The Two Most Popular Copulas

1) Normal (Gaussian) copula is

$$\boxed{C(u_1, u_2) = \Phi_2(\Phi^{-1}(u_1), \Phi^{-1}(u_2))}$$

where

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-\frac{x^2}{2}} du$$

= CDF of standard normal distribution

and

$$\Phi_2(u_1, u_2) = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{u_1} \int_{-\infty}^{u_2}$$

$$e^{-\frac{x^2+y^2-2\rho xy}{2(1-\rho^2)}} dy\, dx$$

= Joint CDF of standard bivariate normal distribution

The normal/Gaussian copula was the most popular copula up until 2008.
Financial data do not follow the normal distribution well.
Hence Gaussian copula is not a good model for financial data.

ASA Excellence in Statistical Reporting Award

The formula that killed Wall Street

Wall Street in the mid-1980s turned to the quants – brainy financial engineers – to invent new ways to boost profits. They and their managers, though laziness and greed, built a huge financial bubble on foundations that they did not understand. It was a recipe for disaster. The journalist **Felix Salmon** won the American Statistical Association's Excellence in Statistical Reporting Award for 2010. We reprint his article, first published as the cover story of *Wired* magazine, because it brilliantly conveys complex statistical concepts to non-specialists.

A formula in statistics, misunderstood and misused, has devastated the global economy

In the years before 2008, it was hardly unthinkable that a math wizard like David X. Li might someday earn a Nobel Prize. After all, financial economists – even Wall Street quants – have received the Nobel in economics before, and Li's work on measuring risk has had more impact, more quickly, than previous Nobel Prize-winning contributions to the field. Today, though, as dazed bankers, politicians, regulators, and investors survey the wreckage of the biggest financial meltdown since the Great Depression, Li is probably thankful he still has a job in finance at all. Not that his achievement should be dismissed. He took a notoriously tough nut – determining correlation, or how seemingly disparate events are related – and cracked it wide open with a simple and elegant mathematical formula, one that would become ubiquitous in finance worldwide.

For five years, Li's formula, known as a Gaussian copula function, looked like an unambiguously positive breakthrough, a piece of financial technology that allowed hugely complex risks to be modeled with more ease and accuracy than ever before. With his brilliant spark of mathematical legerdemain, Li made it possible for traders to sell vast quantities of new securities, expanding financial markets to unimaginable levels.

His method was adopted by everybody from bond investors and Wall Street banks to ratings agencies and regulators. And it became so deeply entrenched – and was making people so much money – that warnings about its limitations were largely ignored.

Then the model fell apart. Cracks started appearing early on, when financial markets began behaving in ways that users of Li's formula hadn't expected. The cracks became full-fledged canyons in 2008 – when ruptures in the financial system's foundation swallowed up trillions of dollars and put the survival of the global banking system in serious peril.

David X. Li, it's safe to say, won't be getting that Nobel anytime soon. One result of the collapse has been the end of financial economics as something to be celebrated rather than feared. And Li's Gaussian copula formula will go down in history as instrumental in causing the unfathomable losses that brought the world financial system to its knees.

How could one formula pack such a devastating punch? The answer lies in the bond market, the multitrillion-dollar system that allows pension funds, insurance companies, and hedge funds to lend trillions of dollars to companies, countries, and home buyers.

A bond, of course, is just an IOU, a promise to pay back money with interest by certain dates. If a company – say, IBM – borrows money by issuing a bond, investors will look very closely over its accounts to make sure it has the wherewithal to repay them. The higher the perceived risk – and there's always *some* risk – the higher the interest rate the bond must carry.

Bond investors are very comfortable with the concept of probability. If there's a 1 percent chance of default but they get an extra two percentage points in interest, they're ahead of the game overall – like a casino, which is happy to lose big sums every so often in return for profits most of the time.

Bond investors also invest in pools of hundreds or even thousands of mortgages. The potential sums involved are staggering: Americans now owe more than $11 trillion on their homes. But mortgage pools are messier than most bonds. There's no guaranteed interest rate, since the amount of money homeowners collectively pay back every month is a function of how many have refinanced and how many have defaulted. There's certainly no fixed maturity date: Money shows up in irregular chunks as people pay down their mortgages at unpredictable times – for instance, when they decide to sell their house. And most problematic, there's no easy way

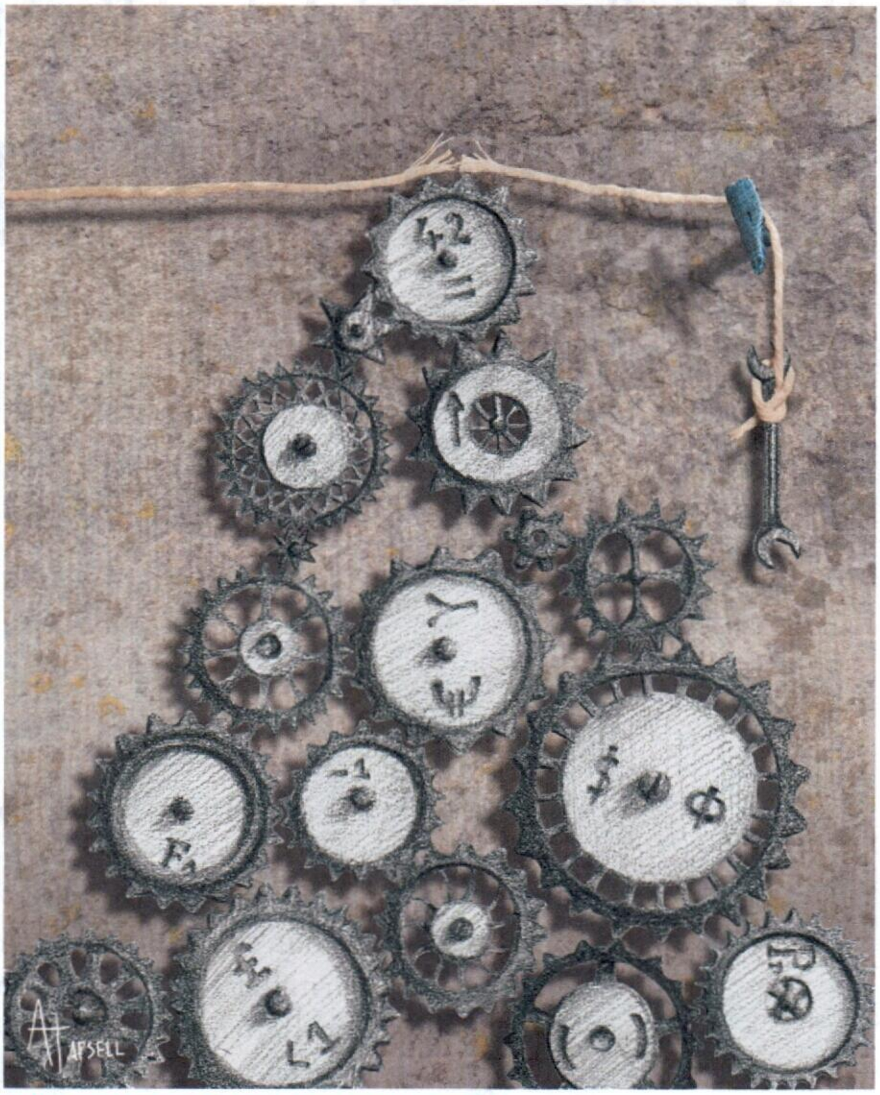

Illustration: Andrew Tapsell, http://andrewtapsell.blogspot.com

to assign a single probability to the chance of default.

Wall Street solved many of these problems through a process called tranching, which divides a pool and allows for the creation of safe bonds with a risk-free triple-A credit rating. Investors in the first tranche, or slice, are first in line to be paid off. Those next in line might get only a double-A credit rating on their tranche of bonds but will be able to charge a higher interest rate for bearing the slightly higher chance of default. And so on.

The reason that ratings agencies and investors felt so safe with the triple-A tranches was that they believed there was no way hundreds of homeowners would all default on their loans at the same time. One person might lose his job, another might fall ill. But those are individual calamities that don't affect the mortgage pool much as a whole: Everybody else is still making their payments on time.

But not all calamities are individual, and tranching still hadn't solved all the problems of mortgage-pool risk. Some things, like falling house prices, affect a large number of people at once. If home values in your neighborhood decline and you lose some of your equity, there's a good chance your neighbors will lose theirs as well. If, as a result, you default on your mortgage, there's a higher probability they will default, too. That's called correlation – the degree to which one variable moves in line with another – and measuring it is an important part of determining how risky mortgage bonds are.

Investors *like* risk, as long as they can price it. What they hate is uncertainty – not knowing how big the risk is. As a result, bond investors and mortgage lenders desperately want to be able to measure, model, and price correlation. Before quantitative models came along, the only time investors were comfortable putting their money in mortgage pools was when there was no risk whatsoever – in other words, when the bonds were guaranteed implicitly by the federal government through Fannie Mae or Freddie Mac.

Yet during the '90s, as global markets expanded, there were trillions of new dollars waiting to be put to use lending to borrowers around the world – not just mortgage seekers but also corporations and car buyers and anybody running a balance on their credit card – if only investors could put a number on the correlations between them. The problem is excruciatingly hard, especially when you're talking about thousands of moving parts. Whoever solved it would earn the eternal gratitude of Wall Street and quite possibly the attention of the Nobel committee as well.

To understand the mathematics of correlation better, consider something simple, like a kid in an elementary school: Let's call her Alice. The probability that her parents will get divorced this year is about 5 percent, the risk of her getting head lice is about 5 percent, the chance of her seeing a teacher slip on a banana peel is about 5 percent, and the likelihood of her winning the class spelling bee is about 5 percent. If investors were trading securities based on the chances of those things happening only to Alice, they would all trade at more or less the same price.

But something important happens when we start looking at two kids rather than one – not just Alice but also the girl she sits next to, Britney. If Britney's parents get divorced, what are the chances that Alice's parents will get divorced, too? Still about 5 percent: The correlation there is close to zero. But if Britney gets head lice, the chance that Alice will get head lice is much higher, about 50 percent – which means the correlation is probably up in the 0.5 range. If Britney sees a teacher slip on a banana peel, what is the chance that Alice will see it, too? Very high indeed, since they sit next to each other: It could be as much as 95 percent, which means the correlation is close to 1. And if Britney wins the class spelling bee, the chance of Alice winning it is zero, which means the correlation is negative: −1.

If investors were trading securities based on the chances of these things happening to both Alice *and* Britney, the prices would be all over the place, because the correlations vary so much.

But it's a very inexact science. Just measuring those initial 5 percent probabilities involves collecting lots of disparate data points and subjecting them to all manner of statistical and error analysis. Trying to assess the conditional probabilities – the chance that Alice will get head lice *if* Britney gets head lice – is an order of magnitude harder, since those data points are much rarer. As a result of the scarcity of historical data, the errors there are likely to be much greater.

In the world of mortgages, it's harder still. What is the chance that any given home will decline in value? You can look at the past history of housing prices to give you an idea, but surely the nation's macroeconomic situation also plays an important role. And what is the chance that if a home in one state falls in value, a similar home in another state will fall in value as well?

Enter Li, a star mathematician who grew up in rural China in the 1960s. He excelled in school and eventually got a master's degree in economics from Nankai University before leaving the country to get an MBA from Laval University in Quebec. That was followed by two more degrees: a master's in actuarial science and a PhD in statistics, both from Ontario's University of Waterloo. In 1997 he landed at Canadian Imperial Bank of Commerce, where his financial career began in earnest; he later moved to Barclays Capital and by 2004 was charged with rebuilding its quantitative analytics team.

Li's trajectory is typical of the quant era, which began in the mid-1980s. Academia could never compete with the enormous salaries that banks and hedge funds were offering. At the same time, legions of math and physics PhDs were required to create, price, and arbitrage Wall Street's ever more complex investment structures.

In 2000, while working at JPMorgan Chase, Li published a paper in *The Journal of Fixed Income* titled "On default correlation:

$$\Pr[T_A<1, T_B<1] = \phi_2(\phi^{-1}(F_A(1)), \phi^{-1}(F_B(1)), \gamma)$$

The formula that killed so many pension plans: David X. Li's Gaussian copula, as first published in 2000. Investors exploited it as a quick – and fatally flawed – way to assess risk.

Probability

Specifically, this is a joint default probability – the likelihood that any two members of the pool (A and B) will both default. It's what investors are looking for, and the rest of the formula provides the answer.

Survival times

The amount of time between now and when A and B can be expected to default. Li took the idea from a concept in actuarial science that charts what happens to someone's life expectancy when their spouse dies.

Equality

A dangerously precise concept, since it leaves no room for error. Clean equations help both quants and their managers forget that the real world contains a surprising amount of uncertainty, fuzziness, and precariousness.

Copula

This couples (hence the Latinate term copula) the individual probabilities associated with A and B to come up with a single number. Errors here massively increase the risk of the whole equation blowing up.

Distribution functions

The probabilities of how long A and B are likely to survive. Since these are not certainties, they can be dangerous: Small miscalculations may leave you facing much more risk than the formula indicates.

Gamma

The all-powerful correlation parameter, which reduces correlation to a single constant – something that should be highly improbable, if not impossible. This is the magic number that made Li's copula function irresistible.

The Wall Street Bull. © iStockphoto.com/Marcio Silva

a copula function approach". (In statistics, a copula is used to couple the behavior of two or more variables.) Using some relatively simple math – by Wall Street standards, anyway – Li came up with an ingenious way to model default correlation without even looking at historical default data. Instead, he used market data about the prices of instruments known as credit default swaps.

If you're an investor, you have a choice these days: You can either lend directly to borrowers or sell investors credit default swaps, insurance against those same borrowers defaulting. Either way, you get a regular income stream – interest payments or insurance payments – and either way, if the borrower defaults, you lose a lot of money. The returns on both strategies are nearly identical, but because an unlimited number of credit default swaps can be sold against each borrower, the supply of swaps isn't constrained the way the supply of bonds is, so the CDS market managed to grow extremely rapidly. Though credit default swaps were relatively new when Li's paper came out, they soon became a bigger and more liquid market than the bonds on which they were based.

When the price of a credit default swap goes up, that indicates that default risk has risen. Li's breakthrough was that instead of waiting to assemble enough historical data about actual defaults, which are rare in the real world, he used historical prices from the CDS market. It's hard to build a historical model to predict Alice's or Britney's behavior, but anybody could see whether the price of credit default swaps on Britney tended to move in the same direction as that on Alice. If it did, then there was a strong correlation between Alice's and Britney's default risks, as priced by the market. Li wrote a model that used price rather than real-world default data as a shortcut (making an implicit assumption that financial markets in general, and CDS markets in particular, can price default risk correctly).

It was a brilliant simplification of an intractable problem. And Li didn't just radically dumb down the difficulty of working out correlations; he decided not to even bother trying to map and calculate all the nearly infinite relationships between the various loans that made up a pool. What happens when the number of pool members increases or when you mix negative correlations with positive ones? Never mind all that, he said. The only thing that matters is the final correlation number – one clean, simple, all-sufficient figure that sums up everything.

The effect on the securitization market was electric. Armed with Li's formula, Wall Street's quants saw a new world of possibilities. And the first thing they did was start creating a huge number of brand-new triple-A securities. Using Li's copula approach meant that ratings agencies like Moody's – or anybody wanting to model the risk of a tranche – no longer needed to puzzle over the underlying securities. All they needed was that correlation number, and out would come a rating telling them how safe or risky the tranche was.

As a result, just about anything could be bundled and turned into a triple-A bond – corporate bonds, bank loans, mortgage-backed securities, whatever you liked. The consequent pools were often known as collateralized debt obligations, or CDOs. You could tranche that pool and create a triple-A security even if none of the components were themselves triple-A. You could even take lower-rated tranches of *other* CDOs, put them in a pool, and tranche them – an instrument known as a CDO-squared, which at that point was so far removed from any actual underlying bond or loan or mortgage that no one really had a clue what it included. But it didn't matter. All you needed was Li's copula function.

The CDS and CDO markets grew together, feeding on each other. At the end of 2001, there was $920 billion in credit default swaps outstanding. By the end of 2007, that number had skyrocketed to more than $62 *trillion*. The CDO market, which stood at $275 billion in 2000, grew to $4.7 trillion by 2006.

At the heart of it all was Li's formula. When you talk to market participants, they use words like *beautiful*, *simple*, and, most commonly, *tractable*. It could be applied anywhere, for anything, and was quickly adopted not only by banks packaging new bonds but also by traders and hedge funds dreaming up complex trades between those bonds.

"The corporate CDO world relied almost exclusively on this copula-based correlation model", says Darrell Duffie, a Stanford University finance professor who served on Moody's Academic Advisory Research Committee. The Gaussian copula soon became such a universally accepted part of the world's financial vocabulary that brokers started quoting prices for bond tranches based on their correlations. "Correlation trading has spread through the psyche of the financial markets like a highly infectious thought virus", wrote derivatives guru Janet Tavakoli in 2006.

The damage was foreseeable and, in fact, foreseen. In 1998, before Li had even invented his copula function, Paul Wilmott wrote that "the correlations between financial quantities are notoriously unstable". Wilmott, a quantitative-finance consultant and lecturer, argued that no theory should be built on such unpredictable parameters. And he wasn't alone. During the boom years, everybody could reel off reasons why the Gaussian copula function wasn't perfect. Li's approach made no allowance for unpredictability: It assumed that correlation was a constant rather than something mercurial. Investment banks would regularly phone Stanford's Duffie and ask him to come in and talk to them about exactly what Li's copula was. Every time, he would warn them that it was not suitable for use in risk management or valuation.

In hindsight, ignoring those warnings looks foolhardy. But at the time, it was easy. Banks dismissed them, partly because the managers empowered to apply the brakes didn't understand the arguments between various arms of the quant universe. Besides, they were making too much money to stop.

In finance, you can never reduce risk outright; you can only try to set up a market in which people who don't want risk sell it to those who do. But in the CDO market, people used the Gaussian copula model to convince themselves they didn't have any risk at all, when in fact they just didn't have any risk 99 percent of the time. The other 1 percent of the time they blew up. Those explosions may have been rare, but they could destroy all previous gains, and then some.

Li's copula function was used to price hundreds of billions of dollars' worth of CDOs filled with mortgages. And because the copula function used CDS prices to calculate correlation, it was forced to confine itself to looking at the period of time when those credit default swaps had been in existence: less than a decade, a period when house prices soared. Naturally, default correlations were very low in those years. But when the mortgage boom ended abruptly and home values started falling across the country, correlations soared.

Bankers securitizing mortgages knew that their models were highly sensitive to house-price appreciation. If it ever turned negative on a national scale, a lot of bonds that had been rated triple-A, or risk-free, by copula-powered computer models would blow up. But no one was willing to stop the creation of CDOs, and the big investment banks happily kept on building more, drawing their correlation data from a period when real estate only went up.

"Everyone was pinning their hopes on house prices continuing to rise", says Kai Gilkes of the credit research firm CreditSights, who spent 10 years working at ratings agencies. "When they stopped rising, pretty much everyone was caught on the wrong side, because the sensitivity to house prices was huge. And there was just no getting around it. Why didn't rating agencies build in some cushion for this sensitivity to a house-price-depreciation scenario? Because if they had, they would have never rated a single mortgage-backed CDO."

Bankers should have noted that very small changes in their underlying assumptions could result in very large changes in the correlation number. They also should have noticed that the results they were seeing were much less volatile than they should have been – which implied that the risk was being moved elsewhere. Where had the risk gone?

They didn't know, or didn't ask. One reason was that the outputs came from "black box" computer models and were hard to subject to a commonsense smell test. Another was that the quants, who should have been more aware of the copula's weaknesses, weren't the ones making the big asset-allocation decisions. Their managers, who made the actual calls, lacked the math skills to understand what the models were doing or how they worked. They could, however, understand something as simple as a single correlation number. That was the problem.

"The relationship between two assets can never be captured by a single scalar quantity", Wilmott says. For instance, consider the share prices of two sneaker manufacturers: When the market for sneakers is growing, both companies do well and the correlation between them is high. But when one company gets a lot of celebrity endorsements and starts stealing market share from the other, the stock prices diverge and the correlation between them turns negative. And when the nation morphs into a land of flip-flop-wearing couch potatoes, both companies decline and the correlation becomes positive again. It's impossible to sum up such a history in one correlation number, but CDOs were invariably sold on the premise that correlation was more of a constant than a variable.

"Very few people understand the essence of the model" – David X. Li

"Anything that relies on correlation is charlatanism" – Nassim Taleb

No one knew all of this better than David X. Li: "Very few people understand the essence of the model", he told *The Wall Street Journal* way back in fall 2005.

"Li can't be blamed", says Gilkes of CreditSights. After all, he just invented the model. Instead, we should blame the bankers who misinterpreted it. And even then, the real danger was created not because any given trader adopted it but because every trader did. In financial markets, everybody doing the same thing is the classic recipe for a bubble and inevitable bust.

Nassim Nicholas Taleb, hedge fund manager and author of *The Black Swan*, is particularly harsh when it comes to the copula. "People got very excited about the Gaussian copula because of its mathematical elegance, but the thing never worked", he says. "Co-association between securities is not measurable using correlation", because past history can never prepare you for that one day when everything goes south. "Anything that relies on correlation is charlatanism."

Li has been notably absent from the current debate over the causes of the crash. In fact, he is no longer even in the US. In 2009 he moved to Beijing to head up the risk-management department of China International Capital Corporation. In a recent conversation, he seemed reluctant to discuss his paper and said he couldn't talk without permission from the PR department. In response to a subsequent request, CICC's press office sent an email saying that Li was no longer doing the kind of work he did in his previous job and, therefore, would not be speaking to the media.

In the world of finance, too many quants see only the numbers before them and forget about the concrete reality the figures are supposed to represent. They think they can model just a few years' worth of data and come up with probabilities for things that may happen only once every 10 000 years. Then people invest on the basis of those probabilities, without stopping to wonder whether the numbers make any sense at all.

As Li himself said of his own model: "The most dangerous part is when people believe everything coming out of it."

Felix Salmon (felix@felixsalmon.com) writes the Market Movers financial blog at Portfolio.com.

2) <u>t copula</u> is

$$\boxed{C_1^1(u_1, u_2) = T_2\left(t_a^{-1}(u_1), t_a^{-1}(u_2)\right)}$$

where

$$t_a(u) = \frac{\Gamma\left(\frac{a+1}{2}\right)}{\sqrt{a\pi}\;\Gamma\left(\frac{a}{2}\right)} \int_{-\infty}^{u} \left(1+\frac{x^2}{a}\right)^{-\frac{a+1}{2}} dx$$

= CDF of Student's t distribution with degree of freedom a.

and

$$T_2(u_1, u_2) = \frac{\Gamma\left(\frac{a+2}{2}\right)}{a\pi\;\Gamma\left(\frac{a}{2}\right)\sqrt{1-\rho^2}}$$

$$\cdot \int_{-\infty}^{u_1}\int_{-\infty}^{u_2}\left[1+\frac{x^2+y^2-2\rho xy}{a(1-\rho^2)}\right]^{-\frac{a+2}{2}} dy\,dx$$

= Joint CDF of a bivariate t distribution with degree of freedom a.

t copula is being increasingly used in the financial sector.