

Universidade de Brasília
Instituto de Exatas
Departamento de Estatística

Value at Risk e Cópula

VaR de um *Portfolio* de ações e Cópula Bivariada

Thiago Assunção de Faria

Brasília
2011

Thiago Assunção de Faria - 10/07521

Value at Risk e Cópula

VaR de um Portfolio de ações e Cópula Bivariada

Relatório apresentado à disciplina Estágio Supervisionado do curso de graduação em Estatística, Departamento de Estatística, Instituto de Exatas, Universidade de Brasília, como parte dos requisitos necessários para o grau de Bacharel em Estatística.

Orientador: *Prof.º Donald Matthew Pianto*

Brasília - DF

junho 2011

Agradecimentos

Primeiramente à Deus e aos meus pais, pelo amor, carinho, suporte e apoio incondicional durante toda a minha vida. Não deve ser fácil ver o filho trocar de curso na Universidade três vezes, por isso obrigado, de verdade. Tiveram muito paciência com as minhas indecisões até que eu encontrasse na Estatística algo que eu realmente gostasse de estudar e fazer.

Aos meus professores e mestres na Universidade de Brasília, pelo conhecimento passado e empenho demonstrado, apesar das dificuldades que a classe docente enfrenta no nosso País. Em especial gostaria de agradecer a três professores: Alan Ricardo, por me mostrar que a Estatística se faz através do suor e do trabalho, não necessariamente da inspiração e genialidade; Afrânio Vieira, pelos conselhos profissionais e experiência transmitida com afincamento e dedicação, tentando desenvolver uma veia de pesquisa e empreendedora em todos os seus alunos; e o mais importante, Donald Pianto, pois além de um orientador foi um amigo, que nunca me deixou ficar desesperado com as dificuldades que surgiram, sempre me transmitindo a calma e a confiança necessária para atingir este objetivo.

À minha namorada, pelo amor e carinho. Foram tantos momentos de descontração e risadas, momentos que me fizeram esquecer os problemas da vida. Sem estes momentos tudo teria tido menos cor e seria mais doloroso.

Aos meus colegas de Universidade, autores e todos que contribuíram (mesmo sem saber) com suas reflexões para o desenvolvimento deste trabalho.

Resumo

O presente trabalho buscou aplicar o conceito de Cópula abordando a dependência entre variáveis, como estudá-la e fazer inferências. A Cópula nada mais é do que a função conjunta, definida pelas marginais, que incorpora em sua construção um conceito de dependência bem geral. Portanto, através da Cópula tem-se como avaliar o comportamento conjunto das variáveis aleatórias abordadas, de maneira clara e simples. Calculou-se o *VaR* de um *Portfolio* composto pelas duas ações mais negociadas (VALE3 e PETR4) na Bolsa de Valores de São Paulo (BOVESPA) da seguinte maneira: estimou-se a distribuição dos *log-retornos* de cada papel com a metodologia GARCH, e em seguida, através de uma modelagem da Cópula Bivariada, pôde-se calcular o *Value at Risk* do *Portfolio* através de Simulações. Ao fim do estudo verificou-se como o cálculo do *VaR* sem considerar tratando as variáveis de forma independente é subestimado quando comparado ao *VaR* da Cópula calculada.

Palavras-chave: GARCH, *Value at Risk*, *VaR*, Cópula, Cramér-von Mises.

Sumário

1	INTRODUÇÃO	1
2	OBJETIVOS	4
3	GESTÃO DO RISCO	5
3.1	O que é Risco?	5
3.2	Risco Financeiro	6
4	VALUE AT RISK	10
4.1	Metodologias de Cálculo do <i>VaR</i>	12
4.2	Simulações de Monte-Carlo	13
4.3	Método Paramétrico	14
5	SÉRIES TEMPORAIS FINANCEIRAS	16
5.1	Modelos ARCH	17
5.2	Modelos GARCH	21
6	CÓPULA	23
6.1	Conceitos Preliminares	24
6.2	Cópula Bivariada	25
6.3	Teorema de Sklar e Variáveis Aleatórias	29
6.4	Cópuas Elípticas	32
6.5	Cópuas Arquimedianas	35
6.6	Medidas de Associação	36

6.7	Teste de <i>Goodness-of-fit</i>	39
7	METODOLOGIA	42
8	RESULTADOS	43
8.1	GARCH dos retornos univariados	44
8.2	Escolha da Cópula	46
8.3	Simulação de Valores e Cálculo do <i>VaR</i>	52
9	CONCLUSÃO	55
	REFERÊNCIAS BIBLIOGRÁFICAS	58
	Apêndice: Outros Gráficos e Tabelas	60

1 INTRODUÇÃO

Hoje em dia existem inúmeros tipos de investimentos em mercados oscilantes cujo resultado esperado pelo investidor é o lucro. Estas movimentações financeiras podem ser através de aplicações mais tradicionais e conhecidas de grande parte da população como ações e fundos de investimento, ou através de investimentos mais complexos como *swaps*, investimentos cambiais, etc. Contudo, é de extrema importância notar que apesar das discrepâncias entre os tipos de investimento e do montante de capital investido, todos colocam o investidor em posição arriscada, pois sempre existe a possibilidade de que o capital investido se reduza.

No decorrer dos últimos anos vem surgindo inúmeras técnicas de análise do risco financeiro, visando prover suporte para a tomada de decisão e algumas perguntas primordiais como “Em que e quanto investir?” e “Quanto posso perder neste investimento em um período x ?”. Algumas técnicas são mais qualitativas e vem caindo em desuso, perdendo cada vez mais espaço no mundo financeiro para técnicas com resultados quantitativos, que auxiliem a tomada de decisão mais rapidamente e objetivamente. Portanto, são diversas as metodologias que dão um suporte matemático à análise de risco e, não é de se espantar, que a grande maioria é construída baseado na teoria estatística e nos conceitos de probabilidade, especialmente envolvendo distribuições conjuntas e probabilidade condicional.

Após o embate gerado pela recessão econômica global, o surgimento das “bolhas de crédito” e o *crash* de 2008 viu-se que a Análise do Risco é cada vez mais fundamental para os aspectos econômicos globais. Reuniões globais de suma importância

já foram feitas para se regulamentar, esclarecer e tornar o Risco cada vez mais pré-avaliado (como o Basileia I e II). Nesta linha, Jorion (1999) apresenta o *Value at Risk* como o modelo mais utilizado por bancos, fundos de investimentos, corretoras e até empresas não vinculadas diretamente com o mercado financeiro a fim de avaliar seus riscos financeiros. A hegemonia do *VaR* como mensurador de risco financeiro mais utilizado se deve principalmente a dois fatores: facilidade de cálculo e resultado em uma simples medida. Mas então, o que é o *VaR*? De maneira intuitiva pode-se dizer que é uma medida que sumariza a maior perda esperada de um ativo ou carteira, para um determinado horizonte de tempo dado um nível de confiança. Logo, o *VaR* é facilmente interpretado por pessoas que não necessariamente necessitam de um vasto conhecimento estatístico.

Um ponto de amplo debate sobre o *Var* é o seu cálculo em uma carteira (*Portfolio*). O retorno de um ativo é uma variável aleatória, logo uma carteira é uma combinação linear de variáveis aleatórias. Como abordar o problema de correlação e dependência entre os elementos do *Portfolio*? Afinal, quando uma ação da Vale despenca ela gera desconfiança no mercado levando a queda de outras ações ou o aumento do valor das ações da concorrência. Cherubini, Luciano e Vecchiato (2004) mostram como pode-se utilizar uma abordagem probabilística bem interessante para resolver este problema de dependência entre as variáveis, envolvendo uma função de dependência entre as variáveis, a Cópula.

A Cópula nada mais é do que a função conjunta, definida pelas marginais, cuja aplicação ainda é muito pouco explorada no mundo estatístico, apesar de alguns

trabalhos bem interessantes e recentes na área de Finanças. É notável verificar que a Cópula possui incorporado em sua construção um conceito de dependência bem geral, que captura relações lineares ou não lineares entre os integrantes¹. Portanto, através da Cópula temos como avaliar o comportamento conjunto das variáveis aleatórias abordadas de maneira clara e simples.

Considerando um *Portfolio* composto por duas ações negociadas na Bolsa de Valores de São Paulo (BOVESPA) pode-se calcular o *VaR* desta carteira bivariada. Como a Vale do Rio Doce e a Petrobras são as duas empresas cujos papéis são os mais líquidos e apresentam o maior peso no IBOVESPA, foram as representantes escolhidas para este estudo. E justamente, graças à grande relevância dessas ações no mercado, iremos abordar o problema de correlação e dependência através da abordagem por Cópula. Este trabalho visa analisar as séries temporais de retornos da VALE3 e da PETR4, estimar sua correlação através de Copulas, para finalmente calcular o *Value at Risk* de um *Portfolio* composto por estas duas ações.

¹Enquanto algumas medidas amplamente utilizadas para se avaliar a dependência, como o Coeficiente de Correlação Linear de Pearson, capturam apenas um ou poucos tipos de dependência.

2 OBJETIVOS

- Objetivo Geral
 - Fazer uma revisão bibliográfica acerca de Cópulas, para entender esta nova abordagem e a sua relação com a dependência entre variáveis aleatórias, a fim de criar subsídios para o cálculo do *VaR* de um *portfolio* composto por ações da VALE3 e PETR4 .

- Objetivos Específicos
 - Estimar as Distribuições de retorno univariadas via metodologias ARCH e GARCH;
 - Verificar o comportamento conjunto do *Portfolio* através dos modelos de Cópula;
 - Estimar o *VaR* do *Portfolio*;

3 GESTÃO DO RISCO

Empresas desenvolvem o seu negócio, independentemente do ramo de atuação, através do gerenciamento do risco. As que conseguem melhor gerenciá-lo diminuem custos fixos, desenvolvem novas tecnologias e tem estruturas financeiras mais robustas e protegidas das adversidades. Algumas empresas simplesmente aceitam passivamente o risco financeiro, enquanto outras tentam tirar vantagens se expondo a riscos calculados. O monitoramento desta exposição ao Risco, de tal maneira que se pode analisar objetivamente quanto vale se expor e quanto se pode perder (ou ganhar) é chamada de Gestão do Risco.

A Gestão do Risco é o processo de identificação, mensuração e controle da exposição ao risco. E neste processo o *Value at Risk* é uma ferramenta de vanguarda, com menos de 20 anos de utilização, com grande aceitação por quase todos os agentes econômicos. Para entender melhor a construção desta poderosa ferramenta, é necessário destrinchar o conceito de risco financeiro.

3.1 O que é Risco?

São muitos os autores e as tentativas de se definir o Risco de maneira didática e aceitável, a forma de mensurá-lo e como identificá-los. O conceito de risco, bem resumido por Andrade (2000), é “uma estimativa do grau de incerteza que se tem com respeito à realização de resultados futuros desejados”. Todas as empresas estão expostas aos inúmeros tipos de risco, que de maneira bem ampla podem ser divididos em Risco *Business* e *Nonbusiness*:

- *Risco Business*: são aqueles que a empresa assume voluntariamente para criar uma vantagem competitiva e somar valor à empresa (inovações tecnológicas, *design* do produto, *marketing*, etc.);
- *Risco Nonbusiness*: são os outros riscos, isto é, aqueles ao qual a empresa não tem controle (alterações na política econômica de um país, alterações na legislação, etc.).

Mas isto levanta uma outra pergunta: de onde vem o risco? São muitas as fontes do Risco, podendo ser criado pelo próprio homem (inflação, guerras, políticas governamentais, etc.) ou através de fenômenos naturais imprevistos (terremotos, variações no clima, etc.). Outro ponto notório sobre o risco é que ele surge intrinsecamente com o crescimento econômico, pois para se criar uma tecnologia inovadora (novas máquinas, materiais, *softwares*, etc.) foi necessário um investimento inicial que, às vezes, poderia não ter gerado nenhuma novidade e nenhum ganho. Logo o risco e, principalmente, a disposição em encarar o risco são essenciais no desenvolvimento econômico.

3.2 Risco Financeiro

Já o Risco Financeiro pode ser visto como o risco relacionado a possíveis perdas no mercado financeiro, como perdas devido às variações na taxa de juros, variações no *portfolio* investido, etc. Qualquer tipo de empresa está sujeita a riscos financeiros, e todas deveriam ter formas de otimizar seus investimentos de maneira

cuidadosa, para que estes riscos não afetem o produto final da empresa. Ela deve realmente estar concentrada em gerenciar os Riscos *Business*.

Geralmente o Risco Financeiro é dividido em 5 categorias:

- *Risco de Mercado*: advém da volatilidade dos preços de mercado;
- *Risco de Crédito*: advém do fato de que algum integrante contratual possa estar relutante ou incapaz de cumprir suas obrigações financeiras;
- *Risco de Liquidez*: advém de problemas de liquidez de ativos ou problemas de fluxo de caixa;
- *Risco Operacional*: advém de erros (acidentes) humanos ou tecnológicos;
- *Risco Legal*: surge quando uma transação realizada deve ser “cancelada” pois era inexequível em lei.

Porém, existem empresas cujo gerenciamento do Risco Financeiro é a sua atividade-fim, as instituições financeiras. Uma instituição financeira tem como propósito assumir, intermediar ou aconselhar sobre Riscos Financeiros. Para realizar tal tarefa elas devem entender o risco associado e ter uma maneira de mensurá-lo. Assim, pode-se planejar melhor perante cenários adversos e estar preparado para as incertezas do mercado.

A indústria da Gestão do Risco vem crescendo muito rapidamente desde o início da década de 1970. Nesse período o mercado financeiro começou a passar por mudanças substanciais em seus mecanismos de funcionamento, graças a um aumento da volatilidade do mercado. Fatos relevantes como a queda das taxas de

câmbio fixas em 1971², a Crise do Petróleo de 1973³ e a “Segunda-Feira Negra” de 1987⁴ mostraram que o único ponto em comum entre todos estes momentos de crise é a imprevisibilidade.

Em momentos de crise, quando ocorreram perdas financeiras substanciais, como as no parágrafo anterior, o mercado observou atônito a rapidez das mudanças, as quedas acentuadas e os prejuízos gigantescos. A Gestão do Risco Financeiro se desenvolveu justamente para isto, para prover uma proteção parcial contra essas perdas. O mercado de seguros e o financeiro se desenvolvem há centenas de anos visando criar formas de dividir os riscos, protegendo seus recursos e evitando uma crise generalizada. O problema do mercado financeiro é a impossibilidade de se proteger de todos os riscos, pois riscos macroeconômicos que criam flutuações no nível de renda e emprego são impossíveis de serem totalmente previstos.

O aumento da volatilidade já citado nas taxas de juros, taxas de câmbio, preços de *commodities* e mercados financeiros em geral, criou uma necessidade global por novos instrumentos financeiros e ferramentas para gerenciar o risco. A Gestão do Risco Financeiro refere-se à criação e implementação de procedimentos de controle dos riscos financeiros. Com os avanços na Computação e na Teoria Financeira surgiram muitas ferramentas novas, principalmente a partir do final da década de 1970. Primeiramente Markowitz (1952) desenvolveu um conceito bem interessante de Análise de Risco, depois a precificação de Black-Scholes, o Modelo de Opções

² A quebra global do sistema de taxas de câmbio fixo levou vários países a adotarem o câmbio flutuante neste ano.

³ O choque dos preços do petróleo gerou uma recessão global acompanhado de inflação elevada e fortes oscilações nas taxas de juros.

⁴ Ações norte-americanas caíram 23% gerando um prejuízo de 1 trilhão de dólares

Binomial e outras ferramentas apareceram para auxiliar a mensurar o Risco. Mas em 1993 surgiu uma metodologia que iria centralizar todos os riscos em uma só medida: o *Value at Risk*.

4 **VALUE AT RISK**

O *Value at Risk* (*VaR*) surgiu para quantificar o Risco de Mercado, um tipo de risco financeiro. A desvantagem de abordar e quantificar apenas um tipo de risco financeiro é que algumas instituições financeiras podem retirar o capital do investimento mensurável, visando migrar para investimentos cujo risco é mais difícil de se avaliar e monitorar. Por isso, surgiram formas de integrar a Gestão do Risco da empresa como um todo, envolvendo todos os Riscos Financeiros.

Atualmente, o *VaR* é considerado a melhor ferramenta de Gestão do Risco. Tanto que Jorion (2001), um estudioso consagrado da Gestão de Risco (especialmente do *VaR*), evidencia a importância e a utilização de tal ferramenta no dia-a-dia de uma instituição financeira escrevendo em seu livro *Value at Risk: The New Benchmark for Managing Financial Risk* (p. 21):

Every morning, Lesley Daniels Webster, head of market risk at Chase Manhattan Bank, receives a neat 30-page report that summarizes the value at risk (VaR) of the bank. The document is generated during the night by computers that quantify the risk of all the trading positions of the bank.

Este trecho evidencia o *VaR* como parte corriqueira da Análise de Risco de Mercado por instituições financeiras privadas, sendo a sua ferramenta mais utilizada hoje em dia por bancos, corretoras mobiliárias, fundos de investimento e até por instituições não financeiras. Sempre tendo em vista ajustar os seus riscos financeiros, maximizando a quantidade de dinheiro investida, com risco contornáveis e buscando não deixar o dinheiro “parado”.

Além do ramo privado, entidades reguladoras dos bancos comerciais, mercados mobiliários e seguradoras⁵ também encaram o *VaR* como endosso para Gestão do Risco. A metodologia do VaR, uma abordagem estatística para gerenciar o risco financeiro, é aceita mundialmente pelos órgãos reguladores dos países que assinaram o Acordo de Basileia de 1988 e suas emendas subsequentes.

De maneira intuitiva, o *VaR* pode ser definido como a sumarização da pior perda sobre um horizonte de tempo pré-determinado dado um nível de confiança.

De maneira mais formal, podemos encarar que o *VaR* descreve o quantil da distribuição de perdas e ganhos (retorno) em um horizonte de tempo. Se c é o intervalo de confiança selecionado, *VaR* corresponde ao nível $(1 - c)$ da cauda inferior. Por exemplo, com um nível de confiança de 95% , o *VaR* deve ser tal que exceda 5% do número total de observações da distribuição.

A maior utilização do *VaR* no mercado, sua grande utilização na Academia e fora dela, pode ser explicada por dois grandes motivos: facilidade de entendimento do cálculo a ser feito e fornecimento do resultado em uma medida, de forma clara e simples. Um gerente de um fundo de investimentos pode agendar uma reunião com os seus investidores, sendo que eles podem ter conhecimentos técnicos e estatísticos baixíssimos, e lhes comunicar que: *sob condições normais do mercado, o máximo que este fundo pode perder no próximo mês é próximo de 1 milhão de reais com um nível de confiança de 95%*. Portanto, o *VaR* descreve a fronteira de perdas potenciais, de forma rápida e didática.

⁵ No Brasil temos o Banco Central, a Comissão de Valores Mobiliários e a Superintendência de Seguros Privados, respectivamente.

4.1 Metodologias de Cálculo do VaR

Surgiram diferentes metodologias para mensurar o VaR no decorrer dos anos. Os dados brutos podem ser difíceis de serem obtidos em alguns casos, mas neste estudo os dados (valores de fechamento de ações negociadas na BOVESPA) estão disponíveis facilmente e não apresentam ruídos ou erros de cálculo. Por se tratar de um cálculo probabilístico existem duas abordagens tradicionais: a paramétrica e a não-paramétrica. Apesar disto, ambas as abordagens seguem uma esquema parecido:

1. Seleção do horizonte de tempo e do nível de confiança;
2. Seleção de fatores relevantes do mercado;
3. Mapeamento dos Riscos;
4. Cálculo do VaR - da pior perda possível.

Para o passo (1), definimos os parâmetros relevantes de acordo com nossas metas e recursos. Nos próximos dois passos, (2) e (3), assume-se algum tipo de modelo, seja apenas um conjunto de fatores relevantes ou um modelo de preços completamente especificadas. Em qualquer caso, um conjunto relativamente pequeno de parâmetros relevantes deve ser definido, e algum método para a avaliação do *portfolio* com base neste conjunto deve ser estabelecido. O passo (4) inclui o cálculo em si, um passo que pode ser muito demorado, especialmente quando empregamos métodos de Simulações de Monte-Carlo (que será melhor definido logo adiante).

É bom salientar que o esquema acima é para o cálculo do VaR de maneira geral. Neste estudo, que visa calcular o VaR de uma carteira de ações, os passos (2) e (3) se referem à hipótese de um determinado modelo para os retornos e da interação entre os elementos da carteira em questão.

São 3 metodologias usuais para se calcular o VaR , que apresentam diferenças em alguns dos passos acima. O método *Paramétrico Clássico* difere dos métodos de *Simulações Históricas* e de *Simulações de Monte-Carlo*, que são métodos não-paramétricos. Outras metodologias existem, algumas semi-paramétricas, que são derivadas destas três grandes abordagens distintas. Porém, a metodologia por *Simulações Históricas* vem caindo em desuso, por ser uma metodologia de otimização linear sem a utilização de variáveis aleatórias. Justamente por isto este estudo visa explicar o *Método de Simulações de Monte Carlo* e o *Método Paramétrico*.

4.2 Simulações de Monte-Carlo

Este método não-paramétrico é a abordagem mais utilizada por usuários experientes por não assumir distribuições específicas para os retornos. O primeiro passo é identificar os importantes fatores de mercado para, em seguida, construir uma distribuição conjunta destes fatores baseado em dados históricos, dados implícitos nos preços observados e dados baseados em cenários econômicos específicos. E para finalizar, simula-se uma enorme quantidade de cenários, obtendo-se um retorno para cada cenário. A partir deste ponto, ordena-se os retornos e o quantil de 5% dos piores resultados é a estimativa do VaR .

Como qualquer metodologia, apresenta vantagens e desvantagens. Seus pontos fortes são o fato de não assumir um modelo específico para os retornos e poder incluir facilmente instrumentos não-lineares. Já a sua principal desvantagem é a convergência muito lenta. Sabe-se que Simulações de Monte Carlo podem convergir para os seus valores verdadeiros no máximo a $\frac{1}{\sqrt{N}}$, onde N é o total de trajetórias simuladas. Isto significa que para aumentar a precisão em um fator de 10, deve-se realizar pelo menos 100 simulações a mais. Outro problema, que surge antes da simulação em si, é a obtenção da distribuição conjunta dos fatores de mercado. Afinal, uma distribuição envolvendo mais de 2 fatores, correlacionados entre si, é uma tarefa complexa e penosa.

Porém o método de Monte-Carlo também pode ser combinado com resultados preliminares de outros métodos. A Simulação histórica pode dar uma primeira aproximação às distribuições, enquanto os métodos paramétricos podem indicar aproximações ainda melhores e evidenciar as correlações entre os fatores, etc. E é justamente por esta possibilidade que o Método de Monte-Carlo é o mais usual no mercado financeiro.

4.3 Método Paramétrico

O cálculo do VaR pode ser simplificado consideravelmente se assumirmos que o retorno da carteira (ou ativo) escolhida pertença a determinada família de distribuições. Neste caso

$$P(z_t < -VaR_t | \Psi_{t-1}) = \theta \quad (4.1)$$

onde $\{z_t\}_{t=1}^T$ é uma série de retornos, T é o tamanho da amostra, Ψ_{t-1} é a informação até o instante $t - 1$ e $(1 - \theta)$ é o nível de confiança dado para calcular o *VaR*. Ao assumir uma determinada distribuição para z_t a equação (4.1) fica bem simples de ser resolvida, seja de maneira analítica ou computacional, revelando assim o *VaR* estimado.

Esta metodologia é também muito utilizada, existindo um amplo debate sobre os seus resultados. Este debate decorre do fato que várias abordagens paramétricas assumem a distribuição como Normal, uma distribuição que não apresenta caudas pesadas, e as caudas pesadas são características de séries temporais financeiras. Existem abordagens para se contornar este problema, principalmente ao assumir diferentes distribuições ou aplicar uma transformação na variável.

Observa-se na equação (4.1) que é necessário ter a série de retornos para o cálculo do *VaR*. Através dos dados históricos de uma série de preços de ativos pode-se obter a distribuição do seu retorno. Esta área da estatística, Séries Temporais, apresenta resultados interessantes e aplicações em inúmeras áreas (como demografia, ecologia, meteorologia, finanças, etc). Neste estudo será utilizado uma metodologia bem conhecida no estudo de Séries Temporais Financeiras para se modelar a volatilidade do retorno, os modelos ARCH e GARCH.

5 SÉRIES TEMPORAIS FINANCEIRAS

O primeiro problema ao tentar estimar o *VaR* de uma carteira de ações é a forma de analisar os dados. Hoje em dia, a coleta destes dados é facilmente obtida pela internet, podendo escolher o período desejado em datas de 10 anos atrás até o dia hoje. Esta facilidade deixa mais tempo para analisar o comportamento da série em si, suas características e propriedades, em vez de se preocupar demasiadamente com a qualidade dos dados e a coleta. Portanto, fica claro que modelar o comportamento univariado da PETR4 e da VALE3 (as duas ações envolvidas neste estudo) nada mais é do que a modelagem de duas séries temporais financeiras. Antes de se modelar os dados reais e apresentar quaisquer resultados é importante evidenciar alguns conceitos sobre séries temporais financeiras⁶, alguns métodos mais utilizados e como trabalhar com essa gama de dados tão importante.

Dados financeiros podem ser obtidos entre intervalos de tempo igualmente espaçados (forma mais usual) ou intervalos irregularmente espaçados. Neste estudo, iremos utilizar dados do preço de fechamento diário das ações na BOVESPA. Como pode ser visto em Morettin e Toloi (2006), a modelagem de séries financeiras não envolve a modelagem dos preços de fechamento nem da modelagem das diferenças entre preços subsequentes ($P_t - P_{t-1}$). Na literatura da área já se verificou que,

⁶Alguns conceitos básicos, porém fundamentais de sobre séries temporais, como estacionariedade e autocovariância, não serão definidos neste estudo e podem ser encontrados em Morettin e Tolói (2006).

devido à propriedades interessantes de normalidade e pelo fato de ficarem centrados em zero, o melhor tipo de dado para se modelar é o *log-retorno*⁷ :

$$r_t = \ln \frac{P_t}{P_{t-1}} \quad (5.1)$$

Contudo, nos deparamos com uma característica muito presente ao tratar de séries de ativos financeiros: a *volatilidade*. A *volatilidade* é a variância condicional de uma variável (neste caso, o retorno) e modelá-la é o principal desafio de modelar uma série temporal financeira. Técnicas utilizadas no estudo de séries temporais de qualquer área são empregadas no caso de séries temporais financeiras, como os procedimentos EWMA, *RiskMetrics*, modelos ARMA, etc. (estes assuntos podem ser aprofundados no livro de Morettin e Tolo (2006)). Neste estudo serão modeladas as séries temporais do retorno de duas ações (PETR4 e a VALE3) através dos métodos ARCH e GARCH, que visam modelar a volatilidade condicional variando no tempo.

5.1 Modelos ARCH

Os modelos ARCH foram introduzidos por Engle, em 1982, e são modelos não-lineares no que se refere à variância. O objetivo aqui será modelar a volatilidade de um retorno cuja série é dada por,

$$r_t = \ln \frac{P_t}{P_{t-1}} = \ln P_t - \ln P_{t-1}$$

⁷Esta definição é a mais usual e será doravante chamada apenas de *retorno*

e sejam

$$\mu_t = E(r_t | \Psi_{t-1}) \quad (5.2)$$

$$h_t = Var(r_t | \Psi_{t-1}) \quad (5.3)$$

respectivamente a média e a variância condicional de r_t , onde Ψ_{t-1} denota a informação até o instante $(t - 1)$, isto é, r_1, r_2, \dots, r_{t-1} . Se supormos que $\mu_t = 0$ então teremos que $h_t = E(r_t^2 | \Psi_{t-1})$.

Os modelos auto-regressivos com heteroscedasticidade condicional (ARCH) surgiram com o objetivo de estimar a variância da inflação. O cerne do seu conceito é que o retorno r_t é não-correlacionado serialmente, mas a volatilidade depende de retornos passados por meio de uma função quadrática. Este debate sobre correlação, auto-correlação e estimação de preços futuros de séries financeiras vinha se acirrando desde o começo dos anos de 1970, quando surgiu a Teoria dos Mercados Eficientes (por Fama, em 1970).

A Teoria dos Mercados Eficientes discorre sobre como os preços assimilam informações disponíveis a todos os participantes do mercado, de tal forma que as variações de preços (e conseqüentemente, dos retornos) sejam imprevisíveis. Dependendo da informação disponível aos participantes, existem três formas de eficiências:

1. *eficiência fraca*: os retornos não podem ser previstos a partir da informação prévia dos preços;
2. *eficiência semi-forte*: o conjunto de informações inclui a história passada dos preços mais toda a informação pública disponível;

3. *eficiência forte*: o conjunto de informações inclui toda a informação pública e privada.

Apesar de, à primeira vista, não parecer existir uma relação entre esta teoria e a modelagem estatística vê-se que ao afirmar que um retorno é imprevisível estamos afirmando que não é possível prever retornos “anormais”. Já a eficiência fraca implica que os retornos ao longo do tempo são independentes, isto é, ficaríamos limitados a um modelo $r_t = \mu + \epsilon_t$. Engle com os modelos ARCH desenvolveu uma metodologia que permite que os erros sejam não-correlacionados, porém dependentes. Logo, a $Cov(r_t^2, r_{t-1}^2)$ pode ser não nula.

Definição 5.1.1. *Um modelo ARCH(m) é definido por*

$$r_t = \epsilon_t \sqrt{h_t} \quad (5.4)$$

$$h_t = \alpha_0 + \alpha_1 r_{t-1}^2 + \dots + \alpha_m r_{t-m}^2 \quad (5.5)$$

onde ϵ_t *i.i.d.* com média zero, $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, \dots, m - 1$ e $\alpha_m > 0$.

Usualmente supõe-se que ϵ_t segue uma distribuição Normal padrão ou alguma outra distribuição que descreva melhor a cauda da série em questão. Outro ponto notório é que, pela própria definição, valores grande de r_t são seguidos por outros valores grandes. Para melhor entendimento dos modelos ARCH nota-se que a média e a variância incondicionais da série para $m = 1$ são:

- i. $E(r_t) = E[E(r_t | \Psi_{t-1})] = 0$;

$$\text{ii. } \text{Var}(r_t) = E(r_t^2) = E[E(r_t^2|\Psi_{t-1})] = E(\alpha_0 + \alpha_1 r_{t-1}^2) = \alpha_0 + \alpha_1 E(r_{t-1}^2) \text{ }^8$$

$$\Rightarrow \text{Var}(r_t) = \frac{\alpha_0}{1-\alpha_1}$$

$$\text{iii. } \text{Cov}(r_t, r_{t+k}) = E(r_t r_{t+k}) = E[E(r_t r_{t+k}|\Psi_{t+k-1})] = E[r_t E(r_{t+k}|\Psi_{t+k-1})] =$$

$$E[r_t E(\sqrt{h_{t+k}} \epsilon_{t+k}|\Psi_{t+k-1})] = 0, \text{ para } k > 0, \text{ pois } r_t \text{ está presente na informação}$$

$$\Psi_{t+k-1} \text{ e } E(\epsilon_{t+k}|\Psi_{t+k-1}) = 0.$$

Define-se a *função de auto-covariância* de um processo estacionário estrito como $\gamma(\tau) = \text{Cov}\{X(t), X(t + \tau)\}$. Por (iii) acima tem-se que, para o caso do modelo ARCH, $\gamma_r(k) = 0$, para $k \geq 1$. Isto indica que $\{r_t\}$ é uma sequência de variáveis não-correlacionadas com média zero e variância dada pelo item (ii) acima.

É possível demonstrar que o modelo ARCH apresenta curtose maior do que três, como pode ser visto em Morettin e Tolói (2006). Isto é, ao admitir que r_t siga um modelo ARCH tem-se um modelo cujas caudas são mais longas do que os da normal, algo extremamente vantajoso. A principal crítica é que a abordagem não faz distinção entre retornos positivos e negativos, já que são os quadrados dos retornos que entram na fórmula. Pois hoje em dia já se sabe que a volatilidade reage de maneira distinta a retornos positivos e negativos.

Já a estimação dos parâmetros do modelo ocorre usualmente através do Método de Máxima Verossimilhança Condicional⁹. Assumindo algumas distribuições (como a normal ou a t de Student) pode-se escrever explicitamente a Verossimilhança, de outras formas não. Mas em ambos os casos teremos que utilizar algum procedimento não-linear de otimização numérica para obter os parâmetros.

⁸se o processo $\{r_t\}$ for estacionário de segunda ordem, então, $E(r_{t-1}^2) = E(r_t^2) = \text{Var}(r_t)$

⁹Os algoritmos implementados atualmente baseiam-se no artigo de Bollerslev e Wooldridge (1992), modelo inicial do algoritmo QMLE, *Quasi Maximum Likelihood Estimation*

5.2 Modelos GARCH

Em 1987, Bollerslev sugeriu uma generalização do modelo ARCH, o modelo GARCH (*generalized auto-regressive conditional heteroskedasticity*), visando calcular o quadrado da volatilidade condicional no tempo t .

Definição 5.2.1. *Um modelo GARCH(m,n) é definido por*

$$r_t = \epsilon_t \sqrt{h_t} \quad (5.6)$$

$$h_t = \alpha_0 + \sum_{i=1}^m \alpha_i r_{t-i}^2 + \sum_{j=1}^n \beta_j h_{t-j} \quad (5.7)$$

em que ϵ_t i.i.d. com média zero; $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, \dots, m-1$, $\alpha_m > 0$; $\beta_j \geq 0$, $j = 1, \dots, n-1$, $\beta_m > 0$; $\sum_{i=1}^q \alpha_i + \beta_i$ e $q = \max(m, n)$.

Supondo que os ϵ_t seguem uma Normal padrão ou uma t de Student, assim como no modelo ARCH. É importante notar que o modelo GARCH($m,0$) é o modelo ARCH(m) (por isso que é considerado uma generalização), que considera também as volatilidades anteriores de maneira explícita no modelo. Na prática utiliza-se bastante o modelo GARCH($1,1$), cuja volatilidade é dada por

$$h_t = \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 h_{t-1} \quad (5.8)$$

As propriedades e características se assemelham bastante às do modelo ARCH: volatilidades altas precedidas de retornos grandes, caudas longas, etc. Duas dificuldades existem ao adotar este modelo: a definição da ordem do modelo (os valores de m e n) e a estimação dos parâmetros. Primeiramente, a identificação da

ordem de um modelo GARCH é usualmente difícil, realizando análises distintas em modelos de ordem baixa, para em seguida comparar os modelos com base em critérios de informação, função perda, log-verossimilhança ou critérios de informação. Já os estimadores dos parâmetros são obtidos assim como no modelo ARCH, através do método de Máxima Verossimilhança Condicional, que demanda métodos numéricos de otimização.

Ao estimar a distribuição dos retornos da PETR4 e da VALE3 através de métodos ARCH e GARCH pode-se calcular o *VaR* para cada retorno, basta utilizar a metodologia do *VaR* paramétrico. No entanto, o objetivo deste estudo é calcular o *VaR* do *Portfolio* e não de maneira individualizada. A correlação entre essas variáveis pode ser mensurada via coeficientes estatísticos (Pearson, Spearman, etc), mas neste estudo a correlação será abordada através da distribuição conjunta das marginais, através da Cópula.

6 CÓPULA

Em análise multivariada tenta-se transformar variáveis para diminuir a correlação, modelar a matriz de correlações e outras maneiras que atendem muito bem alguns problemas. Mas a pergunta básica que alguns estudantes de estatística (e até estatísticos formados), economistas e engenheiros que durante algum trabalho se deparam com uma combinação linear de variáveis aleatórias é: por que em todos os livros só encontramos a soma de variáveis aleatórias independentes? E se existisse uma maneira simples de se escrever a “correlação” entre duas, três, ou até n variáveis, de uma maneira mais simples e eficaz? Pois o coeficiente de correlação de Pearson não se porta bem frente à correlações não-lineares.

Após estimar a distribuição univariada dos retornos poderia-se simular os valores individuais dos ativos para se calcular o *VaR* para cada ação. Mas no caso deste *Portfolio* composto por dois ativos isto estaria incorreto, pois deve existir uma correlação entre a PETR4 e a VALE3. Uma questão bastante complexa para esta análise é de que maneira deve-se considerar a correlação, isto é: como considerar uma variável aleatória que é uma combinação linear de outras duas variáveis aleatórias, porém dependentes? Afinal, o coeficiente de Pearson já teve seus problemas apontados em vários estudos e não é a ferramenta mais recomendada. Como encarar a correlação, a soma de variáveis aleatórias dependentes e outras perguntas (que são muitas vezes ignoradas) podem ser respondidas e sintetizadas no conceito de Cópula¹⁰.

¹⁰O substantivo Cópula no latim significa vínculo, laço.

6.1 Conceitos Preliminares

Os conceitos iniciais, definições, lemas e teoremas encontrados neste capítulo podem ser aprofundados, revistos e suas demonstrações encontram-se no livros Nelsen (2006) e Cherubini et al. (2004). Como o *Portfolio* em questão é composto por dois ativos e para simplificar as demonstrações será mostrada a teoria de Cópula apenas para o caso bivariado. Porém, a sua extensão para n variáveis é clara. Primeiramente evidencia-se a notação que será utilizada a partir de agora. Denotaremos por \mathfrak{R} a reta real $(-\infty, +\infty)$, $\overline{\mathfrak{R}}$ denota a reta real estendida $[-\infty, +\infty]$ e $\overline{\mathfrak{R}}^2$ denota o plano real estendido $\overline{\mathfrak{R}} \times \overline{\mathfrak{R}}$. O produto cartesiano de dois intervalos fechados é um retângulo de $\overline{\mathfrak{R}}^2$, $B = [x_1, x_2] \times [y_1, y_2]$. Chamaremos de I^2 o quadrado unitário, isto é, o produto $I \times I$ onde $I = [0, 1]$. E sabe-se que uma função bivariada H é uma função cujo Domínio, $DomH$, é um subconjunto de $\overline{\mathfrak{R}}^2$ e tem como imagem (ImH) apenas um subconjunto de \mathfrak{R} .

Definição 6.1.1. *Sejam S_1 e S_2 subconjuntos não-vazios de $\overline{\mathfrak{R}}$, e seja H uma função bivariada tal que $DomH = S_1 \times S_2$. Seja um retângulo $B = [x_1, x_2] \times [y_1, y_2]$ cujos vértices pertencem à $DomH$, de tal forma que $x_1 \leq x_2$ e $y_1 \leq y_2$. Então o H -Volume de B é dado por*

$$V_H(B) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1)$$

Definição 6.1.2. *Uma função bivariada real H é 2-crescente se $V_H(B) \geq 0$ para todos os retângulos B cujos vértices pertencem ao $DomH$.*

Quando a função H é 2-crescente, muitos autores se referem ao H-Volume de B como H-Medida de B . Outro ponto interessante é notar que se H é 2-crescente, isto não implica que H é não-decrescente em cada argumento (as derivadas marginais podem ser negativas em algum intervalo do $DomH$).

Lema 6.1.1. *Sejam S_1 e S_2 subconjuntos não-vazios de $\overline{\mathfrak{R}}$, e seja H uma função 2-crescente com o domínio $S_1 \times S_2$. Então a função $t \mapsto H(t, y_2) - H(t, y_1)$ é não-decrescente em S_1 , e a função $t \mapsto H(x_2, t) - H(x_1, t)$ é não-decrescente em S_2 , desde que $x_1, x_2 \in S_1$ com $x_1 \leq x_2$ e que $y_1, y_2 \in S_2$ com $y_1 \leq y_2$.*

Definição 6.1.3. *Seja a_1 o menor elemento de S_1 e a_2 o menor elemento de S_2 . Dizemos que a função H de $S_1 \times S_2$ em \mathfrak{R} é limitada inferiormente se $H(x, a_2) = 0 = H(a_1, y)$ para todo $(x, y) \in S_1 \times S_2$.*

Lema 6.1.2. *Sejam S_1 e S_2 subconjuntos não-vazios de $\overline{\mathfrak{R}}$, e seja H uma função 2-crescente limitada inferiormente com $DomH = S_1 \times S_2$. Então H é não-decrescente em cada argumento.*

Lema 6.1.3. *Sejam S_1 e S_2 subconjuntos não-vazios de $\overline{\mathfrak{R}}$, e seja H uma função 2-crescente limitada inferiormente, com marginais $F(x)$ e $G(y)$, cujo $DomH = S_1 \times S_2$. Sejam (x_1, y_1) e (x_2, y_2) quaisquer pontos pertencentes a $S_1 \times S_2$. Logo*

$$|H(x_2, y_2) - H(x_1, y_1)| \leq |H(x_2, y_2) - H(x_1, y_2)| + |H(x_1, y_2) - H(x_1, y_1)|$$

6.2 Cópula Bivariada

Antes da definição de Cópula define-se a subcópula a partir de elementos vistos anteriormente, como funções 2-crescentes e funções limitadas inferiormente.

A partir deste ponto tem-se subsídios para expandir o conceito, abranger a noção da Cópula Bivariada e explorar algumas de suas propriedades.

Definição 6.2.1. *Uma subcópula bidimensional, ou bivariada, é uma função C' com as seguintes propriedades:*

- i. $DomC' = S_1 \times S_2$, onde S_1 e S_2 são subconjuntos de I contendo 0 e 1;*
- ii. C' é limitada inferiormente e 2-crescente;*
- iii. Para todo $u \in S_1$ e todo $v \in S_2$ teremos*

$$C'(u, 1) = u \text{ e } C'(1, v) = v$$

Observe que, dadas estas propriedades, para todo $(u, v) \in DomC'$, $0 \leq C'(u, v) \leq 1$, e assim a Imagem de C' é subconjunto de I . Agora pode-se estender a definição para o entendimento do que é uma Cópula.

Definição 6.2.2. *Uma Cópula bidimensional (bivariada) é uma subcópula C cujo domínio é o próprio I^2 . Isto é, é uma função C que vai de I^2 para I e tem as seguintes propriedades:*

- i. Para todo $u, v \in I$,*

$$C(u, 0) = 0 = C(0, v)$$

$$C(u, 1) = u$$

$$C(1, v) = v;$$

ii. Para todo u_1, u_2, v_1 e $v_2 \in I$ tal que $u_1 \leq u_2$ e $v_1 \leq v_2$,

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$$

A princípio, a distinção entre Subcópula e Cópula pode parecer algo de menor importância, mas ao estudar o Teorema de Sklar mais a frente nota-se que esta diferença é fundamental. A partir do Teorema de Sklar pode-se verificar propriedades fundamentais da Cópula e a sua relação com o conceito de Variáveis Aleatórias.

Teorema 6.2.1 (Limites de *Fréchet-Hoeffding*). *Seja C' uma subcópula. Então para todo $(u, v) \in \text{Dom}C'$,*

$$\max(u + v - 1, 0) \leq C'(u, v) \leq \min(u, v)$$

Como toda Cópula é uma Subcópula, o teorema acima vale para Cópulas. Sendo assim, para toda Cópula C e todo $(u, v) \in I^2$,

$$\max(u + v - 1, 0) = W(u, v) \leq C(u, v) \leq M(u, v) = \min(u, v)$$

A desigualdade acima é a versão de Cópulas para os limites de *Fréchet-Hoeffding*. Outra Cópula fundamental, que carrega em si o conceito de independência, é a Cópula Produto $\Pi(u, v) = uv$. Outro ponto notável das Subcópulas (e Cópulas por implicação) é a continuidade, estabelecida pelo seguinte teorema:

Teorema 6.2.2. *Seja C' uma subcópula. Então para todo $(u_1, v_1), (u_2, v_2) \in \text{Dom}C'$,*

$$|C'(u_2, v_2) - C'(u_1, v_1)| \leq |u_2 - u_1| + |v_2 - v_1|$$

Definição 6.2.3. *Seja C uma cópula e $a \in I$. A secção horizontal de C em a é uma função de I em I dada por $t \mapsto C(t, a)$; a secção vertical de C em a também é uma função de I em I dada por $t \mapsto C(a, t)$; já a secção diagonal de C é uma função $\delta_c(t)$ de I em I definida por $\delta_c(t) = C(t, t)$.*

Corolário 6.2.1. *As secções vertical, horizontal e diagonal de uma cópula são todas não-decrescentes e contínuas em I .*

É bem interessante verificar que da própria definição de Cópula e de sua continuidade tem-se que o seu gráfico é uma superfície contínua no cubo I^3 . E também, do teorema que se refere aos Limites de *Fréchet-Hoeffding*, verifica-se também que este gráfico está limitado pelas superfícies $z_{inf} = M(u, v)$ e $z_{sup} = W(u, v)$.

Teorema 6.2.3. *Seja C uma cópula. Para qualquer $v \in I$, a derivada parcial $\frac{\partial C(u,v)}{\partial u}$ existe para todo u . Além disso, para tais v e u , $0 \leq \frac{\partial C(u,v)}{\partial u} \leq 1$.*

Semelhantemente, Para qualquer $u \in I$, a derivada parcial $\frac{\partial C(u,v)}{\partial v}$ existe para todo v . Além disso, para tais v e u , $0 \leq \frac{\partial C(u,v)}{\partial v} \leq 1$.

Isto é, as funções $u \mapsto \frac{\partial C(u,v)}{\partial u}$ e $v \mapsto \frac{\partial C(u,v)}{\partial v}$ são bem definidas e não decrescentes em I .

6.3 Teorema de Sklar e Variáveis Aleatórias

Apesar da utilização da Cópula em simulações, na área financeira e nos estudos estatísticos vir se intensificando nos últimos 20 anos para cá o seu conceito já data de mais de 50 anos. O Teorema de Sklar é o centro da teoria de Cópulas, estabelecendo a relação entre a função de distribuição multivariada e suas marginais. Porém, para melhor entendimento, é necessário relembrar alguns conceitos de probabilidade de variáveis aleatórias¹¹.

Definição 6.3.1. *A Função de Distribuição da Variável Aleatória X , representada por F_x (em alguns casos apenas por F), é definida por*

$$F_x = P(X \leq x), \quad x \in \mathfrak{R}$$

onde

- i. F_x é não-decrescente;*
- ii. $F(-\infty) = 0$;*
- iii. $F(+\infty) = 1$.*

Definição 6.3.2. *A variável aleatória X é absolutamente contínua se existe uma função $f(x) \geq 0$ tal que*

$$F_x = \int_{-\infty}^x f(t)dt, \quad \forall x \in \mathfrak{R}$$

¹¹Serão colocadas as definições e resultados para variáveis aleatórias contínuas, mas para variáveis discretas a implicação é análoga

Definição 6.3.3. *Uma função de distribuição conjunta bivariada é uma função H com o domínio em $\overline{\mathbb{R}^2}$ tal que*

- i. H é 2-crescente;*
- ii. $H(x, -\infty) = 0 = H(-\infty, y)$;*
- iii. $H(-\infty, \infty) = 1$.*

Portanto H é limitada inferiormente, e tem marginais F e G dadas por $F(x) = H(x, \infty)$ e $G(y) = H(\infty, y)$, que são as funções de distribuição univariadas F_x e F_y , respectivamente.

Teorema 6.3.1 (Teorema de Sklar Bidimensional). *Seja H uma função de distribuição conjunta bivariada com marginais F e G . Então existe uma Cópula C tal que, para todo $x, y \in \overline{\mathbb{R}}$,*

$$H(x, y) = C(F(x), G(y)) \quad (6.1)$$

E se F e G são absolutamente contínuas, então a cópula C é única. Caso contrário, C é unicamente determinada em $Im F \times Im G$.

De maneira recíproca, se C é uma Cópula e F e G são funções de distribuição, então a função H é a função de distribuição conjunta com marginais F e G .

A Cópula acima, será de agora em diante chamada de Cópula de X e Y , e denotada por C_{XY} .

A partir dos conceitos anteriores e do Teorema de Sklar tem-se subsídios para verificar algumas propriedades envolvendo Cópulas e Variáveis Aleatórias. A

Cópula produto, por exemplo, caracteriza variáveis aleatórias independentes quando as funções de distribuição são contínuas.

Teorema 6.3.2. *Sejam X e Y variáveis aleatórias contínuas. Então X e Y são independentes se, e somente se, $C_{XY} = \pi(F(x), G(y)) = F(x)G(y)$.*

Outra propriedade da Cópula é a sua invariância quando são feitas transformações estritamente monótonas de variáveis aleatórias, como veremos no teorema seguinte¹².

Teorema 6.3.3. *Sejam X e Y variáveis aleatórias contínuas com Cópula C_{XY} . Se h e g são funções estritamente crescentes na $Im X$ e $Im Y$, então $C_{h(X)g(Y)} = C_{XY}$.*

Este teorema significa que C_{XY} é invariante sobre transformações estritamente crescente de X e Y .

Teorema 6.3.4. *Sejam X e Y variáveis aleatórias contínuas com Cópula C_{XY} . Se h e g são funções estritamente monótonas na $Im X$ e $Im Y$, então*

i. Se h é estritamente crescente e g é estritamente decrescente,

$$C_{h(X)g(Y)}(u, v) = u - C_{XY}(u, 1 - v)$$

ii. Se h é estritamente decrescente e g é estritamente crescente,

$$C_{h(X)g(Y)}(u, v) = v - C_{XY}(1 - u, v)$$

¹²se a função de distribuição de uma variável aleatória é contínua, e se h é uma função estritamente monótona cujo domínio está contido na $Im X$, então a função de distribuição da variável aleatória $h(X)$ também é contínua

iii. Se h e g são estritamente decrescente,

$$C_{h(X)g(Y)}(u, v) = u + v - 1 + C_{XY}(1 - u, 1 - v)$$

Com a informação até agora sobre Cópula pode-se verificar a sua forte ligação com a Probabilidade (e por consequência com a Estatística), principalmente através do Teorema de Sklar. Vários estudos surgiram sobre o assunto, evidenciando famílias de Cópulas que se tornaram mais corriqueiras nos estudos sobre o assunto. As duas principais famílias de Cópulas são a família de Cópulas Elípticas e a família de Cópulas Arquimedianas, estudadas nas seções seguintes.

6.4 Cópulas Elípticas

As distribuições Normal, de Cauchy, Logística, de Laplace, t de Student e outras pertencem à classe de Distribuições Elípticas. São distribuições que generalizam a distribuição Normal Multivariada, e assim tem propriedades comuns muito interessantes¹³ que facilitam análises e interpretações mais complexas de maneira mais simples.

Cópulas Elípticas são simplesmente a Cópula de Distribuições Elípticas: distribuições que podem ser escritas na forma $f(x, \tau) = h(x)t(\tau) \exp q(x)s(\tau)I_a(x)$, onde $h(\cdot)$, $t(\cdot)$, $q(\cdot)$ e $s(\cdot)$ são funções conhecidas, e $I_a(x)$ é o suporte da função (que não pode depender de τ). As distribuições elípticas são amplamente estudadas (especialmente em Modelos Lineares Generalizados) e mais informações podem ser

¹³A propriedade mais conhecida de tal grupo é que a combinação linear de distribuições elípticas também é uma distribuição elíptica

encontradas em McCullagh e Nelder (1989). O Teorema de Sklar torna o entendimento e o estudo dessas cópulas menos complexo. Assim como no estudo univariado das distribuições, as Cópulas Elípticas mais utilizadas são a Gaussiana e a t de Student.

A **Cópula Gaussiana** (ou Normal) é a Cópula construída da Normal Multivariada através de mudanças de variáveis. Sabe-se que um vetor aleatório $X = (X_1, X_2, \dots, X_n)$ é uma normal multivariada se as marginais univariadas são Gaussianas, e tem-se uma matriz de correlações entre as distribuições univariadas dada por R ¹⁴. Como visto em Nelsen (2006), é possível provar que a estrutura de dependência entre as marginais é descrita por uma única Cópula:

$$C_R^{Ga}(u_1, \dots, u_n) = \phi_R(\phi^{-1}(u_1), \dots, \phi^{-1}(u_n)) \quad (6.2)$$

onde ϕ_R é a densidade da Normal Padrão Multivariada com matriz de correlação linear R e ϕ^{-1} é a inversa da densidade da Normal Padrão Univariada.

Se $n = 2$, a equação acima (6.2) pode ser reescrita como:

$$C_R^{Ga}(u, v) = \int_{-\infty}^{\phi^{-1}(u)} \int_{-\infty}^{\phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-R_{12}^2}} \exp\left\{-\frac{s^2 - 2R_{12}st + t^2}{2(1-R_{12}^2)}\right\} ds dt \quad (6.3)$$

onde R_{12} é o coeficiente de correlação linear entre as duas variáveis.

¹⁴No caso em que R é uma matriz diagonal, a normal multivariada apresenta marginais independentes

A **Cópula t de Student** é a Cópula da t Multivariada. Isto é, seja X um vetor com distribuição n -variada t -Student com ν graus de liberdade, vetor de médias μ e matriz de covariâncias $\frac{\nu}{\nu-2}\Sigma$, isto é,

$$X \stackrel{d}{=} \mu + \frac{\sqrt{\nu}}{\sqrt{S}}Z$$

em que $\mu \in \mathfrak{R}^n$, $S \sim \mathcal{X}_\nu^2$ e o vetor aleatório $Z \sim N_n(0, \sigma)$ são independentes.

A Cópula deste vetor X pode ser escrita como:

$$C_{\nu,R}^t(u_1, \dots, u_n) = t_{\nu,R}^n(t_\nu^{-1}(u_1), \dots, t_\nu^{-1}(u_n)) \quad (6.4)$$

em que: $R_{ij} = \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}}$; $t_{\nu,R}^n$ é a densidade multivariada do vetor aleatório $Z \frac{\sqrt{\nu}}{\sqrt{S}}$; a variável aleatória S e o vetor aleatório Z são independentes; e t_ν denota as marginais de $t_{\nu,R}^n$.

No caso bivariado, a equação acima (6.4) pode ser reescrita como:

$$C_{\nu,R}^t(u, v) = \int_{-\infty}^{t_\nu^{-1}(u)} \int_{-\infty}^{t_\nu^{-1}(v)} \frac{1}{2\pi\sqrt{(1-R_{12}^2)}} \left\{ 1 + \frac{s^2 - 2R_{12}st + t^2}{\nu(1-R_{12}^2)} \right\} ds dt \quad (6.5)$$

onde R_{12} é o coeficiente de correlação linear da distribuição t -Student bivariada com ν graus de liberdade.

Juntamente com as Cópulas Elípticas, as Cópulas Arquimedianas também são muito utilizadas e estudadas, devido à grande variedade de possibilidades para montá-las, como pode-se ver a seguir.

6.5 Cópulas Arquimedianas

Uma Cópula Arquimediana pode ser escrita da seguinte maneira:

$$C(u_1, \dots, u_n) = \Psi^{-1}[\Psi(u_1) + \dots + \Psi(u_n)] \quad (6.6)$$

para todo $0 \leq u_1, \dots, u_n \leq 1$ e onde Ψ é a função geradora.

Por sua vez, a função geradora tem que satisfazer três critérios:

- i. $\Psi(1) = 0$;
- ii. para todo $t \in (0, 1)$ temos que $\Psi(t)' < 0$, isto é, Ψ é decrescente;
- iii. para todo $t \in (0, 1)$ temos que $\Psi(t)'' \geq 0$, isto é, Ψ é convexa.

A Cópula produto, vista anteriormente, é um exemplo clássico de Cópula Arquimediana. Neste caso a função geradora é $\Psi(t) = -\ln(t)$, e portanto $C(u_1, u_2) = u_1 \cdot u_2$. Outras três funções geradoras muito utilizadas são $\Psi(t) = t^{-\alpha} - 1$ (para $\alpha > 0$), $\Psi(t) = (-\ln t)^\alpha$ (para $\alpha \geq 1$) e $\Psi(t) = -\ln \frac{e^{-\alpha t} - 1}{e^{-\alpha} - 1}$ (para $\alpha \in \mathfrak{R}$). Estas funções geradoras geram, respectivamente, as Cópulas de Clayton, Gumbel e Frank, no caso bivariado.

Até agora pode-se verificar a relação existente entre o conceito de Cópulas, Distribuição Conjunta e função de distribuição univariada. Viu-se também como juntar tudo isto através do Teorema de Sklar, além da introdução de algumas famílias clássicas de Cópula. Isto gera insumos para discutir melhor a relação de dependência entre variáveis aleatórias, que será fundamental para a verificação do comportamento

conjunto das variáveis de um *portfolio* e como simulá-las a fim de obter o *VaR* da carteira em questão.

6.6 Medidas de Associação

De maneira geral, duas variáveis aleatórias X e Y são ditas associadas quando elas não são independentes, isto é, quando a função de distribuição conjunta não é produto das marginais. Porém, graças a um problema de nomenclatura, traduções e mau-uso de termos tornou-se confuso utilizar certos conceitos para estudar a associação entre variáveis aleatórias.

A concordância, a correlação linear e a dependência caudal, são os conceitos mais amplamente utilizados e bem distintos entre si. Já as medidas destes conceitos apresentam características únicas, sendo cada uma melhor para verificar exatamente cada tipo de associação para qual foi criado (tau de Kendall, rho de Spearman, coeficiente linear de Pearson, etc). Todas estas medidas podem ser relacionadas ao conceito de Cópula, já que a Cópula é uma função de distribuição conjunta, contendo toda a relação entre as variáveis (mesmo que a mensuração em si não apareça de maneira direta). Além do mais, como visto na seção anterior, a cópula é invariante sobre transformações estritamente crescente. Para fins didáticos, de agora em diante, assume-se que X e Y são variáveis aleatórias contínuas.

O conceito de **concordância**, de maneira bem simples, visa mensurar o fato que a probabilidade de ter valores “elevados” tanto de X quanto de Y seja alta, enquanto a probabilidade de ter valores elevados de X e pequenos de Y seja baixo. As

propriedades axiomáticas de uma medida de concordância entre variáveis aleatórias X e Y encontram-se no artigo do Scarsini (1984) e são:

Definição 6.6.1. $M_{X,Y} = M_C$ é uma medida de concordância entre as variáveis aleatórias X e Y - com cópula C - se

- é definida para todo par de variáveis aleatórias;
- é uma medida relativa, isto é, $M_{X,Y} \in [-1, 1]$;
- é simétrica, isto é, $M_{X,Y} = M_{Y,X}$;
- se X e Y são independentes, então $M_{X,Y} = 0$;
- $M_{-X,Y} = M_{X,-Y} = -M_{X,Y}$;
- se (X_n, Y_n) é uma sequência de variáveis aleatórias contínuas com Cópulas C_n , e $\lim_{n \rightarrow \infty} C_n(v, z) = C(v, z)$ para todo $(v, z) \in I^2$ então $\lim_{n \rightarrow \infty} M_{X_n, Y_n} = M_{X,Y}$;
- se $C_1 \prec C_2$, então $M_{C_1} \leq M_{C_2}$.

Dois pontos são importantes salientar sobre Medidas de Concordância:

- Independência é condição suficiente, mas não necessária para que ela seja igual a zero;
- São distintas de Medidas de Dependência (como as Medidas de Dependência de Rényi). As Medidas de Dependência assumem o valor mínimo quando X e Y são independentes, não quando são contramonotônicos.

O Coeficiente **Tau de Kendall** - τ - é uma estatística utilizada para se mensurar a associação entre duas quantidades mensuradas. Em vez de utilizar a metodologia usual de cálculo, do número de pares concordantes, este coeficiente pode ser obtido diretamente de uma Cópula:

Definição 6.6.2. *O τ de Kendall entre dois vetores aleatórios X e Y com cópula C é*

$$\tau = 4 \int \int_{I^2} C(v, z) dC(v, z) - 1 \quad (6.7)$$

O **Rho de Spearman** - ρ_s - é uma estatística não-paramétrica que mede a dependência estatística entre duas variáveis. Ela indica o quanto que as variáveis em questão podem ser descritas através de uma função monotônica. Em vez de se utilizar a metodologia corriqueira que considera os postos e varia se existem empates ou não, este coeficiente também pode ser obtido diretamente de uma Cópula:

Definição 6.6.3. *O ρ_s entre dois vetores aleatórios X e Y com cópula C é*

$$\rho_s = 12 \int \int_{I^2} vz dC(v, z) - 3 \quad (6.8)$$

Outras maneiras de mensurar o Tau de Kendall e o Rho de Spearman a partir de Cópulas¹⁵ podem ser vistas em Nelsen (1991). Além disso, pode-se relacionar a Cópula com os conceitos de Dependência Linear, Dependência do Quadrante Positivo e Dependência Caudal de maneira bem interessante, também descritas em Nelsen (1991) e Cherubini, Luciano e Vecchiato (2004).

¹⁵Existem maneiras em que se considera as derivadas parciais da Cópula, ou a Cópula de Sobre-vivência, etc.

6.7 Teste de *Goodness-of-fit*

Para verificar a calibragem de um modelo de Cópula a uma conjunto de dados alguns estudos empregam o uso da Máxima Verossimilhança. Como visto em Nikoloulopoulos et al. (2011) este mecanismo de seleção não garante um ajuste global ótimo, falhando nos valores extremos, pois pode ocorrer um ajuste muito bom nos pontos centrais da Cópula¹⁶. Nikoloulopoulos et al. (2011) propõe a utilização de um teste baseado em previsões, que não é objeto de estudo neste relatório, pois é muito recente e precisa ser melhor avaliado e implementado pela comunidade acadêmica.

Espera-se em um teste de ajuste, de *goodness-of-fit*, verificar a aderência do conjunto de dados a um modelo. No caso da Cópula, quer se testar a hipótese de que a sua cópula C pode ser representada por uma família específica de Cópulas. Isto é, o objetivo é testar $H_0 : C \in C_0$, onde C_0 é a família que se deseja verificar se a Cópula faz parte.

Genest et al (2009) apresenta 5 procedimentos bem novos baseados em postos (em ordenamento de observações) para se testar o ajuste de um conjunto de dados a uma família paramétrica de Cópulas. Seu artigo fundamenta-se em inúmeras simulações de Monte Carlo e mostra evidências do poder dos testes apresentados, não deixando dúvidas sobre a eficiência de tal procedimento (estes testes podem ser utilizados em cópulas n-variadas). Os testes propostos são versões dos famosos

¹⁶Esta precipitação foi verificada neste estudo, onde entre 6 cópulas estudadas, a que apresentou melhor máxima verossimilhança foi a Independente! Enquanto é bem claro que as duas ações escolhidas, PETR4 e VALE3 possuem dependência muito forte. Isto ocorreu pois os valores centrais da Cópula fizeram muito efeito no cálculo da máxima verossimilhança.

testes de Cramér-von Mises e Kolmogorov-Smirnov aplicadas a Cópulas, envolvendo Cópulas empíricas e parâmetros de associação.

Definição 6.7.1. *A estatística de Kolmogorov-Smirnov é*

$$D_n = \sup_x |F_n(x) - F(x)| \quad (6.9)$$

Definição 6.7.2. *A estatística de Cramér-von Mises é*

$$\omega^2 = \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 dF(x) \quad (6.10)$$

Onde, para ambas as definições, $F_n(x)$ é a distribuição empírica dos dados que se deseja testar e $F(x)$ é a função de distribuição acumulada com que se deseja realizar a comparação no teste. Sabem-se as distribuições destas estatísticas, sendo possível calcular os p-valores dos testes e se obter conclusões.

Como visto acima, a versão conhecida destes testes envolve a distribuição empírica e a distribuição que se deseja testar. A Cópula é uma forma de representar a distribuição conjunta de duas variáveis, assim como uma família de Cópulas representa uma família de distribuições com características semelhantes. Portanto, Genest et al. (2009) apresenta a versão do teste para Cópulas e ainda constrói o algoritmo para realizar os 5 testes propostos, através de simulações de Monte Carlo.

O pacote *Copula* do *software* R, em sua última versão, contém este algoritmo que possibilitará o teste para verificar qual Cópula se ajustará melhor à carteira que possui os ativos da VALE3 e PETR4. Pode-se controlar o número de iterações da simulação de Monte Carlo, fixar valores para as famílias envolvidas e outros recursos.

Além disso o resultado do teste (cuja hipótese nula é que a Cópula segue uma família pré-determinada) apresenta o parâmetro da Cópula em questão juntamente com o que se encontra em qualquer teste de hipótese, o p-valor.

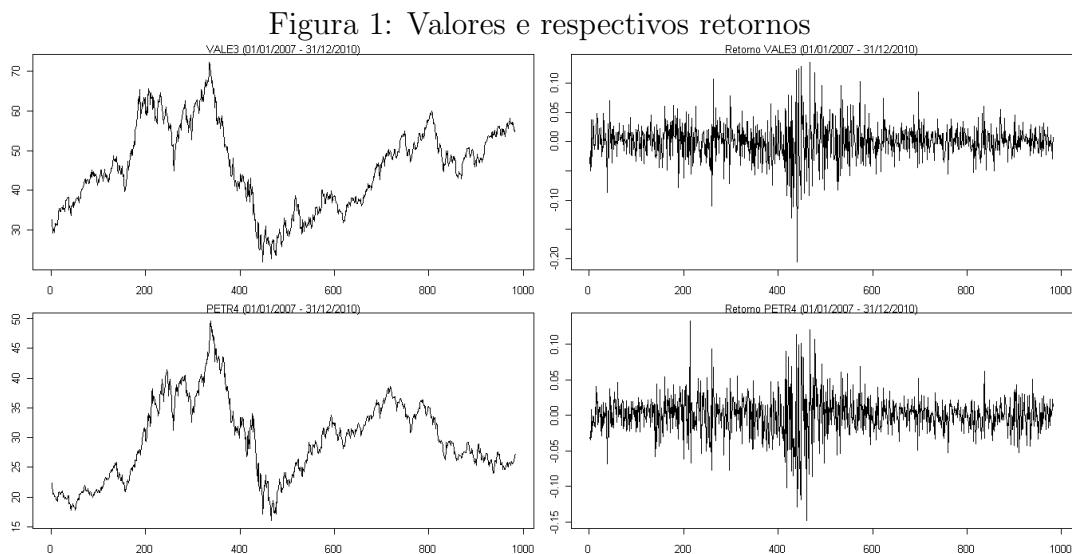
7 METODOLOGIA

A metodologia utilizada na análise da carteira (PETR4 e VALE3) e obtenção do VaR é uma compilação de tudo que foi citado e explicado nas seções anteriores, através da implementação em pacotes no *software* R. Desde conceitos simples, como o porquê da utilização do *log-retorno* e não do preço do ativo, até temas mais densos ,como o Teorema de Sklar e testes de *goodness-of-fit*. Portanto, para melhora do entendimento e melhor assimilação dos resultados e próximos passos, pode-se resumir toda a metodologia em 5 grandes passos:

- i. Estimação dos retornos univariados através da metodologia GARCH, utilizando-se o pacote *fGarch*, escolhendo-se o melhor ajuste através do Critério de Akaike;
- ii. Obtenção das medidas de associação Kendall, Pearson e Spearman entre os resíduos dos modelos GARCH acima;
- iii. Construção de 6 Cópulas dos resíduos da carteira, utilizando as distribuições do modelo GARCH e as medidas de associação, através do pacote *Copula*;
- iv. Teste de Cramèr-von Mises para Cópulas proposto por Genest et al. (2009) para escolher a que melhor se adequa aos valores reais, através do pacote *Copula*;
- v. Simulação de valores para a Cópula escolhida através do pacote *Copula*, a fim de se obter percentis e calcular o VaR da carteira PETR4 e VALE3.

8 RESULTADOS

No intuito de verificar o comportamento das Cópulas, estudar os modelos propostos e avaliar a carteira bivariada, tomaram-se duas séries de retorno: PETR4 e VALE3 durante o 2º Governo Lula (01/01/2007 a 31/12/2010). As séries de preços de fechamento ajustados (cotados em R\$) foram obtidas em <http://br.finance.yahoo.com> e os retornos foram computados como o logaritmo da razão entre os preços de fechamento consecutivos. As séries em questão contém 983 observações. A figura 1 abaixo mostra os gráficos da série original e os respectivos retornos para cada um dos ativos.



A partir destes conjuntos de dados foi feita a estimação dos retornos univariados através da metodologia GARCH já explicitada anteriormente, de onde obteve-se as distribuições dos resíduos. Em um segundo passo, tendo como parâmetros as medidas de concordância já citadas, foram construídas Cópulas bivariadas entre os resíduos, comparadas e escolhida a de melhor ajuste.

8.1 GARCH dos retornos univariados

Com o auxílio do pacote *fGarch* do *software* R foi realizado um estudo para se obter a melhor distribuição dos retornos em questão. Foram utilizados 4 modelos¹⁷, visando-se escolher o melhor através do critério de informação de Akaike. Os modelos utilizados foram:

- ARCH(1) com resíduos seguindo uma Normal;
- ARCH(1) com resíduos seguindo uma *t de student*;
- GARCH(1,1) com resíduos seguindo uma Normal;
- GARCH(1,1) com resíduos seguindo uma *t de student*;

Isto é, tendo 4 possíveis modelos para cada retorno, tem-se 16 possíveis combinações entre os retornos da VALE3 e PETR4 que poderiam ser utilizados na Cópula bivariada escolhida. Ao se adicionar uma outra variável, este número já subiria para 64 possibilidades, mesmo antes de escolher qual Cópula utilizada. Isto só evidencia a dificuldade enfrentada ao trabalhar com variáveis dependentes e as consequência de se tentar modelar esta dependência.

Pelas tabelas 1 e 2, abaixo, observa-se que o melhor modelo tanto para a VALE3 quanto para a PETR4 é o modelo GARCH(1,1) com resíduos *t de student*. Vê-se que o parâmetro da distribuição apresenta graus de liberdade muito pequenos (8, 35 e 4), indicando uma distribuição com cauda pesada, como era esperado¹⁸. Ao

¹⁷Apesar de parecer que são os modelos mais simples, parou-se no GARCH(1,1) pois ele atingiu um patamar satisfatório de ajuste, principalmente quando foi observado o Q-Q Plot dos resíduos.

¹⁸O comando no R otimizou o GARCH(1,1) para a PETR4 obtendo uma *t de student* com 1 grau de liberdade. Porém, com este grau de liberdade a variância seria infinita

se fixar o grau de liberdade como 4 se notou que a perda no Critério de Akaike é insignificante (diminuição de 0,21%) perto do problema que seria enfrentado caso fosse utilizado uma *t de student* com grau 1. O parâmetro μ presente nas duas tabelas é simplesmente para testar a hipótese de que os retornos não possuem média zero, algo que foi rejeitado em todos os modelos, indo de encontro com o conhecimento prévio de que os retornos de um série financeira tem média igual a zero.

Tabela 1: Resultados para o Modelo ARCH(1)

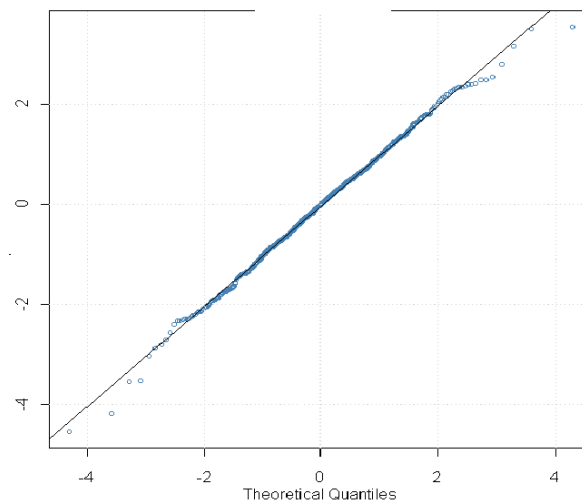
Resíduos Normal	PETR4	VALE3	Resíduos <i>t de student</i>	PETR4	VALE3
Mu	0.0004381	0.0007947	Mu	0.0009107	0.00191
Erro padrão	0.0008357	0.0009307	Erro padrão	0.0007108	0.0007848
P-valor	0.600*	0.3931*	P-valor	0.200085*	0.12915*
Omega	0.0006019	0.0007696	Omega	0.0005977	0.0007841
Erro padrão	0.00003293	0.00004368	Erro padrão	0.00006598	0.00009144
P-valor	≤ 0.0001	≤ 0.0001	P-valor	≤ 0.0001	≤ 0.0001
Alpha1	0.1890	0.1658	Alpha1	0.2613	0.2216
Erro padrão	0.04146	0.04559	Erro padrão	0.07248	0.07631
P-valor	≤ 0.0001	0.000276	P-valor	0.00312	0.00369
g.l.	-	-	g.l.	1	3.85
AIC	-4.408912	-4.184161	AIC	-4.535357	-4.316387

Tabela 2: Resultados para o Modelo GARCH(1,1)

Resíduos Normal	PETR4	VALE3	Resíduos <i>t de student</i>	PETR4	VALE3
Mu	0.0005265	0.00132	Mu	0.001077	0.001302
Erro padrão	0.0007439	0.0009307	Erro padrão	0.0006408	0.0007234
P-valor	0.4276*	0.0760*	P-valor	0.093*	0.0718*
Omega	0.00008929	0.000007046	Omega	0.00001124	0.000007349
Erro padrão	0.000003783	0.000003882	Erro padrão	0.000006609	0.000004473
P-valor	≤ 0.0183	0.0695	P-valor	0.089	0.1004
Alpha1	0.0726	0.06464	Alpha1	0.1079	0.06561
Erro padrão	0.01459	0.01389	Erro padrão	0.02726	0.01641
P-valor	≤ 0.0001	≤ 0.0001	P-valor	≤ 0.0001	≤ 0.0001
Beta1	0.9087	0.9268	Beta1	0.9028	0.9257
Erro padrão	0.01389	0.01556	Erro padrão	0.02406	0.01791
P-valor	≤ 0.0001	≤ 0.0001	P-valor	≤ 0.0001	≤ 0.0001
g.l.	-	-	g.l.	4	8.35
AIC	-4.640787	-4.427742	AIC	-4.665692	-4.449714

Observa-se nas figuras 2 e 3 abaixo (com auxílio das figuras 6 e 7 do apêndice), o comportamento das séries, o bom ajuste do modelo através do Q-Q plot e o gráfico ACF das observações e dos resíduos. Como o objetivo deste estudo é estudar o comportamento conjunto da carteira bivariada através de Cópulas, será feito um estudo das correlações, das medidas de dependência descritas anteriormente. A única ressalva aqui é que será feito um estudo da interdependência entre os resíduos dos modelos GARCH, pois eles seguem uma distribuição conhecida (*t de student* para ambos os modelos escolhidos) e não dos retornos em si. Os retornos podem ser estimados por uma metodologia recursiva dada pela fórmula do GARCH(1,1), possibilitando a simulação de passeios aleatórios condizentes com o modelo.

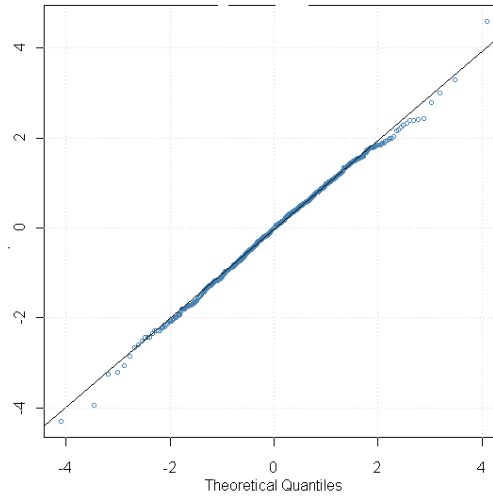
Figura 2: Q-Q plot dos resíduos padronizados da VALE3 pelo GARCH(1,1) - *t de student*



8.2 Escolha da Cópula

Na tabela 10 (Apêndice) tem-se o valor de algumas medidas de dependência entre os resíduos em questão. É importante notar a diferença entre os valores, pois

Figura 3: Q-Q plot dos resíduos padronizados da PETR4 pelo GARCH(1,1) - *t de student*



como citado anteriormente, sabe-se que o Rho de Spearman, o Tau de Kendall e o coeficiente de Pearson mensuram formas diferentes de dependência entre as variáveis.

A associação entre as variáveis é de fundamental importância no passo de calibração e ajuste da cópula em um modelo. Afinal, a Cópula em si é a função que une as variáveis de interesse, representando todas as medidas de associação dentro dela. Justamente por isto, foi evidenciado na seção de Medidas de Dependência a relação íntima entre a Cópula e estas medidas, como se pode mensurar estes coeficientes de correlação e dependência a partir da Cópula e vice-versa.

Neste momento do estudo foi feita uma comparação entre 6 Cópulas bivariadas amplamente utilizadas na literatura da área - 2 Cópulas elípticas, 3 arquimedianas e a Cópula independente. Estas Cópulas foram construídas a partir de 2 fatores (com o auxílio do *software R*): das marginais obtidas pelo método GARCH(1,1) e das medidas de dependência da tabela 10 (que são relacionadas aos parâmetros destas Cópulas).

As 6 Cópulas escolhidas foram:

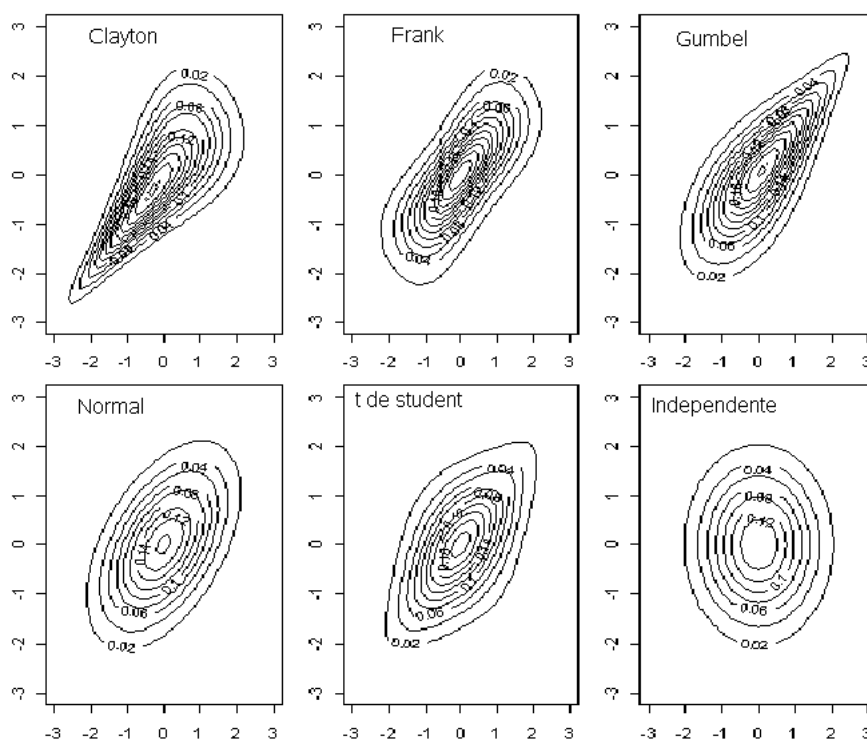
- i. Independente
- ii. Normal
- iii. *t de Student*
- iv. Clayton
- v. Frank
- vi. Gumbel

Estas 6 Cópulas foram escolhidas pois os seus parâmetros podem ser definidos a partir do Tau de Kendall ou do Coeficiente de Pearson, como pode ser verificado em Nelsen (1999), além de terem suas distribuições escritas de maneira fácil e mais intuitiva, como visto nas seções 6.4 e 6.5. Após a montagem das Cópulas fez-se um teste *goodness-of-fit* para verificar qual delas melhor se ajustou ao modelo. Após a escolha da melhor Cópula, foram simulados valores para definir o *VaR* da carteira em questão.

Com o auxílio do *software* R montou-se a figura 4 abaixo, contendo os gráficos de contornos para evidenciar a diferença entre as 6 cópulas, sendo que em todas o Tau de Kendall foi de 0,5 (com exceção, é claro, da Cópula Independente) e as marginais sendo Normal Padrão. Pode-se verificar na figura como a Cópula de Clayton e a de Gumbel apresentam dependência caudal forte, porém opostas. A cópula de Frank apresenta caudas pesadas, tanto positivamente quanto negativamente. A Cópula Normal se apresenta simétrica em relação ao eixo central da dependência (diagonal

positiva no caso), como se espera da distribuição Normal. A Cópula *t de Student* se comporta como presumida, com comportamento semelhante a Normal só que mais achatada na diagonal. Já a Cópula independente se comporta da maneira esperada, totalmente simétrica, tanto na vertical quanto na horizontal.

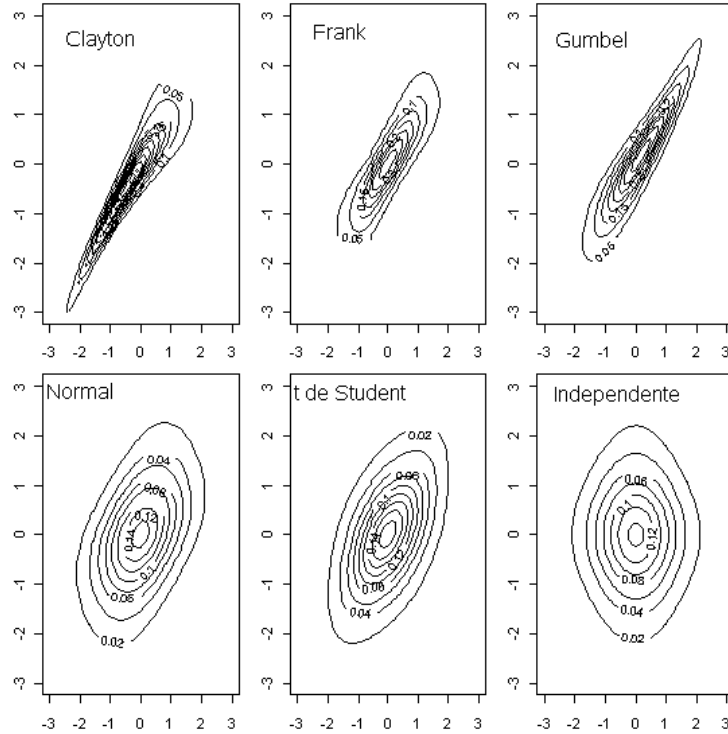
Figura 4: Gráficos de Contorno para Cópulas bivariadas com Tau de Kendall 0,5 e marginais Normal Padrão



Utilizando as marginais obtidas pelo GARCH(1,1) e com as medidas de dependência calculadas (tabela 3) foram construídas as 6 cópulas dos resíduos da como pode ser verificado na figura 5 a seguir. Fica claro como a dependência positiva significativa achata o gráfico na diagonal e como as marginais *t de student*, com graus de liberdade pequeno, prolongam as caudas.

Porém, escolher qual destas Cópulas melhor se ajusta aos dados é um procedimento complexo. Pois ao se levar em conta procedimento de máxima verossimilhança

Figura 5: Gráficos de Contorno para Cópulas bivariadas com Medidas de Dependência dadas na tabela 3 e com marginais *t de student* com graus de liberdade 4 e 8,35



convencionais pode-se obter resultados discrepantes, que deixam de lado os valores extremos pois a Cópula pode-se ajustar muito bem nos pontos centrais e acabar mascarando os valores extremos (como discutido na seção 6.7 baseado no *paper* de Nikoloulopoulos et al. - 2011). Para se ter uma idéia, foi feito um procedimento de máxima verossimilhança para 5 das 6 cópulas abordadas acima (com exceção da *t de student* e a que apresentou melhor resultado no Critério de Máxima Verossimilhança foi a Cópula Independente. Este resultado vai totalmente de encontro à idéia que motivou este estudo e a escolha destes papéis. Por se tratarem dos papéis mais negociados e líquidos da BOVESPA, de empresas que não são concorrentes, esperava-se uma alta correlação entre estes valores. A queda abrupta de valor na VALE3 deveria

gerar desconfiança no mercado e ocasionar queda de preço na PETR4. A recíproca é verdadeira e também nos períodos de otimismo e alta.

Visando enfrentar este problema, utilizou-se o teste proposto por Genest et al (2009), discutido na seção 6.7. A estatística proposta, Cramér-von Mises, testa a hipótese de que os dados se ajustam a uma distribuição específica. Isto é, foram testados se os resíduos se adequam a cada uma das distribuições bivariadas das 6 cópulas específicas. Desta maneira, como pode-se ver pela tabela 3 abaixo, a Cópula t apresenta um p-valor alto, não dando indícios de rejeitar a hipótese nula do teste. Enquanto isto, todas as outras cópulas apresentaram p-valores muito pequeno, dando mais evidências que dentre estas 6 cópulas a mais apropriada para se modelar os resíduos do GARCH(1,1) dos retornos é a Cópula t . Foi feito uma simulação de Monte Carlo com 1000 iterações para cada Cópula a fim de se calcular a integral do teste de Cramèr-von Mises (seção 6.7), e o resultado vai de encontro com *paper* referenciado, dando indícios de que apenas uma distribuição (a da Cópula t se ajusta adequadamente.

Tabela 3: P-valor do teste de Cramèr-von Mises

Cópula	p-valor
Clayton	≤ 0.0001
Frank	≤ 0.0001
Gumbel	≤ 0.0001
Normal	≤ 0.0001
<i>t de Student</i>	0.4540460
Independente	≤ 0.0001

Como citado na seção 6.7, o teste ainda calcula o parâmetro da Cópula que melhor se ajustou. No caso das Cópulas testadas, os parâmetros são as medidas de concordância (coeficiente de *pearson* e o tau de *kendall*) e os graus de liberdade da

Cópula t . Vê-se pela tabela 11 (apêndice), que os parâmetros de concordância se aproximam do valor real (já citado, na tabela 3).

Portanto, com os dados obtidos dos retornos da VALE3 e da PETR4 durante o 2º Governo Lula, tem-se que a Cópula bivariada que modela os resíduos do modelos GARCH(1,1) da VALE3 e da PETR4 é uma Cópula t com 3,4 g.l., cujas marginais são distribuições t com 8,35g.l. e 4g.l., respectivamente.

8.3 Simulação de Valores e Cálculo do VaR

Para se obter o *Value at Risk* do *portfolio* em questão foi utilizado um procedimento de simulação, construído com o auxílio do pacote *Copula* do *softwareR*. Foram simulados 200.000 observações da Cópula escolhida, isto é, 200.000 valores de pares de resíduos para os ativos baseado na distribuição conjunta (e conseqüentemente, baseado na relação de dependência entre eles). Em função dos valores estimados dos parâmetros do GARCH(1,1) da tabela 2 e com estes resíduos simulados, foram estimados 200.000 retornos diários da carteira ¹⁹. Os percentis inferiores obtidos desta simulação encontram-se na tabela 4 abaixo, que nos permitem estimar o *Var* de 1 dia da carteira.

Tabela 4: Percentis inferiores do retorno da carteira PETR4 e VALE3 com investimento igual para cada ativo com Cópula t

%	0,5%	1%	2%	2,5%	5%	10%
Percentil	-0,065271938	-0,053803125	-0,044101127	-0,041028644	-0,032010605	-0,023421322

Obtém-se o valor real da carteira no 1º dia útil do Governo Dilma fazendo-se uma simples transformação do retorno (que na verdade é o log-retorno, equação

¹⁹considerou-se aqui o investimento de 50% do total para cada ativo, logo o retorno diário foi a média entre os retornos.

5.1). Considerando-se um investimento de R\$1.000,00 na carteira, com R\$500,00 em ações da PETR4 e R\$500,00 em ações da VALE3 o *Value at Risk* de acordo com determinados níveis de confiança está na tabela 5 a seguir.

Tabela 5: *VaR* de 1 dia para o 1º dia útil do Governo Dilma de uma carteira com R\$500,00 comprado em VALE3 e R\$500,00 comprado em PETR4 seguindo Cópula t

Grau de Confiança	99,5%	99%	98%	97,5%	95%	90%
<i>VaR</i> (em reais)	63,19	52,38	43,14	40,20	31,50	23,15

Para se obter o *VaR* de 1 mês tendo-se o *VaR* de 1 dia bancos e consultorias multiplicam o *VaR* de 1 dia pela raiz quadrada do número de dias úteis neste mês que se deseja. Esta conta é justificada ao se presumir que os retornos são não-correlacionados dia-a-dia. Uma outra maneira de se fazer isto é se fazer a modelagem da série em valores mensais, em vez de diários, com outro modelo GARCH e outra Cópula. Essa metodologia porém restringe bastante o número de observações, pois teria-se apenas 12 observações no ano e os choques diários seriam deixados de lado na análise do modelo. Calculando-se o *VaR* de um mês (Janeiro de 2011 teve 21 dias úteis) da maneira feita em relatórios bancários e de consultorias, tem-se o resultado da tabela 6 abaixo.

Tabela 6: *VaR* de 1 mês a parti do 1º dia útil do Governo Dilma de uma carteira com R\$500,00 comprado em VALE3 e R\$500,00 comprado em PETR4 seguindo Cópula t

Grau de Confiança	99,5%	99%	98%	97,5%	95%	90%
<i>VaR</i> (em reais)	192,38	159,48	131,35	122,38	95,91	70,48

Outro ponto verificado foi o problema da subestimação do *VaR* quando não se utiliza a correlação evidente entre as variáveis. Desta maneira foram simulados

200.000 valores da Cópula Independente e verificado os percentis e o VaR da carteira desta maneira, como pode ser visto nas tabelas 7, 8 e 9 abaixo.

Tabela 7: Percentis inferiores do retorno da carteira PETR4 e VALE3 com investimento igual para cada ativo de maneira seguindo Cópula Independente

%	0,5%	1%	2%	2,5%	5%	10%
Percentil	-0,046025674	-0,038754417	-0,032341391	-0,030418974	-0,024562819	-0,018366886

Tabela 8: VaR de 1 dia para o 1º dia útil do Governo Dilma de uma carteira com R\$500,00 comprado em VALE3 e R\$500,00 comprado em PETR4 seguindo Cópula Independente

Grau de Confiança	99,5%	99%	98%	97,5%	95%	90%
VaR (em reais)	44,98	38,01	31,82	29,96	24,26	18,20

Tabela 9: VaR de 1 mês a parti do 1º dia útil do Governo Dilma de uma carteira com R\$500,00 comprado em VALE3 e R\$500,00 comprado em PETR4 seguindo Cópula Independente

Grau de Confiança	99,5%	99%	98%	97,5%	95%	90%
VaR (em reais)	136,95	115,73	96,89	91,22	73,87	55,41

É clara a diferença entre os valores levando em conta a correlação e sem levá-la em consideração. O montante de dinheiro em risco cai substancialmente ao se desconsiderar a relação de dependência, um indicativo importante do porquê da necessidade da mitigação do risco e do porquê da necessidade de se montar carteiras de títulos na forma de fundos de *hedge*.

Apresentando-se o valores das tabelas da Cópula t , os investidores pode tomar a decisão que quiser, pois não é apresentado apenas o valor de 99% como é regularmente apresentado. Neste caso investidores mais arrojados podem se basear em valores com grau de confiança reduzido (95% por exemplo), enquanto um investidor desconfiado pode decidir-se baseado no VaR de 99,5%.

9 CONCLUSÃO

Todo este relatório surgiu como uma idéia de se estudar o Risco. Um campo cujo estudo envolve o conhecimento estatístico, mas não apenas. A análise de risco desponta como uma grande área de aplicação da Estatística, devido à sua enorme necessidade e importância no mercado global, complexas análises e escassez de mão-de-obra especializada. Qualquer empresa ou pessoa está sujeita a algum tipo de risco com o intuito de gerar riquezas, financeiras ou não. Ao estudar para um concurso público, a pessoa assume o risco de não passar enquanto poderia estar trabalhando ou dormindo. Em maior escala, um banco tem que escolher aonde investir o dinheiro de seus correntistas, com o risco de todos quererem sacar amanhã, ou de perder o dinheiro investido. São muitos os tipos de risco e as maneiras de enfrentá-los vem crescendo em diversas áreas. Com este intuito surgiu este trabalho, visando a compreensão e a mensuração do risco de mercado através desta nova forma de mensurar o *VaR*: a Cópula.

Porém, ao se analisar o risco de séries de retornos financeiros de um ativo, estuda-se bastante o comportamento do ativo de maneira univariada. O estudo de séries temporais financeiras é de suma importância para a análise do risco, tal como visto neste trabalho. O problema enfrentado aqui é o da carteira bivariada, pois com dois ativos negociados neste mundo de fronteiras muito permeáveis e dinâmicas, espera-se uma correlação entre estes dois ativos e a forma de mensurá-la é complexa. Neste sentido, a Cópula surge como uma ferramenta muito importante para escrever

a função de distribuição conjunta e tornar clara a relação de dependência entre estas variáveis.

A Cópula por si só já apresenta uma grande área da Matemática e Estatística, tendo surgido há muitos anos nos estudos de Espaços Métricos e Probabilidade. Porém, suas aplicações mais diretas apareceram mais nos últimos vinte anos, com mais ênfase de dez anos para cá (destacando-se a área de Finanças). Sua idéia simples *a priori*, de escrever a função conjunta através das marginais deixa um amplo espaço para novas descobertas e o estudo da dependência entre variáveis aleatórias. Dependência que está totalmente definida no conceito de Cópula, uma medida que através de algumas transformações simples (como vistas na seção 6.6) pode evidenciar o valor de medidas de concordância.

Após o aprofundamento destes conceitos, o melhor entendimento da idéia de Risco, Séries Temporais e Cópula, estimou-se o *VaR* do *portfolio* em questão. Com o auxílio do *software* R foi feito um estudo da carteira de ativos contendo papéis da VALE3 e PETR4, estimou-se as distribuições univariadas dos resíduos dos modelos e construiu-se a Cópula para entender a dependência. Assim pôde-se estimar o *VaR* de uma maneira distinta e ainda fazer certas inferências, notar características da carteira de uma forma mais ampla que a usual.

As hipóteses utilizadas aqui ao se calcular o *VaR* não envolvem idéias amplamente arraigadas no estudo de retornos e séries financeiras, como o uso da distribuição normal ou independência entre ativos. Através da modelagem GARCH(1,1) com caudas mais pesadas para os retornos e da inserção da Cópula

obteve-se justamente o contrário, a não utilização da normal e o cálculo do *VaR* baseado na relação de dependência histórica entre os ativos.

De uma maneira mais ampla, mais motivacional para futuros estudos e continuação na carreira acadêmica, verificou-se que o estudo da Cópula, principalmente, é um tema bastante novo e desafiador. Existem poucos e novos artigos sobre inferência envolvendo Cópulas, deixando claro a área de vanguarda que é estudar a relação de dependência entre variáveis através deste fenômeno. Uma área que engloba conceitos avançados de probabilidade, inferência e programação, cujas pesquisas abordam temas diversos. Não são artigos apenas sobre finanças, pois existem estudos de Cópulas georreferenciadas, estudos com redes bayesianas, fenômenos hidrológicos, inferência bayesiana e muitos outros. Em síntese, muitas discussões e artigos sobre Cópulas e relação de dependência entre variáveis surgirão, procurando-se teorias e modelos que melhor expliquem o fenômeno da dependência.

Referências

- [1] ANDRADE, E.L.; **Introdução à Pesquisa Operacional**. 1a Ed. Rio de Janeiro: LTC, 2000.
- [2] BOLLERSLEV T. Generalized Autoregressive Conditional Heteroscedasticity. **Journal of Econometrics**. No.31, pp 307-327, 1986.
- [3] BOLLERSLEV T.; WOOLDRIDGE, J.M. Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariance. **Econometric Reviews**. No.11, pp 143-172, 1992.
- [4] CHERUBINI, U.; LUCIANO, E.; VECCHIATO, W. **Copula Methods in Finance**. 2a Ed. Chichester: Wiley Finance, 2004.
- [5] ELTON, E.J. **Modern Portfolio Theory and Investment Analysis**. 7a Ed. New Delhi: Wiley India, 2007.
- [6] ENGLE, R.F. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. **Econometrica**. No 50, pp 987-1008, 1982.
- [7] GENEST, C.; RÉMILLARD, B.; BEAUDOIN, D. Goodness-of-fit tests for copulas: A review and a power study. **Insurance: Mathematics and Economics**. No. 44, pp. 199-213. New York, 2009.
- [8] GENEST, C.; RIVEST, L.-P. Statistical Inference procedures for bivariate Archimedean Copulas. **Journal of the American Statistical Association**, No.88, pp 1034-1043, 2003.
- [9] JOE, H. **Multivariate Models and Dependence Concepts**. 1a Ed. London: Chapman Hall, 1997.
- [10] JORION, P. **Value at Risk: The New Benchmark for Managing Financial Risk**. 2a Ed. New York: McGraw-Hill, 2001.
- [11] MENDES, B.V.M; MELO, E.F.L. Local Estimation of Copula Based *Value at Risk*. **Revista Brasileira de Finanças**. Vol. 7, No. 1, pp. 29-50. Rio de Janeiro, 2009.
- [12] MCCULLAGH, P.; NELDER, J.A. **Generalized Linear Models**. Chapman and Hall, New York. 2ª Edição, 1989.

- [13] MORETTIN, P.A.; TOLOI, C.M. **Análise de Séries Temporais**. 2a Ed. São Paulo: Blucher, 2006.
- [14] NELSEN, R.B. **An Introduction to Copulas**. 2a Ed. New York: Springer, 2006.
- [15] R Development Core Team. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. Disponível no site <http://www.R-project.org/> , 2010.
- [16] NIKOLOULOPOULOS, A.K.; JOE, H.; LI, H. Vine copulas with asymmetric tail dependence and applications to financial return data. **Computational Statistics and Data Analysis**, In press, Corrected proof. Disponível no site <http://www.sciencedirect.com/science/article/pii/S0167947310002951> desde 24/07/2010.
- [17] SCARSINI, M. On Measures of Concordance. **Stochastica**. Vol.8 ,No. 3, pp. 201-208. New York, 1984.
- [18] YAN, J. **Enjoy the Joy of Copulas: With a Package copula**. Journal of Statistical Software, Vol.21, No. 4, pp. 1-21. Disponível no site <http://www.jstatsoft.org/v21/i04/> , 2007.
- [19] WUERTZ, D.; CHALABI, Y. **fGarch: Rmetrics - Autoregressive Conditional Heteroskedastic Modelling**. R package version 2110.80. Disponível no site <http://CRAN.R-project.org/package=fGarch> , 2009.

Apêndice: Outros Gráficos e Tabelas

Figura 6: ACF das observações e ACF dos resíduos padronizados da VALE3 pelo GARCH(1,1) - *t de student*

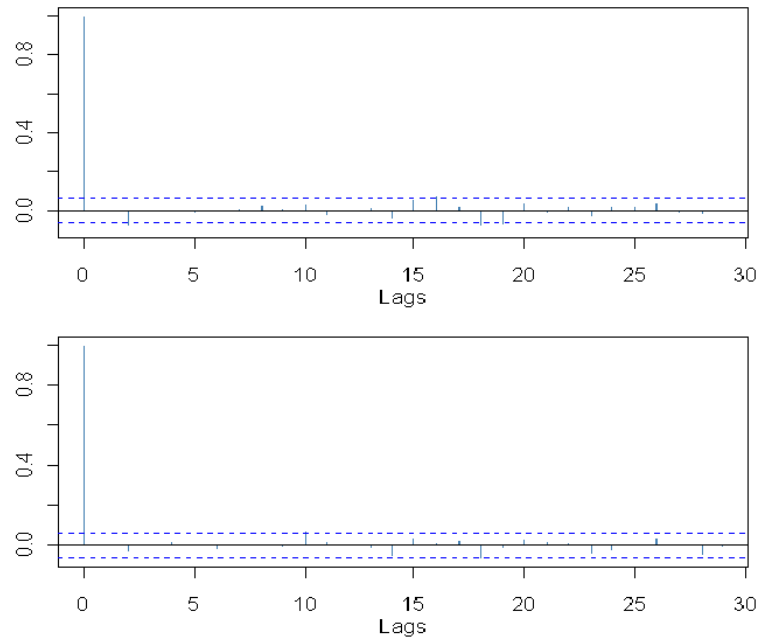


Figura 7: ACF das observações e ACF dos resíduos padronizados da PETR4 pelo GARCH(1,1) - *t de student*

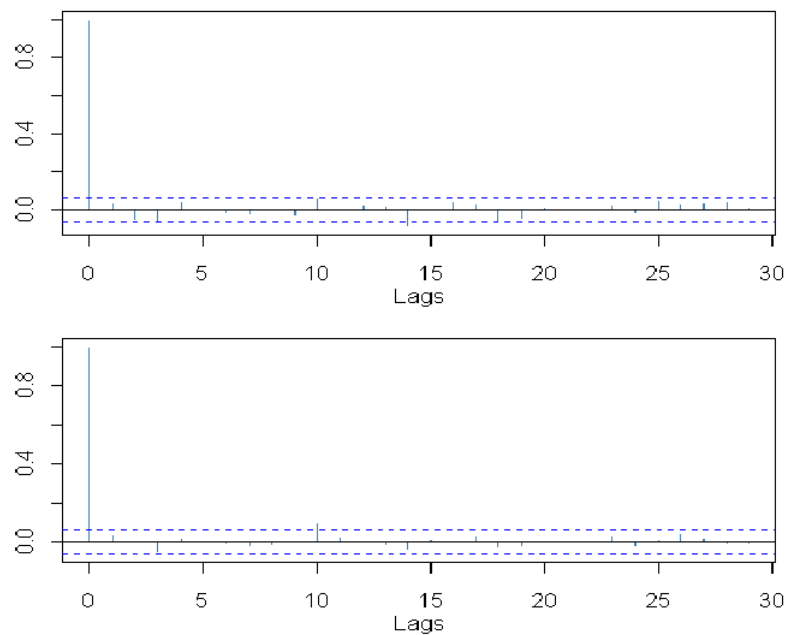


Tabela 10: Medidas de dependência entre os resíduos de modelos GARCH(1,1) - t de student - da VALE3 e PETR4

Pearson	Kendall	Spearman
0.7571642	0.5036346	0.6792452

Tabela 11: Comparação entre os parâmetros estimados pelo teste de *goodness-of-fit* e os parâmetros de concordância real

Cópula	Parâmetro Calculado	Parâmetro de Concordância Real
Clayton	1.614012	2.029290
Frank	4.850487	4.815
Gumbel	1.993852	2.014645
Normal	0.727366	0.7571642
t de Student	0.7158832	0.7571642