The One-Factor Gaussian Copula Applied To CDOs: Just Say NO (Or, If You See A Correlation Smile, She Is Laughing At Your "Results")

Arturo Cifuentes and Georgios Katsaros

Structured Finance Department
R.W. Pressprich & Co.
New York, NY, USA

SUMMARY

The one-factor Gaussian copula method has become the *de facto* standard to analyze most synthetic collateralized debt obligation structures. Unfortunately, this method produces a peculiar phenomenon known as correlation smile (the implied correlation determined by the model depends on the CDO tranche one is considering instead of being tranche-independent). Market participants are divided regarding this issue. Many suspect that the correlation smile is caused by a flaw in the above-mentioned modeling strategy although they have been unable to articulate why. Others insist that the smile is actually correct and reveals important and relevant tranche-dependent characteristics but have failed to produce convincing evidence to support this view.

In this article we present evidence that the correlation smile is really a by-product (artifact) of an unfortunate modeling strategy and has no financial or market-driven interpretation whatsoever. Moreover, we argue that this modeling approach should be abandoned at once.

INTRODUCTION

Collateralized debt obligations (CDOs) in general and synthetic CDOs in particular, specifically bespoke tranches and index tranches such as those of the North American high yield 5-year index (DJ CDX NA HY 5) or the North American investment grade 5-year index (DJ CDX NA IG 5), have become permanent fixtures of the fixed income investment universe.

Therefore, the need to have a sound modeling strategy to assess the risk-reward characteristics of these investments is paramount.

The performance of these vehicles (and to be more precise, the performance of the specific tranche that one is concerned with) depends chiefly on one factor: the credit risk behavior of the underlying assets. Therefore, to assess this risk (a task that one has to tackle using simulation techniques for closed-form solutions seem impossible at this point) the ability to generate realistic default scenarios is essential. To this end, two variables are critical: the default characteristics of the pool of assets (default probability) and the default correlation of such assets. A failure to characterize these variables correctly would be tantamount to building a high-rise on a shaky foundation: only tragedy could ensue.

The remaining of this article is organized as follows: first, we introduce some basic notation and review (for the sake of completion) the key features of the Gaussian copula. Then, we explain briefly the correlation smile problem and present some of the explanations given for it. Afterwards, we discuss in detail the one-factor Gaussian copula as it is currently employed in the context of CDOs and we explain why there are serious flaws with this approach. We end the article with a discussion regarding the implications of the present situation.

PRELIMINARY BACKGROUND

Probability of Default. Consider an asset that is subject to default risk. That is, it can default with a probability p or perform (not default) with a probability $1 - p$. Thus, from a modeling viewpoint, we can simulate the default behavior of this asset by taking successive samples (x) from a univariate normal distribution and comparing x with the appropriate cutoff value, X^* (X^* is such that $\Phi(X^*) = p$, where Φ is the cumulative normal distribution function). For instance, if $X \sim N(0, 1)$ and $p = 30\%$, then a value of $x < X^* = -0.55$ (or $\Phi(x) < 30\%$) indicates a default (see Figure 1).

Default Correlation. Correlation in general, and default correlation in particular, is one of the most misused and misunderstood concepts in structured finance. Thus, before we proceed, we will introduce some formalities for the sake of clarity. First of all, correlation is a precise mathematical concept that only has meaning in reference to random variables. If A and B are random variables, the correlation coefficient, ρ, between A and B is defined as

$$\rho = E[(A-\mu_A)(B-\mu_B)]/(\sigma_A\sigma_B) \tag{1}$$

where μ denotes the mean, σ denotes the standard deviation, and E is the expected value operator. The sub-indexes A and B refer to the random variables.

Thus, statements such as *company X and company Y are correlated* are meaningless. One needs to refer to a quantifiable variable associated with companies X and Y (stock price, revenue growth or credit default spreads, for example) for that statement to make sense.

In the same spirit, stating that *the default behavior of companies 1 and 2 is correlated* does not carry a lot of meaning unless one specifies a random variable that captures what “default behavior” means. In summary, before we can talk about default correlation, we need to define a random variable that somehow captures “default behavior.”

Accordingly, we introduce an index random variable, I, for this purpose. If the asset defaults, then $I = 1$ (otherwise the index is 0).

Two Assets. Suppose we have two assets: 1 and 2. Hence, we can use two random variables Y_1 and Y_2, both univariate normals, to generate default scenarios for assets 1 and 2. If we generate n possible default scenarios we can define two index variables (I_1 and I_2) to capture the default patterns of each asset. (Each index variable, actually a vector, will be a sequence of n *1's* and *0's*).

Using a random sample generated as above we can compute an estimate of the statistic “correlation coefficient” between I_1 and I_2 in order to estimate the default correlation (ρ_D) between assets 1 and 2.

$$\rho_D = \frac{n\sum(v_i w_i) - \sum v_i \sum w_i}{\sqrt{\left(n\sum v_i^2 - \left(\sum v_i\right)^2\right)\left(n\sum w_i^2 - \left(\sum w_i\right)^2\right)}} \tag{2}$$

In this context $v_i = (I_1)_i$ and $w_i = (I_2)_i$. Equation (2) is just a "discrete" version of equation (1).

The Simulation Challenge. We can assume, for the sake of simplicity but without losing generality, that we are dealing with a portfolio in which all the assets have the same default probability, *p*. (This is indeed the case when one deals with index tranches.)

The challenge to do a Monte Carlo relies on the modeler's ability to generate realistic default scenarios. That is, being able to be loyal to *p* (the average default probability of the pool) and ρ_D (the assets' default correlation.) Note that if we are dealing with two assets, the default correlation is captured by one number. In the case of *N* assets, ρ_D is a symmetric matrix. In the simplest case, that is, when all the pair-wise correlations are the same, all the off-diagonals elements are identical.

The One-Factor Gaussian Copula. The one-factor Gaussian copula, leaving CDOs aside for a moment, is nothing but a numerical algorithm to generate samples of normally distributed random variables that have a given pair-wise correlation. This method can be summarized as follows:

[1] Let $Z_1, Z_2, \ldots, Z_M$ be *M* independent random variables, each distributed as $N(0, 1)$

[2] Define random variables $Y_1, Y_2, \ldots, Y_M$ as

$$Y_i = R\sqrt{\rho} + Z_i\sqrt{1-\rho} \quad where \quad i = 1, 2, \ldots, M \tag{3}$$

and *R* is $N(0, 1)$, independent of all Z_i.

By repeatedly drawing instance vectors $(z_1, \ldots, z_M, r)$ of the random variables $Z_1, \ldots, Z_M, R$ we can generate several instance vectors of the form $(y_1, \ldots, y_M)$ with the desired correlation (ρ).

THE STANDARD "ONE-FACTOR GAUSSIAN COPULA" MODELING APPROACH

The quotes in this section's title are *very* intentional. In reality, what is called one–factor Gaussian copula approach in the context of CDO modeling is a bit misleading. In fairness, the method employed (to be described more fully shortly) consists of two steps: (1) the application of the Gaussian copula followed by (2) the creation of an additional random variable (an index variable). As we will see later, this second step creates some problems.

Before describing this approach an additional clarification is required. In principle (as stated before) to characterize properly the credit risk behavior of a pool of assets we need to estimate (and control) two variables: p (default probability) and ρ_D (default correlation).

The default probability is relatively easy to estimate as we can rely on ratings, CDS spreads, fundamental analyses, KMV-like models, etc. The problem is ρ_D. First, there are very little data on default correlation (after all, defaults do not happen that often). Second, the data seem to indicate that this correlation changes over time (it is time-dependent). Worse yet, nobody has proposed an algorithm to generate default scenarios (assuming one knows p and ρ_D) that are consistent with p and ρ_D.

Therefore, most practitioners have given up on estimating default correlations and have concentrated on correlation exhibited by asset prices (normally called asset correlation) since such data abound. The secret hope, of course, is that *asset correlation* has, somehow, something to do with *default correlation*. Asset correlation is denoted as ρ_A.

Consider, again for simplicity, that we have two assets (the extension to many is straightforward) and assume that we know both p (the same for assets 1 and 2) and ρ_A (since we have given up on ρ_D). How do we generate default scenarios "consistent" with p and ρ_A?

The standard one-factor Gaussian copula approach extends the two-step process described in the previous section with a third step that computes the indicators of non-default or default events for each asset. Here is the complete description:

[1] Let R, Z_1, Z_2 be independent random variables, each distributed as $N(0, 1)$

[2] Define random variables Y_1, Y_2 as

$$Y_1 = R\sqrt{\rho_A} + Z_1\sqrt{1-\rho_A}$$
$$Y_2 = R\sqrt{\rho_A} + Z_2\sqrt{1-\rho_A} \qquad (4)$$

[3] Define index random variables I_1 and I_2 with values in $\{0, 1\}$ as

$$I_1 = \begin{cases} 1, & if \;\; \Phi(y_1) \le p, \quad (or \quad alternatively \quad y_1 < Y^*) \\ 0, & otherwise \end{cases}$$
$$I_2 = \begin{cases} 1, & if \;\; \Phi(y_2) \le p, \quad (or \quad alternatively \quad y_2 < Y^*) \\ 0, & otherwise \end{cases} \qquad (5)$$

$\Phi(\cdot)$ is the standard normal cumulative distribution function (note that $\Phi(Y^*) = p$); y_1 and y_2 represent random draws from Y_1 and Y_2, respectively. For example, the event {Asset 1 defaults, Asset 2 does not default) can be expressed as: $\{I_1 = 1, I_2 = 0\}$.

After repeating steps [1], [2] and [3] many times (say L) one has enough default scenarios to calculate whatever is required for the tranche under investigation: price, appropriate spread over LIBOR, estimated rating, default likelihood, etc.

Additionally, one ends up with two index-vectors having L components each (a sequence of *L 1's* and *0's*) that can be used to estimate ρ_D using expression (2). This process is demonstrated more graphically in Figure 2.

Note that up to this point we have no idea what ρ_D is; the only thing we know is that the normal samples we generated in [1] and [2] to simulate the defaults have a given (ρ_A) correlation.

Three observations are relevant at this point. First, although this might seem rather innocuous, calling this approach one-factor Gaussian copula does not seem entirely accurate for the one-factor Gaussian copula approach relates only to steps [1] and [2] above. Step [3], which is an integral and quite important part of the process, has nothing to do with the Gaussian copula.

Second, and this is rather subtle, denoting (as most people do) the Gaussian copula correlation factor ρ_A seems like an *obvious* choice. But is this the *actual* asset (market value) correlation between assets 1 and 2? (Leave aside for a second the issue of whether we have a good estimate of the asset correlation). The truth is, that this is really a Gaussian copula correlation parameter that has nothing to do with the value of any asset. We believe this confusion has arisen due to the common use of Merton's method to estimate the default probability of a company. If one does not know p, Merton's method states, one can estimate it by calculating the likelihood that the market value of the assets of the company will drop below a critical level (value of liabilities) [1]. Such market (asset) value is supposed to follow a normal distribution.

But in the context of the CDO problem at hand we are already assuming that we know p. Thus, the only thing we are doing in Steps [2] and [3] is using a numerical trick to generate default scenarios consistent with such default probability; scenarios which, by the way, just happened to have been generated by normal random variables that have a given correlation coefficient. Whether we have the right to call this parameter asset correlation is far from clear. Thus, in what follows, we will call this parameter "copula correlation factor", or ρ_C.

In any event, the two preceding points are quite irrelevant in comparison with this third observation: What is the relationship between the copula correlation factor (ρ_C) and the pool default correlation (ρ_D). Or, put in a different way: if we choose the copula correlation factor to be some fixed value, what is the value of the default correlation factor we end up with? This is really the critical question at the root of this modeling approach. It arises because what we are controlling here is ρ_C (a variable we think we can estimate and we can use as the driving force for the Gaussian copula algorithm) but we have no control (or clue) as to what the value of ρ_D (the variable of interest) will be. We will return to this point in the subsequent section.

THE CORRELATION SMILE ISSUE

The correlation smile issue can be best described in the context of a CDO structure that has several tranches, from very *safe* to very *risky*, (typically AAA to Equity). Suppose we know the market price of each tranche. Since we know the assets (pool) default probability (p) we can use the one-factor Gaussian copula approach to determine the *implied* correlation. In other words, we work backwards (iterating several times) until we determine the copula correlation factor (ρ_C) such that we can match the tranche price while keeping p fixed. Obviously, p is the same no matter what tranche we are looking at since it is a characteristic of the underlying collateral (not the tranche). Unfortunately, when we do this, we obtain different implied correlation values depending on the tranche we are considering. Figure 3 displays the typical curve, the correlation smile curve. Several explanations have been offered for this "anomaly".[1]

Broadly speaking, market participants are divided in terms of this issue. A small minority believes the correlation smile is a non-issue (annoying but irrelevant). Another group thinks it is "obviously" wrong but cannot pin down the reason for it or they blame it on the normal distribution (tails not fat enough). A third group thinks the correlation smile (different correlation numbers for different tranches) makes sense for it captures demand and supply circumstances, liquidity issues, segmentation among investors, model uncertainty, banks' appetite for selling protection only on certain tranches, and a host of other factors [2, 3, 4, 5].

That said, it is fair to say that this issue has not yet received an explanation that has been fully accepted.

THE GAUSSIAN COPULA DECONSTRUCTED

The best way to tackle the correlation smile issue is by investigating the relationship among the three variables involved: p, ρ_C and ρ_D. For given values for p and ρ_C it is possible to perform a Monte Carlo simulation in order to determine ρ_D. In fact, by covering all possible values of the pair (p, ρ_C) in the range $(0, 1)$ x $(0, 1)$ and then computing the resulting value of ρ_D

[1] Strictly speaking, correlation smile is a special case of correlation skew. From a "purist" point of view correlation skew refers to the fact that the correlation is non-constant (the phenomenon just described in the text) while the "smile" refers to the particular case in which there is a change in slope (from negative to positive). Here, we just use correlation smile as the generic term to refer to the broader anomaly.

via a Monte Carlo/one-factor Gaussian copula method, one can gain great insight into these variables' behavior[2]. Figure 4 depicts a three-dimensional view of these variables' relationship. An immediate observation is that for a given (fixed) value of the default probability, the default correlation ρ_D becomes an increasing function of the copula correlation factor ρ_C. There also appears to be a symmetry around $p = 50\%$, when fixing the value for ρ_C.

A better appreciation for how these variables are interconnected is obtained by looking at the data depicted in Figure 4 in a more selective fashion. To start, consider Figure 5 which shows, for a fixed value of p (arbitrarily chosen at 10%), the relationship between ρ_C and ρ_D. The obvious observation is that for any given value of ρ_C, the value of ρ_D is always less than ρ_C. Clearly, the thought that selecting one specific value for the copula correlation parameter might result in the same value for the pool default correlation (ρ_D) is unwarranted.

Figure 6, again, shows the relationship between ρ_C and ρ_D but for several values of p (10%, 20% and 30%). The fact that the (ρ_C, ρ_D) relationship depends on p (three curves instead of one) is a bit "distressing." Put in a different way, this figure indicates that if we select a value for the copula correlation factor (ρ_C), the value of the default correlation (ρ_D) we end up with, depends on p. This hints at a potential problem. It means that we cannot control the default correlation via the copula correlation factor only (we also need to "control" p). Intuitively, it is clear why the default correlation (ρ_D) is affected by the value of p. This happens in Step [3] of the one-factor Gaussian copula process, when one forms the index variables, which in turn determine the default correlation. (Recall Figure 2). The cutoff value (Y^*), the one factor that determines whether the index is *1* or *0*, depends on p. Incidentally, conventional correlation assumptions state that correlation is a function of either geographical area or industrial sector, but not default probability. The fact that ρ_D depends on both, p and ρ_C, indicates that the one-factor Gaussian copula does not lend itself (at least not *naturally*) to be used in conjunction with these conventional correlation assumptions. Unless, of course, one is willing to accept some strange effects.

[2] As an alternative to the Monte Carlo simulation, one could obtain a semi-analytic expression for ρ_D as a function of the default probability and the copula correlation factor, to examine the relationship among the three variables. This option is discussed in the Appendix.

In addition, the “distortion” introduced in Step [3] when creating the index variables not only depends on p, but it does so in a fairly special way: the “distortion” is “symmetric” with respect to $p = 50\%$. This is fairly obvious just by looking at Step [3] in Figure 2: The results of two binomial processes with probabilities of occurrence q and $1-q$ are essentially mirror images of each other. To illustrate, consider modeling a series of coin flips for two identical coins, where we define the event {Coin 1 is Heads } as 1 for the first coin, and {Coin 2 is Tails } as 1 for the second coin. The two series or 0 and 1 events are, from a statistical point of view, identical. Figure 7 illustrates this point: the (ρ_C, ρ_D) relationship is the same for $p=10\%$ and $p=90\%$ (identical curves).

Figure 8, again, illustrates another interesting point: it shows (like Figure 6) the value of ρ_D as a function of ρ_C for different values of p. Three curves (10%, 20% and 30%) are the same curves depicted in Figure 6; but there is a fourth curve in the graphic ($p=98\%$). The first three curves (10%, 20% and 30%) show that for increasing values of p, as we move in the 0% to 50% direction, the curves move upwards. Then, consistent with the “symmetry with respect to $p=50\%$” situation the order is reverse; the $p=98\%$ curve has moved downward.

Finally, Figure 9 illustrates the relationship between the copula correlation factor (ρ_C) and the default probability. For a fixed default correlation, the copula correlation factor is a convex function of the default probability. Three such iso-contours are displayed in the figure, for three corresponding values of the default correlation. Again, the symmetry of the p-dependence is manifest.

In summary, these graphs show that the relationship among p, ρ_C and ρ_D is not that simple. First, they show that the only way to make sure we end up with a given (and presumably correct) default correlation value is that we select a copula correlation factor (ρ_C) that is not only a function of ρ_D but also a function of p. Second, failure to do this, that is, choosing the same ρ_C regardless of the value of p, can certainly have undesirable effects. And third, it is not clear that working with a functional relationship that imposes on ρ_D (or ρ_C) a symmetry with respect to $p=50\%$ has any realistic basis (or advantage). Although these features by themselves are not sufficient to pass an indictment of the one-factor Gaussian copula approach, they certainly raise some concern as to the suitability of this technique to generate realistic default scenarios.

<u>THE GAUSSIAN COPULA: THE FINAL BLOW</u>

The best way to detect the flaws in the one-factor Gaussian copula approach is by *reductio ad absurdum*. Specifically, we can prove that the correlation smile phenomenon makes no sense (it does not capture any meaningful attribute of the tranches in question and is at odds with reality).

Suppose we have several CDO tranches and further assume that we know the price (given by the market) for each tranche. Invoking the Gaussian copula and noting that p is known we can estimate the "implied correlation value" (ρ_C) associated with each tranche: 1, 2, 3…. Let us denote these values as $(\rho_C)_1$, $(\rho_C)_2$, $(\rho_C)_3$ …. Obviously, these correlation values are all different (correlation smile). Can this be true? Does this make sense? Let us assume that this is indeed the case and let us see where this assumption will take us.

We can invoke now the relationship between ρ_C and ρ_D established in the previous section (a relationship which for a given p will look like the curve shown in Figure 5) and we determine for each value $(\rho_C)_1$, $(\rho_C)_2$, $(\rho_C)_3$ …. the corresponding value of the default correlation $(\rho_D)_1$, $(\rho_D)_2$, $(\rho_D)_3$ …. Clearly, these values ($(\rho_D)_1$, $(\rho_D)_2$, $(\rho_D)_3$ …) will all be different since the implied correlation values ($(\rho_C)_1$, $(\rho_C)_2$, $(\rho_C)_3$ ….) are all different. This would imply that the pool default correlation does not depend on the characteristics of the pool only, it would also depend <u>on the CDO tranches</u>. Or, alternatively, it would imply that the collateral pool has "several" (and correct) default correlation values. This is clearly nonsense for <u>this value has to be unique</u>. Thus, the hypothesis is false. (Unless, of course, one is willing to accept that the market price of the tranches is wrong; but we are assuming that that is not the case). *Ergo,* the Gaussian copula leads to unreasonable conclusions. In a way, the clarity with which one can see the fallacy is due to the relationship presented in Figure 4. This relationship has allowed us to move the discussion from the liabilities side of the CDO (where arguments about the uniqueness of ρ_C or ρ_A might be less obvious) to the asset side of the CDO (where the nonsense associated with having several values for ρ_D becomes indefensible).

DISCUSSION

In the preceding sections we have established that the Gaussian copula leads to results that are at odds with reality. At the root of this incompatibility is the relationship that this method imposes on p, ρ_C and ρ_D, a relationship from which we cannot escape because is part of the intrinsic structure of the copula. Hence, there is no cure for this: the method must be abandoned (cannot be repaired).

In addition, the following observations can be made:

- Some practitioners have voiced the view that the Gaussian copula/ correlation smile issue is overblown because most CDO-tranches (at least in the secondary market) are priced according to supply-demand dynamics and the market "more or less" knows what the correct price is. In other words, who cares about this model? Fair enough. But the fact remains that many investors do analyze newly issued CDO-tranches with this method. Thus, the need to discuss the method (and the need to question the accuracy of the results produced with it) should not be dismissed.

- When it comes to delta hedging (the technique employed to manage the risk in a portfolio of index tranches) the role of the one-factor Gaussian copula is very relevant. In fact, it is the most popular method to determine the so-called hedge ratios. This is crucial for it means that the effectiveness of such hedging strategy, particularly at times of market turmoil (May 2005 comes to mind), could be ineffective. This is anything but trivial.

- Additionally, the rating agencies are increasingly relying on this method to rate CDOs, especially synthetic deals. Although most people are becoming dismissive of ratings and the market is certainly moving to ignoring them, one fact remains: certain investors are subject to ratings-driven requirements in terms of what they can or cannot buy; many SIVs (specialized investment vehicles) are required to liquidate assets whose ratings fall below certain level; and many institutions are subjected to capital requirements that are ratings-driven. Thus, to the extent that ratings are determined with a flawed method, the implications are quite serious. Incidentally, the ratings system is already showing some cracks: for example, in March-April

2007, we have seen newly issued CDO-tranches, with the same rating (BBB/Baa), priced as wide as LIBOR + 1000 and as tight as LIBOR + 120. This is unheard of.

- A byproduct of the correlation smile issue is the so-called *base* correlation. Base correlation is nothing but a cheap numerical trick to re-compute correlation using a different formula with the hope that things will look better. And they do. The base correlation *versus* tranche-rating curve looks just bad instead of terrible: it is not constant (parallel to the x-axis) but at least is monotonic (no change in slope sign). But not much comfort should be derived from this alternative definition no matter what traders say.

- Consider the following thought experiment. Assume we have one asset and also assume that we have a very good estimate of the default probability of this asset. In other words, we have incorporated into this estimate all the information available, including the influence of macro-economic factors, performance of other companies and political events, etc. Suppose now we do the same for several (M) assets: thus, we have $q_1, \ldots, q_M$ all excellent estimates of the default probability of these assets. Clearly, this situation will allow us to make future default projections. Now put these assets into an SPV. Would that change the default probability estimates? (Or, to be more clear, would that change the default probability of these assets?) The answer is a resounding NO (the assets do not realize they are in an SPV; they will behave the same as before for their risk profile is still the same, it does not change as a result of putting them in the SPV). Therefore, there is no need for correlation. In fact, correlation does not exist. What does exist is the need to "correct" our estimate of the default probability of an asset after we put it in an SPV if the initial estimate did not take into account all the relevant factors. Hence, in this case, we throw in "correlation" simply to "improve" an initially bad estimate. In reality, instead of spending time trying to estimate correlation (something that does NOT exist) it would be better to put more effort into estimating default probabilities more accurately. In this sense, a modeling approach assuming no correlation but time-dependent default probabilities (something along the lines of the CreditRisk+ framework [6], for example) seems much more reasonable and promising.

Finally, a curious point: The most common argument that we have heard from people who disagree with us (that is, die-hard Gaussian copula defenders) is that: (i) everybody uses it, and (ii) it is very easy to use. As a matter of principle we endorse the view that those two arguments, in no particular order, are probably the worst possible justifications ever offered for anything: clear evidence that there is something rotten in Denmark! To sum up: the one-factor Gaussian copula method is a flawed technique to model something that does not exist -- two very good reasons to move on and leave all this correlation/copula nonsense behind. Future efforts should be focused on estimating default probabilities better. Period. End of story.

REFERENCES

[1] Robert C. Merton (1974), “On the Pricing of Corporate Debt: The Risk Structure of Interest Rates”, Journal of Finance, Vol. 29.

[2] Jeffery D. Amato and Jacob Gyntelberg (2005), “CDS index tranches and the pricing of credit risk correlations”, BIS Quarterly Review, March 2005.

[3] Dominique Guegan and Julien Houdain (2005), “Collateralized Debt Obligations pricing and factor models: a new methodology using normal inverse Gaussian distributions,” Note de Recherche IDHE-MORA, number 07-2005.

[4] Toby Daglish and Wei Li (2005), “Probability of default as a function of correlation: the problem of non-uniqueness,” available from www.defaultrisk.com.

[5] Julien Turc and Philippe Very (2004), “Pricing and hedging correlation products,” Societe Generale, Credit Research, available from www.defaultrisk.com.

[6] Tom Wilde (1997), “CreditRisk+: A Credit Risk Management framework,” Credit Suisse First Boston, available from www.defaultrisk.com.

[7] Christian Blum, Ludger Overbeck and Christoph Wagner (2003) “An Introduction to Credit Risk Modeling”, Chapman & Hall/CRC, Financial Mathematics Series.

[8] Philipp Schönbucher (2003) “Credit derivatives pricing models: Models, pricing and implementation”, Wiley Finance.

[9] Umberto Cherubini, Elisa Luciano and Walter Vecchiato (2004), “Copula Methods in Finance”, John Wiley & Sons, Ltd.

LIST OF FIGURES

APPENDIX

A Semi-Analytic Derivation of ρ_D

There exists an extensive literature on modeling correlated defaults (see [7, 8] for an introductory treatment and numerous references.) In this section we outline the derivation of the linear correlation of default events in the simplest case of two assets and in the context of the one-factor Gaussian copula described previously but using a semi-analytic approach. The idea is to compare the results obtained for ρ_D using this approach against those obtained with the Monte Carlo method. We will see that both approaches give consistent values for ρ_D.

As in the Bernoulli mixture model developed in [7], we treat the default probabilities p_1, p_2 of the two assets as instances of random variables P_1, P_2 defined in $[0, 1]$. In the context of the previous discussion, where we assumed $p_1 = p_2$, the default events are in turn determined by the value of P_j, if we consider Y_j^* a known, fixed "average" value such that $\bar{p}_j = \Phi\left(Y_j^*\right)$:

$$I_j = \begin{Bmatrix} 1, & if \;\; \Phi(y_j) \le \bar{p}_j, \\ 0, & otherwise \end{Bmatrix}, j = 1,2 \qquad (6)$$

The distribution $F(u_1, u_2)$ of P_1, P_2 can be derived from the conditional probabilities of default events given the value of the common factor R and equation (4), as follows:

$$\begin{aligned}
\Pr\{I_j = 1 | R\} &= \Pr\left\{\Phi\left(y_j\right) \le \bar{p}_j | R\right\} \\
&= \Pr\left\{R\sqrt{\rho_C} + Z_j\sqrt{1-\rho_C} \le \Phi^{-1}\left(\bar{p}_j\right) | R\right\} \qquad (7) \\
&= \Pr\left\{Z_j \le \frac{\Phi^{-1}\left(\bar{p}_j\right) - R\sqrt{\rho_C}}{\sqrt{1-\rho_C}} | R\right\} \qquad j = 1,2
\end{aligned}$$

Z_j is distributed as $N(0,1)$, therefore the conditional probability can be written as a function of the value r of the common factor random variable R:

$$\Pr\{I_j = 1 | R = r\} \equiv q_j(r) = \Phi\left(\frac{\Phi^{-1}(\overline{p}_j) - r\sqrt{\rho_C}}{\sqrt{1-\rho_C}}\right) \qquad (8)$$

The distribution $F(u_1, u_2)$ of P_1, P_2 is the Gaussian copula evaluated at $q_j^{-1}(u_j)$ (see [7] and [9]):

$$F(u_1, u_2) \equiv \Phi_2\left(q_1^{-1}(u_1), q_2^{-1}(u_2); \sqrt{\rho_C}\right) \qquad (9)$$

where $q_j^{-1}(u) = \frac{\Phi^{-1}(\overline{p}_j) - \Phi^{-1}(u)\sqrt{1-\rho_C}}{\sqrt{\rho_C}}$ and $\Phi_2(x_1, x_2; \rho)$ is the bivariate standard normal distribution with correlation coefficient ρ.

The distribution of the default events I_1, I_2 can be defined as:

$$\Delta(d_1, d_2) \equiv \Pr\{I_1 = d_1, I_2 = d_2\}, \quad d_1, d_2 \text{ in } \{0,1\} \qquad (10)$$

The values of the distribution $\Delta(d_1, d_2)$ are determined, given values for $P_1=u_1$, $P_2=u_2$ as follows:

d_1	d_2	$\Delta(d_1, d_2)$
0	0	$(1-u_1)(1-u_2)$
0	1	$(1-u_1)u_2$
1	0	$u_1(1-u_2)$
1	1	u_1u_2

In summary:

$$\Delta(d_1, d_2 | u_1, u_2) = \left(u_1^{d_1}(1-u_1)^{1-d_1}\right) \cdot \left(u_2^{d_2}(1-u_2)^{1-d_2}\right) = \prod_{i=1,2} u_i^{d_i}(1-u_i)^{1-d_i} \qquad (11)$$

In turn, the unconditional distribution of I_1, I_2 given (9) and (11) is:

$$\Delta(d_1, d_2) = \int_0^1\int_0^1 \prod_{i=1,2} u_i^{d_i}(1-u_i)^{1-d_i}\, dF(u_1, u_2) \qquad (12)$$

The following outlines the computation of

$$\rho_D = \mathrm{Corr}(I_1, I_2) = \frac{\mathrm{Cov}(I_1, I_2)}{\sqrt{\mathrm{Var}(I_1)\mathrm{Var}(I_2)}}.$$

The marginal distributions $\Delta_1(d_1)$ and $\Delta_2(d_2)$ can be found as:

$$\Delta_1(d_1) = \sum_{d_2=1,2} \Delta(d_1, d_2) = \Delta(d_1, 0) + \Delta(d_1, 1) = \int_0^1\int_0^1 u_1^{d_1}(1-u_1)^{1-d_1}\, dF(u_1, u_2),$$

and similarly for $\Delta_2(d_2)$.

$$E(I_i) = \sum_{d_i=0,1} d_i \Delta_i(d_i) = \Delta_i(1) = \int_0^1\int_0^1 u_i\, dF(u_1, u_2) = E(P_i) \qquad (13)$$

$$E(I_i^{\,2}) = \sum_{d_i=0,1} d_i^{\,2} \Delta_i(d_i) = \Delta_i(1) = E(P_i) \qquad (14)$$

$$\mathrm{Var}(I_i) = E(I_i^{\,2}) - E(I_i)^2 = E(P_i) - E(P_i)^2 = E(P_i)(1 - E(P_i)) \qquad (15)$$

$$E(I_1 I_2) = \sum_{d_1=0,1}\sum_{d_2=0,1} d_1 d_2 \Delta(d_1, d_2) = \Delta(1,1) = \int_0^1\int_0^1 u_1 u_2\, dF(u_1, u_2) = E(P_1 P_2) \qquad (16)$$

Finally, the covariance of I_1, I_2 is determined through (13) and (16) as:

$$\mathrm{Cov}(I_1, I_2) = E(I_1 I_2) - E(I_1)E(I_2) = \mathrm{Cov}(P_1, P_2) \qquad (17)$$

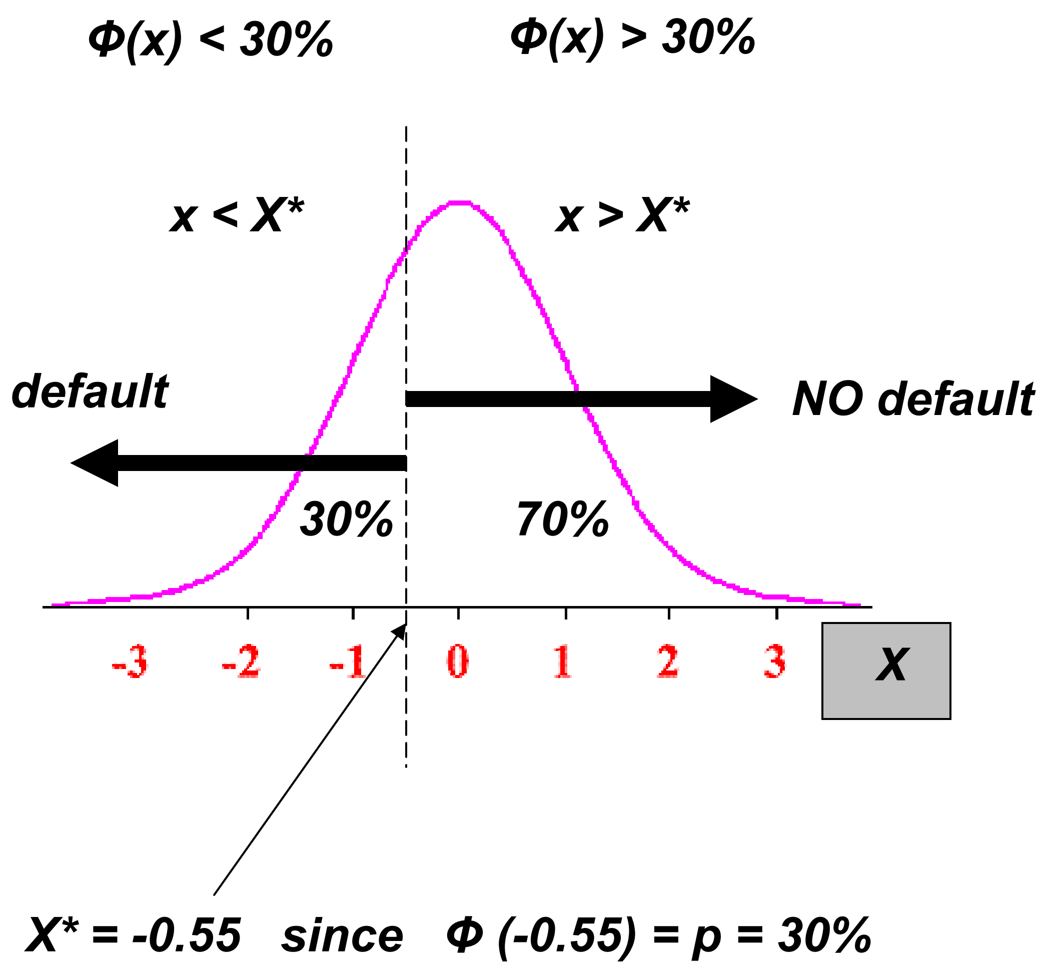
Equations (15) and (17) allow the expression of ρ_D in terms of $F(u_1, u_2)$ of P_1, P_2 as follows:

$$\rho_{\mathrm{D}} = \frac{\mathrm{Cov}(P_1, P_2)}{\sqrt{E(P_1)(1 - E(P_1))}\sqrt{E(P_2)(1 - E(P_2))}} \qquad (18)$$

Equations (9) and (18) can be used to compute ρ_D as a function of $\rho_C, \bar{p}_1, \bar{p}_2$.

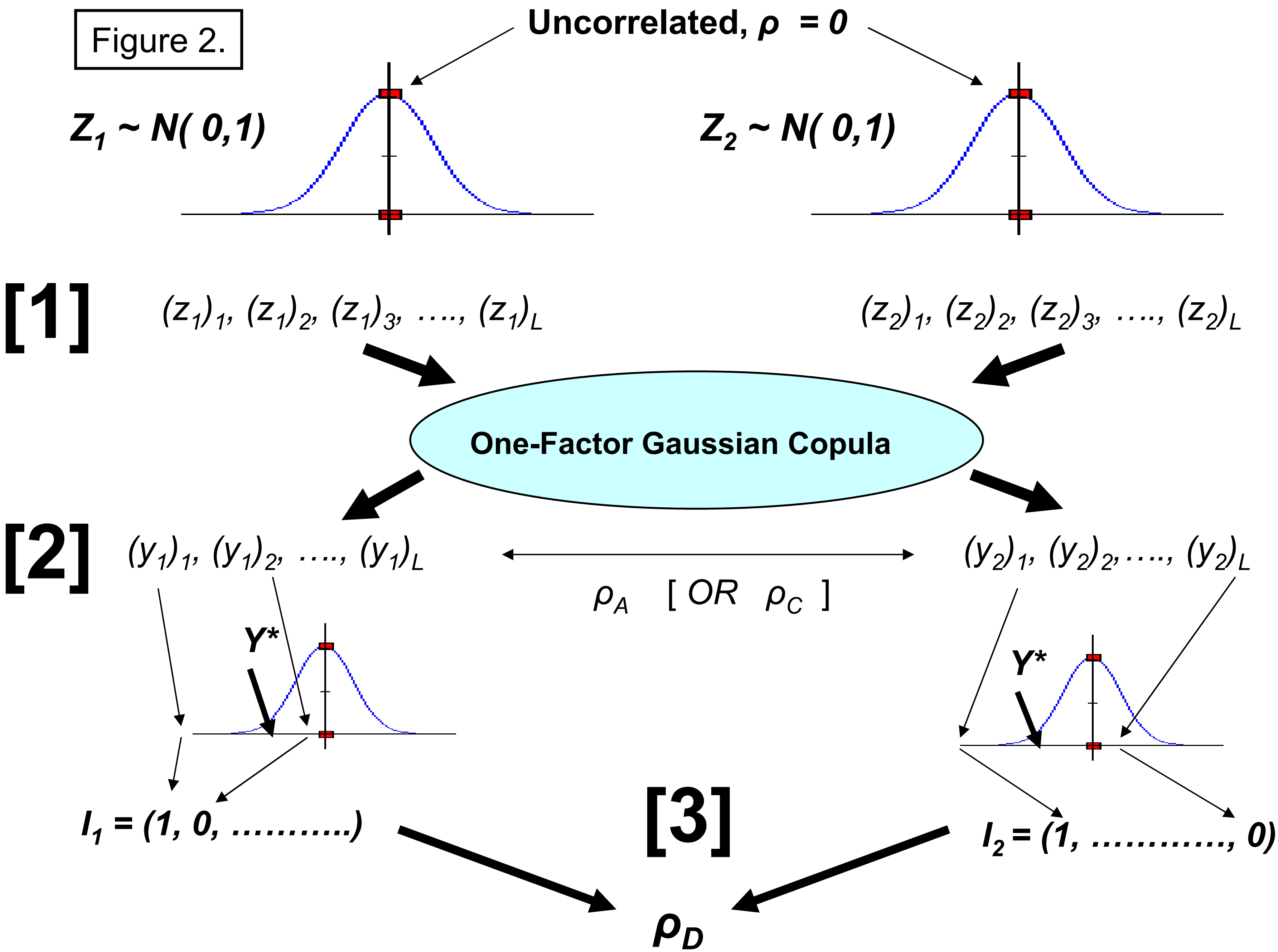
Figure 10 shows a comparison between the values of ρ_D (default correlation) obtained with: (i) the Monte Carlo method; and (ii) the semi-analytic approach outlined in this appendix and a range of the average default probability $p = \bar{p}_1 = \bar{p}_2$. We have restricted the comparison for values of ρ_C < 30% (higher values of ρ_C result in numerically unstable scenarios that are computationally expensive). We note that within the tested range both approaches produce consistent results.

Figure 1.



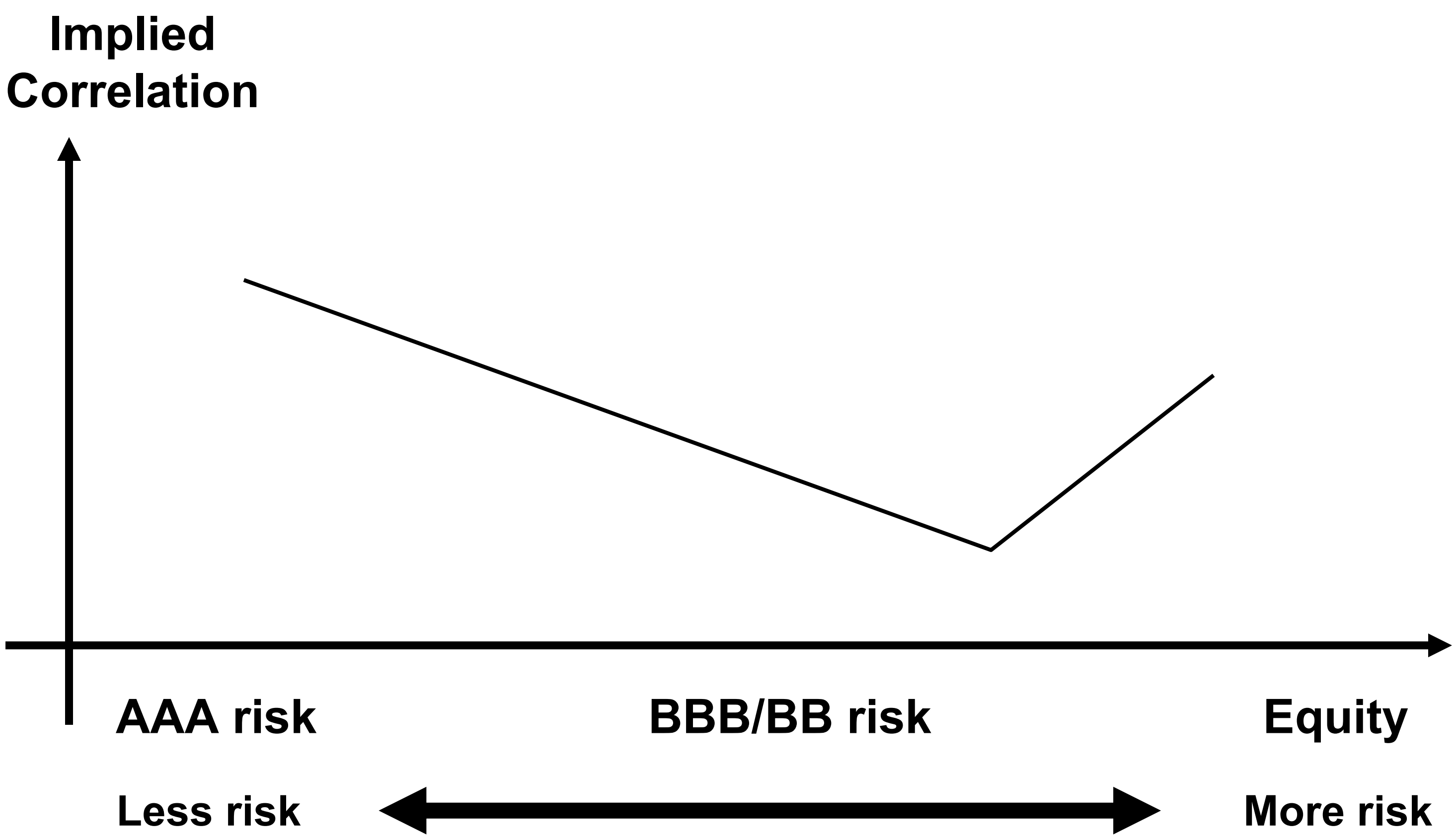
Φ(x) < 30%
Φ(x) > 30%
x < X*
x > X*
default
NO default
30%
70%
-3
-2
-1
0
1
2
3
X
X* = -0.55 since Φ (-0.55) = p = 30%

Figure 2.



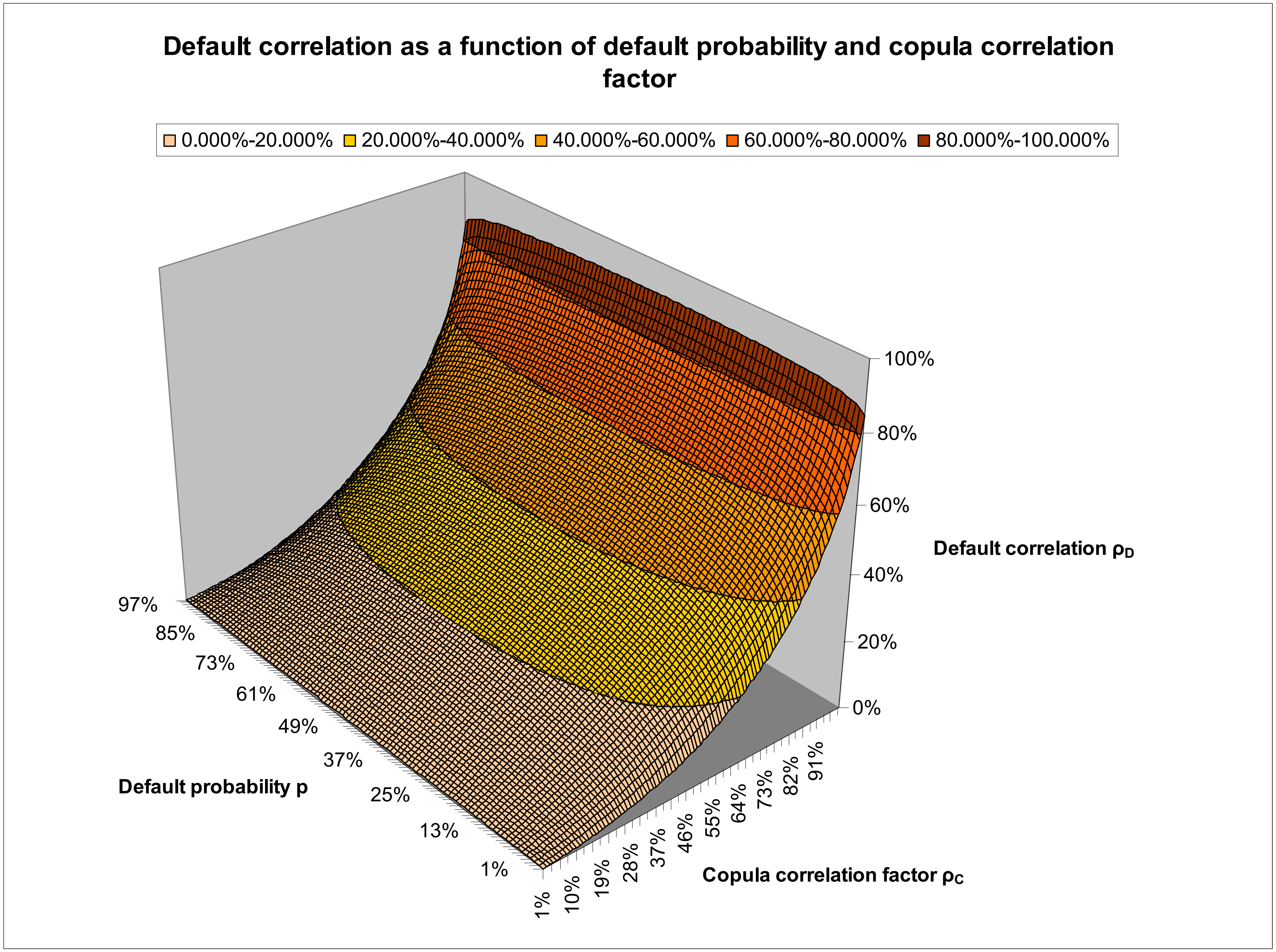
Uncorrelated, ρ = 0
Z1 ~ N(0,1)
Z2 ~ N(0,1)
[1]
(z1)1, (z1)2, (z1)3, …., (z1)L
(z2)1, (z2)2, (z2)3, …., (z2)L
One-Factor Gaussian Copula
[2]
(y1)1, (y1)2, …., (y1)L
(y2)1, (y2)2, …., (y2)L
ρA [OR ρC]
Y*
Y*
I1 = (1, 0, ……….)
[3]
I2 = (1, ………., 0)
ρD

Figure 3.



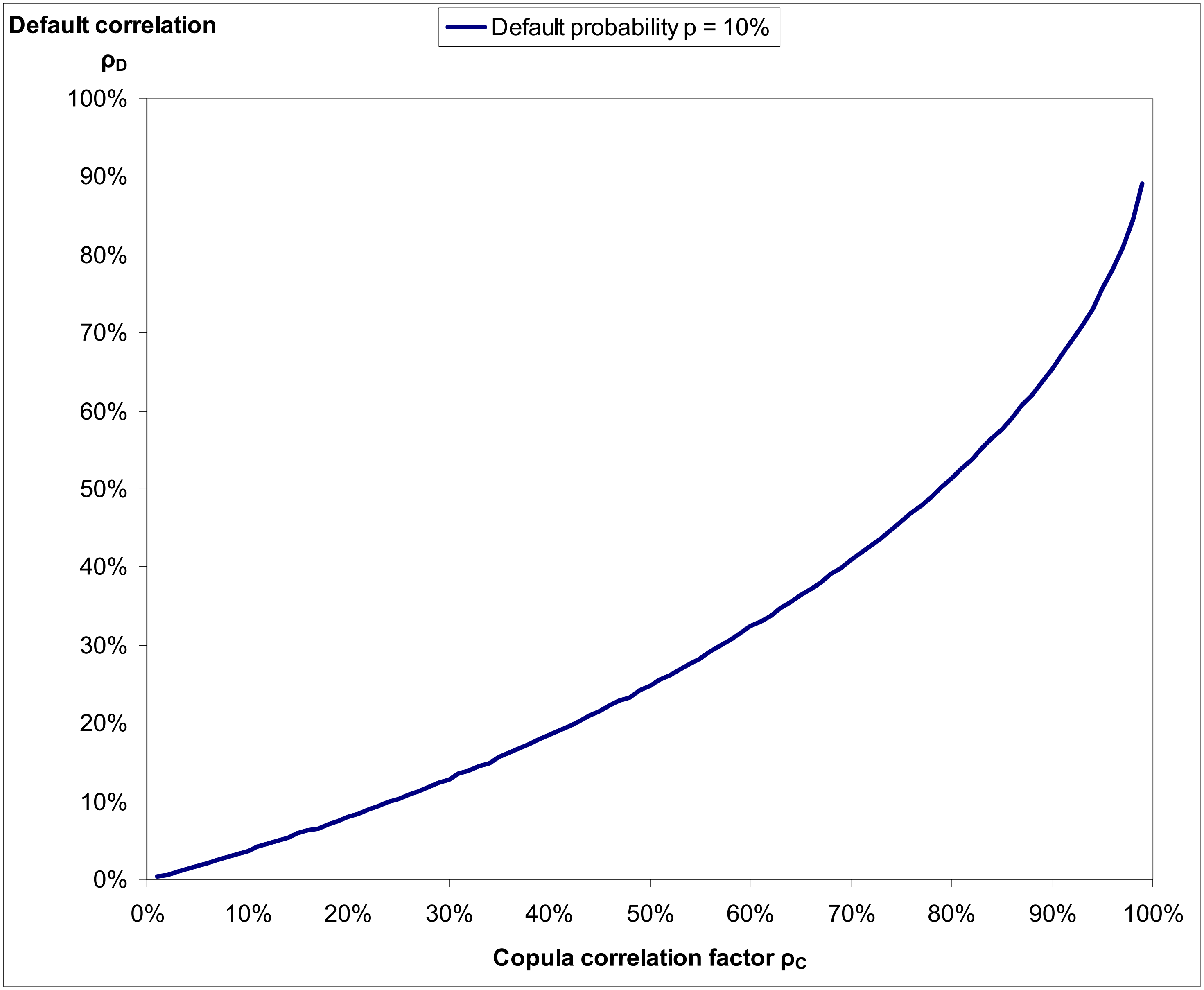
Implied Correlation
AAA risk
BBB/BB risk
Equity
Less risk
More risk

Figure 4.



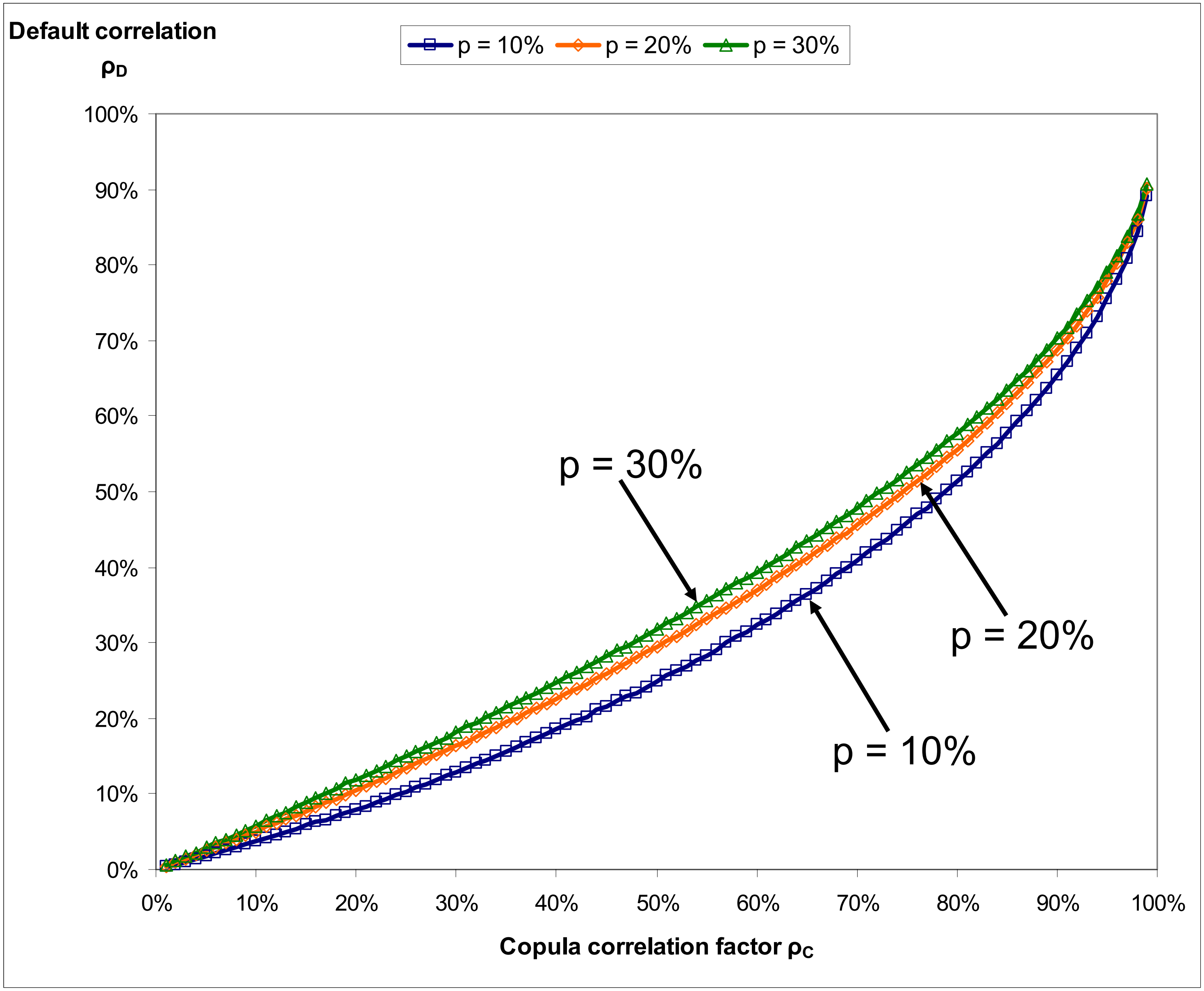
Default correlation as a function of default probability and copula correlation factor
0.000%-20.000%
20.000%-40.000%
40.000%-60.000%
60.000%-80.000%
80.000%-100.000%
100%
80%
60%
40%
20%
0%
Default correlation ρD
97%
85%
73%
61%
49%
37%
25%
13%
1%
Default probability p
1%
10%
19%
28%
37%
46%
55%
64%
73%
82%
91%
Copula correlation factor ρC

Figure 5.



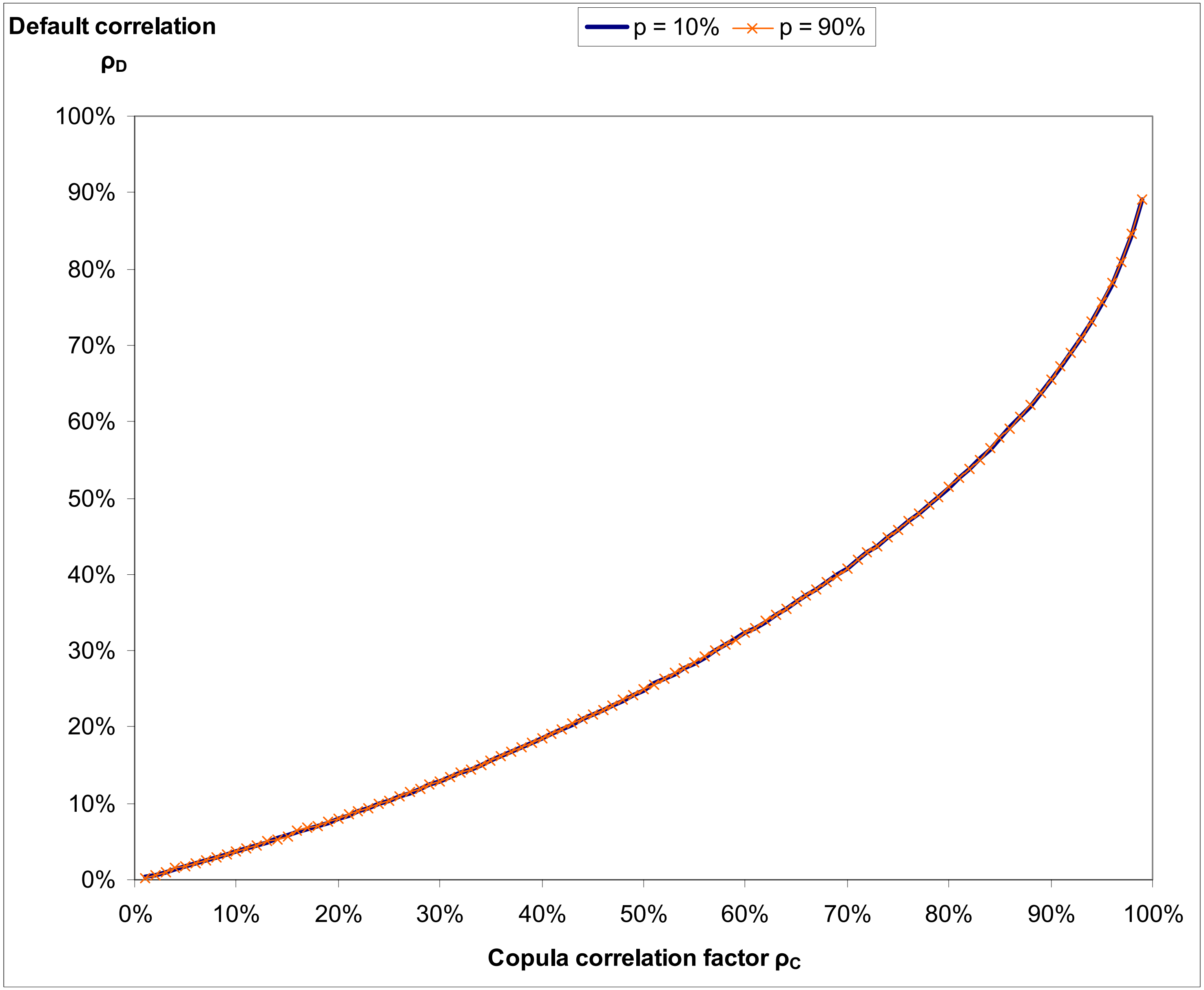
Default correlation
ρD
Default probability p = 10%
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Copula correlation factor ρC

Figure 6.



Default correlation
ρD
p = 10%
p = 20%
p = 30%
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Copula correlation factor ρC
p = 30%
p = 20%
p = 10%

Figure 7.



Default correlation
ρ_D
p = 10%
p = 90%
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Copula correlation factor ρ_C

Figure 8.

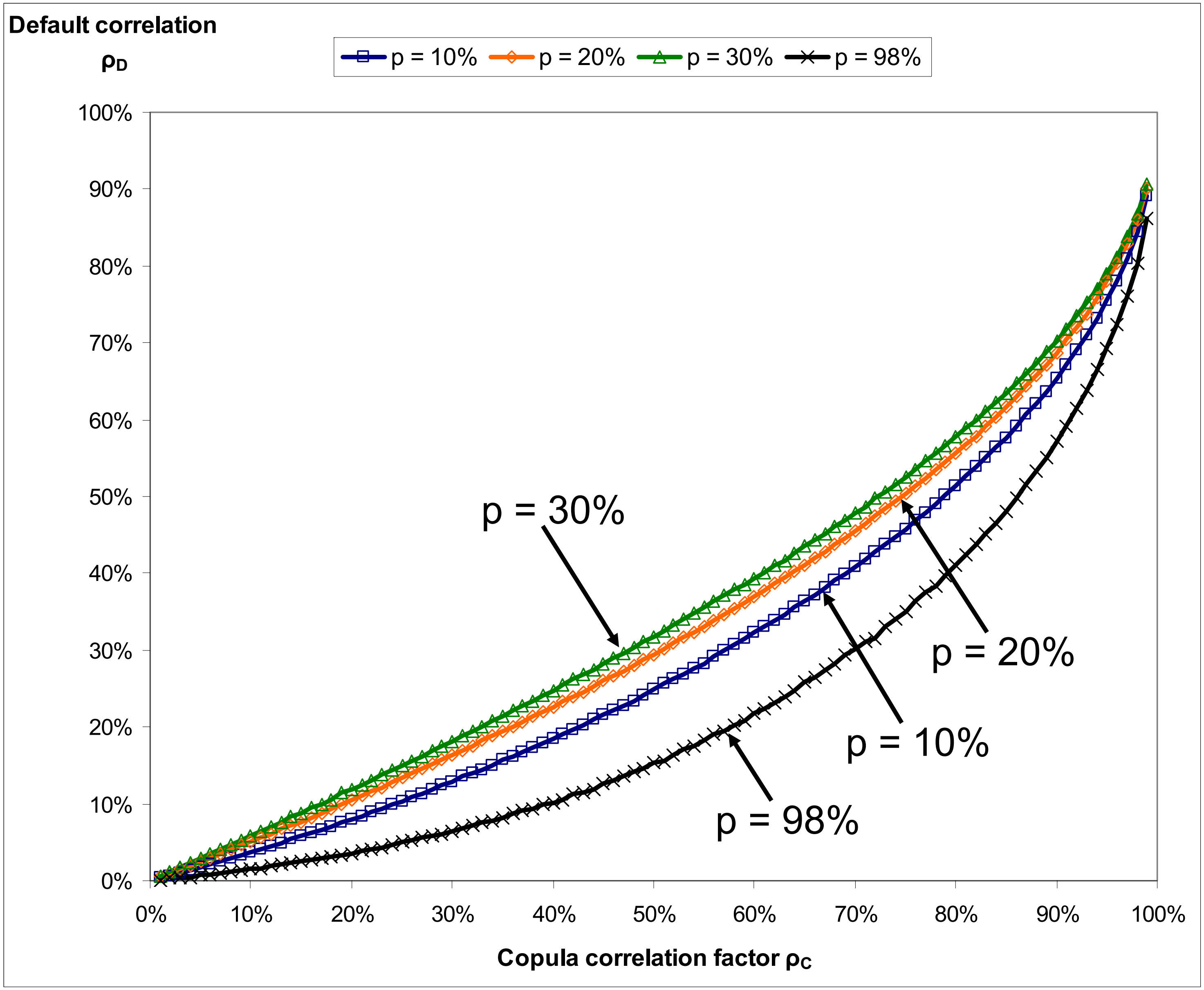
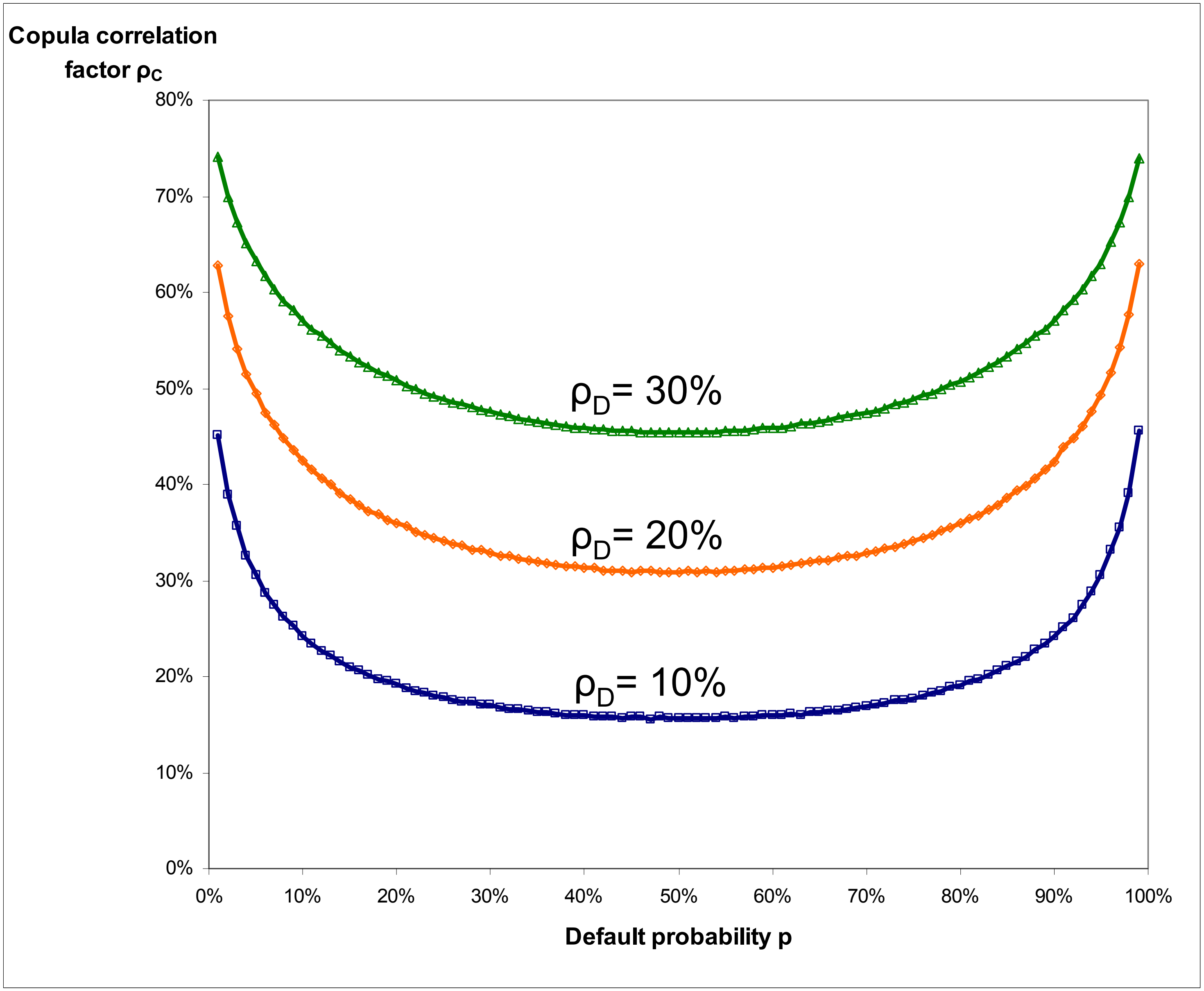
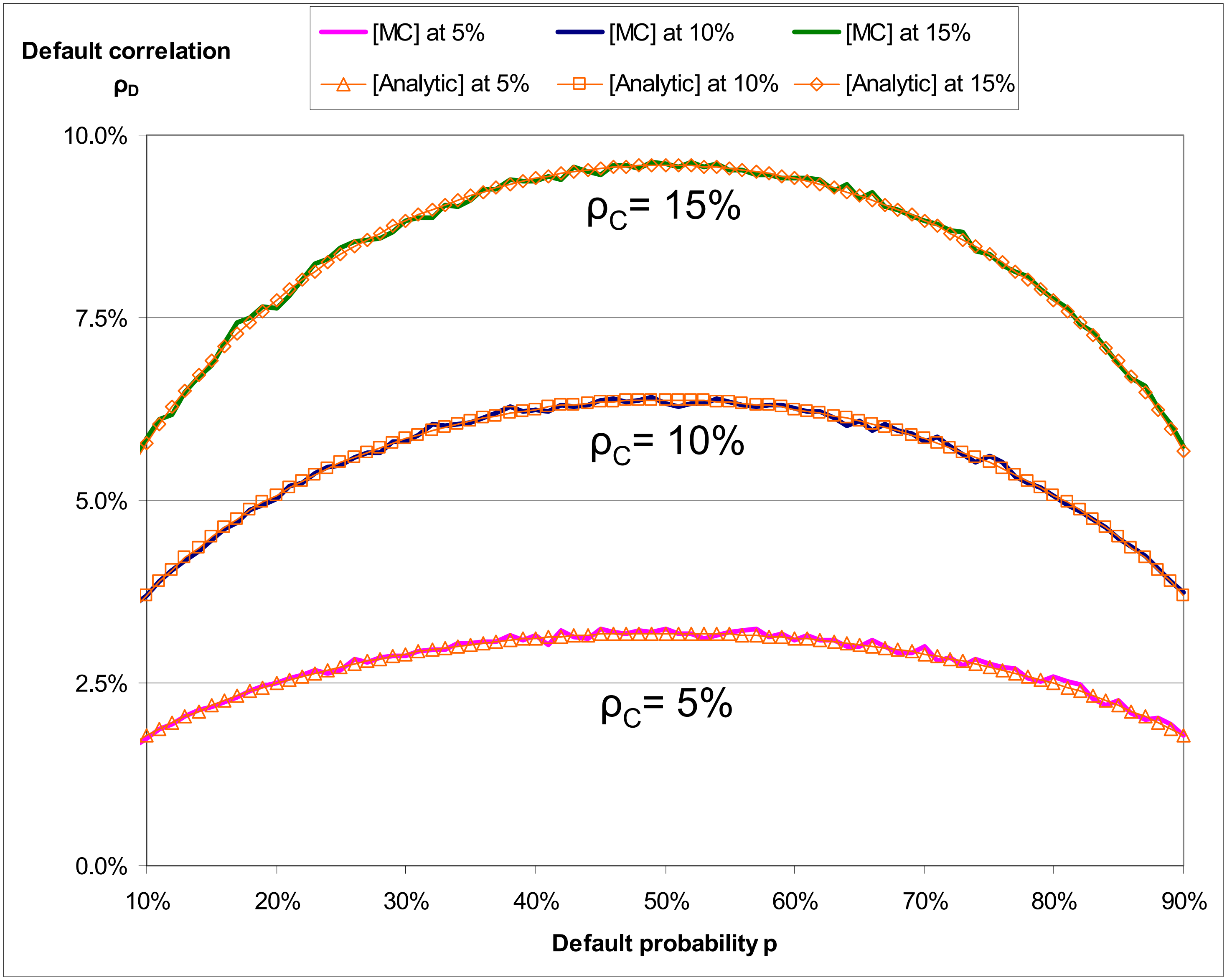

Default correlation
ρD
p = 10%
p = 20%
p = 30%
p = 98%
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Copula correlation factor ρC
p = 30%
p = 20%
p = 10%
p = 98%

Figure 9.



Copula correlation factor ρC
80%
70%
60%
50%
40%
30%
20%
10%
0%
0%
10%
20%
30%
40%
50%
60%
70%
80%
90%
100%
Default probability p
ρD= 30%
ρD= 20%
ρD= 10%

Figure 10.



Default correlation
ρ_D
[MC] at 5%
[MC] at 10%
[MC] at 15%
[Analytic] at 5%
[Analytic] at 10%
[Analytic] at 15%
10.0%
7.5%
5.0%
2.5%
0.0%
ρ_C= 15%
ρ_C= 10%
ρ_C= 5%
10%
20%
30%
40%
50%
60%
70%
80%
90%
Default probability p