Figure 2 illustrates the compound correlation smile in the LHP (market and systematic jump risks are zero) and *Asterion* frameworks.

Thus, one can use the *Asterion* model to back out the values of the three key parameters: the correlation, idiosyncratic jump intensity and market jump intensity implied by the correlation smile observed in the market. Thus, Asterion can help in the assessment of relative value.

[5] As is well-recognized, the losses to the equity tranches are a concave function of portfolio losses and hence an increase in the volatility of losses (resulting from increased correlation of defaults) leads to a decrease in expected losses to equity tranche and an increase in its value. Similarly, the losses to the senior tranche is convex in the portfolio losses (in the relevant range) which means that an increase in correlation will increase the expected losses to the senior tranche and reduce its value.

Figure 2. Correlation smile in the Asterion framework



Source: Lehman Brothers.

To understand the utility of *Asterion* as a relative value tool, it is essential to develop an intuition for the impact of the key parameters of *Asterion* on the correlation smile that will be generated by the model. In this section, we discuss the impact of these parameters on the correlation smile.

3.1. Assumptions

We consider an equity tranche (0% to 3%), a junior mezzanine tranche (3% to 6%) and a senior tranche (12% to 22%). The attachment points reflect the current attachment points of the European iTraxx. We assume the following model parameters:

Figure 3. Asterion parameter assumptions

Asterion parameters	Value
P	45bp
λ(i)	1.5bp
λ	7bp
β	sqrt(20%)
R	40%
T	5

Source: Lehman Brothers.

We fix the parameters as shown in Figure 3 above and modify them one at a time. In particular, the single-name default probability is kept constant, while we change the correlation and the jump intensities.

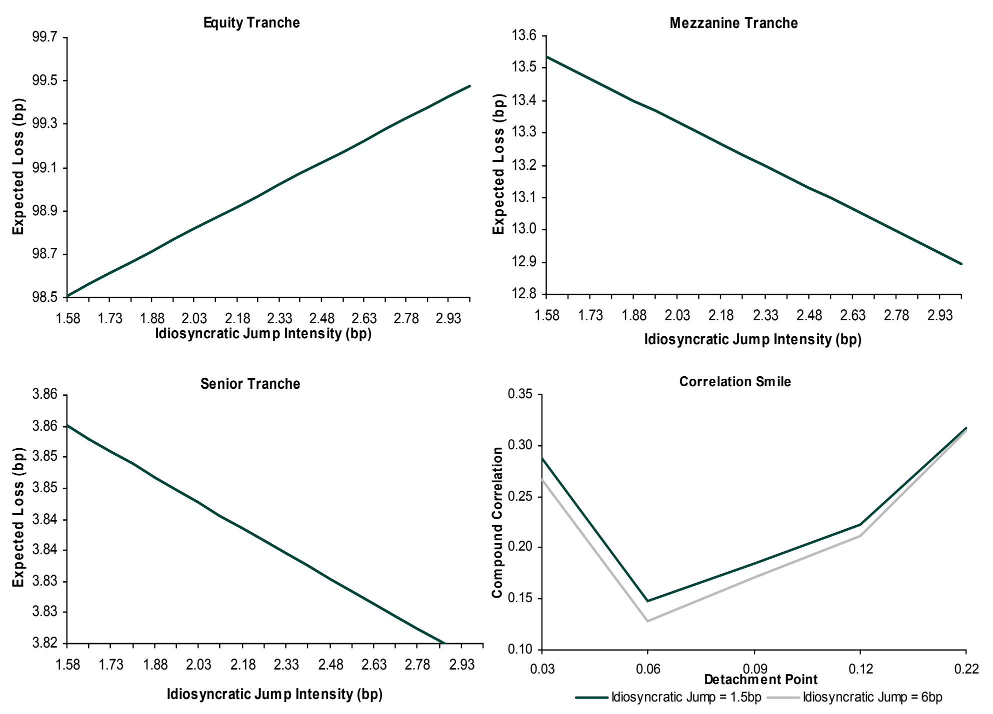
3.2. Idiosyncratic jump-to-default risk

In Figure 4, we plot the expected loss of different tranches and the implied correlation smile for different values of idiosyncratic jump intensity.

The equity tranche holder is short idiosyncratic risk. As idiosyncratic risk increases, expected losses also increase because the equity tranche takes the first loss of any single-name default. Conversely, the senior tranche holder is long idiosyncratic risk. As idiosyncratic risk increases, expected losses decrease because for a given probability of single-name default, the risk of correlated default risk decreases and the default barrier falls to compensate for the higher level of idiosyncratic jump intensity. The mezzanine case is the intermediate case and for the assumed parameter values, it turns out that this tranche is long idiosyncratic risk.

The impact of idiosyncratic jump risk on the correlation smile can therefore be summarised as follows. The idiosyncratic jump risk decreases the value of the equity tranche but increases the value of the senior tranche. Since in the LHP model (used to compute the implied correlation) the equity tranche is long correlation, a lower equity tranche value has to translate into a lower implied correlation. The reasoning is similar for the senior and mezzanine tranches. The correlation smile therefore moves down as shown in Figure 4.

Figure 4. Impact of idiosyncratic jump on expected losses and correlation smile



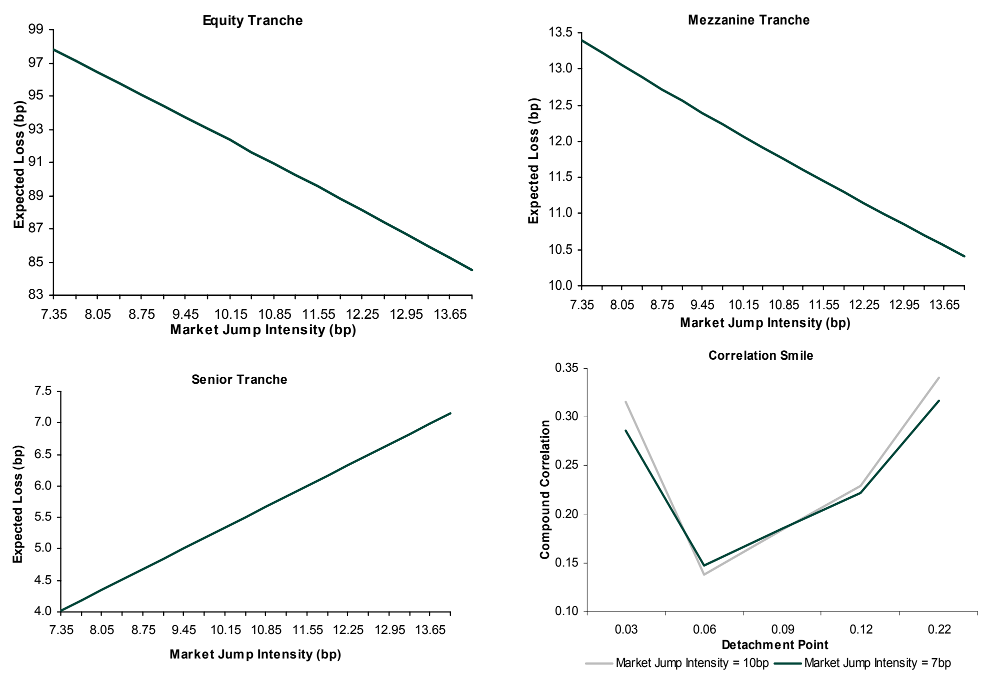
Source: Lehman Brothers.

3.3. Market jump-to-default risk

In Figure 5, we show the effect of changes in systematic jump-to-default risk. As systematic risk increases, expected losses of the equity tranche decrease and those of the senior tranche increase. An increase in systematic risk has the same effect as an increase in volatility of the collateral and therefore increases the value of the equity tranche. The senior tranche holder is short market risk as the senior tranche has limited upside but is exposed to the entire market defaulting.

As explained earlier, the increase in market jump intensity increases the compound correlation of both the equity and senior tranches. Note that the effect of the change in systematic jump risk is exactly the opposite of that of the idiosyncratic risk.

Figure 5. Impact of systematic jump on expected losses and correlation smile

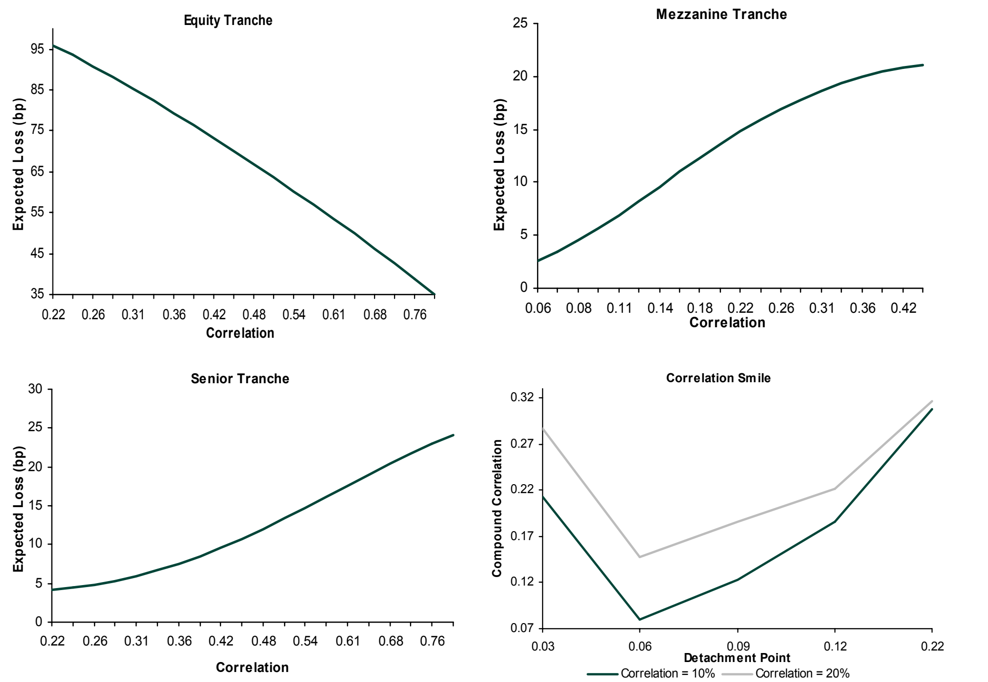


Source: Lehman Brothers.

3.4. Asset correlation

As shown in Figure 6, an increase in the asset correlation has to increase the implied correlation. The equity tranche is long both the flat correlation in *Asterion* and the implied correlation in the LHP model, thus both correlations have to increase together. The same reasoning applies to the mezzanine and the senior tranches. An increase in asset correlation therefore moves the compound correlation smile up.

Figure 6. Impact of correlation on expected losses and correlation smile



Source: Lehman Brothers.

The above results are summarised in Figure 7. These can be used to identify relative value opportunities and take views on the values of the parameters:

Figure 7. Effect of risk on tranche value

Risk	Equity	Junior Mezzanine	Senior
Correlation	Increases	Decreases	Decreases
Idiosyncratic jump risk	Decreases	Increases	Increases
Market jump risk	Increases	Increases	Decreases

Source: Lehman Brothers.

The results on the mezzanine tranche will depend on the particular parameters chosen. The results on the equity and senior tranches are more general.

4. RELATIVE VALUE USING ASTERION

We have shown in the previous sections that the *Asterion* parameters drive the tranche loss distributions, expected losses, present values and hence correlation smiles. The model can be used in two ways:

- Investors with specific views on these parameters can implement those views by trading the different tranches.
- From the correlation smile, an investor can infer the Asterion parameters from the market. Because *Asterion* has three parameters only, typically not every observed price will fit to the model. A discrepancy between the fitted correlation smile and the observed correlation smile could indicate relative value opportunities. This comparison can be done:
  - in cross-section (different tranches, same maturities);
  - in maturity (same tranche, different maturities); or
  - in time (spot tranche vs. forward tranche)

One could force such calibration to fit more liquid tranches more accurately. This would enable investors to get a sense of the relative value between liquid points and illiquid points.

4.1. Views on correlation, market and idiosyncratic jump risks

The model can contribute to structuring trades that express views on the values of the parameters. It enables scenario analyses to evaluate the P&L profiles of these trades by explicitly computing the risks and risk exposures of such trades. In this section, we present two examples of such views and different trades that could be used to implement these views.

Example A: Views on correlation and idiosyncratic risk

Consider the case of an investor who holds the view that correlation is set to rise over the next year and that the idiosyncratic jump intensity stands to decrease over the same horizon. The investor could use the following alternative trades to express this view.

A1. Short Mezzanine tranche, delta hedged by the underlying index; or

A2. Short Mezzanine tranche, delta hedged by Equity Tranche

As mentioned above, the choice of Trade A1 over Trade A2 or vice versa would depend on the P&L profiles of the trades over the maturity and their risk characteristics.

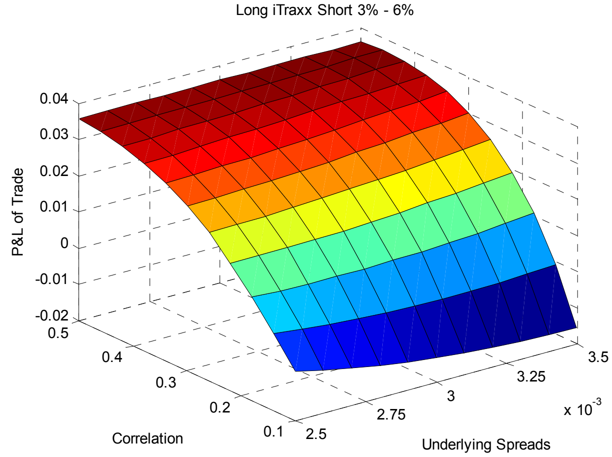
Note that in the above trades investors would need to maintain a delta-neutral position. For simplicity, the P&L profiles we show below assume that the position is delta-neutral only at inception and held to maturity (assumed to be one year).

Trade A1

Figure 8a shows the P&L of Trade A1 for different levels of correlation and underlying spreads assuming zero losses. As expected, we can see that the trade outperforms in high correlation scenarios.

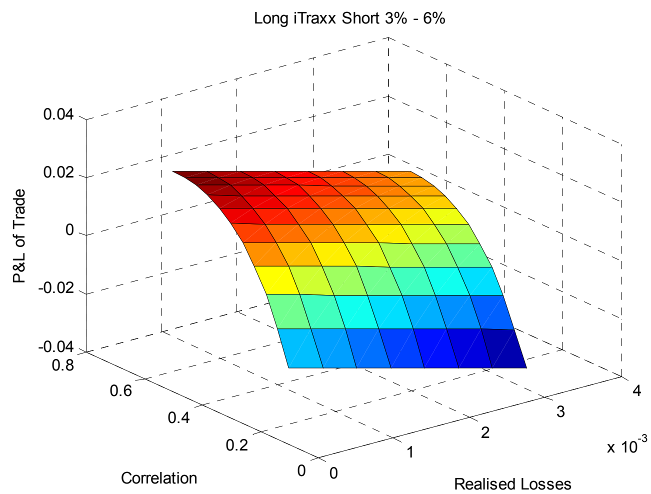
Figure 8b shows the P&L of the trade at different levels of correlation and realised losses of the underlying portfolio over the horizon of the trade. We can see that the trade outperforms in the event of high correlation.

Figure 8a. Long Index Short Mezzanine Trade – correlation, spread vs P&L



Source: Lehman Brothers.

Figure 8b. Long Index Short Mezzanine Trade – correlation, losses vs P&L



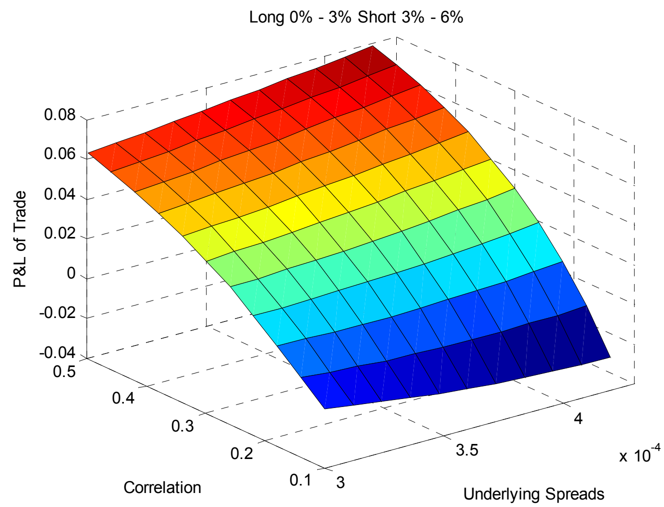
Source: Lehman Brothers.

Trade A2

Figure 9a shows the P&L positions of Trade A2 for different levels of correlation and underlying spread, assuming zero losses over the horizon of the trade. We can see that the trade outperforms in the event of an increase in correlation just as in Trade A1.

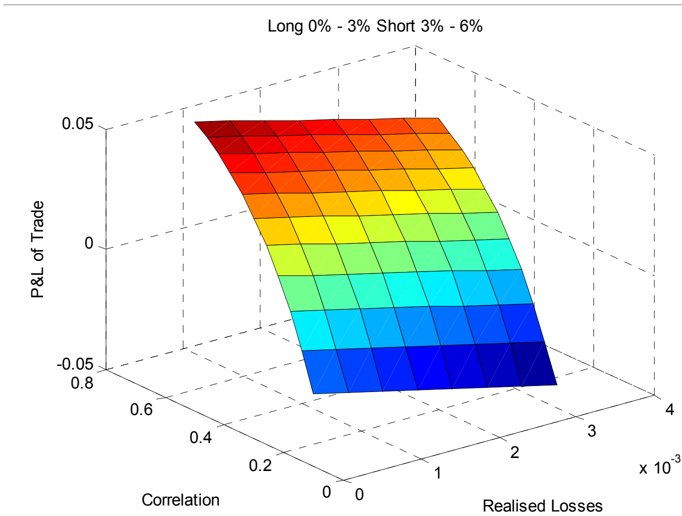
Figure 9b shows the P&L of the trade for different levels of correlation and realised losses on the underlying portfolio. As expected, we can see that the trade outperforms in high correlation low realised loss scenarios.

Figure 9a. Long Equity Short Mezzanine Trade – correlation, spread vs P&L



Source: Lehman Brothers.

Figure 9b. Long Equity Short Mezzanine Trade – correlation, losses vs P&L



Source: Lehman Brothers.

Comparison of the trades

We can see that the outperformance of Trade A2 in high correlation, low spread environments is greater than that of Trade A1. Also, the underperformance is worse in Trade A2 compared with Trade A1. The trades can be compared as shown in Figure 10[6] .

Figure 10. Risk characteristics of trades per $1 notional in mezzanine tranche

Trade	Trade Description	Gamma	λ 01	λi 01	Correl 01[7]
A1	Long Index - Short Mezzanine	0.247	-27.664	-13.632	0.205
A2	Long Equity - Short Mezzanine	0.368	4.0750	-18.904	0.302

Source: Lehman Brothers.

[6] *The values of the Greeks of the trades depend upon the parameters chosen and need not necessarily replicate Greeks observed currently in the market.*

[7] *λ 01, λi 01 and Correl 01 are all defined as change in tranche MTM per 1bp change in the parameter value.*

We can see in Figure 10 that although trades A1 and A2 are based on similar views of correlation and idiosyncratic jump intensity, the trades have different risk characteristics as follows:

a. Trade A1 is short Market jump risk, while Trade A2 is marginally long

b. Trade A2 would be more effective for strong views on the decrease of idiosyncratic jump risk or the increase of correlation

Example B: Views on correlation and market risk

Similarly, the view that the market jump risk and correlation stand to increase over the next year can be expressed using the following trades:

B1. Long Equity Tranche, delta hedged by the underlying index; or

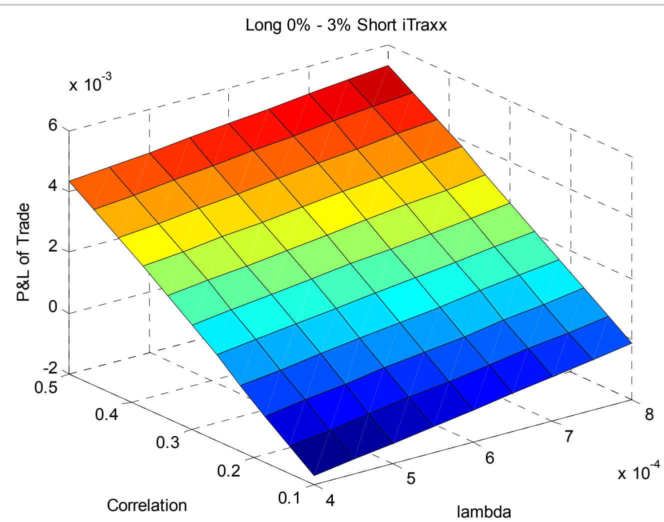
B2. Short Senior Tranche, delta hedged by the underlying index

Trade B1

Figure 11a shows the P&L of the trade for different levels of correlation and idiosyncratic jump intensity, assuming zero losses over the horizon of the trade. We can see that the trade outperforms in high correlation and high market jump intensity scenarios.

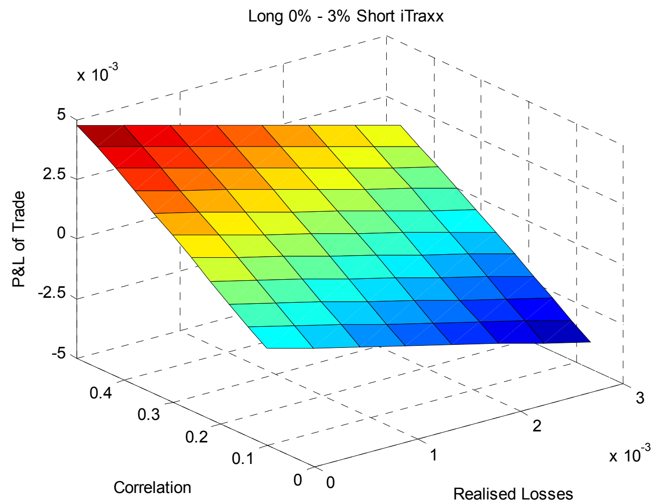
Figure 11b shows the P&L of the trade at different levels of correlation and realised losses on the underlying portfolio over the horizon of the trade. As expected, we see that the trade outperforms in high correlation, low realised loss scenarios.

Figure 11a. Long Equity Short Index Trade – correlation, market jump vs P&L



Source: Lehman Brothers.

Figure 11b. Long Equity Short Index Trade – correlation, realised losses vs P&L

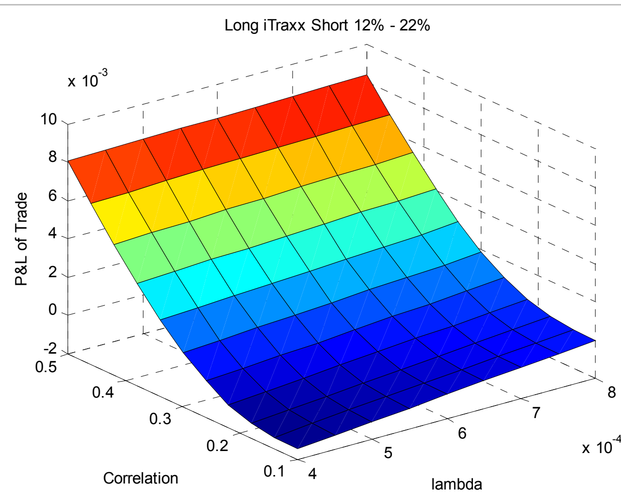


Source: Lehman Brothers.

Trade B2

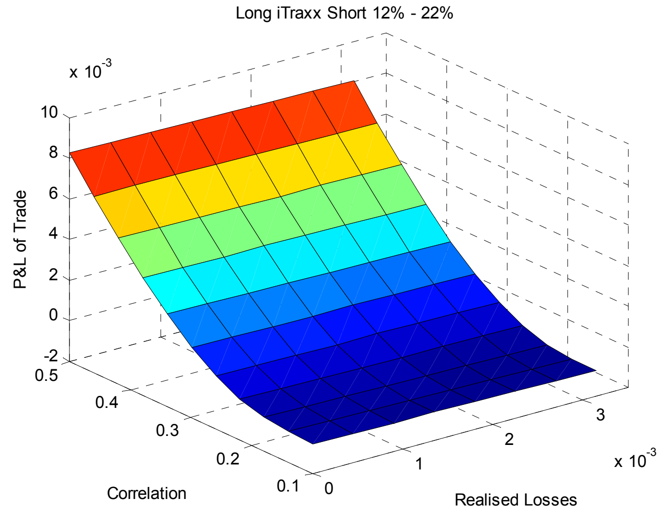
Figure 12a shows the P&L of the trade for different levels of correlation and market jump intensity, assuming zero losses over the horizon of the trade. The performance profile is similar to that of Trade B1.

Figure 12a. Long Index Short Senior Trade – correlation, market jump risk vs P&L



Source: Lehman Brothers.

Figure 12b. Long Index Short Senior Trade – correlation, realised losses vs P&L



Source: Lehman Brothers.

However, the risk characteristics are different as shown in Figure 13:

Figure 13. Risk characteristics of trades per $1 notional in the short leg

Trade	Trade Description	Gamma	λ 01	λi 01	Correl 01
B1	Long Equity - Short Index	0.017	4.467	-0.742	0.014
B2	Long Index - Short Senior	0.006	4.274	-0.134	0.005

Source: Lehman Brothers.

We can see in Figure 13 that although trades B1 and B2 are based on similar views of correlation and market jump intensity, they differ:

a. Trade B1 would be more effective for strong views on the decrease of idiosyncratic jump risk or the increase of correlation

b. Trade B2 would be more effective for weaker views on correlation

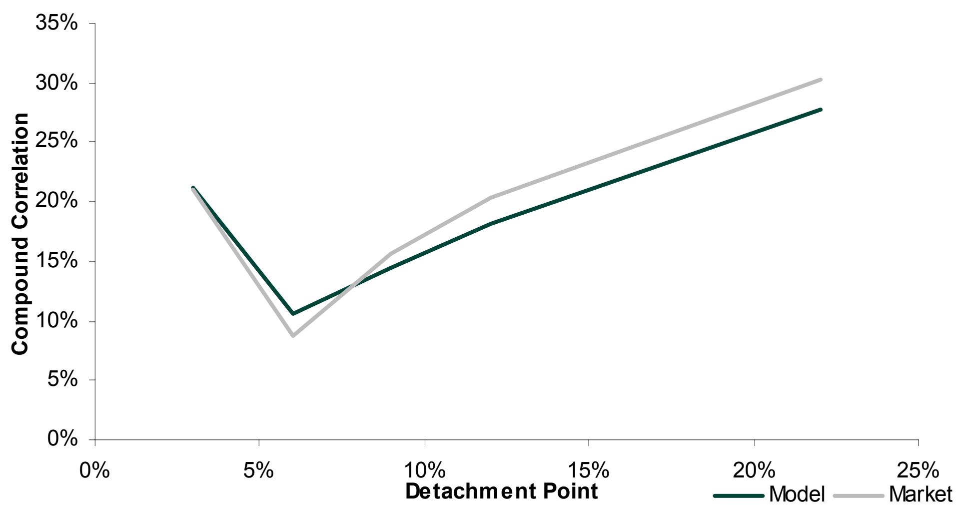
However, these results depend on the values of the parameters chosen and could differ according to the shape of market correlation smile.

The main advantage of the *Asterion* model is that the parameters are readily interpretable and could be inferred from data that are independent of the model. To assess the different risk factors, investors could either assume some dynamics such as mean-reversion or momentum, or infer these parameters from other securities or other markets.

4.2. Relative value views from the correlation smile

In addition to taking views on model parameters, we can calibrate the correlation smile and back up the market jump intensity, the idiosyncratic jump intensity and the implied asset correlation. Because the model has only three parameters, not all observed prices will fit equally well. Tranches whose prices do not fit well might indicate the presence of anomalies relative to those tranches which are well calibrated. Below we present an example of a model calibration to prices of DJ-iTraxx Europe tranches as of 25 February 2005. We consider the iTraxx portfolio with the five standard tranches: 0-3%; 3-6%; 6-9%; 9-12%; 12-22%. The calibration results are shown in Figures 14 and 15. The implied parameters are $\lambda = 6bp; \lambda(i) = 10bp; \beta = \sqrt{18\%}$.

Figure 14. Calibrated vs market compound correlation: 5-year tranches



Source: Lehman Brothers.

Figure 15. Calibration results

5 Year tranche	Market compound correlation	Model compound correlation
0-3%	21.06%	21.17%
3-6%	8.81%	10.66%
6-9%	15.62%	14.46%
9-12%	20.43%	18.25%
12-22%	30.24%	27.72%

Note: Calibration 25 Feb 2005 on DJ iTraxx – Europe.
Source: Lehman Brothers.

Relative value across tranches: tranche trades

From the compound correlation smiles in Figures 14 and 15, the 5-year mezzanine tranche 3-6% seems slightly expensive. This can be explained by the demand for mezzanine tranches in CDO squared deals. Dealers are proposing CDOs on a portfolio of mezzanine tranches causing a demand for these tranches and pulling down the implied correlation.

To exploit this opportunity, an investor could go short a mezzanine tranche, delta hedged by the equity tranche in a trade similar to that shown in Trade A2 in Section 4.1.

Alternatively, the investor could go short the mezzanine tranche and go long the wings. For example, an investor who believes that the mezzanine tranche is rich and has no view on the movement of idiosyncratic or market jump risk over the life of the trade could go short the junior mezzanine tranche, and long the equity and senior mezzanine tranches in such a way that the total exposure of the portfolio to changes in idiosyncratic and market jump risk are zero, while maintaining a negative total *Correl01* as shown in the example in Figure 16.

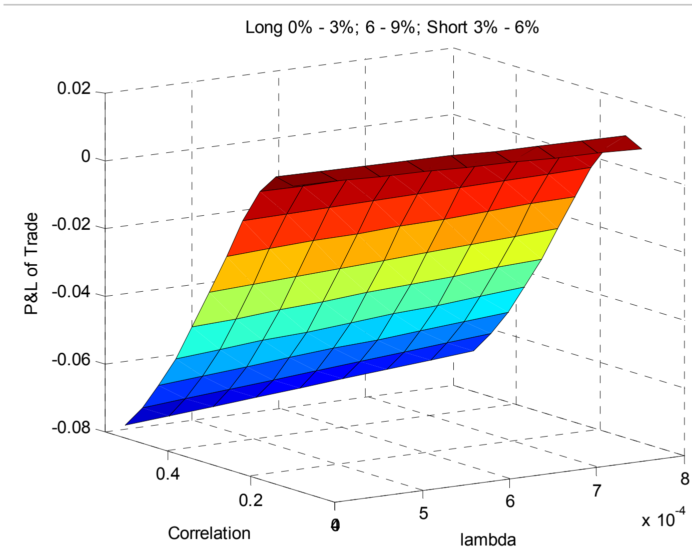
Figure 16. Example correlation trade

Tranche	0-3%	3-6%	6-9%
Weight	0.18	-1	4.73

Source: Lehman Brothers.

Figure 17 shows the P&L profile of the trade against correlation and market jump risk. We can see that the trade outperforms in low correlation scenarios and is relatively unaffected by changes in market jump intensity.

Figure 17. Short Junior Mezzanine; Long Equity and Senior Mezzanine



Source: Lehman Brothers.

5. CONCLUSION

In this article, we have presented a simple extension to the LHP model, named *Asterion.* We have shown how the model can help us understand the correlation smile and the shape of the base correlation curve, and assess the relative cheapness and richness of the tranches. The main model parameters – the asset correlation, the market jump intensity, the idiosyncratic jump intensity and the spread level – drive the value of the tranches and the shape of the compound correlation smile. We have seen how the valuations of the different tranches are influenced by these variables. Because of the presence of jump risks, the equity and senior tranche implied correlations have to be larger than the junior mezzanine tranche correlation to reflect the difference in risk sensitivities.

The model is mostly designed for relative value analysis as it has only three parameters. There are at least two ways of using the model for relative value analysis. First, investors who have specific views on the parameters of the model (say, through their historical behavior or through their economic meaning) can implement those views by trading the different tranches which are priced to imply parameter values different from the investor's views. Second, from the correlation smile, investors can infer the parameters from the market. Any discrepancy between the fitted correlation smile and the observed correlation smile could indicate relative value opportunities. We have also demonstrated how to use the model to analyse the P&L of correlation trades under different scenarios. In future research, we aim to report the empirical performance of our model.

REFERENCES

Greenberg, A., R. Mashal, M. Naldi and L. Schloegl (2004), "Tuning Correlation and Tail Risk to the Market Prices of Liquid Tranches", *Quantitative Credit Research Quarterly*, vol. 2004-Q1, Lehman Brothers.

Isla L., A. Greenberg, and L. Schloegl (2003), *Trading Correlation: A Guide to Assessing Relative Value across CDO Tranches*, Lehman Brothers.

O'Kane, D., and L. Schloegl (2002), "Tail Dependence and Portfolio Risk", *Quantitative Credit Research Quarterly*, vol. 2002-Q3, Lehman Brothers.

O'Kane, D., and L. Schloegl (2003), "An Analytical Portfolio Credit Model with Tail Dependence", *Quantitative Credit Research Quarterly*, vol. 2002-Q4, Lehman Brothers.

O'Kane, D., M. Nadi, S. Ganapati, A. Berd, C. Pedersen, L. Schloegl and R. Mashal (2003), *The Lehman Brothers Guide to Exotic Credit Derivatives*, Risk Waters Group.

Thompson, R., M. Trinh and A. Greenberg (2004), "Portfolio Asset Allocation with CDOs", *Quantitative Credit Research Quarterly*, vol. 2004-Q1, Lehman Brothers.

Trinh, M. (2004), "ORION: A Simple Debt-Equity Model with Unexpected Default", *Quantitative Credit Research Quarterly*, vol. 2004-Q3/4, Lehman Brothers.

Vasicek , O. (1987), *Probability of Loss on Loan Portfolio*, KMV Corporation.

APPENDIX: DERIVATION OF THE ASTERION MODEL

The Asterion model

We can recover the default barrier from the default probability of a single issuer:

$$C(i) = \Phi^{-1}\left(\exp((\lambda(i)+\lambda)T)(p-1+\exp(-(\lambda+\lambda(i))T))\right) \tag{1}$$

Φ^{-1} is the inverse cumulative normal function.

The conditional (on no market default or market jump) default probability of each issuer can be explicitly computed from equations (1):

$$\begin{aligned} &P\left(A(i) \le C(i) \middle| Z, \text{No systematic jump}\right) = \pi(Z) \\ &\pi(Z) = \Phi\left(\frac{C(i)-\beta Z}{\sqrt{1-\beta^2}}\right)\exp(-\lambda(i)T) + [1-\exp(-\lambda(i)T)] \end{aligned} \tag{2}$$

where $\Phi(.)$ is the cumulative normal distribution function. *T* is the time of the cash flow

Similar to the LHP case, as the underlying portfolio becomes large, the fraction of names defaulting is approximately represented by *π(Z)* where the probability *π(Z)* is given by (2). This implies that the expected percentage loss on a homogenous portfolio, conditional on the common market factor, is equal to *π(Z).(1-R)* as well. Therefore, the probability of portfolio losses being less than or equal to a given percentage θ can be computed by inverting the relationship (2):

$$\begin{aligned} &P[L \le \theta \mid \text{No systematic jump}] = P\left[Z \ge \pi^{-1}\left(\frac{\theta}{N(1-R)}\right)\right] = \{1-\Phi(Z_\theta)\} \\ &P[L \le \theta \mid \text{Systematic jump occurs}] = 0 \\ &i.e. P[L \le \theta] = e^{-\lambda T}\{1-\Phi(Z_\theta)\} \end{aligned} \tag{3}$$

We define:

$$Z_\theta = \pi^{-1}\left(\frac{\theta}{N(1-R)}\right) \tag{4}$$

Using the analytical expression for the loss distribution of the collateral given by (3), one can obtain the expected loss of an equity tranche of size θ via integration.

We can write:

$$\begin{aligned} &E[L.1_{\{L<\theta\}}] = N(1-R)\exp(-[\lambda(i)+\lambda]T)\Phi_{2,-\beta(i)}(C(i),-Z_\theta) \\ &+ N(1-R)[1-\exp(-\lambda(i)T)]\Phi(-Z_\theta)\exp(-\lambda T) \end{aligned} \tag{5}$$

Where $\Phi_{2,-\beta(i)}$ is the cumulative bivariate normal distribution function with correlation $-\beta$ and *C(i)*, *β* and Z_θ were defined above.

The expected loss conditional on the loss being smaller than a threshold is composed of two terms. The first term is the expected loss conditional on no jump, the second term is the expected loss conditional on jumps in the idiosyncratic return but not the market return.

The expected loss of an equity tranche of size θ is then:

$$
\begin{aligned}
&E\left[\min(L,\theta)\right] = \theta(1-\exp(-\lambda T)\{1-\Phi(Z_\theta)\}) \\
&+ N(1-R)\exp(-[\lambda(i)+\lambda]T)\Phi_{2,-\beta(i)}(C(i),-Z_\theta) \\
&+ N(1-R)[1-\exp(-\lambda(i)T)]\Phi(-Z_\theta)\exp(-\lambda T)
\end{aligned}
\tag{6}
$$

Using (6), we can then calculate the tranche survival probability of a synthetic tranche with attachment point θ_1 and detachment point θ_2:

$$
Q_{\theta_1,\theta_2}(t) = 1 - \frac{E\left[\min(L,\theta_2)\right] - E\left[\min(L,\theta_1)\right]}{\theta_2-\theta_1} \tag{7}
$$

For each time horizon, equation (7) gives a term structure of tranche survival probabilities. We can then define the premium and protection leg of a tranche:

$$
\text{Protection Leg PV} = (\theta_2-\theta_1)\sum_{i=1}^{K} Z(0,t_i)(Q_{\theta_1,\theta_2}(t_{i-1}) - Q_{\theta_1,\theta_2}(t_i)) \tag{8}
$$

$$
\text{Premium Leg PV} = S_{\theta_1,\theta_2}(\theta_2-\theta_1)\sum_{j=1}^{n} \Delta_j Q_{\theta_1,\theta_2}(T_j) Z(0,T_j)\,, \tag{9}
$$

where $Z(0,t)$ is the discount factor to time t; S_{θ_1,θ_2} is the coupon paid on the tranche and T_j are the payment dates. Δ_j is the accrual period between time T_{j-1} and T_j. Finally,

$$
\text{Tranche MTM} = \text{Premium Leg PV} - \text{Protection Leg PV} \tag{10}
$$

The Shape of Implied Loss Distributions

Lutz Schloegl
44-20 7260 2113
luschloe@lehman.com

Andrei Greenberg
44-20 7260 1285
andgreen@lehman.com

The liquid tranche markets for the iTraxx and CDX portfolios have greatly increased the information available on the market pricing of correlation risk. Market participants communicate implied correlation levels using the base correlation concept. We show how the implied portfolio loss distribution can be calculated explicitly from a base correlation curve in the LHP framework. The relatively simple formulae can be directly implemented to obtain a visualization of the loss distribution under different base correlation assumptions. The redistribution of risk across the different tranches can be tracked and compared, e.g. to a flat Gaussian copula assumption.

1. INTRODUCTION

The establishment of liquid markets for tranches of the standardized iTraxx and CDX portfolios has greatly increased transparency regarding implied correlation. The concept of a correlation smile has become central to the pricing and risk management of portfolio credit derivatives. From a modelling perspective, the market has adjusted to this phenomenon to some extent by modifying the widely-known Gaussian copula model in terms of input correlations, quoting different parts of the capital structure at different correlation levels. The base correlation methodology of decomposing mezzanine and senior tranches into their underlying "base" equity tranches has become popular because it imposes some measure of consistency across the capital structure (see O'Kane and Livesey (2004) for details). While the upward-sloping shape of the base correlation curves for iTraxx and CDX is widely publicized, it is not so clear what this means in terms of the actual loss distribution. In particular, because the expected loss of the portfolio is anchored in the individual default swap spreads, different assumptions about correlation essentially redistribute risk across tranches.

With the methodology introduced by Breeden and Litzenberger (1978), we can use the information contained in the base correlation curve to reconstruct the implied loss distribution. In the large homogeneous portfolio (LHP) framework of Vasicek (1987), we obtain a closed-form solution for the density of the portfolio loss distribution. We set out the methodology and a sketch of the derivation in the next sections. By visualizing the implied loss distribution, it becomes very clear how the market appears to be redistributing tranche risk relative to the Gaussian copula model. Probability mass is moved out of the mezzanine tranche, making this less risky. At the same time, losses within the equity tranche are redistributed. Under the implied loss distribution, the probability of having practically no losses is reduced compared with the Gaussian model. At the same time, there is a greater probability of having losses at the high end of the equity tranche, making that tranche riskier than under the Gaussian model. We illustrate the discrepancy between loss distributions for the iTraxx S2 and CDX IG3 portfolios.

2. LHP AND BASE CORRELATIONS IN CDO TRANCHE PRICING

In the Gaussian LHP model of Vasicek (1987), the assumption of a large homogeneous portfolio is used to invoke the Law of Large Numbers in the context of the Gaussian one-factor model to produce simple valuation formulae for the pricing of CDO tranches, see e.g. O'Kane *et al* (2003). For the percentage portfolio loss to exceed a certain threshold K, the realization of the market factor Z has to be less than $A(K)$ given by:

$$A(K) = \frac{1}{\beta}\left(C - \sqrt{1-\beta^2}\,\Phi^{-1}\left(\frac{K}{1-R} \right) \right) \tag{1}$$

Here R, β, and C correspond to the (identical) values of the recovery rate, correlation with the market factor, and default threshold, respectively, of all credits in the portfolio. The CDF of the normal distribution is denoted by Φ, the default threshold is obtained from the default probability p as $C = \Phi^{-1}(p)$. Fundamental to the concept of base correlation is the fact that the loss L^{tr} on any tranche with percentage strikes K_1 and K_2 can be decomposed as:

$$L^{tr} = [L - K_1]^+ - [L - K_2]^+ \tag{2}$$

For the pricing of a tranche, it is sufficient to compute the expected tranche loss for a series of time horizons. Again, see O'Kane *et al* (2003) for details. The popularity of the LHP framework stems from the fact that one has closed-form solutions for computing the expectation of call option payoffs $F(K) := E\left[[L-K]^+\right]$ on the portfolio loss distribution, these are known as stop-loss transforms in the actuarial literature. In the Gaussian LHP model we have:

$$F(K) = (1-R)\Phi_{2,\beta}\left(C, A(K)\right) - K\Phi\left(A(K)\right) \tag{3}$$

Here $\Phi_{2,\rho}$ denotes the CDF of the bivariate normal distribution with correlation coefficient ρ. In the classical Gaussian model, β is constant. The base correlation approach consists of pricing the different terms in equation (2) at different correlation levels. In effect, we now have a curve *β(K)* for the correlation with the market factor, which introduces an additional dependency on the strike K in equation (3). The values of *β(K)* at the strikes of the traded standard tranches are fitted to the observed tranche spreads. In effect, one observes the curve *β(K)* at five points for the iTraxx and CDX portfolios. Figures 1 and 2 show the implied and base correlations quoted for the iTraxx and CDX portfolios; these are derived from market prices using the LHP framework.

Figure 1. Implied and base correlation of iTraxx S2 at 15 March 2005, swap at 29

Detachment	3%	6%	9%	12%	22%
Implied Corr	23.8%	11.4%	17.0%	21.2%	30.6%
Base Corr	23.8%	32.6%	39.8%	46.1%	60.8%

Source: Lehman Brothers.

Figure 2. Implied and base correlation of CDX IG3 at 15 March 2005, swap at 43

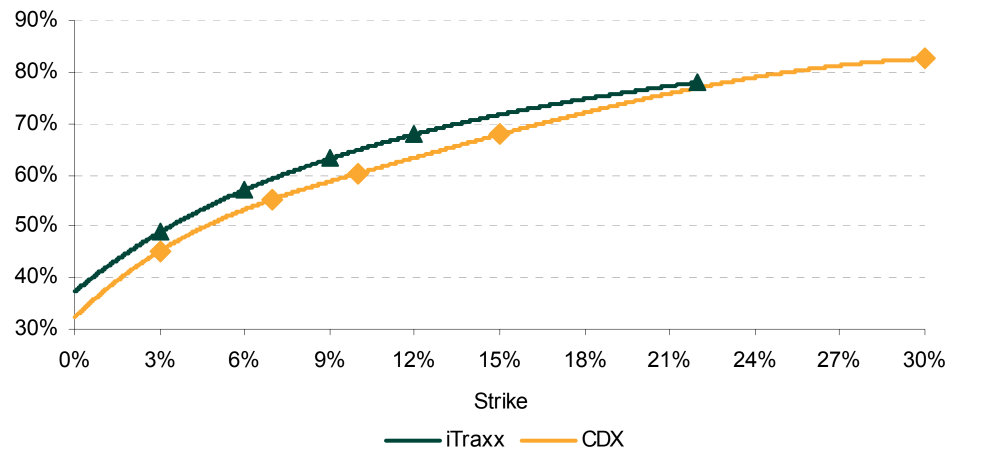
Detachment	3%	7%	10%	15%	30%
Implied Corr	20.4%	8.6%	17.2%	19.5%	29.6%
Base Corr	20.4%	30.5%	36.3%	46.0%	68.2%

Source: Lehman Brothers.

From the given quotes, we construct a curve *β(K)*. Note that the correlation with the market factor is the square root of the quoted base correlation. To generate a full curve, we use a cubic spline to interpolate between the quoted strikes. This ensures sufficient smoothness for our purposes, because we will need to take the second derivative of the curve *β(K)*. However, it should be stressed that no given interpolation method is guaranteed to be arbitrage-free, as the base correlation curve is not generated by a full self-consistent model. Figure 3 shows the interpolated *β(K)* curves; to re-iterate, this is the square root of the base correlation curve.

Because the market for protection above the detachment point of the topmost standard tranche is not very liquid, we have chosen not to extrapolate. This is not a problem as almost all of the portfolio risk is borne by the standardized tranches; above the topmost strike the portfolio becomes almost risk-free.

Figure 3. Interpolated market factor correlation curves



Source: Lehman Brothers.

3. IMPLIED LOSS DENSITIES

The portfolio loss distribution is given on the interval between zero and $1 - R$. To back out the density $f(x)$ of the implied distribution, we use the classic method of Breeden and Litzenberger (1978). This has its origin in the observation that $F(K)$ can be written as:

$$F(K) = \int_0^{1-R} [x - K]^+ f(x)dx \tag{4}$$

Taking distributional derivatives, we have:

$$F'(K) = \int_0^{1-R} H(x - K) f(x)\,dx \tag{5}$$

and:

$$F''(K) = \int_0^{1-R} \delta(x - K) f(x)\,dx = f(K) \tag{6}$$

where H denotes the Heaviside function and δ is the Dirac distribution. Equation (6) gives us the recipe to compute the loss density; we need to calculate the second derivative of $F(K)$ using equation (3), taking into account the dependency of β on K. The rest of the derivation is an exercise in using the chain rule. The structure of the final result is given by:

$$f(x) = \alpha_0 + \left(\alpha_{10} + \alpha_{11} \left.\frac{d\beta}{dK}\right|_{K=x} \right) \left.\frac{d\beta}{dK}\right|_{K=x} + \alpha_2 \left.\frac{d^2\beta}{dK^2}\right|_{K=x} \tag{7}$$

The density is a quadratic polynomial in the first and second derivatives of the function $\beta(K)$. Note, however, that the coefficients still depend on $\beta(K)$. The individual coefficients are:

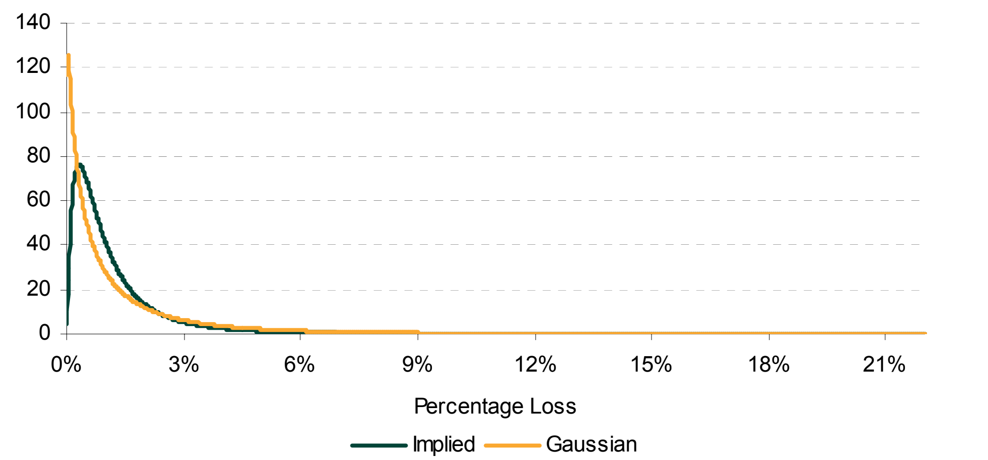
$$
\begin{aligned}
\alpha_0 &= \frac{\varphi(A)\sqrt{1-\beta^2}}{\varphi(B)\beta(1-R)} \\
\alpha_{10} &= -2\varphi(A)\frac{C\beta - A}{\beta\left(1-\beta^2\right)} \\
\alpha_{11} &= \frac{(1-R)\varphi(A)\varphi(B)}{\left(1-\beta^2\right)^{\frac{3}{2}}}\left(\beta - \frac{A}{\beta}(C\beta - A)\right) \\
\alpha_2 &= \frac{(1-R)\varphi(A)\varphi(B)}{\sqrt{1-\beta^2}}
\end{aligned}
\tag{8}
$$

Here A is as defined in equation (1), φ is the density of the standard normal distribution and B is given by $B = \Phi^{-1}\left(\frac{K}{1-R}\right)$. More details of the derivation are given in the Appendix.

4. RESULTS

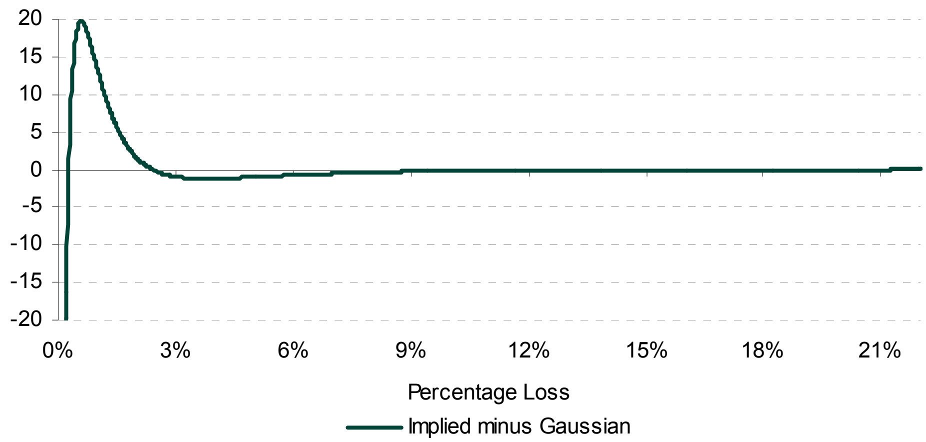
Using equations (7) and (8), we can plot the implied loss density compared with the Gaussian copula. For the flat correlation in the Gaussian case we use the implied correlation of the equity tranche because this is roughly equal to the average of the quoted implied correlations. Figure 4 shows the two loss distributions in the iTraxx case. We clearly see that the implied loss distribution allocates more risk to the higher part of the equity tranche, making the equity riskier than in the Gaussian case. Also, the implied loss density allocates less risk to the mezzanine tranche. This is difficult to see in Figure 4; the difference between the implied and the Gaussian density plotted in Figure 5 makes this somewhat clearer.

Figure 4. Implied and Gaussian loss density for iTraxx



Source: Lehman Brothers.

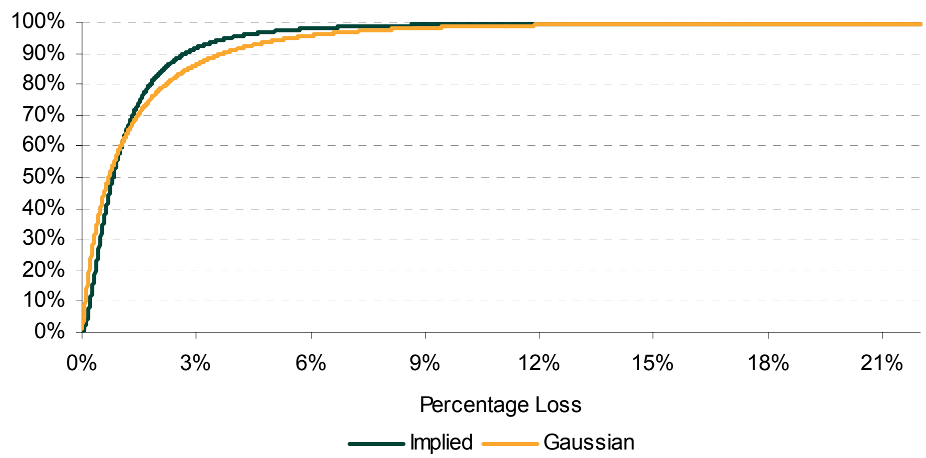
Figure 5. Difference between implied and Gaussian loss density for iTraxx



Source: Lehman Brothers.

Finally, we can numerically integrate the loss density to obtain the cumulative distribution function. This gives us another way to compare the two distributions (Figure 6).

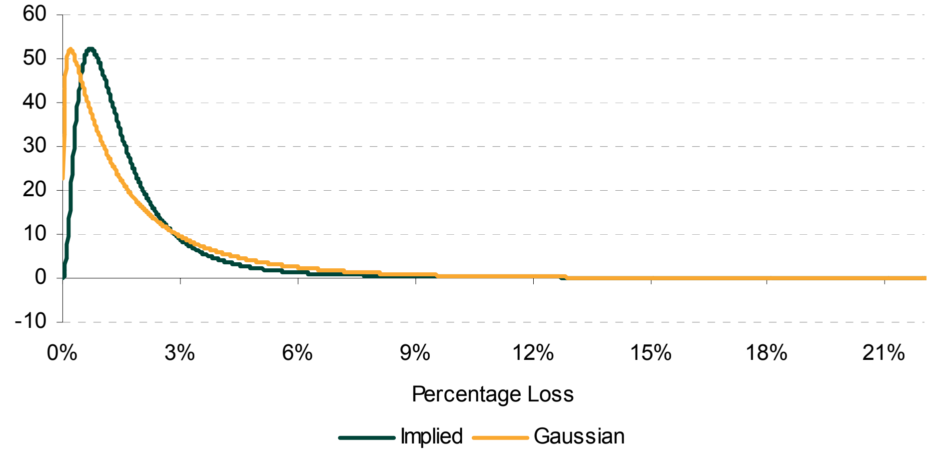
Figure 6. CDF of implied and Gaussian loss distribution for iTraxx



Source: Lehman Brothers.

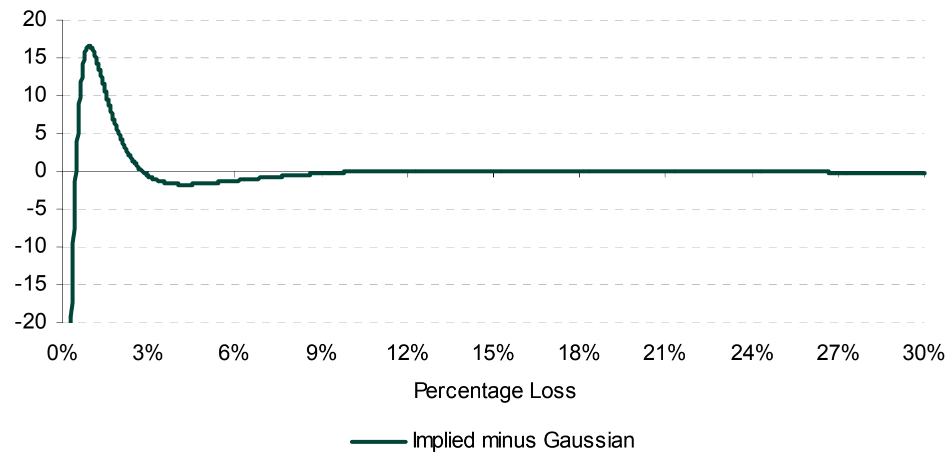
The same analysis on the CDX portfolio gives the loss densities shown in Figure 7. We immediately see the effect of the higher spread of the CDX portfolio compared with iTraxx by the fact that both loss distributions are pushed out to the right. We see a similar redistribution of risk out of the mezzanine and to the higher end of the equity tranche, the difference between densities is given in Figure 8. One should note that the CDX density becomes very mildly negative with values of -0.13 at high levels of the capital structure (around 19.5%). Clearly a negative density is not arbitrage-free. This does not imply that the quoted spreads are not arbitrage-free; rather one has to take care with the interpolation method used. The cubic spline is merely a simple method to interpolate the observed quotes using a twice differentiable function. Having said this, the issue of negative density values appears to be a very minor one in this case, as the negative values are quite small and only occur at a very high level of the capital structure. Figure 9 shows the two cumulative distribution functions for the CDX portfolio.

Figure 7. Implied and Gaussian loss density for CDX



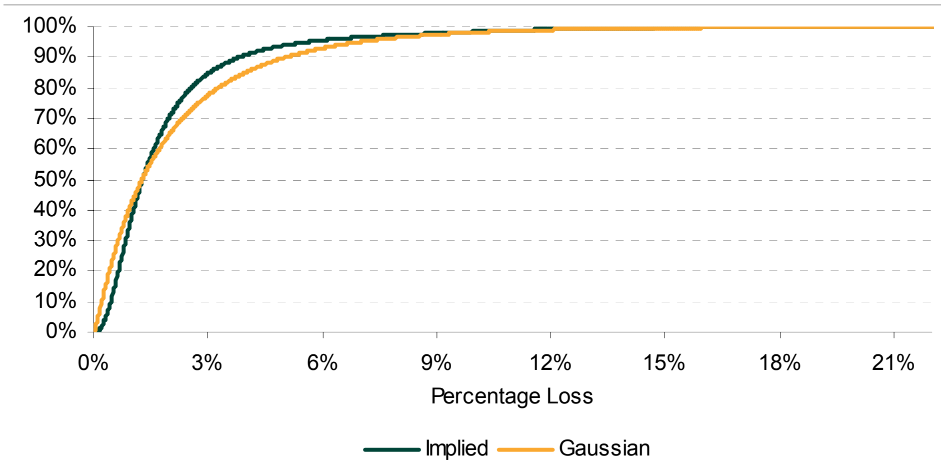
Source: Lehman Brothers.

Figure 8. Difference between implied and Gaussian loss density for CDX



Source: Lehman Brothers.

Figure 9. CDF of implied and Gaussian loss distribution for CDX



Source: Lehman Brothers.

5. CONCLUSIONS

Despite the lengthy calculations needed to compute the second derivative of stop-loss transforms, the final formula for the implied loss density given in equation (7) is surprisingly simple as a function of the slope and curvature of the base correlation curve. We see these closed-form expressions as a useful tool to visualize the impact of different base correlation assumptions on the implied loss distribution. The redistribution of risk across the tranches can be tracked and compared to a flat Gaussian correlation assumption. By showing this redistribution explicitly, we believe this work will also be useful in improving intuition for movements of the base correlation curve.

REFERENCES

Breeden and Litzenberger (1978), "Prices of State-Contingent Claims Implicit in Option Prices", *Journal of Business 51,* 1978

Genz (2004), "Numerical Computation of Rectangular Bivariate and Trivariate Normal and t Probabilities", *Statistics and Computing*, 14, 2004

O'Kane *et al* (2003), *The Lehman Brothers Guide to Exotic Credit Derivatives*, Risk, 2003

O'Kane and Livesey (2004), "Base Correlation Explained", *Quantitative Credit Research Quarterly*, 2004-Q3/Q4, Lehman Brothers

Vasicek (1987), *Probability of Loss on Loan Portfolio*, KMV Corporation, 1987.

APPENDIX

We explain how to obtain the results of equations (7) and (8). We have found it most convenient to compute the derivatives of F by taking partial derivatives with respect to K, β, and A. This gives:

$$\begin{aligned}\frac{d^2F}{dK^2} &= \frac{\partial^2 F}{\partial K^2} + 2\frac{\partial^2 F}{\partial K \partial A}\frac{\partial A}{\partial K} + \frac{\partial F}{\partial A}\frac{\partial^2 A}{\partial K^2} + \frac{\partial^2 F}{\partial A^2}\left(\frac{\partial A}{\partial K}\right)^2 \\ &+ 2\left[\frac{\partial^2 F}{\partial \beta \partial K} + \frac{\partial^2 F}{\partial A \partial K}\frac{\partial A}{\partial \beta} + \frac{\partial F}{\partial A}\frac{\partial^2 A}{\partial K \partial \beta} + \frac{\partial A}{\partial K}\frac{\partial^2 F}{\partial A \partial \beta} + \frac{\partial^2 F}{\partial A^2}\frac{\partial A}{\partial K}\frac{\partial A}{\partial \beta}\right]\frac{d\beta}{dK} \\ &+ \left[\frac{\partial^2 F}{\partial \beta^2} + 2\frac{\partial^2 F}{\partial A \partial \beta}\frac{\partial A}{\partial \beta} + \frac{\partial F}{\partial A}\frac{\partial^2 A}{\partial \beta^2} + \frac{\partial^2 F}{\partial A^2}\left(\frac{\partial A}{\partial \beta}\right)^2\right]\left(\frac{d\beta}{dK}\right)^2 \\ &+ \left[\frac{\partial F}{\partial \beta} + \frac{\partial F}{\partial A}\frac{\partial A}{\partial \beta}\right]\frac{d^2\beta}{dK^2}\end{aligned} \tag{9}$$

The partial derivatives of A are easily computed, to compute the partials of F, one needs to calculate derivatives of the bivariate normal distribution function. Genz (2004) gives the following formula:

$$\frac{\partial \Phi_{2,\rho}(x,y)}{\partial \rho} = \frac{\varphi(y)}{\sqrt{1-\rho^2}}\varphi\left(\frac{x-\rho y}{\sqrt{1-\rho^2}}\right) \tag{10}$$

Also, it is straightforward to show:

$$\frac{\partial \Phi_{2.\rho}(x,y)}{\partial y} = \varphi(y)\Phi\left(\frac{x-\rho y}{\sqrt{1-\rho^2}}\right) \tag{11}$$

Using this, we can establish the following relationships between the partial derivatives of F:

$$\begin{aligned}&\frac{\partial^2 F}{\partial K^2} = \frac{\partial^2 F}{\partial K \partial \beta} = \frac{\partial F}{\partial A} = 0 \quad \frac{\partial^2 F}{\partial K \partial A} = -\varphi(A) \\ &\frac{\partial^2 F}{\partial A^2} = -\beta\frac{\partial F}{\partial \beta} \quad \frac{\partial^2 F}{\partial A \partial \beta} = \frac{C\beta - A}{1-\beta^2}\frac{\partial F}{\partial \beta} \\ &\frac{\partial^2 F}{\partial \beta^2} = \frac{1}{1-\beta^2}\left[\beta - \frac{(C-\beta A)(C\beta - A)}{1-\beta^2}\right]\frac{\partial F}{\partial \beta} \\ &\frac{\partial F}{\partial \beta} = (1-R)\frac{\varphi(A)\varphi(B)}{\sqrt{1-\beta^2}}\end{aligned} \tag{12}$$

With these relationships, equation (9) can be simplified to give the following formulae for the coefficients in equation (7):

$$
\begin{aligned}
\alpha_0 &= 2\frac{\partial^2 F}{\partial K \partial A}\frac{\partial A}{\partial K} + \frac{\partial^2 F}{\partial A^2}\left(\frac{\partial A}{\partial K}\right)^2 \\
\alpha_{10} &= 2\left[\frac{\partial^2 F}{\partial A \partial K}\frac{\partial A}{\partial \beta} + \frac{\partial A}{\partial K}\frac{\partial^2 F}{\partial A \partial \beta} + \frac{\partial^2 F}{\partial A^2}\frac{\partial A}{\partial K}\frac{\partial A}{\partial \beta}\right] \\
\alpha_{11} &= \frac{\partial^2 F}{\partial \beta^2} + 2\frac{\partial^2 F}{\partial A \partial \beta}\frac{\partial A}{\partial \beta} + \frac{\partial^2 F}{\partial A^2}\left(\frac{\partial A}{\partial \beta}\right)^2 \\
\alpha_2 &= \frac{\partial F}{\partial \beta}
\end{aligned}
\tag{13}
$$

Finally making the derivatives explicit, one can then recover the formulae of equation (8).

Estimation of Credit Rating Transition Probabilities on LehmanLive[1]

Jin Liu
212-526-3716
jin.liu@lehman.com

Marco Naldi
212-526-1728
mnaldi@lehman.com

We present our LehmanLive tool for estimation of credit rating transition probabilities. The estimation methodology takes into account the presence of censored data, and allows the user to apply different time-weighting schemes to the history of rating transitions.

INTRODUCTION

Despite the advances of market-implied models of credit risk such as those developed by Moody's KMV (Crosbie and Bohn [2001], Kealhofer [2003]) and RiskMetrics' CreditGrades (Finkelstein, Pan, Lardy, Ta and Thierney [2002]), credit ratings remain a very influential signal in both investment grade and high yield markets. While the rating changes may indeed lag the market developments, the information contained in the act of a downgrade or upgrade as well as the specific language used to characterize its reasons and further credit outlook remains at the forefront of investors' minds and retains some explanatory power even when combined with structural models of credit risk (see Sobehart, Stein, Mikityanskaya and Li [2000] and Sobehart, Stein and Keenan [2000]).

The aggregate market-wide rating transition trends, such as upgrade/downgrade ratios, as well as the realized default rates (we consider default as a particular form of rating transition), are widely followed by portfolio managers as a gauge of the credit cycle. Many long-term credit investors, including those managing bank loan portfolios, insurance assets and CDOs, often pose a question: "what is the current estimate of the N-year rating transition probability?" (where typical values of N can be 1, 5 or 10 years). A reliable estimate of transition probabilities is an important ingredient of their macro investment strategies.

This article is organized as follows. First we present a continuous-time Markov model of rating transitions. We give a brief overview of the commonly used cohort estimators that are reported by rating agencies and followed by many practitioners, and explain their shortcomings. Next, following ideas developed by Lando *et al.* (Lando [2000], Lando and Skodeberg [2002], Christensen, Hansen and Lando [2004] and Fidelius, Lando and Nielsen [2004]), we define a maximum likelihood estimator for transition probabilities and extend the model to allow for time-weighted estimations. Finally, we provide some guidance on how to use the model.

1. A SIMPLE MODEL OF RATING TRANSITIONS

Let us assume that the credit rating transition probability of issuers between time t and s is summarized by an N-by-N transition matrix P(t, s), where N is the number of credit ratings. For example, N=8 for an eight-state letter-grade rating system [Aaa, Aa, …, Caa, D], or N=18 for an eighteen-state notched rating system [Aaa, Aa1, Aa2, …, Caa, D]. Note that we explicitly count default as the last state. The ij-th element of P(t, s) is $p_{i,j}(t, s)$, which is the transition probability between rating i and j from time t to s. We also assume the existence of

[1] *We would like to thank Arthur Berd, Lea Carty, Roy Mashal and Artem Voronov for comments and suggestions, and Timothy Soohoo for creating the LehmanLive application.*

a generator matrix Λ whose ij-th element $\lambda_{i,j}$, sometimes called transition hazard rate, is defined by[2]:

[1] $$\lambda_{i,j} = \lim_{h \to 0+} \frac{p_{i,j}(t, t+h)}{h}$$

This simple model satisfies the Markov property because $p_{i,j}(t,s)$ depends only on the initial rating i at time t, not on the rating history before time t. It is also time-homogeneous because the generator matrix is time independent. The goal of this paper is to use historical rating transition data to estimate the generator matrix and derive the transition probability matrix P(t, s) for different time horizons (s-t).

2. ESTIMATION OF TRANSITION PROBABILITIES

Having defined the rating transition model in the previous section, we now turn to the estimation of the model using historical data. First, we present a brief overview of the commonly used cohort estimators that are reported by rating agencies and explain their shortcomings. Next, we define a maximum-likelihood estimation of the generator matrix, explicitly taking into account the problem of censored data. We then extend the model to allow for a time-weighted estimation procedure.

2.1. Cohort Estimator for the Transition Probability Matrix

The simplest estimator of the historical transition probabilities is obtained by counting the number of times a particular type of transition N_{ij} (from rating *i* to rating *j*) has actually occurred during the observation period, and dividing it by the number of firms N_i that entered the observation period with that particular initial rating *i*. This so-called cohort estimator is regularly published by major credit rating agencies such as Moody's and Standard and Poor's (see for example Carty [1997] and Carty and Fons [1993]). The name "cohort" comes from the particular implementation which tracks the transitions by the "cohort year" in which they occurred – e.g. the transition matrices for the "1990-2000 cohort":

[2] $$T_{ij} = \frac{N_{ij}}{N_i} \qquad \text{for } j \neq i$$

Note that if there are no transitions from rating *i* to rating *j* in the chosen cohort, then the estimate of the corresponding transition probability is equal to 0. This approach has a few important pitfalls:

---

[2] *A few important properties of the generator matrix are:*

$$\lambda_{i,i} = -\sum_{j \neq i} \lambda_{i,j}$$

$$p_{i,i}(t,s) = \exp\left(\lambda_{i,i}(s-t)\right)$$

$$P(t,s) = \exp\left(\Lambda(s-t)\right) = \sum_{m=0}^{\infty} (\Lambda(s-t))^m / m!$$

More details on these properties can be found in Lando and Skoderberg [2002], and Jarrow et al. *[1997].*

- Data grouping: if company X was rated A for 11 months before transitioning to Baa, while company Y was rated A for only 1 month, both will contribute the same to [2].
- Rating transitions are rare events, and a long history is needed to get reliable estimates. This also makes it difficult to incorporate model dynamics since we cannot afford to divide the sample into smaller pieces.
- The choice of cohort time windows is somewhat artificial. Moody's convention for one-year estimates is [Jan. 1 of year T; Jan. 1 of year $T+1$].

Carty and Fons (1993) suggest a remedy to the censored data problem using Weibull distributions. However, even with this modification, the cohort estimator remains very sensitive to the choice of inclusion rules for (rare) transition events.

2.2. Maximum Likelihood Estimation of the Transition Generator Matrix

In order to construct the likelihood function for the continuous-time transition generator matrix, we need to define what constitutes an observation. We will account not only for the event of a transition, but also for the timeline of the rating process, starting from the date when the rating was first observed (either the date of the rating assignment or the first date of the observation period), following with the time interval during which the rating remained unchanged, and finishing with the date when the rating state ceased to be valid, which in turn could be due to transition to another rating, default, rating withdrawal, or the end of the observation period.

Different types of observations are illustrated in Figure 1, which mirrors the schema of the transition events table that we constructed using Moody's default and transition database[3]. We show two examples of fully observed events (a ratings upgrade and a default), and three examples of censored events, where only partial information was observed – a rating withdrawal, a rating surviving until the end of the observation window, and a rating observation starting at the beginning of the observation window. The last event is called "left censored" because the censoring (incomplete information) is at the left (earlier) time boundary, while the preceding two cases are called "right censored" because the incomplete information pertains to the right (latter) time boundary.

Figure 1. Transition events, observation window 1/1/1990 – 10/30/2004

Event Type	Event #	Issuer	Start Date	Start State	End Date	End State
Fully Observed	1	A	1/1/1992	Baa1	2/15/1995	A3
Fully Observed	2	B	5/15/1994	Ba3	12/15/1998	D
Right Censored	3	C	4/10/1996	Aa2	6/15/2001	RW
Right Censored	4	D	10/10/2002	A1	10/30/2004	A1
Left Censored	5	E	1/1/1990	Ba2	6/30/1995	Baa1

Source: Lehman Brothers, sample.

In a continuous-time setting, each of these observations can be further decomposed into a series of mini-observations referring to consecutive short time periods, such that when strung together these mini-observations constitute the chosen line in Figure 1. For example, the observation that issuer A started in rating Baa1 on 1/1/1992 and transited to rating A3 on 2/15/1995 is reinterpreted as a mini-observation that it remained at rating Baa1 for the one-day period [1/1/1992, 1/2/1992], followed by a mini-observation that it remained at rating Baa1 for the

[3] *In 1992, Moody's changed the rating system from letter-grade rating (7 ratings) into notched rating (17 ratings). In Appendix A, we discuss how to deal with this issue. In Appendix B, we show how to convert the transition probability matrix for notched rating system to a letter-grade rating system.*

next period [1/2/1992, 1/3/1992], and so on until the final mini-observation that the issuer transited from rating Baa1 to rating A3 during the last one-day period [2/14/1995, 2/15/1995].

We will assume in what follows a strong independence between the observations of the rating transition process. Namely, we will assume that:

- The rating transition process is Markov, i.e. all transition probabilities for each infinitesimal time period depend only on the ratings at the beginning of the period. Thus we explicitly exclude ratings momentum and other path-dependencies.
- Rating transitions across issuers for each infinitesimal time period are conditionally independent, given their respective rating states and the rating-dependent transition probabilities at the beginning of the period.

The latter assumption is consistent with the recent empirical tests of conditional independence of the default times across issuers reported by Das, Duffie and Kapadia (2004), who show that after including the contemporaneous estimates of hazard rates and macro factors as explanatory variables, there is only very mild evidence for excess clustering of default.

Under these conditional independence assumptions, the complete likelihood function for the entire set of observations can be written as a product of likelihood functions for each event row, which in turn can be written as a product of mini-likelihood functions for each infinitesimal time period. Therefore, the total log-likelihood function is simply a sum of mini-log-likelihood functions across all observations.

Let us enumerate the observation rows by *k,* and denote the initial rating for the row by r_k^s, the final rating for the row by r_k^e. With these notations, we can write down the mini-log-likelihood functions for small time intervals as follows (here $L_t(k, r_k^s)$ corresponds to a time period when no change is observed, $L_t(k, r_k^s, r_k^e)$ corresponds to a time period when a transition was observed, and finally $L_t(k, r_k^s, C)$ corresponds to a time period when a right-censored observation occurred):

$$\text{[3]} \quad \begin{cases} L_t(k, r_k^s) & = & -\sum_{q \neq r_k^s} \lambda_{r_k^s q}(t) \cdot dt \\ L_t(k, r_k^s, r_k^e) & = & \log\left(\lambda_{r_k^s r_k^e}(t) \cdot dt\right) \\ L_t(k, r_k^s, C) & = & -\sum_{q \neq r_k^s} \lambda_{r_k^s q}(t) \cdot dt \end{cases}$$

Taking a sum of these mini-log-likelihood functions across the time periods corresponding to a given observation row, and then summing across rows, we obtain the following result:

$$\text{[4]} \quad L = -\sum_k \left(\int_{T_k^{start}}^{T_k^{end}} \left(\sum_{q \neq r_k^s} \lambda_{r_k^s q}(t) \right) \cdot dt \right) + \sum_k \log\left(\lambda_{r_k^s r_k^e}(t) \cdot dt\right) \cdot \left(1 - I_k^{RC}\right)$$

Here, I_k^{RC} is the indicator for a right-censored event. Note that the left-censoring does not influence the log-likelihood function at all. We denoted by T_k^{start} and T_k^{end} the effective starting and ending times for the row *k,* respectively. Note that effective starting and ending times depend on the chosen observation window, and possibly differ from the starting t_k^{start}

and ending t_k^{end} times of a given rating transition event recorded in the database (this will be important for rolling estimates of the transition probabilities):

$$[5] \quad \begin{cases} T_k^{start} & = & \max\left[t_k^{start}, T_{window}^{start}\right] \\ T_k^{end} & = & \min\left[t_k^{end}, T_{window}^{end}\right] \end{cases}$$

With these definitions, we can rewrite the right-censored indicator in eq. [4] explicitly using indicators for a rating withdrawal and for an incomplete observation:

$$[6] \quad 1 - I_k^{RC} = \left(1 - I_{\left\{r_k^{end} = RW\right\}}\right) \cdot \left(1 - I_{\left\{t_k^{end} > T_{window}^{end}\right\}}\right)$$

Assuming that the transition hazard rates are constant in time, i.e. $\lambda_{ij}(t) = \lambda_{ij}$, we can easily obtain the maximum likelihood estimator of the transition generator matrix (Lando [2000], Lando and Skodeberg [2002], Christensen, Hansen and Lando [2004] and Fidelius, Lando and Nielsen [2004]):

$$[7] \quad \hat{\lambda}_{ij} = \frac{\sum_k I_{\left\{r_k^{start} = i\right\}} \cdot I_{\left\{r_k^{end} = j\right\}} \cdot \left(1 - I_{\left\{r_k^{end} = RW\right\}}\right) \cdot \left(1 - I_{\left\{t_k^{end} > T_{window}^{end}\right\}}\right)}{\sum_k \left(T_k^{end} - T_k^{start}\right) \cdot I_{\left\{r_k^{start} = i\right\}}} \quad \text{for } i \neq j$$

In other words, the maximum likelihood estimator for the element of the generator matrix corresponding to a transition from rating *i* to rating *j* is equal to the count of all non-right-censored transitions within the observation window, divided by the total time that any issuer spent in the state with rating *i* (whether the event was censored or not). Note that since the numerator includes a condition that the final rating state is *j,* then the indicator for rating withdrawal will automatically drop out (we retained it in [7] for completeness).

2.3. Time-Weighted Maximum Likelihood Estimation

In the previous section we derived a maximum likelihood estimator for the transition generator matrix that takes censored data into account. However, empirical studies suggest that one should account for time variation of market conditions in order to explain the observed patterns of rating transitions. Without allowing for such variation, there is a significant clustering of downgrades and defaults in times of market downturns, such as 1990-1992 and 2000-2002 (see Das, Duffie and Kapadia [2004] for a detailed discussion of default risk clustering).

Indeed, equation [7] implies that a company that was downgraded from Baa to Ba in 1996 having survived five years at Baa, is treated similarly to a company that was downgraded in 2002 having survived only one year at Baa. Therefore, the typically longer survival time of companies during the mid-1990s would significantly increase the denominator of [7] and depress the estimate of the corresponding downgrade probability.

For many practitioners it is obvious that the benign credit risk environment in 1996 had little in common with the severe credit downturn during 2002. The shorter lifespan of a company in a given rating category during 2002 was not an exception but the rule. Hence, if we were trying to estimate the rating transition risk in 2002, it might not have been meaningful to dilute the information contained in the recent volatile history of transitions by adding to it a large number of much more tempered transition events from the mid 1990s.

An intuitive way to account for the time variation of credit market conditions is to apply a time-weighting scheme to the maximum likelihood estimator. A common scheme is the geometric time-decay with half-life T_H, which assigns the following weight to an event that occurred at time *t* and is observed at later time *T*:

[8] $$w(t,T) = 2^{-\frac{T-t}{T_H}}$$

Equation [8] implies that an event that occurred T_H ago receives half of the weight received by an event that occurred just an instant ago.

To estimate a weighted log-likelihood function, we remember that it was constructed from independent mini-likelihood functions [3]. Assigning each mini-observation a weight corresponding to the time of the mini-event is the same as taking each of the mini-likelihood functions to the power of the corresponding weight, or equivalently, multiplying each mini-log-likelihood function by the chosen weight. Therefore, we obtain:

[9] $$\begin{aligned} L(T) \;&=\; -\sum_k \left(\int_{T_k^{start}}^{T_k^{end}} w(t,T) \cdot \left(\sum_{q \neq r_k^s} \lambda_{r_k^s q}(T) \right) \cdot dt \right) \\ &\quad + \sum_k w\left(T_k^{end}, T\right) \cdot \log\left(\lambda_{r_k^s r_k^e}(T) \cdot dt\right) \cdot \left(1 - I_k^{RC}\right) \end{aligned}$$

Note that we explicitly acknowledged the dependence of the transition generator estimate on the rolling observation date – as the estimation time $T = T_{window}^{end}$ goes on, the maximum likelihood estimator will change both because of the new events being observed and because of the changing weights. The time-weighted analog of equation [7] is given by:

[10] $$\hat{\lambda}_{ij}(T) = \frac{\sum_k w\left(T_k^{end}, T\right) \cdot I_{\left\{r_k^{start} = i\right\}} \cdot I_{\left\{r_k^{end} = j\right\}} \cdot \left(1 - I_{\left\{t_k^{end} > T\right\}}\right)}{\sum_k I_{\left\{r_k^{start} = i\right\}} \cdot \int_{T_k^{start}}^{T_k^{end}} w(t,T) \cdot dt}$$

Substituting the specific form of the weights in [8], we obtain:

[11] $$\hat{\lambda}_{ij}(T) = \frac{\sum_k 2^{-\frac{T - T_k^{end}}{T_H}} \cdot I_{\left\{r_k^{start} = i\right\}} \cdot I_{\left\{r_k^{end} = j\right\}} \cdot \left(1 - I_{\left\{t_k^{end} > T\right\}}\right)}{T_H \cdot \sum_k I_{\left\{r_k^{start} = i\right\}} \cdot \left(2^{-\frac{T - T_k^{end}}{T_H}} - 2^{-\frac{T - T_k^{start}}{T_H}} \right)}$$

We can see that in the limit $T_H \to \infty$ the estimate converges to the unweighted case [7].

2.4. A Simple Example

To illustrate both the cohort and the continuous-time estimation procedure, consider a simple example, which is a modified version of the one suggested by Lando and Skodeberg (2002).

Let us consider a rating system with two non-default rating categories A and B, and a default category D. Assume that we observe over one year the history of 20 firms, of which 10 start in category A and 10 in category B. Assume that over the year of observation, one A-rated firm changes its rating to category B after three months and stays there for the rest of the year. Finally, assume that over the same period, one B-rated firm is upgraded after nine months and remains in A for the rest of the period, and a firm which started in B defaults after six months and stays there for the remaining part of the period (see Figure 2, where we have highlighted the rows with non-trivial transition activity. Dates are measured in months).

Figure 2. Example transition events table, observation period [0, 12] months

Event #	Issuer	Start Date	Start State	End Date	End State
1	1	0	A	3	B
2	1	3	B	12	B
3	2	0	A	12	A
4	3	0	A	12	A
5	4	0	A	12	A
6	5	0	A	12	A
7	6	0	A	12	A
8	7	0	A	12	A
9	8	0	A	12	A
10	9	0	A	12	B
11	10	0	A	12	B
12	11	0	B	9	A
13	11	9	A	12	A
14	12	0	B	6	D
15	13	0	B	12	B
16	14	0	B	12	B
17	15	0	B	12	B
18	16	0	B	12	B
19	17	0	B	12	B
20	18	0	B	12	B
21	19	0	B	12	B
22	20	0	B	12	B

Source: Lehman Brothers, simulated example.

By straightforward application of formulas [2], [7] and [11], we can now calculate the cohort estimator of the 1-year transition probability matrix, as well as the maximum likelihood estimator of the 1-year generator matrix both under a stationary assumption, and under a time-weighted assumption with a half-life of six months. Having obtained the estimates of the generator matrices, we can also calculate, via a matrix exponential, the estimates of the 1-year transition probability matrices for the latter two cases. The results of this exercise are presented in Figure 3. We show the estimates for the generator matrices on the left and the corresponding estimates for the transition probability matrices on the right. For the cohort method, the generator matrix is calculated from the transition matrix by taking a matrix log.

Figure 3. Estimates for transition matrices under various assumptions

	Transition Generator Matrix				Transition Probability Matrix		
	Cohort Estimator						
	A	**B**	**D**	⇦ mlog	**A**	**B**	**D**
A	-0.1121	0.1183	-0.0063	**A**	0.9000	0.1000	0
B	0.1183	-0.2304	0.1121	**B**	0.1000	0.8000	0.1000
D	0	0	0	**D**	0	0	1
	Stationary Continuous-Time Estimator						
	A	**B**	**D**	mexp ⇨	**A**	**B**	**D**
A	-0.3158	0.3158	0.0000	**A**	0.7412	0.2454	0.0134
B	0.1000	-0.2000	0.1000	**B**	0.0777	0.8312	0.0911
D	0	0	0	**D**	0	0	1
	Time-Weighted Continuous-Time Estimator						
	A	**B**	**D**	mexp ⇨	**A**	**B**	**D**
A	-0.4566	0.4566	0.0000	**A**	0.6544	0.3283	0.0173
B	0.1333	-0.2276	0.0943	**B**	0.0959	0.8190	0.0851
D	0	0	0	**D**	0	0	1

Source: Lehman Brothers.

A couple of observations can be made by comparing the results in Figure 3. First, we note that even in this simple example it becomes apparent that the cohort estimator is not consistent with a continuous-time Markov description of the rating transition process. Indeed, the highlighted cells in the A-D transition rows show that the cohort estimate of the 1-year default probability for category A being exactly zero could only be reconciled with the continuous-time framework if the transition hazard rate from A to D were negative.

We can understand this better by looking at the case of the stationary continuous-time estimate. Here, the transition hazard rate from A to D is zero. However, it still leads to a non-zero 1-year transition probability from A to D, due to the possibility of multi-stage transitions like A-B-D, A-B-A-B-D, etc. The only elements of the 1-year transition matrix that are exactly zero correspond to forbidden transitions from D to A or B. Consequently, it is impossible to find a generator matrix with non-negative transition hazard rates that would produce a zero probability for the 1-year transition A-D.

Second, comparing the stationary and the time-weighted continuous-time generator matrices and their corresponding 1-year transition probabilities, we can see that the impact of the weighting scheme can be quite substantial. Moreover, the way it affects the estimates is driven by a complex interplay between the time of the transition event (more recent transitions increase the corresponding hazard rate) and the duration of the stay in the initial rating (longer survival times decrease the corresponding hazard rate). For example, the A-B transition probability went up by a much larger factor than the B-D transition probability went down, even though both are driven by a single event, because the survival time length (three months vs. six months) had a bigger impact on the hazard rate estimates than the event timing (nine months ago vs. six months ago) for these two transition events.

3. USING HISTORICAL TRANSITION MATRICES ON LEHMANLIVE

The historical transition probabilities on LehmanLive can be found following the path:

Fixed Income > Credit > Quant Toolkit

As we have seen in the previous sections, our credit rating transition model is based on the estimation of a generator matrix which is then "compounded" to produce cumulative transition probabilities for a specific time horizon. On LehmanLive, a user can choose time horizons varying from 1 to 10 years. Users can also set the estimation date in the past; this will give estimates of transition probabilities based on data observed on or before the chosen estimation date.

A second crucial input that a user must specify is the half-life parameter T_H in equation [8]. Rating transitions that occurred T_H years before the estimation date are assigned half the weight of transitions that occurred one instant before the estimation date. Users can set the half-life parameter to be 1, 3, 5, 10 and 20 years. Using a 20-year half-life is for all purposes equivalent to using no time-decay in the estimation. The shorter half-lives of 1, 3, and 5 years are appropriate if the user believes that recent history will be more representative of the chosen time horizon. We feel that a 1-year half-life estimate is only marginally useful since the effective number of transition events is very low and the accuracy of the estimate is poor; we produce it for completeness and for comparison with the popular 1-year cohort estimates. The 3- and 5-year half-life estimates, on the other hand, are generally based on a meaningful amount of information, and may be appropriate for estimating the generator matrix when the final goal is to compute a transition probability matrix for a 1- to 5-year horizon.

In most cases, the application itself will suggest the best way to use our tool. For example, if an investor is interested in the analysis of credit losses for a hypothetical 10-year CLO backed by a diversified pool of loans about which only the ratings distributions are known, then estimating the 10-year half-life generator matrix and deriving the corresponding cumulative default probabilities for a 10-year horizon would seem reasonable. On the other hand, if an asset manager is inquiring about the upgrade-to-downgrade ratio in the next year, then using an estimate based on a short 3-year half-life is more likely to yield a useful answer.

CONCLUSION

In this paper we have described the methodology for estimation of credit rating transition probabilities available on LehmanLive. Building on recent work by Lando and others, we have presented a continuous-time rating transition framework as well as a maximum-likelihood methodology for the estimation of the model's parameters.

Some of the uses of rating transition probabilities include:

- Credit portfolio loss estimates for diversified portfolios pooled by rating. As a particular example, many CLOs report the rating composition of the collateral portfolios but not the specific issuers, in which case our model can be used for estimating the (real-world, not implied) loss rates for the collateral and the tranches.
- Risk management of rating-contingent obligations, such as bonds with step-up coupon provisions.
- Risk management of revolver loan commitments. It is often assumed that a credit line will be drawn if an issuer is downgraded to high yield and is no longer able to access the

capital markets at advantageous rates. Our cumulative downgrade probability estimates can be used for scenario-based risk management of revolvers.

- Counterparty risk management. Many smaller trading counterparties do not have any outstanding public debt or CDS (or even equity) on which to base the estimates of their credit risk and margin requirements. Given credit analyst assessment of "shadow" rating of such counterparties, one can use the transition matrix model for setting and monitoring margin requirements and lines of credit.

REFERENCES

Carty, L. (1997): *Moody's Rating Migration and Credit Quality Correlation 1920-1996*, special comment, Moody's Investor Services

Carty, L. and J. Fons (1993): *Measuring Changes in Corporate Credit Quality*, special comment, Moody's Investor Services

Christensen, J., E. Hansen and D. Lando (2004): "Confidence Sets for Continuous-Time Rating Transition Probabilities", *Journal of Banking and Finance*, vol. 28 (11), p. 2575

Crosbie, P. and J. Bohn (2001): *Modeling Default Risk*, working paper, KMV

Das, S., D. Duffie, and N. Kapadia (2004): *Common Failings: How Corporate Default are Correlated*, working paper, Stanford University and University of Santa Clara

Finkelstein, V., G. Pan, J.-P. Lardy, T. Ta and J. Thierney (2002): *CreditGrades: Technical Document*, RiskMetrics Group

Fidelius, P., D. Lando and J. P. Nielsen (2004): "Non-Parametric Analysis of Rating Transition and Default Data", *Journal of Investment Management*, vol. 2 (2), p. 71

Kealhofer S. (2003): "Quantifying Credit Risk I: Default Prediction", *Financial Analyst Journal*, vol. 59, no. 1 (January/February), p. 30

Lando, D. (2000): "Some Elements of Rating-Based Credit Risk Modeling", in *Advanced Fixed Income Valuation Tools* (N. Jegadeesh and B. Tuckman, eds.), Wiley & Sons, p. 193

Lando, D. and T. M. Skodeberg (2002): "Analyzing Rating Transitions and Rating Drift with Continuous Observations", *Journal of Banking and Finance*, vol. 26, p. 423

Sobehart, J., R. Stein, V. Mikityanskaya and L. Li (2000): *Moody's Public Firm Risk Model: a Hybrid Approach to Modeling Short-Term Default Risk*, special comment, Moody's Investor Services

Sobehart, J., S. Keenan and R. Stein (2000): *Benchmarking Quantitative Default Risk Models: a Validation Methodology*, special comment, Moody's Investor Services

APPENDIX A: LETTER-GRADE RATINGS AND PARTIAL CENSORING

Prior to 1992, Moody's assigned only letter-grade credit ratings (Aaa, Aa, A, Baa, Ba, B, Caa) to issuers. In January 1992, the agency switched to a notched rating system, by splitting each of the letter grade ratings (except Aaa) into three alpha-numeric ratings (for example, the Baa category was split into Baa1, Baa2 and Baa3).

This change of rating denominations presents a challenge for empirical estimation. We confront it by regarding the pre-1992 letter-grade ratings as partially censored versions of the full, "unobservable" notched rating system.

For any transition from a letter-grade rating R to a notched rating *r*, we set the mini-log-likelihood function equal to the equally-weighted average of the values corresponding to transitions from the "unobserved" notched ratings R1, R2, and R3 to the final notched rating *r*. This is consistent with having no information on the unobserved notch state before the transition. Similarly, a transition from a notched rating to a letter-grade rating (if such an event is ever observed) is treated as a partially censored observation corresponding to either one of the "unobserved" final notched ratings. Finally, the pre-1992 transitions from letter-grade to letter-grade rating are treated as partially censored both at the beginning and at the end of the observation.

These rules are summarized in the following equations:

$$[1] \quad \begin{cases} L_t\left(k, R_k^s, r_k^e\right) & = & \frac{1}{3}\cdot\left[L_t\left(k, R1_k^s, r_k^e\right) + L_t\left(k, R2_k^s, r_k^e\right) + L_t\left(k, R3_k^s, r_k^e\right)\right] \\ L_t\left(k, r_k^s, R_k^e\right) & = & \frac{1}{3}\cdot\left[L_t\left(k, r_k^s, R1_k^e\right) + L_t\left(k, r_k^s, R2_k^e\right) + L_t\left(k, r_k^s, R3_k^e\right)\right] \\ L_t\left(k, R_k^s, R_k^e\right) & = & \frac{1}{9}\cdot\left[L_t\left(k, R1_k^s, R1_k^e\right) + L_t\left(k, R2_k^s, R1_k^e\right) + L_t\left(k, R3_k^s, R1_k^e\right)\right. \\ & + & L_t\left(k, R1_k^s, R2_k^e\right) + L_t\left(k, R2_k^s, R2_k^e\right) + L_t\left(k, R3_k^s, R2_k^e\right) \\ & + & \left. L_t\left(k, R1_k^s, R3_k^e\right) + L_t\left(k, R2_k^s, R3_k^e\right) + L_t\left(k, R3_k^s, R3_k^e\right)\right] \end{cases}$$

Note that when the observed transition is from a letter grade to one of its subset notched ratings (such as from Ba to Ba2), then one of the possible "unobserved" initial states is the same as the final state, and therefore corresponds to a "no-transition" event. This is automatically taken into account by properly calculating the mini-log-likelihood function for such events, which does not contain the second non-integral part in equation [9]. A similar statement is valid for a transition from a notched rating to a letter-grade rating.

We have also made a simplification by grouping all Caa1 or below ratings into a single Caa designation. We do not make any adjustments for censored observations corresponding to transitions between these states.

Substituting these terms into the log-likelihood function [9], we obtain the necessary objective function that covers the entire period available in Moody's database.

APPENDIX B: COARSE-GRAINING THE TRANSITION MATRICES

Investors often ask for transition probability estimates on a coarse-grained, letter-grade scale spanning eight rating states [Aaa, Aa, A, Baa, Ba, B, Caa, D]. This can be useful when addressing macro trends in credit markets or broad asset allocation strategies, where it is impractical to maintain a very fine classification with 18 notched rating states [Aaa, Aa1, Aa2, Aa3, A1, A2, A3, Baa1, Baa2, Baa3, Ba1, Ba2, Ba3, B1, B2, B3, Caa, D].

One could potentially answer this question by re-mapping all transition events to their corresponding letter-grade ratings and then applying the same estimation methodology as described in this paper with a smaller set of ratings. The problem with this approach is that it would lead to a dramatic reduction in the number of observed transitions and less efficient estimates of the parameters.

An alternative is to start with the already estimated transition generator matrix for the notched ratings and derive a generator matrix for a coarse-grained rating variable that continues to satisfy the Markov property.

Let R^s denote the coarse-grained initial rating state which has three notched sub-ratings $R1^s, R2^s, R3^s$. Let also r^e denote a final state which has no sub-ratings, i.e. one of [Aaa, Caa, D]. Assuming equally probable initial sub-ratings, the transition intensity from R^s to r^e is equal to the average of the transition intensities from the initial sub-ratings to the final state.

Let now R^e denote the coarse-grained final rating state and r^s denote an initial state which has no sub-ratings. Since the final notched states are mutually exclusive, the transition intensity from r^s to R^e is equal to the sum of the transition intensities from the initial state to the final sub-ratings.

Finally, the transition intensity between two coarse-grained states can be obtained by applying the above rules iteratively. These rules are summarized as follows:

$$[2] \quad \begin{cases} \lambda(R^s, r^e) & = & \frac{1}{3} \cdot \left[\lambda(R1^s, r^e) + \lambda(R2^s, r^e) + \lambda(R3^s, r^e)\right] \\ \lambda(r^s, R^e) & = & \lambda(r^s, R1^e) + \lambda(r^s, R2^e) + \lambda(r^s, R3^e) \\ \lambda(R^s, R^e) & = & \frac{1}{3} \cdot \left[\lambda(R1^s, R1^e) + \lambda(R2^s, R1^e) + \lambda(R3^s, R1^e)\right. \\ & + & \lambda(R1^s, R2^e) + \lambda(R2^s, R2^e) + \lambda(R3^s, R2^e) \\ & + & \left.\lambda(R1^s, R3^e) + \lambda(R2^s, R3^e) + \lambda(R3^s, R3^e)\right] \end{cases}$$

One can see from these equations that the rows of the resulting coarse-grained transition generator matrix sum up to zero, and that its off-diagonal elements are positive, which is sufficient to ensure that the coarse-grained transition generator matrix corresponds to a valid Markov process. Once this generator matrix is constructed, the corresponding transition probability matrices for various horizons can be calculated by the standard matrix exponentiation rules.

The results of this coarse-graining procedure for the 20-year half-life transition generator and the corresponding 1-year transition matrix, estimated as of 1/1/2005, are shown in Figure 4.

Figure 4a. Letter-grade transition generator matrix as of 1/1/2005 with 20-year half-life

Rating	AAA	AA	A	BAA	BA	B	CCC	D
AAA	-0.0856	0.0806	0.0049	0.0000	0.0000	0.0000	0.0000	0.0000
AA	0.0180	-0.0927	0.0728	0.0013	0.0004	0.0003	0.0000	0.0000
A	0.0005	0.0337	-0.1084	0.0716	0.0020	0.0006	0.0000	0.0000
BAA	0.0004	0.0029	0.0571	-0.1442	0.0769	0.0056	0.0007	0.0007
BA	0.0001	0.0009	0.0046	0.0715	-0.1968	0.1091	0.0048	0.0058
B	0.0001	0.0009	0.0021	0.0047	0.0544	-0.2547	0.1464	0.0460
CCC	0.0000	0.0000	0.0003	0.0019	0.0045	0.0658	-0.3219	0.2495
D	0	0	0	0	0	0	0	0.0000

Source: Lehman Brothers.

Figure 4b. Letter-grade 1-year transition probability matrix as of 1/1/2005 with 20-year half-life

Rating	AAA	AA	A	BAA	BA	B	CCC	D
AAA	0.9186	0.0739	0.0072	0.0003	0.0000	0.0000	0.0000	0.0000
AA	0.0165	0.9133	0.0660	0.0035	0.0005	0.0003	0.0000	0.0000
A	0.0008	0.0306	0.9002	0.0633	0.0041	0.0008	0.0001	0.0001
BAA	0.0004	0.0035	0.0507	0.8699	0.0652	0.0081	0.0012	0.0011
BA	0.0001	0.0010	0.0059	0.0608	0.8261	0.0877	0.0099	0.0086
B	0.0001	0.0009	0.0021	0.0057	0.0440	0.7812	0.1102	0.0560
CCC	0.0000	0.0000	0.0004	0.0018	0.0049	0.0496	0.7284	0.2149
D	0	0	0	0	0	0	0	1.0000

Source: Lehman Brothers.