

1 – O PROBLEMA DE PESQUISA

1.1 INTRODUÇÃO

A área de análise e previsão de sucessões cronológicas (séries temporais ou crônicas), hoje campo de pesquisa bem definido dentro dos Métodos Quantitativos, tem sido uma das áreas-chave em Ciências Sociais Aplicadas.

Dadas as observações passadas do comportamento de alguns sistemas, pressupondo ergodicidade, surge a questão de como fazer um intervalo e/ou banda de previsões e identificar ainda o quão preciso esse intervalo pode ser, isto é, qual a porcentagem, via envelopamento, o intervalo conterá das sucessões cronológicas simuladas e/ou reais.

A área de Finanças, por exemplo, vem, ao longo das últimas duas décadas, modificando-se e se transformando em uma ciência eminentemente quantitativa, em que são usadas ferramentas como Estatística, Matemática e Ciência da Computação para que o processamento das informações forneça resultados mais precisos e confiáveis, principalmente ligados aos modelos de previsão (ZOU; YANG, 2004).

As condições presentes determinam, em certo grau, o futuro da forma em que pode haver muitas interações e complexas relações entre as variáveis envolvidas. Existindo informação suficiente disponível sobre o sistema em estudo, então uma abordagem matemática pode ser desejável, sendo que as equações construídas modelam os mecanismos responsáveis pela geração das séries temporais e como o seu comportamento evolui. No entanto, em muitos problemas reais de interesse, não temos informações e/ou condições ideais suficientes para construir equações que governem o comportamento das variáveis que queremos prever. Na ausência de informações suficientes para gerar as equações, é mais atrativo usar uma abordagem baseada em modelos (DE VEAUX et al., 1998).

Com o avanço da tecnologia e da capacidade de armazenagem e processamento dos sistemas computacionais, diversos modelos e técnicas quantitativas de previsão têm sido

pesquisados, complementando e aprimorando as análises qualitativas por uma série de fatores, incluindo maior precisão (HARDIE; FADER ; WISNIEWSKI, 1998).

Modelos de previsão quantitativos utilizam-se basicamente de dados históricos para detectar padrões de comportamento e estimá-los no futuro. Esses modelos empregam um ferramental matemático e estatístico para representar a realidade para a qual foram criados. Diversas técnicas estatísticas têm sido usadas na criação dos modelos, baseadas em diferentes pressupostos assumidos (WINKLHOFER et al., 1996).

Atualmente, uma nova metodologia para decomposição de séries via ondaletas, visando à análise de séries temporais, surgiu como uma área da Matemática Aplicada em rápido desenvolvimento e com diversas aplicações em vários ramos da ciência e da Administração. Ondaletas são funções que consistem em fracionar a série temporal original em duas subséries, uma relativa as altas frequências e a outra às baixas frequências (GENÇAY; SELÇUK; WHITCHER, 2002).

O uso das ondaletas é significativo por melhor detectar o comportamento da série em separado (altas e baixas frequências), visando à redução dos possíveis ruídos existentes na série original.

Já é crescente o número de pesquisas que estão se desenvolvendo com o uso desta nova metodologia nos indicadores de bolsa de valores, como tratam Tak (1995), Lee (1999) e Wong et al. (2003). No Brasil, ainda são poucos os trabalhos, como Chiann (1997) e Homsy, Portugal e Araújo (2000).

Este trabalho visa explorar a melhoria da qualidade das previsões com o uso da decomposição da série temporal via ondaletas, com o uso dos modelos já tradicionais de previsão, com os modelos já conceituados como a metodologia de Box e Jenkins e redes neurais, para sucessões cronológicas lineares e não-lineares simuladas.

1.2 PROBLEMA DE INVESTIGAÇÃO

Previsão, ou em inglês, *forecasting*, que se refere a “ato ou efeito de prever, antevisão, presciência...”, pode ser definida como uma seqüência de passos que o tomador de decisões realiza, seja implícita ou explicitamente, para antever satisfatoriamente um valor futuro (LUXHOJ et al., 1996).

Segundo Harrison e Stevens (1976), uma previsão adequada deve dar suporte a uma decisão minimizadora de risco por parte dos agentes tomadores de decisões. Para Shalizi et al. (2002), o futuro é sempre desconhecido e, até certo ponto, imprevisível e uma previsão é sempre imperfeita, mas não inútil.

Hanke e Reitsch (1995), ao explicarem o porquê de se fazer previsão, afirmam que toda previsão é uma tentativa de prognosticar o futuro através do exame do passado. Consiste em gerar previsões não enviesadas da magnitude de alguma variável, como um índice de Bolsa, com base no conhecimento presente e passado acumulado em bases de dados e na experiência dos gestores e outros profissionais envolvidos.

Conforme Tsay (2002), as previsões variam segundo a sua aplicação em níveis agregados (como na Economia) ou para um componente específico (como para uma campanha individual). Também difere quanto ao horizonte de previsão – curto, médio ou longo prazo – que são conceitos flexíveis, que irão variar conforme a área de aplicação.

Inúmeros autores já pesquisaram sobre a adoção das previsões na área de Finanças, utilizando-se de diferentes técnicas e abordagens. Alguns autores como Keim e Stambough (1986), Chen (1991), O'Connor, Remus e Griggs (1997), Leung, Daouk, Chen (2000), Bressan e De Lima (2002) pesquisaram a adoção de métodos de previsão de indicadores financeiros, desde previsibilidade de retornos de preços de ações até previsões de preços aplicados a contratos futuros na BM&F, através da modelagem de sucessões cronológicas, em que se têm, como variável de entrada, os valores históricos da variável a ser prevista, utilizando abordagens econométricas e redes neurais.

Alguns outros autores exploraram mais especificamente a previsão de índices de Bolsa de Valores. Esses autores, como Kutsurelis (1998), Leung, Daouk e Chen (2000), utilizaram técnicas de modelagem de séries temporais, tomando como base os modelos ARIMA sugeridos por Box e Jenkins (1976), e com uso de redes neurais voltadas para o índice S&P 500.

O surgimento da teoria de decomposição via ondaletas, a partir de meados dos anos 80, que consiste em fracionar a série temporal original em duas subséries, sendo uma relativa às altas frequências e a outra às baixas frequências, por meio de formas de ondas específicas, fez com que alguns autores começassem a incorporar o uso desta teoria em conjunto com outras técnicas já convencionais para fins de previsão (GENÇAY; SELÇUK; WHITCHER, 2002).

Ariño (1995) utilizou as ondaletas para previsão de vendas mensais de carros no mercado espanhol. Tak (1995) fez uso das ondaletas para previsão do S&P500, com a utilização em conjunto de redes neurais, modelos ARIMA e a metodologia proposta por Engle (1982), ambas aplicadas as duas subséries derivadas da série original, sendo que se fez as devidas previsões e após a reconstrução da série, novamente com uma *ondaleta* chegou-se à previsão do índice futuro.

Lee (1999) utilizou ainda a decomposição em diferentes frequências para estudar a relação entre as altas frequências das oscilações dos retornos dos índices das bolsas Dow Jones e KOSPI (*Korea Composite Stock Price Index*) em que detectou a existência de influências da alta volatilidade do mercado americano, nos preços dos ativos, no mercado coreano. Wong et al. (2003) também realizaram previsões com ondaletas em conjunto com modelos ARIMA, na série do dólar americano, e identificaram uma melhora no desempenho das previsões quando do uso das ondaletas para fracionar a série.

No Brasil, alguns poucos trabalhos, utilizando ondaletas apareceram na literatura como Chiann (1997), que trabalha com ondaletas na caracterização estatística de processos estacionários com dados dos componentes espectrais de eletroencefalograma durante o estado de sono de bebês, cujas mães se abstiveram de bebidas alcoólicas durante a gravidez e bebês cujas mães consumiram (de forma moderada) bebidas alcoólicas

durante a gravidez aplicada para detecção de alterações do estado de sono, via distorções nos comportamentos dos escalogramas em escalas altas e baixas.

Em se tratando de aplicações ao mercado brasileiro, Homsy, Portugal e Araújo (2000) também utilizaram a decomposição das ondaleta para tratar séries temporais de produção física industrial, exportações brasileiras e volume de pesca na Groelândia. Utilizou as previsões dentro das subséries extraídas com modelos ARIMA e chegou à conclusão de que a modelagem em separado de subséries de baixas e altas frequências contribui positivamente para a qualidade das previsões.

No caso específico de previsão no mercado de capitais brasileiro, através do índice da Bolsa de Valores de São Paulo, o IBOVESPA, alguns autores como Correia (1998), Morais e Portugal (2000) e Yim (2001) trataram das previsões dos retornos do IBOVESPA, apenas utilizando os modelos tradicionais como ARIMA, os modelos Estruturais, os modelos ARCH-GARCH propostos por Engle (1982) e de redes neurais, encontrando melhor desempenho nos modelos ARIMA e de redes neurais.

Muitos algoritmos neurais para processamento temporal têm sido propostos, entre estes destacam-se o algoritmo de retropropagação através do tempo (BPTT, *back propagation through time*), aprendizado recorrente em tempo real (RTRL, *real-time recurrent learning*) e os algoritmos de redes recorrentes utilizando o filtro de Kalman.

Vê-se, então, uma lacuna na literatura sobre previsão existente no mercado brasileiro, que é o estudo conjunto da utilização dos modelos ARMA-GARCH e de redes neurais, dentro das subséries decompostas por meio de uma ondaleta específica e que, após a previsão feita dentro destas subséries, é reconstruída a série original pela mesma ondaleta. A exemplo da pesquisa de Tak (1995) e Homsy, Portugal e Araújo (2000), o presente estudo investiga a decomposição na série original via uso de formas de ondaletas específicas e também na reconstrução da série original para obtenção da previsão futura.

O diferencial deste trabalho está na realização das previsões dentro das subséries decompostas por uma ondaleta em até dois níveis, e obtendo-se a previsão da série original via reconstrução da série para modelos construídos por processos geradores de

dados e também, apresentar um organograma lógico para guiar passo a passo, as previsões de sucessões cronológicas lineares e não-lineares simuladas.

O problema central da investigação deste trabalho pode ser concentrado na seguinte questão: “comparando os métodos já utilizados de previsão de sucessões cronológicas com a metodologia das ondaletas, qual se mostra mais preciso na previsão por intervalo de um processo estocástico simulado?”

1.3 OBJETIVOS DA PESQUISA

O objetivo geral desta pesquisa é o de explorar a possibilidade de usar uma metodologia capaz de decompor uma série temporal via ondaletas, conjuntamente com os modelos já existentes de previsão e comparar a qualidade de previsões obtidas.

A análise de ondaletas tem se mostrado, a partir de meados dos anos 80, uma metodologia de bastante utilidade em vários campos do conhecimento como Física, Matemática, e, na presente década, em Economia e Administração, podendo-se ressaltar, em econometria, sua utilização como procedimento auxiliar para ajuste e previsão de séries de tempo (GENÇAY; SELÇUK; WHITCHER, 2002).

Entre os objetivos específicos, merecem destaque:

- i) Propor um método de pré-processamento, identificação, estimação, diagnóstico, validação e previsão;
- ii) delinear em um fluxograma as etapas de identificação, estimação, diagnóstico, validação e previsão de sucessões cronológicas;
- iii) verificar se o pré-processamento de uma sucessão cronológica estacionária e ergódica melhora significativamente a previsão estática realizada por meio de algoritmos de aprendizagem; e
- iv) comparar a utilização das ondaletas, como pré-processamento, nas previsões conjuntamente com os métodos já presentes na literatura.

- v) aplicar o método desenvolvido na tese para a sucessão cronológica do IBOVESPA.

1.4 DELINEAMENTO DA TESE

O capítulo 1 trata da apresentação do problema de pesquisa, da motivação que levou ao desenvolvimento do trabalho com seus objetivos gerais e específicos, mostrando o sentido do uso do pré-processamento em força conjunta com os modelos tradicionais de previsão de sucessões cronológicas.

O capítulo 2 mostra a fundamentação teórica dos modelos de previsões. Para tanto, são discutidos os principais testes de classificação para estabelecer uma ordem no processo de previsão para sucessões cronológicas lineares e não-lineares simuladas, o que justifica a necessidade do posicionamento de um fluxograma, em que se mostra o processo de previsão como um todo.

Foram estudados os mais variados modelos de previsão, desde os modelos tradicionais como ARMA até os mais complexos como GARCH e os modelos neurais via algoritmos de retropropagação e das redes recorrentes, apresentando as simulações de processos geradores de dados (DGP).

Diante desses modelos, o capítulo desenvolve ainda a análise de sucessões cronológicas via o pré-processamento através de ondaletas. De forma original, apresenta a idéia central do trabalho de se fazer as previsões dentro das subséries decompostas em alta e baixa frequência, buscando uma melhora na qualidade das previsões. Também, de forma original, o capítulo descreve todo o processo quantitativo de se escrever corretamente os processos ARIMA-GARCH. Trata ainda, da matemática das ondaletas no processo de previsão.

O capítulo 3 destaca o processo metodológico empregado na pesquisa, indicando o fluxograma para os modelos de previsão estudados, levando ao processo final de se verificar, via envelopamento dos dados, a qualidade tanto das previsões pontuais como o estabelecimento de um intervalo de previsão para os dados simulados de redes e

modelos ARMA-GARCH, ARIMA, bilinear e um movimento browniano e para os dados do IBOVESPA.

No capítulo 4, são apresentados todos os resultados experimentais das simulações feitas para alcançar os objetivos da pesquisa.

Finalmente, são traçadas as conclusões da pesquisa, destacando o comportamento de cada modelo simulado, analisando e apontando suas principais características.

2 – FUNDAMENTAÇÃO TEÓRICA

2.1 SUCESSÕES CRONOLÓGICAS

Uma série temporal é qualquer conjunto de variáveis estocásticas eqüiespaçadas e ordenadas no tempo $\{X_t\}_{t=1}^n = \{X_1, X_2, \dots, X_n\}$. Em outras palavras, um sinal que depende do tempo e é medido em pontos particulares no tempo é sinônimo de uma série temporal (KLEIN, 1997). Conforme Morettin (1987), o que se chama de série temporal é uma parte de uma trajetória, dentre muitas que poderiam ter sido observadas de um processo estocástico¹.

Segundo Enders (2003), para que uma série de observações indexadas eqüiespaçadamente no tempo possa ser chamada de sucessão cronológica precisa ter pelo menos 50 observações.

Box, Jenkins e Reinsel (1994) definem série temporal como um conjunto de observações dos resultados de um fenômeno colhido seqüencialmente ao longo do tempo. Observa-se que a definição feita é coerente com a de Newbold (1995): “uma série temporal é uma seqüência de valores, ordenados no tempo, de uma variável de interesse particular”.

Existindo informações disponíveis suficientes sobre o sistema que está em estudo, então uma abordagem matemática pode ser desejável, sendo que as equações construídas modelam os mecanismos responsáveis pela geração das séries temporais e como o seu comportamento evolui.

Tipicamente, esses processos envolvem a consolidação das experiências históricas em sistemas matemáticos que descrevem o comportamento dos eventos através do tempo. Esta tarefa deve ser realizada da forma mais concisa possível e pode ser chamada de construção do modelo. Numa segunda etapa, estes modelos matemáticos são utilizados

¹ Processos estocásticos são processos cuja evolução no tempo é gerado por leis probabilísticas.

para projetar o que é provável que ocorra durante algum curto período de tempo. Este processo é chamado de “previsão”, isto é, uma extrapolação feita para além do processo temporal conhecido. A junção destes dois processos é o que se chama de análise de séries temporais. A estatística é, então, utilizada para se estimar os parâmetros e testar estatisticamente estes modelos. Vale ressaltar que, na regressão, são feitas “previsões” dentro do domínio no qual a sucessão cronológica foi observada.

Dentre os objetivos das análises de sucessões cronológicas, destacam-se: a investigação do mecanismo gerador da série; realização de previsões de valores futuros de curtos e longos prazos; descrição do comportamento da série, com verificação gráfica de existências de tendência, ciclo e variações sazonais, busca de periodicidade nos dados (MORETTIN; TOLOI, 2004).

Se os valores futuros de uma série temporal forem determinados exatamente por alguma função matemática, então esta série é denominada determinística. Caso os valores futuros possam ser descritos apenas em termos de uma distribuição de probabilidade, a série temporal é chamada de não-determinística ou estocástica. Um fenômeno estatístico que envolva leis da probabilidade no tempo é chamado de processo estocástico.

2.1.1 TÉCNICAS ESTATÍSTICAS DE PREVISÃO

Nesta parte do trabalho, é feita uma reunião das técnicas, buscando uma organização do que se deseja fazer em relação à previsão, usando modelos de sucessões cronológicas não-lineares simuladas.

Os problemas de previsão de séries temporais são, geralmente, abordados sob a perspectiva dos modelos matemáticos gerados. Um modelo é uma descrição probabilística para uma série temporal. Os modelos utilizados para descrever sucessões cronológicas são processos estocásticos (DINIZ, 1999).

Winklhofer et al. (1996) destacam os fatores que alavancaram os estudos de previsões:

- aumento da complexidade das organizações e de seu ambiente, destacando o crescimento do número de nichos de mercados e produtos oferecidos;
- adoção, pelas organizações, de procedimentos decisórios mais sistemáticos, que envolvem justificativas explícitas para cada ação tomada. Ter uma idéia da previsão de oscilação do risco de um ativo financeiro é uma forma de suportar esses procedimentos; e
- desenvolvimento contínuo das técnicas de previsão e suas aplicações, permitindo que não só agentes especializados, mas também outros tomadores de decisões possam ter acesso às técnicas mais facilmente.

Observar os dados passados de um fenômeno para tentar antecipar o seu comportamento futuro constitui a essência da previsão. Existindo um modelo matemático capaz de descrever completamente esse fenômeno, diz-se que este fenômeno é determinista. Caso o modelo seja incompleto ou desconhecido, ele é não-determinista e as suas variáveis probabilísticas devem ser consideradas na representação, passando a ser chamado de processo ou modelo estocástico ou aleatório (DINIZ, 1999).

2.1.2 QUANTO AO HORIZONTE DE TEMPO DAS PREVISÕES

Em relação ao horizonte de tempo das previsões feitas, há pouco consenso entre os autores, tanto no que se refere ao desempenho das técnicas quanto às próprias definições de curto, médio e longo prazos. Estudos já foram conduzidos para previsões com horizonte de tempo desde um dia (SAKAI et al., 1999) até 25 anos (NAYLOR, 1981).

A previsão de uma série temporal é a determinação dos prováveis valores que serão assumidos pela variável de previsão, dentro de um horizonte máximo de tempo. A previsão pode ser classificada como de curto, médio e longo prazo, dependendo do valor associado ao horizonte máximo de previsão (DINIZ, 1999).

Souza (1989) ressalta que qualquer que seja o horizonte de previsão (curto, médio ou longo prazo), não existe uma maneira clara de definir o horizonte máximo. Este

dependerá do grau de previsibilidade da série, que, por sua vez, está associado ao erro de previsão.

Os estudos relacionados ao horizonte de tempo e frequência das previsões confirmam, no entanto, a proposição de WINKLHOFER et al. (1996), de que “as companhias utilizam o período que melhor se adapta aos seus produtos, mercado e método de operação. Não há um período ideal para previsões.”

Em relação às previsões feitas para o IBOVESPA, alguns autores como Correia e Perera (1997) e Moraes e Portugal (1999) fizeram previsões no curtíssimo prazo (um passo à frente).

2.1.3 OS MÉTODOS DE PREVISÃO

Segundo Diniz (1999), até o início da década de 70, havia poucas alternativas metodológicas para análise e previsões de séries temporais como, por exemplo, os métodos de decomposição clássica e os métodos de “caixa-preta” (regressão linear simples e múltipla etc.).

Souza (1989) frisa que foi, a partir dos trabalhos de Box e Jenkins, nos anos 70, que a área de previsão começou a despertar grande interesse. Quase que simultaneamente surge, em 1971, a abordagem bayesiana dos modelos de previsão propostos por Harrison e Stevens (1971). Assim, tanto do lado conhecido, também chamado clássico, como o lado bayesiano, toda a década de 70 foi dedicada à implementação prática dos modelos propostos e as conseqüentes melhorias metodológicas sugeridas pela aplicação real destes métodos.

Os métodos de previsão podem ser classificados basicamente em duas grandes famílias, conforme as técnicas que utilizam na elaboração da previsão: técnicas qualitativas e técnicas quantitativas (ABRAHAM; LETHOLTER, 1983).

2.1.3.1 TÉCNICAS QUALITATIVAS

As técnicas de previsão qualitativas, também chamadas de técnicas subjetivas ou baseadas em critérios de juízo, são aquelas que utilizam primordialmente a capacidade humana de estabelecer generalizações e extrapolações (WINKLHOFER et al., 1996). Segundo Pindyck e Rubenfield (1997), essas técnicas são usadas quando dados históricos são inexistentes ou escassos, sendo baseadas em opiniões de especialistas. Para os autores, as técnicas mais difundidas estão associadas à pesquisa de mercado, e incluem a utilização de grupos focados e técnicas de consenso.

Em pesquisa feita por Mentzer e Cox (1984) sobre familiaridade com métodos qualitativos e quantitativos, 39% dos respondentes usavam técnicas qualitativas em seus negócios, contra 61% que afirmavam conhecer pelo menos uma técnica quantitativa de previsão. Na pesquisa de Sanders e Manrodt (1994), o número de pessoas que já trabalhavam com técnicas quantitativas subiu para 76%, contra 24% das técnicas qualitativas. Em recente pesquisa de Jan (2003), apenas 13,95% dos respondentes estavam empregando técnicas qualitativas, 19,69% usavam modelos de causa e efeito (regressão e modelos econométricos) e 63,08% utilizavam técnicas quantitativas (modelos de séries temporais) e 3,28% dos respondentes, utilizavam-se de outros métodos.

Jan (2003) comprovou, ainda, que 66,09% dos que usam modelos de séries temporais são aplicados pelas empresas para previsões no mercado financeiro.

Esses números comprovam os estudos de Luxhoj et al. (1996), Zhou (1999), Subrahmanyam (2000) e Lawrence et al. (2000), sobre a baixa qualidade das previsões feitas por métodos qualitativos, que acabavam por ocasionar distúrbios no planejamento e nos resultados finais das empresas pesquisadas.

Lawrence et al. (2000) sugerem alguns motivos pelos quais as previsões subjetivas, mesmo sendo realizadas por analistas experientes e com informações contextuais do mercado, não possuem bom desempenho:

- as informações contextuais podem não ter valor preditivo;

- enviesamento e ineficiência na interpretação dos relacionamentos podem mascarar as informações contextuais levadas em consideração;
- excesso de dados a serem considerados pelos analistas, fazendo com que ignore ou atribua pesos errados às informações contextuais recebidas;
- perturbações recentes, na série temporal, podem ser entendidas pelo analista como um sinal de tendência, confundindo a interpretação das informações contextuais; e
- acurácia não é o critério mais importante nas previsões das empresas.

Dada a baixa capacidade preditiva das técnicas qualitativas, esta pesquisa está focada em técnicas quantitativas de previsão, e, portanto, apenas apresenta brevemente algumas das técnicas qualitativas mais utilizadas, sem detalhar seu funcionamento.

Segundo Armstrong e Brodie (1999), as duas técnicas qualitativas empregadas mais utilizadas que são:

- método PERT-modificado: que é uma técnica exploratória de dados, baseada em estimativas otimista, pessimista e realista de acontecimentos futuros. Às estimativas, são associadas probabilidades de acontecimento, que são posteriormente avaliadas. Fundamenta-se na teoria das decisões, porém é simplificado quanto à parte estatística;
- técnica delphi: criada em 1950, na Rand Corporation, é uma técnica de grupo baseada em questionários enviados para especialistas que buscam um consenso de opiniões até que se chegue a um resultado final. A premissa básica é a de que o julgamento coletivo, organizado adequadamente, é superior a opiniões individuais.

2.1.3.2 TÉCNICAS QUANTITATIVAS

As técnicas de previsão quantitativas são aquelas que usam dados históricos para calcular matematicamente extrapolações dos dados no futuro (MAKRIDAKIS et al., 1983). Para os autores, o uso das técnicas quantitativas podem ser aplicadas quando:

- informação sobre o passado esteja disponível;

- esta informação possa ser quantificada em termos matemáticos; e
- seja possível assumir que alguns aspectos do padrão verificado no passado continuarão no futuro. Esta colocação é também chamada de pressuposto da continuidade.

Segundo Winklhofer et al. (1996), existem duas grandes famílias de técnicas quantitativas: a análise de séries temporais e os modelos causais e os modelos estruturais. Os autores citam, ainda, que os modelos causais assumem um relacionamento de causa e efeito entre as entradas e as saídas do sistema. O sistema pode ser visto como a economia nacional, o mercado de uma empresa, e assim por diante.

Essa definição é coerente com a encontrada em Pindyck e Rubinfeld (1998), em que estes autores frisam que as técnicas de previsão causais buscam descrever matematicamente as relações de causa e efeito entre a variável que está sendo medida e seus valores constituintes. Como exemplos de modelos causais, os autores citam os modelos de regressão.

Para Yim (2001), os modelos estruturais propostos por Harvey (1989) são aqueles que pressupõem que uma série de tempo resulta da combinação dos seguintes componentes: tendência, cíclica, sazonal e irregular.

A previsão com uso de séries temporais trata o sistema como uma “caixa preta”, sem tentar descobrir os fatores que causam os comportamentos observados (WINKLHOFER et al., 1996).

Segundo Makridakis et al. (1983), existem duas razões básicas para tratar um sistema como uma caixa preta:

1. o sistema não pode ser compreendido, ou mesmo que possa, é difícil medir o relacionamento entre as variáveis que governam seu comportamento; e
2. a preocupação pode ser simplesmente prever, com algum grau de precisão, o que vai acontecer ou não.

A análise de séries temporais é o processo de avaliar os relacionamentos históricos entre uma variável e o tempo. O propósito básico é identificar comportamentos periódicos e utilizá-los para prever séries futuras (CAMPBELL; LO; MACKINLAY, 1997).

Os resultados de Bidarkota (1998), entretanto, mostram que há uma associação positiva entre a importância dos métodos de previsão e a sua acurácia. Isto significa que quanto maior for a acurácia do modelo, maior a importância percebida dele.

Após a geração de um modelo de previsão, deve-se verificar se ele representa, ou não, corretamente os dados conhecidos através de medidas estatísticas convenientes para estabelecer a eficiência do modelo.

2.1.3.3 ANÁLISE DA QUALIDADE DA PREVISÃO PONTUAL REALIZADA PELO MÉTODO

Para Makridakis et al. (1983), a acurácia (previsão em termos pontuais) é vista como o critério primordial na análise para seleção do melhor modelo de previsão. A palavra acurácia refere-se a “o quão bem ajustado” é o modelo, isto é, “quanto” o modelo é capaz de reproduzir os dados já conhecidos.

Dentre as diversas ferramentas possíveis para analisar a eficiência de um modelo, citam-se, a seguir, as principais medidas de desempenho, admitindo o uso de sentenças matemáticas (MAKRIDAKIS et al., 1983).

Adotando a abordagem sugerida por Enders(2003), primeiramente considere um modelo de previsão, a partir de um AR(1), como $X_t = \phi_0 + \phi_1 X_{t-1} + \varepsilon_t$, sendo que ε_t é um ruído branco. Prevendo um passo à frente através da filtragem $I_t = \{X_t, X_{t-1}, \dots\}$, obtém-se:

$$E_t[X_{t+1}] = \phi_0 + \phi_1 X_t \quad (2.1)$$

sendo que $E_t[X_{t+1}]$ é a esperança condicionada de X_{t+1} , dada a informação até período t (inclusive), sendo o estimador ótimo pelo critério do erro quadrático médio condicionado.

Formalmente, $E_t[X_{t+j}] = E[X_{t+j} | X_t, X_{t-1}, \dots]$, sendo I_t o conjunto informacional dado por $I_t = \{X_t, X_{t-1}, \dots\}$. Segundo Enders (2003), pode-se definir o erro de previsão j passos à frente, a partir de t - $e_t(j)$ como sendo a diferença entre a variável aleatória X_{t+j} e o previsor condicionado:

$$e_t(j) \equiv X_{t+j} - E_t[X_{t+j}] \quad (2.2)$$

Assim, o erro da previsão um passo à frente é: $e_t(1) \equiv X_{t+1} - E_t[X_{t+1}] = \varepsilon_{t+1}$, isto é, a parte “não prevista” da porção X_{t+1} , dada a informação contemporânea avaliada no período t e o previsor de X_t dado o conjunto informacional I_{t-1} ficando:

$$\varepsilon_t = e_{t-1}(1) = X_t - E_{t-1}[X_t] \quad (2.3)$$

E ainda, para encontrar o erro de previsão para dois passos à frente, tem-se:

$$e_t(2) \equiv X_{t+2} - E_t[X_{t+2}] \quad (2.4)$$

sendo que $X_{t+2} = \phi_0 + \phi_1 X_{t+1} + \varepsilon_{t+2}$ e

$$E_t[X_{t+2}] = \phi_0 + \phi_1 E_t[X_{t+1}] \quad (2.5)$$

que representa a natureza do modelo ARMA.

Assim:

$$e_t(2) \equiv X_{t+2} - E_t[X_{t+2}] = e_t(2) = \phi_0 + \phi_1 X_{t+1} + \varepsilon_{t+2} - \phi_0 - \phi_1 E_t[X_{t+1}]$$

$$e_t(2) = \phi_1 (X_{t+1} - E_t[X_{t+1}]) + \varepsilon_{t+2}$$

$$e_t(2) = \phi_1 \varepsilon_{t+1} + \varepsilon_{t+2}. \quad (2.6)$$

Generalizando:

$$e_t(j) = \varepsilon_{t+j} + \phi_1 \varepsilon_{t+j-1} + \phi_1^2 \varepsilon_{t+j-2} + \dots + \phi_1^{j-1} \varepsilon_{t+1} \quad (2.7)$$

Como se tem uma série temporal com n períodos de tempo, logo ter-se-ão n termos de erro, e então se pode calcular:

- Erro Percentual Absoluto Médio (*Mean Absolute Percentage Error*) (MAPE): é um valor absoluto médio em percentual, para se verificar a margem de acerto em comparação com o valor previsto. É mais adequado para comparação entre modelos. Quanto menor for o seu valor, melhor é o ajuste do modelo.

$$MAPE_h = \frac{\sum_{i=1}^h \left| \frac{\hat{\varepsilon}_i}{X_i} \right|}{h} \times 100\% \quad MAPE \geq 0 \quad (2.8)$$

- r : Coeficiente de Correlação linear de Pearson. O coeficiente de correlação linear r mede o grau de relacionamento linear entre os valores emparelhados x e y em uma amostra. O coeficiente de correlação linear é chamado, às vezes, coeficiente de correlação momento-produto de Pearson. Varia entre -1 e 1 , sendo estes extremos indicação de associação linear negativa e positiva perfeita, respectivamente.

$$r_h = \frac{\sum_{i=1}^h (X_i - \bar{X})(\hat{X}_i - \bar{\hat{X}})}{S_X S_{\hat{X}}} \in [-1; 1] \quad (2.9)$$

em que \hat{X} é a esperança condicionada de X_{t+1} dada a informação até período t (inclusive).

Quadro 1 : Valor do coeficientes de correlação e sua interpretação

r_h	Interpretação
+ 1,00	positiva perfeita
+ 0,70 a 0,99	positiva muito forte
+ 0,50 a 0,69	positiva substancial
+ 0,30 a 0,49	positiva moderada
+ 0,10 a 0,29	positiva baixa
+ 0,01 a 0,09	positiva ínfima
0,00	Nenhuma
- 0,01 a 0,09	negativa ínfima
- 0,10 a 0,29	negativa baixa
- 0,30 a 0,49	negativa moderada
- 0,50 a 0,69	negativa substancial
- 0,70 a 0,99	negativa muito forte
- 1,00	negativa perfeita

Fonte: Newbold (1995)

- TIC – Coeficiente de Desigualdade de Theil (*Theil Inequality Coefficient*): Este coeficiente sempre estará entre zero e um, sendo que zero indica um ajuste perfeito.

$$TIC_h = \frac{\sqrt{\sum_{i=1}^h \hat{\varepsilon}_i^2}}{\sqrt{\sum_{i=1}^h \hat{X}_i^2} + \sqrt{\sum_{i=1}^h X_i^2}} \in [0,1] \quad (2.10)$$

Outras medidas de acurácia das previsões pontuais ainda podem ser definidas, apesar de não serem tão comuns na prática. A escolha do critério a ser utilizado para medir a acurácia dos modelos não é arbitrária e deve ser feita baseada nas características do problema e das medidas citadas (HARDIE; FADER; WISNIEWSKI, 1998).

No caso das sucessões cronológicas, diversos autores como Tolo (1980), Winklofer, Witt e Diamatopoulos (1996), Diniz (1999) e Yim (2001), utilizaram o RMSE (raíz do

erro quadrático médio) e o MAPE, indicando preferencialmente este último por ser computado por medidas absolutas (e não quadradas) e em porcentagem do valor previsto.

Neste trabalho, é usado MAPE e correlação de Pearson, conforme sugere Qi (1999) e TIC, conforme Tak (1995).

2.2 ANÁLISE DE SUCESSÕES CRONOLÓGICAS

O propósito de descrever um esquema para análise de sucessões cronológicas é explicar como uma sequência de testes estatísticos podem ser aplicados para identificar quais tipos de modelos de sucessões cronológicas univariadas podem ser construídos e aplicados.

O primeiro passo requer testar a estacionariedade do processo. O segundo envolve o teste para normalidade, para prever o tipo de dependência do processo estacionário. No terceiro passo, deve-se determinar se a mesma série é, de fato, estatisticamente independente. Se a hipótese nula de independência é rejeitada, isto implica aceitação da hipótese alternativa da existência de dependência linear ou dependência não-linear. O teste para estas hipóteses ocorre no quarto passo.

Se a série temporal apresentar dependência linear, então se deve especificar os modelos lineares para ajustes com filtro linear ou ARMA. Alternativamente, caso ocorra dependência não-linear, utilizar os modelos específicos para dados não-lineares.

A especificação de modelos não-lineares inicia-se no quinto passo, nos quais modelos não-lineares aditivos ou multiplicativos devem ser construídos. Por exemplo, no caso da existência de não-linearidade na esperança condicionada, podem ser usados os modelos bilinear, TAR (*threshold autoregressive*) ou exponenciais. Alternativamente, se a não-linearidade ocorrer na variância pode ser usados os modelos da família GARCH.

O próximo passo é a análise da qualidade das previsões e encerra-se com a validação dos modelos.

2.2.1 FLUXOGRAMA PARA ANÁLISE DE SUCESSÕES CRONOLÓGICAS

A seguir, apresenta-se um fluxograma, seguido para a análise de sucessões cronológicas unidimensionais lineares e não-lineares simuladas. Trata-se de uma metodologia de análise e previsão a ser adotada para modelos de sucessões cronológicas.

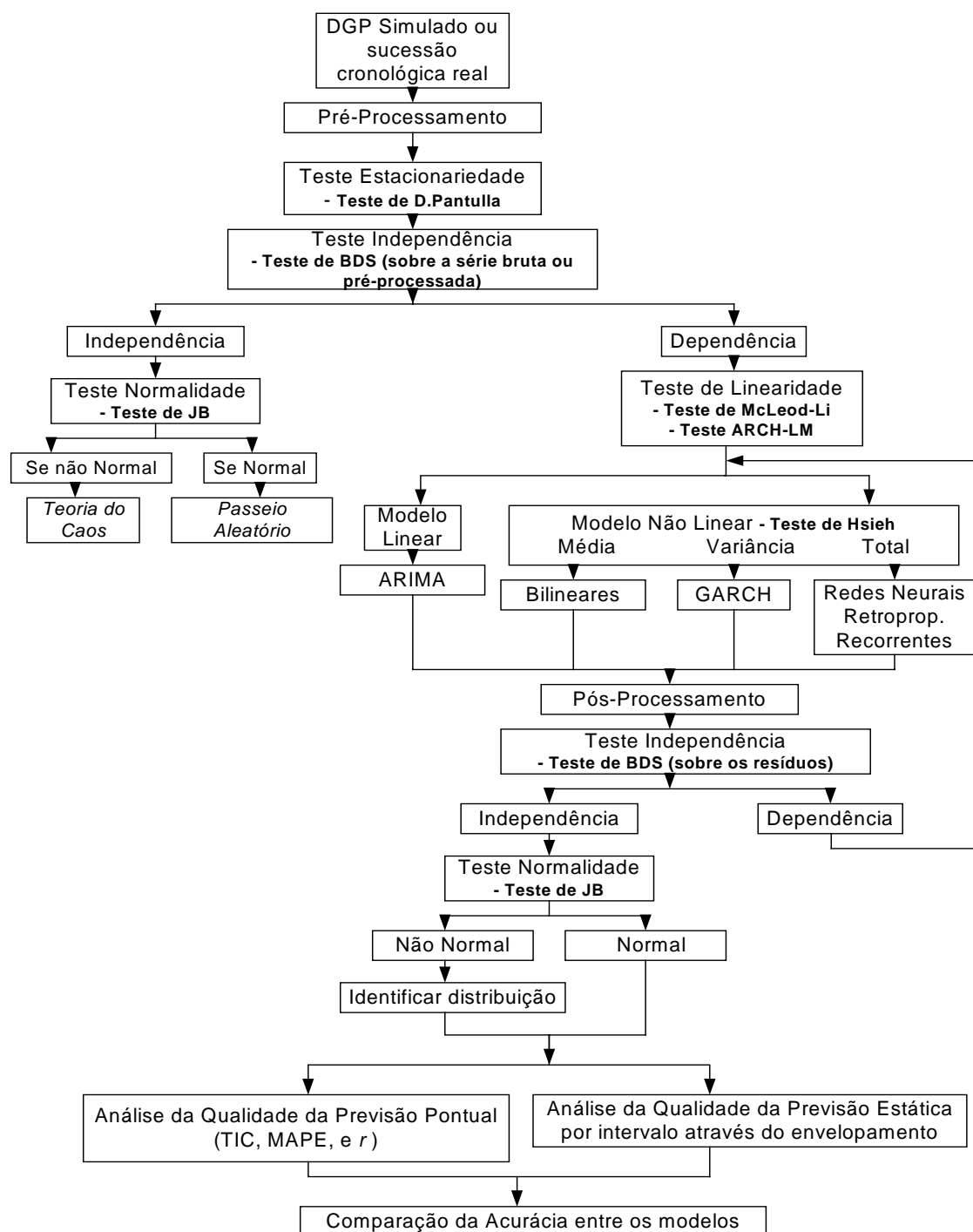


Figura 1: Fluxograma proposto para análise de sucessões cronológicas

A integração entre as ondaletas e a sucessão cronológica prevista, está no fato de que, uma vez realizadas as previsões nas sub-divisões obtidas na etapa de pré-processamento, a transformada inversa da ondaleta na etapa de pós-processamento, que irá trazer de volta a sucessão no nível, isto é, a sucessão original, fará com que transforme também os valores já previstos, em valores previstos no nível.

2.2.2 ESTACIONARIEDADE DE SUCESSÕES CRONOLÓGICAS

Os métodos usados em séries temporais baseiam-se fortemente nos conceitos de estacionariedade, de filtro linear, ergodicidade e independência. Assim, a maioria dos modelos considerados são lineares estacionários fracos, como modelos autoregressivos (AR), de médias móveis (MA) e mistos (ARMA). Em Finanças, existem modelos não-estacionários e não-lineares. Para se descrever séries não estacionárias com comportamento não explosivo, usam-se modelos autoregressivos integrados de médias móveis - ARIMA (MORETTIN e TOLOI, 2004).

Dizer que uma série temporal é estacionária, é dizer que ela se desenvolve no tempo aleatoriamente ao redor de uma média constante (“reversão à média”), refletindo alguma forma de equilíbrio dinâmico estável.

[...] um processo estocástico é estacionário se suas média e variância forem constantes ao longo do tempo e o valor da covariância entre dois períodos de tempo depender apenas da distância ou defasagem entre os dois períodos, e não do período de tempo efetivo em que a covariância é calculada (ENDERS, 2003).

Para Enders (2003), pode-se ter ainda:

- estacionariedade estrita: uma série temporal é dita estritamente estacionária se a distribuição conjunta de qualquer conjunto de n variáveis estocásticas $X(t_1), \dots, X(t_n)$ é a mesma que a distribuição conjunta de $X(t_1 + k), \dots, X(t_n + k)$ para todos n e k .

- estacionariedade fraca: uma série temporal é chamada fracamente estacionária se sua esperança é constante e sua função de autocovariância depender apenas da defasagem, isto é: $E[X(t)] = \mu$, $\mathbb{V}[X(t)] = \sigma^2$ e $C[X(t), X(t+|k|)] = \gamma(|k|)$ se $t \neq t+|k|$.

2.2.2.1 TESTES DE ESTACIONARIEDADE

Box e Pierce (1970, apud MORETTIN, 2002) sugerem um teste para detectar as autocorrelações dos resíduos, que, “apesar de não detectar quebras específicas no comportamento de ruído branco, pode indicar se esses valores são muito altos”. Um aperfeiçoamento deste teste foi feito por Ljung e Box (1978), baseados na função de autocorrelação (FAC). A FAC, no “lag” k , indicada por ρ_k , é definida como:

$$\hat{\rho}_{|k|} = \frac{\sum_{t=1}^n (X_t - \bar{X})(X_{t+|k|} - \bar{X})}{\sum_{t=1}^n (X_t - \bar{X})^2} \quad (2.11)$$

que nada mais é do que a covariância do ativo na defasagem k , dividido pela variância do ativo. Como, tanto a covariância como a variância apresentam as mesmas unidades, ρ_k não tem unidade. Oscila entre -1 e 1 e o gráfico feito colocando $\rho_{|k|}$ contra k é chamado de correlograma amostral da FAC. Na prática, geralmente, somente se trabalha com uma trajetória da série temporal (ergodicidade), pode-se calcular apenas a função de autocorrelação amostral passando a ser denotado por $\hat{\rho}_{|k|}$.

Para realização do teste, tem-se a seguinte hipótese nula:

$$H_0 : \text{todos os } \rho_{|k|} = 0$$

$$H_1 : \text{pelo menos um dos valores de } \rho_{|k|} \neq 0.$$

A estatística teste é:

$$LB = n(n+2) \sum_{k=1}^m \left(\frac{\hat{\rho}_k^2}{n-k} \right) \sim \chi_m^2 \quad (2.12)$$

que segue uma distribuição qui-quadrado com m graus de liberdade (χ_m^2), sendo que k é o tamanho da defasagem e m a defasagem máxima considerada.

A decisão do teste baseia-se no valor da estatística teste, isto é, **se o valor de LB for superior ao valor tabelado para dado nível de significância, deve-se rejeitar H_0 e pelo menos alguns dos $\hat{\rho}_k$ devem ser diferentes de zero.**

Enders (2003) afirma que a análise da FAC amostral requer o conhecimento da distribuição de $\hat{\rho}_{|k|}$, e que para $\rho_{|k|} = 0$ para $k \neq 0$, a variância de $\hat{\rho}_{|k|}$ é aproximadamente igual a $\frac{1}{n}$, sendo que n é o tamanho da amostra. Para n suficientemente grande, tem-se uma distribuição normal para o estimador $\hat{\rho}_{|k|}$.

Bartlett (1946) obteve o seguinte resultado para a variância de $\hat{\rho}_{|k|}$ quando $\rho_{|k|} = 0$ para $k > q$:

$$V(\hat{\rho}_{|k|}) = \frac{1}{n} (1 + 2\rho_1^2 + \dots + 2\rho_q^2) \quad (2.13)$$

para $k = q+1, q+2, \dots$.

De posse da distribuição assintótica de $\hat{\rho}_{|k|}$, pode-se construir um intervalo de confiança e visualmente observar o gráfico da FAC e observar os valores que são significativos (isto é, fora do intervalo de confiança), para um dado nível de significância.

Vandaele (1983) sugere uma “regra prática”: se para k maior do que cinco, o valor do coeficiente de autocorrelação ainda for maior que 0,7, em módulo, a série deve ser considerada não-estacionária. Daí calcula-se a primeira diferença e aplica-se a regra até que se obtenha uma série estacionária.

2.2.2.2 FUNÇÕES DE AUTOCORRELAÇÃO

Box, Jenkins e Reinsel (1994) propuseram a utilização de outro instrumento para facilitar a identificação da estacionariedade de séries temporais que é a função de

autocorrelação parcial (FACP), obtida pelas estimações do valor ϕ_{kk} como uma função de k

$$\phi_{kk} = \frac{|M_k^*|}{|M_k|} \quad (2.14)$$

sendo que M_k é a matriz de autocorrelações para a defasagem k e M_k^* é a matriz M_k com a última coluna substituída pelo vetor de autocorrelações (MORETTIN, 2002).

A função de autocorrelação parcial (FACP) mede a correlação entre os termos X_t e X_{t-k} depois de eliminada a influência dos elementos intermediários, ao passo que a função de autocorrelação (FAC) calcula a correlação apenas entre X_t e X_{t-k} .

O quadro2 apresenta, de forma resumida, o comportamento das FAC e FACP e as características dos respectivos modelos.

Quadro 2: Descrição do comportamento das funções FAC e FACP.

Processo	FAC	FACP
AR(p)	decai, de acordo com exponenciais e/ou senóides amortecidas, infinita em extensão	apresenta um corte em p, sendo que o valor $\phi_{kk} \neq 0$ para os valores de $k \leq p$ e sendo $\phi_{kk} = 0$ para $k > p$
MA(q)	é finita, no sentido de que ela apresenta um truncamento (corte) na defasagem q	decai, de acordo com exponenciais e/ou senóides amortecidas, infinita em extensão
ARMA(p,q)	infinita em extensão, decai, de acordo com exponenciais e/ou senoidais amortecidas após a defasagem q – p	decai, de acordo com exponenciais e/ou senóides amortecidas, infinita em extensão

Fonte: MORETTIN, 2002

A identificação de modelos ARIMA não é uma tarefa muito fácil de ser executada, porque podem ser encontrados diversos modelos candidatos para representar a série quando se observa as FAC e FACP. É comum aparecer certo grau de incerteza na observação, de onde a função é realmente truncada (cortada), ou mesmo verificar se há um decaimento considerável na série.

Empiricamente, pode-se dizer através da análise do gráfico da FAC que a série temporal será não estacionária se o correlograma apresentar uma queda demorada, e ainda, podendo ser regular e contínua, o que dá indícios de não apresentar sazonalidade. Caso essa queda demorada seja descontínua com picos irregulares por um período ou padrão cíclico, é indicativo de presença de sazonalidade.

Enders (2003) afirma que se pode utilizar os critérios AIC (*Akaike's Information Criteria*), proposto por AKAIKE (1973, 1974) e BIC (*Bayesian Information Criteria*) proposto por GEWEKE e MEESE (1981), apud Cromwell, Labys e Terraza (1994). Esses processos são baseados na estimação da variância do $\hat{\varepsilon}_t$, no tamanho da amostra n e nos valores de p e q . Calculam-se, então, as estatísticas:

$$AIC = \ln \widehat{\sigma_{\hat{\varepsilon}_t}^2} + \frac{2(p+q)}{n} \quad (2.15)$$

$$BIC = n \ln \left(\widehat{\sigma_{\hat{\varepsilon}_t}^2} \right) - (n-p-q) \ln \left(1 - \frac{p+q}{n} \right) + \\ + (p+q) \ln(n) + (p+q) \ln \left[(p+q)^{-1} \left(\frac{\sigma_x^2}{\widehat{\sigma_{\hat{\varepsilon}_t}^2}} - 1 \right) \right] \quad (2.16)$$

sendo que $\widehat{\sigma_{\hat{\varepsilon}_t}^2} = E(\hat{\varepsilon}_t^2)$ e $\widehat{\sigma_{\hat{\varepsilon}_t}^2}$ é a máxima verossimilhança estimada para os resíduos.

Escolhe-se aquela especificação que apresentar o menor valor para AIC ou BIC. Os autores recomendam ainda que os critérios AIC e BIC sejam considerados apenas como procedimentos complementares e não alternativos, as FAC e FACP.

2.2.2.3 TESTE DE RAIZ UNITÁRIA – TESTE DE DICKY-PANTULA

Também conhecido como teste de Dickey-Pantula, sendo que não se sabe se a sucessão cronológica possui mais de uma raiz unitária, pois o número de termos de diferenças

defasadas é, muitas vezes, determinado empiricamente. Segundo Cromwell, Labys e Terraza (1994), a idéia é incluir termos suficientes, de modo que o termo do erro seja serialmente independente. As hipóteses ficam então, por exemplo, para testar a presença de três raízes:

Hipóteses Testadas	
$H_0 : 3RU$	
$H_1 : 2RU$	
$H_0 : 2RU$	
$H_1 : 1RU$	
$H_0 : 1RU$	
$H_1 : 0RU$	

A estatística do teste ADF tem a mesma distribuição assintótica que a estatística de DF, de modo que podem ser usados os mesmos valores críticos, conforme a Tabela 1 do Apêndice A.

2.2.3 TESTE DE NORMALIDADE – TESTE DE JARQUE-BERA (JB)

O conceito de distribuição normal é usado, neste trabalho, em combinação com o conceito de independência para prover métodos para especificar os modelos a serem construídos para as sucessões cronológicas. Se a série passar pelo conceito de estacionariedade e pelo teste de normalidade, então é possível especificar um modelo linear que descreva o comportamento da sucessão cronológica (CROMWELL; LABYS e TERRAZA, 1994).

O teste mais comum aplicado para verificar a normalidade é o teste de Jarque e Bera (1987), que é um teste assintótico, isto é, para grandes amostras. A estatística de cálculo é:

$$JB = n \left[\frac{\hat{A}^2}{6} + \frac{(\hat{C} - 3)^2}{24} \right] \quad (2.17)$$

sendo que n é o tamanho da série, \hat{A} é o coeficiente de assimetria dado por

$$\hat{A} = \frac{1}{n\hat{\sigma}^3} \sum_{t=1}^n (X_t - \bar{X})^3 \text{ e } \hat{C} \text{ representa a curtose dado por } \hat{C} = \frac{1}{n\hat{\sigma}^4} \sum_{t=1}^n (X_t - \bar{X})^4.$$

As hipóteses a serem testadas ficam:

H_0 : X_t tem análise de assimetria estatisticamente nula e excesso de curtose zero

H_1 : X_t não tem análise de assimetria estatisticamente nula e excesso de curtose zero

A hipótese nula é H_0 : a série é normal, tem distribuição qui-quadrado com dois graus de liberdade ($\chi^2(2)$). Para uma distribuição normal, o valor da assimetria é zero e a curtose é 3. O valor ($\hat{C} - 3$) é chamado de excesso de curtose.

Assim, para testar a normalidade de uma sucessão cronológica, basta estimar os valores de \hat{A} e \hat{C} , calcular a estatística JB e comparar com o valor tabelado dado um nível de significância previamente adequado. **Outra forma é analisar o nível de significância adotado do teste, sendo que se o p-valor for suficientemente baixo, rejeita-se H_0 , caso contrário se o p-valor for razoavelmente alto, não se rejeita H_0 , ou seja, nada se pode concluir acerca da normalidade da série.**

2.2.4 TESTE BDS – TESTE DE BROCH-DECHERT-SCHEINKMAN

Segundo Cromwell, Labys e Terraza (1994), a independência estatística requer que uma distribuição de probabilidade de uma sucessão cronológica, para um conjunto de variáveis aleatórias, possa ser escrita como um produto das distribuições marginais para cada uma das variáveis aleatórias. Por exemplo, $f(x_1, x_2) = f(x_1) \times f(x_2)$. Na prática, o caso da independência é examinado através dos momentos da distribuição, ou seja, se $t = 2$, por exemplo, o valor esperado de duas variáveis aleatórias é o produto dos valores esperados de cada uma delas, e então, as duas variáveis são parcialmente independentes e também não correlacionadas no caso onde X_t tem variância finita.

O teste para independência pode ser aplicado para testar se a sucessão cronológica apresenta independência estatística, pois, se for, então o processo é estatisticamente um ruído branco independente, que será definido a seguir.

Broch, Dechert e Scheinkman (1987) desenvolveram um método alternativo melhor adaptado às sucessões cronológicas econômico-financeiras. Inicialmente organiza-se os dados em vetores n -dimensionais como $\{X_t\}_{t=1}^n = \{X_{t-n+1}, \dots, X_t\}$, sendo que o parâmetro n é chamado dimensão de imersão. Por questão de simplificação de notação, neste estudo, T será o tamanho total da amostra. A estatística do teste BDS é baseada na integral de correlação entre dois vetores de sucessões cronológicas, definidas através das seguintes quantidades:

$$C_n(k) \equiv \lim_{T \rightarrow \infty} C_{n,T}(k) \quad (2.18)$$

$$C_{n,T}(k) \equiv \frac{\sum_{s=1}^T \sum_{t=s}^T k_{st}}{T(T-1)/2} \quad (2.19)$$

$$k_{st} = \begin{cases} 1, & \text{se } \max |X_{s-i} - X_{t-1}| < k \\ 0, & \text{caso contrário} \end{cases} \quad (2.20)$$

sendo que i é o tamanho da sucessão cronológica e então:

$$C_n(k) = C_1^n(k) \quad (2.21)$$

Para entender melhor este resultado, a razão:

$$\frac{C_{n+1}(k)}{C_n(k)} = \Pr \left(\max_{i=0,1,\dots,n} |X_{s-i} - X_{t-1}| < k \mid \max_{i=1,\dots,n} |X_{s-i} - X_{t-1}| < k \right) = \Pr \left(|X_s - X_t| < k \mid \max_{i=1,\dots,n} |X_{s-i} - X_{t-1}| < k \right)$$

pode ser interpretada como uma probabilidade condicionada. Fixando

$$C_{n+1}(k) / C_n(k) = C_1(k) \text{ para todo } n \text{ positivo, obtém-se a equação (2.21).}$$

Os autores propuseram a seguinte estatística para o teste BDS:

$$J_{n,T}(k) = \sqrt{T} \frac{C_{n,T}(k) - C_{1,T}^n(k)}{\hat{\sigma}_{n,T}(k)} \quad (2.22)$$

sendo que:

$$\hat{\sigma}_{n,T}(k) = \left(4 \left[w^n + 2 \sum_{j=1}^{n-1} w^{n-j} c^{2j} + (n-1)^2 c^{2n} - n^2 w c^{2n-2} \right] \right)^{\frac{1}{2}} \quad \text{em que}$$

$$c = C_{1,n}(k) \quad (2.23)$$

$$w = \frac{6}{n(n-1)(n-2)} \sum_{t=1}^n \sum_{s=t+1}^n \sum_{r=s+1}^n h_k(X_t, X_s, X_r)$$

$$h_k(i, j, k) = \frac{1}{3} \left[k_{i,j} k_{j,w} + k_{i,w} k_{w,j} + k_{j,i} k_{i,w} \right] \quad (2.24)$$

e ainda $C_{n,T}(k)$ e $C_{1,T}(k)$ são as integrais de correlação amostral definidas em (2.21) e $\hat{\sigma}_{n,T}(k)$ é o estimador do desvio-padrão assintótico de $C_{n,T}(k) - C_{1,T}^n(k)$.

As hipóteses do teste BDS são:

H_0 : o processo X_t é estocasticamente independente

H_1 : o processo X_t é estocasticamente dependente

A estatística do teste BDS definida em (2.22) é assintoticamente uma normal padrão (N(0,1)) usando um teste bicaudal e a hipótese nula é a de independência identicamente distribuída é rejeitada ao nível de significância de 5% se $|J_{n,T}(k)| > 1,96$.

Entretanto, é nos resultados dos testes subseqüentes, que ficará claro que a rejeição da hipótese de que as variáveis aleatórias que compõem uma sucessão cronológica apresentam distribuição independente e está mais fortemente associada à existência de dependência linear ou não-linear nas sucessões cronológicas simuladas.

Vale lembrar que foi através das pesquisas em torno da Teoria do Caos determinística que se agregaram procedimentos úteis para tratamento da não-linearidade, especialmente no teste BDS para verificação de não-linearidade em sucessões cronológicas (CORREA; PEREIRA, 1997).

2.2.5 TESTES DE DEPENDÊNCIA LINEAR E NÃO-LINEAR

Na análise de sucessões cronológicas, antes mesmo de se começar a procurar uma especificação geral, para encontrar os devidos modelos que se ajustem ao seu conjunto particular de dados, é importante testar a presença de não-linearidade. Um dos testes mais conhecidos é o de McLeod-Li, o qual visa determinar se existem correlações significantes nos resíduos quadrados de uma equação linear.

2.2.5.1 TESTE DE MCLEOD-LI

O teste é baseado na autocorrelação amostral dos quadrados dos resíduos de uma sucessão cronológica estacionária gaussiana. Para realizar o teste, deve-se estimar a série, estimando o melhor modelo de ajuste linear e chamar os resíduos de $\hat{\varepsilon}_t$.

A hipótese a ser testada é:

$H_0 : X_t^2$ é independente, isto é, a sucessão cronológica é linear

$H_1 : X_t^2$ é dependente, isto é, a sucessão cronológica é não-linear

Considerando então (X_1, X_2, \dots, X_n) os resíduos ajustados de um modelo ARMA e denotando $\hat{\rho}_{X^2}$ o coeficiente de correlação entre os resíduos X_t^2 e X_{t-i}^2 dado por:

$$\hat{\rho}_{X^2}(k) = \frac{\sum_{j=1}^{n-k} (X_j^2 - \hat{\sigma}^2)(X_{j+k}^2 - \hat{\sigma}^2)}{\sum_{j=1}^n (X_j^2 - \hat{\sigma}^2)^2} \quad (2.25)$$

em que $\widehat{\sigma}^2 = \frac{\sum_{j=1}^T X_j^2}{n}$.

A hipótese nula de as primeiras k autocorrelações do quadrado dos retornos serem simultaneamente zero, pode ser testada, utilizando a estatística de Ljung-Box (LB). A rejeição da hipótese nula indica a existência de dependência não-linear, que pode ocorrer na média ou na variância. A estatística LB tem uma distribuição χ^2 assintótica com n graus de liberdade se a sequência $\{X_t^2\}_{t=1}^n$ é não correlacionada.

2.2.5.2 TESTE PARA VERIFICAR A PRESENÇA DE HETEROCEDASTICIDADE CONDICIONADA

Para verificar se a sucessão cronológica apresenta heterocedasticidade condicionada, pode-se aplicar os seguintes testes:

- 1) Teste de Box-Ljung (LB) para a série X_t^2
- 2) Teste dos Multiplicadores de Lagrange (LM), onde a hipótese nula $H_0: \alpha_i = 0$ para todo $i = 1, \dots, r$ na regressão $X_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_r X_{t-r}^2$ para $t = r+1, \dots, n$. A estatística de teste é

$$S = nR^2 \quad (2.26)$$

que segue uma distribuição assintótica $\chi^2(r)$ com r graus de liberdade em que n é o tamanho da amostra e R^2 é o quadrado do coeficiente de correlação múltipla da regressão acima (MORETTIN, 2002).

O teste verifica a existência de correlação dos resíduos quadrados. As hipóteses são:

H_0 : não existe efeitos ARCH no modelo

H_1 : existe efeitos ARCH no modelo

A hipótese nula indica a não necessidade de um modelo com não-linearidade na variância condicionada para os dados. A estatística de teste ARCH-LM é calculada, a partir do teste auxiliar de uma regressão como descrito anteriormente que indica a

hipótese nula de que existe linearidade na variância condicionada, isto é, a variância condicionada poderia ser escrita como combinação linear dos ε_t^2 na forma

$$\varepsilon_t^2 = \alpha_0 + \left(\sum_{s=1}^q \alpha_s \varepsilon_{t-s}^2 \right) + \nu_t, \text{ sendo } \{\nu_t\} \text{ um processo de ruído branco.}$$

2.2.5.3 TESTE PARA DETECTAR NÃO-LINEARIDADE NA ESPERANÇA OU NA VARIÂNCIA CONDICIONADA: TESTE DE HSIEH

Uma vez estabelecido que existe algum tipo de não-linearidade em uma sucessão cronológica, é preciso filtrar o tipo de não-linearidade existente. Hsieh (1989) propôs um teste para detectar os dois tipos diferentes de não-linearidade: aditiva (não-linearidade na média) ou multiplicativa (não-linearidade na variância).

A dependência multiplicativa implica que a esperança condicionada dos resíduos dadas as defasagens anteriores da variável X_t e X_{t-1} é zero:

$$E[X_t | X_{t-1}, \dots, X_{t-k}] = 0.$$

A dependência aditiva implica que a mesma esperança condicionada seja diferente de zero:

$$E[X_t | X_{t-1}, \dots, X_{t-k}] \neq 0$$

As hipóteses do teste de Hsieh são:

$$H_0 : X_t \text{ é não-linear na variância condicionada}$$

$$H_1 : X_t \text{ é não-linear na esperança condicionada}$$

A estatística do teste de Hsieh para a hipótese nula é:

$$\hat{\rho}(i, j) = \frac{\left(\frac{1}{n} \right) \sum_t X_t X_{t-i} X_{t-j}}{\left[\left(\frac{1}{n} \right) \sum_t X_t^2 \right]^{\frac{3}{2}}} \quad (2.27)$$

sendo $i, j > 0$ representam as defasagens do processo $\{X_t\}$.

O teste segue assintoticamente uma distribuição normal padrão $N(0,1)$, sendo que, por exemplo, o valor $|\hat{\rho}(i, j)| < 1,96$ com 95% de confiança não se rejeita a hipótese nula de não-linearidade na variância condicionada.

2.2.6 MODELOS LINEARES – METODOLOGIA DE BOX-JENKINS

A análise de sucessões cronológicas, segundo enfoque de Box, Jenkins e Reinsel (1994), tem por objetivo principal a realização de previsões. Essa metodologia permite que valores futuros de uma série possam ser previstos com base apenas em seus valores presentes e passados. O feitiço deste processo se dá pela exploração da correlação temporal que existe, geralmente, entre os valores exibidos pela série. Por envolverem apenas uma série de tempo, são classificados como modelos paramétricos lineares univariados (MORETTIN, 2002).

Para Enders (2003), a relação temporal desenvolvida por Box-Jenkins é representada formalmente por um conjunto de processos estocásticos genericamente denominados modelos ARMA.

Muitas das idéias hoje conhecidas nos modelos ARIMA foram desenvolvidas, nos anos 70, por George Box e Gwilym Jenkins² e por estas razões os modelos ARIMA são chamados de modelagem ou metodologia Box-Jenkins. Vale lembrar que uma série de tempo pode conter os três “filtros” ou apenas um subconjunto destes, resultando aí em várias alternativas de construção dos modelos passíveis de análise pela metodologia de Box e Jenkins (1976).

Os modelos de esperanças condicionadas (*Autoregressive Moving Average – ARMA*³) fazem parte de uma classe genérica de métodos capazes de gerar modelos, através da

² Para mais detalhes vide BOX, JENKINS e REINSEL, 1994.

³ Há uma certa confusão ocasionada pela nomenclatura dada para ARIMA. Existem duas ambigüidades: 1ª) O termo *Integrated*, representado pela letra I no nome da técnica, refere-se ao processo de diferença fíntia, a série pelo fato dela não ser estacionária, e não pela intenção de integrar a série; 2ª) o termo *moving average* – médias móveis – tem neste método um significado diferente daquele usado nos

combinação de três técnicas matemático-estatísticas, também chamados “filtros⁴”: modelos autoregressivos, médias móveis e diferenciação (MAKRIDAKIS et al., 1983).

Para Enders (2003), um modelo ARIMA pode ter qualquer componente, ou ambos os componentes sazonais e não sazonais. Existem diferentes tipos de modelos ARIMA e a forma geral para se descrever um modelo de esperança condicionada é o SARIMA (p, d, q):(P, D, Q)_s onde, ao primeiro parênteses, cabe a parte não sazonal e o segundo a parte sazonal do modelo que pode ser aditivo ou multiplicativa. No caso dos modelos com a parte sazonal, estes costumam ser encontrados na literatura com o nome de SARIMA, sendo que o S representa periodicidade sazonal. E ainda:

- p se refere à ordem do processo AR – auto-regressivo, não sazonal, incorporado no modelo ARIMA, e P se refere à ordem do processo AR sazonal;
- d se refere à ordem de integração ou diferenciação da parte não sazonal e D a parte sazonal da integração ou diferenciação;
- q se refere à ordem do processo de médias móveis, não sazonal, incorporado no modelo e Q a ordem da parte sazonal de médias móveis.

A construção dos modelos estocásticos discretos para séries temporais é feita via ruído branco. A sequência $\{\varepsilon_t\}_{t=1}^n$ é dita ser um ruído branco se cada valor da sequência tiver esperança nula, variância constante e forem serialmente não correlacionados (HAMILTON, 1994).

Formalmente, pode-se definir, segundo Hamilton (1994), ruído branco (RB) como uma série de erros estocásticos $\{\varepsilon_t\}_{t=1}^n$ se:

- (i) A esperança dos erros é nula:

$$\{E[\varepsilon_t] = 0\}_{t=1}^n \quad (2.28)$$

- (ii) Os erros entre instantes de tempo distintos são não correlacionados e todas as suas variâncias são iguais (homocedasticidade):

métodos de suavização, referindo-se à série dos termos dos erros em diferentes períodos de tempo considerados: $Y_t = \mu + b_1 e_{t-1} + b_2 e_{t-2} + \dots + b_k e_{t-k} + e_t$.

⁴ Entende-se por filtragem ou tratamento, um caminho antes de analisar os dados, visando melhorias na interpretação e utilização dos dados além de resumir os efeitos na autocovariância (HAMILTON, 1994).

$$\left\{ \begin{array}{l} C[\varepsilon_t; \varepsilon_s] = 0, s \neq t \\ V[\varepsilon_t] = \sigma^2 \end{array} \right\}_{t=1}^n$$

$$\therefore \{\varepsilon_t\}_{t=1}^n \sim \text{RB} \quad (2.29)$$

Pode-se definir ainda ruído branco independente (RBI) como uma série de erros estocásticos $\{\varepsilon_t\}_{t=1}^n$ se:

(i) A esperança dos erros é nula:

$$\{E[\varepsilon_t] = 0\}_{t=1}^n \quad (2.30)$$

(ii) os erros são estocasticamente independentes entre instantes de tempo distintos:

$$\begin{aligned} \varepsilon_t &= e_t \in \mathbb{R} \\ f_{\varepsilon_1; \varepsilon_2; \dots; \varepsilon_n} \left(e_1; e_2; \dots; e_n \right) &= \prod_{t=1}^n f_{\varepsilon_t} \left(e_t \right) \\ &\vdots \\ \{\varepsilon_t\}_{t=1}^n &\sim \text{RBI} \end{aligned} \quad (2.31)$$

Define-se ruído branco independente normal (RBIN) como uma série de erros estocásticos $\{\varepsilon_t\}_{t=1}^n$ se:

(i) esperança dos erros é nula:

$$\{E[\varepsilon_t] = 0\}_{t=1}^n \quad (2.32)$$

(ii) os erros são estocasticamente independentes e normalmente distribuídos entre instantes de tempo distintos:

$$\begin{aligned} \{\varepsilon_t \sim N(0; \sigma^2)\}_{t=1}^n \\ \varepsilon_t &= e_t \in \mathbb{R} \\ f_{\varepsilon_t} \left(e_t \right) &= \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\frac{e_t^2}{\sigma^2} \right) \right\} \\ f_{\varepsilon_1; \varepsilon_2; \dots; \varepsilon_n} \left(e_1; e_2; \dots; e_n \right) &= \prod_{t=1}^n f_{\varepsilon_t} \left(e_t \right) \end{aligned}$$

$$\begin{aligned}
\varepsilon_n &= \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \sim N_n \left(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}; \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \ddots & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & \sigma^2 \end{bmatrix} \right) \\
&= N_n \left[0; \sigma^2 I_n \right] \\
&\therefore \{\varepsilon_t\}_{t=1}^n \sim \text{RBIN}
\end{aligned} \tag{2.33}$$

2.2.6.1 MODELOS AUTO-REGRESSIVOS (AR)

Segundo Box, Jenkins e Reinsel (1994), a parte auto-regressiva dos modelos ARMA usa variáveis independentes para prever a variável dependente. O termo auto-regressivo indica que as séries de valores a serem previstos provem dos valores passados para prever o valor futuro quando há autocorrelação entre as observações. Quando se cita ordem do modelo, faz-se referência à diferença entre o valor a ser previsto e o número de defasagens usadas como preditores. Se a série de valores utilizar 2 valores passados para prever o valor atual, tem-se um modelo auto-regressivo de segunda ordem, indicado por AR(2), isto é, o valor X_t da série dependerá dos valores X_{t-1} e X_{t-2} .

A equação de previsão para um modelo AR(2) fica:

$$E_t[X_t] = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t \tag{2.34}$$

os valores de ϕ_1 e ϕ_2 são os coeficientes (parâmetros) a serem estimados pelo modelo, bem como a componente do erro $\varepsilon_t \sim N(0, \sigma^2)$, que indica o erro ou o termo aleatório não, correlacionado com média aritmética zero e variância constante. Respeitadas aqui as condições de estacionariedade do modelo.

De um modo geral, tem-se para um modelo auto-regressivo de ordem p :

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t \quad (2.35)$$

Da mesma forma, poder-se-ia escrever via um operador de atraso L :

$$LX_t = X_{t-1} \quad (2.36)$$

$$L^2 X_t = LLX_t = LX_{t-1} = X_{t-2} \quad (2.37)$$

$$L^n X_t = X_{t-n} \quad (2.38)$$

sendo que,

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p} = \varepsilon_t \quad (2.39)$$

$$X_t - \phi_1 LX_t - \phi_2 L^2 X_t - \dots - \phi_p L^p X_t = \varepsilon_t \quad (2.40)$$

embora o operador L não seja um número, pode ser tratado algebricamente como se fosse. Ao colocar X_t em evidência:

$$X_t(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) = \varepsilon_t \quad (2.41)$$

Aparece, então, um polinômio no operador L , que se pode chamar simplesmente de $\Theta_p(L)$, e então o modelo fica resumido a:

$$\Theta_p(L)X_t = \varepsilon_t \quad (2.42)$$

2.2.6.2 MODELOS DE MÉDIAS MÓVEIS (MA)

De acordo com Box, Jenkins e Reinsel (1994), os termos de médias móveis dos modelos ARIMA representam as defasagens dos valores dos erros passado e presente, muitas vezes, chamado de ruído branco⁵, usado quando há autocorrelação entre os resíduos. Pode-se interpretar também os termos de médias móveis como eventos externos ou choques no sistema, isto é, uma mudança nos valores não previstos no ambiente pode influenciar tanto os valores passados como os valores futuros. Assim, o

⁵ O termo do erro é chamado de “ruído branco” se ele possui média zero, variância constante σ^2 e é não autocorrelacionado. Ao se estimar um modelo, deseja-se que o erro seja um “ruído branco”.

componente de erro para o período de tempo atual se relaciona ao valor da série no futuro. A ordem q – do termo de médias móveis se refere ao tamanho da defasagem entre o erro e a variável da série. Dessa forma, se a série é influenciada pelo erro do modelo defasado de 1 período, tem-se um processo MA(1), cuja equação seria:

$$X_t = \theta_1 \varepsilon_{t-1} + \varepsilon_t \quad (2.43)$$

sendo que θ_1 é parâmetro, isto é, o coeficiente do primeiro termo do erro de média móvel e o componente do erro $\varepsilon_t \sim N(0, \sigma^2)$ que indica o erro ou termo aleatório não correlacionado com média aritmética zero e variância constante.

De um modo geral, tem-se para um modelo de médias móveis de ordem q :

$$X_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (2.44)$$

Da mesma forma que no processo auto-regressivo, poder-se-ia escrever via um operador de atraso L :

$$X_t = \theta_1 L \varepsilon_t + \theta_2 L^2 \varepsilon_t + \dots + \theta_q L^q \varepsilon_t + \varepsilon_t \quad (2.45)$$

Ao colocar ε_t em evidência e atribuindo $\Phi_q(L)$ a $1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q$, fica:

$$X_t = \varepsilon_t \Phi_q(L) \quad (2.46)$$

Uma característica importante de um modelo MA(q), e conseqüentemente do modelo ARMA(p, q), é que as variáveis na equação (2.44), ou seja, ε_{t-1} até ε_{t-q} , são não observáveis e têm de ser estimadas, usando a amostra de dados disponível. Para que isso não cause problemas, é usual manter o valor de q o tanto menor quanto possível. Na prática, esse valor de q é freqüentemente tomado como sendo 0, 1 ou 2. À primeira vista, pode parecer que X_t não depende de seu próprio passado quando um modelo MA(q) descreve esta variável.

2.2.6.3 MODELOS ARIMA

São modelos que são aplicados no caso específico de séries não estacionárias ou estacionarizadas e são compostos pelos três filtros: auto-regressivo, média móvel e diferenciação. Para se montar um modelo ARIMA para uma sucessão cronológica, há três estágios a se considerar: identificação, estimação e verificação (diagnóstico) (MORETTIN, 2002).

Box, Jenkins e Reinsel (1994) frisam que a componente de integração ou diferenciação no modelo ARIMA provê um meio de representar a tendência dentro de um modelo de sucessão cronológica. Diferenciar a série, significa subtrair os valores dos valores adjacentes para avaliar os componentes restantes do modelo. A diferenciação pode ocorrer tanto no nível sazonal quanto não sazonal, podendo ocorrer diferenciações sucessivas, embora não seja, muitas vezes, necessário. A forma de diferenciar a série em nível não sazonal é:

$$Z_t = X_t - X_{t-1} \quad (2.47)$$

e, assim, sucessivamente.

A metodologia proposta por Box e Jenkins (1976) define três passos para a construção do modelo de previsão:

1) identificação: consiste em descobrir qual entre as várias versões possíveis de um modelo ARIMA descreve o comportamento da série. O objetivo desta etapa é encontrar os filtros que irão compor o modelo, estabelecendo os valores de p , d e q . A escolha é feita, principalmente, com base o gráfico das funções de autocorrelações (FAC) e na função autocorrelações parciais (FACP) proposta por Box, Jenkins e Reinsel (1994), as quais se espera que representem adequadamente as respectivas quantidades teóricas, que são desconhecidas no modelo. De maneira geral, a FAC é usada para identificação dos processos autoregressivos (AR) e a FACP, para identificar os processos de médias móveis (MA). A parte de integração dependerá da estacionariedade da série, e caso não seja estacionária, usar-se-á o processo de diferenciação para tornar a série estacionária.

Séries temporais não estacionárias apresentam fortes correlações seriais, ou seja, os valores dos coeficientes de autocorrelação declinam lentamente à medida que o valor da defasagem k aumenta. Assim sendo, o gráfico de uma FAC com valores muito altos inicialmente e que declinam lentamente para um valor estatisticamente igual a zero, à medida que o valor da defasagem k aumenta, é um indicativo de que a série é não-estacionária e precisa ser diferenciada (HAMILTON, 1994).

Caso contrário, a série será estacionária, o valor de d será zero e passa-se para a identificação dos demais filtros AR e MA. Cabe ressaltar aqui, que nem sempre a diferenciação é suficiente para estacionarizar a série, sendo que se esta tiver tendência determinística, deverá ser removida antes ou se usar outro modelo. Se a não estacionariedade estiver associada à variância da série, pode-se optar por usar transformações como logaritmo neperiano ou raiz-quadrada (VASCONCELLOS e ALVES, 2000).

2) estimação: o segundo passo consiste em estimar os parâmetros p , d e q do modelo e a variância do ruído branco σ^2 , observando sempre a minimização dos critérios AIC e BIC ⁶.

3) verificação: consiste em verificar a eficácia da estimação dos parâmetros do modelo. Caso falhe na eficácia dos parâmetros, o processo é refeito desde a identificação para se obter o ajuste ótimo do modelo.

Genericamente, um modelo ARMA (p,q) seria:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (2.48)$$

Ou ainda:

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (2.49)$$

$$X_t - \phi_1 L X_t - \phi_2 L^2 X_t - \dots - \phi_p L^p X_t = \varepsilon_t + \theta_1 L \varepsilon_t + \theta_2 L^2 \varepsilon_t + \dots + \theta_q L^q \varepsilon_t \quad (2.50)$$

Ou, simplesmente:

⁶ Critérios AIC (*Akaike's Information Criteria*) proposto por AKAIKE (1973, 1974) e BIC (*Bayesian Information Criteria*) proposto por GEWEKE e MEESE(1981), apud Cromwell, Labys e Terraza (1994).

$$\Theta_p(L)X_t = \varepsilon_t \Phi_q(L) \quad (2.51)$$

sendo $Y_t = \ln\left(\frac{X_t}{X_{t-1}}\right) = \Delta \ln(X_t)$ e

$$\begin{aligned} \varepsilon_t^{ARIMA(p,1,q)} &= \frac{\Theta_p(L)}{\Phi_q(L)} Y_t = \Theta_p(L) \Phi_q^{-1}(L) \Delta \ln X_t \\ \varepsilon_t^{ARIMA(p,1,q)} &\sim RB \end{aligned} \quad (2.52)$$

Da equação (2.2) com a equação (2.52), tem-se:

$$\varepsilon_t^{ARIMA} = e_{t-1}(1) \equiv Y_t - E_{t-1}[Y_t] \quad (2.53)$$

Poderia ter-se ainda caso a série seja não estacionária, definir uma nova variável Z_t como sendo a primeira diferença de X_t , isto é:

$$Z_t = X_t - X_{t-1} = \Delta X_t \quad (2.54)$$

Se X_t não é estacionária, mas Z_t é, diz-se que X_t é integrada de ordem 1, ou I(1). Às vezes, a primeira diferença não é suficiente para estacionarizar a série, e então temos de tomar a segunda diferença, isto é, a diferença da diferença:

$$Z_t = \Delta^2 X_t = \Delta X_t - \Delta X_{t-1} \quad (2.55)$$

Agora, se X_t não é uma variável estacionária, mas sua d -ésima diferença é, um modelo ARIMA (p,d,q) , ficaria:

$$Z_t = \Delta^d X_t \quad (2.56)$$

$$Z_t = \phi_1 \Delta^d X_{t-1} + \phi_2 \Delta^d X_{t-2} + \dots + \phi_p \Delta^d X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (2.57)$$

Indicando $W_t = \Delta^d X_t$ tem-se:

$$Z_t = \phi_1 W_{t-1} + \phi_2 W_{t-2} + \dots + \phi_p W_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (2.58)$$

Alternativamente, utilizando o operador L de defasagem, tem-se:

$$(1 - \phi_1 L - \dots + \phi_p L^p) w_t = (1 - \theta_1 L - \dots - \theta_q L^q) \varepsilon_t \quad (2.59)$$

$$W_t = (1 - L)^d X_t \quad (2.60)$$

$$(1 - L)^d \Theta_p(L) = \Phi_q(L) \varepsilon_t \quad (2.61)$$

Nesse caso, $(1-L)^d \Theta_p(L) = 0$ apresenta d raízes sobre o círculo unitário (d raízes unitárias) e p raízes fora do círculo unitário.

Um caso especial do processo não estacionário homogêneo é o ARIMA(0,1,0), denominado *random walk* ou passeio aleatório: $X_t = X_{t-1} + \varepsilon_t$.

A série X_t é integrada de ordem 1 e sua primeira diferença é ARMA(0,0), ou seja, um RB.

2.2.7 MODELOS NÃO-LINEARES

Em muitas situações, uma sucessão cronológica pode exibir comportamentos incompatíveis com a formulação de um processo linear. Os comportamentos incluem, por exemplo, mudanças repentinas, variância condicionada, evoluindo no tempo e irreversibilidade no tempo (MORETTIN, 2002).

Para Enders (2003), a teoria econômica sugere que um grande número de sucessões cronológicas econômico-financeiras forneça evidências de exibir um comportamento não-linear, conforme sugerido por Morettin (2002), visto que, as séries financeiras apresentam instabilidades no tempo e peculiaridades institucionais que fazem com que os modelos de coeficientes fixos lineares e gaussianos sejam inadequados para aproximar tais séries. O objetivo passa a ser então, modelar o que se chama de volatilidade, que é a variância condicional de uma variável, comumente um retorno.

Para Campbell, Lo e Mackinlay (1997), a volatilidade não é medida diretamente e se manifesta numa série temporal financeira da seguinte maneira:

- a volatilidade aparece em grupos, de maior ou de menor variabilidade;
- a volatilidade evolui continuamente no tempo, podendo ser considerada estacionária;
- reage a valores positivos ou negativos de maneira diferente.

Tsay (2002) define um modelo não-linear como sendo aquele em que a condição de primeira ordem para a estimação dos parâmetros por mínimos quadrados gera funções não-lineares nos parâmetros. Dessa forma, a não-linearidade é definida em termos das técnicas necessárias para estimar os parâmetros, e não da forma funcional da regressão.

Para Campbell, Lo e Mackinlay (1997), alguns modelos de sucessões cronológicas típicas relacionam umas séries de tempo observadas a uma sucessão subjacentes de choques aleatórios ε_t . Na análise de sucessões cronológicas lineares, assume-se que estes choques são não correlacionados, mas não necessariamente assume-se que sejam independentes e identicamente distribuídos.

Segundo Hamilton (1994), pelo teorema da representação de Wold, qualquer série temporal pode ser escrita como uma média móvel linear, de ordem finita para tais choques estocásticos e esta representação resume a variância não condicionada e autocovariâncias da série.

Na análise de modelos não-lineares, Campbell, Lo e Mackinlay (1997) afirmam que os valores de ε_t são, em geral, supostos independentes e identicamente distribuídos (I.I.D), mas buscam uma função não-linear relacionando à série de X_t com os choques estocásticos, cuja representação geral é:

$$X_t = f\left(\{\varepsilon_s\}_{s=t}^{-\infty}\right) = f\left(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots\right) \quad (2.62)$$

sendo que se assume que os choques possuem esperança nula e variância unitária, ε_t é um choque estocástico contemporâneo e $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$ são os choques estocásticos passados, isto é $\{\varepsilon_s\}_{s=t}^{-\infty} \sim \text{RBI}$.

Pode-se, então, aproximar X_t por Taylor em torno de $\varepsilon_t^0 = 0$:

$$\begin{aligned} X_t = & f\left(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots\right) + \left(\varepsilon_t - \varepsilon_t^0\right) \frac{\partial}{\partial \varepsilon_t} f\left(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots\right) \Big|_{\varepsilon_t = \varepsilon_t^0} + \\ & + \left(\varepsilon_t - \varepsilon_t^0\right)^2 \frac{\partial^2}{\partial \varepsilon_t^2} f\left(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots\right) \Big|_{\varepsilon_t = \varepsilon_t^0} + \dots \end{aligned} \quad (2.63)$$

Segundo Campbell, Lo e Mackinlay (1997), considerando até a segunda expansão onde se tem para $p \geq 2$, $E\left[\left(X_t - E_{t-1}[X_t]\right)^p\right] \equiv h^p(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots) E[\varepsilon_t^p]$, chamado momento centrado condicionado de ordem p , e definindo $g(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots) \equiv E_{t-1}[X_t]$, pois $E_{t-1}[X_t]$ é o preditor ótimo de X_t pelo critério do erro quadrático médio, fazendo com que a equação (2.63) fique:

$$X_t \cong f(0, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) + \varepsilon_t \frac{\partial}{\partial \varepsilon_t} f(0, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) \quad (2.64)$$

$$X_t \cong g(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots) + \varepsilon_t h(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = g(\cdot) + \varepsilon_t h(\cdot) \quad (2.65)$$

de modo que a função g representa a média condicional e h^2 é a variância condicionada. Se a função g for não-linear, o modelo diz-se não-linear na esperança, enquanto se h for não-linear, o modelo diz-se não-linear na variância.

Considerando $\{\varepsilon_t\} \sim RBIN$, então $E[\varepsilon_t^2] = 1$, de onde se tem:

$$E\left[\left(X_t - E_{t-1}[X_t]\right)^2\right] = \mathbb{V}_{t-1}[X_t] = h^2(\cdot) E[\varepsilon_t^2] \quad (2.66)$$

$$\mathbb{V}_{t-1}[X_t] = h^2(\cdot) \times 1 \quad (2.67)$$

$$h(\cdot) = \sqrt{\mathbb{V}_{t-1}[X_t]} \quad (2.68)$$

Dessa forma, juntando-se a equação (2.65) com o resultado de (2.68), tem-se:

$$X_t \cong g(\cdot) + \varepsilon_t h(\cdot) = E_{t-1}[X_t] + \varepsilon_t \sqrt{\mathbb{V}_{t-1}[X_t]} \quad (2.69)$$

$$X_t - E_{t-1}[X_t] \cong \varepsilon_t \sqrt{\mathbb{V}_{t-1}[X_t]} \quad (2.70)$$

Da equação (2.2), tem-se $e_{t-1}(1) \equiv X_t - E_{t-1}[X_t]$ e juntamente com a equação (2.70):

$$e_{t-1}(1) \equiv X_t - E_{t-1}[X_t] = \varepsilon_t \sqrt{\mathbb{V}_{t-1}[X_t]} \quad (2.71)$$

$$e_{t-1}(1) \cong \varepsilon_t \sqrt{\mathbb{V}_{t-1}[e_{t-1}(1)]} \quad (2.72)$$

sendo que $e_{t-1}(1) \equiv X_t - E_{t-1}[X_t]$ representa o erro de previsão um passo à frente e lembrando que $\mathbb{V}_{t-1}[e_{t-1}(1)] = \mathbb{V}_{t-1}[X_t - E_{t-1}[X_t]] = \mathbb{V}_{t-1}[X_t]$. (2.73)

Segundo Tsay (2002), se os modelos com não-linearidade na esperança condicionada se mostrarem adequados para séries financeiras, eles podem ser utilizados em previsão como concorrentes ou complementares a outros métodos. Nos conglomerados de valores extremos, um dos fatos estilizados de dados financeiros pode ser explicado por modelos não-lineares, embora possam ser decorrência de *outliers* e/ou mudanças estruturais.

Para Enders (2003), os modelos ARCH e GARCH são muito aplicados a estes tipos de dados por captarem os fatos estilizados citados pelos autores acima. Existem ainda outras classes de modelos não-lineares como o modelo de volatilidade estocástica que modela a variância através de um processo não observado, tentando captar a notícia que chega ao mercado, os modelos TAR (Limiar Auto-regressivo) que incluem não-linearidades na esperança e os bilineares.

2.2.7.1 VARIÂNCIA CONDICIONADA

Segundo Soares e Castelar (2003), dentre as hipóteses do modelo clássico de regressão linear, tem-se a que estabelece que os distúrbios estocásticos têm a mesma variância em todas as observações, o que recebe o nome de homocedasticidade. A violação deste pressuposto recebe o nome de heterocedasticidade.

Muitas sucessões cronológicas econômico-financeiras apresentam informações adicionais discrepantes e *outliers*, provindos de inovações que não acontecem com frequência. Uma vez que os dados refletem resultados de negócios entre compradores e vendedores, por exemplo, o mercado de ações, várias fontes de eventos exógenos podem ter um impacto sobre o padrão dos preços dos ativos. Algumas notícias podem levar a várias interpretações diferentes e também que certos eventos econômicos específicos, tal como uma crise de petróleo, pode durar por algum tempo. Frequentemente, observa-se que grandes quantidades positivas e negativas de observações de séries temporais financeiras podem aparecer em *clusters*

(agrupamentos). O autor frisa, ainda, que uma abordagem para tratar esta situação e explorar o fato de que os *outliers* aparecem em *clusters* e tentar construir um modelo de série temporal para estes próprios *outliers*. Ou seja, devido à presença de conjuntos de valores discrepantes, a variância das séries temporais financeiras deverão variar em função do tempo e, conseqüentemente, os intervalos de previsão também deverão para cada nível. Logo, os modelos não-lineares são aplicados quando as séries são não serialmente correlacionadas, mas dependentes.

2.2.7.2 MODELOS NÃO-LINEARES NA MÉDIA

Para os modelos da equação (2.65) vista anteriormente, pode-se gerar, existe uma grande variedade de modelos não-lineares disponíveis na literatura. Destacam-se, a seguir, alguns destes modelos. Para mais detalhes, ver Tsay (2002).

2.2.7.2.1 MODELOS POLINOMIAIS

Para este modelo, a função $g(\cdot)$ fica:

$$g(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = \sum_{i=1}^{\infty} \alpha_i \varepsilon_{t-i} + \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \beta_{ij} \varepsilon_{t-i} \varepsilon_{t-j} + \dots \quad (2.74)$$

Segundo Morettin e Toloi (2004), esta é uma expansão envolvendo termos lineares, bilineares etc. dos ε_t .

2.2.7.2.2 MODELOS BILINEARES

Para este modelo, tem-se:

$$g(\varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = \sum_{i=1}^{\infty} \alpha_i \varepsilon_{t-i} + \sum_{i=1}^{\infty} \beta_i X_{t-i} + \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \gamma_{ij} X_{t-i} \varepsilon_{t-j} + \dots \quad (2.75)$$

2.2.7.2.3 MODELOS LINEARES POR PARTES

Segundo Cromwell, Labys e Terraza (1994), a idéia básica da classe dos modelos lineares por partes, também chamados de modelos TAR (“*threshold autoregressive*”) é a linearização por partes por meio da introdução de limiares.

Um modelo TAR(2,1,1) fica:

$$X_t = \begin{cases} \alpha_1 X_{t-1} + \varepsilon_1, & \text{se } X_{t-1} < d \\ \alpha_2 X_{t-1} + \varepsilon_2, & \text{se } X_{t-1} > d \end{cases} \quad (2.76)$$

Aqui o processo tem dois regimes, dependendo de X_{t-1} ser maior ou menor do que o limiar d . Segundo Morettin (2004), os modelos com quebras estruturais podem ser incluídos nessa classe de modelos e um dos objetivos é estimar os pontos de mudanças de regimes.

2.2.7.2.4 MODELO AUTOREGRESSIVO EXPONENCIAL

Segundo Cromwell, Labys e Terraza (1994), um modelo autoregressivo exponencial (EXPAR (p)) tem a forma:

$$X_t = \left\{ \varphi_1 + \pi_1 e^{(-\tau(X_{t-1}^2))} \right\} X_{t-1} + \dots + \left\{ \varphi_p + \pi_p e^{(-\tau(X_{t-p}^2))} \right\} X_{t-p} + \varepsilon_t \quad (2.77)$$

onde ε_t é um processo independente e identicamente distribuído com média zero e variância finita (RBI) e φ_p e π_p são os parâmetros e $\tau > 0$.

2.2.7.3 MODELOS NÃO-LINEARES NA VARIÂNCIA CONDICIONADA – MODELOS GARCH

É uma generalização dos modelos ARCH (*Autoregressive Conditional Heteroskedasticity*) isto é, de heterocedasticidade autoregressiva condicionada, proposto inicialmente por Engle (1982), que é conhecido por GARCH foi proposto por Bollerslev (1986), e pode ser usado para descrever a volatilidade com menos parâmetros do que um modelo ARCH. Apresentam as mesmas vantagens e desvantagens dos modelos ARCH, conforme se observa no quadro a seguir. Altas volatilidades são precedidas de

grandes retornos ou volatilidades, onde se observam os grupos de volatilidades presentes nas séries financeiras, isto é, é comum o fato de que grandes valores num determinado instante de tempo sejam seguidos também por valores elevados nos períodos subseqüentes, não necessariamente na mesma direção.

Quadro 3: Vantagens e desvantagens dos Modelos ARCH

Vantagens dos modelos ARCH	Desvantagens do modelo ARCH
A distribuição das taxas de retorno é leptocúrtica.	Tratar os retornos positivos e negativos da mesma forma, já que os quadrados dos retornos entram na forma da volatilidade e na prática, sabe-se que a volatilidade reage de forma diferente a retornos positivos e negativos. Também devido ao fato de trabalhar com retornos ao quadrado, alguns retornos grandes e isolados podem levar a super-previsões.

Fonte: Morettin, 2002

O modelo ARMA especifica uma expressão para a esperança condicionada. O modelo de heterocedasticidade condicionada auto-regressivo generalizado (GARCH) especifica uma expressão para a variância condicionada e modela movimentos persistentes na volatilidade de maneira parcimoniosa. O GARCH é significativo para pesquisadores de “Economia Estocástica” ou “da Incerteza”. Modelos teóricos, usando variância como uma medida de incerteza, como, por exemplo, a análise de média-variância, podem ser testados, usando esperança e variância condicionadas por um conjunto informacional.

Define-se um Conjunto informacional ordem p até o instante t (CI) como: uma série de variáveis estocásticas p -dimensionais indexadas por instantes de tempo equiespaçados

$$\mathcal{F}_t^p = \left\{ \Phi_i \right\}_{i=-\infty}^{\infty} \quad \text{é conjunto informacional de ordem } p=1 \text{ até o instante } t \text{ inclusive.}$$

Φ_i is a $(p \times 1)$ vector

E também pode-se definir um Conjunto informacional ordem 1 até o instante t (CI): uma série de variáveis estocásticas unidimensionais indexadas por instantes de tempo

equiespaçados $\mathcal{F}_t = \{\Phi_i\}_{i=-\infty}^{\infty}$ é conjunto informacional de ordem 1 até o instante t inclusive.

O \mathcal{F}_t pode ser entendido como um conjunto de informações disponíveis e selecionadas pelo pesquisador até o instante t inclusive. Intuitivamente, espera-se que o conjunto informacional auxilie na diminuição da incerteza na previsão de uma variável estocástica por meio do condicionamento. Uma propriedade importante é a de filtragem: $\mathcal{F}_{t-1} \subset \mathcal{F}_t, \forall t$.

Um modelo GARCH(p,q) para o $\varepsilon_{t-1}^{ARIMA(p,1,q)}(1) = \varepsilon_t^{ARIMA(p,1,q)}$ tem-se:

$$\varepsilon_t^{ARIMA(p,1,q)} = \frac{\Theta_p(L)}{\Phi_q(L)} Y_t = \Theta_p(L) \Phi_q^{-1}(L) \Delta \ln X_t \text{ da equação (2.56) é definido a partir da}$$

equação (2.72):

$$\varepsilon_t \cong \nu_t \sqrt{\mathbb{V}_{t-1}[\varepsilon_t]} \quad (2.78)$$

sendo que:

$$\{\nu_t\}_{t=1}^n \sim \text{RBIN}(0;1) \quad \text{e} \quad \{\varepsilon_t\} \sim \text{RB} \\ \{\nu_t\} \text{ e } \{\varepsilon_t\} \text{ independentes}$$

$$\mathbb{V}_{t-1}[\varepsilon_t^{ARIMA(p,1,q)}] \equiv \omega + \sum_{i=1}^s \beta_i \mathbb{V}_{t-i-1}[\varepsilon_{t-i}^{ARIMA(p,1,q)}] + \sum_{j=1}^r \alpha_j (\varepsilon_{t-j}^2)^{\text{GARCH}(r,s)} > 0 \\ B_r(L) \mathbb{V}_{t-1}[\varepsilon_t] = A_s(L) (\varepsilon_t^2)^{\text{GARCH}(r,s)} \quad (2.79)$$

$$B_r^{-1}(L) A_s(L) (\varepsilon_t^{ARIMA(p,1,q)})^2 = \mathbb{E}_{t-1} \left[(\varepsilon_t^{ARIMA(p,1,q)})^2 \right]$$

$$\omega > 0$$

$$\beta_i > 0, i = 1, 2, \dots, r \text{ (parâmetros GARCH)} \quad (2.80)$$

$$\alpha_j > 0, j = 1, 2, \dots, s \text{ (parâmetros ARCH)}$$

$$\mathbb{V}[\varepsilon_t^{ARIMA(p,1,q)}] = \frac{\omega}{1 - \sum_{i=1}^r \alpha_i - \sum_{i=1}^s \beta_i} = \sigma^2 > 0 \quad (2.81)$$

$$1 - \sum_{i=1}^r \alpha_i - \sum_{i=1}^s \beta_i > 0 \Rightarrow \mathbb{V}[\varepsilon_t^{ARIMA(p,1,q)}] < \infty \quad (2.82)$$

Dessa forma, tem-se:

$$X_t - \underbrace{\mathbb{E}_{t-1}[X_t]}_{\text{ARIMA}} \cong \nu_t \sqrt{\underbrace{\mathbb{V}_{t-1}[\varepsilon_t^{ARIMA(p,1,q)}]}_{\text{GARCH}}} \quad (2.83)$$

Da equação (2.79), tem-se:

$$\begin{aligned}
 \Theta_p(L)\Phi_q^{-1}(L)\Delta \ln X_t &\approx v_t \sqrt{B_r^{-1}(L)A_s(L)\left(\varepsilon_t^{\text{ARIMA}(p,1,q)}\right)^2} \\
 \Theta_p^{-2}(L)\Phi_q^2(L)(\Delta \ln X_t)^2 &\cong v_t^2 B_r^{-1}(L)A_s(L)\left(\varepsilon_t^{\text{ARIMA}(p,1,q)}\right)^2 \\
 \left[\left(\varepsilon_t^{\text{ARIMA}(p,1,q)}\right)^2\right]^{\text{GARCH}(r,s)} &\cong v_t^{-2}\Theta_p^{-2}(L)\Phi_q^2(L)B_r(L)A_s^{-1}(L)\left[(\Delta \ln X_t)^2\right]
 \end{aligned} \tag{2.84}$$

E ainda:

$$\begin{aligned}
 \mathbb{E}_{t-1}\left[\varepsilon_t^{\text{ARIMA}(p,1,q)}\right] &= \mathbb{E}\left[\varepsilon_t^{\text{ARIMA}(p,1,q)}\right] = 0 \\
 \mathbb{V}_{t-1}\left[\varepsilon_t^{\text{ARIMA}(p,1,q)}\right] &= \mathbb{E}_{t-1}\left[\left(\varepsilon_t^{\text{ARIMA}(p,1,q)}\right)^2\right] \\
 \mathbb{E}\left[\mathbb{V}_{t-1}\left[\varepsilon_t^{\text{ARIMA}(p,1,q)}\right]\right] &= \mathbb{E}\left[\mathbb{E}_{t-1}\left[\left(\varepsilon_t^{\text{ARIMA}(p,1,q)}\right)^2\right]\right] = \sigma^2
 \end{aligned} \tag{2.85}$$

Os parâmetros são estimados, por exemplo, por meio de métodos de maximização da chamada função de máxima verossimilhança condicionada. A descrição do processo não será feita, neste trabalho, por não estar entre os objetivos. Mais detalhes podem ser vistos em BOLLERSLEV (1986, 1987) e MILLS (1999).

Toloi (1980) comparou as técnicas Box-Jenkins, alisamento exponencial, método de filtragem adaptativa, método de regressão para 8 séries temporais distintas para previsão de curto, médio e longo prazos. A autora identificou, com exceção da previsão a longo prazo, que o método de Box-Jenkins é o mais eficiente por apresentar, na maioria das comparações, um menor erro quadrático médio. O método de Holt-Winters foi considerado sempre o pior, quando da análise das séries de manchas solares, embarque de metal, concentração de um processo químico e imigração. Todavia, o método de Holt-Winters apresenta um comportamento significativo, para ocupar um segundo lugar, nas séries de produção de leite e vendas de champanhe, para qualquer distância de previsão à sua origem. E devido à sua simplicidade, o método de Alisamento Exponencial Simples apresentou um bom comportamento, em alguns casos esporádicos, e, na maioria das vezes, quando trabalhamos com séries não sazonais.

Bollerslev, Chou e Kroner (1992) mostraram que, na prática, é típico encontrar que o modelo GARCH (1,1) produz uma descrição adequada de muitas séries temporais financeiras.

Expressões explícitas para curtose e as autocorrelações de ε_t gerada pelas equações (2.68) podem ser geradas, mostrando que um modelo GARCH (r,s) pode descrever séries temporais com seqüências de grandes observações, tanto positivas quanto negativas. A FAC de $\hat{\varepsilon}_t^2$ pode ser encontrada, aplicando as mesmas técnicas dos modelos ARCH (BOLLERSLEV, 1997).

Diniz (1999) utilizou modelos ARIMA na modelagem da série dos retornos diários dos preços mínimo e máximo da Telebrás, no período de 01/08/94 a 13/03/97. Obteve melhores resultados nas previsões um passo-a-frente. Estudou as funções FAC para identificação dos modelos e chegou em ARIMA (1,1,0) para a série dos preços mínimo e ARIMA (2,1,0) para os valores máximo. Mediu a acurácia dos modelos pelo MAPE, que forneceu 2,29% para os valores mínimos e 1,97% para os valores máximos.

Soncin e Corrente (2003) analisaram a série de fechamento da Embraer durante 24 meses, num total de 128 dados. Após ajustes indicados pelas funções de autocorrelação, gráficos dos resíduos e verificações das pressuposições para o ajuste, bem como uma análise de diagnóstico, o modelo adotado e que melhor explicou os dados foi o modelo ARIMA (2,1,1). Cabe ressaltar que o objetivo dos autores era apenas a modelagem da série.

2.2.8 REDES NEURAIAS

Redes neurais artificiais são sistemas de processamento de informações distribuídas, compostas por muitos elementos computacionais simples que interagem através de conexões com pesos distintos. Inspiradas na arquitetura do cérebro humano, elas exibem características como a habilidade de “aprender” padrões complexos de informação e generalizar a informação aprendida (ZHANG; PATUWO; HU, 1998).

Cada elemento computacional não-linear é chamado de nó, e é densamente interconectado através de conexões diretas. Os nós operam em passos discretos, de forma análoga a uma função de dois estágios: o primeiro estágio calcula a soma dos

sinais de entrada, atribuindo pesos aos sinais; o segundo estágio consiste da aplicação de uma função de saída, chamada de função de ativação (HILL; O'CONNOR; REMUS,1996).

Segundo Haykin (2001), as unidades básicas da rede são os neurônios artificiais. Os neurônios se agrupam em camadas. Existem três categorias de camadas: a camada de entrada, a intermediária que pode também conter mais de uma camada e a camada de saída. Os neurônios entre as camadas são conectados por sinapses ou pesos, os quais refletem a relativa importância de cada entrada com o neurônio. A camada de entrada é responsável pelas variáveis de entrada do modelo; a camada de saída contém um ou mais nós, representando os resultados finais do processamento e as camadas intermediárias, ou camadas ocultas, que podem existir uma ou mais que irão tornar o processamento mais refinado e não-linear.

A Figura, a seguir, apresenta o modelo de um neurônio, que forma a base para o projeto de redes neurais (artificiais):

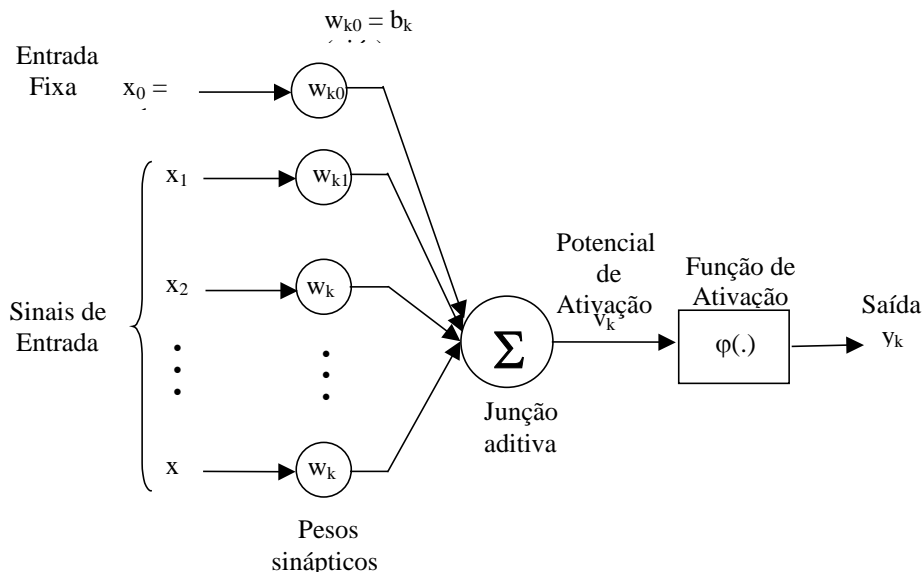


Figura 2: Modelo de um neurônio artificial
Fonte: HAYKIN, 2001.

No modelo neural da Figura 2, temos:

- Um conjunto de *sinapses* ou *elos de conexão*, sendo que cada sinapse é caracterizada por um peso ou força própria. Especificamente, um sinal

x_j , na entrada da sinapse j conectada ao neurônio k , é multiplicado pelo peso sináptico w_{kj} . É importante notar como são escritos os índices do peso sináptico w_{kj} . O primeiro índice se refere ao neurônio em questão e o segundo ao terminal de entrada da sinapse à qual o peso se articula.

- Um *somador ou integrador* para somar os sinais de entrada, ponderados pelas respectivas sinapses do neurônio; as operações descritas, neste caso, constituem um combinador linear.
- Uma *função de ativação* para restringir a amplitude da saída de um neurônio. A função de ativação é também referida como *função restritiva* já que restringe (limita) o intervalo permissível de amplitude do sinal de saída a um valor finito.

Tipicamente, o intervalo normalizado da amplitude da saída de um neurônio é escrito como intervalo unitário fechado $[0,1]$ ou alternativamente $[-1,1]$.

O modelo neural da Figura 2 inclui também um viés aplicado externamente, representado por b_k . O viés b_k tem o efeito de aumentar ou diminuir a entrada líquida da função de ativação, dependendo se ele é positivo ou negativo, respectivamente.

Em termos matemáticos, pode-se descrever um neurônio k , a partir das seguintes equações:

$$u_k = \sum_{j=1}^m w_{kj} x_j \quad (2.86)$$

$$y_k = \varphi(u_k + b_k) \quad (2.87)$$

sendo que x_1, x_2, \dots, x_m são os sinais de entrada; $w_{k1}, w_{k2}, \dots, w_{km}$ são os pesos sinápticos do neurônio k ; u_k é a *saída do combinador linear* devido aos sinais de entrada; b_k é o viés; $\varphi(\bullet)$ é a função de ativação e y_k é o sinal de saída do neurônio. O uso do viés b_k tem o efeito de aplicar um *transformação afim* à saída u_k do combinador linear no modelo da Figura 2, como mostrado por:

$$v_k = u_k + b_k \quad (2.88)$$

Em particular, dependendo se o viés b_k é positivo ou negativo, a relação entre o *campo local induzido* ou *potencial de ativação* v_k do neurônio k e a saída do combinador linear u_k pode ser alterada.

O modelo linear tem a propriedade útil de ter uma solução de forma fechada para resolver o problema de uma regressão, ou seja, de minimizar a diferença quadrática entre o_i e o valor previsto \hat{p}_i . Para a previsão de um período curto, o modelo linear é um ponto de partida razoável, ou um padrão reconhecido, visto que, em muitos mercados, observam-se apenas pequenas mudanças simétricas na variável a ser predita, ao longo de uma tendência de longo prazo. No entanto, isto pode não ser preciso o bastante para mercados financeiros voláteis em que pode haver um processo não-linear dos dados. Movimentos lentos no aumento dos preços dos ativos, seguidos por colapsos repentinos conhecidos como “bolhas” são comuns de acontecer. Dessa forma, o modelo linear pode falhar em capturar ou prever pontos de mudança acentuada nos dados. Por esta razão, é que são utilizadas técnicas não-lineares de previsão.

A rede neural é uma alternativa aos modelos lineares e a algumas abordagens não-paramétricas para a aproximação de sistemas não-lineares. A razão para o uso de uma rede neural é simples e direta. O objetivo é encontrar uma abordagem ou método o qual realize bem previsões para dados gerados para processos que, freqüentemente, são desconhecidos e altamente não-lineares, com uma quantidade pequena de parâmetros e o qual seja mais fácil de estimar que os modelos não-lineares paramétricos (DE OLIVEIRA, 2003).

2.2.8.1 ALGORITMO DE RETROPROPAGAÇÃO (BACKPROPAGATION)

Suponha que se tenha uma rede com N neurônios na camada de entrada ($n = 1, 2, \dots, N$), J neurônios na camada intermediária ($j = 1, 2, \dots, J$) e K na camada de saída ($k = 1, 2, \dots, K$). Considerando $W1_{jn}$ como sendo o peso entre o j -ésimo neurônio na camada intermediária e o n -ésimo neurônio da camada de entrada, e ainda, $W2_{kj}$ o peso

entre o k -ésimo neurônio na camada de saída e o j -ésimo neurônio da camada intermediária, conforme a figura a seguir:

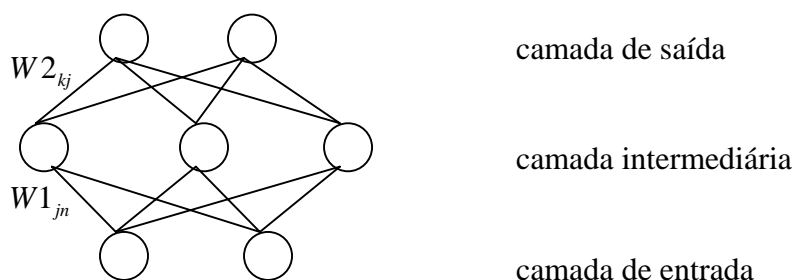


Figura 3: Rede neural com 3 camadas

Fonte: TAK, 1995.

Usando I para indicar a camada de entrada (*input*) e O para saída (*output*), e ainda o sobrescrito i para camada de entrada, h (*hidden*) para intermediária e o para camada de saída, onde se poderá indicar I^h , denotando a entrada do neurônio da camada intermediária, O^i indicando a saída do neurônio da camada de entrada.

A camada de entrada mostra que estes nós é que recebem os valores dos dados de entrada. As linhas de interconexão indicam que o valor de saída de um neurônio que é passado ao longo desta linha até o próximo neurônio. Quando todos os dados da camada de entrada tiverem passado através da última camada, conhecida como camada de saída, um ciclo ou época terá sido realizado. Cada linha de interconexão tem um valor, chamado de peso, os pesos são parâmetros que operam sobre os dados associados a cada linha (ou seja, é realizada uma multiplicação do valor do dado pelo peso). Estes pesos são incrementalmente ajustados durante a fase de treinamento de tal forma a alcançar o resultado de saída desejado para um determinado dado de entrada. Inicializações típicas para estes pesos encontram-se no intervalo de $[-1;1]$ ou menos.

A segunda camada e todas as camadas subseqüentes contêm nós de processamento, conhecidos como neurônios artificiais e são chamadas de intermediárias (ou ocultas). O processamento realizado em um neurônio artificial pode ser dividido em três passos:

1- os dados que passam, ao longo das linhas de entrada para os neurônios, são multiplicados pelos de pesos;

- 2- todos esses dados que foram multiplicados pelos pesos são somados dentro do neurônio;
- 3- o valor total dessa soma é passado através de uma função de transferência, cuja saída representa o valor de saída do neurônio.

A função de transferência mais simples é a função linear (adequada para um espaço linearmente separável), a qual iguala a saída à entrada. A utilização de funções de transferência não-lineares (adequadas para um espaço não-linearmente separável) é uma das características principais da rede neural.

A seguir, são apresentadas todas as etapas do algoritmo de retropropagação (AZOFF, 1994).

1. Definir a função de pesos da rede, $E(w)$, conhecida como a energia ou função de custo, a qual se deseja minimizar. Considerando a metade da soma quadrática dos erros (SSE – *sum of square errors*), tem-se que:

$$E_{total} = \sum_{\mu} \sum_o E(w) \quad (2.89)$$

$$E(w) = \frac{1}{2} [T - O]^2 \quad (2.90)$$

$$\frac{\partial E}{\partial O} = -(T - O) \quad (2.91)$$

sendo que a soma em (2.89) é realizada sobre os padrões μ e os neurônios de saída o , T é o valor alvo e O é a saída observada da rede, uma função dos pesos da rede w . A expressão (2.89) significa que se deve minimizar o erro de todos os processadores da camada de saída para todos os padrões que são apresentados à rede.

2. Calcular a saída do neurônio da camada intermediária O^h , multiplicando os pesos $W1$ pelas saídas do neurônio da camada de entrada O^i

$$I^h = \sum_{n=1}^N W1_n O^i$$

3. Esta entrada I^h é o nível de ativação do neurônio que determina se o neurônio pode produzir uma saída. A relação entre o nível de ativação e a saída pode ser linear ou não-linear (HAYKIN, 2001).

Considerando a função logística $\varphi(x) = \frac{1}{1+e^{-x}}$ que é facilmente derivável:

$$\frac{d\varphi}{dx} = \varphi'(x) = \frac{e^{-x}}{(1+e^{-x})^2} = \varphi(x)(1-\varphi(x)) \quad (2.92)$$

A saída do neurônio da camada intermediária O^h é:

$$O^h = \varphi(I^h) = \frac{1}{1+e^{-I^h}}$$

4. Calculando a saída do neurônio da camada de saída O^o , multiplicando os pesos $W2$ pela saída do neurônio da camada intermediária O^h :

$$I^o = \sum_{j=1}^J W2_j \cdot O^h = \sum_{j=1}^J W2_j \cdot \left(\frac{1}{1+e^{-I^h}} \right) \quad (2.93)$$

e

$$O^o = \varphi(I^o) = \frac{1}{1+e^{-I^o}} \quad (2.94)$$

5. Calculando o erro:

$$E = \sum_{i=1}^n (T - O^o) = E = \sum_{i=1}^n \left(T - \frac{1}{1+e^{-I^o}} \right) \quad (2.95)$$

As derivadas são importantes porque se conhecerem as derivadas parciais do erro com respectivos pesos, conhecer-se-á o caminho que os pesos deveriam percorrer para reduzir o erro.

$$\Delta W2 = \alpha \frac{\partial E}{\partial W2} \quad (2.96)$$

$$\begin{aligned} \frac{\partial E}{\partial W2} &= \frac{\partial E}{\partial I^o} \cdot \frac{\partial I^o}{\partial W2} \\ \frac{\partial E}{\partial I^o} &= \frac{\partial E}{\partial O^o} \cdot \frac{\partial O^o}{\partial I^o} = (T - O^o) \varphi'(I^o) \end{aligned} \quad (2.97)$$

$$\text{Usando (2.92): } \frac{\partial E}{\partial I^o} = (T - O^o) O^o (1 - O^o) = \delta^o, \text{ e ainda } \frac{\partial I^o}{\partial W2} = O^h \quad (2.98)$$

$$\text{E assim: } \Delta W2 = \alpha \frac{\partial E}{\partial W2} = \alpha \delta^o O^h \quad (2.99)$$

6. Calculando o erro da camada intermediária:

$$\Delta W1 = \alpha \frac{\partial E}{\partial W1} \quad (2.100)$$

$$e \quad \frac{\partial E}{\partial W1} = \frac{\partial E}{\partial I^h} \cdot \frac{\partial I^h}{\partial W1} \quad (2.101)$$

$$\frac{\partial E}{\partial I^h} = \frac{\partial E}{\partial O^o} \cdot \frac{\partial O^o}{\partial W1} = O^o(1-O^o) \sum \delta^o W2 = \delta^h$$

$$e \text{ ainda } \frac{\partial I^h}{\partial W1} = O^i \quad (2.102)$$

$$E \text{ assim: } \Delta W1 = \alpha \frac{\partial E}{\partial W1} = \alpha \delta^h O^i \quad (2.103)$$

7. Ajustando $W2$, isto é, os pesos entre a camada intermediária e a camada de saída:

$$W2(t+1) = W2(t) + \alpha \delta^o O^h + \Theta(W2(t) - W2(t-1)) \quad (2.104)$$

onde α é a taxa de aprendizagem e $\Theta(W2(t) - W2(t-1))$ o termo de momento onde, Θ é o coeficiente de momento que recebe um valor no intervalo $[0,1]$, usado para dar à atualização dos pesos uma memória de sua última atualização, fornecendo uma suavização nas forças que afetam as mudanças nos pesos.

8. Ajustando $W1$:

$$W1(t+1) = W1(t) + \alpha \delta^h O^i + \Theta(W1(t) - W1(t-1)) \quad (2.105)$$

Se a lista de épocas foi exaurida, deve-se ir até o próximo passo, caso contrário, apresentar o próximo conjunto de treinamento à camada de entrada e voltar ao passo inicial.

Tendo completado uma época, e acumulado os erros de todos os padrões de entrada, os erros obtidos são avaliados e comparados com os níveis de tolerância.

O que dá à rede neural seu poder de previsão é o processamento paralelo. Adicionalmente ao processamento sequencial dos sistemas lineares típicos, no qual apenas entradas observadas são utilizadas para prever uma saída observada, pela ponderação dos neurônios de entrada, os neurônios na camada oculta processam as entradas de forma paralela, de tal forma a melhorar as previsões. Esta rede de camada simples alimentada adiante, ou multiperceptron, com uma camada escondida é a rede neural mais básica e comumente utilizada em aplicações econômicas e financeiras. Mais genericamente, a rede representa o modo como o cérebro humano processa os dados

sensoriais de entrada, recebidos como neurônios de entrada, e internamente reconhecidos como um neurônio de saída. Com o desenvolvimento do cérebro, mais e mais neurônios são interconectados e mais sinapses, e os sinais de diferentes neurônios, trabalham de forma paralela, em mais e mais camadas escondidas, são combinados pelas sinapses para produzir *insights* e reações.

2.2.8.2 REDES NEURAIS RECORRENTES

Redes neurais recorrentes são redes neurais com um ou mais laços de realimentação, podendo ser local ou global (HAYKIN, 2001).

Talvez, os algoritmos mais amplamente utilizados sejam o algoritmo de aprendizagem recorrente em tempo real (RTRL – *real time recurrent learning*) e o algoritmo de retropropagação através do tempo (BPTT – *backpropagation through time*) (WILLIAM e ZIPSER, 1995).

Uma rede neural pode ser classificada, de acordo com a arquitetura da rede. A forma mais simples de rede é a rede alimentada adiante – *feedforward*. Essas redes constituem-se de múltiplas camadas, onde a entrada de cada camada é a saída da camada predecessora e a interconexão dos neurônios é acíclica.

As redes recorrentes tendem a ser mais difíceis de treinar do que as redes alimentadas adiante, principalmente em função dos ciclos. As redes recorrentes mais utilizadas são a rede de Jordan, a rede de Elman e a rede de Elman estendida, conforme podem ser vistas nas figuras a seguir (DE OLIVEIRA, 2003).

A diferença entre a rede de Elman e a rede de Jordan é que, na rede de Jordan, a realimentação ocorre apenas dos nós de saída para os nós de entrada, enquanto que na rede de Elman a realimentação sai dos nós das camadas escondidas. Existe também a rede de Elman estendida, na qual a realimentação sai dos nós da camada escondida e também dos nós (ou neurônios) de saída.

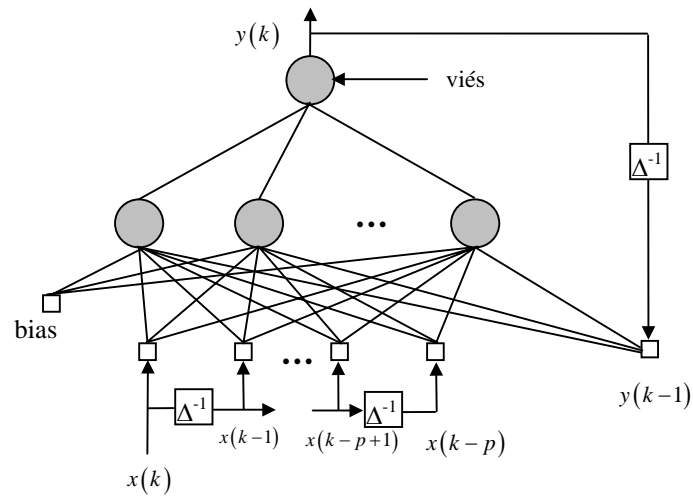


Figura 4: Rede de Jordan

Fonte: DE OLIVEIRA, 2003.

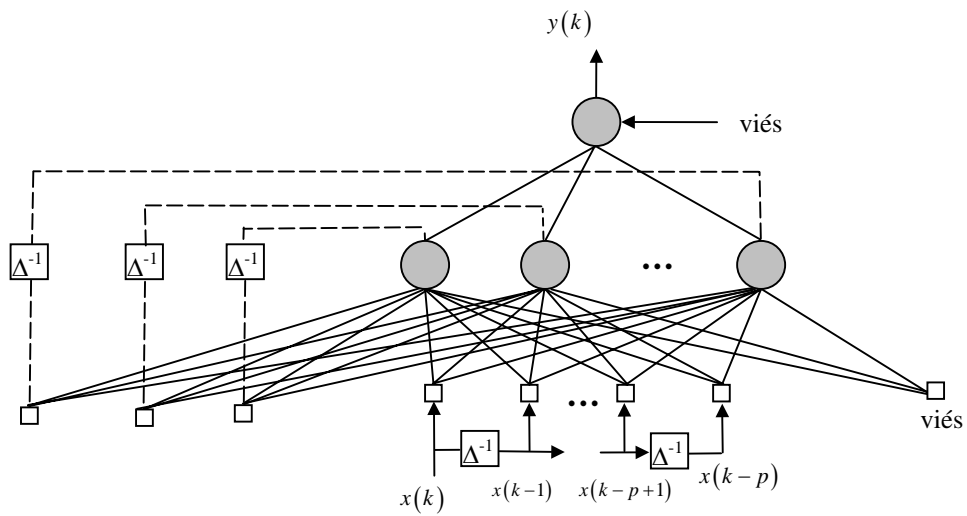


Figura 5: Rede de Elman

Fonte: DE OLIVEIRA, 2003

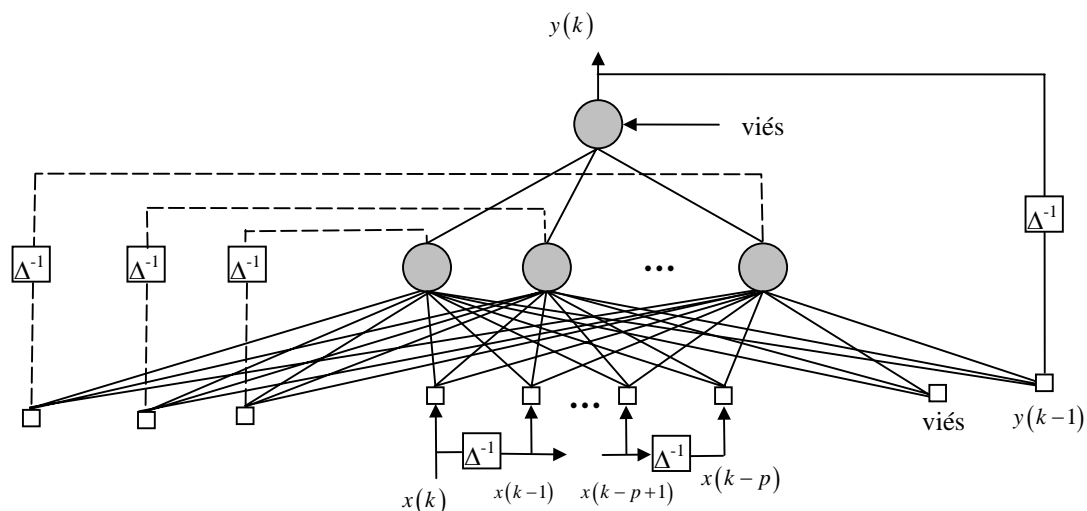


Figura 6: Rede de Elman estendida

Fonte: DE OLIVEIRA, 2003

A rede de Elman pode ser treinada com qualquer algoritmo de aprendizagem para MLPs, tal como o retropropagação. De forma semelhante à rede de Jordan, a rede de Elman pertence à classe das redes recorrentes simples (SRN – *simple recurrent networks*). Embora ela contenha conexões de realimentação ela não pode ser vista como um sistema dinâmico nas quais as ativações são espalhadas indefinidamente. Em vez disso, as ativações para cada camada são computadas, apenas uma vez, a cada passo de tempo (cada apresentação de um vetor de seqüências).

Freeman (1994) as redes de mapeamento, tais como redes com retropropagação, e memórias associativas, geralmente trabalham com padrões estáticos ou espaciais. Elman (1990) discute o uso das redes recorrentes simples, apresentada na Figura 5, para descobrir fronteiras entre palavras em uma sucessão de fonemas. Haykin (2001) comparou quatro arquiteturas principais de redes pertencentes à classe das redes recorrentes com realimentação global: as redes auto-regressivas não-lineares com entradas exógenas (NARX), usando realimentação da camada de saída para a camada de entrada; as redes recorrentes totalmente conectadas com realimentação da camada oculta para a camada de entrada; o perceptron de múltiplas camadas recorrentes com mais de uma camada oculta, usando realimentação da saída de cada camada

computacional para a sua própria entrada e as redes recorrentes de segunda ordem, usando neurônios de segunda ordem.

Williams e Zipser (1995) afirmam que, em se tratando de previsão de sucessões cronológicas, as redes recorrentes não sofrem nem de escassez nem de excesso de ajuste, sendo que as redes recorrentes de tempo real são as mais adequadas para dados dinâmicos, por aprender a dinâmica subjacente de um ambiente não-estacionário, armazenando o conhecimento obtido da amostra de treinamento em um conjunto de pesos.

Enquadrando as redes neurais no contexto dos métodos estatísticos tradicionais, elas podem ser consideradas como uma técnica de estimação não paramétrica, não-linear, multivariada e múltipla, ou seja, é uma função dos dados sem um modelo de descrição específico. Ser multivariada indica que a entrada da rede neural contém duas ou mais variáveis distintas, cujas interdependências e causalidades são exploradas para prever o comportamento futuro de uma ou mais variáveis intertemporais. Não paramétrica e sem modelo específico é uma consequência de algumas suposições referentes às relações entre as variáveis de entrada e as extrapolações dentro de um horizonte de tempo futuro (DE OLIVEIRA, 2003).

Nas abordagem de redes neurais, ao invés de explicitamente deduzir equações que descrevam o comportamento subjacente do sistema em estudo, um capturador automático de um modelo não-linear, usando um algoritmo de aprendizagem neural é construído para aproximar as equações ideais. Muitos algoritmos neurais para processamento temporal têm sido propostos, entre estes se destacam o algoritmo de retropropagação (BPTT, *backpropagation through time*), aprendizado recorrente, em tempo real (RTRL, *real-time recurrent learning*) e os algoritmos de redes recorrentes, utilizando o filtro de Kalman. As redes neurais são atrativas para modelar sistemas não-lineares porque elas são inerentemente não-lineares, devido às funções de ativação não-lineares, e também são bem estabelecidas como funções de aproximação (DE OLIVEIRA; SIQUEIRA, 2003).

Segundo Hill, O'Connor e Remus (1996), a modelagem através de redes neurais artificiais é uma das técnicas que é defendida como alternativa às já tradicionais

técnicas de previsões estatísticas. Os autores frisam, ainda, que um aspecto significativo das redes neurais é a sua capacidade de construir relações não-lineares entre os dados de entrada e as saídas desejadas. Os modelos lineares (por exemplo, $y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots$ sendo que y_t é uma série e os α_i são os parâmetros) apenas descrevem a relação mais simples que pode ocorrer em uma série temporal.

Kutsurelis (1998) trabalhou com o desenvolvimento de redes neurais para previsão no mercado de capitais e chegou a um nível de eficiência, em sua rede, de 88,07% na previsão do S&P 500 com redes MLP e concluíram que as redes neurais são eficientes previsores para séries financeiras.

De Oliveira e Siqueira (2003) justificam o uso de redes neurais para previsão de séries financeiras, principalmente por dois motivos: primeiro, a utilização de camadas intermediárias de neurônios que capturam todas as relações indiretas entre variáveis explanatórias e variáveis dependentes; segundo, a aplicação de uma função de ativação que possibilita a aproximação de qualquer função não-linear.

Essa afirmação também é encontrada em Vojinovic, Kecman e Seidel (2001) que frisa, ainda, que com o uso de redes neurais, um conhecimento, a priori, da relação entre as variáveis não é requerido já que elas são auto adaptativas, ou seja, elas são capazes de gerar modelos não-lineares, sem um conhecimento prévio sobre os relacionamento entre as variáveis de entrada e saída, podendo ser, portanto, mais genéricas e flexíveis que metodologias estatísticas tradicionais.

Em particular, as redes neurais artificiais oferecem potencial significativo para a abordagem do problema de previsão de séries financeiras, como sugerem pesquisas em diferentes áreas do mercado financeiro, dentre outras:

- Mercado de câmbio: utilizando marco e dólar (KLIMASAUSKAS, 1992); dólar com franco suíço (LEQUARRÉ, 1994);
- Cotações de ações: Ações da *Apple Computer* (DEBOECK, 1992), da *Kyokuyo Company*, uma empresa japonesa de porte médio

(HUTCHINSON, 1994); Hochtief, empresa alemã (SCHUMANN; LOHRBACH, 1996), Telebrás (DINIZ et al., 1998);

- Indicadores macroeconômicos: índice de desemprego americano (DRAISMA, KAASHOEK; DIJK, 1993); (SILVA; FERREIRA; SÁFADI, 2000);
- Índices de Bolsas de Valores: ATX – Índice da Bolsa de Valores da Austrália (HAEFKE e HELMENSTEIN, 1996), DAX- Índice da Bolsa de Valores da Alemanha (SIEGLER; STEURER, 1998), S&P500 (MOODY; SAFFELL, 1998) e KUTSURELIS(1998).

Yang (1999) concorda com a visão anterior, afirmando ter conseguido resultados promissores na previsão de sucessões cronológicas financeiras, com o uso de redes neurais probabilísticas, que são conceitualmente construídas com base na classificação dos métodos bayesianos.

Testando a eficácia de técnicas de previsão probabilísticas da direção/sinal do retorno de índices de bolsa de valores (S&P 500, FTSE 100 e NIKEI 225), Leung, Daouk e Chen (2000) compararam dois tipos de modelos: de classificação e de modelos de estimação em nível. Os modelos de classificação testados foram de análise discriminante linear, *logit*, *probit* e redes neurais probabilísticas (PNN) e os modelos de estimação em nível foram o de alisamento exponencial, função de transferência multivariada, vetor autoregressivo com filtros de Kalman e redes neurais com alimentação adiante (*feedforward*). Os resultados sugeriram uma superioridade na performance das previsões feitas pelos modelos de classificação. De acordo com os autores, os melhores desempenhos, nos testes de verificação, ocorreram a partir dos modelos de redes neurais com média de 64%.

Diniz (1999) também usou as redes neurais, para efeitos de previsão um passo à frente das ações dos preços mínimo e máximo. Testou diversas combinações de atrasos das camadas intermediária e de saída, e a melhor rede MLP selecionada possuía dois neurônios na camada de entrada e apenas um neurônio, tanto na camada de entrada quanto na de saída. Os erros medidos pelo MAPE indicaram para os a série dos valores mínimos 1,09% e para os valores máximo 1,83%.

Yim (2001) encontrou resultados que mostram os modelos ARIMA melhores no curto prazo (um passo à frente) do que os modelos de redes neurais, chegando à conclusão de que, embora as redes neurais tenham acompanhado melhor a direção dos valores reais observados nos retornos do IBOVESPA, para todas as frequências, a hipótese de que os modelos de redes neurais é superior aos modelos econométricos tradicionais, não pode ser verificada, pois os resultados das estatísticas de previsão medidos pelo MAPE ficaram muito próximos aos obtidos nos modelos de séries temporais.

Moshiri e Cameron (2000) compararam o uso das redes neurais MLP com algoritmo de retropropagação, na comparação da previsão da inflação canadense com os modelos ARIMA e estruturais. Os dados utilizados foram mensais de janeiro de 1970 a dezembro de 1994 e divididos em duas partes: inicialmente de janeiro de 1970 a dezembro de 1990, para estimação e treinamento, e o restante para previsão. O período de estimação começava com baixas taxas de inflação, em torno de 5% e terminava com altas taxas de inflação. O período de previsão coincide com períodos de mais baixas taxas de inflação. De maneira geral, a diferença no trabalho destes autores é que utilizaram a estratégia de previsão recursiva, isto é, realizaram a previsão para um período (um mês), re-estimavam o modelo, e faziam novas previsões para mais um período à frente. O processo foi repetido para três períodos (um trimestre) de previsão e para doze (um ano) períodos de previsão. O procedimento para medir a acurácia dos modelos foi o MAE e os modelos gerados no Eviews® e Matlab®. Os resultados mostraram que as redes neurais foram superiores aos modelos ARIMA, nos períodos de longo prazo (1 ano). No caso das previsões de um e três passos à frente, os modelos ARIMA levaram uma pequena vantagem sobre as redes neurais. Os piores resultados foram todos para os modelos estruturais.

De Oliveira (2003) afirma que um dos algoritmos mais amplamente utilizados é o de aprendizagem recorrente em tempo (*real time recurrent learning-RTRL*), conforme já citado anteriormente. Esse algoritmo é baseado na computação do gradiente de uma medida de erro de saída em relação aos pesos da rede, seguindo o trabalho de Williams e Zipser (1995). Para que os dados fossem inseridos na rede, utilizou o processo de normalização proposto por Azoff (1994). As previsões de um passo à frente da série do IBOVESPA obtiveram pela rede com o uso do algoritmo RTRL um resultado

satisfatório, no que concerne à aproximação dos valores futuros dada uma sequência histórica de dados.

O autor também explica que as previsões foram realizadas através de modelos ARIMA-GARCH e redes neurais artificiais (RNAs) que utilizam o algoritmo de aprendizagem recorrente em tempo real (ARTR). As metodologias estudadas foram aplicadas separadamente para cada série. Assim se pode avaliar que a sequência empregada no modelamento ARIMA-GARCH dispõe, por construção, de rigor estatístico, em cada etapa do processo, para realizar a previsão, no caso das redes neurais, do trabalho principal constitui-se basicamente na implementação de um algoritmo de otimização. Uma vez que o algoritmo da rede neural foi transformado em um programa de computador, os dados de entrada são inseridos e tem-se um valor de resposta, ou seja, uma saída. Em muitos casos, uma determinada rede neural implementada pode ser utilizada para a tarefa de previsão de séries temporais com comportamentos distintos, por exemplo, tanto para séries de alta, quanto de baixa frequência. Ou seja, não existem critérios bem definidos para escolha do tipo de rede neural que deve ser utilizada para previsão de séries temporais que apresentam comportamentos específicos.

Uma diferença significativa que deve ser destacada entre as RNAs recorrentes e os modelos ARIMA-GARCH é a capacidade das RNA de capturar padrões não-lineares subjacentes ao comportamento da série, enquanto que isso não acontece na modelagem ARIMA-GARCH.

No caso da série de retornos do IBOVESPA, pode-se observar que o modelo da RNA gerou uma curva de valores de previsão que acompanhou, de forma mais próxima, os valores desejados. As previsões obtidas pela RNA, para a série de retornos do IBOVESPA, apresentaram estatística MAPE menor que o modelo ARIMA-GARCH e o TIC da RNA também indica um ajuste mais perfeito que o obtido pelo modelamento que usou as técnicas de volatilidade.

Especificamente, em relação aos trabalhos que compararam as previsões com modelos ARIMA e de redes neurais, destaca-se que, mesmo com metodologias diferentes de previsões, as redes neurais se mostraram potencialmente superiores aos modelos

ARIMA, por apresentarem maior versatilidade e não requererem especificações de modelos formais.

De acordo com Viera e Thomé (2000), a previsão de sucessões cronológicas econômico-financeiras é considerada uma tarefa de grande dificuldade por vários motivos: pequena quantidade de observações da série; presença de intenso ruído; caráter não estacionário e da típica não-linearidade inerente a essas séries.

Um resultado coerente foi encontrado em De Oliveira e Siqueira (2003), ao explicarem que séries temporais financeiras como preços de ações, crescimento econômico, taxas de câmbio e curva de Phillips são não-lineares ou caóticas.

Para Correia e Pereira (1997), os modelos não-lineares podem “explicar” melhor as séries financeiras, especialmente em períodos de turbulência, e eventualmente, apresentar previsões melhores do que os modelos lineares, que são incapazes de captar mudanças bruscas e outros efeitos não-lineares.

No mercado brasileiro, o estudo realizado por Throstensen (1976) sobre a eficiência fraca do mercado de capitais, com dados obtidos do índice BOVESPA, no período de 1971 a 1975, demonstrou os seguintes pontos: nos testes de correlação serial para diferentes tipos de períodos, os resultados mostraram pouca evidência de dependência para as variações do índice; já, nos testes de sequência efetuados, apesar de, nos anos de 1971 e 1972, terem demonstrado evidências de padrões de dependência, anos considerados atípicos, estes não foram suficientes para se obter ganhos significativos. Nos demais períodos de análise, o estudo demonstrou ausência de padrões de dependência, evidenciando que o Índice BOVESPA se adapta ao modelo de passeio aleatório.

Correia e Pereira (1997) realizaram uma discussão sobre a previsibilidade e não-linearidade do IBOVESPA. No teste de normalidade, os autores encontraram assimetria negativa e coeficiente de curtose ligeiramente inferior a três unidades nos retornos diários do IBOVESPA, no período de 17/04/1991 a 29/11/1993, não rejeitando a hipótese de distribuição normal. O teste de McLeod e Li (1983) apresenta indícios de não-linearidade nos retornos diários, e o teste de Hisieh (1989) rejeita a hipótese de

linearidade na média. Adicionalmente, os autores verificam um ganho de previsibilidade na consideração do efeito do dia da semana, em função da tendência dos retornos a serem ligeiramente menores na segunda feira, tal qual observado nos estudos norte-americanos. Finalmente, os autores verificaram que modelos não-lineares, como GARCH (1,1), geraram melhores previsões de retornos em relação a modelos lineares.

Outros autores como Kutsurelis (1998), Leung, Daouk e Chen (2000) utilizaram outras técnicas de modelagem de séries temporais, tomando como base os modelos ARIMA, sugeridos por Box e Jenkins (1976) e com uso de redes neurais voltados para o índice S&P 500, indicando boa performance para as redes neurais e modelos ARIMA em detrimento de outras técnicas tradicionais como alisamento exponencial.

A análise de séries temporais, para efeitos de previsão, é o enfoque deste trabalho, por isso, é dada maior ênfase a estas técnicas. Uma visão destes modelos é apresentada, a seguir, com a discussão dos trabalhos encontrados na literatura sobre previsão.

2.3 DA ANÁLISE DE FOURIER À ANÁLISE DE ONDALETAS

2.3.1 ANÁLISE DE FOURIER

A análise de séries temporais já tem, à sua disposição, um impressionante arsenal de ferramentas (WINKLHOFER et al., 1996). Talvez a mais conhecida de todas seja a análise de Fourier, ou análise harmônica, a qual quebra a série em seus componentes senoidais, cada uma com dada frequência (combinação linear de harmônicos). Uma outra maneira de se definir a análise de Fourier é como uma técnica matemática para transformar a série temporal em uma base de frequências. Essa técnica foi criada por Jean Baptiste Joseph Fourier e publicada, em 1822, num livro intitulado *Théorie Analytique de la Chaleur* (OLIVEIRA; FAVERO, 2003).

Chiann (1997) afirma que a análise de Fourier é fundamental em áreas onde o interesse básico é a procura de periodicidades nos dados. Para muitas séries, a análise de Fourier é extremamente usada porque a frequência contém grande importância no estudo e é apropriada para processos estacionários.

Uma importante ferramenta da análise das séries temporais é a decomposição espectral das séries em séries de Fourier (SILVA; FERREIRA; SÁFADI, 2000). Para estes autores, essa análise, utilizando o termo espectro como uma metáfora, é como colocar a série em um prisma com a finalidade de identificar o tamanho das ondas e a importância dos componentes cíclicos.

Fourier acreditava ser possível representar os mais diferentes tipos de funções através da soma das funções seno e cosseno. Dessa forma, pode-se observar que uma função genérica $f(x)$ pode ser representada da seguinte forma:

$$f(x) = A + a_1 \cos(1x) + a_2 \cos(2x) + a_3 \cos(3x) + \dots \\ \dots + b_1 \text{sen}(1x) + b_2 \text{sen}(2x) + b_3 \text{sen}(3x) + \dots \quad (2.106)$$

ou ainda, pode-se representar a equação (2.106) por:

$$f(x) = A + \sum_{k=1}^{\infty} a_k \cos(kx) + b_k \text{sen}(kx) \quad (2.107)$$

Todavia, surge então a seguinte questão: “dada uma função $f(x)$, definida em um certo intervalo, quais seriam os valores dos coeficientes A , a_k e b_k de modo que a equação (2.107) a represente?”

Esses valores foram encontrados por Fourier para uma função definida no intervalo $[-\pi, \pi]$.

Tem-se:

$$\int_{-\pi}^{\pi} f(x) dx = \int_{-\pi}^{\pi} \left(A + \sum_{k=1}^{\infty} a_k \cos(kx) + b_k \text{sen}(kx) \right) dx \quad (2.108)$$

ou ainda:

$$\int_{-\pi}^{\pi} f(x) dx = \int_{-\pi}^{\pi} A dx + \int_{-\pi}^{\pi} \sum_{k=1}^{\infty} a_k \cos(kx) + b_k \text{sen}(kx) dx \quad (2.109)$$

$$\int_{-\pi}^{\pi} f(x) dx = \int_{-\pi}^{\pi} A dx + \sum_{k=1}^{\infty} \int_{-\pi}^{\pi} a_k \cos(kx) + b_k \text{sen}(kx) dx \quad (2.110)$$

Donde:

$$\int_{-\pi}^{\pi} f(x) dx = A(x) \Big|_{-\pi}^{\pi} + \sum_{k=1}^{\infty} \left(a_k \frac{1}{k} \text{sen}(kx) - b_k \frac{1}{k} \cos(kx) \right) \Bigg|_{-\pi}^{\pi} \quad (2.111)$$

$$\int_{-\pi}^{\pi} f(x)dx = A(\pi - (-\pi)) + \sum_{k=1}^{\infty} \left(a_k \frac{1}{k} \text{sen}(k\pi) - b_k \frac{1}{k} \cos(k\pi) - a_k \frac{1}{k} \text{sen}(-k\pi) + b_k \frac{1}{k} \cos(-k\pi) \right) \quad (2.112)$$

Como a função seno é uma função ímpar ($\text{sen}(-x) = -\text{sen}(x)$) e a função cosseno é par ($\cos(-x) = \cos(x)$), portanto:

$$\int_{-\pi}^{\pi} f(x)dx = 2\pi A + \sum_{k=1}^{\infty} \left(a_k \frac{1}{k} \text{sen}(k\pi) - b_k \frac{1}{k} \cos(k\pi) + a_k \frac{1}{k} \text{sen}(k\pi) - b_k \frac{1}{k} \cos(k\pi) \right) \quad (2.113)$$

$$\int_{-\pi}^{\pi} f(x)dx = 2\pi A + 0 \quad (2.114)$$

e assim:

$$A = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)dx \quad (2.115)$$

Os valores de a_k e b_k ficariam:

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(kx)dx \quad (2.116)$$

$$b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \text{sen}(kx)dx \quad (2.117)$$

Cabe ressaltar ainda que: para funções pares tem-se $a_k \neq 0$ e $b_k = 0$ para funções ímpares $a_k = 0$ e $b_k \neq 0$.

Para períodos genéricos, a função de Fourier pode se expressa por:

$$f(x) = A + \sum_{k=1}^{\infty} a_k \cos\left(\frac{k\pi x}{L}\right) + b_k \text{sen}\left(\frac{k\pi x}{L}\right) \quad (2.118)$$

E para esse período genérico, definindo como intervalo $[-L; L]$, tem-se:

$$A = \frac{1}{2L} \int_{-L}^L f(x)dx \quad (2.119)$$

$$a_k = \frac{1}{L} \int_{-L}^L f(x) \cos\left(\frac{k\pi x}{L}\right)dx \quad (2.120)$$

$$b_k = \frac{1}{L} \int_{-L}^L f(x) \text{sen}\left(\frac{k\pi x}{L}\right)dx \quad (2.121)$$

Para medir a frequência e encontrar as frequências contidas em um sinal, é necessário utilizar a Transformada de Fourier (FT). Se a frequência de um sinal no domínio do tempo é dada, a representação da amplitude do sinal é obtida. Ao traçar o gráfico

frequência×amplitude, sendo a frequência indo de zero a infinito, para toda frequência ter-se-á uma amplitude. Em outras palavras, a FT dá a informação de frequência do sinal, a qual significa que se tem o quanto de cada frequência existe no sinal, mas isso não nos diz quando no tempo estas componentes frequências existem, isto é, a FT fornece quais os componentes de frequência (componentes espectrais) que existem no sinal.

Segundo Shiavi (1999), na maioria das vezes, uma informação que não pode ser lida no domínio do tempo pode ser extraída no domínio da frequência. Um exemplo disso é o eletrocardiograma, onde qualquer alteração na forma de onda conhecida pode indicar a presença de alguma patologia clínica. Os cardiologistas usam os sinais do eletrocardiograma no domínio do tempo, os quais são registrados e analisados posteriormente para possíveis constatações de outras patologias.

Todavia, se esta informação é necessária, isto é, se pretende saber qual o componente espectral que ocorre em qual tempo (intervalo), então a transformada de FT não é a transformada correta para ser utilizada. Surge, então, que a transformada de *ondaleta* é capaz de fornecer a informação de tempo de frequência simultaneamente, conduzindo a representação de frequência-tempo do sinal.

Todavia, a análise de Fourier tem uma séria desvantagem, pois quando transforma a série para o domínio da frequência, a informação do tempo é perdida. Ao se observar a transformada de Fourier de uma série é impossível de se saber quando determinado evento ocorreu (GENÇAY; SELÇUK; WHITCHER, 2002).

Segundo Misiti et al. (1997), Dennis Garbor (1946) adaptou a transformada de Fourier para analisar apenas uma pequena seção da série no tempo, que chamou de *Short-time Fourier Transform* (STFT), que conseguia mapear a série com uma função de duas dimensões: o tempo e a frequência. Essa transformada conseguiria captar algumas informações, tanto no tempo, como na frequência, porém com certa limitação, pois a janela que era criada para se fazer a análise era de um tamanho fixo, captando, assim, sempre a mesma frequência, como se observa na figura a seguir:

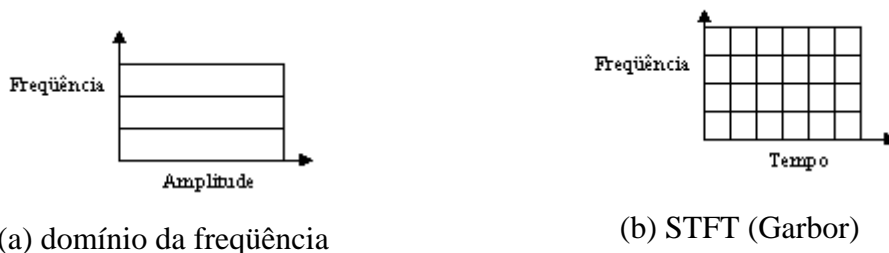


Figura 7: Análise de Fourier no domínio do tempo e a STFT proposta por Garbor

Fonte: Misiti et al., 1997.

Dessa forma, o próximo passo então era obter uma janela de análise com tamanho flexível, e as ondaletas conseguem esse feito, pois, em um mesmo intervalo de tempo, conseguem detectar informações de baixa e de alta frequência, que podem assim ser mais bem ajustadas em cada trecho da série. A figura, a seguir, ajuda a ilustrar o papel da ondaleta:

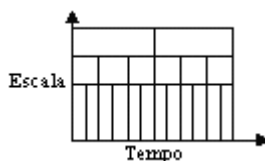


Figura 8: Análise por Ondaletas

Fonte: Misiti et al., 1997.

Para Gençay, Selçuk e Whitcher (2002), apesar da funcionalidade da transformada de Fourier, especialmente no que diz respeito à obtenção da análise espectral de um sinal, existem muitas falhas nesta técnica como:

- a incapacidade da transformada de Fourier em representar funções que têm componentes não periódicas, que estão localizadas no tempo e no espaço, tais como pulsos breves;
- a incapacidade de fornecer qualquer informação sobre a dependência temporal de um sinal, visto que ela faz a média sobre todo o tempo de duração do sinal;
- as funções seno e cosseno usadas não são locais e, portanto, desempenham uma tarefa muito pobre de aproximação de sinais muito localizados;
- a análise de Fourier é altamente instável em relação à perturbações na série; e

- pequenas mudanças em algumas das observações podem provocar mudanças em todas as componentes de uma expansão de Fourier, o que não acontece nas ondaletas.

Diante desses pressupostos, a contrapartida a ser usada seriam as ondaletas, vistas como um mecanismo para quebrar a série em duas partes constituintes, permitindo-nos analisar os dados em diferentes domínios de frequência. Dessa forma, apresenta-se, a seguir, uma discussão sobre a teoria das ondaletas.

2.3.2 ANÁLISE DE SUCESSÕES CRONOLÓGICAS VIA ONDALETAS

Ondeletas, Ondelettes, Ôndulas, Ondaletas podem ser traduzidas respectivamente nas línguas inglês, francês, português de Portugal e português do Brasil. Morettin (1999) sugere, ainda, as formas alternativas para o vocábulo que seriam ondícula ou ondinha. A grosso modo, representa um algoritmo matemático para se criar funções bases para representações de funções em outros modelos (MISITI et al., 1997).

A primeira menção sobre o termo ondaletas foi dada em 1909, por Alfred Haar. O conceito de ondaletas, na forma atual, foi primeiramente proposto por Jean Morlet e sua equipe no *Marseille Theoretical Physics Center*, trabalhando em *Alex Grossmann*, na França (MISITI et al., 1997).

Por falta de embasamento matemático, Morlet inicialmente encontrou muitos opositores. Esses, como lembra Daubechies (1992), teciam críticas do seguinte tipo: “se isso tivesse correto, já estaria nos livros de matemática. Como não está, provavelmente é inútil.” Posteriormente, os métodos de análise de ondaletas foram desenvolvidos, principalmente por Yves Meyer, um matemático que ajudou a enriquecer e amadurecer a nova teoria e que assegurou a disseminação dos métodos em diversos campos da matemática. Logo em seguida, em 1988, Stéphane Mallat, um estudante de processamento de imagens, desenvolveu um algoritmo para calcular a transformada de ondaleta, de forma computacionalmente eficiente, abrindo as portas à comunidade de processamento de sinais (GALVÃO et al., 2001).

A análise de dados, de acordo com escalas variáveis, no domínio do tempo e da frequência, é a idéia básica da utilização da teoria das ondaletas. As ondaletas são funções matemáticas que ampliam intervalos de dados, separando-os em diferentes componentes de frequência, permitindo a análise de cada componente em sua escala correspondente. Essa característica distingue métodos de ondaletas de outros métodos em estatística, baseados em séries ortogonais, tais como as usuais séries de Fourier. (MISITI et al., 1997).

Para Torrence e Compo (1998), a vantagem das ondaletas sobre os métodos tradicionais de Fourier, é que a base das funções de Fourier são dependentes da frequência, mas não do tempo, ou seja, pequenas alterações no domínio da frequência produzem alterações em todo o domínio do tempo. Segundo os autores, as ondaletas são dependentes de ambos os domínios, da frequência (via dilação) e do tempo (via translação), o que é uma vantagem em diversos casos. Por muitas décadas, os cientistas procuraram funções mais apropriadas do que senos e cossenos para análise de sinais. As bases das funções de Fourier são impróprias para o tratamento local de dados, pois são séries infinitas e não se adaptam a análise de dados descontínuos. As ondaletas não só se prestam a aproximação de funções finitas, como também servem para análise de dados descontínuos.

Comparando as ondaletas com as ondas senoidais, essas não têm duração limitada, indo de menos ao mais infinito. Onde as ondas senoidais são lisas e suaves, as ondaletas tendem a ser irregulares e assimétricas como se pode observar a seguir: (MISITI et al., 1997):

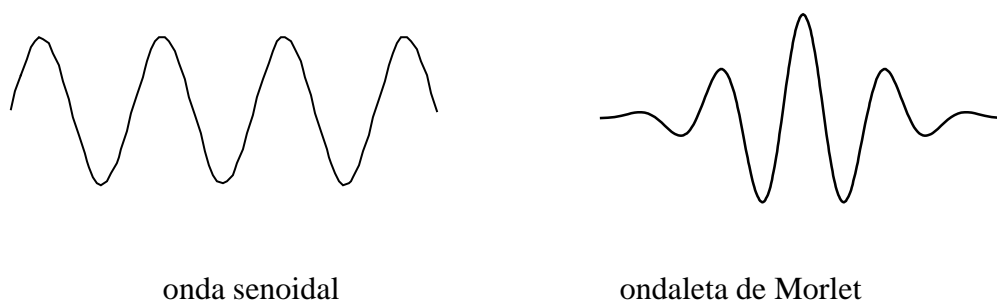


Figura 9: Formas de ondas: senoidal e de uma ondaleta

Para Moura (2002), a análise de Fourier consiste em quebrar uma série em suas componentes de senos e cossenos, em suas várias frequências. Similarmente à análise de ondaleta, considera dilatações (compressões) e translações de uma onda original, chamada de ondaleta mãe.

De uma maneira geral, a filtragem dos dados via ondaletas, em contraste com a filtragem via transformada de Fourier, não desloca nem deforma pontos notáveis (máximos, mínimos e inflexões) do sinal a ser filtrado, apresentando ainda a vantagem quando a série possui baixo fator de preenchimento (*filling factor*), isto é, apresenta picos alternados com regiões de baixa intensidade (GALVÃO et al., 2001).

A problemática em questão consiste em obter uma frequência crítica de maneira que os componentes da série original relacionados a frequências menores ou iguais a uma frequência crítica para que formem uma Subsérie $\{A\}_{t=1}^{\frac{n}{2}}$ relativa a baixas frequências, que também recebem o nome de série “aproximada” (*approximation part*) e os demais componentes formem outra Subsérie relativa a altas frequências, denotada $\{D\}_{t=1}^{\frac{n}{2}}$, também chamada de série “detalhe” (*detail part*) (HOMSY; PORTUGAL; ARAÚJO, 2000).

Para Misiti et al. (1997), este processo de decomposição que, a partir de uma série X_t , passando por um filtro de uma ondaleta, dá origem a duas novas séries, uma chamada aproximação (baixas frequências) e outra detalhe (altas frequências). O autor cita ainda que o processo de decomposição da série pode ser iterativo, com sucessivas decomposições, formando uma árvore de decomposição com 2^n caminhos diferentes para a codificação da série. A árvore de decomposição fica da seguinte forma:

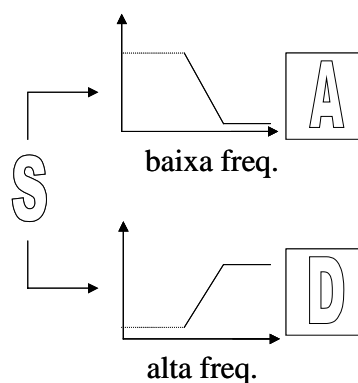


Figura 10: Decomposição via ondaletas

Fonte: MISITI et al., 1997.

O autor cita, ainda, que o processo de decomposição da série pode ser iterativo, com sucessivas decomposições, formando uma árvore de decomposição apresentada na figura a seguir.

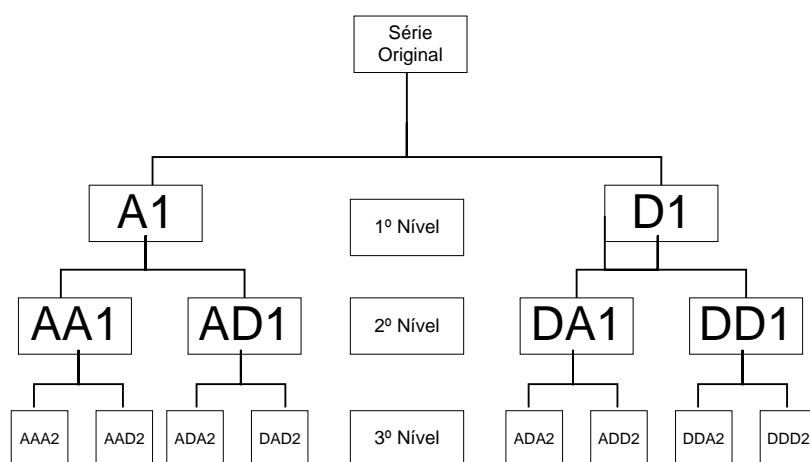


Figura 11: Árvore de decomposição em 3 níveis de uma série temporal via decomposição por ondaleta.

Fonte: MISITI et al., 1997.

Polikar (1994) afirma que, nas ondaletas, as altas frequências são mais bem resolvidas no tempo, e baixas frequências são mais bem resolvidas em frequência. Isto significa que um certo componente de alta frequência pode ser localizado melhor no tempo (com menor erro relativo) que um componente de baixa frequência. Do mesmo modo, um componente de baixa frequência pode ser localizado melhor em frequência se comparado ao componente de alta frequência.

Galvão et al. (2001) frisam que, em cada nível, cada subconjunto de dados passa por uma nova divisão, de maneira a gerar novos coeficientes e assim sucessivamente. Nesses casos, o resultado da transformada de ondaleta é um conjunto de coeficientes indexados por um fator de escala e por um fator de translação.

2.3.3 A MATEMÁTICA DAS ONDALETAS

Uma base de um espaço vetorial V é um conjunto de vetores linearmente independentes, tal que um vetor $v \in V$ pode ser escrito como uma combinação linear desses vetores bases. Pode haver mais do que uma base para um espaço vetorial. Porém, todos têm o mesmo número de vetores, e este número é conhecido como dimensão do espaço vetorial.

O conceito de vetores pode ser generalizado para funções, da seguinte forma:

$$f(t) = \sum_k \mu_k \phi_k(t) \quad (2.122)$$

No caso das FT, as funções exponenciais complexas senos e cossenos são as devidas funções base, sendo funções ortogonais, as quais fornecem algumas propriedades necessárias para reconstrução (POLIKAR, 1994).

Seja $f(t)$ e $g(t)$ duas funções em $L^2[a, b]$ ⁷. O produto interno das duas funções é:

$$\langle f(t), g(t) \rangle = \int_a^b f(t) \cdot g^*(t) dt \quad (2.123)$$

De acordo com a definição acima, a transformada contínua de *ondaleta* (CWT) pode ser interpretada como o produto interno do sinal de teste com as funções base $\Psi(\tau, s)(t)$:

$$CWT_x^\psi(\tau, s) = \Psi_x^\psi(\tau, s) = \int x(t) \cdot \psi_{\tau, s}^*(t) dt \quad (2.124)$$

$$\text{onde } \psi_{\tau, s} = \frac{1}{\sqrt{s}} \psi\left(\frac{t - \tau}{s}\right).$$

Esta definição de CWT mostra que a análise de ondaleta é uma medida de similaridade entre as funções base (ondaleta) e o próprio sinal. Aqui similaridade está no sentido de

⁷ Denota o conjunto de funções quadráticas integráveis no intervalo $[a, b]$.

conteúdo de frequência similar. Os coeficientes CWT calculados referem o quanto próximo do sinal está a ondaleta na escala atual. Se o sinal tem maior componente de frequência correspondente à escala atual, então a ondaleta (funções base) na escala atual serão similares ou próximas ao sinal na posição particular, onde esta componente de frequência ocorre. Dessa forma, o coeficiente CWT calculado naquele ponto no plano escala-tempo será um número relativamente grande (POLIKAR, 1994).

Segundo Morettin (1999), toda função periódica, de período 2π , de quadrado integrável, ou seja, de $L^2(0, 2\pi)$, é gerada por uma superposição de exponenciais complexas $W_n(x) = e^{inx}$, $n = 0, \pm 1, \pm 2, \dots$, obtidas por dilatações da função $W_n(x) = e^{ix}$.

O objetivo, segundo o autor, é estender essa função para $L^2(\mathbb{R})$, isto é, gerar um espaço, a partir de uma função Ψ , que pode ser conseguida por dilatações (parâmetro “ a ”) ou compressões e translações (parâmetro “ b ”) de Ψ , dada por:

$$\Psi_{a,b}(x) = |a|^{-\frac{1}{2}} \Psi\left(\frac{x-b}{a}\right), a, b \in \mathbb{R}, a \neq 0 \quad (2.125)$$

Sendo que Ψ é chamada de ondaleta mãe e os parâmetros a e b tomam os seguintes valores especiais, $a = 2^{-j}$ e $b = k2^{-j}$, que compõem a equação:

$$\Psi_{j,k}(x) = 2^{\frac{j}{2}} \Psi(2^j x - k), j, k \in \mathbb{Z} \quad (2.126)$$

obtida pela dilatação binária 2^{-j} e pela translação diática $k2^{-j}$, conforme sugere Morettin (1999).

Uma maneira de gerar ondaletas é pela função escala, ou chamada de ondaleta pai, Φ , que é uma solução da equação

$$\Phi(t) = \sqrt{2} \sum_k l_k \Phi(2t - k) \quad (2.127)$$

onde $l_k = \sqrt{2} \int_{-\infty}^{\infty} \Phi(t) \Phi(2t - k) dt$

Essa função gera uma família ortonormal de $L^2(\mathbb{R})$,

$$\Phi_{j,k}(x) = 2^{\frac{j}{2}} \Phi(2^j x - k), j, k \in \mathbb{Z} \quad (2.128)$$

Nessas condições, Ψ pode ser obtida de Φ por:

$$\Psi(t) = \sqrt{2} \sum_k h_k \Phi(2t - k) \quad (2.129)$$

onde

$$h_k = (-1)^k l_{1-k} \quad (2.130)$$

que é chamada de *quadrature mirror filter relation* (MORETTIN, 1999).

Segundo Gençay, Selçuk e Whitcher (2002), uma série temporal financeira pode ser decomposta por uma análise de ondaletas, por uma sequência de projeções de ondaletas pai e mãe, a partir das funções Φ e Ψ , como seguem as equações (2.126) e (2.128).

A representação de uma série temporal X_t em $L^2(\mathbb{R})$ pode ser dada por:

$$X_t = \sum_k a_{j,k} \Phi_{j,k}(t) + \sum_k d_{j,k} \Psi_{j,k}(t) + \sum_k d_{j-1,k} \Psi_{j-1,k}(t) + \dots + \sum_k d_{1,k} \Psi_{1,k}(t) \quad (2.131)$$

$$X_t = \sum_k a_{j,k} \Phi_{j,k}(t) + \sum_i \sum_k d_{i,k} \Psi_{j-i+1,k}(t) \quad (2.132)$$

onde j é o número de componentes e k que varia de 1 ao número de coeficientes do componente específico. Os coeficientes $a_{j,k}, d_{j,k}, \dots, d_{1,k}$ são os coeficientes das transformadas de ondaletas dadas pelas projeções

$$a_{j,k} = \int \Phi_{j,k}(t) \cdot X_t dt, \text{ chamada parte de aproximação} \quad (2.133)$$

e

$$d_{j,k} = \int \Psi_{j,k}(t) \cdot X_t dt, \text{ chamada parte de detalhe} \quad (2.134)$$

E no nível j tem-se a metade do número de coeficientes do nível $j+1$, donde o nome piramidal.

Para entender como a função (2.126) atua, considere, por exemplo, a ondaleta mais simples, chamada de ondaleta de Haar, cuja equação é:

$$\Psi(x) = \begin{cases} 1, & \text{se } 0 \leq x < \frac{1}{2} \\ -1, & \text{se } \frac{1}{2} \leq x < 1 \\ 0, & \text{caso contrário} \end{cases} \quad (2.135)$$

Se $j = k = 0$, de (2.126) tem-se:

$$\Psi_{0,0}(x) = 2^{\frac{0}{2}} \Psi(2^0 x - 0) = \Psi(x) \quad (2.136)$$

que é a própria ondaleta primária (2.135).

Se $j = 1, k = 0$ tem-se:

$$\Psi_{1,0}(x) = 2^{\frac{1}{2}} \Psi(2^1 x - 0) = \sqrt{2} \Psi(2x) \quad (2.137)$$

e

$$\Psi(2x) = \begin{cases} 1, \text{ se } 0 \leq 2x < \frac{1}{2} \\ -1, \text{ se } \frac{1}{2} \leq 2x < 1 \\ 0, \text{ caso contrário} \end{cases} = \begin{cases} 1, \text{ se } 0 \leq x < \frac{1}{4} \\ -1, \text{ se } \frac{1}{4} \leq x < \frac{1}{2} \\ 0, \text{ caso contrário} \end{cases} \quad (2.138)$$

e assim:

$$\Psi_{1,0}(x) = \sqrt{2} \Psi(2x) = \begin{cases} \sqrt{2}, \text{ se } 0 \leq x < \frac{1}{4} \\ -\sqrt{2}, \text{ se } \frac{1}{4} \leq x < \frac{1}{2} \\ 0, \text{ caso contrário} \end{cases} \quad (2.139)$$

Se $j = 2, k = 0$ tem-se:

$$\Psi_{2,0}(x) = 2^{\frac{2}{2}} \Psi(2^2 x - 0) = 2 \Psi(4x) \quad (2.140)$$

e

$$\Psi(4x) = \begin{cases} 1, \text{ se } 0 \leq 4x < \frac{1}{2} \\ -1, \text{ se } \frac{1}{2} \leq 4x < 1 \\ 0, \text{ caso contrário} \end{cases} = \begin{cases} 1, \text{ se } 0 \leq x < \frac{1}{8} \\ -1, \text{ se } \frac{1}{8} \leq x < \frac{1}{4} \\ 0, \text{ caso contrário} \end{cases} \quad (2.141)$$

e assim:

$$\Psi_{2,0}(x) = 2 \Psi(4x) = \begin{cases} 2, \text{ se } 0 \leq x < \frac{1}{8} \\ -2, \text{ se } \frac{1}{8} \leq x < \frac{1}{4} \\ 0, \text{ caso contrário} \end{cases} \quad (2.142)$$

As ondaletas $\Psi_{0,0}(x)$, $\Psi_{1,0}(x)$ e $\Psi_{2,0}(x)$ podem ser vistas na figura seguintes:

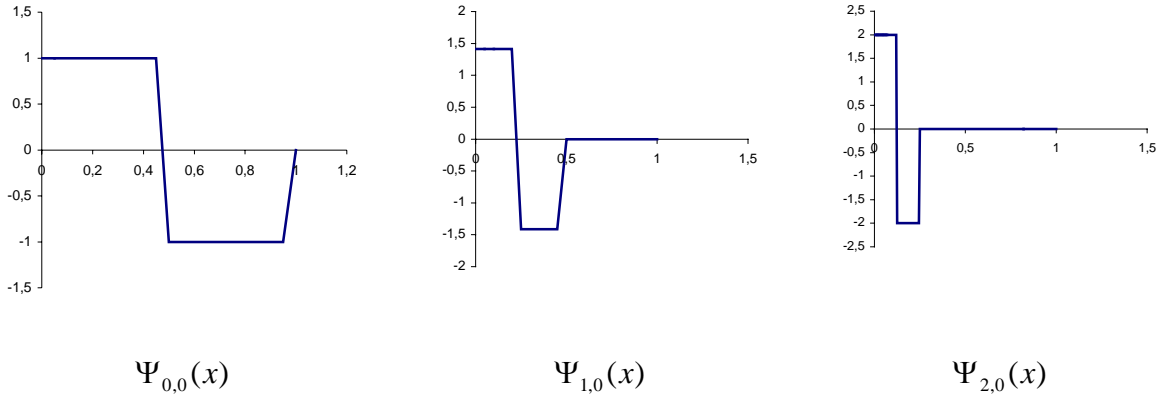


Figura 12: ondaletas $\Psi_{0,0}(x)$, $\Psi_{1,0}(x)$ e $\Psi_{2,0}(x)$

Fazendo agora as translações com $k = 1, 2, 3, \dots$ e considerando $j = 0$, tem-se:

Para $j = 0, k = 1$

$$\Psi_{0,1}(x) = 2^{\frac{0}{2}} \Psi(2^0 x - 1) = \Psi(x - 1) \quad (2.143)$$

e

$$\Psi(x - 1) = \begin{cases} 1, & \text{se } 0 \leq x - 1 < \frac{1}{2} \\ -1, & \text{se } \frac{1}{2} \leq x - 1 < 1 \\ 0, & \text{caso contrário} \end{cases} = \begin{cases} 1, & \text{se } 1 \leq x < \frac{3}{2} \\ -1, & \text{se } \frac{3}{2} \leq x < 2 \\ 0, & \text{caso contrário} \end{cases} \quad (2.144)$$

e assim:

$$\Psi_{0,1}(x) = \Psi(x - 1) = \begin{cases} 1, & \text{se } 1 \leq x < \frac{3}{2} \\ -1, & \text{se } \frac{3}{2} \leq x < 2 \\ 0, & \text{caso contrário} \end{cases} \quad (2.145)$$

E ainda, para $j = 0, k = 3$

$$\Psi_{0,3}(x) = 2^{\frac{0}{2}} \Psi(2^0 x - 3) = \Psi(x - 3) \quad (2.146)$$

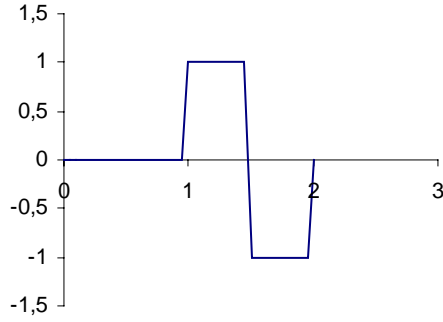
e

$$\Psi(x - 3) = \begin{cases} 1, & \text{se } 0 \leq x - 3 < \frac{1}{2} \\ -1, & \text{se } \frac{1}{2} \leq x - 3 < 1 \\ 0, & \text{caso contrário} \end{cases} = \begin{cases} 1, & \text{se } 3 \leq x < \frac{7}{2} \\ -1, & \text{se } \frac{7}{2} \leq x < 4 \\ 0, & \text{caso contrário} \end{cases} \quad (2.147)$$

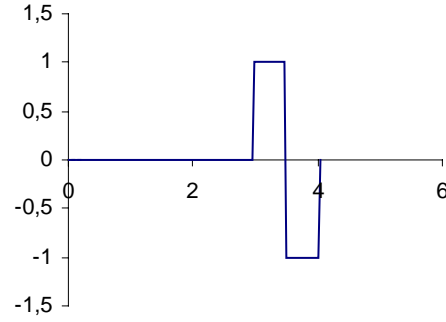
e assim:

$$\Psi_{0,3}(x) = \Psi(x-3) = \begin{cases} 1, & \text{se } 3 \leq x < 7/2 \\ -1, & \text{se } 7/2 \leq x < 4 \\ 0, & \text{caso contrário} \end{cases} \quad (2.148)$$

As ondaletas $\Psi_{0,1}$ e $\Psi_{0,3}$ ficam:



$\Psi_{0,1}(x)$



$\Psi_{0,3}(x)$

Figura 13: Formas de *Ondaletas* $\Psi_{0,1}$ e $\Psi_{0,3}$

Como visto anteriormente, a análise de ondaletas usa as ondaletas como funções bases para representar uma série temporal.

Para uma sucessão cronológica, utilizando-se um modelo ARIMA e as decomposições via ondaletas para prever a Subséries chamada aproximação e denotada $A1_n$ e a Subséries detalhe, denotada $D1_n$, tem-se:

$$A1_n = \sum_{i=0}^p \phi_{Ai} A1_{n-1} + \sum_{i=0}^q \theta_{Ai} \varepsilon_{A1,n-i} \quad (2.149)$$

$$D1_n = \sum_{i=0}^p \phi_{Di} D1_{n-1} + \sum_{i=0}^q \theta_{Di} \varepsilon_{D1,n-i} \quad (2.150)$$

Considerando, agora, a ondaleta chapéu mexicano e combinando com redes neurais com algoritmo de retropropagação, por exemplo, tem-se:

$$\Psi(x) = (1 - x^2) e^{\left(-\frac{x^2}{2}\right)}, x \in \mathbb{R} \quad (2.151)$$

e substituindo a equação (2.151), no neurônio da camada de entrada no desenvolvimento de (2.89), tem-se:

$$I_j^h = W1_{j,1}(1-x^2)e^{\left(\frac{x^2}{2}\right)} + W1_{j,2}\left(1-\left(\frac{x}{3}\right)^2\right)e^{\left(\frac{\frac{x}{3}}{2}\right)^2} \quad (2.152)$$

E ainda:

$$O_j^h = \frac{1}{1 + e^{-W1_{j,1}(1-x^2)e^{\left(\frac{x^2}{2}\right)} - W1_{j,2}\left(1-\left(\frac{x}{3}\right)^2\right)e^{\left(\frac{\frac{x}{3}}{2}\right)^2}}} \quad (2.153)$$

E assim, obtém-se a entrada do neurônio da camada de saída:

$$I^o = \sum_{j=1}^3 W2_j \frac{1}{1 + e^{-W1_{j,1}(1-x^2)e^{\left(\frac{x^2}{2}\right)} - W1_{j,2}\left(1-\left(\frac{x}{3}\right)^2\right)e^{\left(\frac{\frac{x}{3}}{2}\right)^2}}} \quad (2.154)$$

E, conseqüentemente, a saída do neurônio da camada de saída:

$$O^o = \frac{1}{1 + e^{-\sum_{j=1}^3 W2_j \frac{1}{1 + e^{-W1_{j,1}(1-x^2)e^{\left(\frac{x^2}{2}\right)} - W1_{j,2}\left(1-\left(\frac{x}{3}\right)^2\right)e^{\left(\frac{\frac{x}{3}}{2}\right)^2}}}}} \quad (2.155)$$

Assim, fazendo a decomposição como proposta na Figura 11, em 1º nível, tem-se para uma sucessão cronológica X_t :

$$X_t = \frac{1}{1 + e^{-\sum_{j=1}^3 W2_j \frac{1}{1 + e^{-W1_{j,1}S(t-1) + W1_{j,2}S(t-2)}}}} + \varepsilon \quad (2.156)$$

E usando a decomposição de ondaletas, usando duas redes neurais, uma para prever a série de aproximação (A_1) e outra para a parte detalhe (D_1):

$$A_1(t) = \frac{1}{1 + e^{-\sum_{j=1}^3 W2_j \frac{1}{1 + e^{-W1_{j,1}A_1(t-1) + W1_{j,2}A_1(t-2)}}}} + \varepsilon \quad (2.157)$$

e

$$D_1(t) = \frac{1}{1 + e^{-\sum_{j=1}^3 W2_j \frac{1}{1 + e^{-W1_{j,1}D_1(t-1) + W1_{j,2}D_1(t-2)}}}} + \varepsilon \quad (2.158)$$

Torrence e Compo (1998) citam diferentes tipos de ondaletas $\Psi_{a,b}(x)$, que podem ser aplicadas nestas decomposições: Haar, Daubechies (que possui 9 formas distintas), Biorthogonal (com 14 formas distintas), Coifles (com 5 formas distintas), Symlets (com 7 formas distintas), Morlet, Mexican Hat e Meyer. Ambas as formas de ondaletas citadas são usadas tanto na decomposição como na reconstrução da série. Os autores destacam, ainda, que estudos comparativos mostraram que a mudança na escolha da função ondaleta gera apenas pequenos efeitos nos resultados.

2.3.4 APLICAÇÕES DE PREVISÕES BASEADO NA DECOMPOSIÇÃO DE ONDALETAS

Tak (1995) desenvolveu, a partir dos dados mensais do índice *Standard and Poor's 500* (S&P 500), de maio de 1980 a dezembro de 1990, baseado na metodologia de decomposição via ondaletas, em dois subníveis, modelos de previsão, dentro destas séries, via modelos ARIMA e de redes neurais, fazendo posteriormente às previsões, a reconstrução da série original. Seu objetivo foi de determinar se a decomposição da série traria ou não um melhor desempenho para as previsões do que se fosse feito diretamente na série original.

O autor utilizou testes de Dickey-Fuller para raízes unitárias, para testar estatisticamente a existência de estacionariedade das séries tanto na série original como nas decomposições das subséries. Observou que a decomposição não altera a estacionariedade da série, pois o S&P 500 era não estacionária, bem como suas duas decomposições pelas ondaletas de Morlet's e chapéu mexicano. Como resultados, o autor encontrou que as redes neurais com decomposição, via ondaletas, obteve 26,3% de redução dos erros medidos pelo MAPE. Para os modelos econométricos com decomposição via ondaletas, a melhora foi de 17,3%, melhorando a qualidade das previsões.

Ariño (1995) analisou a previsão de vendas de carro no mercado espanhol, no período de janeiro de 1974 a dezembro de 1994. Utilizando as primeiras 240 observações e testando para as últimas 12, referentes a ano de 1994, ou seja, fez a previsão para doze

passos à frente. Utilizando os modelos ARIMA, na série original, obteve erros medidos pelo RMSE de 16,964%. Com a decomposição, via ondaletas pela ondaleta DAUB8 (Ondaleta de Daubechies), o erro caiu para 12,895%, quando fizeram as previsões em ambas as subséries pelos mesmos modelos e reconstruindo pela mesma onda.

Homsy, Portugal e Araújo (2000) comparam, sob forma de estudo de casos, previsões relativas a três diferentes métodos de modelagem de séries de tempo, os quais consistem na aplicação da metodologia ARIMA, tanto da forma usual quanto amparada por dois procedimentos auxiliares, baseados na análise de ondaletas para as séries da produção industrial, exportações brasileiras, volume de pesca na Groelândia. Utilizou também o procedimento de alisamento exponencial das séries. Com base na medida MAPE de desempenho econômico, o autor mostrou que o procedimento de decomposição via ondaletas mostrou-se superior em 3,25% contra 6,88% para a série da produção industrial; para a série das exportações, indicou 9,9% contra 15,52% e para a série da produção de pesca, indicou 17,47% contra 27,35% a favor, em todos os casos, da decomposição via ondaletas. A forma de ondaleta utilizada foi a ondaleta de Haar. Isto indica, segundo o autor, ser um indício de que a modelagem em separado (ondaletas) de subséries de baixas e altas frequências contribui positivamente para a qualidade das previsões.

Incorporando dados mais recentes, Wong et al. (2003) trabalharam com a taxa de câmbio para previsões de dez passos à frente e na construção dos modelos de tendência, utilizando uma amostra de 512 observações de 1º de agosto de 1989 a 31 de julho de 1991. Utilizou a ondaleta de Daubechies (DAUB7), e comparando com os modelos ARIMA, o menor erro, 0,95% ficou com a decomposição via ondaletas, contra 2,2% para os modelos ARIMA, também medidos pelo MAPE.

Como visto, recentemente, o desenvolvimento da teoria das ondaletas vem rapidamente se confirmando nos vários campos da ciência, devido à aplicação de suas ferramentas matemáticas de suporte a análise de séries temporais. Essa metodologia representa um vasto campo de pesquisa quantitativa (WONG et al. 2003).

Todavia, desde seu início, a teoria de ondaletas se mostrou interdisciplinar, o que, em parte, explica a grande popularidade adquirida nos últimos anos. Essas conclusões

fortalecem a carência latente dos segmentos ligados à previsão de séries financeiras em desenvolver melhor esta metodologia das ondaletas, aplicadas ao mercado brasileiro. Também corroboram para a idéia de que se pretende diagnosticar a qualidade das previsões, provendo a alta administração de investimentos com análises de suporte a decisões acionáveis no mercado.

2.4 PROCESSO DE GERAÇÃO DE DADOS (DGP – *DATA GENERATION PROCESS*)

Segundo Doornik e Hendry (2001), somente em alguns casos mais simples, a teoria econométrica é capaz de deduzir as propriedades estatísticas de uma distribuição amostral. Na maior parte das vezes, a teoria é forçada a fazer uso da álgebra assintótica, produzindo resultados que se aplicam apenas às grandes amostras. Mesmo que, em alguns casos, estes resultados assintóticos tenham apresentados boas aproximações para as distribuições amostrais associadas a amostras não muito grandes, ainda não se pode ter certeza dos valores obtidos. Por este motivo, utilizam-se ferramentas computacionais para encontrar as propriedades estatísticas da distribuição amostral em pequenas amostras usando o chamado Método de Simulação.

O Método de Simulação é usado para simular um processo gerador de dados (DGP) criando milhares de exemplos típicos, criando para cada um deles o valor da estatística em questão, e em seguida, usando-os na caracterização da distribuição amostral. Tais métodos de simulação consistem geralmente nos seguintes experimentos:

- assumir que o verdadeiro DGP é conhecido;
- especificar o conjunto de variáveis explicativas;
- criar um conjunto artificial de dados, via seleção aleatória de observações, a partir de um DGP e da população de regressores;
- estimar o modelo, a partir do conjunto de dados artificiais gerados, usando o estimador ou modelo sob investigação; e
- gerar um novo conjunto artificial de dados e obter as novas estimativas e testes estatísticos para os dados.

Para os autores, cabe salientar que, para assumir um gerador de dados específico, a verdadeira função populacional nunca é conhecida (pois se fosse, não precisaria ser estimada). Portanto, um DGP nada mais é do que um modelo que se espera caracterizar com certa acurácia algumas características dos dados. Entretanto, a simulação visa construir uma situação hipotética com propriedades conhecidas e então comparar os resultados com a verdade.

3 – METODOLOGIA DA PESQUISA

3.1 METODOLOGIA

No referencial teórico, foram apresentadas e discutidas as abordagens sobre previsão de sucessões cronológicas e seus principais resultados, visando refletir sobre o uso desta nova metodologia, capaz de decompor uma sucessão cronológica, via ondaletas, conjuntamente com os modelos já existentes de previsão e comparar a qualidade de previsões obtidas para assessorar as decisões de investimentos.

Este trabalho está fundamentalmente calcado em uma abordagem quantitativa, a fim de atender ao objetivo da pesquisa. Os estudos quantitativos são aqueles em que os pesquisadores estabelecem hipóteses e as testam por meio da mensuração de variáveis operacionais definidas, quantificando o resultado com o uso de variados métodos quantitativos e estatísticos. É um tipo de pesquisa tradicionalmente aceito como tendo confiabilidade, fidedignidade e validade, desde que realizada seguindo uma metodologia rigorosa. A validação dá-se pelo nível de significância adotado nos testes estatísticos (MARCONI; LAKATOS, 2002).

A metodologia utilizada fundamenta-se na análise e na construção de modelos univariados de previsão de sucessões cronológicas lineares e não-lineares, com base em dados simulados, conjuntamente com a decomposição de ondaletas. Há uma grande variedade de modelos aplicáveis a estudos dessa natureza. Para fins desta pesquisa, optou-se por selecionar os modelos ARIMA-GARCH, ARIMA, bilineares e de Redes Neurais com o algoritmo de aprendizagem recorrente em tempo real, conforme sugerido por Williams e Zipser (1995) e De Oliveira (2003).

Utilizou-se dados de séries simuladas para maior controle dos dados, isto é, trabalhou-se com DGP, utilizando ondaletas em até dois níveis para obtenção de um intervalo de previsão estática para os modelos construídos. Para a etapa de construção dos DGP's, foram construídos programas, que se encontram no Apêndice C, no software Mathematica® 5.0, com rotinas auxiliares nas *Packages*, desenvolvidos pela Wolfram

Inc. para o Mathematica® 5.0, especificamente para tratamento de sucessões cronológicas, redes neurais e ondaletas.

Entendem-se as ondaletas como uma etapa inicial de pré-processamento, para fins de realizar as previsões dentro das subséries de baixa e alta frequência que a ondaleta as decompõe. Posteriormente, fez-se um fluxograma de análise de sucessões cronológicas, visto como essencial para se fazer previsões.

A partir das sucessões cronológicas simuladas, conduziu-se aos testes para confirmação dos modelos, passando-se à análise dos dados propriamente dita. Trabalhou-se com dados de sucessões cronológicas lineares e não-lineares. É razoável que, antes de se utilizar metodologias para se lidar com sistemas não-lineares, seja comprovada a não-linearidade subjacente aos dados da sucessão. As não-linearidades também podem ter outras origens que não as não estacionariedades. Os algoritmos de detecção de não-linearidades são utilizados para quantificar o nível de interações não-lineares encontradas nos dados. Billings e Voon (1986 apud FREITAS, SECURATO e ANDRADE NETO, 2004) afirmam que a relação:

$$\Theta_{u,u^2}(\tau) = E\left[\left(u_t - E[u_t]\right)\left(u_{t-\tau}^2 - E[u_t^2]\right)\right] = 0 \quad \forall \tau \quad (3.1)$$

é válida se e somente se o sistema original for linear. Assim, pode-se identificar não-linearidades e, mais ainda, estabelecer os limites de um intervalo de confiança de, por exemplo, 95% por meio da fórmula $\pm 1.96n$, sendo que n é o comprimento do registro de dados disponíveis. A função de correlação dada pela equação (3.1) pode ser estimada, utilizando os dados disponíveis. Logo, se os valores da função sair fora dos limites estabelecidos, o sistema que gerou os dados deverá ser representado por um modelo não-linear.

Freitas, Securato e Andrade Neto (2004) afirmam que um critério relevante para a escolha deste tipo de teste é sua capacidade de discriminação. A capacidade de discriminação do teste é definida como a probabilidade de se rejeitar a hipótese nula quando realmente a mesma é falsa. Obviamente, também depende de como quão intensamente os dados atuais se desviam da hipótese nula. Um indicador cujo poder de discriminação é particularmente eficiente para este fim é o de detecção de não-linearidades, via assimetrias, sob reversão no tempo, apresentado a seguir:

$$\phi^{rev}(\tau) = \frac{\sum_{n=\tau+1}^N (X_n - X_{n-\tau})^3}{\left[\sum_{n=\tau+1}^N (X_n - X_{n-\tau})^2 \right]^{3/2}} \quad (3.2)$$

Segundo os autores, esta estatística permite detectar possíveis assimetrias sob reversão no tempo. Schreiber e Schmitz (1999), compararam quantitativamente os testes mais populares de não-linearidades e concluiu que esta métrica apresenta melhores resultados empíricos.

3.2. PREVISÃO

Em sucessões cronológicas simuladas é comum encontrar séries com tendência na média, de modo que em se tomando diferenças finitas obtêm-se séries estacionárias. Todavia, segundo Pankratz e Dudley (1987), se a variância aumentar com o tempo, só as diferenças podem não ser suficientes para estacionarizar a série.

Se, por exemplo, X_t é uma sucessão cronológica original e

$$y_t = g(X_t) \quad (3.3)$$

uma transformação em X_t , que no caso dos retornos é a função logarítmica, conforme Campbell, Lo e Mackinlay (1997).

O problema passa a ser, então, buscar previsões para X_{t+h} , dado que se tem um modelo de previsão para de y_t e também para as previsões de

$$y_{t+h} = g(X_{t+h}) \quad (3.4)$$

Segundo Pankratz e Dudley (1987), uma maneira “ingênua” de se proceder para obtenção da previsão da série original é considerar a equação (3.4) e substituir as previsões por valores futuros:

$$\hat{y}_{t+h} = g(\hat{X}_{t+h}) \quad (3.5)$$

e posteriormente tentar obter \hat{X}_{t+h} em função de \hat{y}_{t+h} , a partir da equação (3.5). Isso poderia ser feito se a função g possuir inversa, e conseqüentemente:

$$\hat{X}_{t+h} = g^{-1}(\hat{y}_{t+h}) \quad (3.6)$$

Para Campbell, Lo e Mackinlay (1997), a logaritmização dos valores observados é uma das transformações mais usuais para estacionarizar sucessões cronológicas econômico-financeiras e permite, quer a linearização de tendências exponenciais, quer a estabilização de variâncias.

No caso das previsões com os log-retornos, tem-se:

$$y_t = \ln(X_t) \quad (3.7)$$

e então

$$e^{y_t} = e^{\ln(X_t)} = X_t \quad (3.8)$$

e as previsões para X_{t+h} ficariam:

$$X_t = e^{\hat{y}_{t+h}} \quad (3.9)$$

Com efeito, tem-se:

$$X_{t+h} = E[X_{t+h} | X_t, X_{t-1}, \dots] \neq e^{[y_{t+h}]} = e^{[E[y_{t+h} | y_t, y_{t-1}, \dots]]} \quad (3.10)$$

sendo que y_t segue a equação (3.5).

Todavia, Granger e Newbold (1976) afirmam que as previsões feitas com a equação (3.7) conduzem a previsões viciadas e, como consequência, o erro quadrático médio da previsão aumentará.

Prankratz e Dudley (1987) frisam que no caso de y_t ser gaussiana, a previsão ótima deverá ser dada por:

$$X_t = e^{\left[\hat{y}_{t+h} + \frac{1}{2} Var_{y_{t+h}} \right]} \quad (3.11)$$

onde $Var(h) = \text{Variância}[\widehat{e_t(h)}]$, sendo que, $e_t(h) \equiv y_{t+h} - E_t[y_{t+h}]$, conforme define Enders (2003).

Segundo Morettin (2003), se $y_t = \ln(X_t)$ seguir, por exemplo, um modelo ARIMA, sabe-se que a distribuição condicional de y_{t+h} , dado o passado, é uma distribuição

$N(\hat{y}_t(h), \mathbb{V}_y[h])$ é um intervalo de confiança para y_{t+h} , com nível de confiança de 95% será $\hat{y}_t(h) \pm 1,96\sqrt{\hat{\mathbb{V}}_y[h]}$.

Uma vez que um ou mais modelos de séries temporais foram selecionados, devemos gerar previsões 1-passo ou h-passos à frente de previsão para y_t (onde y_t denota um variável diferenciada adequadamente). As previsões referem-se às observações y_{t+h} e conseqüentemente será denotada por \hat{y}_{t+h} , com $h = 1, 2, 3, \dots, m$. Em geral, a previsão um passo à frente é definida por \hat{y}_{t+1} , a qual é baseada no conjunto de informação y_n . No caso onde t observações foram registradas para avaliar uma previsão um passo à frente, avança-se o conjunto de informação com cada previsão. Por exemplo, \hat{y}_{t+2} é baseado em y_{t+1} e \hat{y}_{t+3} é baseado em y_{t+2} . A motivação intuitiva por trás dessa lógica é que não é sempre possível ser capaz de re-estimar os parâmetros do modelo de séries temporais. Neste caso, são mantidos os estimadores para os parâmetros fixados para a amostra $t = 1, 2, \dots, n$, enquanto faz-se previsões um passo à frente para y_{t+1} até y_{t+h} . Se possível, podemos também decidir estimar os parâmetros para cada um dos conjuntos de informação y_{t+h-1} . Além disso, podemos então deletar os primeiros h dados pontuais, de tal forma que a amostra de estimação tem o mesmo tamanho para as m previsões. Finalmente, uma previsão h -passos (ou dinâmica) denota uma previsão de y_{t+h} (para $h = 1, 2, \dots, m$), a qual é baseada em y_t . Obviamente, quando h é maior que a memória do modelo (ou, por exemplo, o comprimento do polinômio AR), a previsão \hat{y}_{t+h} provavelmente inclui previsões para dados pontuais prévios como \hat{y}_{t+h-1} e também os erros de previsão tendem a ser maiores (ENDERS, 2003).

3.3 DESENVOLVIMENTO DA PESQUISA

Conforme mencionado anteriormente, a linha mestra desta pesquisa é estudar e comparar os métodos de previsão de sucessões cronológicas com a decomposição via ondaletas. Para tanto, dividiu-se o trabalho nas seguintes etapas:

Num primeiro momento da pesquisa, levantou-se, na literatura, teses e dissertações, via *UMI - Proquest*, artigos em periódicos da área, em busca de referencial de trabalhos já divulgados na área, principalmente em busca de aplicações das técnicas econométricas que serão utilizadas. Posteriormente, estabeleceu-se o organograma da Figura 14, a seguir, para apresentação passo a passo das etapas de uma análise de sucessões cronológicas, a fim de pôr ordem nos procedimentos, conforme apresentados a seguir.

Dentre as etapas descritas no organograma, tem-se o pré-processamento, onde se retém *outliers*, faz-se a decomposição por ondaletas, diferenciação, logaritmização e normalização da série. Parte-se, então, para testar a estacionariedade do processo. A seguir, o teste para normalidade para prever o tipo de dependência do processo estacionário. Logo após, deve-se determinar se a mesma série é, de fato, estatisticamente independente. Se a hipótese nula de independência é rejeitada, isto implica a aceitação da hipótese alternativa da existência de dependência linear ou dependência não-linear.

Caso a sucessão cronológica apresente dependência linear, então se deve especificar os modelos lineares para ajustes. Alternativamente, caso ocorra dependência não-linear, utilizar os modelos específicos para dados não-lineares. A especificação de modelos não-lineares inicia-se no posteriormente, onde modelos não-lineares aditivos ou multiplicativos devem ser construídos. Faz-se, então, o pós-processamento que consiste na desnormalização, deslogaritmização, integração e reconstrução da sucessão cronológica original para obter a sucessão no nível. O próximo passo é a análise da qualidade das previsões e encerra-se com a validação dos modelos, estimando os testes de independência e normalidade dos resíduos.

O fluxograma, a seguir, resume e ilustra todo o processo descrito:

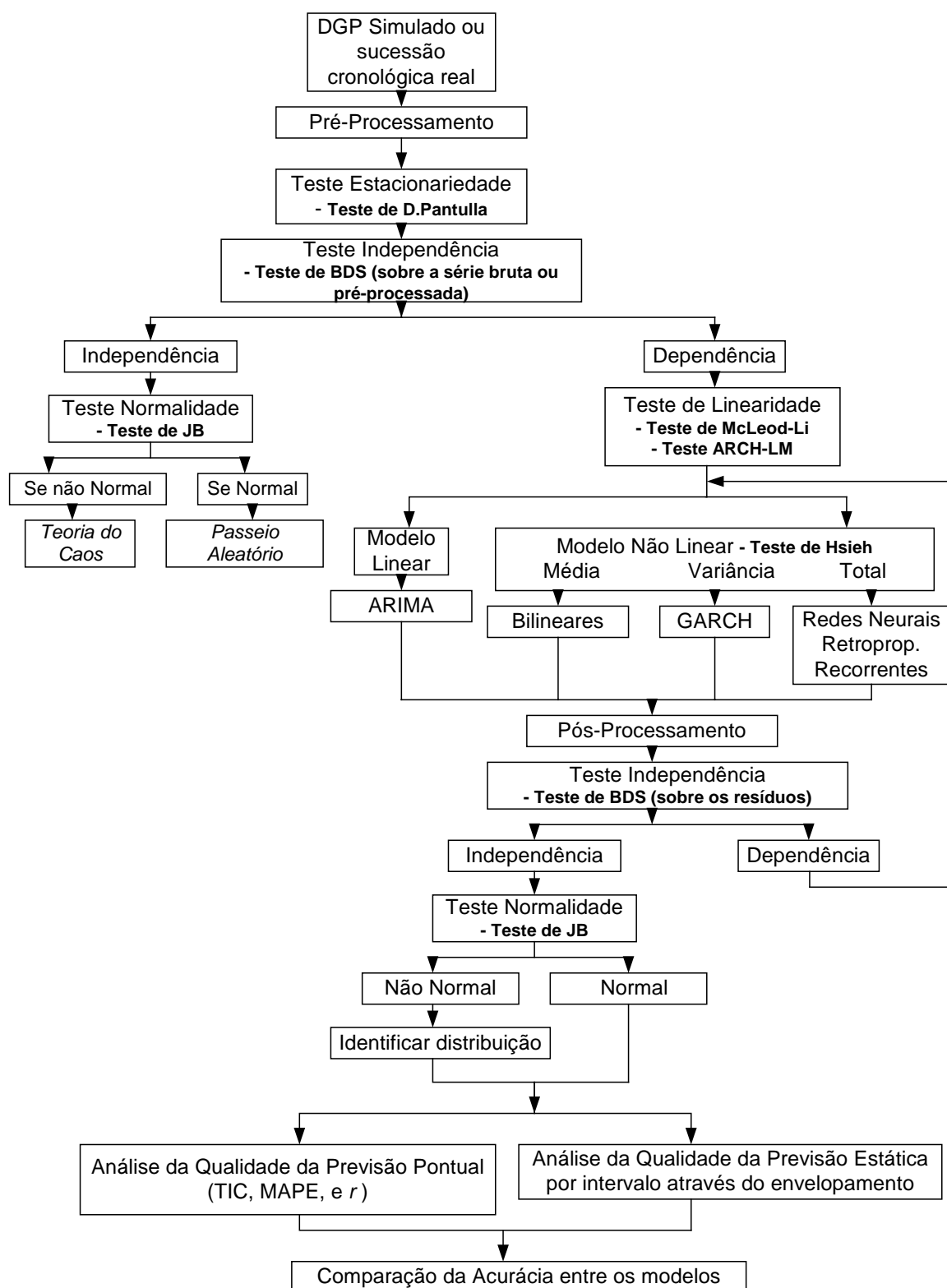


Figura 14: Fluxograma proposto para análise de sucessões cronológicas

Seguindo o fluxograma da Figura 14, depois de se estimar um modelo de uma sucessão cronológica, passa-se a testar se a especificação estava correta. A pressuposição foi de

que os termos de erros aleatórios ε_t , no processo efetivo, têm distribuição normal e são independentes. Então, se o modelo foi especificado corretamente, os resíduos $\hat{\varepsilon}_t$ deveriam ser semelhantes a um processo ruído branco. Em particular, tem-se a expectativa de que os resíduos quase não tenham correlação entre si, de modo que a FAC amostral dos resíduos seja próxima de zero para uma defasagem $k \geq 1$.

Seguindo os passos descritos, aplica-se novamente os testes de independência estatística (Teste BDS) e teste de Jarque-Bera de normalidade. Caso mantenha dependência nos resíduos, o modelo deve ser novamente gerado e testado. Se houver independência, deve-se estimar a distribuição dos resíduos para, posteriormente, estimar-se as bandas de previsão para o envelopamento.

4 – RESULTADOS EXPERIMENTAIS

Neste capítulo analisar-se-ão alguns resultados empíricos obtidos, a partir da utilização de metodologias econométricas tradicionais e de redes neurais com algoritmos de retropropagação e redes recorrentes para modelagem de sucessões cronológicas, visando à utilização do fluxograma proposto.

A pesquisa foi conduzida por meio de simulações de dados através de DGPs gerados pelo software MATHEMATICA® 5.0 através de construção de pacotes para gerar os dados que também foram usados para tratamento das decomposições e reconstruções, via ondaletas e todo o tratamento das previsões, dentro das subséries, via combinação com pacotes específicos do próprio software. Usou-se também o EVIEWS® 5.0, para tratamento dos testes e previsões. Tais programas se encontram no apêndice C. Para as redes neurais, utilizou-se o MATLAB® 6.5, para as redes recorrentes e o software Statistica da StatSoft© 6.1, para as redes com retropropagação.

4.1 SIMULAÇÃO DE PROCESSO ARIMA-GARCH

Considerando um DGP para o modelo ARIMA-GARCH da forma ARIMA(2,0,0)-GARCH(1,1), cujos parâmetros são: $\phi_1 = 0,2$ e $\phi_2 = -0,5$ e $\alpha_0 = 0,06, \alpha_1 = 0,04, \beta_1 = 0,7$.

Utilizando-se o programa 1, disponível no apêndice C, foram geradas 1200 observações para o modelo que foram divididas em duas partes: a primeira com mil observações foi utilizada para gerar o modelo para previsão e as 200 restantes, utilizadas para validar o modelo.

A seguir, tem-se seu gráfico com 1000 instantes de tempo:

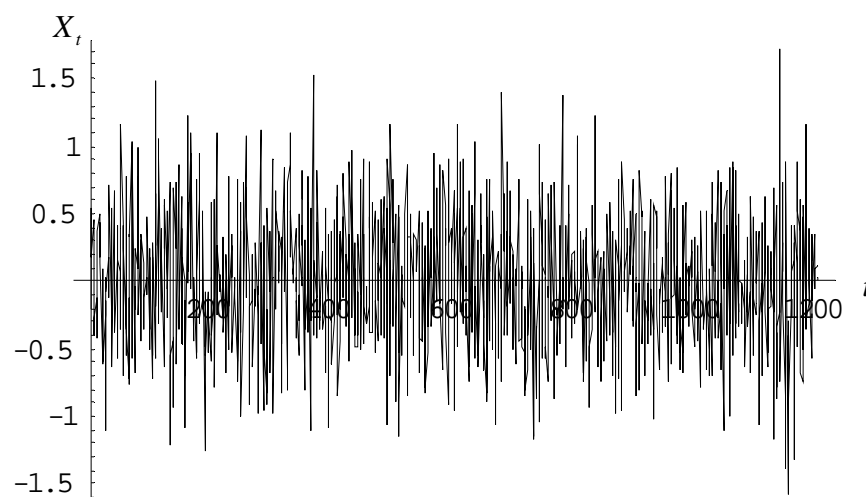


Figura 15: DGP para o modelo ARIMA(2,0,0)-GARCH(1,1)

Verificando a sucessão cronológica como um todo, tem-se:

Teste de raiz unitária

DGP ARIMA(2,0,0)-GARCH(1,1) $\phi_1 = 0,2$ e $\phi_2 = -0,5$ e $\alpha_0 = 0,06, \alpha_1 = 0,04, \beta_1 = 0,7$		
Hipóteses	Estatística de Teste	Valores Críticos a 5%
$H_0 : 3RU$ $H_1 : 2RU$	-69,47651	-2,76
$H_0 : 2RU$ $H_1 : 1RU$	-22,28469 -38,53137	-3,37 -3,77
$H_0 : 1RU$ $H_1 : 0RU$	-31,02768 -14,84719 -17,64252	-3,42 -3,80 -4,16

Considerando um nível de confiança de 95%, se rejeita a todas as hipóteses nulas, o que indica que ambas as subséries são estacionárias.

Teste de independência estatística (BDS) sobre a sucessão cronológica

Modelo ARIMA-GARCH				
Dimensão	BDS	Erro-padrão	Estatística Z	p-valor
2	-6,94E-06	1,46E-06	-4,769876	0,0000
3	-9,57E-07	3,60E-08	-26,56382	0,0000
4	-9,41E-09	6,70E-10	-14,04558	0,0000
5	-9,26E-11	1,09E-11	-8,491502	0,0000
6	-9,07E-13	1,65E-13	-5,509733	0,0000

Calculou-se até a sexta dimensão para os dados padronizados e concluiu-se por rejeitar a hipótese nula ao nível de significância de 5%, indicando que os dados da sucessão cronológica são dependentes, baseado na análise do p-valor, pois pelo menos um dos p-valores teve seu valor menor que 5%.

Teste de Normalidade de Jarque-Bera

A figura, a seguir, ilustra a distribuição da sucessão cronológica.

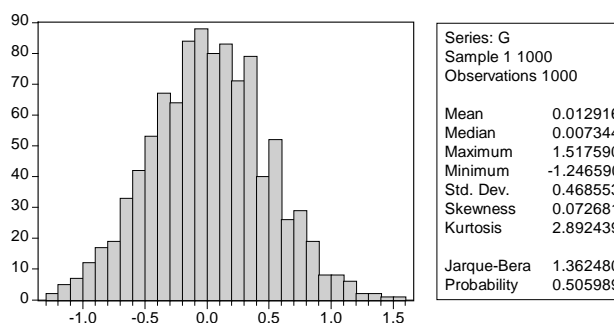


Figura 16: Distribuições do modelo ARIMA(2,0,0)-GARCH(1,1) simulado

De acordo com o p-valor, tem-se uma distribuição normal.

Teste de McLeod-Li

O resultado da estatística de teste calculado no Eviews® 5.0 é:

Modelo ARIMA-GARCH			
Estatística F	11,97598	p-valor	0,00000

Dessa forma, o teste de McLeod-Li, para 5 defasagens, rejeitou a hipótese nula pelo fato do p-valor ser inferior a 5%. Isto indica que a sucessão cronológica simulada possui comportamento não-linear.

Teste de Hsieh

Uma vez estabelecido que existe algum tipo de não-linearidade em uma sucessão cronológica, é preciso identificar o tipo de não-linearidade existente, se na média ou na variância condicionadas.

Para a realização do teste, fez-se o programa 4 (ver apêndice C) que irá calcular os coeficientes (i,j) para cada uma das subséries. Os resultados desta etapa vem a seguir:

Modelo ARIMA-GARCH		
Coeficiente	Estatística de Hsieh	p-valor
1 1	-0,0333	0,4867
1 2	-0,6631	0,2536
1 3	-2,4526	0,0071*
1 4	-0,8405	0,2003
1 5	0,5051	0,6933
2 2	0,408	0,6584
2 3	-1,6157	0,0531
2 4	-0,9803	0,1635
2 5	-2,2876	0,0111*
3 3	-0,0189	0,4925
3 4	2,1549	0,9844
3 5	1,828	0,9662
4 4	0,8702	0,8079
4 5	-1,7137	0,0433*
5 5	-0,5708	0,2841

Conforme se observa no quadro anterior, para quinze pares (i,j), os valores dos coeficientes do teste de Hsieh amostrais para as duas subséries e o correspondente p-valor para um nível de confiança de 95%, apresenta apenas três coeficientes significantes (indicado com *), podendo-se dizer que, de maneira geral, não se rejeita a hipótese nula dos coeficientes iguais a zero, o que indica uma não-linearidade na variância condicionada.

Seguindo o fluxograma proposto no trabalho, a primeira etapa corresponde ao pré-processamento da decomposição via ondaletas. Foi escolhida a ondaleta de Daubesch número 1 (DB1), por ser a mais trabalhada nos artigos da revisão da literatura.

Através do programa 2 do apêndice C, que foi desenvolvido para estas aplicações, pode-se gerar o pré-processamento do DGP gerado acima, em apenas um nível de subdivisões.

A figura a seguir ilustra a decomposição (pré-processamento) em primeiro nível da série no nível original:

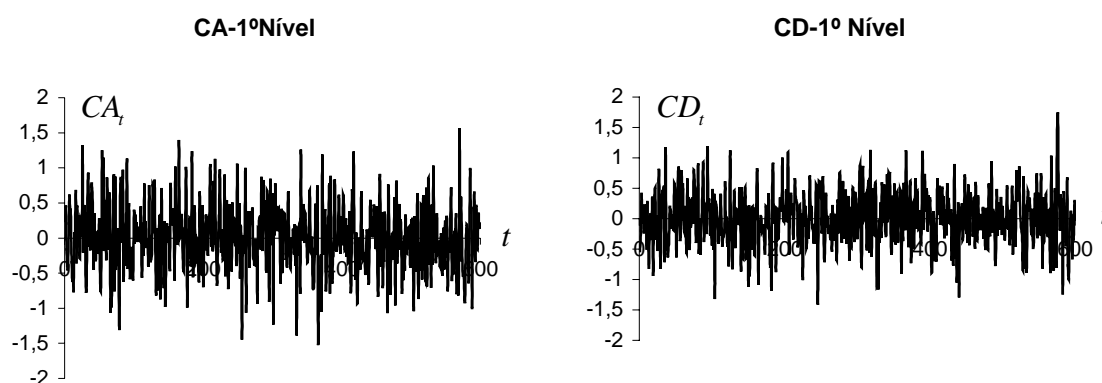


Figura 17: Pré-processamento em 1º Nível da série original do modelo ARIMA(2,0,0)-GARCH(1,1)

Testes de raiz unitária (*RU*):

O teste de DFA de raízes unitárias foi aplicado no Eviews® 5.0. A seguir, são apresentados os testes para verificar se a série apresenta uma ou múltiplas raízes unitárias através das hipóteses:

H_0 : a série é não-estacionária (apresenta raiz unitária)

H_1 : a série é estacionária (sem raiz unitária)

Para o teste de raízes unitárias, foi utilizado o teste de DF aumentado, conhecido como teste de Dikey-Pantula, que apresentou os seguintes resultados com 5% de significância para as subséries decompostas no pré-processamento:

Subsérie: CA			Subsérie: CD		
Hipóteses	Estatística de Teste	Valores Críticos a 5%	Hipóteses	Estatística de Teste	Valores Críticos a 5%
$H_0 : 3RU$ $H_1 : 2RU$	-54,6783	-2,76	$H_0 : 3RU$ $H_1 : 2RU$	-52,1822	-2,76
$H_0 : 2RU$ $H_1 : 1RU$	-17,63888 -29,18961	-3,37 -3,77	$H_0 : 2RU$ $H_1 : 1RU$	-17,22859 -29,60606	-3,37 -3,77
$H_0 : 1RU$ $H_1 : 0RU$	-25,43932 -9,790482 -14,86584	-3,42 -3,80 -4,16	$H_0 : 1RU$ $H_1 : 0RU$	-25,43257 -9,322556 -15,29174	-3,42 -3,80 -4,16

Considerando um nível de confiança de 95%, se rejeita a todas as hipóteses nulas, o que indica que ambas as subséries são estacionárias.

Teste para Independência Estatística – Teste BDS

Como visto anteriormente, o teste para independência pode ser aplicado para ver se a sucessão cronológica apresenta independência, pois, se for, estão o processo é RBI.

Suas hipóteses são:

H_0 :o processo CA é estocasticamente independente

H_1 : o processo CA é estocasticamente dependente

H_0 :o processo CD é estocasticamente independente

H_1 : o processo CD é estocasticamente dependente

Subsérie: CA					Subsérie: CD				
Dimensão	BDS	Erro Padrão	Estatística Z	p-valor	Dimensão	BDS	Erro Padrão	Estatística Z	p-valor
2	-3,00E-06	4,85E-05	-0,061797	0,9507	2	4,77E-05	3,97E-05	1,199797	0,2302
3	2,87E-05	5,75E-06	4,989138	0,0000	3	-1,82E-06	4,68E-06	-0,389464	0,6969
4	1,81E-05	5,12E-07	35,39495	0,0000	4	-6,13E-06	4,14E-07	-14,81243	0,0000
5	7,84E-06	3,99E-08	196,4425	0,0000	5	-3,04E-07	3,20E-08	-9,493997	0,0000
6	-1,52E-08	2,88E-09	-5,280915	0,0000	6	-1,49E-08	2,29E-09	-6,514062	0,0000

O software Eviews® 5.0 calcula a estatística BDS para todas as dimensões de dois até o valor especificado, variando a dimensão de correlação para maximizar o poder do teste. Neste caso, calculou-se até a sexta dimensão para os dados padronizados e concluiu-se por rejeitar a hipótese nula ao nível de significância de 5%, indicando que os dados da

sucessão cronológica são dependentes, baseado na análise do p-valor, pois pelo menos um dos p-valores deve ser menor que 5%.

Teste de Normalidade

Na figura a seguir, têm-se as estatísticas das duas subséries do pré-processamento, onde a hipótese nula e alternativa são:

H_0 : a série segue uma distribuição normal

H_1 : a série não segue uma distribuição normal

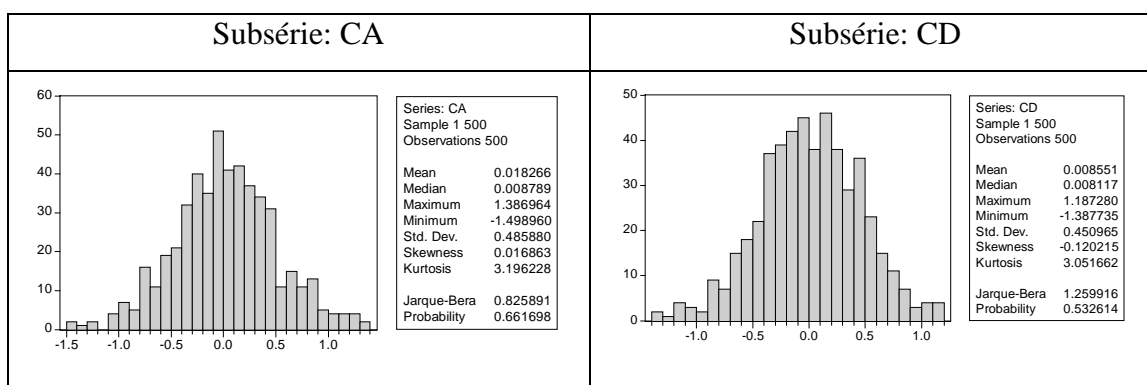


Figura 18: Distribuições das subséries em 1º nível do modelo ARIMA-GARCH

Para as duas subséries, não há evidências de não rejeição da hipótese nula pelo p-valor que é maior que 5%, nível de significância adotado no teste. Portanto, pode-se considerar que as duas subséries apresentam comportamento normal. O teste de Jarque-Bera também foi realizado no Eviews ® 5.0.

Teste de McLeod-Li

O teste é baseado na autocorrelação amostral dos quadrados dos resíduos de uma sucessão cronológica estacionária gaussiana.

A hipótese a ser testada é:

H_0 : X_t^2 é independente, isto é, a sucessão cronológica é linear

H_1 : X_t^2 é dependente, isto é, a sucessão cronológica é não-linear

O resultado da estatística de teste calculado no Eviews® 5.0 é:

Subsérie: CA				Subsérie: CD			
Estatística F	20,41297	p-valor	0,0000000	Estatística F	13,87677	p-valor	0,000000

Dessa forma, o teste de McLeod-Li para 5 defasagens rejeitou a hipótese nula pelo fato do p-valor ser inferior a 5%. Isto indica que ambas as subséries possuem comportamento não-linear.

Matematicamente, este resultado pode ser entendido que a série não pode ser escrita como combinação linear dos termos de erro da forma $X_t = f\left(\{\varepsilon_{t-i}\}_{i=1}^{\infty}\right) = \sum_{i=1}^{\infty} \alpha_i \varepsilon_{t-i}$, sendo que $\{\varepsilon_t\} \sim RB$, pelo Teorema de Wold (HAMILTON, 1994).

Freitas, Securato e Andrade Neto (2004) implementaram um algoritmo, com base nos trabalhos de Billings e Voon (1986) e Scheiber (1998), para estudar o comportamento de uma sucessão cronológica ao longo de sua trajetória. Tal implementação ilustrada, a seguir, indica que quando a sucessão cronológica se encontra dentro da banda construída, o seu comportamento tem características lineares e, fora desta banda, comportamento não-linear.

A figura, a seguir, mostra o comportamento para as sub-divisões do primeiro nível de pré-processamento do modelo ARIMA-GARCH, feito com base no programa 7 do Apêndice C.

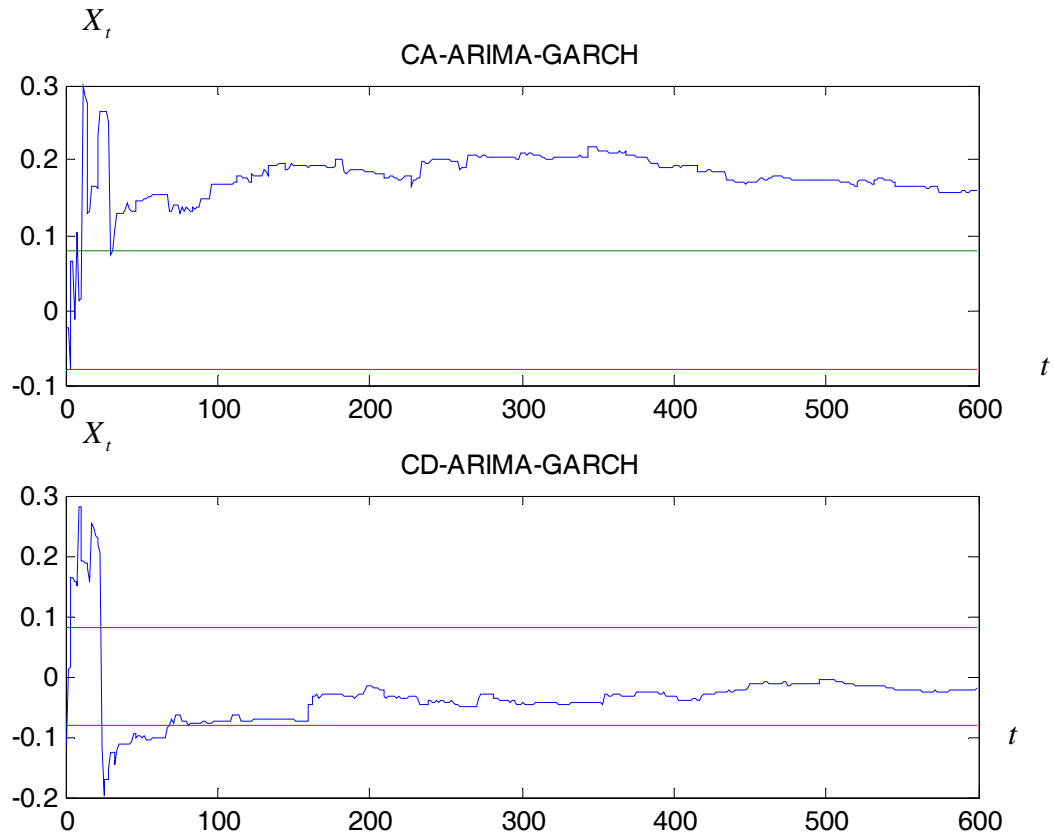


Figura 19: Comportamento da série ao longo do tempo para o modelo ARIMA-GARCH em 1 nível de pré-processamento.

Observa-se que a parte relativa às baixas frequências (CA) manteve o comportamento não-linear durante toda a trajetória, ao passo que a parte de alta frequência da decomposição, oscila no início e depois apresenta comportamento linear.

Teste de Hsieh

Uma vez estabelecido que existe algum tipo de não-linearidade, em uma sucessão cronológica, é preciso identificar o tipo de não-linearidade existente, se na média ou na variância condicionadas.

As hipóteses do teste de Hsieh são:

H_0 : CA apresenta não-linearidade na variância condicionada

H_1 : CA apresenta não-linearidade na esperança condicionada

H_0 : CD apresenta não-linearidade na variância condicionada

H_1 : CD apresenta não-linearidade na esperança condicionada

Para a realização do teste, fez-se o programa 4 que irá calcular os coeficientes (i,j) para cada uma das subséries. Os resultados desta etapa vêm a seguir:

Subsérie: CA			Subsérie: CD		
Coeficiente	Estatística de Hsieh	p-valor	Coeficiente	Estatística de Hsieh	p-valor
1 1	-0,662	0,25399	1 1	0,6549	0,74373
1 2	-1,37	0,08534	1 2	-0,474	0,31775
1 3	-0,6574	0,25546	1 3	0,2844	0,61195
1 4	-2,5114	0,00601*	1 4	1,5583	0,94042
1 5	0,1633	0,56486	1 5	-0,0346	0,4862
2 2	0,6884	0,7544	2 2	0,1942	0,57699
2 3	0,3119	0,62244	2 3	0,5378	0,70464
2 4	0,0571	0,52277	2 4	-0,0107	0,49573
2 5	-0,0592	0,4764	2 5	-0,2428	0,40408
3 3	0,6083	0,72851	3 3	-0,1283	0,44896
3 4	1,9814	0,97623	3 4	-0,1268	0,44955
3 5	-0,7488	0,22699	3 5	-0,1189	0,45268
4 4	-0,028	0,48883	4 4	0,1248	0,54966
4 5	0,0619	0,52468	4 5	1,7597	0,96077
5 5	-0,2244	0,41122	5 5	0,097	0,53864

Conforme se observa, no quadro anterior, para quinze pares (i,j) , os valores dos coeficientes do teste de Hsieh amostrais para as duas subséries e o correspondente p-valor para um nível de confiança de 95%, apresenta apenas um coeficiente significativo (indicado com *), podendo-se dizer que, de maneira geral não se rejeita a hipótese nula dos coeficientes iguais a zero, o que indica uma não-linearidade na variância. Para a sucessão de alta frequência (CD), isto já não ocorre e todos os pares apontam pela dependência na variância condicionada.

De posse das informações descritas acima, gerou-se um modelo ARIMA-GARCH, para cada uma das subséries descritas a seguir:

$$ca_t = \phi_1 ca_{t-1} + \phi_2 ca_{t-2} + \varepsilon_t$$

$$cd_t = \phi_1 cd_{t-1} + \phi_2 cd_{t-2} + \varepsilon_t$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1} \quad \varepsilon_t^2 | I_{t-1} \sim N(0, h_t)$$

Parâmetros Subsérie CA	Coefficiente	Erro Padrão	Estatística t	Significância estimada
ϕ_1	0,009764	0,046617	0,209463	0,8341
ϕ_2	-0,010070	0,045200	-0,222781	0,8237
α_0	0,026695	0,062264	0,428739	0,6681
α_1	0,009319	0,024975	0,373132	0,7091
β_1	0,878487	0,278042	3,159551	0,0016
Modelo	Log de Verossimilhança	AIC	BIC	
AR(2)-GARCH(1,1)	-346,9275	1,413363	1,455639	
Parâmetros Subsérie CD	Coefficiente	Erro Padrão	Estatística t	Significância estimada
ϕ_1	0,002285	0,044694	0,051127	0,9592
ϕ_2	-0,079776	0,042094	-1,895189	0,0581
α_0	0,014863	0,008863	1,677016	0,0935
α_1	-0,025768	0,014765	-1,745226	0,0809
β_1	0,954012	0,040681	23,45115	0,0000
Modelo	Log de Verossimilhança	AIC	BIC	
AR(2)-GARCH(1,1)	-306,4353	1,250744	1,293019	

Após a etapa de identificação e previsão dos modelos nas subséries de baixa e alta frequência, tem-se, a seguir, as estatísticas de previsão estática feitas pelo para 100 passos à frente dentro de cada subsérie. Na etapa do pós-processamento, como para cada um valor previsto em cada sub-nível, na junção novamente para obter a série no nível via transformada inversa da ondaleta, resulta em 200 pontos na série no nível. Como os erros ficaram demasiadamente grandes, considerou-se também a estatística de acurácia pontual para os dez primeiros elementos previstos, para fins de comparação. A previsão foi feita, da forma estática, através do Eviews ® 5.0.

A reconstrução da série no nível na etapa de pós-processamento do fluxograma proposto, foi feita com o auxílio do programa 3, disponível no Apêndice C.

Após a etapa de pós-processamento, tomou-se os valores previstos e através de uma planilha do Microsoft Excel, programou-se as equações (2.8) do MAPE, (2.9) do coeficiente de correlação e (2.10) do TIC. Os resultados obtidos aparecem a seguir:

Identificação com ARIMA-GARCH – Modelo ARIMA(2,0,0)-GARCH(1,1)			
Estatísticas de Previsão Pontual para os 200 pontos	Valor	Estatísticas de Previsão Pontual para os 10 primeiros pontos	Valor
TIC	0,99	TIC	0,83
MAPE	110,00%	MAPE	10,00%
Correlação	-0,11	Correlação	0,16

Utilizando Redes Neurais

A seguir, são apresentados os resultados obtidos, utilizando-se redes neurais artificiais para o tratamento do DGP do modelo ARIMA-GARCH. A rede neural utilizada, primeiramente, foi com o algoritmo de retropropagação e, posteriormente, com a rede recorrente proposta por Williams e Zipser (1995). Foi utilizado, nesta etapa, o Matlab® 6.5.

Para que os dados sejam inseridos na rede, foi utilizado o processo de normalização proposto por Azoff (1994), que considera um vetor de entrada de dimensão n , ou seja, os elementos no vetor variam de $i = 1, \dots, n$. Os elementos t_i^j , para a sinapse i e a amostra total de tamanho N , do vetor indo de $j = 1, \dots, N$ são normalizados e utilizados para análise pela expressão:

$$Z_t = \frac{Y_t - \min \{Y_t\}_{t=1}^N}{\max \{Y_t\}_{t=1}^N - \min \{Y_t\}_{t=1}^N} \in [0;1]$$

Dessa forma, os 500 pontos de cada subsérie da etapa de pré-processamento dos dados do DGP do modelo ARIMA-GARCH são divididos em três partes, a primeira com 200 valores que são utilizados para treinar os pesos sinápticos da rede, a segunda parte para realizar a validação com 100 valores e a terceira para previsão com 100 valores pelo algoritmo de retropropagação.

Tem-se, a seguir, as estatísticas de previsão estática feitas pelo para 100 passos à frente em cada subsérie, que resulta em 200 pontos na série no nível e para apenas os 10 primeiros, via rede neural, com algoritmo de retropropagação:

Redes Neurais: retropropagação – Modelo ARIMA(2,0,0)-GARCH(1,1)			
Estatísticas de Previsão Pontual para os 200 pontos	Valor	Estatísticas de Previsão Pontual para os 10 primeiros pontos	Valor
TIC	0,86	TIC	0,75
MAPE	72,76%	MAPE	3,39%
Correlação	-0,47	Correlação	-0,04

Usando redes neurais recorrentes, tem-se:

Redes Neurais: recorrente – Modelo ARIMA(2,0,0)-GARCH(1,1)			
Estatísticas de Previsão Pontual para os 200 pontos	Valor	Estatísticas de Previsão Pontual para os 10 primeiros pontos	Valor
TIC	0,77	TIC	0,85
MAPE	26,22%	MAPE	1,14%
Correlação	-0,18	Correlação	0,01

A figura, a seguir, mostra a banda de previsão de 95% de confiança para as redes recorrentes para a previsão pontual com 200 pontos com base na expressão $\hat{Y}_t(h) \pm 1,96\sqrt{\hat{V}_Y[h]}$, indicado pelos seus respectivos valores máximos e mínimos nas bandas acima. Posteriormente, fez-se os mesmos para $h=10$, isto é, os 10 primeiros valores.

Como foram simuladas 100 sucessões cronológicas de 1200 observações, foi feito todo o trabalho descrito anteriormente para a primeira delas, isto é, tomou-se a amostra dos mil primeiros elementos para estimação e testada com as 200 observações restantes, construindo um intervalo de previsão para as redes e para o modelo econométrico. Posteriormente, fez-se o mesmo trabalho para as 99 restantes, com os respectivos intervalos de confiança para os valores máximos e mínimos. Tomou-se, então, a média dos 100 valores mínimos do intervalo de confiança $LI = \hat{Y}_t(h) - 1,96\sqrt{\hat{V}_Y[h]}$ e a média dos 100 valores máximos $LS = \hat{Y}_t(h) + 1,96\sqrt{\hat{V}_Y[h]}$. Assim, teve-se um intervalo de confiança médio de toda a simulação $\left[\hat{Y}_t(h) \pm 1,96\sqrt{\hat{V}_Y[h]} \right]$. Verificou-se, a partir daí, apenas para a parte prevista que foram os 200 pontos previstos, quantos de cada uma das 100 sucessões estavam dentro deste envelope criado pelas bandas médias. Para isso,

deu-se o nome de envelopamento, objetivando perceber a média de pontos que caíam dentro deste intervalo como uma medida comparativa de previsão do modelo.



Figura 20: Intervalo de previsão em 1º nível para redes recorrentes para a amostra de teste do modelo ARIMA-GARCH

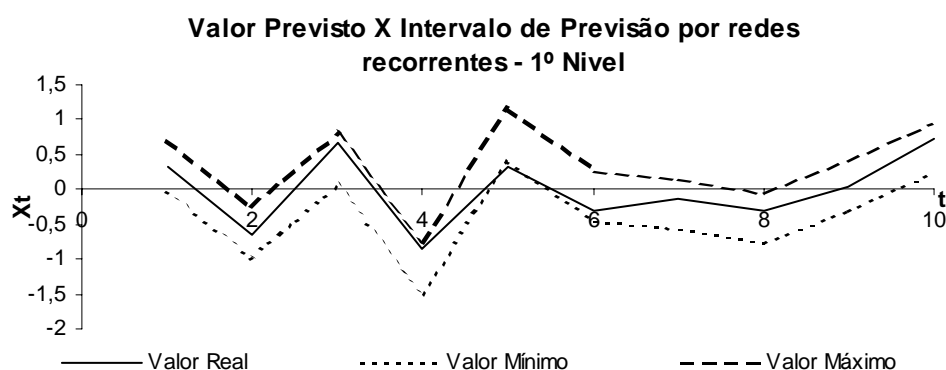


Figura 21: Intervalo de previsão em 1º nível para redes recorrentes para os primeiros 10 elementos do modelo ARIMA-GARCH.



Figura 22: Intervalo de previsão em 1º nível feito por redes neurais com retropropagação do modelo ARIMA-GARCH

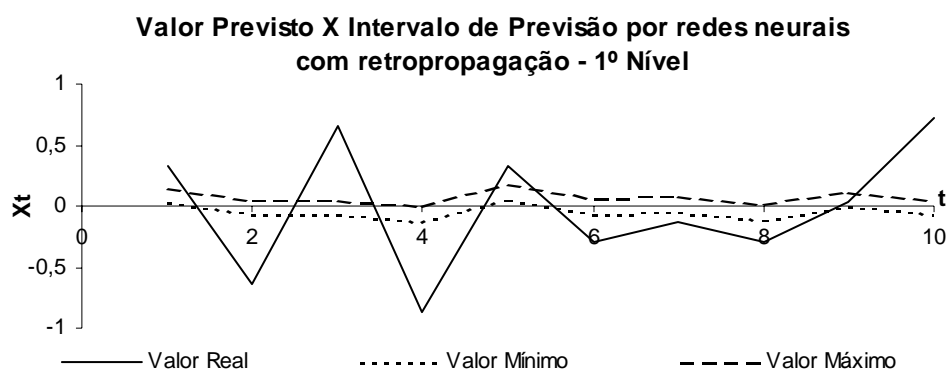


Figura 23: Intervalo de previsão em 1º nível para redes neurais com retropropagação para os primeiros 10 elementos do modelo ARIMA-GARCH.

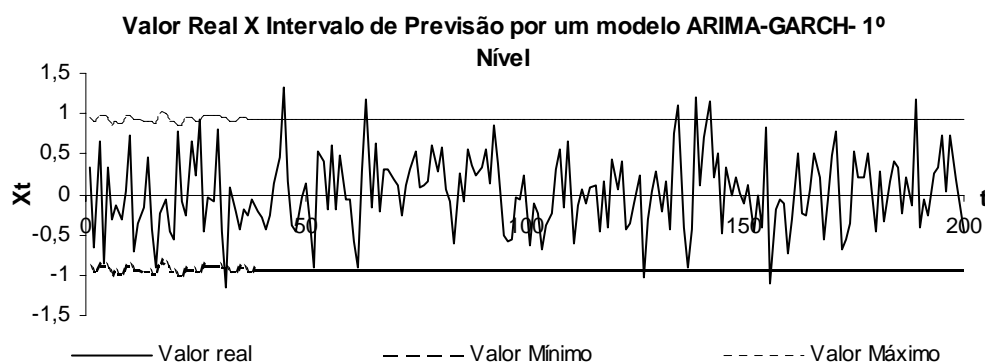


Figura 24: Intervalo de previsão em 1º nível feito por um modelo ARIMA-GARCH

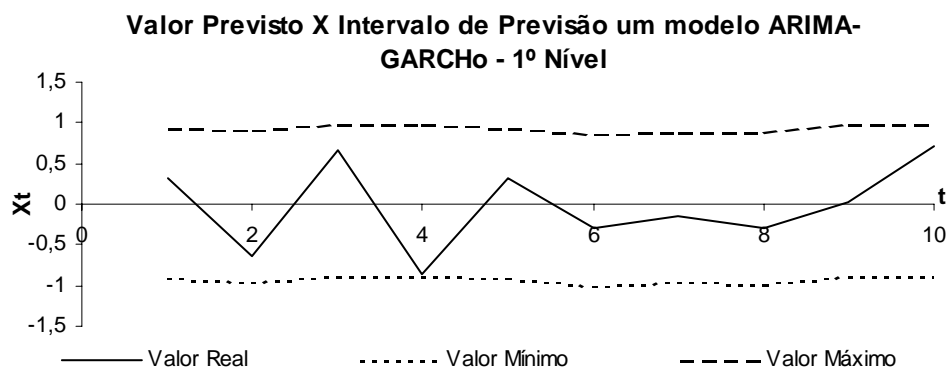


Figura 25: Intervalo de previsão em 1º nível por um modelo ARIMA-GARCH para os primeiros 10 elementos.

Nas figuras anteriores, valor real significa o valor simulado pelo DGP, o valor mínimo é o limite inferior da banda de previsão e o valor máximo o correspondente limite superior.

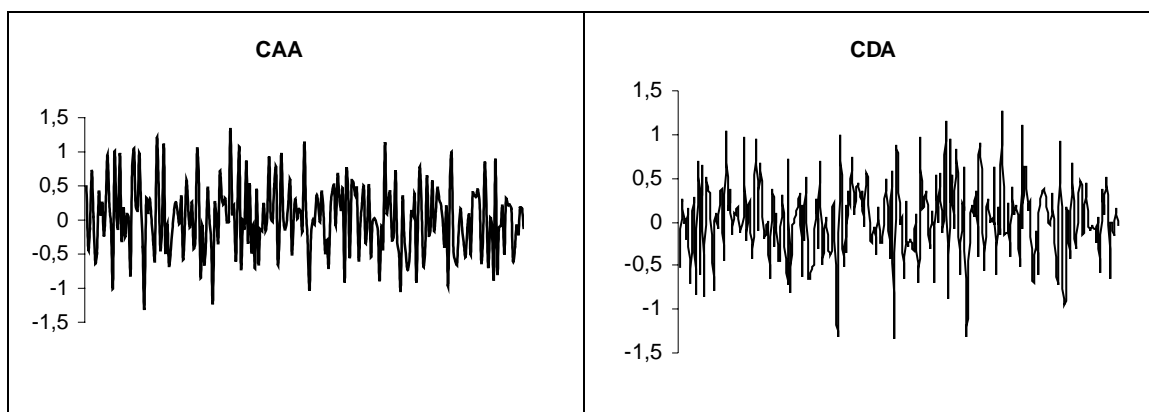
Fazendo o cálculo do envelopamento para 200 séries simuladas e verificando quais estão contidas completamente na banda de confiança de 95%, tem-se:

Porcentagem de Envelopamento	
Modelo	Percentual
Rede Neural Recorrente	96,45%
Rede Neural Retropropagação	93,74%
ARIMA-GARCH	48,42%

O resultado anterior mostra uma vantagem das redes recorrentes sobre os demais modelos no critério do envelopamento da sucessão cronológica prevista.

Simulação no 2º Nível do Pré-Processamento

De acordo com a proposta do trabalho, apresenta-se, a seguir, o pré-processamento feito em dois níveis do DGP para o modelo ARIMA-GARCH com o auxílio do programa 2, aplicado sobre as duas primeiras decomposições:



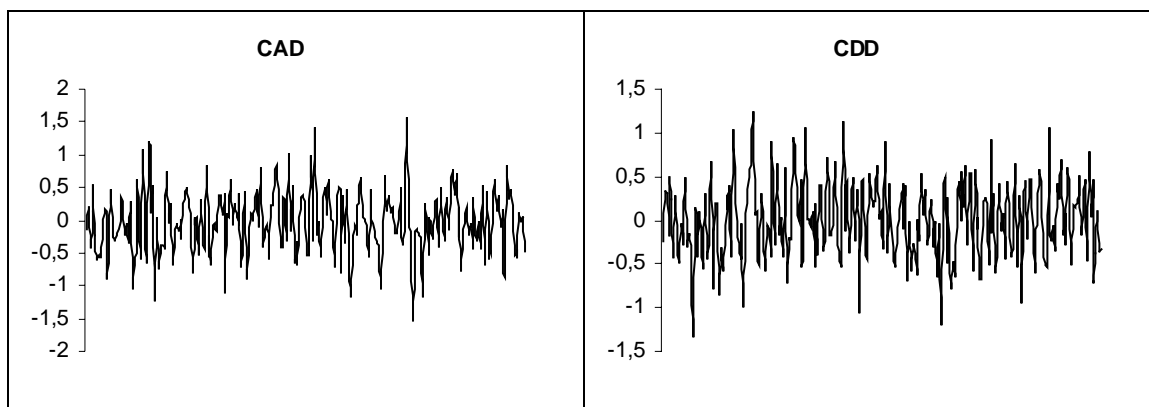


Figura 26: Pré-processamento em 2º nível do modelo ARIMA-GARCH

Seguindo o mesmo roteiro do modelo anterior, apresentam-se, a seguir, os resultados:

Teste de Raízes Unitárias					
Subsérie: CAA			Subsérie: CAD		
Hipóteses	Estatística de Teste	Valores Críticos a 5%	Hipóteses	Estatística de Teste	Valores Críticos a 5%
$H_0 : 3RU$ $H_1 : 2RU$	-31,95246	-2,76	$H_0 : 3RU$ $H_1 : 2RU$	-37,73648	-2,76
$H_0 : 2RU$ $H_1 : 1RU$	-10,32123 -19,99126	-3,37 -3,77	$H_0 : 2RU$ $H_1 : 1RU$	-10,56451 -20,71348	-3,37 -3,77
$H_0 : 1RU$ $H_1 : 0RU$	-14,66594 -7,184838 -8,716525	-3,42 -3,80 -4,16	$H_0 : 1RU$ $H_1 : 0RU$	-15,96954 -6,150885 -9,942485	-3,42 -3,80 -4,16

Teste de Raízes Unitárias					
Subsérie: CDA			Subsérie: CDD		
Hipóteses	Estatística de Teste	Valores Críticos a 5%	Hipóteses	Estatística de Teste	Valores Críticos a 5%
$H_0 : 3RU$ $H_1 : 2RU$	-35,01270	-2,76	$H_0 : 3RU$ $H_1 : 2RU$	-34,24383	-2,76
$H_0 : 2RU$ $H_1 : 1RU$	-10,72136 -19,88246	-3,37 -3,77	$H_0 : 2RU$ $H_1 : 1RU$	-11,73372 -17,81364	-3,37 -3,77
$H_0 : 1RU$ $H_1 : 0RU$	-15,26169 -7,034603 -8,978980	-3,42 -3,80 -4,16	$H_0 : 1RU$ $H_1 : 0RU$	-15,87592 -7,299054 -8,577802	-3,42 -3,80 -4,16

Considerando um nível de confiança de 95%, não se aceita a hipótese nula, o que indica que ambas as subséries são estacionárias.

Teste para Independência Estatística – Teste (BDS)

Subsérie: CAA					Subsérie: CAD				
Dimensão	BDS	Erro Padrão	Estatística Z	p-valor	Dimensão	BDS	Erro Padrão	Estatística Z	p-valor
2	-0,000216	9,45E-05	-2,286002	0,0223	2	-0,00034	6,10E-05	-5,630532	0,0000
3	-4,09E-05	1,19E-05	-3,432417	0,0006	3	-3,99E-0	7,14E-06	-5,590338	0,0000
4	-7,11E-06	1,13E-06	-6,303094	0,0000	4	-5,08E-0	6,28E-07	-8,085309	0,0000
5	-3,77E-07	9,37E-08	-4,018874	0,0001	5	-2,41E-0	4,83E-08	-4,979506	0,0000
6	-2,04E-08	7,21E-09	-2,837173	0,0046	6	-1,14E-0	3,44E-09	-3,310096	0,0009
Subsérie: CDA					Subsérie: CDD				
Dimensão	BDS	Erro Padrão	Estatística Z	p-valor	Dimensão	BDS	Erro Padrão	Estatística Z	p-valor
2	-4,01E-05	8,12E-05	-0,494265	0,6211	2	0,000460	6,98E-05	6,586959	0,0000
3	-6,89E-05	1,02E-05	-6,777076	0,0000	3	7,44E-05	8,55E-06	8,698174	0,0000
4	-6,86E-06	9,55E-07	-7,186312	0,0000	4	-5,99E-06	7,87E-07	-7,611436	0,0000
5	-3,60E-07	7,86E-08	-4,577967	0,0000	5	-2,92E-07	6,34E-08	-4,595844	0,0000
6	-1,83E-08	6,00E-09	-3,059027	0,0022	6	-1,43E-08	4,74E-09	-3,024861	0,0025

O teste de Hsieh rejeita a hipótese de independência de todas as subséries, pela análise do p-valor para os dados padronizados, sendo que, pelo menos um deles, deve ser menor que 5%.

Teste de Normalidade

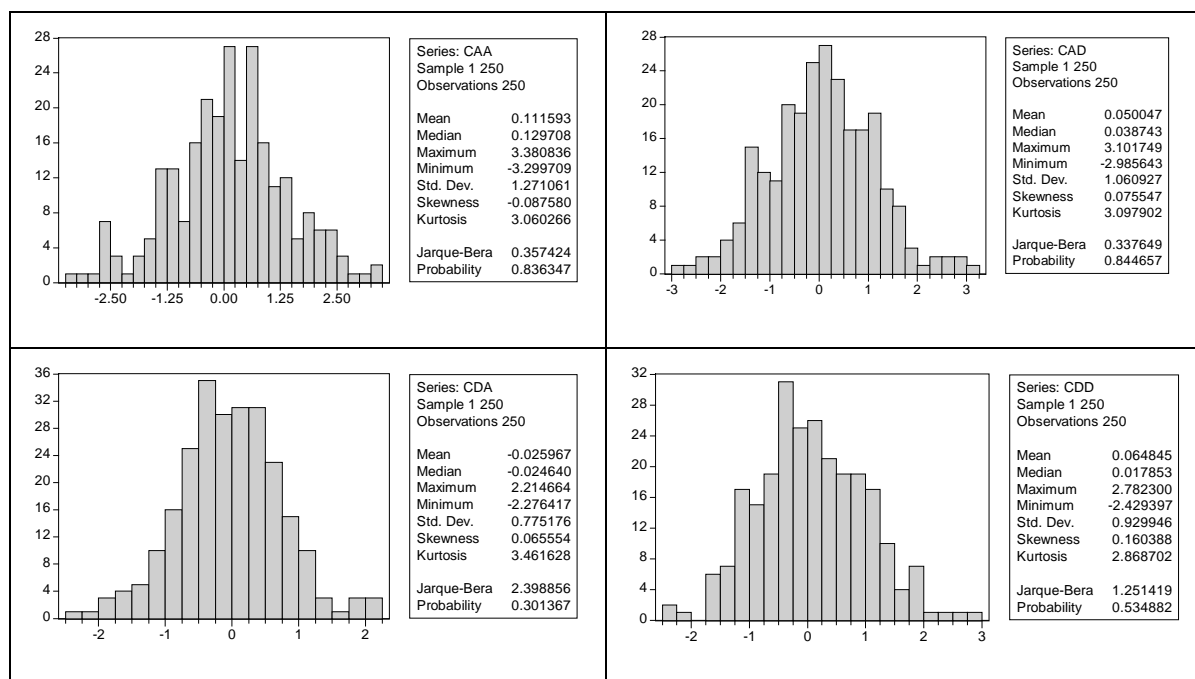


Figura 27: Distribuições das subséries em 2º nível do modelo ARIMA-GARCH

Não se rejeita a hipótese de normalidade pelo teste de Jarque-Bera e nota-se um excesso de curtose, isto é, os dados seguem uma distribuição normal e há provável indicação da necessidade de um modelo não-linear.

Teste de McLeod-Li

Subsérie: CAA				Subsérie: CAD			
Estatística F	9,37709	p-valor	0,0457191	Estatística F	10,58417	p-valor	0,0384177
Subsérie: CDA				Subsérie: CDD			
Estatística F	17,43084	p-valor	0,0125519	Estatística F	7,70173	p-valor	0,0798300

Para 95% de confiança, o teste de McLeod-Li indica a necessidade de um modelo não-linear para representar os dados.

Com base no algoritmo de Billings e Voon (1986 apud FREITAS, SECURATO e NETO, 2004), tem-se:

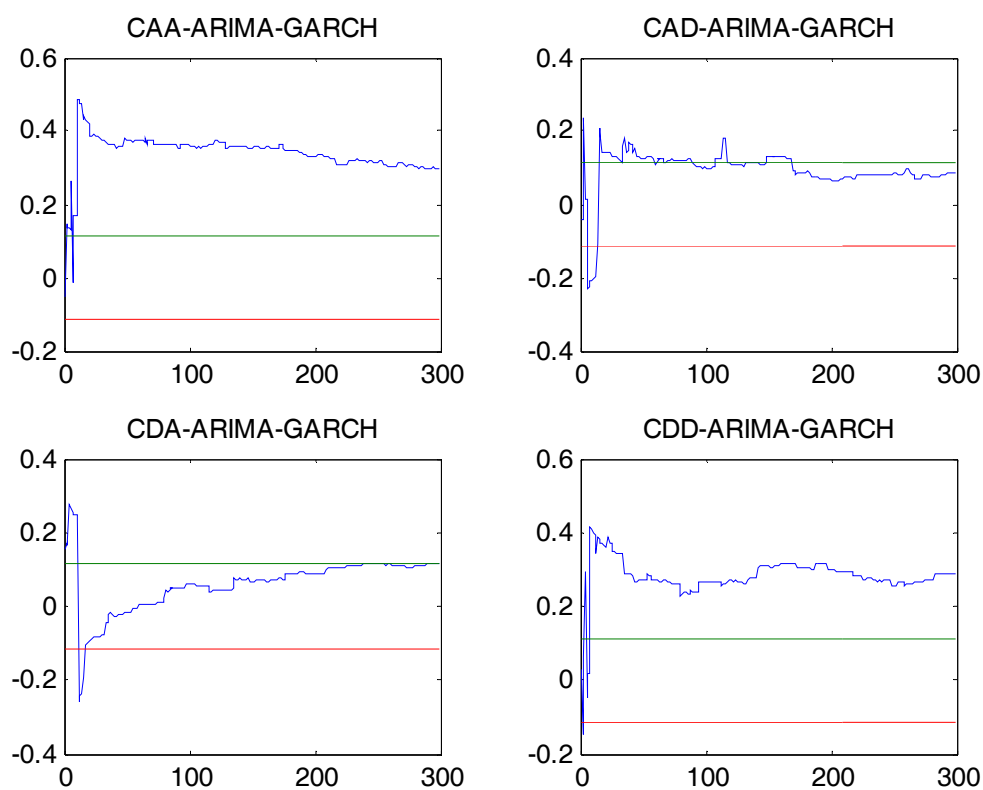


Figura 28: Comportamento da série, ao longo do tempo, para o modelo ARIMA-GARCH em 2 níveis de pré-processamento.

Observa-se, na figura anterior, que, na segunda decomposição, a parte de baixa frequência e a parte de alta frequência da primeira subdivisão tiveram comportamento não-linear, ao longo do tempo, enquanto que as demais apresentaram oscilações, mas com maior comportamento linear.

Teste de Hsieh

Subsérie: CAA			Subsérie: CAD		
Coeficiente	Estatística de Hsieh	p-valor	Coeficiente	Estatística de Hsieh	p-valor
1 1	0,4698	0,68075	1 1	-0,4292	0,33389
1 2	-1,1475	0,12559	1 2	1,8214	0,96573
1 3	2,8339	0,9977	1 3	-2,0257	0,0214*
1 4	0,1524	0,56056	1 4	0,1589	0,56313
1 5	-0,8364	0,20146	1 5	-0,9621	0,168
2 2	-0,2085	0,41742	2 2	0,4216	0,66334
2 3	0,2751	0,60838	2 3	1,3729	0,91511
2 4	0,4183	0,66214	2 4	-1,4502	0,0735
2 5	-2,4288	0,00757*	2 5	0,6314	0,73611
3 3	-0,6885	0,24557	3 3	-0,3166	0,37577
3 4	-0,0774	0,46915	3 4	-0,6647	0,25312
3 5	1,8667	0,96903	3 5	0,2207	0,58734
4 4	0,3276	0,62839	4 4	-0,7103	0,23876
4 5	1,1642	0,87783	4 5	1,1564	0,87624
5 5	-0,568	0,28502	5 5	-0,3742	0,35413

Subsérie: CDA			Subsérie: CDD		
Coeficiente	Estatística de Hsieh	p-valor	Coeficiente	Estatística de Hsieh	p-valor
1 1	0,3661	0,64285	1 1	0,0159	0,50634
1 2	0,4026	0,65638	1 2	-0,3975	0,3455
1 3	2,02	0,97831	1 3	-0,3184	0,37509
1 4	-0,4836	0,31433	1 4	-0,39	0,34827
1 5	-0,361	0,35905	1 5	-1,0603	0,1445
2 2	0,278	0,60949	2 2	0,3411	0,63349
2 3	-0,8093	0,20917	2 3	1,7178	0,95708
2 4	0,3801	0,64806	2 4	-0,549	0,2915
2 5	2,193	0,98585	2 5	-1,6716	0,0473*
3 3	0,2333	0,59224	3 3	-0,1566	0,43778
3 4	-1,6869	0,04581*	3 4	1,1567	0,8763
3 5	1,2024	0,8854	3 5	-1,1896	0,1171
4 4	-0,0308	0,48771	4 4	-0,479	0,31597
4 5	0,2495	0,59851	4 5	1,575	0,94237
5 5	0,2613	0,60307	5 5	-0,1297	0,4484

Conforme se observa, na tabela anterior, para quinze pares i,j , os valores dos coeficientes do teste de Hsieh amostrais para as quatro subséries e o correspondente p-valor para um nível de confiança de 95%, apresenta apenas um coeficiente significativo (indicado com *) em cada uma delas, podendo-se dizer que, de maneira geral, não se rejeita a hipótese nula dos coeficientes iguais a zero, o que indica uma não-linearidade na variância condicionada.

Estimando a equação para a subsérie CAA, tem-se:

$$caa_t = \phi_1 caa_{t-1} + \phi_2 caa_{t-2} + \varepsilon_t$$

$$\varepsilon_t^2 | I_{t-1} \sim N(0, h_t)$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$$

Parâmetros Subsérie CAA	Coeficiente	Erro Padrão	Estatística t	Significância
ϕ_1	-1,319948	0,052072	-25,34834	0,0000
ϕ_2	-0,866083	0,051420	-16,84337	0,0000
α_0	0,130362	0,053318	2,445007	0,0145
α_1	-0,161613	0,040130	-4,027200	0,0001
β_1	0,626123	0,208442	3,003817	0,0027
Modelo	Log de Verossimilhança		AIC	BIC
AR(2)-GARCH(1,1)	-170,6000		1,432258	1,531427

Para a subsérie CAD:

$$cad_t = \phi_1 cad_{t-1} + \phi_2 cad_{t-2} + \varepsilon_t$$

$$\varepsilon_t^2 | I_{t-1} \sim N(0, h_t)$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$$

Parâmetros Subsérie CAD	Coeficiente	Erro Padrão	Estatística t	Significância
ϕ_1	1,034446	0,070684	14,63470	0,0000
ϕ_2	-0,692425	0,051489	-13,44811	0,0000
α_0	1,109032	0,954899	1,161413	0,0245
α_1	-0,106960	0,030310	-3,528851	0,0004
β_1	0,040028	0,904893	0,044235	0,0464
Modelo	Log de Verossimilhança		AIC	BIC
ARMA(2,2)-GARCH(1,1)	-166,5539		1,383499	1,454334

Estimando a equação para a subsérie CDA, tem-se:

$$cda_t = \phi_1 cda_{t-1} + \phi_2 cda_{t-2} + \varepsilon_t$$

$$\varepsilon_t^2 | I_{t-1} \sim N(0, h_t)$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$$

Parâmetros Subsérie CDA	Coeficiente	Erro Padrão	Estatística t	Significância
ϕ_1	0,877888	0,103388	8,491157	0,0000
ϕ_2	-0,792456	0,060697	-13,05591	0,0000
α_0	-0,955782	0,078134	-12,23262	0,0000
α_1	0,224620	0,063608	3,531325	0,0004
β_1	0,199012	0,098891	2,012431	0,0442
Modelo	Log de Verossimilhança		AIC	BIC
AR(2)-GARCH(1,1)	-148,9386		1,257569	1,356739

Estimando a equação para a subsérie CDD, tem-se:

$$cdd_t = \phi_1 cdd_{t-1} + \phi_2 cdd_{t-2} + \varepsilon_t$$

$$\varepsilon_t^2 | I_{t-1} \sim N(0, h_t)$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$$

Parâmetros Subsérie CDD	Coeficiente	Erro Padrão	Estatística t	Significância
ϕ_1	0,071763	1,727515	0,041541	0,0000
ϕ_2	-0,011981	1,988571	-0,006025	0,0000
α_0	0,071700	0,033239	2,157094	0,0310
α_1	-0,125130	0,011606	-10,78123	0,0000
β_1	0,766429	0,160036	4,789103	0,0000
Modelo	Log de Verossimilhança		AIC	BIC
AR(2)-GARCH(1,1)	-145,8158		1,232385	1,331555

As redes recorrentes foram geradas com a arquitetura 10-2R-1. Estimando-se os resultados por redes neurais com retropropagação, tem-se, a seguir, um quadro resumindo o tipo de rede utilizado e as respectivas performances de seleção, teste e validação. Por exemplo, o modelo neural para a subsérie CAA apresentou uma rede MLP (*Multilayer Perceptron*) - MLP cuja arquitetura, identificação e seleção indicou como a melhor rede, a MLP 1:5-8-1:1, com um neurônio na camada de entrada, nas camadas ocultas cinco e oito, e um neurônio na camada de saída:

Subséries	Rede	Desempenho de treinam.	Performance Seleção	Performance Teste	Erro Treinamento	Erro Seleção	Erro Teste	Características
CAA	MLP s5 1:5-8-1:1	0,999861	0,996576	0,994204	0,161463	0,175151	0,152884	BP100,CG4b
CAD	MLP s5 1:5-8-1:1	0,999861	0,996576	0,994204	0,161463	0,175151	0,152884	BP100,CG4b
CDA	MLP s5 1:5-2-1:1	0,972494	0,993032	1,028907	0,201680	0,209264	0,204046	BP100,CG21b
CDD	MLP s5 1:5-8-1:1	0,981154	1,002502	0,990287	0,217439	0,242021	0,249169	BP67b

Estimando-se os valores e calculando as estimativas pontuais de erros para a previsão em 2 níveis de pré-processamento, tem-se:

MODELO ARMA-GARCH – Modelo ARIMA(2,0,0)-GARCH(1,1)			
Estatísticas de Previsão Pontual para os 200 pontos	Valor	Estatísticas de Previsão Pontual para os 10 primeiros pontos	Valor
TIC	0,99	TIC	0,96
MAPE	99,99%	MAPE	11,37%
Correlação	-0,07	Correlação	-0,47

MODELO RN-BACKPROPAGATION - Modelo ARIMA(2,0,0)-GARCH(1,1)			
Estatísticas de Previsão Pontual para os 200 pontos	Valor	Estatísticas de Previsão Pontual para os 10 primeiros pontos	Valor
TIC	0,99	TIC	0,86
MAPE	98,46%	MAPE	13,05%
Correlação	-0,11	Correlação	-0,28

MODELO RN RECORRENTE - Modelo ARIMA(2,0,0)-GARCH(1,1)			
Estatísticas de Previsão Pontual para os 200 pontos	Valor	Estatísticas de Previsão Pontual para os 10 primeiros pontos	Valor
TIC	0,69	TIC	0,14
MAPE	89,40%	MAPE	5,80%
Correlação	0,37	Correlação	0,96

Fazendo o cálculo do envelopamento para 200 séries simuladas e verificando quais estão completamente contidas dentro do intervalo de confiança de 95% para a série prevista via procedimento estático, tem-se:

Envelopamento – Modelo	Percentual
Rede Neural Recorrente	93,62%
Rede Neural Retropropagação	91,59%
ARIMA-GARCH	48,15%

A seguir, tem-se os resultados das previsões feitas e após a etapa de pós-processamento.

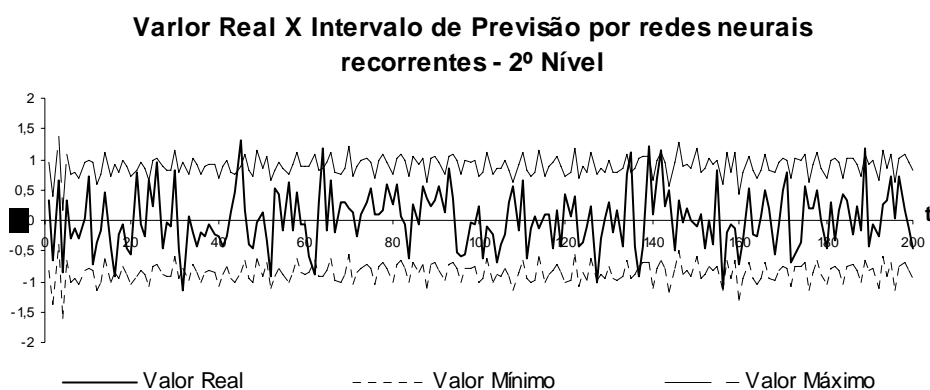


Figura 29: Intervalo de previsão em 2º nível para redes recorrentes para a amostra de teste do modelo ARIMA-GARCH

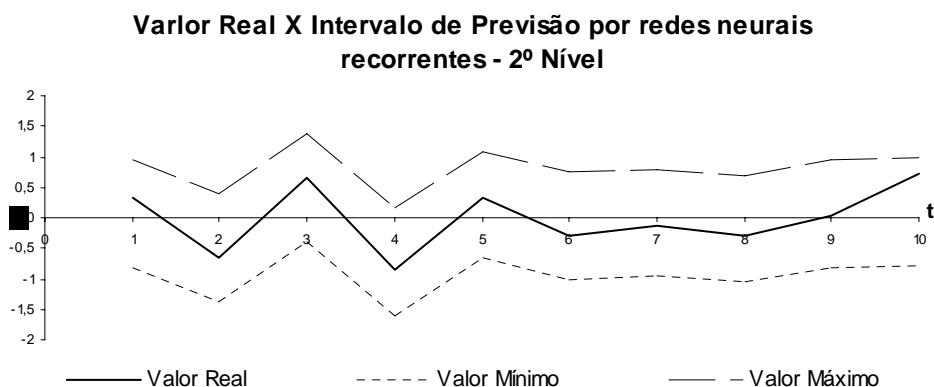


Figura 30: Intervalo de previsão em 2º nível para redes recorrentes para os primeiros 10 elementos do modelo ARIMA-GARCH.

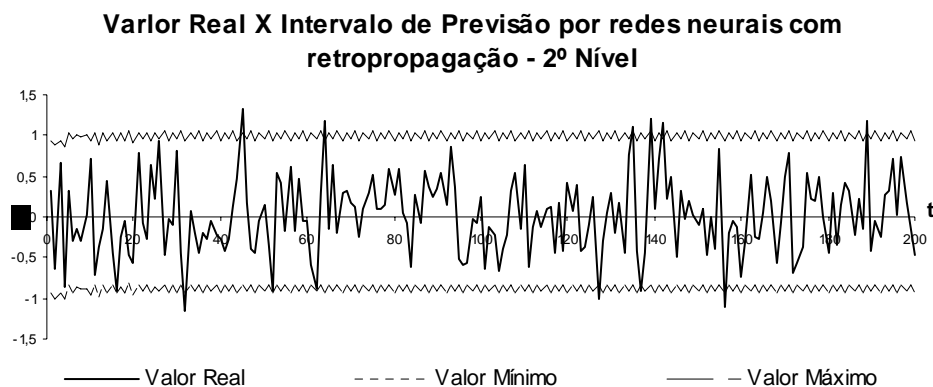


Figura 31: Intervalo de previsão em 2º nível feito por redes neurais com retropropagação do modelo ARIMA-GARCH

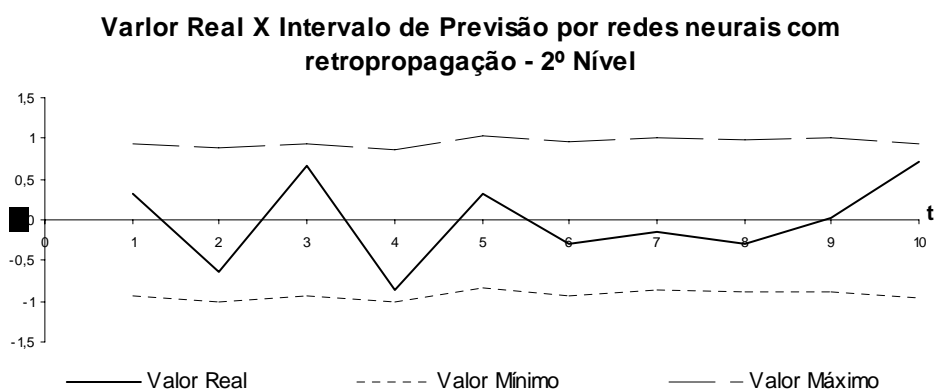


Figura 32: Intervalo de previsão em 2º nível para redes neurais com retropropagação para os primeiros 10 elementos do modelo ARIMA-GARCH.

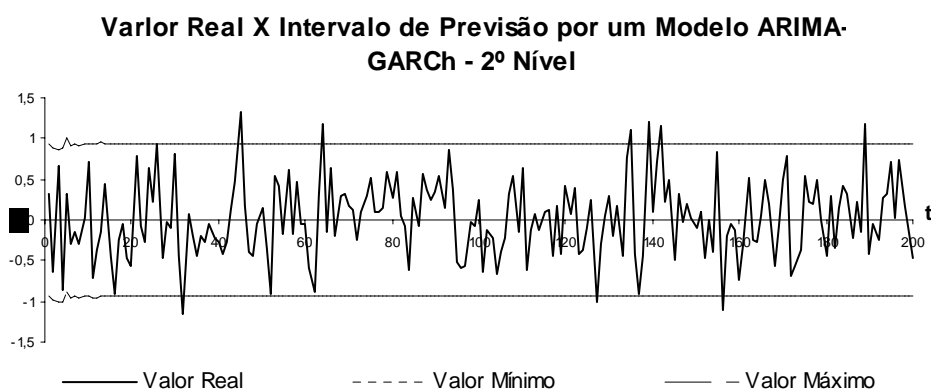


Figura 33: Intervalo de previsão em 2º nível feito por um modelo ARIMA-GARCH

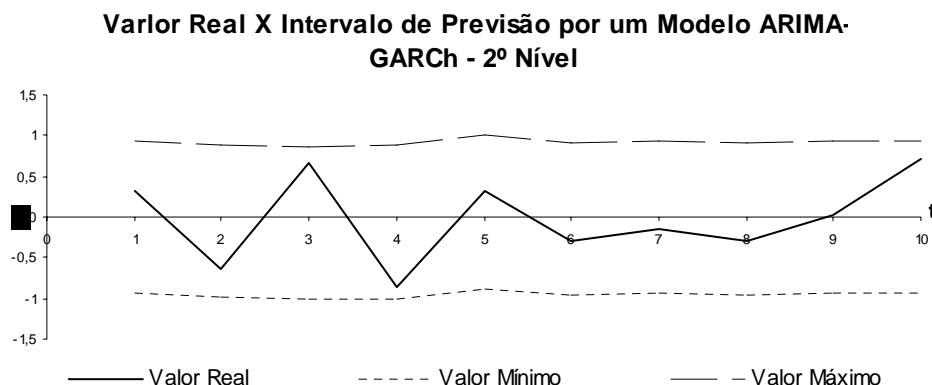


Figura 34: Intervalo de previsão em 2º nível por um modelo ARIMA-GARCH para os primeiros 10 elementos.

Nos gráficos anteriores, o valor real corresponde ao valor no nível do processo simulado. O valor mínimo é o termo inferior da banda de previsão média do envelopamento e o valor máximo é o termo superior da banda.

Teste para Independência Estatística dos Resíduos– Teste BDS

Como visto anteriormente, o teste para independência deve ser aplicado para ver se a sucessão cronológica apresenta independência nos resíduos. Suas hipóteses são:

H_0 : os resíduos são estocasticamente independentes

H_1 : os resíduos são estocasticamente dependentes

Resíduos Modelo ARIMA-GARCH – Previsto pelo GARCH com 1 Nível de pré-processamento					Resíduos Modelo ARIMA - GARCH – Previsto pelo GARCH com 2 Níveis de pré-processamento				
Dimensão	BDS	Erro Padrão	Estatística Z	p-valor	Dimensão	BDS	Erro Padrão	Estatística Z	p-valor
2	-0,000944	0,004561	-0,206852	0,8361	2	-0,000648	0,004541	-0,142673	0,8865
3	0,003871	0,007238	0,534773	0,5928	3	0,003913	0,007202	0,543379	0,5869
4	0,006691	0,008603	0,777761	0,4367	4	0,006235	0,008557	0,728710	0,4662
5	0,006365	0,008950	0,711227	0,4769	5	0,006221	0,008897	0,699292	0,4844
6	0,006514	0,008613	0,756320	0,4495	6	0,006410	0,008558	0,748946	0,4539
Resíduos Modelo ARIMA-GARCH – Previsto RN com retrop. com 1 Nível de pré-processamento					Resíduos Modelo ARIMA-GARCH – Previsto RN com retrop. com 2 Níveis de pré-processamento				
Dimensão	BDS	Erro Padrão	Estatística Z	p-valor	Dimensão	BDS	Erro Padrão	Estatística Z	p-valor
2	0,103531	0,003402	30,43279	0,0000	2	-0,000859	0,004344	-0,197658	0,8433
3	0,143677	0,005346	26,87661	0,0000	3	0,002777	0,006920	0,401302	0,6882
4	0,240527	0,006291	38,23391	0,0000	4	0,005129	0,008258	0,621073	0,5346
5	0,306458	0,006478	47,30565	0,0000	5	0,005429	0,008623	0,629619	0,5289
6	0,355816	0,006172	57,65411	0,0000	6	0,006309	0,008331	0,757276	0,4489

Resíduos Modelo ARIMA-GARCH – Previsto RN recorrentes com 1 Nível de pré-processamento					Resíduos Modelo ARIMA-GARCH – Previsto RN recorrentes com 2 Níveis de pré-processamento				
Dimensão	BDS	Erro Padrão	Estatística Z	p-valor	Dimensão	BDS	Erro Padrão	Estatística Z	p-valor
2	1,90E-05	8,50E-06	2,231815	0,0256	2	-0,000614	0,000965	-0,636124	0,5247
3	-7,48E-07	2,79E-07	-2,675869	0,0075	3	0,000890	0,002063	0,431336	0,6662
4	-6,75E-09	6,91E-09	-0,977534	0,3283	4	0,002581	0,003301	0,781933	0,4343
5	-5,91E-11	1,50E-10	-0,394381	0,6933	5	0,002600	0,004621	0,562700	0,5736
6	-5,09E-13	3,01E-12	-0,168993	0,8658	6	0,003773	0,005985	0,630418	0,5284

O software Eviews®-5.0 calcula a estatística BDS para todas as dimensões de dois até o valor especificado variando a dimensão de correlação para maximizar o poder do teste. Neste caso, calculou-se até a sexta dimensão e optou-se por não se rejeitar a hipótese nula ao nível de significância de 5%, indicando que os dados dos resíduos padronizados do modelo ARIMA-GARCH são independentes para os modelos gerados baseados na análise do p-valor, com exceção do primeiro nível de pré-processamento na rede neural com retropropagação.

Teste de Normalidade

Na figura, a seguir, tem-se as estatísticas dos resíduos do pós-processamento onde a hipótese nula e alternativa são:

H_0 : os resíduos seguem uma distribuição normal

H_1 : os resíduos não seguem uma distribuição normal

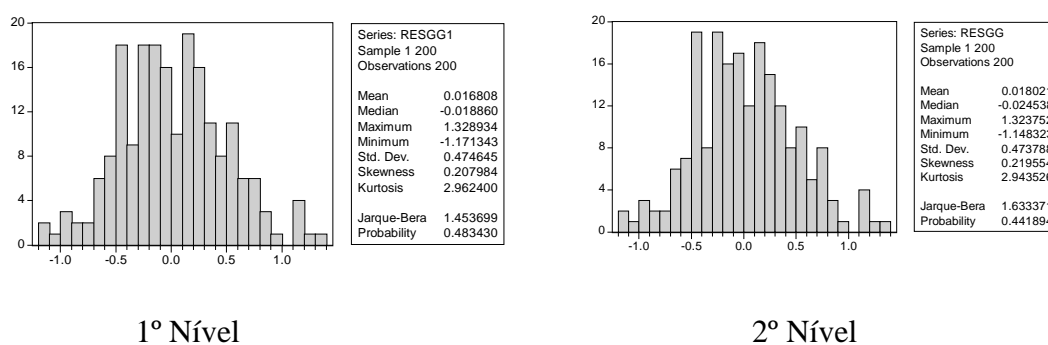
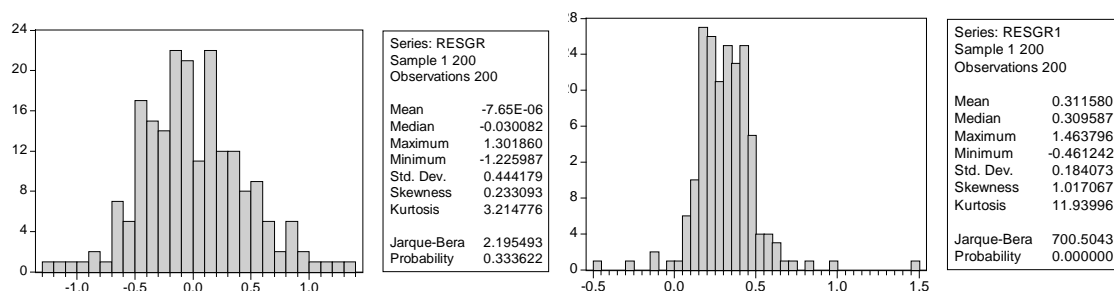
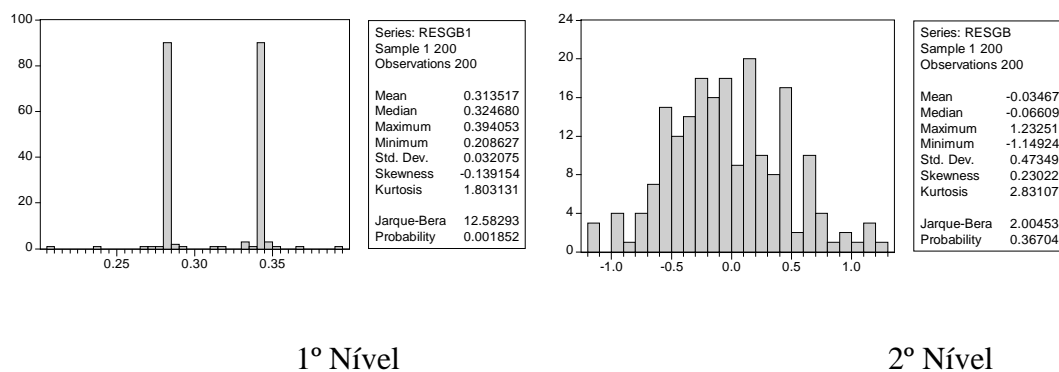


Figura 35: Distribuição dos resíduos do modelo ARIMA-GARCH previsto por um modelo ARIMA-GARCH, em 1 e 2 níveis de pré-processamento.



Observa-se que os dados dos resíduos são normais pela estatística de Jarque-Bera, com exceção no primeiro nível de pré-processamento das redes neurais com retropropagação, e no segundo nível das redes neurais recorrentes.

4.2 SIMULAÇÃO DE UM PROCESSO ARIMA

Considere agora um DGP para o modelo ARIMA, da forma ARIMA(1,0,0), com o seguinte parâmetro: $\phi_1 = 0,3$.

Utilizando-se o programa 5, disponível no Apêndice C, foram geradas 1200 observações para o modelo, sendo divididas em duas partes: a primeira com mil observações foi utilizada para gerar o modelo para previsão e as 200 restantes utilizadas para validar o modelo e seu gráfico com 1000 pontos seria:

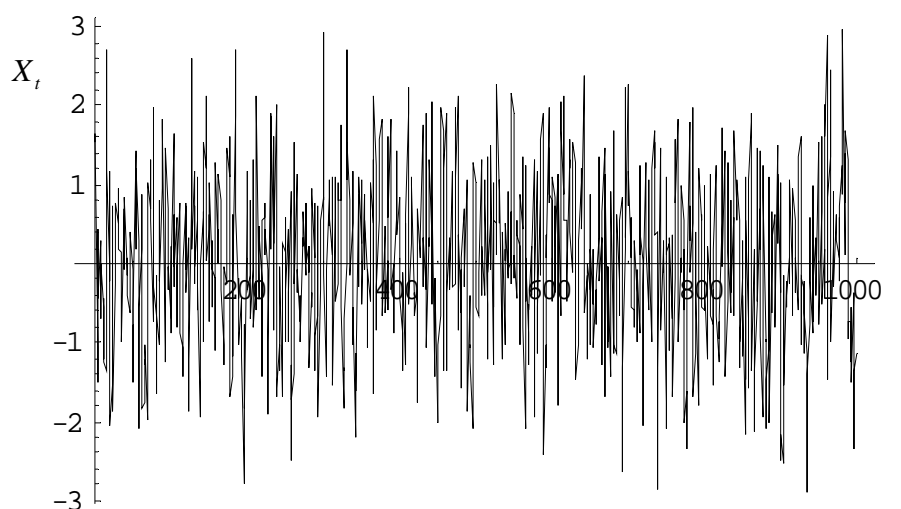


Figura 38: DGP para um modelo AR(1)

Verificando a sucessão cronológica como um todo, tem-se:

Teste de raiz unitária

Modelo AR(1)		
Hipóteses	Estatística de Teste	Valores Críticos a 5%
$H_0 : 3RU$ $H_1 : 2RU$	-61,28628	-2,76
$H_0 : 2RU$ $H_1 : 1RU$	-25,32047 -31,97124	-3,37 -3,77
$H_0 : 1RU$ $H_1 : 0RU$	-32,93413 -14,96314 -16,43593	-3,42 -3,80 -4,16

Considerando um nível de confiança de 95%, se rejeita a todas as hipóteses nulas, o que indica que ambas as subséries são estacionárias.

Teste de independência estatística – BDS

Modelo AR(1)				
Dimensão	BDS	Erro Padrão	Estatística Z	p-valor
2	-9,45E-06	1,62E-06	-5,826504	0,0000
3	-9,97E-07	4,06E-08	-24,53405	0,0000
4	-1,00E-08	7,65E-10	-13,11928	0,0000
5	-1,00E-10	1,26E-11	-7,930219	0,0000
6	-1,01E-12	1,93E-13	-5,243348	0,0000

Calculou-se até a sexta dimensão e concluiu-se por rejeitar a hipótese nula ao nível de significância de 5%, indicando que os dados da sucessão cronológica são dependentes, baseado na análise do p-valor, sendo que pelo menos um dos p-valores deve ser menor que 5%.

Teste de Normalidade

A figura, a seguir, ilustra a distribuição da sucessão cronológica.

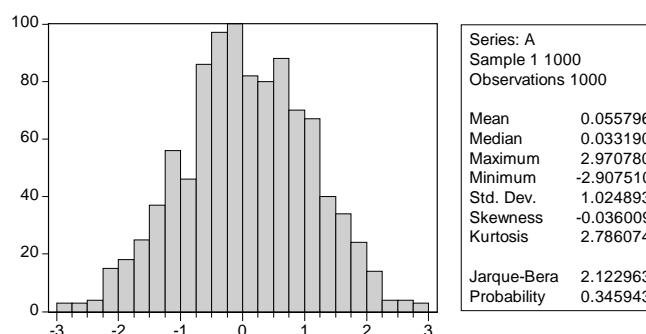


Figura 39: Distribuições do modelo AR(1) simulado

Teste de McLeod-Li

O resultado da estatística de teste calculado no Eviews® 5.0 é:

Modelo AR(1)			
Estatística F	1,388241	p-valor	0,229348

Dessa forma, o teste de McLeod-Li para 5 defasagens não se rejeita a hipótese nula pelo fato do p-valor ser inferior a 5%. Isto indica que o modelo apresenta um comportamento linear

E seguindo o organograma, foi escolhida novamente a ondaleta de Daubesch número 1 (DB1). A figura, a seguir, ilustra a decomposição (pré-processamento), em primeiro nível, da série no nível original feita com o auxílio do programa 2 do Apêndice C:

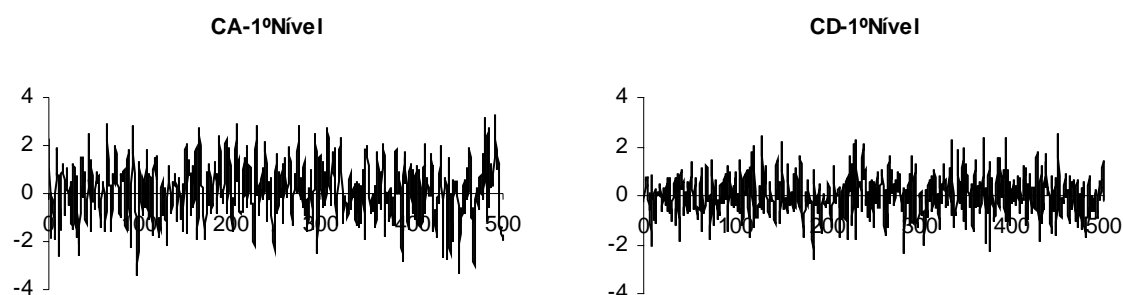


Figura 40: Pré-processamento em 1º nível para o DGP do modelo AR(1)

Teste de raízes unitária

Subsérie: CA			Subsérie: CD		
Hipóteses	Estatística de Teste	Valores Críticos a 5%	Hipóteses	Estatística de Teste	Valores Críticos a 5%
$H_0 : 3RU$ $H_1 : 2RU$	-36,79883	-2,76	$H_0 : 3RU$ $H_1 : 2RU$	-50,94641	-2,76
$H_0 : 2RU$ $H_1 : 1RU$	-15,34845 -20,82436	-3,37 -3,77	$H_0 : 2RU$ $H_1 : 1RU$	-14,25024 -24,79960	-3,37 -3,77
$H_0 : 1RU$ $H_1 : 0RU$	-22,75377 -6,216736 -13,30905	-3,42 -3,80 -4,16	$H_0 : 1RU$ $H_1 : 0RU$	-20,90149 -7,698684 -12,44013	-3,42 -3,80 -4,16

Considerando um nível de confiança de 95%, se rejeita a todas as hipóteses nulas, o que indica que ambas as subséries são estacionárias.

Teste de independência estatística - BDS

Subsérie: CA					Subsérie: CD				
<u>Dimensão</u>	<u>BDS</u>	<u>Erro Padrão</u>	<u>Estatística Z</u>	<u>p-valor</u>	<u>Dimensão</u>	<u>BDS</u>	<u>Erro Padrão</u>	<u>Estatística Z</u>	<u>p-valor</u>
2	-3,00E-06	4,85E-05	-0,061797	0,9507	2	4,77E-05	3,97E-05	1,199797	0,2302
3	2,87E-05	5,75E-06	4,989138	0,0000	3	-1,82E-06	4,68E-06	-0,389464	0,6969
4	1,81E-05	5,12E-07	35,39495	0,0000	4	-6,13E-06	4,14E-07	-14,81243	0,0000
5	7,84E-06	3,99E-08	196,4425	0,0000	5	-3,04E-07	3,20E-08	-9,493997	0,0000
6	-1,52E-08	2,88E-09	-5,280915	0,0000	6	-1,49E-08	2,29E-09	-6,514062	0,0000

O teste BDS rejeita a hipótese de independência das subséries; todos com 95% de confiança, já que pelo menos um dos valores é menor que 5%.

Teste de normalidade

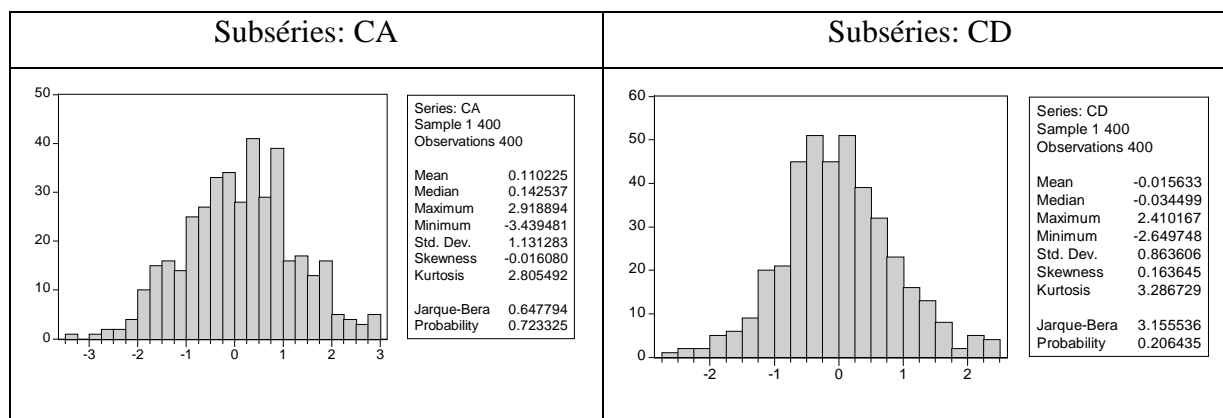


Figura 41: Distribuição das subséries para o modelo AR(1)

Não se rejeita a hipótese de normalidade pelo teste de Jarque-Bera, com 95% de confiança para as subséries CA e CD.

Teste de McLeod-Li

Subsérie: CA				Subsérie: CD			
Estatística F	0,949063	p-valor	0,448828	Estatística F	0,951046	p-valor	0,447565

Para 95% de confiança, o teste de McLeod-Li aceita-se a hipótese de linearidade do modelo simulado, isto é, há evidências de que um modelo linear é melhor ajustado às subséries.

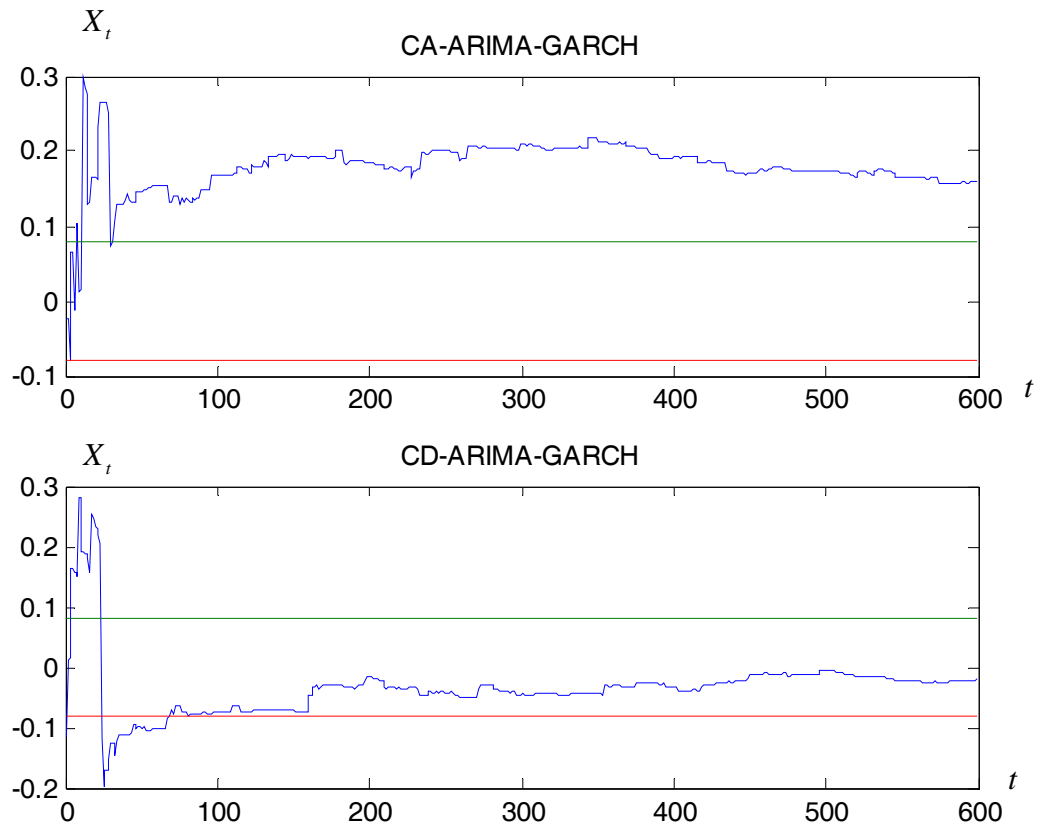


Figura 42: Comportamento da série ao longo do tempo para o modelo ARIMA em 1 nível de pré-processamento.

Observa-se que a subsérie de baixa frequência apresenta comportamento não-linear, enquanto que a subsérie de alta frequência tem comportamento linear ao longo da trajetória.

Verificando-se o correlograma da série simulada no pré-processamento, tem-se:

CA							CD												
Modelo	ARIMA(1,0,0)						Modelo	ARIMA(3,0,0)											
FAC e FACP	Autocorrelation		Partial Correlation		AC	PAC	Q-Stat	Prob	FAC e FACP	Autocorrelation		Partial Correlation		AC	PAC	Q-Stat	Prob		
					1	0.218	0.218	19.163		0.000					1	-0.171	-0.171	11.757	0.001
					2	-0.108	-0.164	23.899		0.000					2	0.061	0.033	13.263	0.001
					3	-0.191	-0.138	38.675		0.000					3	-0.066	-0.052	15.011	0.002
					4	-0.052	0.009	39.784		0.000					4	0.031	0.011	15.411	0.004
					5	-0.006	-0.039	39.800		0.000					5	0.012	0.025	15.472	0.009
					6	0.067	0.048	41.618		0.000					6	-0.008	-0.007	15.496	0.017
					7	0.085	0.056	44.606		0.000					7	0.026	0.026	15.777	0.027
					8	0.043	0.018	45.379		0.000					8	-0.077	-0.069	18.196	0.020
					9	-0.025	-0.008	45.634		0.000					9	0.076	0.051	20.579	0.015
					10	-0.076	-0.044	48.016		0.000					10	0.049	0.080	21.565	0.017
					11	0.033	0.076	48.472		0.000					11	-0.068	-0.064	23.454	0.015
					12	0.042	0.001	49.188		0.000					12	0.034	0.019	23.935	0.021
					13	-0.048	-0.078	50.126		0.000					13	-0.060	-0.042	25.417	0.020
					14	-0.020	0.026	50.285		0.000					14	-0.067	-0.104	27.275	0.018
					15	0.038	0.031	50.902		0.000					15	0.072	0.063	29.467	0.014
					16	0.049	0.023	51.926		0.000					16	0.069	0.089	31.434	0.012
					17	-0.029	-0.038	52.276		0.000					17	-0.024	-0.008	31.679	0.016
					18	-0.038	-0.013	52.893		0.000					18	-0.022	-0.014	31.885	0.023
					19	0.022	0.042	53.092		0.000					19	0.114	0.106	37.413	0.007
					20	-0.038	-0.072	53.698		0.000					20	-0.008	0.030	37.442	0.010
															21	0.014	0.005	37.530	0.015
															22	0.057	0.067	38.896	0.015
															23	-0.098	-0.063	42.991	0.007
															24	-0.033	-0.073	43.454	0.009

Figura 43: Funções FAC e FACP para o DGP do modelo AR(1)

Parâmetros Subséries CA	Coefficiente	Erro Padrão	Estatística t	Significância
ϕ_1	0,225422	0,048613	4,637076	0,0000
Modelo	Log de Verossimilhança		AIC	BIC
AR(1)	-604,7926		3,036554	3,046552

Parâmetros Subséries CD	Coefficiente	Erro Padrão	Estatística t	Significância
ϕ_1	-0,170420	0,049392	-3,450396	0,0006
Modelo	Log de Verossimilhança		AIC	BIC
AR(1)	-501,8275		2,520439	2,530436

Estimando-se os valores e calculando as estimativas de erros pontuais para a previsão das subséries e fazendo o pós-processamento com o auxílio do programa 3, tem-se:

MODELO ARIMA - Modelo ARIMA(1,0,0)			
Estatísticas de Previsão Pontual para os 200 pontos	Valor	Estatísticas de Previsão Pontual para os 10 primeiros pontos	Valor
TIC	1,00	TIC	0,90
MAPE	100,00%	MAPE	10,90%
Correlação	-0,05	Correlação	-0,34

MODELO ARIMA – Rede Neural Recorrente			
Estatísticas de Previsão Pontual para os 200 pontos	Valor	Estatísticas de Previsão Pontual para os 10 primeiros pontos	Valor
TIC	0,99	TIC	0,64
MAPE	100,00%	MAPE	8,76%
Correlação	0,46	Correlação	0,47

MODELO ARIMA – Rede Neural Retropropagação			
Estatísticas de Previsão Pontual para os 200 pontos	Valor	Estatísticas de Previsão Pontual para os 10 primeiros pontos	Valor
TIC	0,99	TIC	0,86
MAPE	110,11%	MAPE	10,24%
Correlação	-0,03	Correlação	-0,12

Fazendo o cálculo do envelopamento para 200 séries simuladas e verificando quais se enquadram dentro do intervalo de confiança de 95%, para a série prevista, via procedimento estático, tem-se:

Envelopamento – modelo	Percentual
Rede Neural Recorrente	93,50%
Rede Neural Retropropagação	91,00%
Modelo ARIMA	92,67%

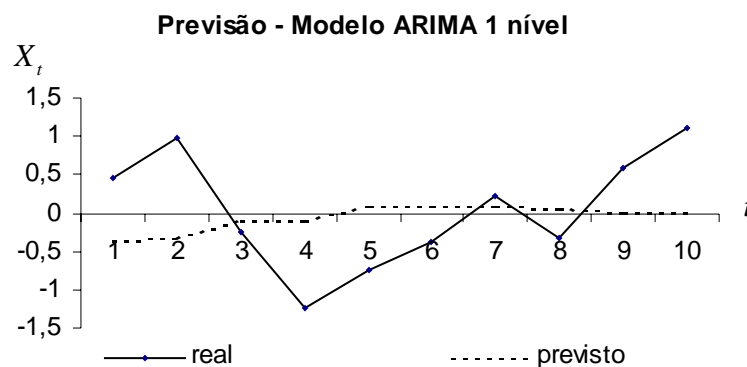


Figura 44: Previsão em 1º nível para o modelo AR(1) por para os primeiros 10 elementos.

Simulação no 2º Nível do Pré-Processamento

Apresenta-se, a seguir, o gráfico da decomposição em 2 níveis do DGP para o modelo ARIMA com o auxílio do programa 2, sobre as duas primeiras subséries anteriores:

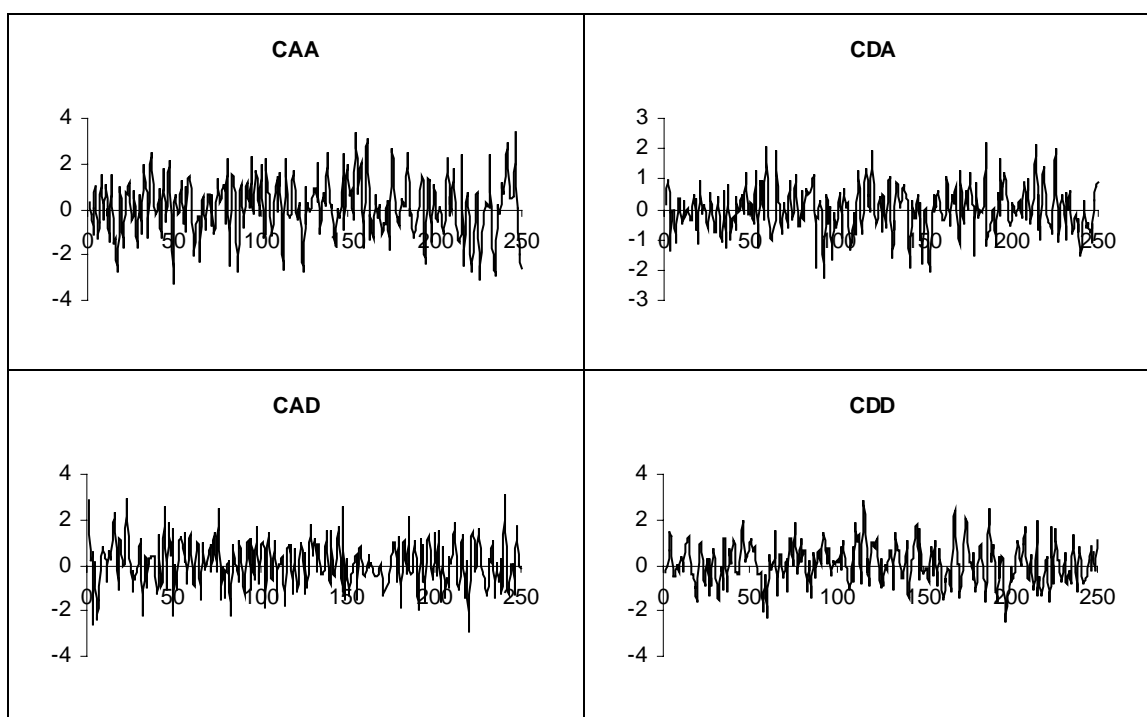


Figura 45: Pré-processamento em 2º nível para o DGP do modelo AR(1)

A partir do mesmo roteiro do modelo anterior, apresenta-se, a seguir, os resultados obtidos na tabela a seguir:

Teste de Raízes Unitárias

Subsérie: CAA			Subsérie: CAD		
Hipóteses	Estatística de Teste	Valores Críticos a 5%	Hipóteses	Estatística de Teste	Valores Críticos a 5%
$H_0 : 3RU$ $H_1 : 2RU$	-31,95246	-2,76	$H_0 : 3RU$ $H_1 : 2RU$	-37,73648	-2,76
$H_0 : 2RU$ $H_1 : 1RU$	-10,32123 -19,99126	-3,37 -3,77	$H_0 : 2RU$ $H_1 : 1RU$	-10,56451 -20,71348	-3,37 -3,77
$H_0 : 1RU$ $H_1 : 0RU$	-14,66594 -7,184838 -8,716525	-3,42 -3,80 -4,16	$H_0 : 1RU$ $H_1 : 0RU$	-15,96954 -6,150885 -9,942485	-3,42 -3,80 -4,16

Teste de Raízes Unitárias					
Subsérie: CDA			Subsérie: CDD		
Hipóteses	Estatística de Teste	Valores Críticos a 5%	Hipóteses	Estatística de Teste	Valores Críticos a 5%
$H_0 : 3RU$ $H_1 : 2RU$	-35,01270	-2,76	$H_0 : 3RU$ $H_1 : 2RU$	-34,24383	-2,76
$H_0 : 2RU$ $H_1 : 1RU$	-10,72136 -19,88246	-3,37 -3,77	$H_0 : 2RU$ $H_1 : 1RU$	-11,73372 -17,81364	-3,37 -3,77
$H_0 : 1RU$ $H_1 : 0RU$	-15,26169 -7,034603 -8,978980	-3,42 -3,80 -4,16	$H_0 : 1RU$ $H_1 : 0RU$	-15,87592 -7,299054 -8,577802	-3,42 -3,80 -4,16

Observa-se que, com 95% de confiança, ambas as quatro subséries definidas no pré-processamento são estacionárias.

Teste de independência estatística – BDS

Subsérie: CAA					Subsérie: CAD				
<u>Dimensão</u>	<u>BDS</u>	<u>Erro Padrão</u>	<u>Estatística Z</u>	<u>p-valor</u>	<u>Dimensão</u>	<u>BDS</u>	<u>Erro Padrão</u>	<u>Estatística Z</u>	<u>p-valor</u>
2	-0,000216	9,45E-05	-2,286002	0,0223	2	-0,000344	6,10E-05	-5,630532	0,0000
3	-4,09E-05	1,19E-05	-3,432417	0,0006	3	-3,99E-05	7,14E-06	-5,590338	0,0000
4	-7,11E-06	1,13E-06	-6,303094	0,0000	4	-5,08E-06	6,28E-07	-8,085309	0,0000
5	-3,77E-07	9,37E-08	-4,018874	0,0001	5	-2,41E-07	4,83E-08	-4,979506	0,0000
6	-2,04E-08	7,21E-09	-2,837173	0,0046	6	-1,14E-08	3,44E-09	-3,310096	0,0009

Subsérie: CDA					Subsérie: CDD				
Dimensão	BDS	Erro Padrão	Estatística Z	p-valor	Dimensão	BDS	Erro Padrão	Estatística Z	p-valor
2	-4,01E-05	8,12E-05	-0,494265	0,6211	2	0,000460	6,98E-05	6,586959	0,0000
3	-6,89E-05	1,02E-05	-6,777076	0,0000	3	7,44E-05	8,55E-06	8,698174	0,0000
4	-6,86E-06	9,55E-07	-7,186312	0,0000	4	-5,99E-06	7,87E-07	-7,611436	0,0000
5	-3,60E-07	7,86E-08	-4,577967	0,0000	5	-2,92E-07	6,34E-08	-4,595844	0,0000
6	-1,83E-08	6,00E-09	-3,059027	0,0022	6	-1,43E-08	4,74E-09	-3,024861	0,0025

O teste de independência não aceita a hipótese nula da independência das observações.

Teste de Normalidade

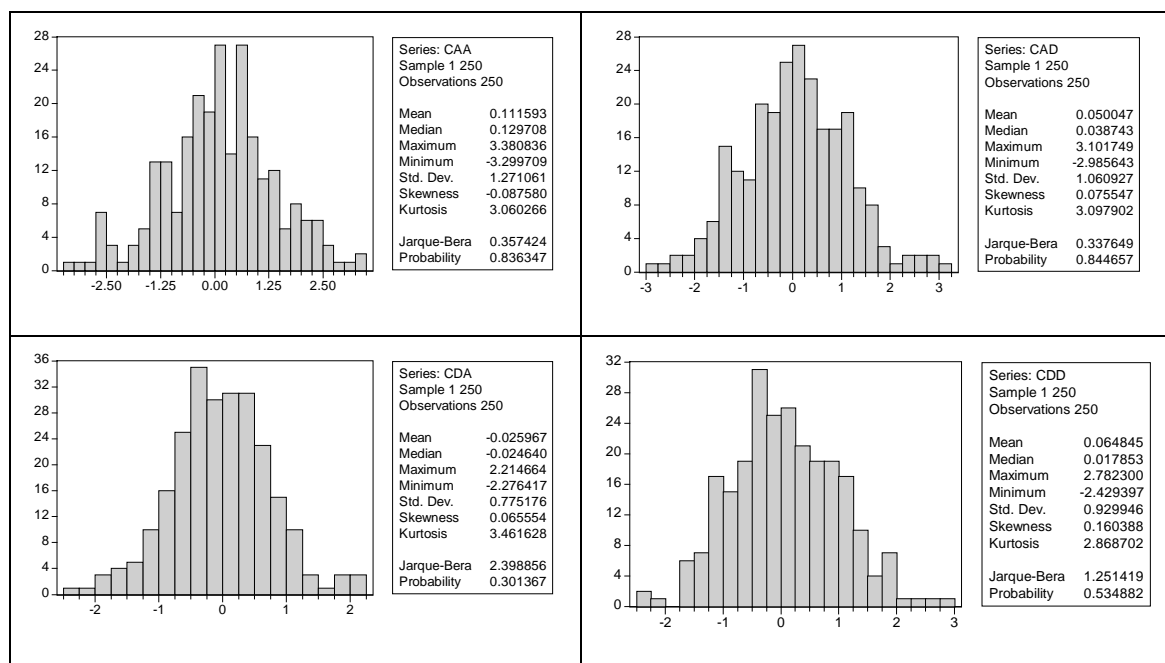


Figura 46: Distribuição em 2º nível para o pré-processamento do DGP do modelo AR(1)

O mesmo ocorre com o teste de normalidade dos dados, o que indica que todas as subséries apresentam comportamento estatisticamente normal.

Teste de McLeod-Li

Subsérie: CAA				Subsérie: CAD			
Estatística F	0,937709	p-valor	0,457191	Estatística F	1,058417	p-valor	0,384177
Subsérie: CDA				Subsérie: CDD			
Estatística F	1,743084	p-valor	0,125519	Estatística F	0,470173	p-valor	0,798300

Pelo teste, aceita-se a hipótese nula que indica que o modelo é linear.

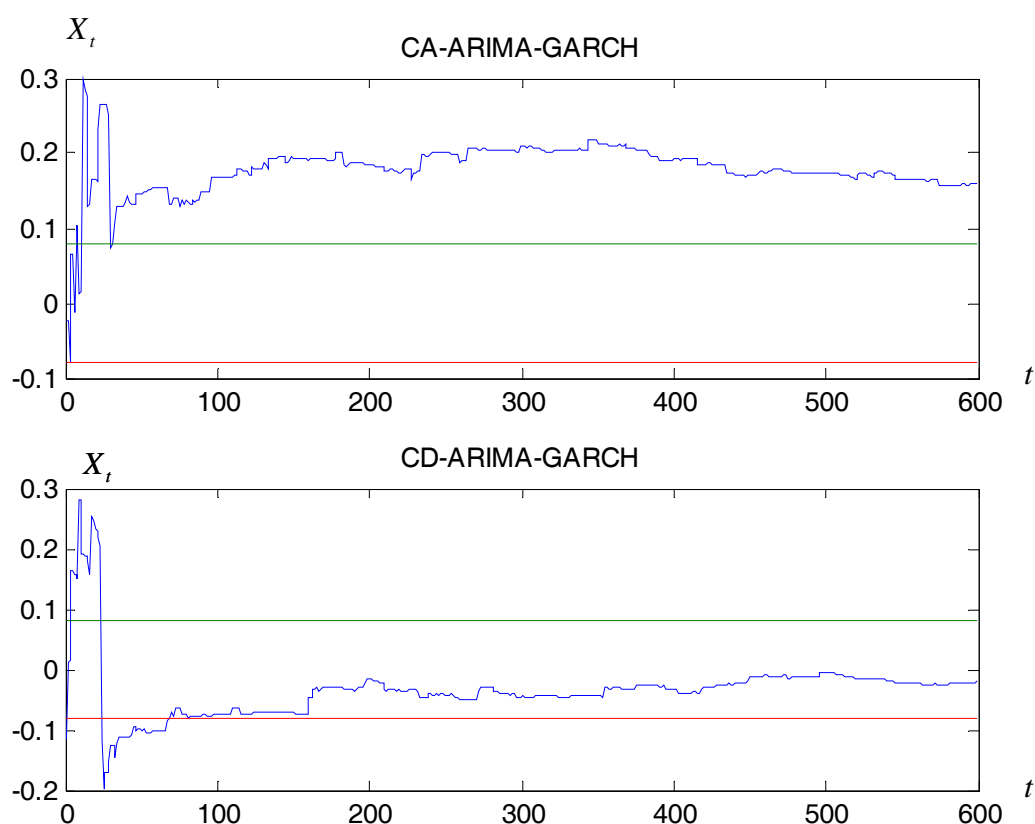


Figura 47: Comportamento da série ao longo do tempo para o modelo ARIMA em 2 níveis de pré-processamento.

A subsérie de baixa frequência apresenta comportamento não-linear, ao passo que a de alta frequência tem comportamento linear ao longo da trajetória.

Estimando-se os parâmetros para cada modelo tem-se:

Parâmetros Subsérie CAA	Coefficiente	Erro Padrão	Estatística t	Significância
ϕ_1	-0,710816	0,215902	-3,292299	0.0011
Modelo	Log de Verossimilhança		AIC	BIC
AR(1)	-410,7312		3,344606	3,401274

Parâmetros Subsérie CAD	Coefficiente	Erro Padrão	Estatística t	Significância
ϕ_1	-0,171287	0,061596	-2,780803	0.0058
Modelo	Log de Verossimilhança		AIC	BIC
AR(1)	-360,7360		2,905510	2,919636

Parâmetros Subsérie CDA	Coefficiente	Erro Padrão	Estatística t	Significância
ϕ_1	-0,223134	0,064951	-3,435409	0.0007
Modelo	Log de Verossimilhança		AIC	BIC
AR(1)	-253,0027		2,295989	2,341825

Parâmetros Subsérie CDD	Coefficiente	Erro Padrão	Estatística t	Significância
ϕ_1	0,156319	0,067640	2,311057	0.0218
Modelo	Log de Verossimilhança		AIC	BIC
AR(1)	-284,0033		2,700965	2,779610

Estimando-se os valores e calculando as estimativas de erros pontuais para a previsão em 2 níveis de pré-processamento e fazendo o pós-processamento, tem-se:

MODELO ARIMA			
Estatísticas de Previsão Pontual para os 200 pontos	Valor	Estatísticas de Previsão Pontual para os 10 primeiros pontos	Valor
TIC	0,97	TIC	0,91
MAPE	99,68%	MAPE	9,19%
Correlação	0,11	Correlação	-0,36

MODELO ARIMA – Redes Neurais Recorrentes			
Estatísticas de Previsão Pontual para os 200 pontos	Valor	Estatísticas de Previsão Pontual para os 10 primeiros pontos	Valor
TIC	0,90	TIC	0,53
MAPE	112,19%	MAPE	7,88%
Correlação	0,32	Correlação	0,64

MODELO ARIMA – Redes Neurais Retropropagação			
Estatísticas de Previsão Pontual para os 200 pontos	Valor	Estatísticas de Previsão Pontual para os 10 primeiros pontos	Valor
TIC	0,90	TIC	0,82
MAPE	108,67%	MAPE	28,99%
Correlação	-0,01	Correlação	-0,26

Nota-se que a previsão feita para o curto prazo apresenta menos oscilações nas estatísticas pontuais, e que as redes neurais recorrentes também apresentam um significativo comportamento, principalmente em relação à correlação.

Fazendo o cálculo do envelopamento para 200 séries simuladas e verificando quais se enquadram dentro do intervalo de confiança de 95% para a série prevista, via procedimento estático, tem-se:

Envelopamento – modelo	Percentual
Redes Neurais – Recorrente	95,00%
Redes Neurais – Retropropagação	95,00%
Modelo ARIMA	90,69%

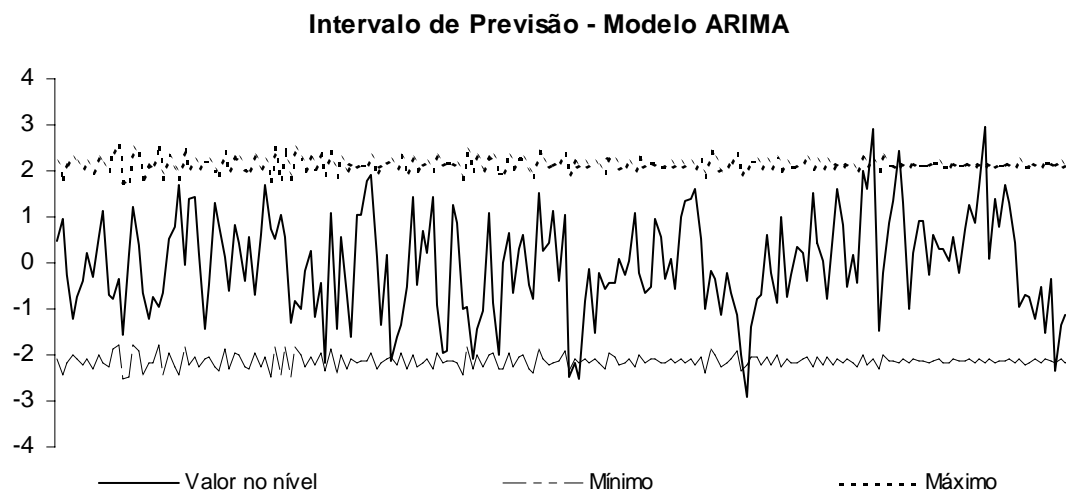


Figura 48: Intervalo de previsão em 2º nível para o DGP do modelo AR(1)

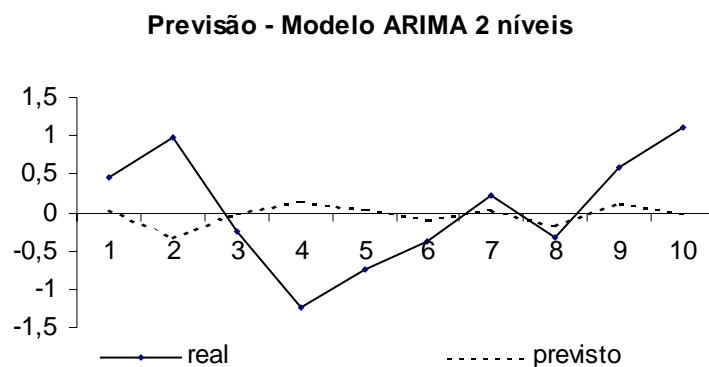


Figura 49: Previsão dos primeiros 10 pontos para o 2º nível para o DGP do modelo AR(1)

Fazendo a troca da ondaleta de Daubesch número um pela ondaleta chapéu mexicano, têm-se os seguintes resultados:

MODELO ARIMA					
Estatísticas de Previsão Pontual para os 200 pontos	Valor 1 nível de pré-processamento	Valor 2 níveis de pré-processamento	Estatísticas de Previsão Pontual para os 10 primeiros pontos	Valor 1 nível de pré-processamento	Valor 2 níveis de pré-processamento
TIC	0,99	0,98	TIC	0,40	0,18
Correlação	0,12	0,16	Correlação	0,756602	0,97
MAPE	100,00%	100,82%	MAPE	7,59%	5,69%

Envelopamento – modelo	Percentual
Modelo ARIMA 1 nível	89,67%
Modelo ARIMA 2 níveis	83.25%

Fazendo a troca da ondaleta de Daubesch, número um, pela ondaleta biortogonal 3.1, tem-se o seguinte resultado:

MODELO ARIMA					
Estatísticas de Previsão Pontual para os 200 pontos	Valor 1 nível de pré-processamento	Valor 2 níveis de pré-processamento	Estatísticas de Previsão Pontual para os 10 primeiros pontos	Valor 1 nível de pré-processamento	Valor 2 níveis de pré-processamento
TIC	1	0,99	TIC	0,92	0,62
Correlação	-0,04	0,08	Correlação	-0,27	0,29
MAPE	100,00%	100,23%	MAPE	10,55%	15,01%

Envelopamento – modelo	Percentual
Modelo ARIMA 1 nível	92,70%
Modelo ARIMA 2 níveis	90,03%

Teste para Independência Estatística dos Resíduos– Teste BDS

Como visto anteriormente, o teste para independência deve ser aplicado para ver se a sucessão cronológica apresenta independência nos resíduos. Suas hipóteses são:

H_0 :os resíduos são estocasticamente independentes

H_1 : os resíduos são estocasticamente dependentes

Resíduos Modelo ARIMA – Previsto pelo AR(1) com 1 Nível de pré-processamento					Resíduos Modelo ARIMA – Previsto pelo AR(1) com 2 Níveis de pré-processamento				
Dimensão	BDS	Erro Padrão	Estatística Z	p-valor	Dimensão	BDS	Erro Padrão	Estatística Z	p-valor
2	-3,24E-06	1,07E-06	-3,011273	0,2625	2	-2,73E-06	9,36E-07	-2,916037	0,3522
3	-5,65E-09	1,44E-08	-0,393125	0,6942	3	-4,35E-09	1,23E-08	-0,354429	0,7230
4	-9,66E-12	1,45E-10	-0,066870	0,9467	4	-6,77E-12	1,21E-10	-0,055943	0,9554
5	-1,77E-14	1,27E-12	-0,013934	0,9889	5	-1,03E-14	1,04E-12	-0,009815	0,9922
6	-2,96E-17	1,04E-14	-0,002848	0,9977	6	-1,33E-17	8,36E-15	-0,001590	0,9987

Neste caso, calculou-se também até a sexta dimensão e concluiu-se por não se rejeitar a hipótese nula ao nível de significância de 5%, indicando que os dados dos resíduos do modelo ARIMA são independentes para os modelos gerados baseados na análise do p-valor.

Teste de Normalidade

Na figura, a seguir, têm-se as estatísticas dos resíduos do pós-processamento, sendo que a hipótese nula e alternativa são:

H_0 : os resíduos seguem uma distribuição normal

H_1 : os resíduos não seguem uma distribuição normal

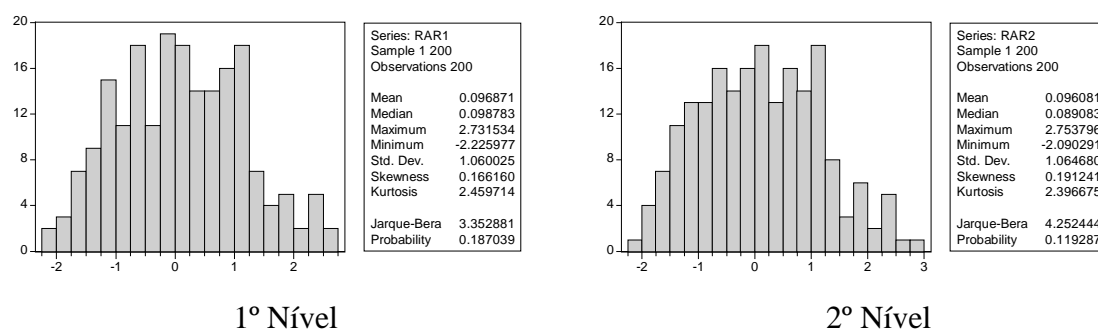


Figura 50: Distribuição dos resíduos do modelo AR(1) previsto por um modelo ARIMA, em 1 e 2 níveis de pré-processamento.

Observa-se que os dados dos resíduos são normais pela estatística de Jarque-Bera.

4.3 SIMULAÇÃO DE UM MOVIMENTO BROWNIANO GEOMÉTRICO

Gerando-se um DGP com a seguinte expressão, $X_t = X_{t-1} + X_{t-1}(\mu\Delta t + \sigma\sqrt{\Delta t}Z_t)$, $\{Z_t\} \sim N(0;1)$, para o movimento browniano geométrico, com parâmetros $\mu = 0,09$ e $\sigma = 0,40$. Tem-se mil pontos no gráfico a seguir:



Figura 51: Simulação de um movimento browniano geométrico

Verificando a sucessão cronológica como um todo, tem-se:

Teste de raiz unitária

Movimento Browniano		
Hipóteses	Estatística de Teste	Valores Críticos a 5%
$H_0 : 3RU$ $H_1 : 2RU$	-56,15525	-2,76
$H_0 : 2RU$ $H_1 : 1RU$	-31,81716 -22,45761	-3,37 -3,77
$H_0 : 1RU$ $H_1 : 0RU$	-31,80359 -22,43190 -0,324796*	-3,42 -3,80 -4,16

Nota-se a não rejeição da hipótese nula de não estacionariedade para a presença de uma raiz unitária na sucessão cronológica do movimento browniano geométrico.

Teste de independência estatística – BDS

Movimento Browniano Geométrico				
Dimensão	BDS	Erro Padrão	Estatística Z	p-valor
2	-3.56E-08	1.19E-08	-2.978556	0.2901
3	-6.75E-12	3.21E-11	-0.209812	0.8338
4	-1.28E-15	6.49E-14	-0.019788	0.9842
5	-2.46E-19	1.15E-16	-0.002140	0.9983
6	-4.72E-23	1.88E-19	-0.000251	0.9998

O teste de independência, revela a não-rejeição da hipótese nula pelo p-valor que são superiores ao nível de significância de 5%, concluindo, pela independência das observações, indicando que os dados representam um RBI.

Teste de Normalidade

A figura, a seguir, ilustra a distribuição da sucessão cronológica do movimento browniano geométrico.

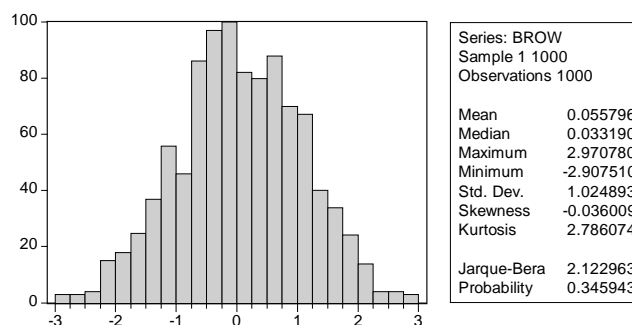


Figura 52: Distribuição do movimento browniano geométrico simulado

E seguindo o fluxograma, foi escolhida novamente a ondaleta de Daubesch número 1 (DB1). A figura, a seguir, ilustra a decomposição (pré-processamento) em primeiro nível da série no nível original:

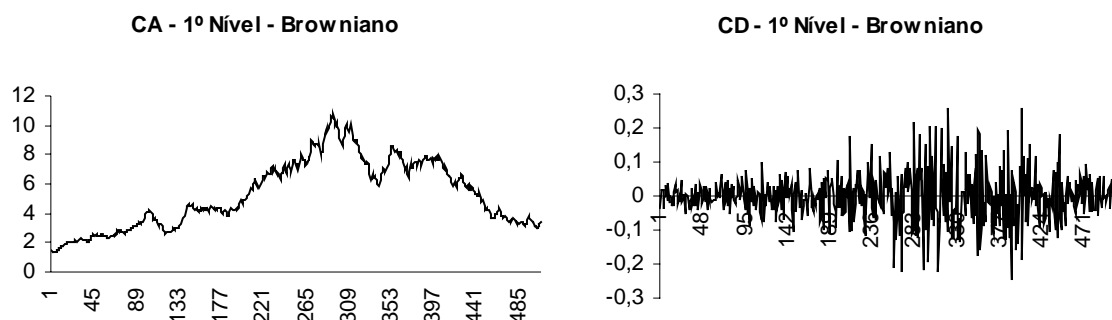


Figura 53: Pré-processamento em 1º nível do movimento browniano geométrico

Teste de raízes unitária

Subsérie: CA			Subsérie: CD		
Hipóteses	Estatística de Teste	Valores Críticos a 5%	Hipóteses	Estatística de Teste	Valores Críticos a 5%
$H_0 : 3RU$ $H_1 : 2RU$	-30,08746	-2,76	$H_0 : 3RU$ $H_1 : 2RU$	-50,20512	-2,76
$H_0 : 2RU$ $H_1 : 1RU$	-18,06209 -14,72879	-3,37 -3,77	$H_0 : 2RU$ $H_1 : 1RU$	-13,58794 -26,37183	-3,37 -3,77
$H_0 : 1RU$ $H_1 : 0RU$	-17,97134 -14,70409 0,546188*	-3,42 -3,80 -4,16	$H_0 : 1RU$ $H_1 : 0RU$	-19,45983 -9,214053 -11,54976	-3,42 -3,80 -4,16

Considerando um nível de confiança de 95%, se rejeita a todas as hipóteses nulas, o que indica que ambas as subséries são estacionárias. Nota-se ainda, que a sucessão cronológica do movimento browniano na subsérie de baixa frequência, apresenta comportamento semelhante à sucessão cronológica inicialmente simulada.

Teste de independência estatística – BDS

Subséries: CA					Subséries: CD				
<u>Dimensão</u>	<u>BDS</u>	<u>Erro Padrão</u>	<u>Estatística Z</u>	<u>p-valor</u>	<u>Dimensão</u>	<u>BDS</u>	<u>Erro Padrão</u>	<u>Estatística Z</u>	<u>p-valor</u>
2	-3,30E-08	4,93E-08	-0,668203	0,5040	2	-1,77E-08	3,91E-08	-0,452892	0,6506
3	-6,07E-12	2,93E-10	-0,020752	0,9834	3	-2,40E-12	2,28E-10	-0,010519	0,9916
4	-8,58E-16	1,30E-12	-0,000659	0,9995	4	-3,27E-16	9,92E-13	-0,000329	0,9997
5	-1,50E-19	5,06E-15	-2,97E-05	1,0000	5	-2,80E-20	3,78E-15	-7,39E-06	1,0000
6	-2,67E-23	1,82E-17	-1,46E-06	1,0000	6	-3,54E-24	1,33E-17	-2,65E-07	1,0000

O teste de independência revela a não-rejeição da hipótese nula pelo p-valor que são superiores ao nível de significância de 5%, concluindo pela independência das observações, indicando que os dados representam um processo independente.

Teste de normalidade

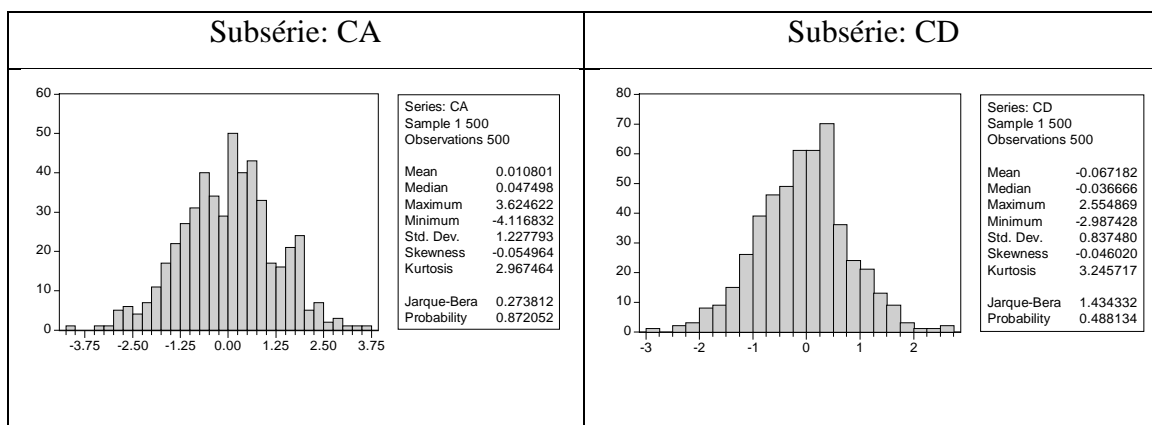
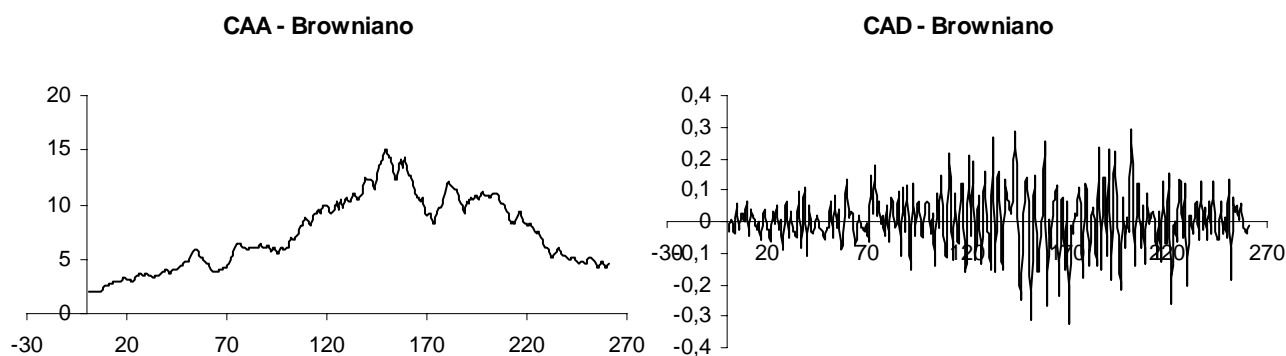


Figura 54: Distribuições em 1º nível do movimento browniano

Não se rejeita a hipótese de normalidade pelo teste de Jarque-Bera, para a sucessão cronológica de CA e CD, o que confirma ser um movimento browniano geométrico.

Simulação no 2º Nível do Pré-Processamento

Após aplicar o pré-processamento novamente para as duas primeiras sub-divisões, tem-se os gráficos das subséries decompostas. Utilizou-se a mesma ondaleta.



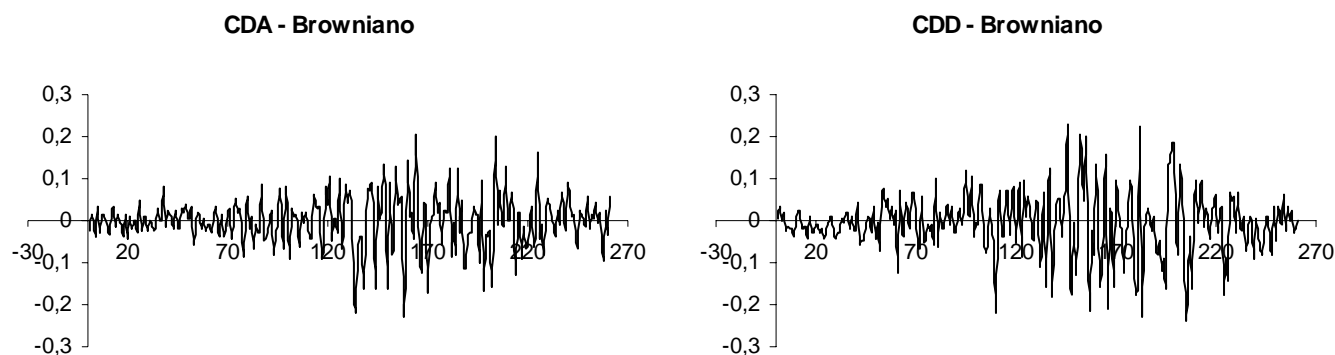


Figura 55: Pré-processamento em 2º nível do movimento browniano geométrico

Teste de raízes unitária

Subsérie: CAA			Subsérie: CAD		
Hipóteses	Estatística de Teste	Valores Críticos a 5%	Hipóteses	Estatística de Teste	Valores Críticos a 5%
$H_0 : 3RU$ $H_1 : 2RU$	-24,87385	-2,76	$H_0 : 3RU$ $H_1 : 2RU$	-32,61729	-2,76
$H_0 : 2RU$ $H_1 : 1RU$	-15,34300 -9,025265	-3,37 -3,77	$H_0 : 2RU$ $H_1 : 1RU$	-9,160192 -19,56653	-3,37 -3,77
$H_0 : 1RU$ $H_1 : 0RU$	-15,22935 -8,970041 0,245851*	-3,42 -3,80 -4,16	$H_0 : 1RU$ $H_1 : 0RU$	-15,53842 -3,709442 -10,59386	-3,42 -3,80 -4,16
Subsérie: CDA			Subsérie: CDD		
Hipóteses	Estatística de Teste	Valores Críticos a 5%	Hipóteses	Estatística de Teste	Valores Críticos a 5%
$H_0 : 3RU$ $H_1 : 2RU$	-29,37741	-2,76	$H_0 : 3RU$ $H_1 : 2RU$	-27,85526	-2,76
$H_0 : 2RU$ $H_1 : 1RU$	-9,627443 -17,91034	-3,37 -3,77	$H_0 : 2RU$ $H_1 : 1RU$	-9,468162 -18,09463	-3,37 -3,77
$H_0 : 1RU$ $H_1 : 0RU$	-14,61356 -5,231595 -9,147208	-3,42 -3,80 -4,16	$H_0 : 1RU$ $H_1 : 0RU$	-13,51133 -6,717338 -8,064806	-3,42 -3,80 -4,16

Observa-se claramente que a subsérie de baixa frequência do 2º nível de pré-processamento se apresenta com a mesma forma do modelo simulado originalmente, ou seja, a transformada de ondaleta de baixa frequência não decompõe o movimento browniano geométrico. Considerando um nível de confiança de 95%, se rejeita a todas as hipóteses nulas, o que indica que ambas as subséries são estacionárias.

Teste de independência estatística – BDS

Subséries: CAA					Subséries: CAD				
<u>Dimensão</u>	<u>BDS</u>	<u>Erro Padrão</u>	<u>Estatística Z</u>	<u>p-valor</u>	<u>Dimensão</u>	<u>BDS</u>	<u>Erro Padrão</u>	<u>Estatística Z</u>	<u>p-valor</u>
2	2,89E-05	1,06E-06	27,32995	0,4101	2	2,97E-05	7,33E-07	40,45374	0,4214
3	-6,79E-09	1,42E-08	-0,478657	0,6322	3	-4,62E-09	9,42E-09	-0,490176	0,6240
4	-1,33E-11	1,43E-10	-0,092749	0,9261	4	-7,94E-12	9,08E-11	-0,087440	0,9303
5	-2,64E-14	1,27E-12	-0,020840	0,9834	5	-1,39E-14	7,67E-13	-0,018099	0,9856
6	-4,81E-17	1,04E-14	-0,004630	0,9963	6	-2,47E-17	6,00E-15	-0,004114	0,9967

Subséries: CDA					Subséries: CDD				
<u>Dimensão</u>	<u>BDS</u>	<u>Erro Padrão</u>	<u>Estatística Z</u>	<u>p-valor</u>	<u>Dimensão</u>	<u>BDS</u>	<u>Erro Padrão</u>	<u>Estatística Z</u>	<u>p-valor</u>
2	2,69E-05	1,03E-06	26,21474	0,3841	2	-3,17E-06	8,75E-07	-3,627919	0,3996
3	-1,25E-08	1,49E-08	-0,836821	0,4027	3	-5,79E-09	1,16E-08	-0,499994	0,6171
4	-2,98E-11	1,62E-10	-0,184312	0,8538	4	-9,98E-12	1,15E-10	-0,086680	0,9309
5	-7,26E-14	1,54E-12	-0,047096	0,9624	5	-1,85E-14	1,00E-12	-0,018419	0,9853
6	-1,65E-16	1,36E-14	-0,012133	0,9903	6	-3,11E-17	8,09E-15	-0,003841	0,9969

O teste de independência revela a não-rejeição da hipótese nula pelo p-valor que são superiores ao nível de significância de 5%, concluindo pela independência das observações.

4.4 SIMULAÇÃO DE UM PROCESSO BILINEAR

Considere agora um DGP para o modelo bilinear da equação (2.76) com a seguinte equação: $X_t = 0.3X_{t-1} + 0.8X_{t-1}\varepsilon_{t-1} + \varepsilon_t$. O DGP possui um grande coeficiente associado ao termo bilinear e apresenta um grande número de explosões de grande amplitude.

Foram gerados, pelo programa 6 do Apêndice C, 1400 valores para o modelo, descartados os 200 primeiros para que os valores iniciais atribuídos $x_0 = \varepsilon_0 = 0$ não tenham influência no comportamento do modelo. Seu gráfico com 1000 pontos pode ser visto a seguir:

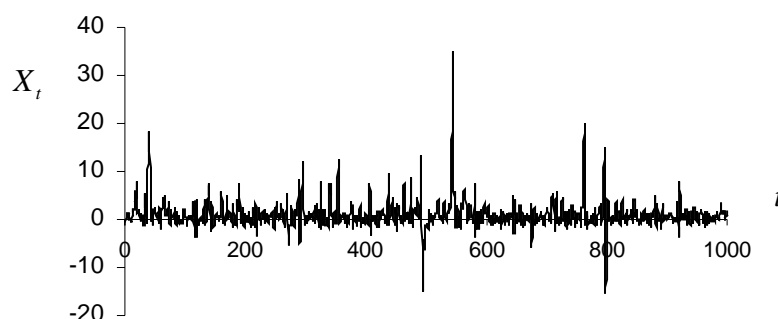


Figura 56: DGP para um modelo bilinear.

Verificando a sucessão cronológica como um todo, tem-se:

Teste de raiz unitária

Modelo Bilinear		
Hipóteses	Estatística de Teste	Valores Críticos a 5%
$H_0 : 3RU$ $H_1 : 2RU$	-55,61291	-2,76
$H_0 : 2RU$ $H_1 : 1RU$	-24,91645 -31,05414	-3,37 -3,77
$H_0 : 1RU$ $H_1 : 0RU$	-30,14238 -18,63510 -13,32537	-3,42 -3,80 -4,16

Considerando um nível de confiança de 95%, se rejeita a todas as hipóteses nulas, o que indica que ambas as subséries são estacionárias.

Teste de independência estatística – BDS

Modelo Bilinear				
Dimensão	BDS	Erro Padrão	Estatística Z	p-valor
2	0,074498	0,003227	23,08473	0,0000
3	0,119117	0,005131	23,21576	0,0000
4	0,139180	0,006114	22,76362	0,0000
5	0,144287	0,006378	22,62265	0,0000
6	0,138812	0,006156	22,54743	0,0000

Calculou-se até a sexta dimensão e concluiu-se por rejeitar a hipótese nula ao nível de significância de 5%, indicando que os dados da sucessão cronológica são dependentes, baseados na análise do p-valor, sendo que pelo menos um dos p-valores deve ser menor que 5%.

Teste de Normalidade

A figura, a seguir, ilustra a distribuição da sucessão cronológica.

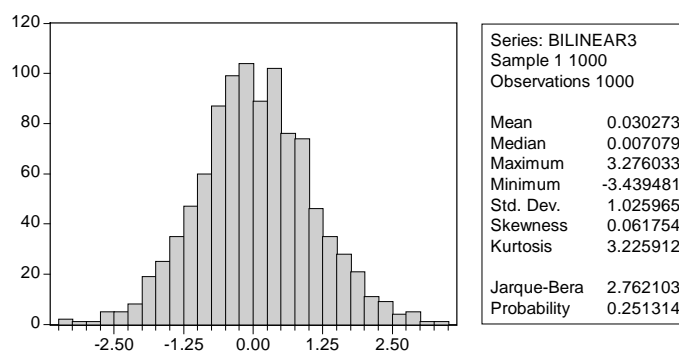


Figura 57: Distribuições do modelo bilinear simulado

Observa-se, pelo p-valor, que a distribuição da sucessão cronológica bilinear segue uma distribuição normal pela não rejeição da hipótese nula.

Teste de McLeod-Li

O resultado da estatística de teste calculado no Eviews® 5.0 é:

Modelo Bilinear			
Estatística F	205,6457	p-valor	0,00000

Dessa forma, o teste de McLeod-Li para 5 defasagens rejeitou a hipótese nula pelo fato do p-valor ser inferior a 5%. Isto indica que a sucessão cronológica do modelo bilinear possui estatisticamente comportamento não-linear.

Teste de Hsieh

Uma vez estabelecido que existe algum tipo de não-linearidade em uma sucessão cronológica, é preciso identificar o tipo de não-linearidade existente, se na média ou na variância condicionadas.

Para a realização do teste, fez-se o programa 4 que irá calcular os coeficientes (i,j) para cada uma das subséries. Os resultados desta etapa vem a seguir:

Modelo Bilinear		
Coeficiente	Estatística de Hsieh	p-valor
1 1	-1,8707	0,0307
1 2	-1,406	0,0799*
1 3	-2,6029	0,0046
1 4	-1,7474	0,0403
1 5	-1,9897	0,0233
2 2	-2,4454	0,0072
2 3	-1,689	0,0456
2 4	-1,9339	0,0266
2 5	-1,688	0,0457
3 3	-1,6685	0,0476
3 4	-1,9953	0,0230
3 5	-2,1541	0,0156
4 4	-2,2459	0,0124
4 5	1,0084	0,8434*
5 5	3,0644	0,9989*

Conforme se observa no quadro anterior, para quinze pares (i,j) , os valores dos coeficientes do teste de Hsieh amostrais para o modelo bilinear e o correspondente p-valor para um nível de confiança de 95%, apresenta apenas três coeficientes significantes (indicado com *), podendo-se dizer que, de maneira geral, rejeita-se a hipótese nula dos coeficientes iguais a zero, o que indica uma não-linearidade na esperança condicionada.

E seguindo o fluxograma de análise e testes, foi escolhido novamente a ondaleta de Daubesch número 1 (DB1). A figura, a seguir, ilustra a decomposição (pré-processamento) em primeiro nível da série no nível original:

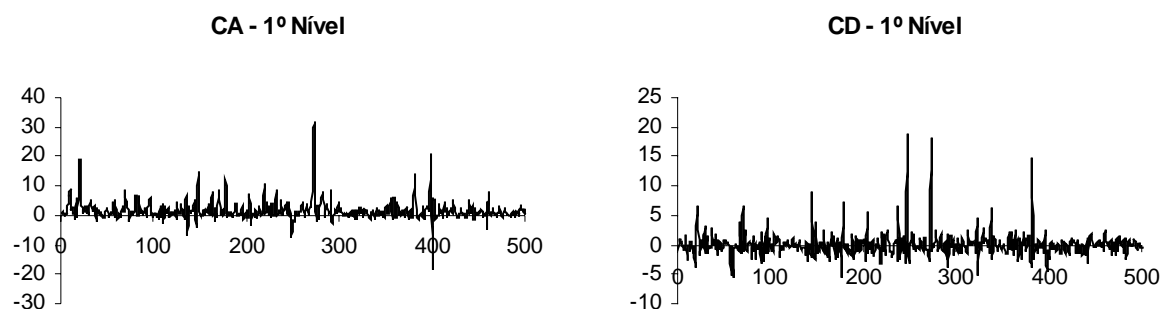


Figura 58: Pré-processamento em 1º nível para o DGP do modelo bilinear

Teste de raízes unitárias

Subséries: CA			Subséries: CD		
Hipóteses	Estatística de Teste	Valores Críticos a 5%	Hipóteses	Estatística de Teste	Valores Críticos a 5%
$H_0 : 3RU$ $H_1 : 2RU$	-40.14459	-2.76	$H_0 : 3RU$ $H_1 : 2RU$	-50.94260	-2.76
$H_0 : 2RU$ $H_1 : 1RU$	-15.48948 -25.77090	-3.37 -3.77	$H_0 : 2RU$ $H_1 : 1RU$	-14.33825 -30.38486	-3.37 -3.77
$H_0 : 1RU$ $H_1 : 0RU$	-18.97206 -15.03338 -9.008961	-3.42 -3.80 -4.16	$H_0 : 1RU$ $H_1 : 0RU$	-22.40184 -8.158893 -14.47807	-3.42 -3.80 -4.16

Considerando um nível de confiança de 95%, rejeitam-se a todas as hipóteses nulas, o que indica que ambas as subséries são estacionárias.

Teste para Independência Estatística – BDS

Subsérie: CA					Subsérie: CD				
Dimensão	BDS	Erro Padrão	Estatística Z	p-valor	Dimensão	BDS	Erro Padrão	Estatística Z	p-valor
2	0,000507	9,04E-05	5,611479	0,0000	2	0,000311	6,40E-05	4,853707	0,0000
3	0,000114	1,16E-05	9,821680	0,0000	3	7,86E-05	7,78E-06	10,10195	0,0000
4	2,43E-05	1,12E-06	21,75208	0,0000	4	1,52E-06	7,10E-07	2,137668	0,0325
5	-4,33E-07	9,46E-08	-4,577848	0,0000	5	-3,32E-07	5,69E-08	-5,833533	0,0000
6	-2,34E-08	7,43E-09	-3,151588	0,0016	6	-1,65E-08	4,22E-09	-3,903729	0,0001

O teste de BDS rejeita a hipótese de independência das subséries todos com 95% de confiança, o que também implica na existência de não-linearidade.

Teste de normalidade

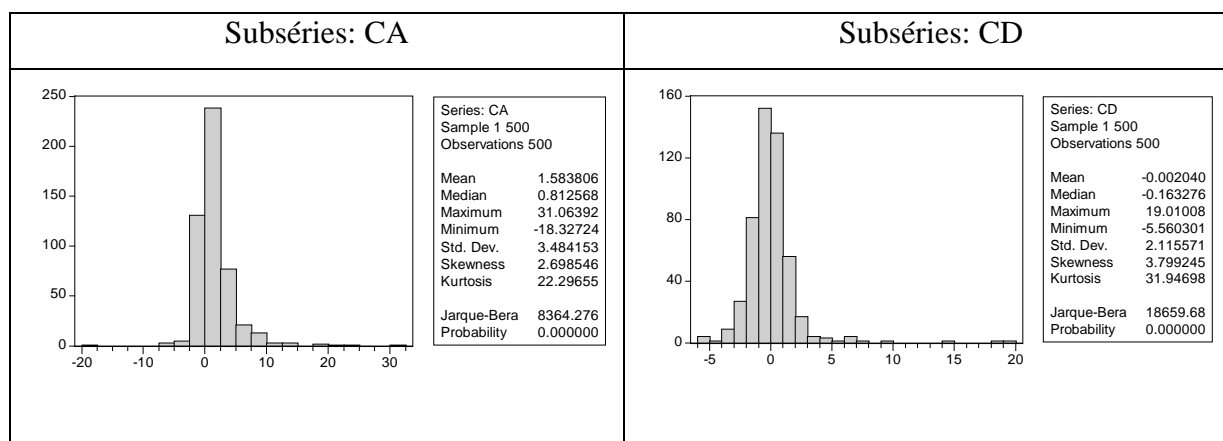


Figura 59: Distribuição das subséries para o modelo bilinear

Rejeita-se a hipótese de normalidade pelo teste de Jarque-Bera com 5% de significância.

Teste de McLeod-Li

Subsérie: CA				Subsérie: CD			
Estatística F	31.90431	p-valor	0.000000	Estatística F	40.08631	p-valor	0.000000

Dessa forma, o resultado do teste de McLeod-Li para 5 defasagens rejeitou a hipótese nula pelo fato do p-valor ser inferior a 5%. Isto indica que ambas as subséries estatisticamente possuem comportamento não-linear.

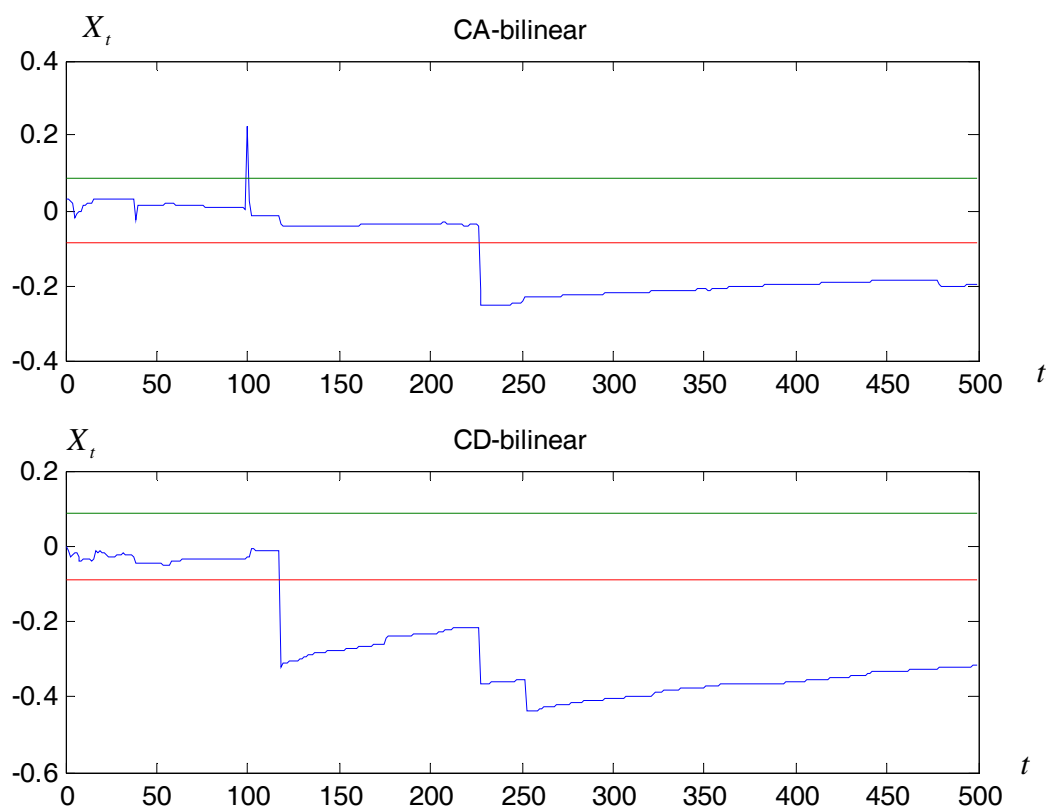


Figura 60: Comportamento da série ao longo do tempo para o modelo bilinear em 1 níveis de pré-processamento.

Nota-se que, nas duas decomposições, ocorreram a não-linearidade, o que não vinha acontecendo nos demais processos.

Teste de Hsieh

Uma vez estabelecido que existe algum tipo de não-linearidade, em uma sucessão cronológica, é preciso filtrar o tipo de não-linearidade existente, se na média ou na variância.

Subsérie: CA			Subsérie: CD		
Coeficiente	Estatística de Hsieh	p-valor	Coeficiente	Estatística de Hsieh	p-valor
1 1	0,069	0,5275	1 1	-0,7257	0,23401
1 2	0,0386	0,5154	1 2	13,2497*	0,00000
1 3	1,94	0,0674	1 3	1,6951	0,95497
1 4	5,474*	0,0000	1 4	-4,8641*	0,00000
1 5	-8,125*	0,0000	1 5	-5,2335*	0,00000
2 2	0,0044	0,5018	2 2	0,1219	0,54851
2 3	-1,997*	0,0000	2 3	7,4856*	0,00000
2 4	-2,423*	0,0000	2 4	2,1278*	0,00000
2 5	-2,4157*	0,0000	2 5	-6,646*	0,00000

Subsérie: CA			Subsérie: CD		
3 3	-1,748	0,0847	3 3	1,1512	0,87517
3 4	1,434	0,4145	3 4	4,4668*	0,00000
3 5	7,145*	0,0000	3 5	4,2653*	0,00000
4 4	-1,435	0,0756	4 4	0,6609	0,74566
4 5	4,095*	0,0000	4 5	6,3659*	0,00000
5 5	-4,567*	0,0000	5 5	1,072	0,85814

Conforme se observa, no quadro anterior, para quinze pares i,j , os valores dos coeficientes do teste de Hsieh amostrais para as duas subséries e o correspondente p-valor para um nível de confiança de 95%, apresenta nove coeficientes significantes (indicados com *) na série CD e oito na série CA, podendo-se dizer que, de maneira geral, rejeita-se a hipótese nula dos coeficientes iguais a zero, o que indica uma não-linearidade na esperança condicionada.

Estimando-se os valores e calculando as estimativas de erros pontuais para a previsão em 1 nível de pré-processamento, via redes neurais, com algoritmo de retropropagação, tem-se:

Previsão em 1 nível de pré-processamento via redes neurais com algoritmo de retropropagação			
MODELO BILINEAR			
Estatísticas de Previsão Pontual para os 200 pontos	Valor	Estatísticas de Previsão Pontual para os 10 primeiros pontos	Valor
TIC	0,63	TIC	0,68
MAPE	342,27%	MAPE	81,86%
Correlação	-0,03	Correlação	-0,66

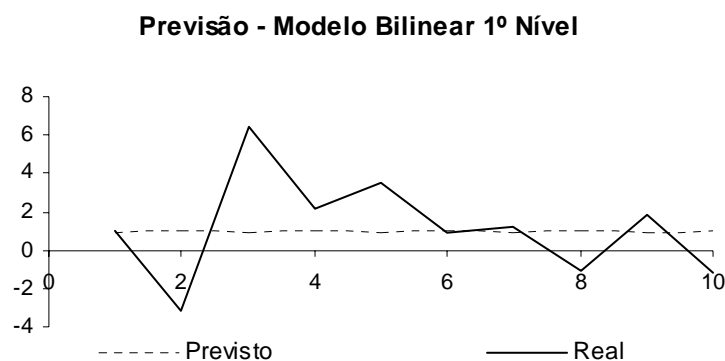


Figura 61: Previsão em 1º nível para o modelo bilinear por redes neurais com retropropagação para os primeiros 10 elementos.

Estimando-se agora os valores e calculando as estimativas de erros pontuais para a previsão em 1 nível de pré-processamento, via redes neurais com algoritmo de redes recorrentes, tem-se:

Previsão em 1 nível via redes neurais com algoritmo de redes recorrentes			
MODELO BILINEAR			
Estatísticas de Previsão Pontual para os 200 pontos	Valor	Estatísticas de Previsão Pontual para os 10 primeiros pontos	Valor
TIC	0,52	TIC	0,24
MAPE	91,15%	MAPE	128,70%
Correlação	0,32	Correlação	0,91

Fazendo o cálculo do envelopamento para 200 séries simuladas e verificando quais se enquadram dentro do intervalo de confiança de 95%, para a série prevista via procedimento estático, tem-se:

Envelopamento – modelo	Percentual
Modelo Bilinear – Rede Neural recorrente	96,38%
Modelo Bilinear – Rede Neural com retropropagação	90,35%

Previsão - Modelo Bilinear 1º Nível

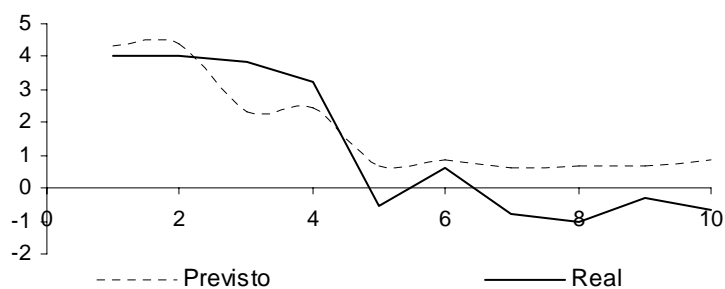


Figura 62: Previsão em 1º nível para o modelo bilinear por redes neurais com redes recorrentes para os primeiros 10 elementos. As linhas estão muito lisas...usar linhas retas para ligar os pontos.

Simulação no 2º Nível do Pré-Processamento

Apresenta-se, a seguir, o gráfico da decomposição em 2 níveis do DGP para o modelo bilinear:

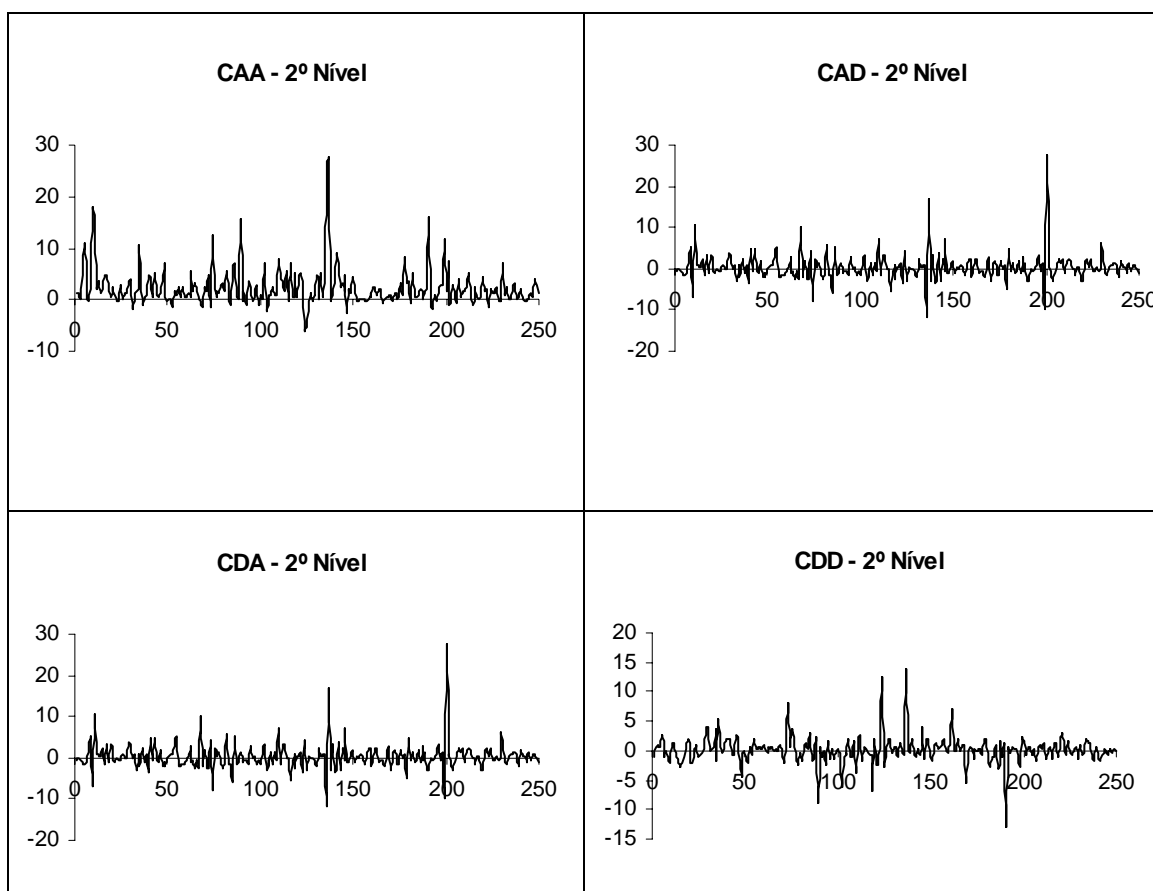


Figura 63: Pré-processamento em 2º nível para o DGP do modelo bilinear.

Seguindo o mesmo roteiro do modelo anterior, apresentam-se, a seguir, os resultados obtidos:

Teste de Raízes Unitárias

Subsérie: CAA			Subsérie: CAD		
Hipóteses	Estatística de Teste	Valores Críticos a 5%	Hipóteses	Estatística de Teste	Valores Críticos a 5%
$H_0 : 3RU$ $H_1 : 2RU$	-29,80782	-2,76	$H_0 : 3RU$ $H_1 : 2RU$	-44,43612	-2,76
$H_0 : 2RU$ $H_1 : 1RU$	-11,48298 -17,39684	-3,37 -3,77	$H_0 : 2RU$ $H_1 : 1RU$	-11,08275 -20,16610	-3,37 -3,77
$H_0 : 1RU$ $H_1 : 0RU$	-13,64584 -10,79305 -5,973635	-3,42 -3,80 -4,16	$H_0 : 1RU$ $H_1 : 0RU$	-17,02308 -5,154023 -10,56126	-3,42 -3,80 -4,16

Teste de Raízes Unitárias

Subsérie: CDA			Subsérie: CDD		
Hipóteses	Estatística de Teste	Valores Críticos a 5%	Hipóteses	Estatística de Teste	Valores Críticos a 5%
$H_0 : 3RU$ $H_1 : 2RU$	-38,62259	-2,76	$H_0 : 3RU$ $H_1 : 2RU$	-33,04524	-2,76
$H_0 : 2RU$ $H_1 : 1RU$	-10,56866 -20,65877	-3,37 -3,77	$H_0 : 2RU$ $H_1 : 1RU$	-10,02628 -21,07824	-3,37 -3,77
$H_0 : 1RU$ $H_1 : 0RU$	-15,35957 -6,859079 -9,253903	-3,42 -3,80 -4,16	$H_0 : 1RU$ $H_1 : 0RU$	-14,00581 -8,350655 -8,252479	-3,42 -3,80 -4,16

Observa-se, que com 95% de confiança, ambas as quatro subséries definidas no pré-processamento são estatisticamente estacionárias.

Teste de Normalidade

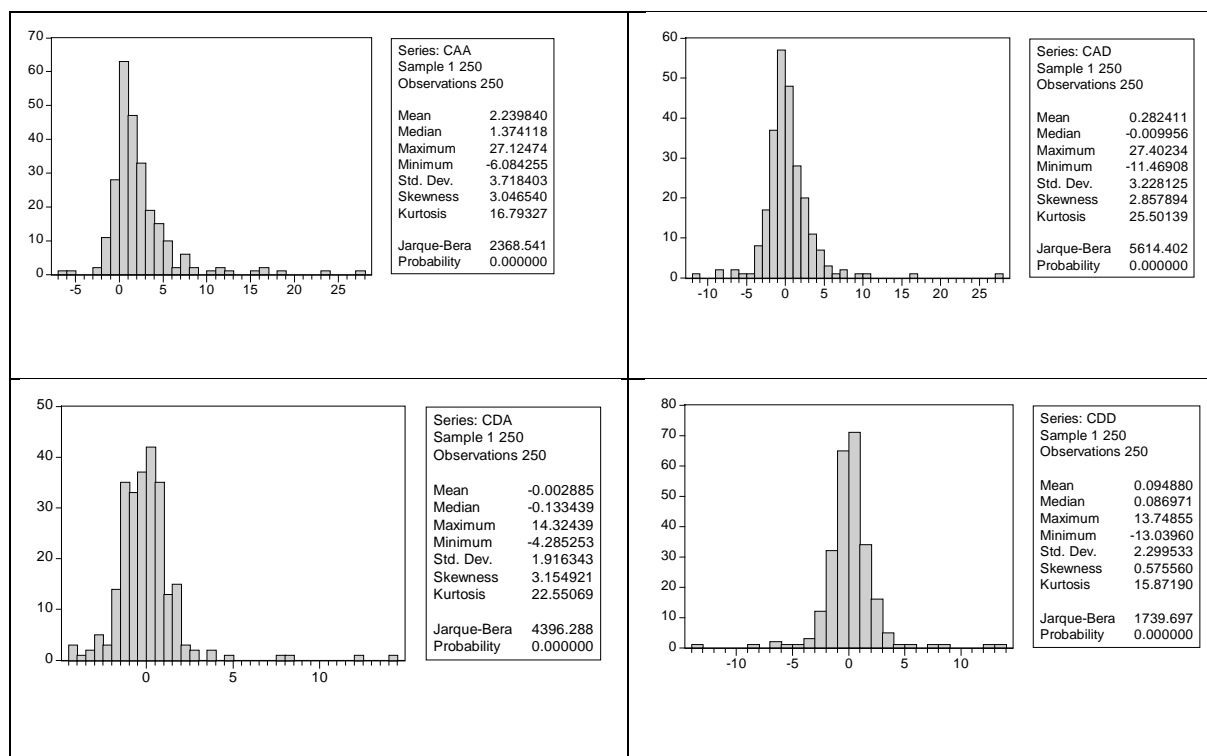


Figura 64: Distribuição em 2º nível para o pré-processamento do DGP do modelo bilinear

O mesmo ocorre com o teste de normalidade dos dados, o que indica que todas as subséries apresentam comportamento não normal.

Teste para independência estatística – BDS

Subsérie: CAA					Subsérie: CAD				
<u>Dimensão</u>	<u>BDS</u>	<u>Erro Padrão</u>	<u>Estatística Z</u>	<u>p-valor</u>	<u>Dimensão</u>	<u>BDS</u>	<u>Erro Padrão</u>	<u>Estatística Z</u>	<u>p-valor</u>
2	0,000374	0,000100	3,736277	0,0002	2	-4,10E-05	0,000137	-0,298182	0,7656
3	-5,22E-07	1,27E-05	-0,041172	0,9672	3	0,000149	1,79E-05	8,322030	0,0000
4	-6,67E-06	1,20E-06	-5,537094	0,0000	4	5,80E-05	1,75E-06	33,21924	0,0000
5	-3,30E-07	1,00E-07	-3,283252	0,0010	5	-3,97E-07	1,50E-07	-2,650488	0,0080
6	-1,64E-08	7,76E-09	-2,109405	0,0349	6	-2,04E-08	1,19E-08	-1,712293	0,0868

Subsérie: CDA					Subsérie: CDD				
<u>Dimensão</u>	<u>BDS</u>	<u>Erro Padrão</u>	<u>Estatística Z</u>	<u>p-valor</u>	<u>Dimensão</u>	<u>BDS</u>	<u>Erro Padrão</u>	<u>Estatística Z</u>	<u>p-valor</u>
2	1,26E-05	7,43E-05	0,169738	0,8652	2	0,000200	0,000140	1,435249	0,1512
3	-2,60E-05	9,18E-06	-2,830488	0,0046	3	-5,79E-05	1,75E-05	-3,302999	0,0010
4	-6,17E-06	8,51E-07	-7,247715	0,0000	4	-6,10E-06	1,66E-06	-3,682260	0,0002
5	-3,07E-07	6,91E-08	-4,434827	0,0000	5	-2,98E-07	1,38E-07	-2,162280	0,0306
6	-1,50E-08	5,20E-09	-2,891670	0,0038	6	-1,45E-08	1,06E-08	-1,368179	0,1713

O teste de independência não aceita a hipótese nula independência das observações.

Teste de McLeod-Li

Subséries: CAA				Subséries: CAD			
Estatística F	7,351095	p-valor	0,000002	Estatística F	3,279797	p-valor	0,0006965
Subséries: CDA				Subséries: CDD			
Estatística F	1,2626232	p-valor	0,0386371	Estatística F	3,010427	p-valor	0,0079300

Pelo teste, rejeita-se a hipótese nula, o que indica que o modelo é estatisticamente não-linear.

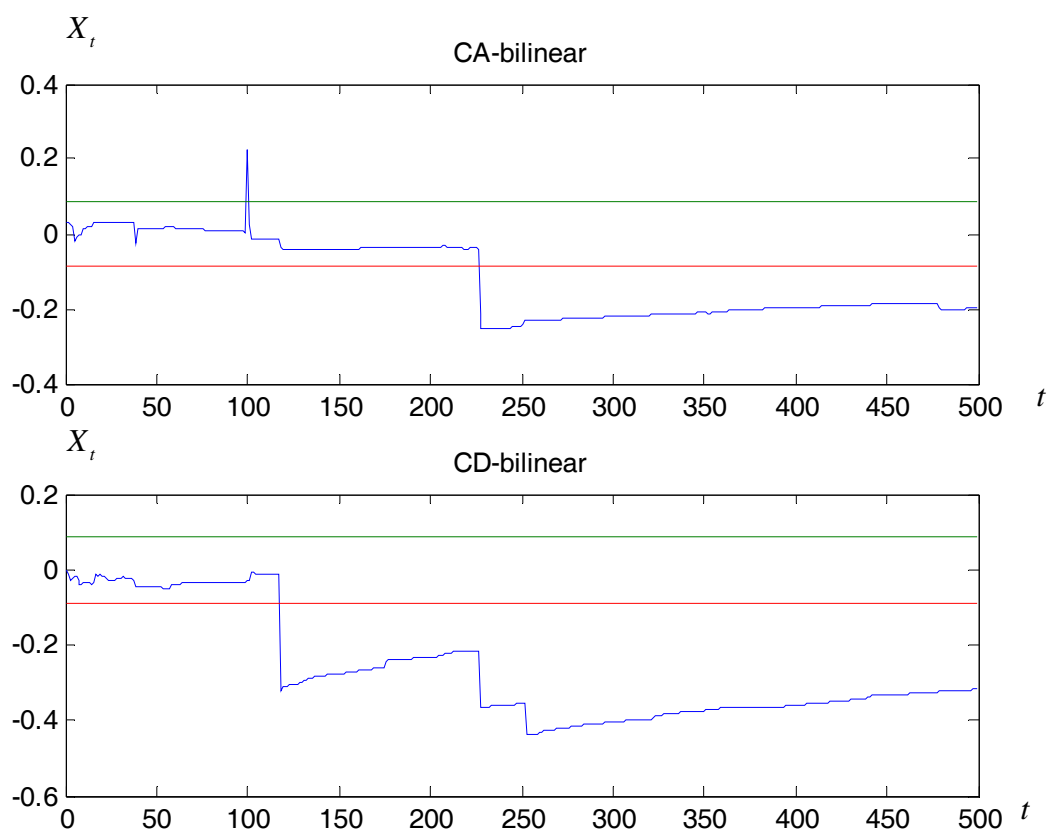


Figura 65: Comportamento da série ao longo do tempo para o modelo bilinear em 1 níveis de pré-processamento.

As redes recorrentes foram geradas com a arquitetura 10-3R-1. Estimando-se os resultados por redes neurais com retropropagação, tem-se, a seguir, um quadro resumindo o tipo de rede utilizado e as respectivas performances de seleção, teste e validação.

Subséries	Redes	Performance de treinam.	Performance Seleção	Performance Teste	Erro Treinamento	Erro Seleção	Erro Teste	Características
CAA	MLP s5 1:5-6-2-1:1	0,846273	1,065622	1,209212	0,084252	0,082395	0,093471	BP100,CG21b
CAD	MLP s5 1:5-1-1:1	0,960315	0,959103	1,044380	0,095606	0,074169	0,079927	BP100,CG21b
CDA	MLP s3 1:3-1-1:1	0,966962	0,992191	0,980613	0,078589	0,105044	0,100650	BP100,CG25b
CDD	MLP s5 1:5-1-1:1	0,962363	0,992466	0,971795	0,078215	0,105076	0,099702	BP100,CG39b

As estatísticas de qualidade das previsões foram:

MODELO BILINEAR - RN com retropropagação			
Estatísticas de Previsão Pontual para os 200 pontos	Valor	Estatísticas de Previsão Pontual para os 10 primeiros pontos	Valor
TIC	0,62	TIC	0,71
MAPE	346,47%	MAPE	86,93%
Correlação	-0,07	Correlação	-0,50

Estimando-se os valores e calculando as estimativas de erros pontuais para a previsão em 2 níveis de pré-processamento, por meio de redes neurais recorrentes, tem-se:

MODELO BILINEAR - RN recorrente			
Estatísticas de Previsão Pontual para os 200 pontos	Valor	Estatísticas de Previsão Pontual para os 10 primeiros pontos	Valor
TIC	0,85	TIC	0,65
MAPE	199,69%	MAPE	118,56%
Correlação	0,16	Correlação	-0,41

Fazendo o cálculo do envelopamento para 200 séries simuladas e verificando quais se enquadram dentro do intervalo de confiança de 95%, para a série prevista via procedimento estático do modelo bilinear com 2 níveis de pré-processamento, tem-se:

Envelopamento – modelo	Percentual
Modelo bilinear - Rede Neural Recorrente	95,29%
Modelo bilinear – Rede Neural Retropropagação	91,11%

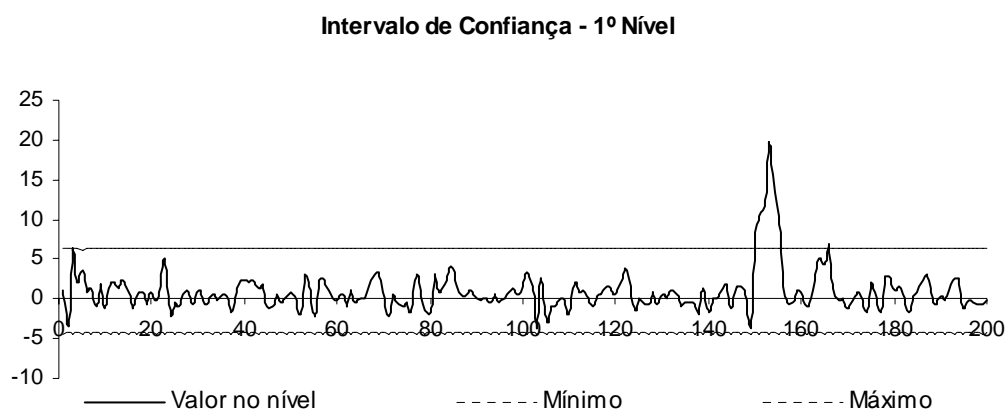


Figura 66: Intervalo de previsão em 1º nível para o DGP do modelo bilinear via RN com retropropagação.

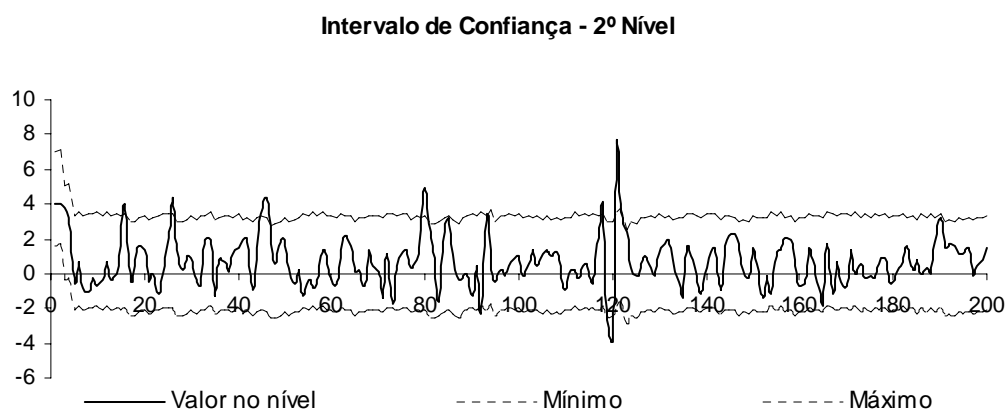


Figura 67: Intervalo de previsão em 2º nível para o DGP do modelo bilinear via redes neurais com retropropagação.

Teste para Independência Estatística dos Resíduos– Teste BDS

Como visto anteriormente, o teste para independência deve ser aplicado para ver se a sucessão cronológica apresenta independência nos resíduos. Suas hipóteses são:

H_0 :os resíduos são estocasticamente independentes

H_1 :os resíduos são estocasticamente dependentes

Resíduos Modelo Bilinear – Previsto RN com retrop. com 1 Nível de pré-processamento					Resíduos Modelo Bilinear – Previsto RN com retrop. com 2 Níveis de pré-processamento				
<u>Dimensão</u>	<u>BDS</u>	<u>Erro</u> <u>Padrão</u>	<u>Estatística</u> <u>Z</u>	<u>p-valor</u>	<u>Dimensão</u>	<u>BDS</u>	<u>Erro</u> <u>Padrão</u>	<u>Estatística</u> <u>Z</u>	<u>p-valor</u>
2	0,014389	0,006935	2,074836	0,0880	2	6,61E-05	3,39E-06	19,52304	0,1540
3	0,018640	0,011068	1,684062	0,0922	3	-1,71E-07	7,47E-08	-2,288941	0,2104
4	0,016867	0,013241	1,273817	0,2027	4	-9,13E-10	1,24E-09	-0,736845	0,4612
5	0,018619	0,013868	1,342564	0,1794	5	-5,07E-12	1,80E-11	-0,281401	0,7784
6	0,023703	0,013441	1,763576	0,0778	6	-2,53E-14	2,43E-13	-0,104304	0,9169

Resíduos Modelo Bilinear – Previsto RN recorrentes com 1 Nível de pré-processamento					Resíduos Modelo Bilinear – Previsto RN recorrentes com 2 Níveis de pré-processamento				
<u>Dimensão</u>	<u>BDS</u>	<u>Erro</u> <u>Padrão</u>	<u>Estatística</u> <u>Z</u>	<u>p-valor</u>	<u>Dimensão</u>	<u>BDS</u>	<u>Erro</u> <u>Padrão</u>	<u>Estatística</u> <u>Z</u>	<u>p-valor</u>
2	-1,86E-05	2,54E-06	-7,308696	0,0877	2	0,004820	0,006237	0,772786	0,4396
3	-7,82E-08	4,83E-08	-1,619223	0,1054	3	0,008734	0,009926	0,879949	0,3789
4	-3,25E-10	6,92E-10	-0,469927	0,6384	4	0,006174	0,011838	0,521503	0,6020
5	-1,44E-12	8,68E-12	-0,165516	0,8685	5	0,007244	0,012360	0,586084	0,5578
6	-6,46E-15	1,01E-13	-0,063943	0,9490	6	0,007796	0,011941	0,652919	0,5138

Calculou-se a estatística BDS para todas as dimensões de dois até o valor especificado, variando a dimensão de correlação para maximizar o poder do teste. Neste caso, calculou-se até a sexta dimensão e concluiu-se por não se rejeitar a hipótese nula ao nível de significância de 5%, indicando que os dados dos resíduos do modelo Bilinear são independentes para os modelos gerados baseados na análise do p-valor.

Teste de Normalidade

Na figura a seguir, tem-se as estatísticas dos resíduos do pós-processamento, onde a hipótese nula e alternativa são:

H_0 : os resíduos seguem uma distribuição normal

H_1 : os resíduos não seguem uma distribuição normal

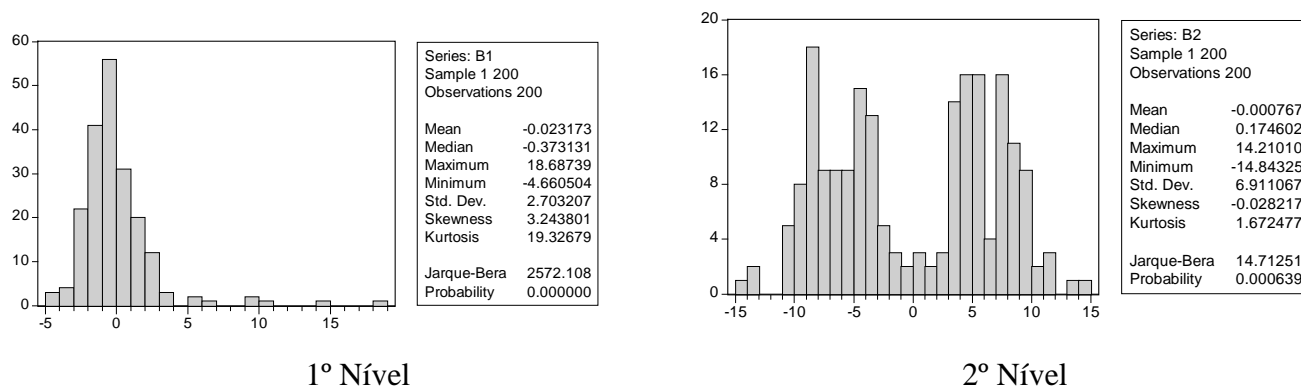


Figura 68: Distribuição dos resíduos do modelo Bilinear previsto por redes neurais com retropropagação, em 1 e 2 níveis de pré-processamento.

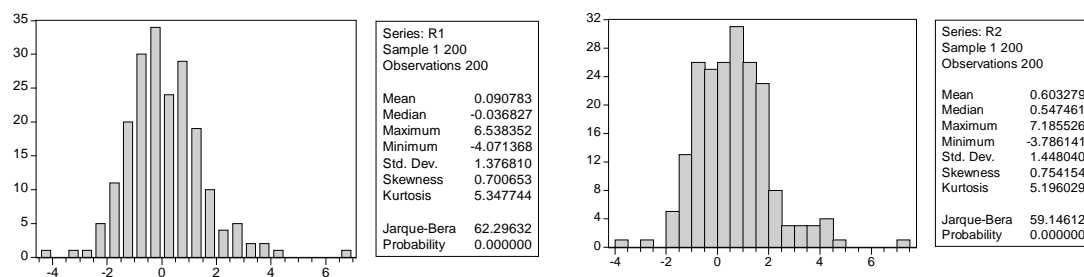


Figura 69: Distribuição dos resíduos do modelo bilinear previsto redes neurais recorrentes, em 1 e 2 níveis de pré-processamento.

Observa-se que os dados dos resíduos não são normais pela estatística de Jarque-Bera. Existe um teste para detectar a bilinearidade nos resíduos. Para mais informações, consultar Cromwell, Labys e Terraza (1994).

A seguir, tem-se um quadro comparativo da acurácia entre os modelos.

Quadro 4: Resumo das estatísticas da qualidade da previsão pontual e do envelopamento das sucessões – modelo ARIMA-GARCH

MODELO SIMULADO	PREVISÃO ESTATICA FEITA COM MODELO ARIMA-GARCH						
	Estatística de Análise	Sem pré-processamento		Com pré-processamento em 1 nível		Com pré-processamento em 2 níveis	
		Previsão de 200 pontos	Previsão de 10 pontos	Previsão de 200 pontos	Previsão de 10 pontos	Previsão de 200 pontos	Previsão de 10 pontos
ARIMA- GARCH	TIC	1,00	0,96	0,99	0,83	0,99	0,96
	MAPE	100%	9,49%	110,00%	10,0,%	99,99%	11,37%
	Correlação	-0,02	0,18	-0,11	0,16	-0,07	-0,47
	Envelopamento	48,06%		48,42%		48,15%	
	PREVISÃO ESTATICA FEITA COM REDE NEURAL RETROPROPAGAÇÃO						
	TIC	0,93	0,97	0,86	0,75	0,99	0,86
	MAPE	98,12%	9,25%	72,76%	3,39%	98,46%	13,05%
	Correlação	0,01	0,10	-0,47	-0,04	-0,11	-0,28
	Envelopamento	49,44%		93,74%		91,59%	
	PREVISÃO ESTATICA FEITA COM REDE NEURAL RECORRENTE						
	TIC	0,60	0,82	0,77	0,85	0,69	0,14
	MAPE	98,90%	9,60%	26,22%	1,14%	89,40%	5,80%
	Correlação	-0,02	0,36	-0,18	0,01	0,37	0,96
	Envelopamento	50,87%		96,45%		93,62%	

Quadro 5: Resumo das estatísticas da qualidade da previsão pontual e do envelopamento das sucessões – modelo AR(1).

MODELO SIMULADO	PREVISÃO ESTÁTICA FEITA COM MODELO ARIMA						
	Estatística de Análise	Sem pré-processamento		Com pré-processamento em 1 nível		Com pré-processamento em 2 níveis	
		Previsão de 200 pontos	Previsão de 10 pontos	Previsão de 200 pontos	Previsão de 10 pontos	Previsão de 200 pontos	Previsão de 10 pontos
ARIMA	TIC	1,00	0,95	1,00	0,90	0,97	0,91
	MAPE	99,09%	9,54%	100%	10,90%	99,68%	9,19%
	Correlação	0,18	0,22	-0,05	-0,34	0,11	-0,36
	Envelopamento	45,54%		92,67%		90,96%	

Quadro 6: Resumo das estatísticas da qualidade da previsão pontual e do envelopamento das sucessões – modelo bilinear com rede neural e retropropagação.

MODELO SIMULADO	PREVISÃO ESTÁTICA FEITA COM MODELO BILINEAR COM REDE NEURAL – Retropropagação						
	Estatística de Análise	Sem pré-processamento		Com pré-processamento em 1 nível		Com pré-processamento em 2 níveis	
		Previsão de 200 pontos	Previsão de 10 pontos	Previsão de 200 pontos	Previsão de 10 pontos	Previsão de 200 pontos	Previsão de 10 pontos
BILINEAR	TIC	1,00	0,82	0,63	0,68	0,62	0,71
	MAPE	100%	9,60%	342,27%	81,86%	346,47%	86,93%
	Correlação	0,03	0,36	-0,03	-0,66	-0,07	-0,50
	Envelopamento	40,54%		90,35%		91,11%	

Quadro 7: Resumo das estatísticas da qualidade da previsão pontual e do envelopamento das sucessões – modelo bilinear com redes recorrentes.

MODELO SIMULADO	PREVISÃO ESTÁTICA FEITA COM MODELO BILINEAR COM REDE NEURAL RECORRENTE						
	Estatística de Análise	Sem pré-processamento		Com pré-processamento em 1 nível		Com pré-processamento em 2 níveis	
		Previsão de 200 pontos	Previsão de 10 pontos	Previsão de 200 pontos	Previsão de 10 pontos	Previsão de 200 pontos	Previsão de 10 pontos
BILINEAR	TIC	0,49	0,68	0,52	0,24	0,85	0,65
	MAPE	146,08%	113,48%	91,15%	128,70%	199,69%	118,56%
	Correlação	0,43	0,51	0,32	0,91	0,16	-0,41
	Envelopamento	92,47%		96,38%		95,29%	

4.5 TESTE PARA O IBOVESPA

A série do IBOVESPA, apresentada na Figura 70, correspondente aos valores nominais de fechamento do índice da Bolsa de Valores de São Paulo, no período de 03/01/2000 até 29/10/2004, é claramente não estacionária. Nota-se a presença de oscilações em determinados períodos, como por exemplo, em setembro de 2001 (atentados terroristas aos Estados Unidos) e outubro de 2002.

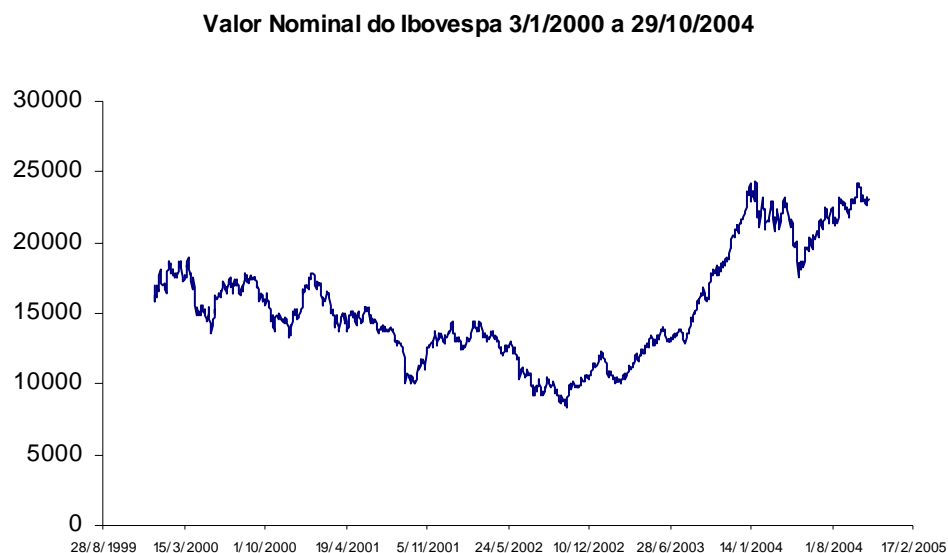


Figura 70: Valores nominais do IBOVESPA diário em pontos no período de 03/01/2000 a 29/10/2004.

Fonte: Economática.

Aplicando a primeira diferença ao logaritmo da série, obtemos a série de retornos do IBOVESPA para dados diários que, nesse caso, chamamos de DLIBOVESPA.

O comportamento da série DLIBOVESPA é apresentado na Figura 71. Pode-se identificar, nesta figura, que flutuações de mesma magnitude do retorno tendem a se agrupar, ou seja, grandes variações tendem a ser seguidas por outras grandes variações, o mesmo acontecendo para o caso de pequenas variações, sendo seguidas por outras variações.

Retorno do Ibovespa - 01/01/2000 a 30/9/2004

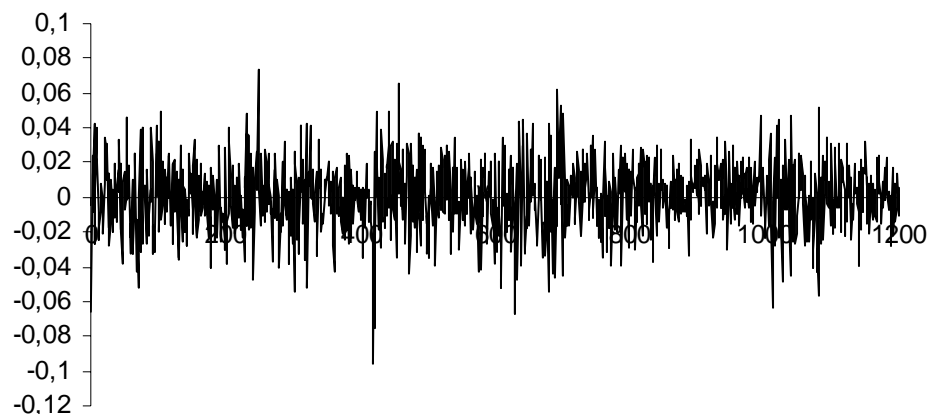


Figura 71: Log-Retornos do IBOVESPA diário no período de 03/01/2000 a 29/10/2004.

Para realizar as análises de desempenho de previsão de cada um dos métodos estudados, a série foi dividida em duas partes: uma para aprendizagem, correspondendo a 1000 valores e uma parcela para previsão no total de 200 valores.

Verificando a sucessão cronológica do DLIBOVESPA como um todo, tem-se:

Teste de raiz unitária

Ibovespa		
Hipóteses	Estatística de Teste	Valores Críticos a 5%
$H_0 : 3RU$ $H_1 : 2RU$	-68,24733	-2,76
$H_0 : 2RU$ $H_1 : 1RU$	-22,38750 -38,46974	-3,37 -3,77
$H_0 : 1RU$ $H_1 : 0RU$	-31,44229 -14,39437 -18,01330	-3,42 -3,80 -4,16

Considerando um nível de confiança de 95%, se rejeita a todas as hipóteses nulas, o que indica que ambas as subséries são estacionárias.

Teste de independência estatística – BDS

IBOVESPA				
Dimensão	BDS	Erro Padrão	Estatística Z	p-valor
2	0,001521	0,000486	3,132866	0,0017
3	0,005164	0,001038	4,975060	0,0000
4	0,009762	0,001660	5,879685	0,0000
5	0,015237	0,002324	6,557137	0,0000
6	0,020615	0,003008	6,852887	0,0000

Calculou-se até a sexta dimensão e concluiu-se por rejeitar a hipótese nula ao nível de significância de 5%, indicando que os dados da sucessão cronológica são dependentes, baseados na análise do p-valor, sendo que pelo menos um dos p-valores deve ser menor que 5%.

Teste de Normalidade

A figura, a seguir, ilustra a distribuição da sucessão cronológica.

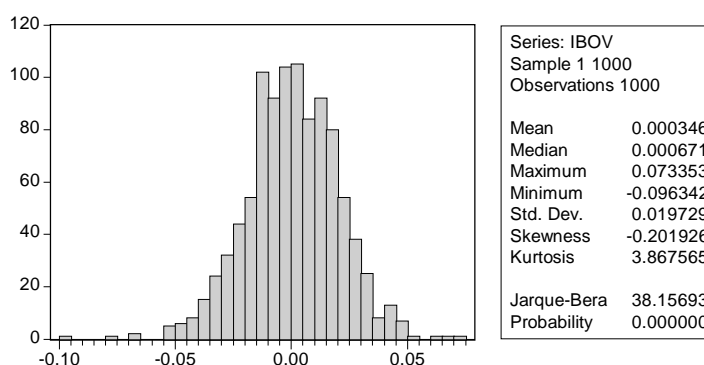


Figura 72: Distribuição do IBOVESPA

Observa-se, pelo p-valor, que a distribuição da sucessão cronológica IBOVESPA não segue uma distribuição normal pela rejeição da hipótese nula. Observa-se excesso de curtose e assimetria negativa.

Teste de McLeod-Li

O resultado da estatística de teste calculado no Eviews® 5.0 é:

Modelo IBOVESPA			
Estatística F	11,01762	p-valor	0,00000

Dessa forma, o teste de McLeod-Li para 5 defasagens rejeitou a hipótese nula pelo fato do p-valor ser inferior a 5%. Isto indica que a sucessão cronológica do modelo IBOVESPA possui estatisticamente comportamento não-linear.

Teste de Hsieh

Uma vez estabelecido que existe algum tipo de não-linearidade em uma sucessão cronológica, é preciso identificar o tipo de não-linearidade existente, se na média ou na variância condicionadas.

Para a realização do teste, fez-se o programa 4 que calcula os coeficientes (i,j) para cada uma das subséries. Os resultados desta etapa vêm a seguir:

IBOVESPA		
Coeficiente	Estatística de Hsieh	p-valor
1 1	-0,1171	0,4534
1 2	0,7191	0,7640
1 3	0,1492	0,5593
1 4	0,1298	0,5516
1 5	-0,0195	0,4922
2 2	0,0744	0,5297
2 3	-0,5456	0,2927
2 4	0,3259	0,6277
2 5	-0,2027	0,4197
3 3	0,2956	0,6162
3 4	-0,6942	0,2438
3 5	-0,9021	0,1835
4 4	0,4649	0,6790
4 5	0,1655	0,5657
5 5	-0,0273	0,4891

Conforme se observa, no quadro anterior, para quinze pares (i,j) , os valores dos coeficientes do teste de Hsieh amostrais para o IBOVESPA e o correspondente p-valor para um nível de confiança de 95%, podendo-se dizer que não se rejeita a hipótese nula dos coeficientes iguais a zero para nenhum valor, o que indica uma não-linearidade na variância condicionada.

E seguindo o fluxograma de análise e testes, foi escolhida novamente a ondaleta de Daubesch número 1 (DB1). A figura, a seguir, ilustra a decomposição (pré-processamento) em primeiro nível da série no nível original:

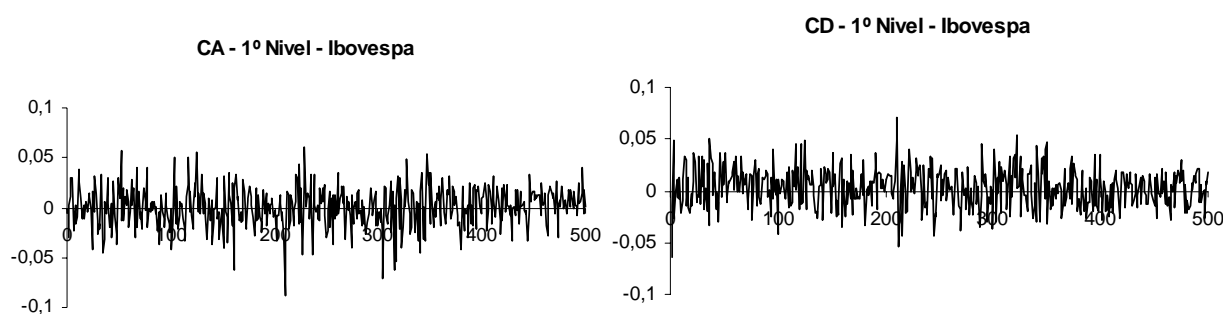


Figura 73: Pré-processamento em 1º nível para o IBOVESPA.

Teste de raízes unitárias

Subsérie: CA			Subsérie: CD		
Hipóteses	Estatística de Teste	Valores Críticos a 5%	Hipóteses	Estatística de Teste	Valores Críticos a 5%
$H_0 : 3RU$ $H_1 : 2RU$	-50,35651	-2,76	$H_0 : 3RU$ $H_1 : 2RU$	-52,92045	-2,76
$H_0 : 2RU$ $H_1 : 1RU$	-16,23697 -26,49827	-3,37 -3,77	$H_0 : 2RU$ $H_1 : 1RU$	-16,55276 -27,39209	-3,37 -3,77
$H_0 : 1RU$ $H_1 : 0RU$	-23,13910 -9,200660 -13,32963	-3,42 -3,80 -4,16	$H_0 : 1RU$ $H_1 : 0RU$	-22,69907 -10,56710 -12,47816	-3,42 -3,80 -4,16

Considerando um nível de confiança de 95%, rejeitam-se todas as hipóteses nulas, o que indica que ambas as subséries são estacionárias.

Teste para Independência Estatística – BDS

Subsérie: CA					Subsérie: CD				
<u>Dimensão</u>	<u>BDS</u>	<u>Erro Padrão</u>	<u>Estatística Z</u>	<u>p-valor</u>	<u>Dimensão</u>	<u>BDS</u>	<u>Erro Padrão</u>	<u>Estatística Z</u>	<u>p-valor</u>
2	5,32E-05	2,98E-06	17,82401	0,0000	2	5,53E-05	3,26E-06	16,98216	0,0000
3	-9,89E-07	8,18E-08	-12,09364	0,0000	3	-1,08E-06	9,16E-08	-11,82881	0,0000
4	-9,76E-09	1,68E-09	-5,791813	0,0000	4	-1,10E-08	1,93E-09	-5,682562	0,0000
5	-9,85E-11	3,04E-11	-3,240053	0,0012	5	-1,14E-10	3,58E-11	-3,194547	0,0014
6	-9,73E-13	5,09E-13	-1,912310	0,0008	6	-1,18E-12	6,14E-13	-1,923652	0,0044

O teste de BDS rejeita a hipótese de independência das subséries, com 95% de confiança, o que também implica na existência de não-linearidade.

Teste de normalidade

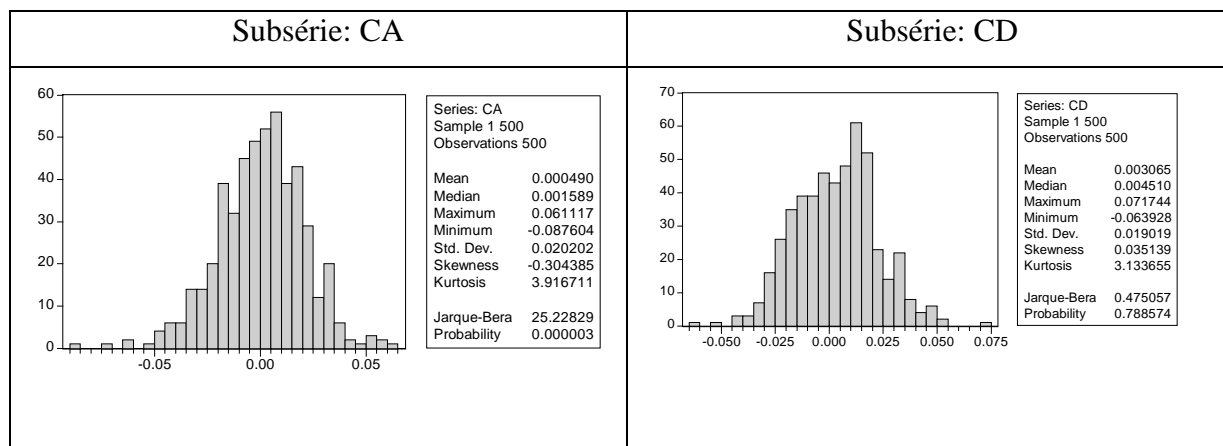


Figura 74: Distribuição das subséries para o IBOVESPA.

Rejeita-se a hipótese de normalidade pelo teste de Jarque-Bera com 5% de significância para a subsérie de alta frequência (CD) e não se rejeita para a subsérie de baixa frequência.

Teste de McLeod-Li

Subséries: CA				Subséries: CD			
Estatística F	15,58455	p-valor	0,000000	Estatística F	5,530925	p-valor	0,0000000

Dessa forma, o resultado do teste de McLeod-Li para 5 defasagens rejeitou a hipótese nula pelo fato do p-valor ser inferior a 5%. Isto indica que ambas as subséries estatisticamente possuem comportamento não-linear.

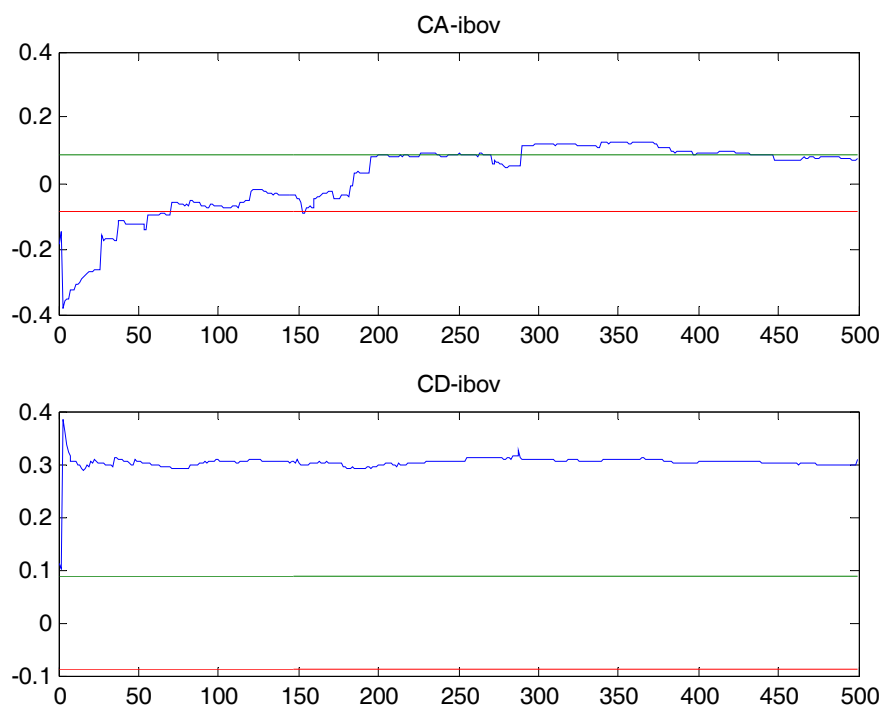


Figura 75: Comportamento da série ao longo do tempo para o IBOVESPA em 1 níveis de pré-processamento.

Nota-se que nas duas decomposições, ocorreu a não-linearidade ao longo do comportamento das subséries.

Teste de Hsieh

Uma vez estabelecido que existe algum tipo de não-linearidade em uma sucessão cronológica, é preciso filtrar o tipo de não-linearidade existente, se na média ou na variância.

Subsérie: CA			Subsérie: CD		
Coeficiente	Estatística de Hsieh	p-valor	Coeficiente	Estatística de Hsieh	p-valor
1 1	0,0199	0,5079	1 1	-0,7673	0,2215
1 2	-0,3985	0,3451	1 2	12,2915	1,0000
1 3	-0,5165	0,3028	1 3	5,8233	1,0000
1 4	-2,5064	0,00618	1 4	-1,0921	0,1374
1 5	1,1543	0,8758	1 5	-1,3654	0,0861
2 2	0,6902	0,7550	2 2	1,3649	0,9139
2 3	1,9596	0,9750	2 3	-1,1688	0,1212
2 4	0,3939	0,6532	2 4	-3,7453	0,0001*
2 5	-0,3271	0,3718	2 5	-1,4606	0,0721
3 3	-0,2684	0,3942	3 3	14,5133	1,0000

Subsérie: CA			Subsérie: CD		
3 4	0,1726	0,5685	3 4	3,3806	0,9996
3 5	0,045	0,5179	3 5	-2,4782	0,0066*
4 4	0,4792	0,6841	4 4	0,7403	0,7704
4 5	-1,3619	0,0866	4 5	2,6633	0,9961
5 5	-0,3465	0,3645	5 5	0,478	0,6837

Conforme se observa no quadro anterior, para quinze pares i,j , os valores dos coeficientes do teste de Hsieh amostrais para as duas subséries e o correspondente p-valor para um nível de confiança de 95%, aceita-se a hipótese nula dos coeficientes iguais a zero, o que indica uma não-linearidade na variância condicionada. A subsérie de alta frequência tem apenas dois coeficientes significativos marcados com (*).

Análise dos parâmetros estimados para o modelo AR(||7||)-GARCH(1,1)

$$CA_t = \phi_7 CA_t + \varepsilon_t$$

$$\varepsilon_t^2 | I_{t-1} \sim N(0; h_t)$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta h_{t-1}$$

Parâmetros	Coeficiente	Erro-padrão	Estatística z	Significância
ϕ_7	0,111503	0,046572	2,394198	0,0167
α_0	3,14E-05	2,80E-05	1,122401	0,2617
α_1	0,037427	0,030131	1,242159	0,2142
β	0,884750	0,096367	9,181047	0,0000

Análise dos parâmetros estimados para o modelo AR(||9||)-GARCH(1,1) da subsérie CD.

$$CD_t = \phi_9 CD_t + \varepsilon_t$$

$$\varepsilon_t^2 | I_{t-1} \sim N(0; h_t)$$

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta h_{t-1}$$

Parâmetros	Coeficiente	Erro-padrão	Estatística z	Significância
ϕ_9	-0,030409	0,042890	-0,708995	0,4783
α_0	8,91E-05	5,61E-05	1,589941	0,1118
α_1	0,105876	0,041375	2,558965	0,0105
β	0,645691	0,171914	3,755889	0,0002

Estimando-se os valores e calculando as estimativas de erros pontuais para a previsão em 1 nível de pré-processamento, via redes neurais, com algoritmo de retropropagação, tem-se:

IBOVESPA – MODELO ARIMA-GARCH			
Estatísticas de Previsão Pontual para os 200 pontos	Valor	Estatísticas de Previsão Pontual para os 10 primeiros pontos	Valor
TIC	1,00	TIC	0,93
MAPE	100,00%	MAPE	98,82%
Correlação	-0,03	Correlação	-0,05

Trabalhando com redes neurais com retropropagação e com redes recorrentes, a rede recorrente foi gerada com 10-2R-1, com função de ativação do tipo sigmóide, taxa de aprendizagem de 0,005 e número de ciclos de treinamento de 20000. Para as redes com retropropagação, tem-se:

Subsérie	Redes	Performance de treinam.	Performance Seleção	Performance Teste	Erro Treinamento	Erro Seleção	Erro Teste	Características
CA	MLP s5 1:5-1-1:1	1,007509	0,997100	1,003204	0,16006	0,16454	0,14668	BP100,CG21b
CD	MLP s5 1:5-1-1-1:1	1,001132	0,987027	1,002208	0,15449	0,13768	0,14730	BP100,CG22b

Efetuada o pós-processamento e calculando-se as estatísticas pontuais, tem-se:

IBOVESPA – REDES NEURAIIS com retropropagação			
Estatísticas de Previsão Pontual para os 200 pontos	Valor	Estatísticas de Previsão Pontual para os 10 primeiros pontos	Valor
TIC	0,85	TIC	0,87
MAPE	131,89%	MAPE	91,60%
Correlação	0,09	Correlação	0,06

IBOVESPA – REDES NEURAIIS recorrentes			
Estatísticas de Previsão Pontual para os 200 pontos	Valor	Estatísticas de Previsão Pontual para os 10 primeiros pontos	Valor
TIC	0,99	TIC	0,99
MAPE	207,48%	MAPE	99,76%
Correlação	0,07	Correlação	0,20

Fazendo o cálculo do envelopamento para a própria série do IBOVESPA e verificando quais se enquadram dentro do intervalo de confiança de 95%, para a série prevista, via procedimento estático, tem-se:

Envelopamento – modelo	Percentual
IBOVESPA por Redes recorrentes	66%
IBOVESPA por Redes com retropropagação	28%
IBOVESPA por ARIMA-GARCH	5%

Simulação no 2º Nível do Pré-Processamento

Apresenta-se, a seguir, o gráfico da decomposição em 2 níveis do DGP para o IBOVESPA:

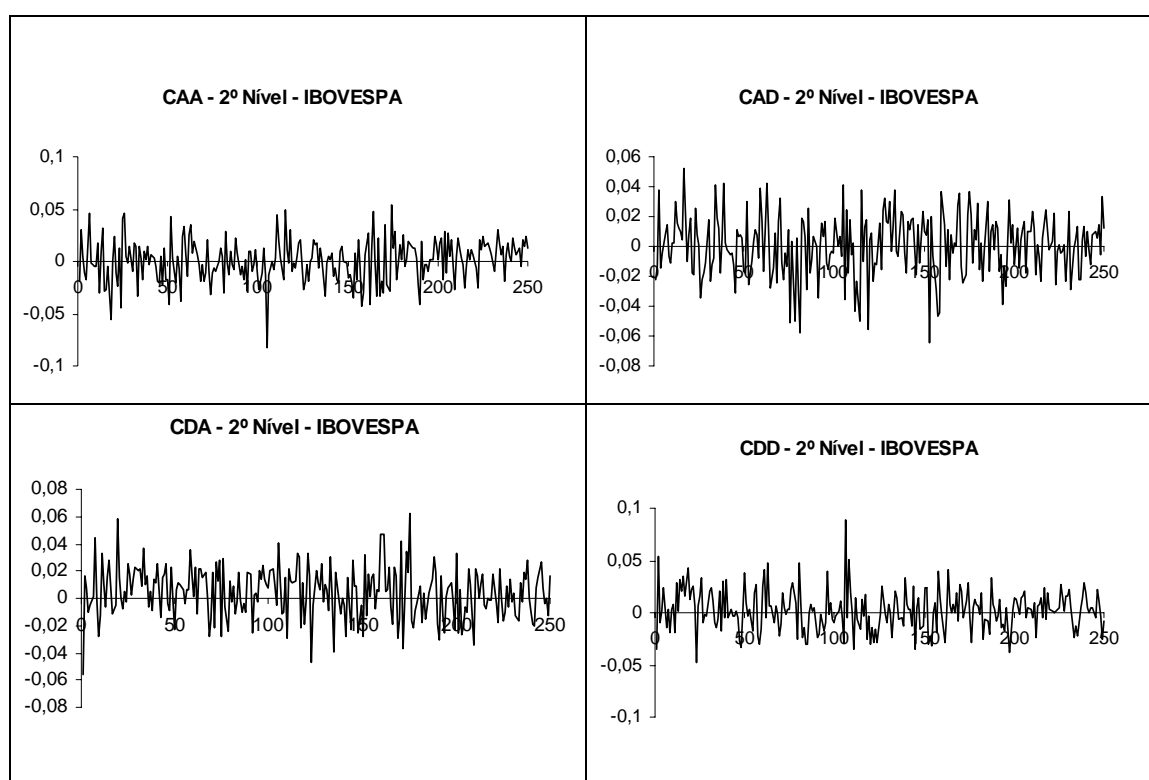


Figura 76: Pré-processamento em 2º nível para o IBOVESPA.

Seguindo o mesmo roteiro do modelo anterior, apresentam-se, a seguir, os resultados obtidos na tabela a seguir:

Teste de Raízes Unitárias

Subsérie: CAA			Subsérie: CAD		
Hipóteses	Estatística de Teste	Valores Críticos a 5%	Hipóteses	Estatística de Teste	Valores Críticos a 5%
$H_0 : 3RU$ $H_1 : 2RU$	-31,40144	-2,76	$H_0 : 3RU$ $H_1 : 2RU$	-35,64922	-2,76
$H_0 : 2RU$ $H_1 : 1RU$	-9,646718 -21,86531	-3,37 -3,77	$H_0 : 2RU$ $H_1 : 1RU$	-11,21028 -19,33772	-3,37 -3,77
$H_0 : 1RU$ $H_1 : 0RU$	-13,54963 -8,558362 -8,118091	-3,42 -3,80 -4,16	$H_0 : 1RU$ $H_1 : 0RU$	-15,79216 -6,904285 -9,062377	-3,42 -3,80 -4,16

Teste de Raízes Unitárias

Subsérie: CDA			Subsérie: CDD		
Hipóteses	Estatística de Teste	Valores Críticos a 5%	Hipóteses	Estatística de Teste	Valores Críticos a 5%
$H_0 : 3RU$ $H_1 : 2RU$	-32,99632	-2,76	$H_0 : 3RU$ $H_1 : 2RU$	-35,93773	-2,76
$H_0 : 2RU$ $H_1 : 1RU$	-10,73305 -20,15705	-3,37 -3,77	$H_0 : 2RU$ $H_1 : 1RU$	-12,35845 -17,62230	-3,37 -3,77
$H_0 : 1RU$ $H_1 : 0RU$	-14,87379 -7,917262 -8,511772	-3,42 -3,80 -4,16	$H_0 : 1RU$ $H_1 : 0RU$	-16,68212 -7,021329 -8,819183	-3,42 -3,80 -4,16

Observa-se que, com 95% de confiança, ambas as quatro subséries definidas no pré-processamento são estatisticamente estacionárias.

Teste de Normalidade

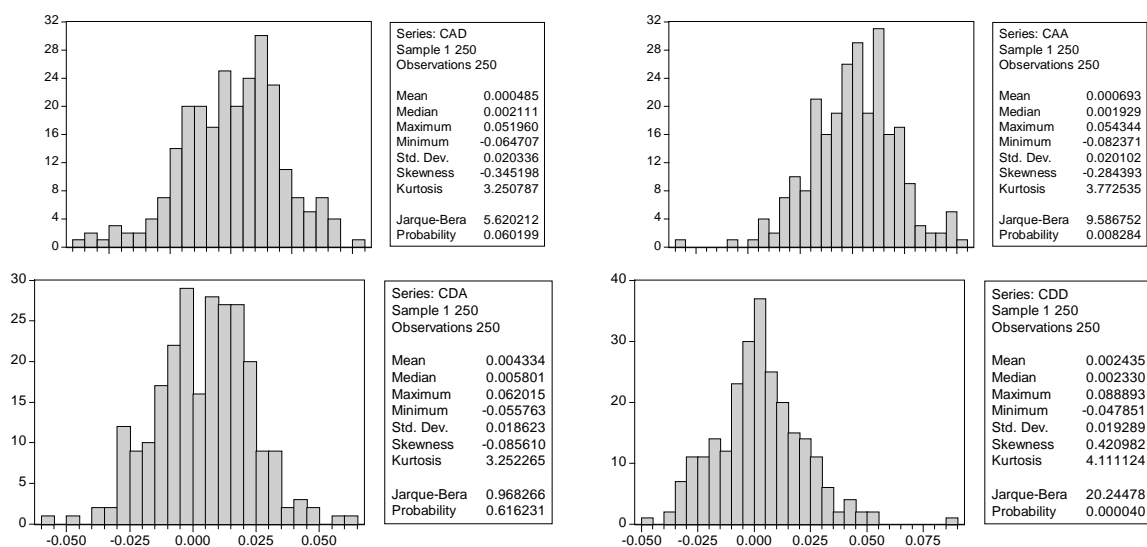


Figura 77: Distribuição em 2º nível para o pré-processamento do IBOVESPA.

O mesmo ocorre com o teste de normalidade dos dados, o que indica que as subséries de alta frequência da primeira decomposição e a baixa frequência também da primeira decomposição apresentam comportamento normal, as demais apresentam comportamento não normal.

Teste para independência estatística – BDS

Subsérie: CAA					Subsérie: CAD				
Dimensão	BDS	Erro Padrão	Estatística Z	p-valor	Dimensão	BDS	Erro Padrão	Estatística Z	p-valor
2	0.000687	0.000250	2.742139	0.0061	2	-0.000627	0.000241	-2.598822	0.0094
3	0.000328	6.06E-05	5.422063	0.0000	3	-0.000139	5.62E-05	-2.467741	0.0136
4	0.000155	1.10E-05	14.06307	0.0000	4	-6.06E-05	9.81E-06	-6.182160	0.0000
5	-1.11E-05	1.75E-06	-6.368636	0.0000	5	-9.01E-06	1.50E-06	-5.999156	0.0000
6	-1.13E-06	2.57E-07	-4.386083	0.0000	6	-8.67E-07	2.13E-07	-4.079398	0.0000

Subsérie: CDA					Subsérie: CDD				
Dimensão	BDS	Erro Padrão	Estatística Z	p-valor	Dimensão	BDS	Erro Padrão	Estatística Z	p-valor
2	0.000618	6.51E-05	9.492246	0.0000	2	0.000167	9.55E-05	1.753199	0.0196
3	7.34E-05	7.95E-06	9.231371	0.0000	3	-3.49E-05	1.20E-05	-2.920055	0.0035
4	-6.05E-06	7.29E-07	-8.302491	0.0000	4	-6.65E-06	1.13E-06	-5.900760	0.0000
5	-2.96E-07	5.86E-08	-5.046648	0.0000	5	-3.42E-07	9.31E-08	-3.670094	0.0002
6	-1.43E-08	4.35E-09	-3.290212	0.0010	6	-1.77E-08	7.13E-09	-2.484924	0.0130

O teste de independência rejeita a hipótese nula de independência das observações.

Teste de McLeod-Li

Subsérie: CAA				Subsérie: CAD			
Estatística F	4,25148	p-valor	0,00000	Estatística F	3,35478	p-valor	0,0005
Subsérie: CDA				Subsérie: CDD			
Estatística F	2,1244	p-valor	0,00150	Estatística F	3,21010	p-valor	0,00012

Pelo teste, rejeita-se a hipótese nula, o que indica que o modelo é estatisticamente não-linear.

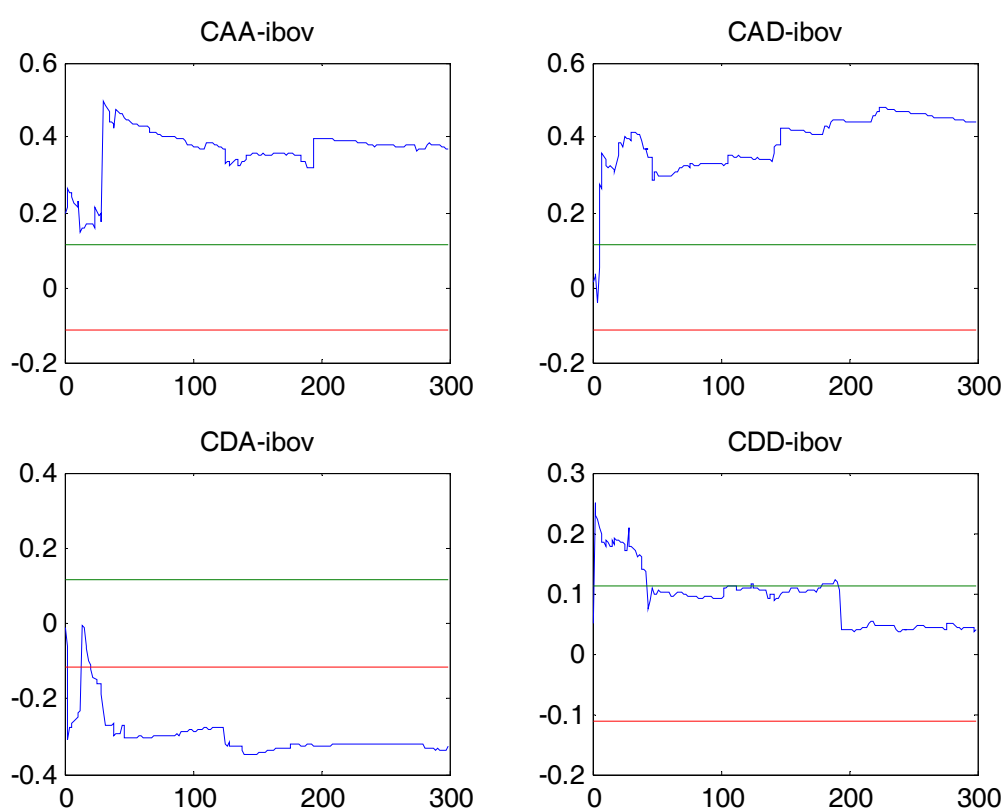


Figura 78: Comportamento da série, ao longo do tempo, para o IBOVESPA em 2 níveis de pré-processamento.

Observa-se que somente a subsérie de alta frequência da decomposição da própria subsérie de alta frequência do 1º nível apresenta comportamento linear, depois de um certo período. As demais apresentam claramente comportamento não-linear.

Estimando-se os valores e calculando as estimativas de erros pontuais para a previsão em 2 níveis de pré-processamento, por meio de redes neurais com retropropagação, tem-se que as redes recorrentes foram geradas com a arquitetura 10-2R-1. Estimando-se os resultados por redes neurais com retropropagação, tem-se, a seguir, um quadro resumindo o tipo de rede utilizado e as respectivas performances de seleção, teste e validação.

Subséries	Redes	Performance de treinam.	Performance Seleção	Performance Teste	Erro Treinamento	Erro Seleção	Erro Teste	Características
CAA	MLP s5 1:5-1-1:1	0,997355	0,998706	0,998381	0,15345	0,15632	0,14804	BP100,CG21b
CAD	MLP s5 1:5-8-1:1	1,005360	1,029760	1,013637	0,19837	0,18489	0,19656	BP8b
CDA	MLP s5 1:5-1-1:1	0,992644	1,001492	1,007899	0,16622	0,18870	0,19275	BP100,CG21b
CDD	MLP s5 1:5-1-2-1:1	0,983534	0,972308	1,035553	0,14664	0,11320	0,14462	BP100,CG23b

As estatísticas de qualidade das previsões foram:

MODELO IBOVESPA - RN com retropropagação			
Estatísticas de Previsão Pontual para os 200 pontos	Valor	Estatísticas de Previsão Pontual para os 10 primeiros pontos	Valor
TIC	0,73	TIC	0,78
MAPE	175,35%	MAPE	102,03%
Correlação	0,16	Correlação	0,30

Estimando-se os valores e calculando as estimativas de erros pontuais para a previsão em 2 níveis de pré-processamento por meio de redes neurais recorrentes, tem-se:

MODELO IBOVESPA - RN recorrente			
Estatísticas de Previsão Pontual para os 200 pontos	Valor	Estatísticas de Previsão Pontual para os 10 primeiros pontos	Valor
TIC	0,99	TIC	0,98
MAPE	102,52%	MAPE	208,34%
Correlação	0,13	Correlação	0,27

Fazendo o cálculo do envelopamento para o IBOVESPA, com 2 níveis de pré-processamento através das fórmulas:

$$NDLIBOVESPA_t = \frac{DLIBOVESPA_t - \min(DLIBOVESPA_t)}{\max(DLIBOVESPA_t) - \min(DLIBOVESPA_t)}$$

$$DLIBOVESPA_t = NDLIBOVESPA_t \left[\max(DLIBOVESPA_t) - \min(DLIBOVESPA_t) \right] + \min(DLIBOVESPA_t)$$

$$\ln(IBOVESPA_t) - \ln(IBOVESPA_{t-1}) = NDLIBOVESPA_t \left[\max(DLIBOVESPA_t) - \min(DLIBOVESPA_t) \right] + \min(DLIBOVESPA_t)$$

$$\ln(IBOVESPA_t) = NDLIBOVESPA_t \left[\max(DLIBOVESPA_t) - \min(DLIBOVESPA_t) \right] + \min(DLIBOVESPA_t) + \ln(IBOVESPA_{t-1})$$

$$IBOVESPA_t = e^{NDLIBOVESPA_t \left[\max(DLIBOVESPA_t) - \min(DLIBOVESPA_t) \right] + \min(DLIBOVESPA_t) + \ln(IBOVESPA_{t-1})}$$

Envelopamento – modelo	Percentual
IBOVESPA - Rede Neural Recorrente	83%
IBOVESPA – Rede Neural Retropropagação	34%
IBOVESPA – ARIMA-GARCH	7%

Embora o envelopamento tenha dado 83%, isso não significa que a previsão seja adequada, conforme as estatísticas de acurácia pontuais.

Teste para Independência Estatística dos Resíduos– Teste BDS

Como visto anteriormente, o teste para independência deve ser aplicado para ver se a sucessão cronológica apresenta independência nos resíduos. Suas hipóteses são:

H_0 :os resíduos são estocasticamente independentes

H_1 : os resíduos são estocasticamente dependentes

Resíduos IBOVESPA – Previsto RN com retrop. com 1 Nível de pré-processamento					Resíduos IBOVESPA – Previsto RN com retrop. com 2 Níveis de pré-processamento				
Dimensão	BDS	Erro Padrão	Estatística Z	p-valor	Dimensão	BDS	Erro Padrão	Estatística Z	p-valor
2	0.014896	0.005151	2.891971	0.0888	2	0.01667	4.80E-05	34.72782	0.0950
3	0.018453	0.008213	2.246835	0.0747	3	0.03333	0.000107	31.09772	0.0814
4	0.020576	0.009810	2.097390	0.0560	4	0.05000	0.000179	27.90881	0.0762
5	0.019249	0.010256	1.876818	0.0605	5	0.06667	0.000262	25.44965	0.0774
6	0.017827	0.009920	1.796982	0.0723	6	0.08333	0.000354	23.52168	0.0969

Resíduos IBOVESPA – Previsto RN recorrentes com 1 Nível de pré-processamento					Resíduos IBOVESPA – Previsto RN recorrentes com 2 Níveis de pré-processamento				
<u>Dimensão</u>	<u>BDS</u>	<u>Erro Padrão</u>	<u>Estatística Z</u>	<u>p-valor</u>	<u>Dimensão</u>	<u>BDS</u>	<u>Erro Padrão</u>	<u>Estatística Z</u>	<u>p-valor</u>
2	0.002820	0.004442	0.634819	0.5255	2	-5.05E-07	5.33E-07	-0.947302	0.3435
3	0.010294	0.007040	1.462196	0.1437	3	-3.70E-10	6.89E-09	-0.053687	0.9572
4	0.014577	0.008359	1.743854	0.0812	4	-2.77E-13	6.68E-11	-0.004142	0.9967
5	0.014577	0.008687	1.678068	0.0933	5	-2.11E-16	5.66E-13	-0.000373	0.9997
6	0.013164	0.008352	1.576136	0.1150	6	-1.64E-19	4.45E-15	-3.70E-05	1.0000

Resíduos IBOVESPA – Previsto ARIMA-GARCH com 1 Nível de pré-processamento					Resíduos IBOVESPA – Previsto ARIMA-GARCH com 2 Níveis de pré-processamento				
<u>Dimensão</u>	<u>BDS</u>	<u>Erro Padrão</u>	<u>Estatística Z</u>	<u>p-valor</u>	<u>Dimensão</u>	<u>BDS</u>	<u>Erro Padrão</u>	<u>Estatística Z</u>	<u>p-valor</u>
2	-1.74E-06	9.65E-07	-1.803930	0.0712	2	-2.32E-06	8.86E-07	-2.616872	0.0891
3	-2.37E-09	1.39E-08	-0.170852	0.8643	3	-3.64E-09	1.31E-08	-0.277813	0.7812
4	-3.29E-12	1.49E-10	-0.022009	0.9824	4	-5.83E-12	1.45E-10	-0.040131	0.9680
5	-4.66E-15	1.41E-12	-0.003304	0.9974	5	-9.53E-15	1.41E-12	-0.006757	0.9946
6	-6.75E-18	1.23E-14	-0.000547	0.9996	6	-1.59E-17	1.27E-14	-0.001255	0.9990

Calculou-se a estatística BDS para todas as dimensões de dois até o valor especificado, variando a dimensão de correlação para maximizar o poder do teste. Neste caso, calculou-se até a sexta dimensão e concluiu-se por não se rejeitar a hipótese nula ao nível de significância de 5%, indicando que os dados dos resíduos do IBOVESPA são independentes para os modelos gerados baseados na análise do p-valor. Pode-se concluir que os três modelos são plausíveis estatisticamente.

Teste de Normalidade

Na figura a seguir, tem-se as estatísticas dos resíduos do pós-processamento onde as hipóteses nulas e alternativas são:

H_0 : os resíduos seguem uma distribuição normal

H_1 : os resíduos não seguem uma distribuição normal

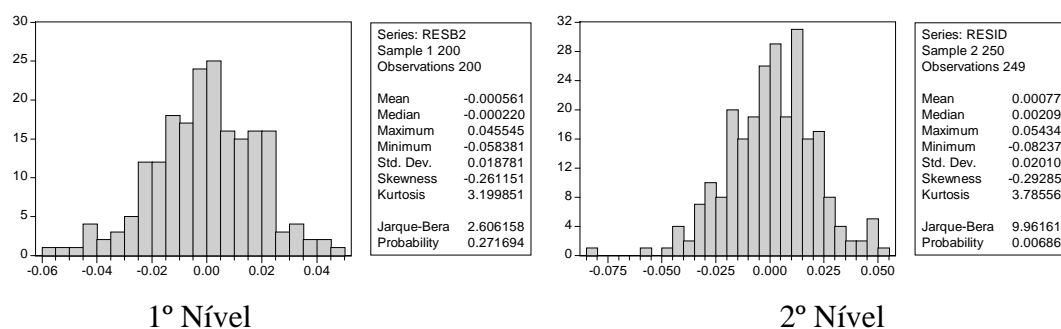


Figura 79: Distribuição dos resíduos do IBOVESPA previsto por redes neurais com retropropagação, em 1 e 2 níveis de pré-processamento.

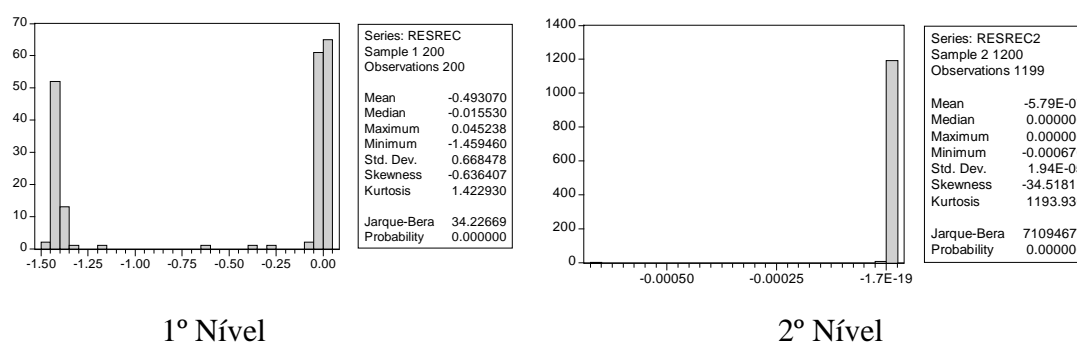


Figura 80: Distribuição dos resíduos do modelo IBOVESPA previsto por redes neurais recorrentes, em 1 e 2 níveis de pré-processamento.

Observa-se que os dados dos resíduos não são normais pela estatística de Jarque-Bera.

CONCLUSÃO

Nesta pesquisa, o objetivo principal foi o de explorar a possibilidade de usar uma metodologia capaz de decompor uma sucessão cronológica via ondaletas, conjuntamente com os modelos econométricos e de redes neurais já existentes de previsão e comparar a qualidade de previsões obtidas para sucessões cronológicas lineares e não-lineares simuladas. Teve também a proposta de traçar um fluxograma para tratamento das previsões e para propor um rigor quantitativo mais adequado.

O diferencial deste trabalho está na realização das previsões dentro das subséries decompostas por uma ondaleta em até dois níveis, obtendo-se a previsão da série original, via reconstrução da série para modelos construídos, por processos geradores de dados de sucessões cronológicas lineares e não-lineares.

Foram simuladas sucessões cronológicas de um processo ARIMA-GARCH, um ARIMA, um bilinear e um movimento browniano geométrico. O trabalho principal constituiu-se na elaboração da fase de pré-processamento e das previsões estáticas, em separado para cada uma das subséries encontradas, sendo feitas para 10 e 200 observações futuras. Além das previsões pontuais, foi verificado também o envelopamento dos dados, que consiste em comparar o modelo de previsão através de uma banda de confiança, para os valores previstos em 200 sucessões cronológicas simuladas. Considerou-se que um envelopamento significativo foi superior a 66%, descartando os demais.

No caso da simulação feita para um modelo ARIMA(1,0,0)-GARCH(1,1), pode-se observar que o pré-processamento pela ondaleta foi melhor para apenas uma etapa de separação de altas e baixas frequências, tanto pela correlação quanto pelos critérios do TIC, sendo este reduzido, e pelo MAPE menor para as previsões de curto prazo.

Para os modelos de previsão com redes neurais, ressalta-se uma diferença entre as redes neurais recorrentes e as redes com algoritmo de retropropagação: é a capacidade de previsão das redes recorrentes para dados não-lineares com 2 níveis de pré-processamento e para previsões de curto prazo. Todavia, para o critério do envelopamento, o modelo ARIMA-GARCH apresentou melhores resultados para o pré-processamento com 1 nível e por rede

recorrente. Todos os dados também foram comparados com as previsões feitas sem pré-processamento, as quais se mostraram impróprias com MAPE, perto de 100% para previsões de longo prazo.

Na modelagem feita para a sucessão cronológica AR(1), o melhor modelo ajustado, segundo o TIC, foi com pré-processamento em uma etapa; mas as previsões são melhores com duas etapas. O envelopamento, para este processo, foi melhor com duas etapas de decomposição.

Também se verificaram, neste trabalho, as alterações que a mudança da escolha de uma ondaleta por outra poderia impactar nos resultados das previsões futuras. Constatou-se que a troca da forma de onda, no pré-processamento que se pareça mais visualmente com a forma dos dados da série, aumenta as medidas de acurácia, deixando evidências de que pode haver melhoras nos resultados.

Para a simulação do movimento browniano, salienta-se que, na subsérie de baixa frequência, o pré-processamento em um e dois níveis, reproduz a série original no nível, o que acaba por provocar a não separação das altas e baixas frequências, induzindo a sucessão cronológica ao critério de independência nas observações.

Com os dados do modelo bilinear, as redes recorrentes se mostraram mais adequadas tanto para as previsões de curto prazo, quanto para o envelopamento, muito embora, neste último caso, a diferença ficou pouco significativa na capacidade de envelopar.

Da análise dos resíduos resultantes dos modelos simulados, observou-se que, para os modelos ARIMA-GARCH e AR(1), seus valores foram normais e independentes. Já, nos dados do modelo bilinear, detectou-se apenas a independência dos resíduos e a não normalidade dos mesmos.

Testando os algoritmos para a série do IBOVESPA, também com 1200 observações, os resultados das previsões não foram satisfatórios, apresentaram altos índices de erros nas estatísticas pontuais e o envelopamento pouco significativo, se comparado com os obtidos, via simulação. Um fato verificado está na correlação dos dados, em que houve aumento no que diz respeito às previsões dos log-retornos da sucessão, não se podendo dizer, ainda, se o pré-

processamento favorece de forma significativa a previsão, mas sim ajuda a diminuir o erro no curto prazo.

Destacou-se também o aumento da correlação com o pré-processamento e o envelopamento ficou mais significativo em apenas um nível. As previsões estáticas são melhores e também apresentam redução do erro por intervalo em um nível de pré-processamento.

Outrossim, da análise deste trabalho, emerge a necessidade do fluxograma implementado para as previsões e o papel das previsões, em separado por ondaletas, como redutores dos erros nos processos estocásticos e da implementação das bandas de previsões, para redes recorrentes para sucessões cronológicas lineares e não-lineares. A capacidade das redes neurais de lidarem com sucessões não-lineares é uma vantagem, pois as previsões feitas com pré-processamento convalidam uma indicação dessa característica.

REFERÊNCIAS

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**. AC – 19., 1974. p. 716-723.

_____. Maximum likelihood identification of Gaussian autoregressive moving average models. **Biometrika**. v. 60, p. 255-265, 1973.

ALBUQUERQUE, J. R. **Previsibilidade de retorno de ações no mercado brasileiro**. Instituto de Ciências Sociais, 2003. 57 f. Dissertação (Mestrado em Economia) – Instituto de Ciências Sociais. Brasília: Universidade de Brasília, 2003.

AMARAL, L. A. N. et al. Scale-independent measures and pathologic cardiac dynamics. **Physical Review Letters**, New York: v. 81, n. 11, p. 2388-2392, S&Pt 1998.

ARMSTRONG, J. S., COLLOPY, F. Error measures for generalizing about forecasting methods: empirical comparisons. **International Journal of Forecasting**. n. 08, 1992. p. 69-80.

ARIÑO, M. A. Time series forecasts via ondaletas: an application to car sales in the spanish market. **Discussion Paper** 95-30, ISDS, Duke University.

ARMOSTRONG, J. S., BRODIE, R. J. **Quantitative methods in marketing**. 2. ed. London: International Thompson Business Press, 1999.

AZOFF, E. M. **Neural network time series forecasting of financial markets**. John Wiley & Sons. 1994.

BARTLETT, M. S. On the theoretical specification and sampling properties of autocorrelated time series. **Journal of the Royal Statistical Society**. B2, 1946, p. 27-41.

BIDARKOTA, P. V. The comparative forecast performance of univariate and multivariate models: an application to real interest rate forecasting. **International Journal of Forecasting**. n. 14. p. 457- 468, 1998.

BILLINGS, S. A.; VOON, W. S. F. Correlation based model validity tests for nonlinear models. **International Journal of Control**. v. 44, n. 1, p. 235-244, 1986.

BLACKMAN, R. B.; TUKEY, J. W. **The measurement of power spectrum from the point of view of communications engineering**, New York: Dover, 1959.

BOLLERSLEV, T. A conditional heteroscedasticity time series model for speculative process and rates of return. **Review of Economics and Statistics**, v. 69, p. 542-547, 1987.

BOLLERSLEV, T. Generalized autoregressive conditional heteroscedasticity. **Journal of Econometrics**. v. 31, p. 307-327, 1986.

BOLLERSLEV, T; CHOU, R.; KRONER, K. ARCH modeling in finance: a review of the theory and empirical evidence. **Journal of Econometrics**. v. 52, p. 5-59, 1992.

BOVESPA. **Mercado de capitais**. 1999. Disponível em: <<http://www.bovespa.com.br>>. Acesso em 22 mar. de 2003.

_____. **Índice Bovespa, definição e metodologia**. 2002. Disponível em: <<http://www.bovespa.com.br>> . Acesso em 22 mar. de 2003.

BOX, G. E. P; PIERCE, D. A. Distribution of autocorrelations in autoregressive moving average models. **Journal of the American Statistical Association**, v. 65,1970. p. 1509-1526, 1970.

BOX, G. E. P., JENKINS, G. M., REINSEL, G. C. **Time series analysis: forecasting and control**. 3. ed. New York : Prentice Hall, 1994.

BOX, G. E. P., JENKINS, G. M. **Time series analysis: forecasting e control**. 3. ed. New York: Prentice Hall, 1994.

BRITO, N. O.; DE MANAZES, J. C. F. A eficiência informacional fraca do mercado à vista na Bolsa de Valores do Rio de Janeiro no período 1973-1980. **Relatório de Pesquisa 22**. (Setembro), COPPEAD, UFRJ, 1981.

BRUNI, A. L., FAMÁ, R. Eficiência, previsibilidade dos preços, anomalias em mercados de capitais: teoria e evidências. **Caderno de Pesquisas em Administração**, São Paulo, v. 1, n. 7, p. 71-85, 1998.

CAMARGOS, M. A. de., BARBOSA, F. V. Teoria e evidência da eficiência informacional do mercado de capitais brasileiro. **Caderno de Pesquisas em Administração**. São Paulo, v. 10, n. 1, p. 41-55, 2003.

CAMPBELL, J.Y.; LO, A.W.; MACKINLAY, A. C. **The econometrics of financial markets**. New Jersey: Princeton University Press, 1997.

CAO, L.; SOOFI, A. S. Non-linear deterministic forecasting of daily dollar exchange rates. **International Journal of Forecasting**. Milwaukee, n. 15, p. 421–430, 1999.

CERETTA, P. S. Investigando a presença de Caos no Ibovespa. **Revista Eletrônica de Administração**, 29. v. 8, n. 5. Disponível em <<http://read.adm.ufrgs.br>>. Acesso em 19: mar. 2003.

CERETTA, P. S., COSTA JÚNIOR, N. C. A. Quantas ações tornam um portfólio diversificado no mercado de capitais brasileiro. In: **Mercado de capitais**. São Paulo: Atlas, 2000. cap. 1. p. 19-33.

CHANG, C. **Análise de ondaletas em séries temporais**. São Paulo: IME, 1997. 117 f. Tese (Doutorado em Matemática Aplicada) – Instituto de Matemática e Estatística, Universidade de São Paulo, 1997.

CHOW, G. C. Tests of equality between sets of coefficients in two linear regression. **Econometrica**. v. 28., n. 3, p. 591-605, 1960.

CHUI, C. K. **Ondaletas: a tutorial in theory and aplicattions**. Boston: Academic Press, 1992.

CORREA, M. M. R. da L.; PEREIRA, P. L. V. **Modelos não-lineares em finanças: previsibilidade em mercados financeiros e aplicações a gestão de risco**. 1997. Disponível em: <<http://www.ibmec.br/pvalls>> Acesso em: 10 dez. 2002.

COWLES, A. Can stock market forecasters forecast? **Econometrica** **I**. 1993. p. 309-324.

COSTA, P. H. S.; BAIDYA, T. K. N. Propriedades estatísticas das séries de retornos das principais ações brasileiras. **Pesquisa Operacional**, v. 21. n. 1. p. 38-78, 2001.

CROMWELL, J. B.; LABYS, W. C; TERRAZA, M. **Univariate tests for time series models**. Sage University Paper series on Quantitative Applications in the Social Sciences. n. 7-99. Thousand Oaks, CA: Sage.

DAMODARAN, A., BERNSTEIN, P. L. **Administração de investimentos**. Porto Alegre: Bookman, 2000.

DAUBECHIES, I. **Ten lectures on ondeletas**, Society for Industrial and Applied Mathematics. Philadelphia: Pennsylvania, 1992.

DE OLIVEIRA, M. A. **Previsão de sucessões cronológicas econômico-financeiras por meio de redes neurais artificiais recorrentes de tempo real e de processos ARMA-GARCH: um estudo comparativo quanto à eficiência de previsão**. Faculdade de Economia, Administração e Contabilidade. São Paulo, 2003. 171 f. Dissertação (Mestrado em Administração). Faculdade de Economia, Administração e Contabilidade: Universidade de São Paulo, 2003.

DEBOECK, G. J. Pre-processing and evaluation of neural nets for trading stocks. **Advanced Technology for Developers**, USA, v. 1, p. 1-13, aug. 1992.

DE GRAUWE, P.; DEWACHTER, H.; EMBRECHTS, M. **Exchange rate theory**. Oxford: Blackwell Publishers, 1993.

DE VEAUX, R.; SCHWEINSBERG, J.; SCHUMI, J.; UNGAR, L. H. Prediction intervals for neural networks via nonlinear regression. **Technometrics**. v. 40, n. 4. p. 273-282, nov, 1998.

DINIZ, H. A.; L. CARVALHO, A.; ANDRADE, M. Previsão de séries temporais utilizando redes neurais artificiais e modelos de Box e Jenkins. **Proceedings of the V Brazilian Symposium on Neural Networks**, Belo Horizonte: p. 173-178, dez. 1998.

ECHEVERRY, G. E. S. **Métodos de comparação de séries temporais**. São Paulo: IME, 1999. 131 f. Dissertação (Mestrado em Estatística) – Instituto de Matemática e Estatística, Universidade de São Paulo, 1999.

ELMAN, J. L. Finding structure in time. **Cognitive Science**. v. 14. p. 179-211, 1990.

ENDERS, W. **Applied econometric time series** 2. ed. John Wiley & Sons, 2003.

ENGLE, R. Autoregressive conditional heteroscedasticity with estimates of the variances of U.K. inflation. **Econometrica**, v. 50. p. 987-1008, 1982.

FAMA, E. Efficient capital markets: a review of theory and empirical work. **The Journal of Finance**. 25(1). p. 383-470, 1970.

_____. Efficient capital markets II. **The Journal of Finance**. 46(5). p. 1575-1617, 1991.

FAMA, E.; FRENCH, K. Permanent and temporary components of stock prices. **Journal of Political Economy**. v. 26, p. 246-273, 1988.

FRANCIS, J. C. **Investments: analysis and management**. 5. ed. New York: MacGraw-Hill, 1991.

FREEMAN, J. A. **Simulating neural networks with mathematica**. New York: Addison-Wesley Publishing Company, 1994.

FREITAS, A. A. C.; SECURATO, J. R.; NETO, M. A. Non-linear autoregressive (NPAR) neural network (NN) methodology to predict brazilian exchange rates. **IV Encontro Brasileiro de Finanças**, Rio de Janeiro. **Anais do IV Congresso da SBFIn**, Rio de Janeiro, 2004.

GENÇAY, R.; SELÇUK, F.; WHITCHER, B. **An introduction to wavelets and other filtering methods in finance and economics**. New York: Academic Press, 2002.

GONÇALVES, A. Técnicas de previsão em finanças: uma aplicação na administração de carteiras de ativos. **III SEMEAD**, São Paulo, 2000. **Anais do III SEMEAD**, São Paulo, 2000.

GRANGER, C. W. J.; MORGENSTERN, O. **The predictability of stock market prices**. Lexington: Heath-Lexington. 1970.

GROFF, G. K. Empirical comparison of models for short range forecasting. **Management Science**. 20: p. 22-31, 1973.

GUJARATI, D. N. **Econometria básica**. 3. ed. São Paulo: Makron Books, 2000.

HAEFKE, C.; HELMENSTEIN, C. Predicting stock market averages to enhance profitable trading strategies. Society of Computational Economics. **Second International Conference on Computing in Economics and Finance**. Geneva, Switzerland, 26-28, jun. 1996.

HAMILTON, J. D. **Time series analysis**. Princeton: Princeton University Press, 1994.

HANNAN, E. J. The estimation of the order of an ARMA process. **The Annals of Statistics**. v. 8, p. 1071-1081. 1980.

HARDIE, B. G. S.; FADER, P. S.; WISNIEWSKI, M. An empirical comparison of new product trial forecasting methods. **International Journal of Forecasting**. n. 17, p. 209-229, 1998.

HARRISON, P. J.; STEVENS, C. F. Bayesian Forecasting. **Journal of the Royal Statistical Society: series B**, v. 38, n. 3, p.81-135, 1976.

HARVEY, A. C. **Forecasting, structural time series models and the Kalman filter.** Cambridge: Cambridge Academic Press:1989.

HAYKIN, S. **Redes neurais: princípios e prática.** 2. ed. Porto Alegre: Bookman, 2001.

HILL, T.; O'CONNOR, M.; REMUS, W. Neural network models for time series forecasts. **Management Science.** v. 42, n. 7, p. 1082 – 1092, jul. 1996.

HISIEH, D. Testing for nonlinear dependence in daily foreign exchange rates. **Journal of Business,** 62, n. 3, p. 339-368, 1989.

HUTCHINSON, J. M. **A radial basis function approach to financial time series analysis.** 1994. 159 f. PhD Thesis (Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science), Febr. 1994.

HOMSY, G. V.; PORTUGAL, M. S.; ARAÚJO, J. P., Ondas e previsões de séries de tempo: uma análise empírica. **XXII Encontro Brasileiro de Econometria.** Campinas, 2000.

JAN, C. L. Benchmarking forecasting models. **The Journal of Business Forecasting,** 2003. p. 16-18.

JARQUE, C. M., e BERA, A. K. A test for normality of observations and regression residuals. **International Statistical Review.** v. 55, p. 163-172, 1987.

JARRETT, J. **Business forecasting methods.** 2. ed. The Alden Press, 1991.

KEIM, D.; STAMBAUGH, R. Predicting returns in stock and bond markets. **Journal of Financial Economics.** v. 17, 1986, p. 357-390.

KLEIN, J. L. **Statistical visions in time: a history of time series analysis, 1662-1938.** New York: Cambridge University Press, 1997.

KLIMASAUŠKAS, C. C. Accuracy and profit in trading systems. **Advanced Technology for Developers.** p 10-14. 1992.

KUTSURELIS, J. E. **Forecasting financial markets using neural network: an analysis of methods and accuracy.** 1998. 123 f. (Tese em Administração). Naval Postgraduate School, Monterey, California. 1998.

LEQUARRÉ, J.Y. Foreign currency dealing: a brief introduction. In WEINGEND, A. S.; GERSHENFELD, N. A. Eds. **Time series prediction: forecasting the future and understanding the past.** Addison-Wesley, 1994.

LJUNG, G. M., BOX, G. P. E. On a measure of lack of fit in time series models. **Biometrika.** v. 66, p. 66-72, 1978.

LUDLOW, J.; ENDERS, W. Estimating non-linear ARMA models using Fourier coefficients. **International Journal of Forecasting.** 16, 2000, p. 333 –347.

- McALEER, M., OXLEY, L. The econometrics of financial time series. **Journal of Economic Surveys**. v. 16. n. 3, 2002. p. 237-249.
- MCLEOD, A. J. e W. K. Li, Diagnostic checking ARMA times series models using squared residuals correlations. **Journal of Time Series Analysis**, n. 4, 1983, p. 269-73.
- MAKRIDAKIS, S.; WHEELWRIGHT, S. C.; McGEE, V. E. **Forecasting: methods and applications**. 2 .ed. John Wiley e Sons, 1983.
- MARCONI, M. de A., LAKATOS, E. M. **Técnicas de pesquisa**. 5. ed. São Paulo: Atlas, 2002.
- MEYER, Y. **Ondaletas and operators**. Tradução de D.H. Salinger. Cambridge: Cambridge University Press, 1993.
- MISITI, M. et al. **Ondaleta toolbox: for use with Matlab**. The Math Works, Inc. 1997. Disponível em: <<http://www.mathworks.com>> . Acesso em: 14 ago. 2003.
- MILLS, T. C. **The econometric modeling of financial time series**. New York: Cambridge University Press, 1996.
- MOODY, J. e SAFFELL, M. Reinforcement learning for trading system and portfolios. In: AGRAWAL, R.; PIATETSKY-SHAPIO, G. Ed. KDD98 - **Proceedings of The Fourth International Conference on Knowledge Discovery and Data Mining**. AAAI Press, California, 1998, p. 279-283.
- MORAIS, I. A. C., PORTUGAL, M. S. Modelagem e previsão de volatilidade determinista e estocástica para a série do IBOVESPA. 1999. Disponível em: <<http://www.ufrgs.br/ppge>>. Acesso em: 20 jul. 2003.
- MORETTIN, P. A. **Séries temporais em finanças**. São Paulo, 2002. 156 p. (Texto para um Curso na Universidade Católica de Lima, Peru).
- _____. **Ondas e ondaletas**. São Paulo: Edusp, 1999.
- MORETTIN, P. A.; TOLOI, C. M. de C. **Análise de séries temporais**. São Paulo: Edgard Blücher, 2004.
- MOSHIRI, S., CAMERON, N. Neural network versus econometric models in forecasting inflation. **Journal of Forecasting**, v. 19, 2000. p. 201-217.
- MOURA, C. A. de. **Introdução às ondeletas**. 2002. Disponível em: <<http://www.uerj.br/ime/cademoura>> . Acesso em 10 de jan. 2003.
- NEFTCI, S. N. Naïve trading rules in financial markets e Wiener-Kolmogorov prediction theory: a study of technical analysis. **Journal of Business**. v. 64, n. 4, 1991.
- OLIVEIRA, M. A.; FAVERO, L. P. L. Uma breve descrição de algumas técnicas para análise de séries temporais: séries de Fourier, ondaletas, ARIMA, modelos estruturais para séries de tempo e redes neurais. **Anais do VI SEMEAD**, São Paulo, 2003.

- OSBORNE, M.F.M. Brownian motion in the stock market. **Operations Research VII**. 1959
- PANKRATZ, A.; DUDLEY, U. Forecasts of power-transformed series. **Journal of Forecasting**. 6, 1987. p. 239-248.
- PASSARI, A. F. L. **Exploração de dados automatizados para previsão de vendas no varejo utilizando redes neurais**. Faculdade de Economia, Administração e Contabilidade. São Paulo, 2003. 143 f. Dissertação (Mestrado em Administração). Faculdade de Economia, Administração e Contabilidade: Universidade de São Paulo, 2003.
- PINDYCK, R. S.; RUBINFELD, D. L. **Econometric models and economic forecasts**. 4. ed. New York:McGraw-Hill International, 1997.
- POLIKAR, R. **The wavelet tutorial**. 1994. Disponível em: <<http://www.site.uottawa.ca/~qingchen/wavelet.htm>> . Acesso em 10 de jan. 2002.
- PRIESTLEY, M.B. **Spectral analysis and time series**. London: Unknown, 1983.
- QI, Min. Nonlinear predictability of stock returns using financial and economic variables. **Journal of Business & Economic Statistics**. v. 64, n. 4, 1999, p. 419-429.
- SAKAI, H.; NAKAJIMA, H.; YASUDA, M.; HIGASHIHARA, M.; OOSUMI, M. **Development of a fuzzy sales forecasting system for vending machines**. Computer and Industrial Engineering. Elsevier Science. n. 36, p. 427-449. 1999.
- SCHAWZ, G. Estimating the dimension of a model. **The Annals of Statistics**. v. 6, 1978. p. 461-464.
- SCHREIBER, T.; SCHMITZ, A. Interdisciplinary application of nonlinear time series methods. Physical Review. E 55, 5443, D-42097 Wuppertal, German, 1999.
- SCHUMANN, M.; LOHRBACH, T. **Comparing artificial neural networks with statistical methods within the field of stock market prediction**. 0-7803-2566-4/96. IEEE. 1996.
- SHALIZI, C. R.; SHALIZI, K. L; CRUTCHFIELD, J. P. **An algorithm for pattern discovery in time series**. Santa Fe Institute. Working Paper – Nov. 2002.
- SHIAVI, R. **Introduction to applied statistical signal analysis**. 2.ed. San Diego: Academic Press, 1999.
- SIEGLER, W.; STEURER, E. Forecasting of the german stock index DAX with neural networks: using daily data for experiments with input variable reduction and a modified Error Function. p 265-273. In: EBECKEN, N. F. F. Editor. **Proceedings of the international conference on data mining**, Rio de Janeiro, S&P. 1998. p. 265-273
- SIQUEIRA, J. O. Introdução à gestão de risco. **FACEF Pesquisa**, Franca, v. 6. n. 3, p. 19–25, 2003.

SONCIN, C. A., CORRENTE, J. E. Uso de modelos de séries temporais para análise de dados de preços de ações. **48ª Reunião da RBRAS e 10ª SEAGRO**. Universidade Federal de Lavras. Lavras, 2003.

SOUZA, R. C. **Modelos estruturais para previsão de séries temporais: abordagem clássica e bayesiana**. Rio de Janeiro: IMPA, 1989.

STRUZIK, Zbigniew R. Ondaleta methods in financial time series processing. **Physica A**. Amsterdam, 296: 307 – 319, 2001.

SWANSON, N., WHITE, H., A model selection approach to real-time macroeconomic forecasting under linear models and artificial neural networks. **The Review of Economics and Statistics**. 1997.

TAK, B. **A new method for forecasting stock prices using artificial neural network and ondaleta theory**. 1995. 107 f. Tese (Doutorado em Economia). Universidade de Pensilvânia, Estados Unidos, 1995.

TOLOI, C. M. de C. **Comparação de métodos de previsão de séries temporais**. São Paulo: IME, 1980. 230 f. Dissertação (Mestrado em Matemática). Instituto de Matemática e Estatística, Universidade de São Paulo, 1980.

TORRENCE, C., COMPO, G. A. A practical guide to ondaletas analysis. **Bulletin of the American Meteorology Society**, v. 79, n.1, 1998. p. 61-78.

TORRES, R., BONOMO, M. e FERNANDES, C. A aleatoriedade do passeio na Bovespa: testando a eficiência do mercado acionário brasileiro. In **XXIII ENCONTRO BRASILEIRO DE ECONOMETRIA**, Salvador. Disponível em: <<http://www.sbe.org.br>>. Acesso em: 20 dez. 2003.

TSAY, R. S. **Analysis of financial time series**. Wiley series in probability and statistics. 2002.

THROSTENSEN, V. H. **A teoria de eficiência no mercado de capitais: uma revisão da literatura e dos trabalhos empíricos: o modelo de random walk aplicado ao índice de ações da Bovespa**. 1976, 121 f. Dissertação (Mestrado em Administração). Escola de Administração de Empresas de São Paulo da Fundação Getúlio Vargas, São Paulo. 1976.

VANDAELE, W. **Applied time series and Box-Jenkins models**. New York: Academic Press, 1983.

VASCONCELLOS, M. A. S., ALVES, D. **Manual de econometria**. São Paulo: Atlas, 2000.

WINKLHOFFER H.; WITT, S. F.; DIAMATOPOULOS, A. Forecasting practice: a review of the empirical literature and an agenda for the future research. **International Journal of Forecasting**, Elsevier Science B, v. 1, n.12, p. 193-221, 1996.

WILLIAMS, R. J.; ZIPSER, D. A learning algorithm for continually running fully recurrent neural networks. **Neural Computation**. v. 1, p. 270-280. 1989.

_____. Gradient-based learning algorithm for recurrent networks and their computational complexity, 1995.

WONG, H., IP. W., XIE, Z., LUI, X.. Modelling and forecasting by ondaletas, and the application to exchange rates. **Journal of Applied Statistics**, v. 30, n. 5, 2003. p. 537-553.

YANG, Z. Probabilistic neural network in bankruptcy prediction. **Journal of Business Research**, 44, 1999. p. 67-74.

YIM, J. Previsão de séries de tempo: modelos ARIMA, estruturais e redes neurais artificiais. São Paulo: IME, 2001. 87 f. Dissertação (Mestrado em Matemática Aplicada) – Instituto de Matemática e Estatística, Universidade de São Paulo, 2001.

ZHANG, G.; PATUWO, B. E.; HU, M. Y. Forecasting with artificial neural networks: the state of the art. **International Journal of Forecasting**, Kent(Ohio) 14, p. 35–62, 1998.

ZIEGELMANN, F. A., PERERA, P. L. V. Modelos de volatilidade estocástica com deformação temporal: um estudo empírico para o índice BOVESPA. **Política e Planejamento Econômico**. v. 27, n. 2, p. 353–376, 1997.

ZOU, W. Integration of different forecasting models. **The Journal of Business Forecasting**. p. 26-32, 1999.

ZOU, H., YANG, Y. Combining time series models for forecasting. **International Journal of Forecasting**. n. 20, p. 69-84. 2004.

APÊNDICE A

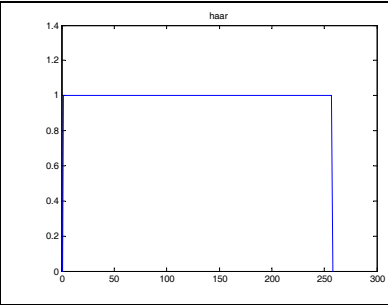
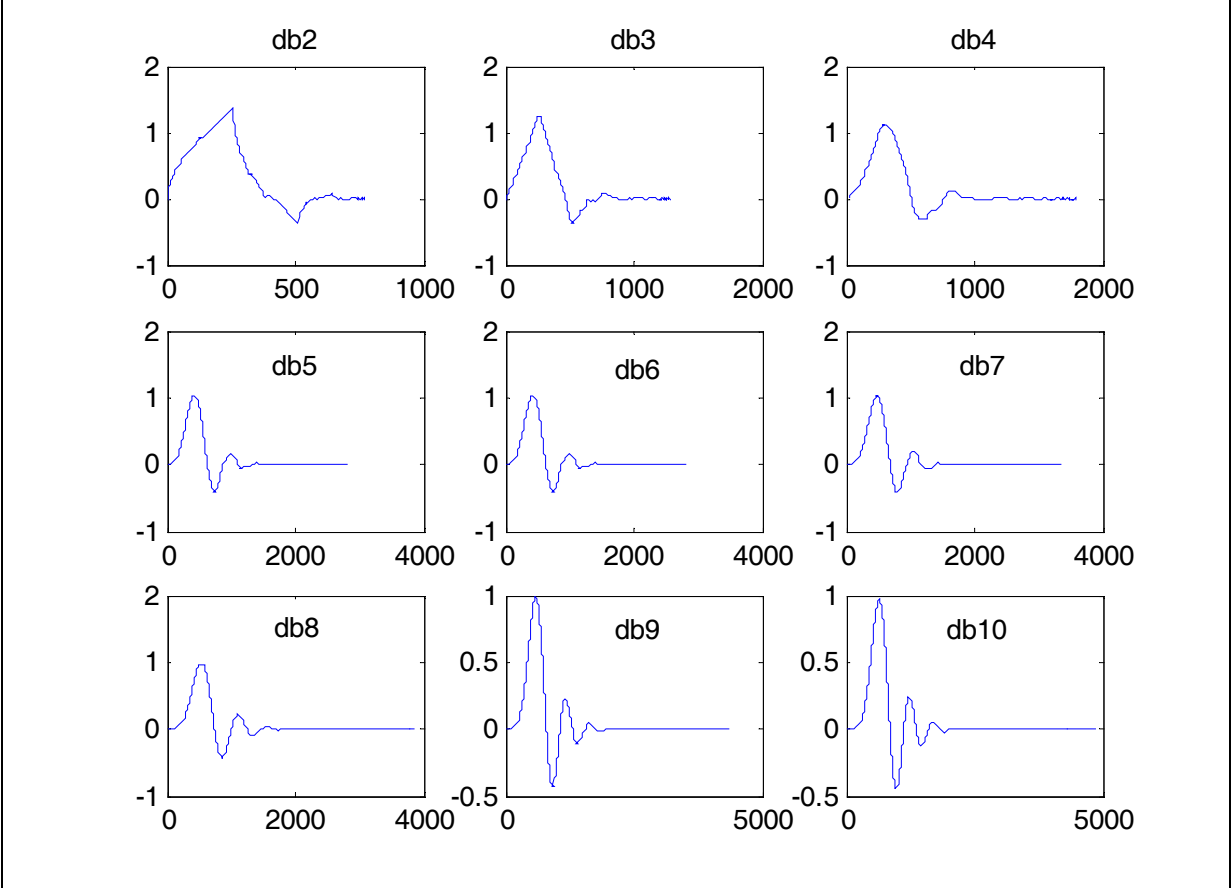
Tabela 1: Valores críticos para a estatística do teste de ADF

Tamanho	Nível de significância							
	0.01	0.025	0.05	0.10	0.90	0.95	0.975	0.99
Sem intercepto e sem constante								
25	-2.66	-2.26	-1.95	-1.60	0.92	1.33	1.70	2.16
50	-2.62	-2.25	-1.95	-1.61	0.91	1.31	1.66	2.08
100	-2.60	-2.24	-1.95	-1.61	0.90	1.29	1.64	2.03
250	-2.58	-2.23	-1.95	-1.62	0.89	1.29	1.63	2.01
300	-2.58	-2.23	-1.95	-1.62	0.89	1.28	1.62	2.00
∞	-2.58	-2.23	-1.95	-1.62	0.89	1.28	1.62	2.00
Com intercepto e sem constante								
25	-3.75	-3.33	-3.00	-2.62	-0.37	0.00	0.34	0.72
50	-3.58	-3.22	-2.93	-2.60	-0.40	-0.03	0.29	0.66
100	-3.54	-3.17	-2.89	-2.58	-0.42	-0.05	0.26	0.63
250	-3.46	-3.14	-2.88	-2.57	-0.42	-0.06	0.24	0.62
300	-3.44	-3.13	-2.87	-2.57	-0.43	-0.07	0.24	0.61
∞	-3.43	-3.12	-2.86	-2.57	-0.44	-0.07	0.23	0.60
Com intercepto e com constante								
25	-4.38	-3.95	-3.60	-3.24	-1.14	-0.80	-0.50	-0.15
50	-4.15	-3.80	-3.50	-3.18	-1.19	-0.87	-0.58	-0.24
100	-4.04	-3.73	-3.45	-3.15	-1.22	-0.90	-0.62	-0.28
250	-3.99	-3.69	-3.43	-3.13	-1.23	-0.92	-0.64	-0.31
300	-3.98	-3.68	-3.42	-3.13	-1.24	-0.93	-0.65	-0.32
∞	-3.96	-3.66	-3.41	-3.12	-1.25	-0.94	-0.66	-0.33

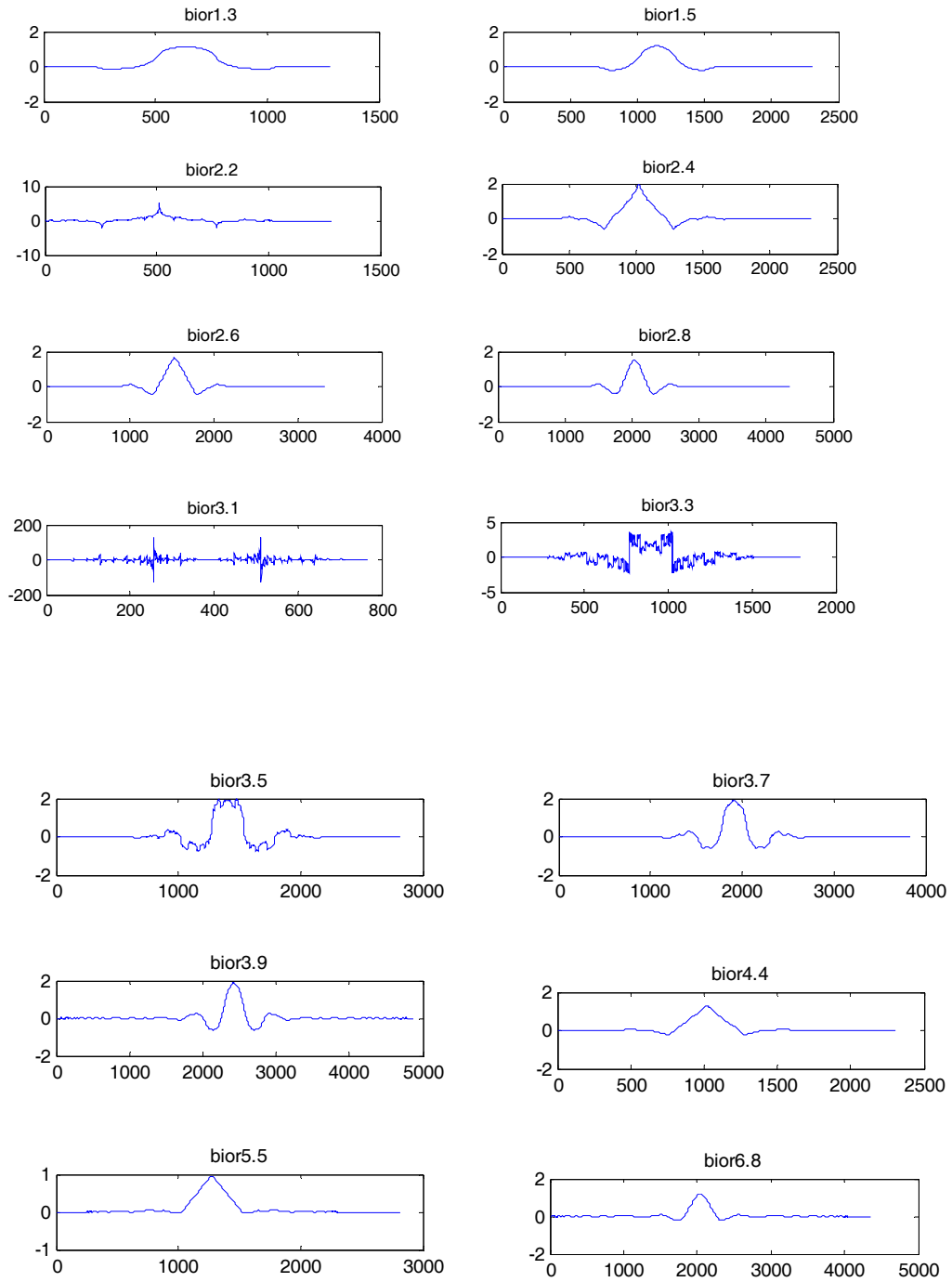
Fonte: CROMWELL, LABYS, TERRAZA (1994).

APÊNDICE B

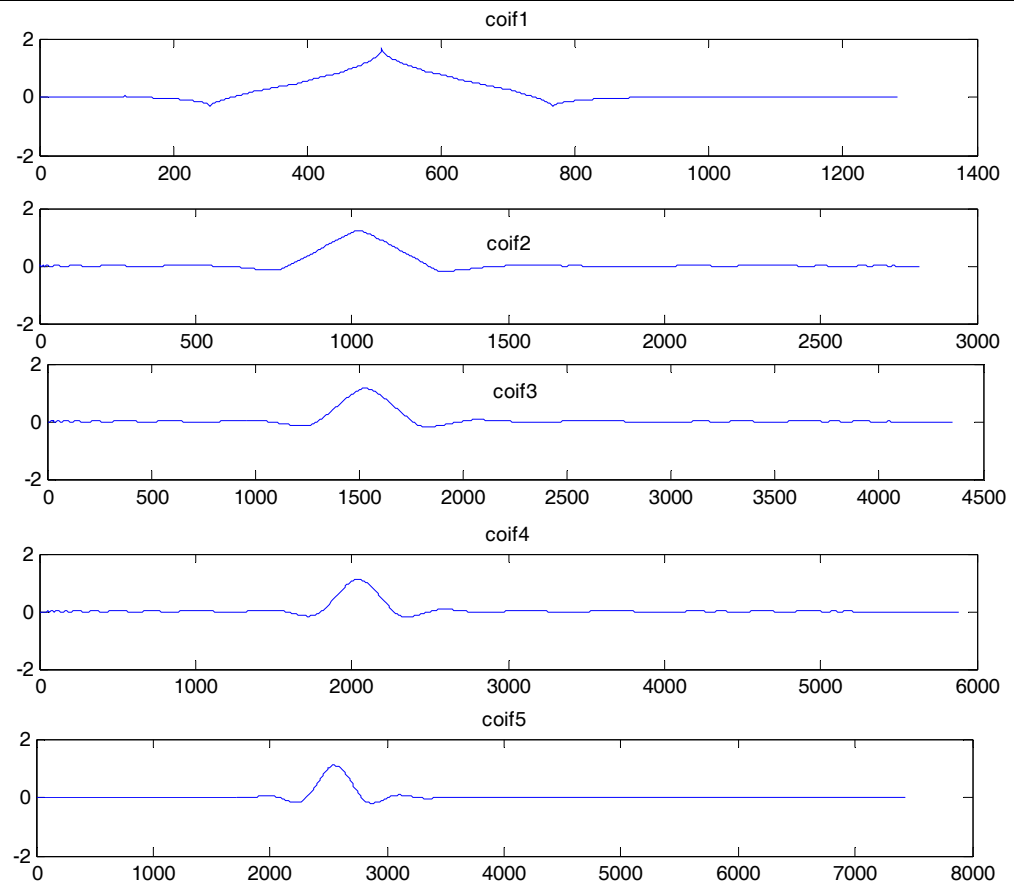
TIPOS DE WAVELETS – (Misiti et al., 1997)

Haar	$\Psi(x) = \begin{cases} 1, se\ 0 \leq x \leq \frac{1}{2} \\ -1, se\ \frac{1}{2} \leq x < 1 \\ 0, caso\ contrário \end{cases}$	
Daubechies		

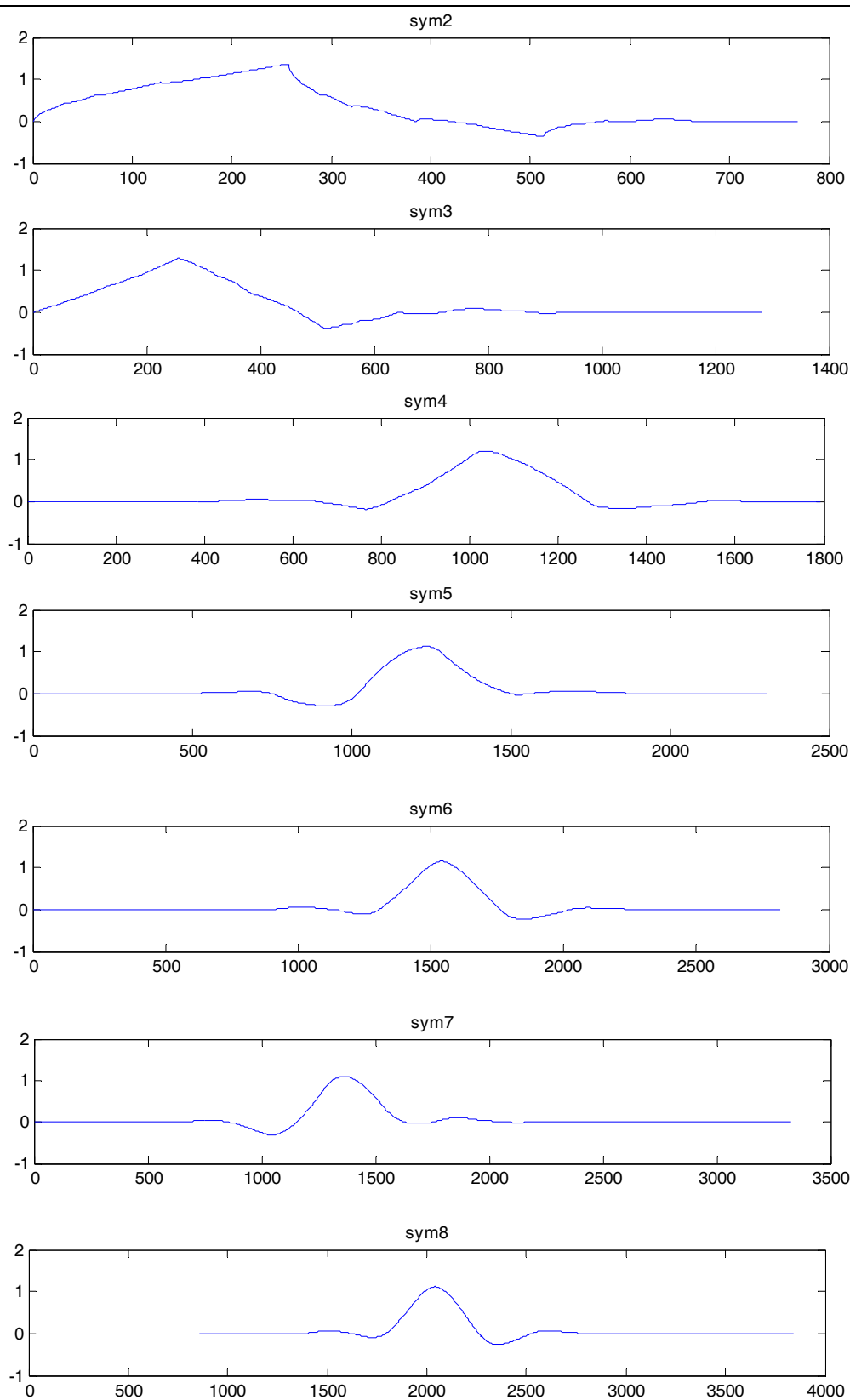
Biorthogonal

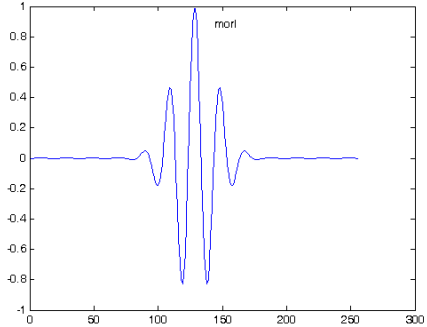
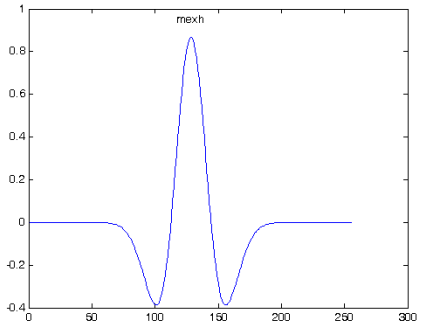
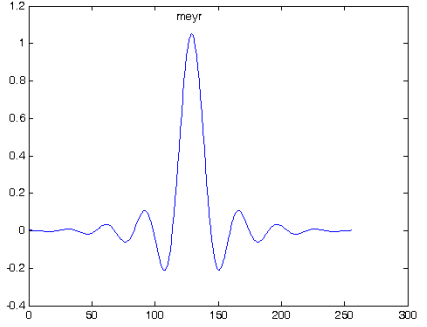


Coiflets



Symlets



Morlet	$\Psi(x) = Ce^{\frac{-x}{2}} \cdot \cos(5x)$ <p>sendo que C é uma constante.</p>	
Mexican Hat	$\Psi(x) = \left(\frac{2}{\sqrt{3}} \pi^{-\frac{1}{4}} \right) (1 - x^2) e^{-\frac{x}{2}}$	
Meyer		

APÊNDICE C

PROGRAMAS USADOS AO LONGO DO TRABALHO

PROGRAMA 1 – Sequência de comandos para se simular o modelo ARIMA-GARCH no Mathematica ® 5.0.

Autor: Fabiano Guasti Lima

```
Needs["TimeSeries`TimeSeries`"]
```

```
SeedRandom[4584715];
```

```
Dados=Table[x[i]=Table[TimeSeries[GARCHModel[{0.06,0.04},{0.7}],1200]],{i,1,200}];TimeSeries[ARModel[{0.2,-0.5}],z,{0,0}]
```

```
(* Gera 200 observações com 1200 pontos *)
```

```
ListPlot[x[1],PlotJoined->True]
```

```
(* Faz o gráfico da primeira sucessão cronológica simulada *)
```

```
writeMatrixSpreadSheet[filename_String, data_List] :=
```

```
Block[{file = OpenWrite[filename]},
```

```
Scan[(WriteString[file,First[#]];
```

```
Scan[WriteString[file,"\t",#]&,Rest[#]];
```

```
WriteString[file,"\n"]&,data];
```

```
Close[file]]
```

```
(* Rotina para escrever os dados no formato de planilha *)
```

```
writeMatrixSpreadSheet["serieteste.txt",Transpose[dados]]
```

```
(* Salva as observações numa matrix de 200 colunas com 1200 linhas *)
```

PROGRAMA 2 – Pacote (Ondaleta) com os comandos para se decompor uma sucessão cronológica em um nível de pré-processamento via Mathematica ® 5.0.

Deve-se carregar o pacote intitulado Ondaleta no Front End do Mathematica ® 5.0, que deve estar salvo em um dos arquivos dos pacotes disponíveis – no caso, no de estatística – com o comando:

```
<< Statistics/Ondaleta.m

(* Título: Ondaleta 1 nível*)
(* Autor: Fabiano Guasti Lima *)
(* Mathematica version: 5.0 *)

BeginPackage["Ondaleta`"]
Ondaleta::usage =
"Ondaleta faz o pré-processamento de uma sucessão cronológica
decompondo a série em duas subséries, uma de baixa (ca) e outra
de alta (cd) frequencia."
ca::usage =
"ca faz o pré-processamento da parte de baixa frequencia da
decomposição em 1 nível."
cd::usage =
"cd faz o pré-processamento da parte de alta frequencia da
decomposição em 1 nível."

Needs["Wavelets`Wavelets`"];
Ondaleta[arquivo_]:=ListPlot[arquivo, PlotJoined -> True, AxesLabel
-> {"t", "X(t)"}];
(* Plota apenas o gráfico do arquivo da sucessão cronológica*)
ca[arquivo_]:=ListPlot[Flatten[Take[WaveletTransform[arquivo,
DaubechiesFilter[1], 1], 1]],
PlotJoined -> True]
(* Faz o pré-processamento da parte de baixa frequencia da
decomposição em 1 nível*)
cd[arquivo_]:=ListPlot[Drop[Flatten[WaveletTransform[arquivo,Daubech
iesFilter[1], 1]],
Length[Flatten[WaveletTransform[arquivo, DaubechiesFilter[1],
1]]]/2], PlotJoined -> True]
(* Faz o pré-processamento da parte de alta frequencia da
decomposição em 1 nível*)
EndPackage[ ]
```

Obs: O Pacote possui um help próprio. Caso o usuário queira saber o que é, por exemplo “ca”, basta digitar na linha de comando:

```
?ca << Shift Enter>>
```

PROGRAMA 3 – Pacote (Ondaletapos) com os comandos para se reconstruir uma sucessão cronológica em um nível de pré-processamento via Mathematica ® 5.0.

Deve-se carregar o pacote intitulado Ondaletapos no Front End do Mathematica ® 5.0, que deve estar salvo em um dos arquivos dos pacotes disponíveis – no caso, no de estatística – com o comando:

```
<< Statistics/Ondaletapos.m

(* Título: Ondaletapos 1 nível*)
(* Autor: Fabiano Guasti Lima *)
(* Mathematica version: 5.0 *)

BeginPackage["Ondaletapos`"]

Ondaletapos::usage =
"Ondaletapos faz o pós-processamento de uma sucessão cronológica em
primeiro nível dadas as duas subséries, uma de baixa (ca) e outra
de alta (cd) frequencia."

Needs["Wavelets`Wavelets`"];

(* Fazer a leitura primeiro das duas subséries já previstas *)
(* ca = ReadList["c:\ca.txt", Real];*)
(*cd = ReadList["c:\cd.txt", Real];*)

Ondaletapos[arquivo1_,arquivo2_]:=InverseWaveletTransform[{arquivo1,
arquivo2}, DaubechiesFilter[1]]

(* Faz a reconstrução pela transformada de ondaleta inversa *)
EndPackage[ ]
```

PROGRAMA 4 – Hsieh – Segue o programa com os comandos para calcular a estatística de teste para o teste de Hsieh escrito no Matlab ® 6.5.

```
(Neste exemplo, calcula-se o valor dos coeficientes (i,j) para a
Subséries ca.
load ca.txt
valor=ca;
soma=0;soma2=0;
n=length(valor);

%cálculo da média dos dados

for i=1:n
soma=soma+valor(i);
end;
media=soma/n;

%cálculo do desvio padrão
for i=1:n
soma2=soma2+(valor(i)-media)*(valor(i)-media);
end;
dp=sqrt(soma2/(n-1));

%cálculo do vetor normalizado
for i=1:n
nvalor(i)=(valor(i)-media)/dp;
end;

%cálculo do vetor normalizado ao quadrado
for i=1:n
nvalor2(i)=(nvalor(i))*(nvalor(i));
end;

%cálculo da variância do vetor normalizado ao quadrado
soma3=0;soma4=0;
for i=1:n
soma3=soma3+nvalor2(i);
end;
medial=soma3/n;
for i=1:n
soma4=soma4+(nvalor2(i)-medial)*(nvalor2(i)-medial);
end;
var=soma4/(n-1);

q=input('quantos coeficientes quer calcular = ');

aux=0;
for i=1:q
soma5=0;soma6=0;
for j=i:q
for t=(i+j-aux):n
soma5=soma5+(nvalor(t))*(nvalor(t-i))*(nvalor(t-j));
soma6=soma6+(nvalor2(t))*(nvalor2(t-i))*(nvalor2(t-j));
end;
end;
end;
x=((1/(n))*soma5)/((1/(n))*soma3)^(3/2);
```

```
y=((1/(n))*soma6)/((1/(n))*soma3)^(3);  
  
end;  
soma5=0;  
soma6=0;  
hsieh=sqrt(n)*x/y  
%if((hsieh<=1.96) & (hsieh>=-1.96))  
%disp ('Aceito Ho - nao linearidade na variancia')  
%else  
%disp ('Reieito Ho - nao linearidade na media')  
%end;  
end;  
aux=aux+1;  
end;  
  
clear
```

PROGRAMA 5 – Sequência de comandos para se simular o modelo ARIMA(1,0,0) no Mathematica ® 5.0.

Autor: Fabiano Guasti Lima

Needs["TimeSeries`TimeSeries`"]

`dados =Table[x[i]=TimeSeries[ARModel[{0.3},1],1200],{i,1,200}];`

`(* Gera 200 observações com 1200 pontos *)`

`ListPlot[x[1],PlotJoined->True]`

`(* Faz o gráfico da primeira sucessão cronológica simulada *)`

`writeMatrixSpreadSheet[filename_String, data_List] :=`

`Block[{file = OpenWrite[filename]},`

`Scan[(WriteString[file,First[#]];`

`Scan[WriteString[file,"\t",#]&,Rest[#]];`

`WriteString[file,"\n"]&,data];`

`Close[file]]`

`(* Rotina para escrever os dados no formato de planilha *)`

`writeMatrixSpreadSheet["serieteste.txt",Transpose[dados]]`

`(* Salva as observações numa matrix de 200 colunas com 1200 linhas *)`

PROGRAMA 6 – Seqüência de comandos para se simular o modelo Bilinear no Mathematica ® 5.0.

O modelo gerado foi: $X_t = 0.3X_{t-1} + 0.8X_{t-1}\varepsilon_{t-1} + \varepsilon_t$

```
Autor: Fabiano Guasti Lima
Needs["TimeSeries`TimeSeries`"]
```

```
Table[erro[i] = Random[Real, {0, 1}], {i, 1, 1400}];
(* Gera os números aleatórios*)
```

```
erro[0] = 0;
```

```
x[0] = 0;
(* inicialização das variáveis*)
```

```
serie = Table[x[j] = (0.3)*x[j - 1] +
(0.8)*(x[j - 1])*(erro[j - 1]) + erro[j] , {j, 1, 1400}];
```

```
seriefinal = Drop[serie, 200];
```

```
(* Descarta-se os 200 primeiros valores do início da série
para que os valores iniciais atribuídos não tenham
influência na série*)
```

```
writeMatrixSpreadSheet[filename_String, data_List] :=
  Block[{file = OpenWrite[filename]},
    Scan[WriteString[file, First[#]]];
    Scan[WriteString[file, "\t", #]&, Rest[#]];
  WriteString[file, "\n"]&, data];
  Close[file]]
```

```
(* Rotina para escrever os dados no formato de planilha *)
```

```
writeMatrixSpreadSheet["serieteste.txt", Transpose[dados]]
```

```
(* Salva as observações numa matrix de 200 colunas com 1200 linhas
*)
```

PROGRAMA 7 – Teste de Não-linearidade por intervalo. FREITAS, SECURATO, ANDRADE NETO (2004).

Segue o programa com os comandos para calcular o teste escrito no Matlab ® 6.5 para uma sub-divisão.

```

u=ca/max(ca);
t=length(u);
for dtau=1:t-1
soma1=0;soma2=0;
for dtau1=1:dtau

    soma1=soma1+(u(t)-u(t-dtau1))^3;soma2=soma2+((u(t)-u(t-
dtau1))^2)*(3/2);

end

ft(dtau)= soma1/soma2;
lim_max(dtau) = (1.96/sqrt(t));lim_min(dtau) = -(1.96/sqrt(t));

end

t1=[1:t-1];

subplot(2,1,1)
plot(t1,ft,t1,lim_max,'-',t1,lim_min,'-')
title('CA-ARIMA-GARCH')

clear

u=cd/max(cd);
t=length(u);
for dtau=1:t-1
soma1=0;soma2=0;

for dtau1=1:dtau

    soma1=soma1+(u(t)-u(t-dtau1))^3;soma2=soma2+((u(t)-u(t-
dtau1))^2)*(3/2);

end

ft(dtau)= soma1/soma2;
lim_max(dtau) = (1.96/sqrt(t));lim_min(dtau) = -(1.96/sqrt(t));

end

t1=[1:t-1];
subplot(2,1,2)
plot(t1,ft,t1,lim_max,'-',t1,lim_min,'-')
title('CD-ARIMA-GARCH')

```


PROGRAMA 8 – Programa para rodar a rede recorrente. De Oliveira (2003).

```

%                               APRENDIZAGEM RECORRENTE EM TEMPO REAL (ARTR)
%
%                               ALGORITMO - ARTR
load c:\bilinear\Inputs.txt
load c:\bilinear\DesiredOutputs.txt
H = 4;InitialState = [0.0 zeros(1,H)];
sigmoid_type = 0;
alpha = 1.0;
learning_rate = 0.005;
epochs = 20000;epsilon = 0.001;Inputs = Inputs(801:1000);
DesiredOutputs = DesiredOutputs(801:1000);
[WeightMatrix] =
artr_treino(Inputs,DesiredOutputs,H,InitialState,sigmoid_type,alpha,
learning_rate,epochs,epsilon);
[Outputs] =
artr_perf(Inputs,InitialState,H,WeightMatrix,sigmoid_type,alpha);
[samples columns] = size(Inputs);
t = (1:samples)';
figure
hold on
plot(t,DesiredOutputs,'b-.');
plot(t,Outputs,'r');
legend('Valor Desejado','Valor Previsto',0)
xlabel('Janela de Previsao')
ylabel('Serie d(t) versus Serie y(t)')
figure
hold on
plot(t, DesiredOutputs - Outputs)
legend('Erro de previsao',0)
xlabel('Janela de Previsao')
ylabel('Erro de previsao, e(t)')
figure
hold on
plot(t, Outputs)
legend('Valores Previstos',0)
xlabel('Janela de Previsao')
ylabel('Valores Previstos, e(t)')
wklwrite('c:\bilinear\bilinearrec',Outputs);
figure
hold on
histfit(DesiredOutputs - Outputs)
e = DesiredOutputs - Outputs
perf_mae = mae(e)
% perf2_mse = mse(e)
rmse = sqrt(mse(e))
eq = (DesiredOutputs - Outputs)*(DesiredOutputs - Outputs)'
dq = (DesiredOutputs)*(DesiredOutputs)'
oq = (Outputs)*(Outputs)'
deq = diag(eq)
ddq = diag(dq)
doq = diag(oq)
theil =
(sqrt(sum(diag(eq))/10))/((sqrt(sum(diag(oq))/10))+(sqrt(sum(diag(dq)
)/10)))

```