1
Introdução

Nas últimas décadas, a análise dos dados oriundos das áreas financeiras, ambientais e biomédicas, dentre outras, mostra que os conjuntos de dados para os quais podemos assumir a hipótese de normalidade são exceções e não regra. Por exemplo, é fato que as séries de retornos financeiros possuem características não condizentes com a hipótese de normalidade, como a assimetria e a presença de caudas pesadas (curtose).

Apesar de ser um conceito onipresente na teoria moderna de finança e atuária, a correlação linear [1] é um conceito compreendido por poucos. Grande parte da confusão no uso e interpretação desse conceito vem do fato do termo correlação, literalmente, ser usado para descrever todo o tipo de estrutura de dependência entre v.a.'s ((11)) quando, de fato, mede um tipo particular de dependência estocástica: a dependência linear.

A distribuição normal foi, e ainda é, a base para muitas análises de dados. A grande utilização da distribuição normal pode ser explicada pelo fato de ser matematicamente tratável, ser fechada por soma, marginalização e dependência e, em especial, pelo seu papel de destaque no Teorema Central do Limite ((7)). A distribuição normal multivariada tem sua estrututra de dependência completamente especificada pela matriz de correlação associada, que é uma matriz positiva-definida na qual os elementos da diagonal principal são iguais a 1 e os outros elementos nos dão os valores de ρ entre as diversas v.a.'s em análise.

Tomar por hipótese a distribuição normal multivariada implica restringir o tipo de associação entre as v.a.'s envolvidas a ser linear. Esta é uma restrição drástica, e não parece razoável ser a associação linear o único tipo de dependência que poderia ser observada entre séries financeiras ((19), por exemplo).

[1] **Definição (Coeficiente de correlação linear).** Seja (X,Y) um vetor aleatório no qual as variáveis aleatórias (v.a.'s) X e Y são definidas em um mesmo espaço de probabilidade, integráveis e com variâncias positivas e finitas ($0 < \sigma_X^2 < \infty$, $0 < \sigma_Y^2 < \infty$). O coeficiente de correlação linear, ρ, entre X e Y é definido por

$$\rho := \frac{\mathbb{C}ov(X,Y)}{\sigma_X \cdot \sigma_Y}, \tag{1-1}$$

onde $\mathbb{C}ov(X,Y)$ denota a covariância entre X e Y.

Embora existam alternativas de distribuições assimétricas na literatura, como por exemplo, a distribuição gama, nas últimas décadas cresceu o interesse por distribuições que melhor modelam não somente a assimetria, mas também outras características não condizentes com a normalidade, como a curtose e a multimodalidade ((2), (3), (6), (14), (19), dentre outros).

Tomando o mercado de seguros como exemplo, vemos que tanto no ramo vida quanto no ramo não vida, uma questão fundamental para a análise de risco é saber qual a estrutura de dependência (e não somente dependência linear) existente entre todos os riscos aos quais a seguradora está exposta.

A estrutura de dependência das v.a.'s reais é completamente determinada pela função de distribuição acumulada (f.d.a.) conjunta ((25)).

Com o objetivo de analisar a estrutura de dependência existente entre as várias v.a.'s envolvidas, as cópulas ganharam papel de destaque nos últimos anos.

As cópulas expressam a f.d.a. conjunta como função de suas f.d.a.'s marginais; além disso, as cópulas são invariantes por transformações estritamente crescentes quase-certamente (q.c.) das v.a.'s envolvidas. Portanto, a representação e interpretação claras da estrutura de dependência presente em vetores aleatórios, em particular em vetores bivariados, podem ser feitas com o uso do conceito de cópulas. Por isso, diz-se que as cópulas capturam toda a estrutura de dependência envolvida.

Na análise bivariada, os coeficientes de dependência de cauda, ((15) e (21)), têm por objetivo estudar uma medida de dependência quando as variáveis assumem valores extremos. Tais coeficientes de dependência de cauda foram definidos, inicialmente, com o objetivo de mensurar a dependência das variáveis nos extremos dos quadrantes ímpares de $\mathbb{R}^2$. Para aumentar o espectro de análise dos dados, Zhang ((26)) introduziu o conceito de dependência de cauda total, que permite analisar a dependência das variáveis nos extremos de todos os quadrantes de $\mathbb{R}^2$. Quando os coeficientes de dependência de cauda são iguais a zero, dizemos que o par de v.a.'s em análise é assintoticamente independente.

Esses coeficientes nos dão um exemplo da limitação do coeficiente de correlação linear na descrição da estrutura de dependência presente em eventos extremos: as distribuições normais bivariadas, com $|\rho| < 1$, são necessariamente assintoticamente independentes.

Se as f.d.a.'s das v.a.'s envolvidas são contínuas, podemos expressar os coeficientes de dependência de cauda como funções da cópula. Assim, especificar uma cópula implica determinar os coeficientes de cauda. Esse fato será determinante na escolha das cópulas usadas no Capítulo 5.

No Capítulo 2 apresentamos definições e resultados básicos relacionados às cópulas e, em particular, relacionados às cópulas Arquimedianas; neste Capítulo e no Apêndice A são apresentadas as cópulas que serão usadas no Capítulo 5.

Os conceitos de cópulas de sobrevivência e vetor aleatório radialmente simétrico em torno de um ponto são estudados com o objetivo de obter resultados que simplifiquem o cálculo dos coeficientes de dependência de cauda de distribuições importantes, como a Gaussiana e a distribuição t de Student, como veremos no Capítulo 3.

Como contribuição para a literatura, apresentamos o cálculo do coeficiente de correlação linear para a f.d.a. bivariada determinada pela cópula Farlie-Gumbel-Morgenstern (FGM) e por distribuições marginais normal-padrão usando resultados relacionados à distribuição normal assimétrica.

No Capítulo 3 definimos a dependência homogênea e heterogênea de cauda. São calculados os coeficientes de dependência de cauda para as cópulas usadas no Capítulo 5; os cálculos desses coeficientes para as cópulas Gumbel, BB7 e Clayton são dados no Apêndice A.

A expressão dos coeficientes de dependência de cauda quando dispomos da expressão "fechada"para a cópula associada ao vetor aleatório em análise já é conhecida.

Porém, cópulas importantes, como a Gaussiana e a t de Student, não possuem expressões "fechadas"; para contornar esse problema, obtemos as expressões dos coeficientes de dependência heterogênea de cauda em função da f.d.a. condicional, que são resultados inéditos.

No Capítulo 4 definimos as distribuições normais assimétricas univariadas e multivariadas. É calculado o excesso de curtose em uma distribuição normal assimétrica supondo o parâmetro de escala diferente de 1. A demonstração de que os coeficientes de dependência homogênea de cauda de uma distribuição normal assimétrica são iguais a zero é um resultado inédito.

Para o cálculo dos coeficientes de dependência homogênea de cauda associados a uma distribuição normal assimétrica bivariada, faremos uso de um resultado que fornece a esperança da v.a. obtida da composição entre a f.d.a. normal padrão e uma variável normalmente distribuída. Tal resultado já foi enunciado na literatura científica ((25)), mas sua correta demonstração é dada na Seção 4.3.

Uma informação de grande interesse para o Departamento de Operações de Mercado Aberto (Demab) do Banco Central do Brasil (Bacen) é a estrutura a termo das taxas de juros pré-fixada, que reflete a expectativa do mercado em relação à economia brasileira como um todo e, em particular, reflete a expectativa do mercado em relação a algumas variáveis da economia.

Dado que não são negociadas no mercado brasileiro taxas de juros pré-fixadas para todos os prazos, um grupo de estudo do Demab implementou um modelo de interpolação que fornece a estrutura a termo das taxas de juros pré-fixada para todos os prazos a partir dos títulos negociados no mercado ((1)).

Além disso, o Bacen dispõe das Séries Temporais das Expectativas de Mercado, que são planilhas com os dados diários das expectativas de mercado para as principais variáveis da economia, dentre as quais citamos a taxa de câmbio de referência do dólar dos Estados Unidos, conhecida no mercado como a taxa PTAX.

No Capítulo 5, de modo inédito, estudamos a estrutura de dependência entre as seguintes variáveis: (i) log-retornos das taxas, interpoladas, para a estrutura a termo pré-fixada de 1 ano e de 2 anos; (ii) log-retorno das taxas para a estrutura a termo pré-fixada de 1 (um) ano e log-retorno do índice do Ibovespa; e (iii) log-retorno das taxas para a estrutura a termo pré-fixada de 1 (um) ano e log-retorno da expectativa da taxa PTAX, 6 meses a frente. Inicialmente, modelaremos as f.d.a.'s marginais dos log-retornos das variáveis citadas acima através de um modelo semi-paramétrico GPD, ou seja, ajustamos a distribuição generalizada de Pareto (GPD) para as caudas da distribuição, a partir de limiares pré-definidos, e calculamos a distribuição empírica para o centro da distribuição. Com base nas f.d.a.'s marginais ajustadas e na amostra coletada calcularemos a pseudo-amostra da cópula associada. Tal pseudo-amostra permite-nos obter os estimadores de máxima verossimilhança (EMV) dos parâmetros desconhecidos das cópulas a serem ajustadas. A f.d.a. conjunta estimada é obtida com a aplicação do Teorema de Sklar.

A partir do ajuste de uma cópula aos dados, estudamos a estrutura de dependência existente entre as variáveis listadas no parágrafo anterior. Vale ressaltar que o uso dos coeficientes de dependência de cauda em tal análise permite o estudo do comportamento das variáveis quando da ocorrência de eventos extremos.