

Universidade de Lisboa
Faculdade de Ciências
Departamento de Matemática

Instituto Universitário de Lisboa
ISCTE-IUL- Business School
Departamento de Finanças



Ciências
ULisboa



Previsão de séries temporais financeiras: o caso PSI 20

Elsa Maria Alves Garcia

Mestrado em Matemática Financeira

Dissertação orientadora por:

Professora Doutora Diana Aldea Mendes

Departamento de Métodos Quantitativos para Gestão e Economia

ISCTE – IUL

2017

AGRADECIMENTOS

Em primeiro lugar, agradeço à professora Dr. ^a Diana Mendes, pelo incentivo, disponibilidade, paciência, orientação e dedicação que teve durante este projeto.

Agradeço a todos os meus amigos e à minha família pelo apoio e incentivo durante este trabalho. Principalmente à minha mãe que, com sua paciência, amor, apoio, dedicação, disciplina tornou me uma pessoa que sou hoje e ao meu irmão, por ser, além de irmão, um verdadeiro amigo e uma pessoa com a qual poderei contar e agradeço todo o apoio que me transmitiu.

RESUMO

O objetivo deste trabalho é a previsão de Séries Temporais univariados através de uma Análise Econométrica que nos permite estimar e analisar o valor de uma série num determinado intervalo de tempo. Para o efeito considerou-se uma amostra de valores do fecho (*close*) do principal índice bolsista português (PSI20) cotado na Bolsa de Valores de Lisboa (BVL).

As Séries Temporais apresentam um comportamento bastante complexo ao longo do tempo o que os torna difícil de modelar e prever. Existem vários modelos de análise de séries temporais. Seguindo a literatura científica dos últimos anos nesta área, proponha-se um estudo comparativo dos resultados obtidos pela abordagem clássica, utilizando modelos lineares (modelos ARMA) e dos modelos heterocedásticos, os modelos GARCH. Neste sentido vamos verificar qual dos métodos é o mais eficiente quando tais modelos são aplicados a séries que tem um comportamento bastante irregular.

Deste modo, vamos apresentar de uma forma detalhada a análise dos modelos econométricos em estudo para a obtenção de previsões pretendidas.

Palavras-chave: Bolsa de valores de Lisboa, PSI 20, Modelos ARMA, Modelos GARCH

ABSTRACT

The main purpose of this thesis is the forecast of uni-variate time series through and Econometric Analysis that allows us to estimate and analyze the value of a series in a certain time interval. For this purpose, a sample of the closing values of the main Portuguese stock exchange index (PSI20) listed on the Lisbon Stock Exchange (BVL) was considered.

The time series have a quite complex behavior over time, which makes them difficult to model and predict. There are several models of time series analysis. Following the scientific literature of recent years in this area, we propose a comparative study of the results obtained by the classical approach, using linear models (ARMA models) and the heteroscedastic models, that is, GARCH models. In this sense we will check which of these methods is most efficient when such models are applied to series which present a rather irregular behavior.

This way, we will present with some detail the analysis of the mentioned econometric models in order to obtain out-of-sample predictions.

ÍNDICE GERAL

| | |
|---|-----------|
| AGRADECIMENTOS | i |
| RESUMO | ii |
| ABSTRAT | iii |
| ÍNDICE GERAL | iv |
| ÍNDICE DE TABELAS | v |
| ÍNDICE DE FIGURAS | vi |
| INTRODUÇÃO | 1 |
| ESTRUTURA DA TESE | 2 |
| 1. REVISÃO BIBLIOGRÁFICA | 3 |
| 2. BOLSA DE VALORES | 7 |
| 2.1. BOLSA DE VALORES DE LISBOA | 7 |
| 2.1.1. HISTÓRIA DA BVL | 8 |
| 2.2. ÍNDICE BOLSISTA | 11 |
| 2.3. PSI 20 | 12 |
| 2.3.1 PRINCIPAIS ACONTECIMENTOS | 14 |
| 3. CONCEITOS E MODELOS ECONOMÉTRICOS | 16 |
| 3.1. ESTUDO DE SÉRIES TEMPORAIS | 16 |
| 3.1.1. ESTACIONARIDADE | 19 |
| 3.2. MODELOS DE SÉRIES TEMPORAIS | 23 |
| 3.2.1. MODELOS ARMA | 24 |
| 3.2.2. MODELOS ARCH/GARCH | 28 |
| 3.2.2.1. MODELO GARCH | 28 |
| 4. ANÁLISE EMPÍRICA DOS DADOS | 31 |
| 4.1. ESTIMAÇÃO DOS MODELOS ECONOMÉTRICOS | 34 |
| 4.1.1. ESTIMAÇÃO DO MODELO ARMA (0,1) | 36 |
| 4.1.2 ESTIMAÇÃO DO MODELO GARCH | 41 |
| COMPARAÇÃO DOS MODELOS ARCH e GARCH | 44 |
| 5. CONCLUSÃO | 46 |
| BIBLIOGRAFIA | 47 |
| ANEXOS | 49 |

ÍNDICE DE TABELAS

| | |
|--|----|
| Tabela 1. Composição de PSI 20 a 30 de dezembro de 2016..... | 14 |
| Tabela 2. Diferentes hipóteses para os testes DF e ADF..... | 23 |
| Tabela 3. Valores do AIC dos modelos. | 36 |

ÍNDICE DE FIGURAS

| | |
|---|----|
| Figura 1. Principais acontecimentos do PSI 20 | 14 |
| Figura 2.Exemplo de série temporal | 17 |
| Figura 3. Exemplo de série não estacionária | 20 |
| Figura 4.. Exemplo de série estacionária | 20 |
| Figura 5.Representação gráfica da série PSI20..... | 31 |
| Figura 6. Histograma da série PSI20 | 32 |
| Figura 7. Representação gráfica da série rpsi | 33 |
| Figura 8. Estatística descritiva da série rpsi..... | 33 |
| Figura 9.Teste de Raiz Unitária aplicado a rpsi..... | 34 |
| Figura 10. Correlograma da série rpsi..... | 35 |
| Figura 11. Gráfico dos resíduos da série rpsi..... | 37 |
| Figura 12. Gráfico e histograma dos resíduos da série rpsi20 | 37 |
| Figura 13. Representação gráfica da previsão da série psi20 do ano 2016..... | 39 |
| Figura 14. Representação gráfica dos erros do modelo MA (1)..... | 40 |
| Figura 15. Gráfico dos resíduos do modelo GARCH (1,1). | 41 |
| Figura 16. Correlograma dos resíduos aplicado ao modelo GARCH..... | 42 |
| Figura 17. Previsão do ano 2016 utilizando o modelo GARCH (1, 1)..... | 43 |
| Figura 18. Representação gráfica dos erros do modelo GARCH (1, 1) | 44 |
| Figura 21. Representação gráfica da previsão da série psi20 aplicando o modelo MA (1) | 51 |
| Figura 22. Representação gráfica da previsão da série psi20 aplicando o modelo GARCH (1, 1)..... | 51 |

INTRODUÇÃO

A análise de séries temporais é fundamental no apoio à tomada de decisões e definição estratégicas, a nível empresarial, financeiro, e cada vez mais organizações usam previsões de séries temporais na fase de tomada de decisões. A grande quantidade de dados armazenados nas bases de dados das mais diversas áreas permite-nos realizar estudos com base adquirido no passado para podermos extrapolar para o futuro. A previsão de determinados fatos auxilia a tomada de decisões que poderão melhorar o desempenho das empresas ou até mesmo minimizar prejuízos.

As séries temporais financeiras e, em particular, as séries temporais geradas por negociação em bolsas de valores exibem uma série de características que as tornam especialmente difíceis de modelar e prever. O seu comportamento altera-se ao longo do tempo, ou seja, apresentam períodos de tempo em que se verifica grandes variações e períodos em que não se verifica variações. Este comportamento não-estacionário deve-se a alguns fatores entre os quais destacam-se:

- Alterações ao nível dos investidores – ao longo do tempo, alguns investidores deixam de atuar nos mercados, enquanto outros iniciam, isso provoca alterações no número de investidores nos mercados;
- Alteração no funcionamento das bolsas – diferenças ao nível dos períodos de abertura, dos sistemas informáticos de negociação, das regras de formação das cotações;
- Alterações económicas, políticas, ou tecnológicas globais;
- Alterações ao nível das empresas presentes nos mercados – no caso dos mercados acionistas ou obrigacionistas, importantes alterações podem resultar, por exemplo, da fusão entre duas empresas concorrentes, ou da falência de uma empresa.

Dadas algumas observações passadas do comportamento de uma série, podem fazer previsões sobre o seu comportamento futuro. O estudo do comportamento das séries temporais já existe a algum tempo e tem recebido uma permanente atenção ao longo das últimas décadas utilizando modelos econométricos. No entanto, os modelos estão sujeitos a erros de previsão. Qualquer que seja a metodologia utilizada o objetivo é minimizar estes erros (Cortez, 2000). Nesta tese, vamos fazer abordagem aos métodos

mais utilizados, os modelos ARMA de Box & Jenkins (1970) e o modelo GARCH de Bollerslev (1986). Os modelos ARMA são modelos muito utilizados para a previsão de séries temporais, contudo na sua utilização, a tarefa de identificação do melhor modelo a ser considerado não é uma tarefa fácil. Tendo isso em conta, vamos utilizar o modelo GARCH para comparar os resultados.

Em termos gerais, pretende-se comparar modelos lineares univariados (ARMA) com modelos heterocedásticos, os modelos GARCH, verificando qual dos métodos é o mais eficiente e em que condições são aplicados a séries que tem um comportamento bastante irregular.

ESTRUTURA DA TESE

Esta dissertação divide-se em diversos capítulos, estruturados da seguinte forma:

Capítulo 1. Revisão Bibliográfica - neste capítulo é apresentada um histórico no qual são citados alguns dos principais trabalhos encontrados na literatura sobre análise de séries temporais e trabalhos relacionados ao tema desta tese.

Capítulo 2. Bolsa de Valores - neste capítulo é feita um enquadramento sobre bolsa de valores de Lisboa, tendo em conta a sua história. Também é feita uma revisão do principal índice nacional, PSI 20, e os seus principais marcos históricos.

Capítulo 3. Conceitos e Modelos Econométricos – neste capítulo são apresentados uma breve introdução sobre a análise de séries temporais, assim como a metodologia para a previsão de séries temporais.

Capítulo 4. Análise empírica dos dados - é feita uma análise e confronto dos modelos obtidos ao longo do nosso estudo. Na implementação prática, foi utilizado um suporte em termos de software, o Eviews, pelo que todas as representações gráficas e os outros outputs apresentados foram obtidos a partir deste software.

Capítulo 5. Conclusão - Completada a investigação e análise de todos os dados, este capítulo apresentará todas as conclusões retiradas com a realização da dissertação.

1. REVISÃO BIBLIOGRÁFICA

O principal objetivo da análise de séries temporais é prever valores futuros da série que pode ser a curto prazo ou a longo prazo (Morettin; Tolo, 1985). Durante as últimas décadas o estudo de séries temporais tem sido foco de discussões mediante distintas abordagens. Um número crescente de métodos e técnicas estatísticas tem sido desenvolvido e testado com o objetivo de obter previsões de séries temporais e minimizar os erros.

Até 1920 a previsão de séries temporais era efetuada através da extrapolação de ajustes lineares dos dados históricos conhecidos, realizada manualmente com base em representações gráficas. Em 1927, Yule criou o modelo Autorregressivo (AR), segundo este modelo o valor futuro de uma série depende de valores passados. Desde então durante um período longo os pesquisadores consideravam que o comportamento de uma série dependia dos modelos lineares e do ruído. Com o passar do tempo, os pesquisadores achavam que este método não era suficiente para a análise de séries temporais, uma vez que a maioria das séries apresentam tendências de não linearidade (Chaves, 1991).

Novos estudos foram realizados a partir de 1950 e destacou-se a técnica do Alisamento Exponencial com poucos cálculos e fácil de utilizar. Nesta mesma época com o aparecimento do computador, iniciou uma evolução acentuada nos métodos de previsão.

Em 1965, Fama defende a hipótese da eficiência do mercado, afirmando que as decisões são tomadas de modo a que todos os investidores possuem as mesmas informações e estão ao mesmo tempo no mercado. Colocou ênfase na natureza estocástica do comportamento dos mercados financeiros, caracterizados por uma dependência não-linear e a partir daqui vários estudos surgiram com esta abordagem.

Em 1975, surgiu a metodologia Box & Jenkins, os modelos ARMA e ARIMA (Auto-Regressivos e de Médias Móveis e Auto-Regressivos e de Médias Móveis Integrados). Nesta metodologia trata-se de uma técnica de identificação do melhor modelo que se ajuste aos dados, da estimação dos parâmetros e da verificação da validade do modelo escolhido. Estes modelos fornecem previsões através de um modelo linear e apresentam algumas limitações, como por exemplo, a escolha do melhor modelo não ser fácil. A metodologia Box & Jenkins é um dos resultados de maior

importância e reconhecimento na área de Previsão de Séries Temporais. O estudo feito pelos cientistas George Box e Gwilyn Jenkins (1970) é baseado no importante resultado de Wold (1954) que provou que qualquer série temporal pode ser representada por uma estrutura de médias móveis infinita (CHAVES, 1991). A metodologia proposta consiste em ajustar modelos Auto-Regressivos e de Médias Móveis, ARMA (p, q) à série temporal. Desde então, várias técnicas têm sido testadas com o objetivo de efetuar previsões mais precisas.

Em 1982, surgiu o modelo ARCH (*Autoregressive Conditional Heteroskedasticity*) proposto por Engle. O modelo surge como uma alternativa as distribuições empíricas de dados e pretende responder a evidência de que a variância de dados não é constante ao longo do tempo.

Por sua vez, Bollerslev em 1986 desenvolveu o modelo GARCH (*Generalized Autoregressive Conditional Heteroskedasticity*). Este modelo constitui uma generalização do modelo ARCH e consiste em modelar a variância condicionada não só em função do quadrado dos erros passados, mas também em função dos próprios valores passados da variância condicionada.

Posteriormente, com o objetivo de melhorar a análise da variância condicional das séries temporais financeiras, surgiram variações que tem como base os modelos ARCH e GARCH. Black (1976), constatou que os preços das ações tendem a ser negativamente correlacionadas com as variações na volatilidade (variância), ou seja, períodos de queda nos preços são frequentemente seguidos por períodos de grande volatilidade, do que nos períodos, cujos preços de volatilidade sejam elevados, a volatilidade não é tão intensa. Este efeito é denominado de *leverage* (efeito de alavancagem) ou efeito assimétrico. Neste contexto, foram desenvolvidas algumas extensões que incorporam o problema da distribuição assimétrica. Um dos primeiros modelos da GARCH assimétrico foi o EGARCH (*Exponential Generalized Autoregressive Conditional Heteroskedasticity*), proposto por Nelson em 1991 e Glosten, Jagannathan e Runkle (1993) desenvolveram o modelo TGARCH (*Threshold Generalized Autoregressive Conditional Heteroskedasticity*).

Cortez (2002) propôs uma abordagem híbrida entre Programação Genética e Redes Neurais na previsão de séries temporais, em sua tese de doutorado. Os resultados foram comparados com métodos de previsão tradicionais, como alisamento exponencial e a metodologia Box & Jenkins, tendo concluído que os modelos inspirados na natureza

produzem melhores resultados. Segundo Cortez “o desejo de compreender o passado e prever o futuro impulsiona a procura por leis que expliquem o comportamento de certos fenômenos ou acontecimentos. Se as equações que os determinam são conhecidas, é possível utilizar as mesmas para prever o resultado de um determinado experiência, desde que sejam conhecidas as condições iniciais. Na ausência de regras que definam o comportamento de um sistema, procura-se determinar o seu comportamento futuro a partir de observações concretizadas no passado”.

Segundo Bentes (2011), a evolução registada dos modelos ARCH (q) para os GARCH (p, q) pode comparar-se à ocorrência dos modelos AR (p) para os ARMA (p, q) de Box e Jenkins (1976).

Segundo Brooks (2002) as séries financeiras podem passar por episódios onde o comportamento da série muda. Esta mudança pode ser em termos da média, da volatilidade, ou o padrão de autocorrelação. As causas podem ser múltiplas, desde mudanças macroeconómicas e políticas, até mudanças com intervalos regulares ou de maior frequência. Resultados da literatura relacionada apontam ainda que o retorno pode apresentar um comportamento distinto dependendo da evolução do volume negociado.

Alguns dos trabalhos encontrados na literatura que estão relacionados com a previsão de séries temporais são:

- Marta Oliveira (2014) desenvolveu uma análise econométrica, para determinar um modelo que permite estimar e analisar o valor do fundo de pensões. Considerou uma amostra de variáveis provenientes da Associação Portuguesa de Fundos de Investimentos, Pensões e Patrimónios (APFIPP), num determinado momento de tempo, utilizando os modelos ARMA e ARIMA.

- Carla Amaral (2015) fez uma análise de modelos de previsão do Value- At-Risk aplicados ao índice PSI 20. O objetivo é descrever modelos heterocedásticos aplicados na previsão e modelação do risco de mercado e, conseqüentemente, fazer a previsão do *VaR* associado ao principal índice bolsista português. Recorreu-se, assim, aos modelos GARCH (1,1), EGARCH (1,1).

- Sandra Barão (2008). Esta tese pretende contribuir para uma melhor compreensão do problema relacionado com a previsão de variáveis provenientes de mercados financeiros (Portugal, Espanha, França, Alemanha, Itália, Grécia, Reino Unido, Japão, Estados Unidos). Para tal comparou modelos lineares (por exemplo, ARIMA) com modelos não lineares de forma a perceber quais os modelos que obtêm uma maior *performance*/poder preditivo.

- Ana Furiel (2011) descreveu em sua tese de Mestrado os modelos Heterocedásticos - ARCH e GARCH como forma de modelar e prever o valor médio condicional (rendimento) e a variância (volatilidade) de séries financeiras.

- Paula Moldovan (2015) fez uma análise de valores futuros da taxa de câmbio mensal entre o Euro e o Dolar Americano. Para obter a previsão utilizou modelos econométricos lineares e não-lineares, nomeadamente, ARMA (Auto Regressive Moving Average) e STAR (Smooth Transition Auto Regression). Conclui que para o curto e médio prazo os modelos lineares têm uma *performance* melhor do que os modelos não-lineares. A qualidade de forecast foi avaliada pelo valor do erro quadrático médio (RMSE).

2. BOLSA DE VALORES

A Bolsa de Valores é um mercado organizado onde se transacionam (compra/venda) ações de sociedades de capital aberto (públicas ou privadas) e outros valores mobiliários, como as opções e obrigações. As compras e as vendas de títulos são realizadas por corretores e por sociedades de corretagem (que só podem atuar no mercado por conta de nutrem) e as sociedades financeiras de corretagem (que podem atuar por conta própria gerindo carteiras de cliente. A bolsa de valores tem as seguintes funções:

- Assegurar e proporcionar liquidez aos títulos;
- Analisar e avaliar os valores mobiliários;
- Canalizar a poupança para o mercado primário;
- Formar os preços dos valores mobiliários no mercado.

Em Portugal, a entidade encarregue de supervisionar e regular os mercados de valores mobiliários e instrumentos financeiros derivados e a atividade de todos os agentes que neles atuam é a Comissão do Mercado de Valores Mobiliários (CMVM), sendo um organismo público independente, com autonomia administrativa e financeira (CMVM, 2014).

2.1. BOLSA DE VALORES DE LISBOA

A Bolsa de Valores de Lisboa (BVL) também designada Euronext Lisbon é uma das seis bolsas europeias que em conjunto formam a Euronext. A Euronext Lisbon foi constituída em 2002, quando as ações da Bolsa de Valores de Lisboa e Porto foram adquiridas pela Euronext.

Nos últimos anos, a bolsa tem conhecido importantes e amplas mutações que vão desde o próprio enquadramento legal, à sua estrutura funcional e aos sistemas de negociação também como a integração na plataforma internacional Euronext,

supervisionada e fiscalizada pela CMVM (Comissão do Mercado dos Valores Mobiliários).

A Euronext, foi formada em 2000 pela fusão das Bolsas de Paris, Amesterdão, Bruxelas. Em 2002 o grupo alargou-se, com a entrada da Bolsa de Valores Lisboa e de Porto e a Bolsa de Londres LIFFE. Em 2007 fundiu-se com a NYSE, gerando NYSE Euronext e em junho de 2014, completou uma oferta pública inicial tornando novamente uma empresa autónoma. A Euronext é o primeiro mercado **pan-europeu** e um dos maiores mercados bolsistas mundiais. Veio permitir que o mercado de capitais português acompanhasse o desenvolvimento das bolsas internacionais, e permitisse aos investidores e às empresas cotadas acesso a mercados dotados de elevada liquidez e profundidade (Justino:1994).

A *Euronext Lisbon* tem como objetivo principal a gestão das bolsas, podendo ainda:

- Gerir outros mercados de valores mobiliários;
- Gerir sistemas de liquidação de valores mobiliários;
- Prestar outros serviços relacionados com a emissão e a negociação de valores mobiliários que não constituam atividade de intermediação;
 - Prestar aos membros dos mercados por si gerados os serviços que se revelem necessários à intervenção desses membros em mercados geridos por entidade congénere de outro Estado, com quem tenha celebrado acordo.

2.1.1. HISTÓRIA DA BVL

As referências mais remotas relativas ao aparecimento das bolsas centram-se na Idade Média. Com o desenvolvimento do comércio originou-se um maior contato entre os comerciantes e a presença frequente de negociantes estrangeiros atraiu os corretores que facilitavam as transações de mercadorias.

A primeira tentativa de regular a atividade dos corretores de Lisboa surgiu em 1495, e na segunda metade do século XVIII surgiram as primeiras emissões de ações e os títulos de dívida pública.

Em Portugal, a primeira bolsa de valores foi criada em 1769, como a Assembleia dos Homens de Negócio na Praça do Comércio, no centro de Lisboa. Com o regulamento de 1837, que consagra a existência de corretores de valores, e com a aprovação, em setembro de 1833, do *Código Commercial Portuguez*, de Ferreira Borges, estabeleceu as regras de funcionamento das bolsas.

O Código Comercial Português, de Veiga Beirão, aprovado em 28 de junho de 1888, e o Regulamento das Bolsas de 1889, deram um impulso decisivo na criação das Bolsas de Valores, para se negociarem valores mobiliários e com este código regulamentou as operações a prazo.

Em janeiro de 1891 e em outubro de 1901, respetivamente, são criadas as Bolsas de Valores do Porto (BVP) e de Lisboa (BVL). Em 1891 surgiu a primeira crise financeira, atingindo o mercado de capitais português.

Em outubro de 1901 são aprovados o Regimento do ofício do Corretores e o Regulamento de Bolsa, que admite a especificidade jurídica do mercado de valores mobiliários. São regulamentadas as operações a contado e a prazo, as datas de liquidação das transações, os lotes máximos e mínimos, os horários de negociações e são previstas operações sobre produtos derivados. Esses dois documentos aprovados assumem uma importância no mercado de capitais até à Revolução de Abril de 1974. Durante este período, o mercado de título mostra pouco ativo, devido ao fato de a grande maioria dos títulos no serem objeto de negociação.

Após o período de conturbação que se sentiu devido a implementação da Republica e da I Guerra Mundial, segue a retoma da atividade bolsista no pós-guerra.

Como resposta à grande depressão, nos Estados Unidos da América são publicados em 1933 e 1934, respetivamente, o *Securities Act* e o *Securities Exchange Act* e é criada a instituição de supervisão e regulação dos mercados e dos valores mobiliários: a *Securities and Exchange Commission*.

Em Portugal, 57 anos depois, surge a Comissão do Mercado dos Valores Mobiliários (CMVM).

A partir de 1932 o mercado de capitais português experimenta um período de recuperação e é interrompido pela II Guerra Mundial.

Em 1948 marca uma nova crise com efeitos negativos no mercado de capitais em Portugal. Em janeiro de 1950 é realizado em Bolsa o último conjunto de negociações a prazo e retomados com a negociação dos contratos de futuros OT-10 e PSI-20 na Bolsa de Derivados do Porto, em junho de 1996, aquando da negociação oficial.

A tomada do poder político pelos militares em 25 de Abril de 1974 constitui um grande impacto para as Bolsas de Valores de Lisboa e do Porto. Estas bolsas encerram de imediato, sendo que BVL reabre em 12 de janeiro de 1976 para as realizações de transações de Obrigações e em 28 de fevereiro para os negócios com ações. A BVP reabre em 12 de janeiro de 1981.

Em 1985 o Ministro das Finanças, Miguel Cadilhe, convida um conjunto de empresas a participar na Bolsa marcando uma grande mudança no mercado de capitais português.

Em julho de 1993 a BVP pede formalmente permissão à CMVM para implementar um mercado organizado de futuros e opções, e em março do ano seguinte é estabelecido um acordo de especialização de mercados, ficando a BVL com a responsabilidade de desenvolvimento do mercado a contado e a BVP, mais tarde redenominada Bolsa de Derivados do Porto, com o do mercado de produtos derivados.

Em junho de 1994 é criado o Mercado Especial de Operações por Grosso destinado ao registo de transações de grandes lotes de obrigações.

O Conselho Diretivo da CMVM aprovou em 8 de outubro de 1998 o Regulamento que contempla normas relativas à admissão direta à negociação, no mercado de cotações oficiais, de ações emitidas por entidades estrangeiras. Trata-se de um passo importante para o mercado bolsista português, não só por ser suscetível de reforçar a capitalização da Bolsa de Lisboa e aumentar a sua visibilidade, mas também por se inserir no processo de integração e globalização dos mercados financeiros que se intensificará com a introdução do Euro.

Em 1999, é aprovado o novo regime jurídico que reestrutura e reorganiza as entidades gestoras de mercado de valores mobiliários, procedendo à sua desmutualização, dando-se assim origem à BVLP - Sociedade Gestora de Mercados

Regulamentados, SA., em resultado da fusão entre a Associação da BVL e a Associação da BDP.

Em julho de 2000 as transações com obrigações da dívida pública passam a ser efetuadas no Mercado Especial de Dívida Pública.

2.2. ÍNDICE BOLSISTA

Os índices bolsistas representam a evolução da cotação de um determinado mercado, ou seja, representam a sua evolução em termos do valor que representa, incluindo as ações mais representativas desse mesmo mercado. Indicam aos investidores se, em média, as cotações apresentam uma tendência de subida ou de descida. Isto quer dizer que a subida do índice não significa que uma determinada ação, em particular, se tenha valorizado. Do mesmo modo, a valorização de um só título pode não ter um efeito significativo no índice.

Para calcular um índice de bolsa, podem considerar-se todas as ações cotadas numa Bolsa. Contudo, também é possível limitar o índice às ações mais importantes ou mais representativas de uma bolsa, que são ponderadas de forma a refletir a sua importância no mercado, ou calcular vários índices consoante os vários setores de atividade.

Existem vários índices bolsistas no mundo, mas os maiores e principais encontram-se nos Estados Unidos da América, na Europa e na Ásia.

Nos Estados Unidos encontra-se a bolsa de ações de New York que regista os maiores negócios de ações, servindo de referência aos outros mercados bolsistas mundiais. O mais antigo índice bolsista do mundo é o Dow Jones, na New York Stock Exchange, que reúne as ações das 30 empresas líderes neste país.

Outros índices bolsistas importantes nos Estados Unidos são o NASDAQ-100, o S&P 500 e o Russell 2000. Ainda no grande continente americano se localizam os índices IBOVESPA (Brasil), Merval (Argentina), IPSA (Chile) e IPC (México).

Na Europa destaca-se o índice de Londres, o segundo maior mercado bolsista do mundo. Trata-se do FTSE 100 na bolsa de valores de Londres. Sublinhe-se também o CAC 40 na Euronext Paris, o EUROSTOXX 50 da Zona Euro, o AEX na Euronext

Amsterdão, o BEL 20 na Euronext Bruxelas, o DAX em Frankfurt, o IBEX 35 na Bolsa de Madrid e o PSI 20 na Euronext Lisboa.

Na Ásia encontra-se o mercado de ações de Tóquio, o Nikkei 225. É de relevar também os índices Kospi (Coreia do Sul), BSE SENSEX (Índia), SSE 180 (China), Hang Seng (Hong Kong).

2.3. PSI 20

O PSI 20 é a abreviatura de *Portuguese Stock Index* e representa o principal índice bolsista português. Acompanha os preços das 20 empresas com maior capitalização de mercado e volume de negócios das ações no PSI Geral, mercado de ações geral da Bolsa de Lisboa. Devido às suas características, o índice PSI-20 tem vindo a ser selecionado pelo mercado para servir de subjacente a produtos estruturados, cuja rentabilidade depende do comportamento do mercado bolsista português.

Foi lançado com dupla finalidade: atuar como base para contratos de futuros e opções e atuar como referência para o mercado de ações nacional. É um dos principais índices do grupo pan-europeu de bolsa de valores Euronext ao lado da BEL 20 de Bruxelas, do CAC 40 de Paris e do AEX de Amesterdão.

Como acontece com outros índices acionistas mais populares, existem investidores profissionais que investem em contratos de futuros ou opções do PSI 20, em vez de ganharem com a subida de uma ação ganham com a valorização do índice.

O índice português é revisto a cada três meses, sendo que a revisão anual é feita em março e conhecida em janeiro. O índice é calculado durante o período de sessão regular e divulgado de 15 em 15 segundos. A capitalização bolsista das empresas que compõe o PSI 20 é ajustada pelo *free float market Capitalization*, empresas que estão dispersas em bolsas.

Se uma empresa quiser entrar no PSI 20, tem de cumprir os seguintes requisitos:

- O valor do *free float*, percentagem de ações disponíveis para negociação em bolsa, não pode ser inferior a 15%

- O valor da capitalização bolsista tem de ser superior a 100 milhões de euros
- As ações de uma empresa devem ter um *free float velocity* de pelo menos 25%

Com o desmoronamento do BES e da Portugal Telecom em 2014, dois títulos emblemáticos da bolsa nacional, o PSI 20 passou a contar com 18 empresas, número mínimo de cotadas que pode ter, e não 20 empresas como é suposto. Nenhuma das empresas que se encontra fora do índice cumpriu os requisitos mínimos para ser incluída.

Segundo os dados apresentados pela Euronext a 30 de dezembro de 2016, as 18 empresas que compõem o PSI 20 atualmente apresentavam os seguintes valores do peso (%) e do *float*, sendo que as principais eram a Galp Energia, Jerónimo Martins, Energias de Portugal e a NOS, todas com um peso superior a 10%.

A tabela seguinte mostra a composição do índice a 30 de dezembro de 2016:

| Nome | Setor | Peso (%) | Float |
|---------------------------------|-------------------------------------|----------|-------|
| Altri | Produtos industriais gerais | 1.98 | 0.35 |
| Banco Comercial Português | Bancos | 4.22 | 0.8 |
| Banco Português de Investimento | Bancos | 3.53 | 0.3 |
| Corticeira Amorim | Bens de consumo | 1.62 | 0.2 |
| CTT | Transporte industrial | 6.56 | 0.95 |
| Energias de Portugal | Eletricidade | 12.49 | 0.85 |
| EDP renováveis | Eletricidade | 9.41 | 0.25 |
| Galp | Produção de Petróleo e Gás | 16.55 | 0.65 |
| Jerónimo Martins | Retalho alimentar e de medicamentos | 13.63 | 0.4 |
| Montepio | Bancos | 0.42 | 0.35 |
| Mota-Engil | Construção e materiais | 1.09 | 0.4 |
| NOS | Media | 10.38 | 0.5 |
| Pharol | Holding | 0.93 | 0.7 |
| Redes Energéticas Nacionais | Eletricidade | 4.63 | 0.45 |

| | | | |
|--------------------|-------------------------------------|------|------|
| Semapa | Silvicultura e papel | 2.35 | 0.3 |
| Sonae | Retalho alimentar e de medicamentos | 4.37 | 0.35 |
| Sonae Capital | Serviços Financeiros | 0.47 | 0.35 |
| The Navigator Comp | Silvicultura e papel | 5.37 | 0.3 |

Tabela 1. Composição de PSI 20 a 30 de dezembro de 2016

2.3.1 PRINCIPAIS ACONTECIMENTOS



Figura 1. Principais acontecimentos do PSI 20

O PSI 20 foi criado em 31 de dezembro de 1992 com um valor base de 3.000 pontos de índice. O seu historial começou a partir de 1998.

O índice experimentou uma volatilidade consideravelmente maior do que os principais mercados financeiros mundiais entre 1998 e 2000, causado pela incerteza nos mercados emergentes mundiais. Atingiu o seu máximo histórico, 14.822, em março de 2000.

No período entre 2000 e 2001, as bolsas mundiais enfrentavam os efeitos negativos de um *bear market*, grande queda, causado pelo estouro da bolha tecnológica.

O PSI20 desvalorizou, atingindo um mínimo nos meados de 2003, o que não acontecia desde 1996. A este momento de maior instabilidade, sucedeu-se um período de forte recuperação das ações, sobretudo verificado a partir da segunda metade do ano de 2003, período esse que também contagiou a bolsa portuguesa.

Entre 1 de janeiro de 2004 e meados de julho de 2007, o principal índice da bolsa português, acumulou ganhos positivos, onde em meados de 2007 atingiu valores próximos do máximo histórico de março de 2000.

Em julho de 2007 surgiu a crise do "*subprime*" nos Estados Unidos a dar origem à crise financeira de 2008. Esta crise levou as principais praças mundiais a uma queda vertiginosa. Em resultado da crise, a falência do banco de investimento norte-americano *Lehman Brothers*, em setembro de 2008, desencadeou uma série de eventos que se transformou nos anos seguintes na pior crise económica internacional, desde a grande depressão de 1929. Embora a crise tenha suas origens no mercado imobiliário dos Estados Unidos, os seus reflexos afetaram os indicadores económicos das principais economias mundiais. A transmissão da crise para a Europa afetou principalmente o grupo de países formado por Portugal, Irlanda, Itália, Grécia e Espanha. A crise que se iniciou nos mercados bancários, evoluiu em 2009 para uma crise das dívidas soberanas dos governos nacionais europeus. No entanto, a origem dos desequilíbrios económicos que contribuíram para a formação da crise europeia, decorreu principalmente das elevadas dívidas privadas, contraídas pelos bancos e instituições financeiras destes países, com o sistema financeiro internacional.

Muitos índices ainda não conseguiram recuperar deste então. E, desta vez, as consequências para a economia nacional foram mais nefastas e duradouras, já que, em 2011, Portugal foi obrigado a pedir ajuda externa para evitar a bancarrota.

A estes cenários juntou-se, em 2014, o caso problemático do BES e da Portugal Telecom. Com estes acontecimentos a confiança dos investidores no mercado nacional diminuíram. Nos mercados financeiros, a confiança é fundamental, sobretudo numa bolsa tão periférica como a de Lisboa.

Desde sempre notou-se que a entrada/saída de investidores estrangeiros da praça nacional tem um papel muito importante na evolução das cotações. Com efeito, a reduzida dimensão do mercado e, consequentemente a sua baixa liquidez, leva a que quando há investidores de maior dimensão a entrar ou sair do mercado, isso tenha um

impacto relevante na subida ou descida do mercado na sua globalidade. Logo, para uma subida mais sustentada e duradoura da bolsa de Lisboa é importante que a economia nacional retome o caminho do crescimento económico e que a sua imagem seja melhorada. Só assim será restaurada a confiança dos investidores que permitirá que voltem a investir em força no mercado nacional.

Apesar das quedas dos últimos anos, a bolsa de Lisboa está corretamente avaliada. Porém, há empresas que, sobretudo devido à desvalorização que sofreram, estão baratas e constituem uma boa oportunidade de investimento.

3. CONCEITOS E MODELOS ECONOMETRÍCOS

A econometria consiste no estudo da relação entre variáveis económicas utilizando instrumentos estatísticos e matemáticos. O método mais popular da econometria é a análise de regressão, dando a conhecer os efeitos que uma variável pode exercer sobre outra (s) ao longo do tempo. Tem como objetivo estimar, com base em dados históricos, relações económicas através da criação de modelos de modo a caracterizar o seu comportamento e fazer previsões através da comparação de modelos econométricos.

Neste capítulo pretende-se fazer uma abordagem ao estudo de séries temporais financeiras, tendo o objetivo principal usar metodologias para previsão.

3.1. ESTUDO DE SÉRIES TEMPORAIS

Uma série temporal (sucessão cronológica) é um conjunto de observações, y_1, y_2, y_3, \dots , ordenadas com intervalos de tempo iguais, t_1, t_2, t_3, \dots . Numa série temporal os dados são registados sequencialmente, num determinado período do tempo, portanto pressupõem-se a existência de correlação serial ao longo do tempo.

Um primeiro passo para modelar uma série consta na apresentação do seu gráfico ao longo do tempo. Permite-nos o estudo da sua tendência e sazonalidade, se sua variabilidade aumenta ou não com o tempo, ou seja, fornece informações sobre a estacionaridade da série. Abaixo temos um exemplo da representação gráfica de uma

série temporal, definida pelos valores y_1, y_2, \dots de uma variável y nos momentos t_1, t_2, \dots . Portanto, a variável y é uma função de t e é definida por $y = F(t)$.

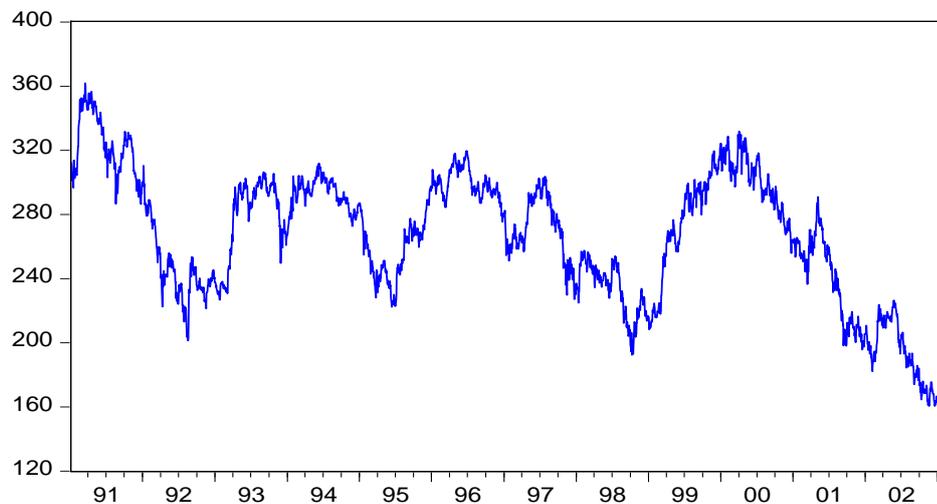


Figura 2.Exemplo de série temporal

Existem quatro tipos de componentes principais de uma série temporal: **Movimentos de tendência**, que é a direção geral pelo o qual o gráfico da série se desenvolve ao longo do tempo; **Movimentos ou variações cíclicas** que são oscilações a longo prazo ou a desvios em torno da reta ou da curva de tendência, essas variações podem ser ou não periódicas, ou seja, podem seguir ou não padrões exatamente análogos ao longo de intervalo de tempo iguais; **Movimentos ou variações por estações (sazonalidade)** que padrões idênticos a que uma série parece obedecer durante os mesmos períodos sucessivos e **Movimentos aleatórios** - deslocamentos esporádicos das sucessões cronológicas, provocados por acontecimentos casuais.

O fator de correlação é talvez a uma das mais poderosas ferramentas de análise que podemos utilizar no estudo de séries temporais. O conceito de covariância entre duas variáveis pretende avaliar até que ponto é que a variação de uma série está associada à variação da outra. Por sua vez, para duas variáveis aleatórias X e Y , a covariância mede a variabilidade conjunta dessas variáveis, e é dada por:

$$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)],$$

onde μ_X e μ_Y são as médias das respectivas variáveis X e Y .

O **coeficiente de correlação Pearson** entre as variáveis aleatórias X e Y mede a dependência linear entre elas e é definido por:

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

onde $Var(X)$ e $Var(Y)$ são as variâncias das respectivas variáveis X e Y .

Se $X = Y$, então temos autocorrelação. A sequência de coeficientes de autocorrelação designa-se de **função de autocorrelação** (ACF – *autocorrelation function*) e define-se por:

$$\rho(t,s) = \frac{Cov(y_t, y_{t-s})}{\sqrt{Var(y_t)Var(y_{t-s})}} = \frac{\gamma(t,s)}{\sqrt{\gamma(0)\gamma(0)}} = \frac{\gamma(t,s)}{\gamma(0)}$$

Se a correlação é estável ao longo do tempo, então ACF tem a forma:

$$\rho(t,s) = \rho(s) = \frac{\gamma(s)}{\gamma(0)}, \forall s \in T$$

O coeficiente de **autocorrelação parcial** de ordem s denota-se por $p(s)$ e é definido pelo coeficiente y_{t-s} no seguinte modelo linear:

$y_t = a + a_1 y_{t-1} + \dots + a_{s-1} y_{t-s+1} + r(s) y_{t-s} + \varepsilon_t$ onde ε_t não é correlacionado com y_{t-1}, \dots, y_{t-s} .

Mede a intensidade da relação entre duas observações da série:

$$y_t = a_0 + a_1 y_{t-1} + \varepsilon_t \Rightarrow p(1) = a_1$$

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \varepsilon_t \Rightarrow p(1) = a_2$$

A função de autocorrelação parcial (PACF) $p(s)$ é a correlação entre y_t e y_{t-s} , mantendo os y_h constantes para $t - s < h < t$.

Os coeficientes de autocorrelação para diferentes defasamentos temporais de uma variável, podem ser usados para identificar padrões de comportamento expressos por uma série cronológica. Se a série tem tendência, y_t e y_{t-1} são fortemente correlacionados e se a série tem forte componente sazonal, então um coeficiente de autocorrelação tende a ocorrer para o defasamento apropriado. As funções ACF e

PACF ajudam identificar as ordens dos modelos ARMA, também como visualizar propriedades dos resíduos resultantes destes modelos.

3.1.1. ESTACIONARIDADE

Para desenvolver e utilizar modelos capazes de testar e estudar dados económicos e financeiros, temos que verificar se as respectivas séries temporais são ou não estacionárias. Uma série é estacionária se seguir um processo estocástico com média e variância constantes ao longo do tempo e a covariância entre os valores desfasados (lag^1) da série depende apenas do valor do desfasamento.

Uma série $\{y_t\}_{t \in T}$ diz-se fracamente estacionária (estacionária em covariância) se:

- A média (valor esperado) é constante, isto é:

$$E(y_t) = E(y_{t-s}) = \mu, \forall t, t-s \in T$$

- A variância é constante e finita, isto é:

$$E((y_t - \mu)^2) = E((y_{t-s} - \mu)^2) = \sigma_y^2 < \infty, \forall t, t-s \in T$$

- A covariância é constante, isto é:

$$\begin{aligned} E((y_t - \mu)(y_{t-s} - \mu)) &= E(y_{t-s} - \mu)(y_{t-s-j} - \mu) = \\ &= \gamma(s), \forall t, t-s, t-s-j \in T \end{aligned}$$

A estacionaridade implica que os y_t são identicamente distribuídos e que a natureza de qualquer correlação entre termos adjacentes é a mesma ao longo de todos os períodos. Quando provamos estacionaridade referimos a estacionaridade fraca, mas contrário não se verifica, apenas para séries com uma distribuição Gaussiana. Uma série que não verifica as condições referidas acima, diz-se série não estacionária. As figuras

¹ Define-se operador lag como $L(y_t) = y_{t-1}$ e as suas iteradas de ordem superior $L^2 = y_{t-2}$. Especificamente, operamos sobre uma série não com o operador lag mas com um polinómio operador lag , isto é:

$$P(L)y_t = a_0y_1 + a_1y_{t-1} + \dots + a_ny_{t-n} = \sum_{k=1}^n a_ky_{t-k}$$

seguintes são exemplos de série estacionária (Figura 4) e serie não estacionária (Figura 3).

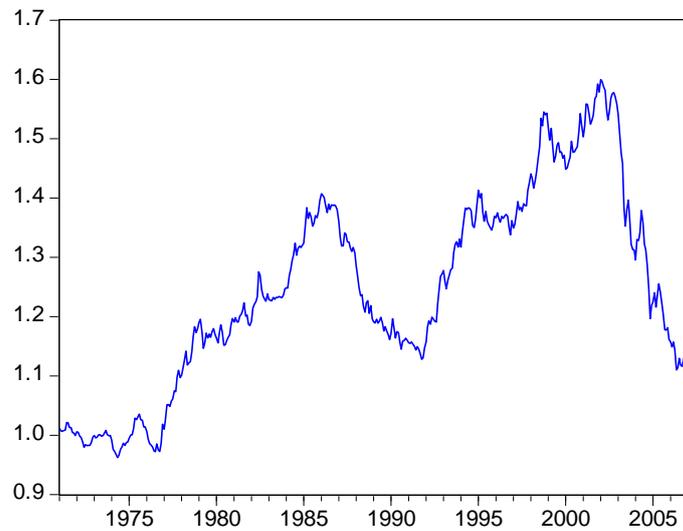


Figura 3. Exemplo de série não estacionária

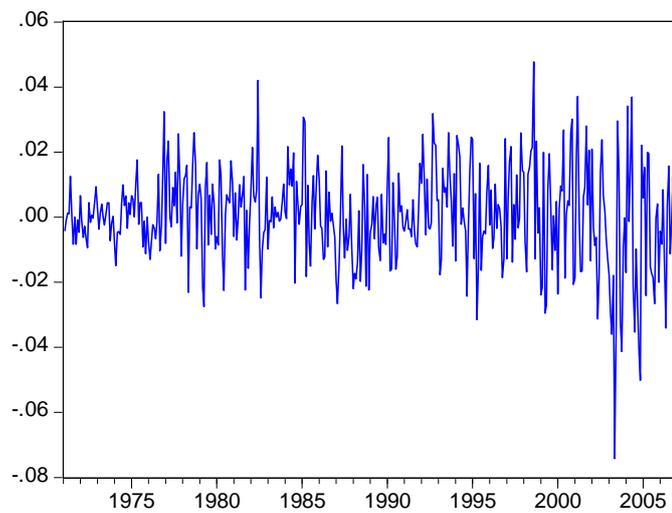


Figura 4.. Exemplo de série estacionária

Geralmente, as séries financeiras apresentam-se como séries não-estacionárias, sendo assim, vamos estudar a forma de estabilizar uma série não-estacionária. Uma série pode ser estabilizada (estacionarizada) fazendo:

- Diferenças sucessivas:

$$\nabla y_t = y_t - y_{t-1}; \quad \Delta^2 y_t = y_t - 2y_{t-1} + y_{t-2}$$

- Diferenças dos logaritmos dos dados:

$$\nabla \log y_t = \log y_t - \log y_{t-1}$$

Os dados originais, y_1, y_2, y_3, \dots , podem ser transformados em:

- Retornos simples:

$$R_t = \frac{y_t - y_{t-1}}{y_{t-1}} \times 100\%$$

- Retornos logaritmizados (rendibilidades)

$$R_t = \ln\left(\frac{y_t}{y_{t-1}}\right) \times 100\%$$

Uma série não estacionária pode ser transformada numa série estacionária após d diferenças sucessivas, sendo a nova série designada por série integrada e é associada a sua ordem de integração de ordem d , denotada por $y_t \sim I(d)$. Uma série estacionária diz-se integrada de ordem 0, $I(0)$.

A maioria das séries temporais financeiras apresentam-se como séries não estacionárias, ou seja, tem uma raiz unitária. Podem apresentar tendência determinística ou estocásticas. Para determinar se a série tem tendência determinística ou estocásticas vamos considerar o seguinte modelo:

$$z_t = \beta_0 + \beta_1 t + \rho z_{t-1} + \mu_t$$

Podemos ter em conta os seguintes casos. Se:

- $\rho = 0$, existe tendência determinística (processo estacionário em tendência puro), ou seja, processo estacionário em torno de uma tendência linear (media não constante e variância constante);
- $\rho = 1$ e $\beta_1 = 0$, existe tendência estocástica, ou seja, passeio aleatório com deriva- *drift*;
- $\rho = 1$, passeio aleatório com *drift* e tendência determinística;
- $0 < \rho < 1$, existe correlação serial em z , mas não existe tendência estocástica.

Verificar se uma série é não estacionária, é verificar se $\rho = 1$, ou seja, tem raiz unitária. Deve-se realizar testes estatísticos para testar a presença ou não de raiz unitária na série. Os testes de raízes unitárias são capazes de detetar se a série foi suficientemente diferenciada para se tornar estacionária.

Teste de raiz unitária:

$$H_0 : \rho = 1 \text{ (tem raiz unitária)}$$

$$H_1 : \rho < 1 \text{ (não tem raiz unitária)}$$

Se $\rho = 1$, temos que existe uma raiz unitária, pelo que a série é não estacionária.

Se $\rho < 1$, rejeitamos a hipótese nula, logo a série é estacionária.

Diversos testes de raiz unitária foram desenvolvidos sendo que as principais são: teste de **Dickey- Fuller (DF)** e teste de **Dickey- Fuller Aumentado (ADF- Augmented Dickey- Fuller)**. Abaixo é apresentado uma breve descrição dos dois testes.

Utilizamos o teste de DF quando os erros μ_t não tem correlação serial e o teste de ADF quando os erros μ_t tem correlação serial.

Teste de Dickey- Fuller (DF)

Subtraindo z_{t-1} no modelo de regressão $z_t = \beta_0 + \beta_1 t + \rho z_{t-1} + \mu_t$, obtém-se:

$$z_t - z_{t-1} = (\rho - 1)z_{t-1} + \beta_0 + \beta_1 t + \mu_t$$

A hipótese nula para o teste é dada por:

$$H_0 : \alpha = \rho - 1 = 0 \text{ (não- estacionária)}$$

$$H_1 : \alpha \leq 0 \text{ (Estacionária)}$$

Não se rejeita a hipótese nula se o valor do teste for superior aos valores críticos determinados para o nível de confiança 1%, 5% e 10%.

Teste de Dickey- Fuller aumentado (ADF)

O teste ADF considera valores desfasados ∇z das variáveis independentes (neste caso adicionamos 2 lags), ou seja:

$$\nabla z = z_t - z_{t-1} = (\rho - 1)z_{t-1} + \beta_0 + \beta_1 t + \beta_2 \nabla z_{t-1} + \beta_3 \nabla z_{t-2} + \mu_t$$

A hipótese nula é idêntica ao teste de DF e portanto, não se rejeita a hipótese nula se o valor do teste for superior aos valores críticos para o nível de confiança 1%, 5% e 10%.

O teste ADF para os resíduos é dado da seguinte forma:

$$\nabla e_t = (\rho - 1)e_{t-1} + \beta_0 + \beta_1 \nabla e_{t-1}$$

Os testes DF e ADF devem ser aplicados considerando as seguintes hipóteses:

| | |
|---|--|
| $H_0: z_t$ é um passeio aleatorio | $H_1: z_t$ é uma série estacionária com media zero |
| $H_0: z_t$ é um passeio aleatório com <i>drift</i> | $H_1: z_t$ é uma série estacionária com media $\frac{\beta_0}{1-\rho}$ |
| $H_0: z_t$ é um passeio aleatório com <i>drift</i> em torno de uma tendência determinística | $H_1: z_t$ é série estacionária em torno de uma tendência determinística |

Tabela 2. Diferentes hipóteses para os testes DF e ADF

3.2. MODELOS DE SÉRIES TEMPORAIS

Nesta secção vai ser abordado a metodologia e os modelos a serem utilizados para a previsão de séries temporais.

Nos últimos anos, foram desenvolvidas várias técnicas estatísticas de modelagem para a previsão de séries temporais, porém todas possuem algum erro de previsão, cuja grandeza depende maioritariamente da natureza e das características dos dados em estudo. Os modelos mais utilizados são modelos de séries temporais lineares que pertencem á família dos autorregressivos (AR - *Auto Regressive*), médias móveis (MA - *Moving Average*) e os modelos ARMA (*Auto Regressive Moving Average*), combinação dos modelos AR e MA. A metodologia de Box-Jenkins (1970) é a mais utilizada para a previsão de séries, no entanto a tarefa de identificação do melhor

modelo a ser utilizado não é fácil. Estes modelos assumem que a variância dos erros é constante ao longo do tempo (homocedasticidade) e que os erros são independentes. Neste sentido, pesquisas vêm sendo desenvolvidas, no sentido de obter modelos de previsão com o menor erro possível.

Nas abordagens econométricas mais recentes destacam-se os modelos da família ARCH/GARCH que permitem detetar e modelar manifestações de heterocedasticidade condicionada nos resíduos das séries temporais. Esses modelos procuram captar a volatilidade de autocorrelações, onde o risco de hoje depende do risco observado no passado, permitindo que a variância condicionada não seja constante e varie ao longo do tempo como uma função linear dos quadrados dos erros do passado.

De forma breve, a metodologia das séries temporais consiste nos seguintes passos: análise do gráfico (tendência, sazonalidade, etc), estatística descritiva (média, variância, normalidade, curtose), teste de raiz unitária (estacionaridade), estimar os parâmetros, avaliar o modelo, re-avaliar o modelo adequado e usar o modelo para a previsão.

3.2.1. MODELOS ARMA

O modelo ARMA foi introduzido combinando os modelos autorregressivos- $AR(p)$ e os modelos Média Móveis – $MA(q)$ com o objetivo de mostrar que o modelo $ARMA(p, q)$ pode modelar series temporais estacionarias desde que as ordens p e q estivessem devidamente especificados.

Box & Jenkins (1976) tornaram este modelo muito popular e a maioria das series temporais utilizavam este modelo na previsão.

Este modelo assume que a variância é constante ao longo do tempo. Esta hipótese, contudo, raramente se verifica, uma vez que as séries financeiras apresentam comportamentos alterados ao longo do tempo.

Para uma melhor compreensão do modelo ARMA vamos fazer uma breve abordagem aos modelos $AR(p)$ e $MA(q)$, para depois então combinando os dois modelos obter o modelo $ARMA(p, q)$.

MODELOS AUTOREGRESSIVOS AR (p)

Os modelos Autorregressivos de ordem p , $AR(p)$, tem a seguinte forma:

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \varepsilon_t, \quad \text{com } \varepsilon_t \sim WN(0, \sigma^2)$$

onde p é a ordem de desfasamento (*lag order*) e ε_t é um processo de ruído branco que assinala a nova informação recebida no passo t . Podemos escrever a equação, $AR(p)$, de forma equivalente se utilizamos o operado *lag*, como:

$$\Phi(L)y_t = a_0 + \varepsilon_t \quad \text{onde } \Phi(L) = 1 - a_1 L - a_2 L^2 - \dots - a_p L^p$$

O processo $AR(p)$ é estacionário se e só se todas as raízes $z = \frac{1}{\alpha_i}$ do polinómio

$$\Phi(z) = 1 - a_1 z - a_2 z^2 - \dots - a_p z^p = (1 - \alpha_1 z)(1 - \alpha_2 z) \dots (1 - \alpha_p z)$$

são situadas no exterior do círculo unitário, $|\alpha_i| < 1$.

A função PACF de um processo $AR(p)$ é dada por:

$$r(p) \begin{cases} a_h \neq 0 \text{ para } h < p \text{ (existe PACF)} \\ 0, \text{ para } h > p \text{ (não existe PACF)} \end{cases}$$

A função PACF denota a convergência brusca para zero ao fim de p passos. A função ACF tem uma convergência lenta para zero.

MODELOS MÉDIA MOVEL MA (q)

Os modelos Média Móvel (*Moving Average model*) de tipo $MA(q)$, são definidos da seguinte forma:

$$\begin{aligned} y_t &= \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \\ &= \varepsilon_t + \theta_1 L \varepsilon_t + \theta_2 L^2 \varepsilon_t + \dots + \theta_q L^q \varepsilon_t \end{aligned}$$

O modelo do tipo $MA(1)$, é dada por:

$$y_t = \epsilon_t + \theta\epsilon_{t-1} = (1 + \theta L)\epsilon_t \text{ onde } \epsilon_t \sim WN(0, \sigma^2)$$

Num modelo $MA(1)$ o valor atual da série observada é dado como uma função de ruído branco atuais e passados. A média e a variância são dados como:

$$E(y_t) = E(\epsilon_t) + \theta E(\epsilon_{t-1}) = 0$$

$$Var(y_t) = Var(\epsilon_t) + \theta^2 Var(\epsilon_{t-1}) = \sigma^2 + \theta^2 \sigma^2 = \sigma^2(1 + \theta^2)$$

Podemos observar que a variância cresce à medida que o coeficiente θ cresce.

No modelo $MA(1)$ a função ACF é definida por:

$$p(t) = \frac{\gamma(t)}{\gamma(0)} = \begin{cases} \frac{\theta}{1 + \theta^2}, & t = 1 \\ 0, & \text{caso contrário} \end{cases}$$

A função ACF denota a convergência abrupta para zero ao fim do primeiro passo e a função PACF converge gradualmente para zero.

Generalizando, para o modelo $MA(q)$, temos que a média, a variância e a função de autocorrelação são dados como:

- $E(y_t) = 0$
- $Var(y_t) = (1 + \theta_1^2 + \dots + \theta_q^2)\sigma^2$
- $\gamma(s) = \begin{cases} (\theta_s + \theta_{s+1}\theta_1 + \dots + \theta_q\theta_{q-s})\sigma^2 & \text{para } s = 1, 2, \dots \\ 0 & \text{para } s > q \end{cases}$

A função ACF de um modelo $MA(q)$, anula-se no lag $(q + 1)$ e a função PACF converge suavemente para zero. O modelo $MA(q)$ é fracamente estacionária para todos os valores de $\{\theta_u\}$.

MODELOS ARMA (p, q)

Combinando os modelos $AR(p)$ e $MA(p)$ obtem-se o modelo $ARMA(p, q)$. O modelo é descrito da seguinte forma:

$$\Phi(L)y_t = a_0 + \Theta(L)\epsilon_t$$

onde

$$\Phi(L) = 1 - a_1L - a_2L^2 - \dots - a_pL^p \text{ e}$$

$$\Theta(L) = 1 + \theta_1L + \theta_2L^2 + \dots + \theta_qL^q$$

ou

$$y_t = a_0 + a_1y_{t-1} + \dots + a_py_{t-p} + \epsilon_t + \theta_1\epsilon_{t-1} + \dots + \theta_q\epsilon_{t-q} + \epsilon_t,$$

$$\text{com } E(\epsilon_t) = 0, \quad E(\epsilon_t^2) = \sigma^2, \quad E(\epsilon_t \epsilon_s) = 0, \quad t \neq s$$

A média de uma serie $ARMA(p, q)$ é dada por:

$$E(y_t) = \frac{a_0}{1 - a_1 - a_2 - \dots - a_p}$$

O modelo $ARMA(p, q)$ é estacionário se $|1 - a_1 - a_2 - \dots - a_p| < 1$. Se a série não é estacionária, utilizam-se os modelos $ARMA$ integrados, ou seja, modelos $ARIMA$. Um modelo $ARIMA(p, d, q)$ é um modelo $ARMA(p, q)$ aplicado a uma série que foi integrada (diferenciada) d vezes para obter a estacionaridade.

A função de autocorrelação de um modelo $ARMA(p, q)$ combina comportamentos dos modelos $AR(p)$ e $MA(q)$ e quando estamos a considerar *lags* inferiores a q , a função é idêntica ao do modelo $AR(p)$.

METODOLOGIA DE BOX- JENKINS

A metodologia de Box- Jenkins segue os seguintes passos:

Identificação do modelo através de métodos gráficos da série temporal, ACF, PACF para determinar as ordens p e q do modelo e através de critérios de informação: o objetivo é escolher o modelo que minimiza o valor do critério de informação utilizando os valores de AIC (*Akaike Information Criterion*).

Estimação dos parâmetros do modelo: pelo método dos mínimos quadrados, OLS (*Ordinary Least Square*).

Diagnostico: Após a estimação dos parâmetros, deve-se verificar se o modelo se ajusta bem aos dados, para isso é feita uma análise dos seus resíduos (correlação entre os resíduos, homocedaticidade dos resíduos e normalidade). O correlograma ajuda em identificar o tipo de modelo.

Previsão: após escolher o melhor modelo para a série, o modelo será usado para fazer a previsão de valores futuros da série.

3.2.2. MODELOS ARCH/GARCH

Até ao início da década de 80 os modelos (ARMA, ARIMA) utilizados na previsão de séries centravam-se na média condicionada. Como vimos anteriormente, estes modelos assumem que a variância dos erros é constante ao longo do tempo (homocedasticidade), isto é, $var(u_t) = \sigma_u^2$, o que falha bastantes vezes quando trabalhamos com dados. Assim a via mais correta para analisar o comportamento das séries consiste em modelar não só a média como também a variância condicionada (heterocedasticidade condicional).

Neste contexto, surge o modelo ARCH (*Autoregressive Conditional Heteroskedasticity*), proposto por Engle em 1982. O modelo ARCH pretende responder à evidência empírica de que a variância não é constante ao longo do tempo. Posteriormente, Bollerslev em 1986 desenvolveu o modelo GARCH (*Generalized Autoregressive Conditional Heteroskedasticity*) que incorpora a própria variância condicionada, observada no passado, ao modelo ARCH.

3.2.2.1. MODELO GARCH

Os modelos *GARCH* (p, q) consideram que a variância condicionada é parametrizada como função linear dos quadrados dos erros passados e das variâncias condicionadas passadas e são definidos por:

$$y_t = b_1 + b_2 y_{t-1} + \dots + b_k y_{t-k} + \mu_t, \quad \mu_t \sim N(0, \sigma_t^2) \quad \text{equação da média}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \mu_{t-1}^2 + \dots + \alpha_q \mu_{t-q}^2 + \beta_1 \sigma_{t-1}^2 + \dots + \beta_q \sigma_{t-q}^2 \quad \text{equação da variância}$$

onde σ_t^2 é a variância condicionada, α_0 é a constante (*intercept*), μ_{t-q}^2 é o termo de volatilidade (informação da volatilidade do período anterior) e σ_{t-q}^2 é a variância prevista do último período.

Geralmente, um modelo *GARCH* (1, 1) é suficiente para estudar a volatilidade dos dados, é entendida como a média ponderada da variância incondicional, do desvio do último período previsto da variância incondicional e o desvio do último período dos quadrados dos desvios da variância. É um tipo de modelo *ARMA* (1, 1) para a equação da variância. Define-se por:

$$y_t = b_1 + b_2 y_{t-1} + \dots + b_k y_{t-k} + \mu_t, \quad \mu_t \sim N(0, \sigma_t^2) \quad \text{equação da média}$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \mu_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad \text{equação da variância}$$

onde σ_t^2 é a volatilidade condicionada, α_0 é a constante (*intercept*), α_1 mede a forma como um choque no retorno de hoje afeta a volatilidade desse mesmo retorno no dia seguinte, μ_{t-1}^2 é o termo da volatilidade (fornece informação sobre a volatilidade do período anterior), σ_{t-1}^2 é a variância prevista do último período.

A soma ($\alpha_1 + \beta_1$) reflete a maneira como o impacto de um choque no retorno hoje propaga-se ao longo do tempo, sobre a volatilidade dos retornos futuros.

Os efeitos dos *shocks* ao longo do tempo são obtidos a partir da variância não-condicional de u_t e é definida por

$$Var(u_t) = \frac{\alpha_0}{1 - (\alpha_1 + \beta_1)}$$

quando $\alpha_1 + \beta_1 < 1$ garante estacionaridade e variância finita, quando $\alpha_1 + \beta_1 \geq 1$ temos não-estacionaridade em variância (os valores previstos da variância condicional não convergem para os seus valores não condicionados) e quando $\alpha_1 + \beta_1 = 1$ designa-se *GARCH* integrada (os *shocks* do passado persistem ao longo de períodos de longos tempo).

Para estimar um modelo *GARCH* temos de:

Especificar a equação para a média e para a variância, por exemplo, para um modelo *GARCH* (1,1) teremos:

$$y_t = \mu + \Phi y_{t-1} + \mu_t, \quad \mu_t \sim N(0, \sigma_t^2)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \mu_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

Especificar a função log-verossimilhança, L , e maximizar a função L , fornecendo os valores estimados dos parâmetros e os erros padrão, sendo L definida por:

$$L = -\frac{T}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^T \log(\sigma_t^2) - \frac{1}{2} \sum_{t=1}^T \left(\frac{(y_t - \mu - \Phi y_{t-1})^2}{\sigma_t^2} \right)$$

No entanto, o modelo GARCH apresenta algumas limitações: o modelo não consegue captar o efeito de alavancagem (*leverage*), a existência de uma estrutura assimétrica para a volatilidade gera distribuições enviesadas na previsão de preços, o modelo GARCH permite apenas identificar uma função para descrever a evolução da volatilidade através da sua modelação, não fornecendo informações sobre as suas causas.

4. ANÁLISE EMPÍRICA DOS DADOS

Neste capítulo vamos iniciar o nosso estudo sobre os dados considerados. Os dados referem-se a registos diários (5 dias uteis por semana) representativo dos valores do fecho do índice bolsista, PSI20. Esta série é denotada ao longo do trabalho por *psi20*. Os dados respeitam ao período entre 01 de novembro de 2004 e 18 de agosto de 2016 num total de 3079 observações (fonte: <https://www.euronext.com/pt-pt/products/indices/PTING0200002-XLIS>).

A análise foi feita utilizando o software Eviews.

Relativamente a análise da série, vamos iniciar o nosso estudo com a representação gráfica da série original em níveis (Figura 5), seguida posteriormente da observação de algumas estatísticas descritivas da mesma série (Figura 6).



Figura 5. Representação gráfica da série PSI20

Através de uma análise ao gráfico, podemos verificar que a série apresenta um padrão não linear com média e variância não constantes, indicando ser não estacionária, com tendências alternadas (crescente e decrescente). Podemos observar um crescimento acentuado nos primeiros anos, entre 2004 e 2007, com o índice a atingir um máximo histórico em julho de 2007. Entre meados de 2007 e 2008 verifica-se um decrescimento

bastante acentuado, em consequência da crise ‘*subprime*’ como referimos anteriormente. Entre os anos 2008 e 2016, nota-se uma alternância entre quedas e valorização do PSI20, com o índice a assumir um mínimo em 2012.

A figura seguinte apresenta o histograma e a estatística descritiva da série:

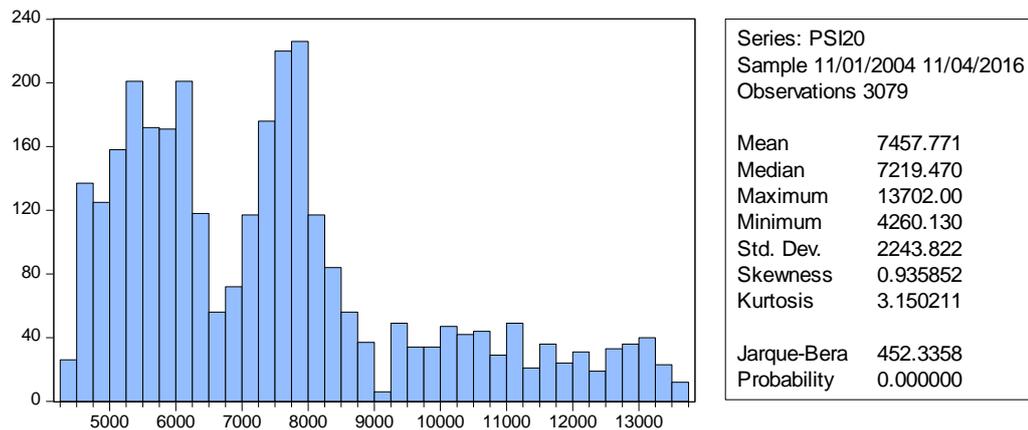


Figura 6. Histograma da série PSI20

Tendo em conta as estatísticas observadas, o histograma sugere a não normalidade da série, uma vez que o teste de Jarque-Bera apresenta um *p-value* aproximadamente igual a zero, sendo que rejeita-se a hipótese nula de existência de normalidade para o nível de significância 5%. Verifica-se um coeficiente de curtose próximo de 3 e assimetria moderada (*skewness*) positiva.

Vamos verificar a estacionaridade da série *psi20* através do teste de raiz unitária utilizando o teste ADF com a hipótese nula: a série tem raiz unitária (ou seja, serie *psi20* não é estacionaria).

| Null Hypothesis: PSI20 has a unit root | | |
|---|------------------|---------------|
| Exogenous: None | | |
| Lag Length: 1 (Automatic - based on SIC, maxlag=28) | | |
| | t-Statistic | Prob.* |
| Augmented Dickey-Fuller test statistic | -0.789097 | 0.3744 |
| Test critical values: | 1% level | -2.565714 |
| | 5% level | -1.940927 |
| | 10% level | -1.616630 |

Quadro 1. Teste de raiz unitária para a psi20

Como o *p-value* é maior que o nível de significância 5%, não rejeitamos a hipótese nula. Temos que a série tem raiz unitária, logo, não é estacionária.

Como a série *psi20* é não estacionaria, vamos trabalhar com os retornos (diferença de 1ª ordem dos logaritmos, que em geral, tornam as series estacionárias) da série logaritmizada, isto é: $rpsi = d(\log(\psi20))$. Podemos visualizar o gráfico da série *rpsi* e o respectivo histograma, nas Figuras 7 e 8.

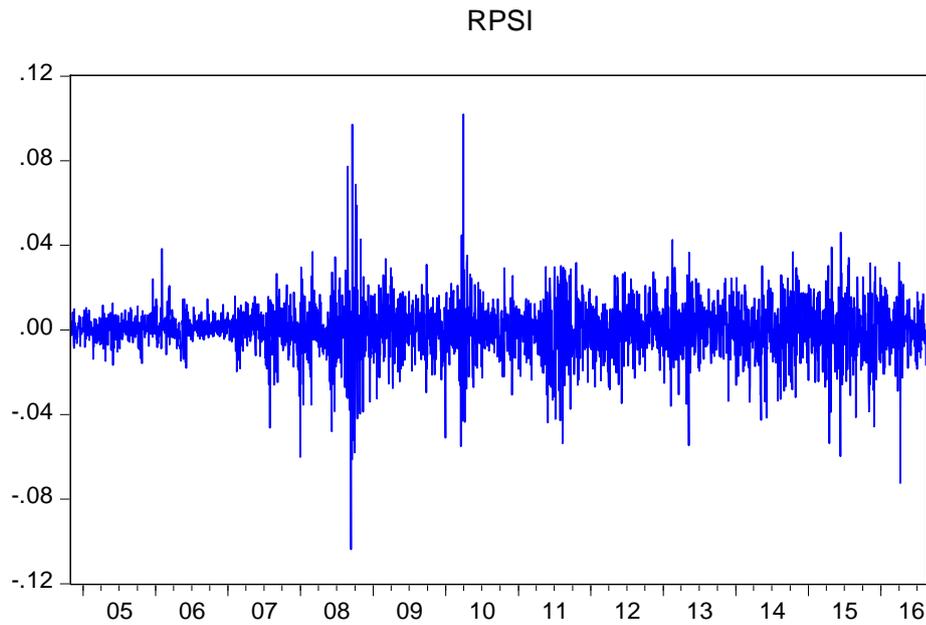


Figura 7. Representação gráfica da série rpsi

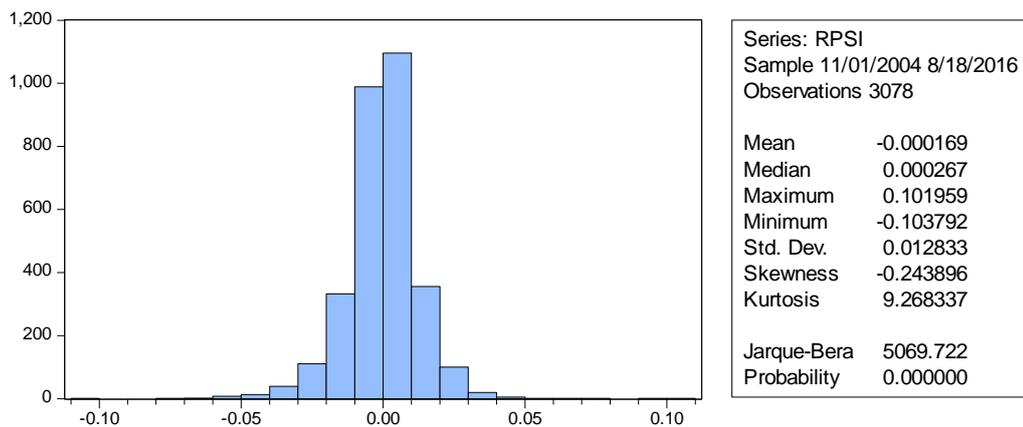


Figura 8. Estatística descritiva da série rpsi

Relativamente a esta série, pode-se concluir que a hipótese da normalidade se rejeita (o p -value no teste de Jarque-Bera é inferior que o nível de significância 5%), no entanto, o valor da média e do desvio padrão são próximos de zero. A série dos retornos logaritmizados, $rpsi$, apresenta uma assimetria negativa e é bastante leptocúrtica (o coeficiente de curtose é superior a 3). Observam-se clusters de volatilidade e alguns picos na variância que correspondem a crise de 2008.

A transformação da série original em retornos logaritmizados permitiu estabilizar o valor médio e diminuir a variabilidade dos dados. Temos agora o intuito que a série $rpsi$ é estacionária.

Para confirmar se a série $rpsi$ é estacionária, vamos ver se a série tem alguma raiz unitária, ou seja, vamos fazer o teste de raiz unitária (teste ADF) com a hipótese nula: a série tem raiz unitária (serie $rpsi$ não é estacionaria).

| Null Hypothesis: RPSI has a unit root | | |
|---|-------------|--------|
| Exogenous: None | | |
| Lag Length: 0 (Automatic - based on SIC, maxlag=28) | | |
| | t-Statistic | Prob.* |
| Augmented Dickey-Fuller test statistic | -50.98027 | 0.0001 |
| Test critical values: | | |
| 1% level | -2.565714 | |
| 5% level | -1.940927 | |
| 10% level | -1.616630 | |

Figura 9. Teste de Raiz Unitária aplicado a $rpsi$

Tendo em conta o teste de raiz unitária ADF, temos o p -value muito pequeno, o que conduz a rejeição da hipótese nula para os níveis de significância (H_0 : a série tem uma raiz unitária), logo, a série não tem uma raiz unitária, portanto é estacionária.

4.1. ESTIMAÇÃO DOS MODELOS ECONOMETRICOS

O correlograma informa os valores das funções de autocorrelação (ACF) e autocorrelação parcial (PACF) para os $lags$ especificados (para a série $rpsi$ o número de $lags$ é 36) e dá-nos alguma informação sobre o tipo de modelo a utilizar. Estas funções

caracterizam o padrão da dependência temporal da série e determinam a ordem dos modelos. Abaixo, apresentamos o correlograma da serie *rpsi*:

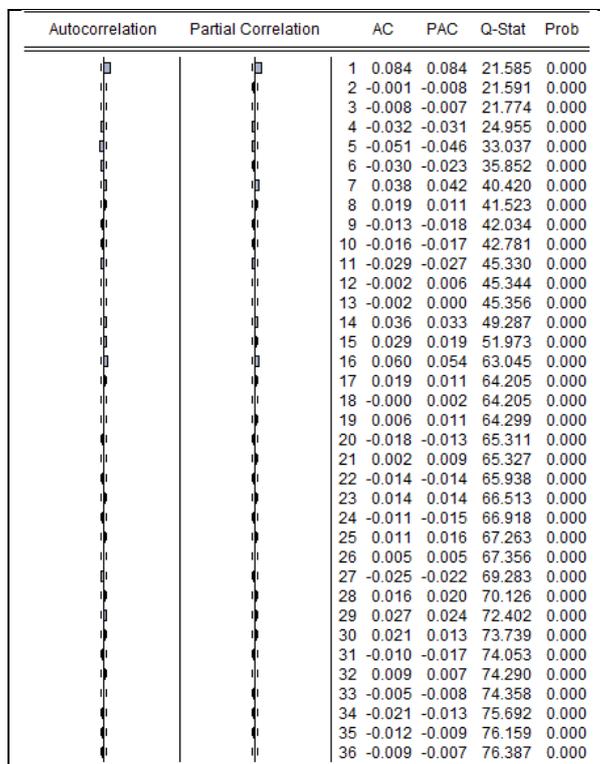


Figura 10. Correlograma da série *rpsi*

Da análise do correlograma, observamos que as funções ACF e PACF convergem para zero ao fim do primeiro *lag*, o que sugere a possibilidade de utilizar um modelo $ARMA(1, 1)$. Para escolher o melhor modelo que melhor se ajusta aos dados, deve-se experimentar vários modelos $AR(p)$, $MA(q)$ e $ARMA(p, q)$ e depois comparar o valor de AIC (*Akaike Information Criterion*). Considera-se o melhor modelo o que tem o valor do AIC mais pequeno.

Neste caso vamos estudar os seguintes modelos: $AR(1) = ARMA(1, 0)$, $MA(1) = ARMA(0, 1)$, $ARMA(1, 1)$, $ARMA(1, 2)$, e $ARMA(2, 1)$ para ver qual é o melhor que se ajusta a série. Os modelos são estimados utilizando o método da estimação dos mínimos quadrados. A tabela seguinte apresenta os valores do AIC para cada modelo em estudo, de modo a escolher o melhor que se ajusta aos dados:

| Modelos | AR(1) | MA(1) | ARMA (1, 1) | ARMA (1, 2) | ARMA(2, 1) |
|---------|-----------|-----------|-------------|-------------|------------|
| AIC | -5.879287 | -5.879639 | -5.878705 | -5.878701 | -5.878389 |

Tabela 3. Valores do AIC dos modelos.

De acordo com a tabela em cima, podemos observar que o $MA(1)$ apresenta o valor do AIC mais pequeno relativamente aos outros modelos estudados. Desta forma temos que o modelo $MA(1)$ é o melhor que se ajusta aos dados.

4.1.1. ESTIMAÇÃO DO MODELO ARMA (0,1)

A tabela seguinte (Quadro 2) é o *output* que resulta da estimação do modelo $MA(1) = ARMA(0,1)$:

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|--------------------|-------------|-----------------------|-------------|-----------|
| C | -0.000169 | 0.000250 | -0.676386 | 0.4988 |
| MA(1) | 0.084463 | 0.017969 | 4.700409 | 0.0000 |
| R-squared | 0.007072 | Mean dependent var | | -0.000169 |
| Adjusted R-squared | 0.006749 | S.D. dependent var | | 0.012833 |
| S.E. of regression | 0.012790 | Akaike info criterion | | -5.879639 |
| Sum squared resid | 0.503194 | Schwarz criterion | | -5.875719 |
| Log likelihood | 9050.764 | Hannan-Quinn criter. | | -5.878231 |
| F-statistic | 21.90725 | Durbin-Watson stat | | 1.999720 |
| Prob(F-statistic) | 0.000003 | | | |
| Inverted MA Roots | -0.08 | | | |

Quadro 2. Estimação do modelo MA (1)

Podemos observar que o valor do *p-value* associado ao modelo (pela estatística t e F) é muito pequeno o que sugere que a hipótese nula é rejeitada. Podemos considerar o modelo aceitável em termos globais, sendo o MA (1) estatisticamente significativo. O valor da estatística de Durbin-Watson é muito próximo de 2, sendo que os resíduos são independentes (para desfasamento 1) e as raízes invertidas são menores que 1 o que confirma a estabilidade do modelo. Deste modo, vamos analisar os resíduos para a validação do modelo $MA(1)$.

Para validar o modelo é necessário a análise dos resíduos e a validação dos mesmos. No entanto, se não conseguimos validar todos os pressupostos o modelo não é considerado válido para a previsão da série em estudo. Vamos visualizar o gráfico dos resíduos e o respectivo histograma:

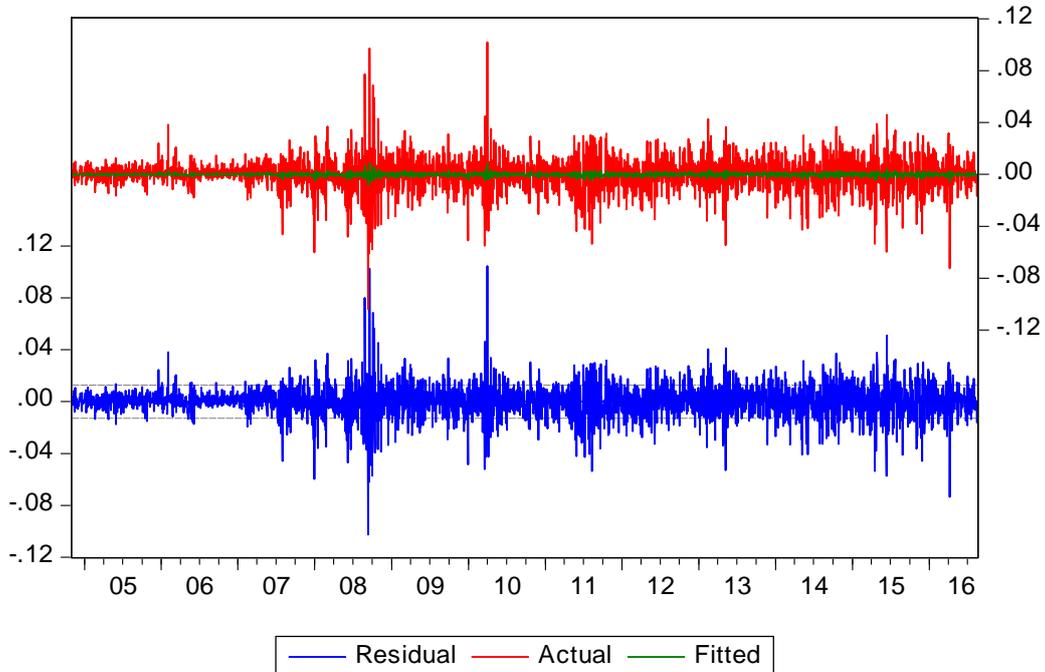


Figura 11. Gráfico dos resíduos da série rpsi

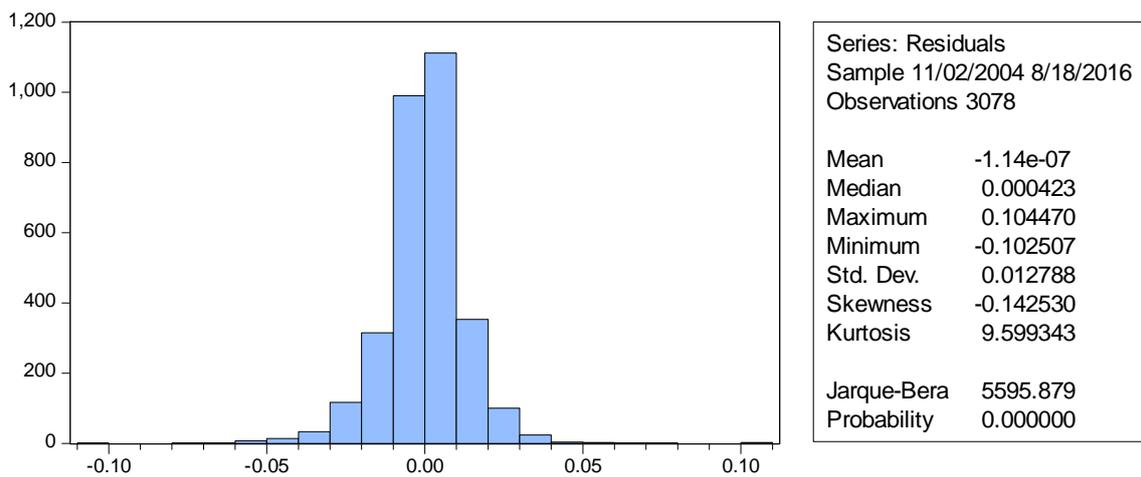


Figura 12. Gráfico e histograma dos resíduos da série rpsi20

Da análise do gráfico e do histograma, podemos verificar que os resíduos têm uma variação moderada, com existência de alguns *outliers*. Os resíduos têm media nula,

o desvio padrão é bastante próximo de zero. Apresentam assimetria negativa e o coeficiente de curtose é superior a 3 o que leva os resíduos a não seguirem uma distribuição normal. Pode-se concluir que a hipótese na normalidade se rejeita (o *p-value* no teste de Jarque-Bera é inferior ao nível de significância 5%).

O output seguinte mostra o teste de autocorrelação entre os resíduos, utilizando o teste *Breusch-Godfrey Serial Correlation LM Test*, onde a hipótese nula é definida por:

H_0 : Resíduos independentes vs

H_1 : Resíduos autocorrelacionados (não-independentes)

| F-statistic | 0.008797 | Prob. F(2,3074) | 0.9912 | |
|---|-------------|-----------------------|-------------|--------|
| Obs*R-squared | 0.017616 | Prob. Chi-Square(2) | 0.9912 | |
| Test Equation: | | | | |
| Dependent Variable: RESID | | | | |
| Method: Least Squares | | | | |
| Date: 09/11/17 Time: 05:32 | | | | |
| Sample: 11/02/2004 8/18/2016 | | | | |
| Included observations: 3078 | | | | |
| Presample missing value lagged residuals set to zero. | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | 1.10E-07 | 0.000250 | 0.000439 | 0.9996 |
| MA(1) | -0.306453 | 2.490637 | -0.123042 | 0.9021 |
| RESID(-1) | 0.306391 | 2.490683 | 0.123015 | 0.9021 |
| RESID(-2) | -0.026776 | 0.211139 | -0.126815 | 0.8991 |
| R-squared | 0.000006 | Mean dependent var | -1.14E-07 | |
| Adjusted R-squared | -0.000970 | S.D. dependent var | 0.012788 | |
| S.E. of regression | 0.012794 | Akaike info criterion | -5.878345 | |
| Sum squared resid | 0.503191 | Schwarz criterion | -5.870506 | |
| Log likelihood | 9050.773 | Hannan-Quinn criter. | -5.875529 | |
| F-statistic | 0.005864 | Durbin-Watson stat | 1.999710 | |
| Prob(F-statistic) | 0.999383 | | | |

Quadro 3. Teste Breusch-Godfrey Serial Correlation LM Test aplicado aos resíduos

Podemos verificamos que pelo teste *Breusch-Godfrey Serial Correlation LM Test* não rejeitamos a hipótese nula (*p-value* da estatística de F maior que o nível de significância 5%), logo os resíduos são independentes (não existe auto-correlação entre os resíduos). Depois de se ter verificado a independência dos resíduos, vamos verificar a existência de efeitos ARCH que nos indica se a variância é constante. Utilizou-se o teste de Heterocedasticidade cuja hipótese nula é não existirem efeitos ARCH na série (a variância é constante), obtendo-se o seguinte output:

| Heteroskedasticity Test: ARCH | | | |
|-------------------------------|----------|---------------------|--------|
| F-statistic | 102.0017 | Prob. F(1,3075) | 0.0000 |
| Obs*R-squared | 98.79097 | Prob. Chi-Square(1) | 0.0000 |

Quadro 4. Teste de heterocedasticidade aplicado ao modelo MA (1)

Podemos verificar que para o nível de significância 5%, rejeita-se a hipótese nula, concluindo-se que existem efeitos ARCH na série dos resíduos, ou seja, a variância não é constante. Deste modo, temos indícios claros que podemos aplicar o modelo GARCH na nossa série, pois, os resíduos são heterocedásticos.

Uma vez terminada a verificação dos pressupostos dos resíduos, e, apesar de o modelo não ser validado, procedemos com a previsão *in-sample* da série PSI20. A Figura 13 apresenta o output com a previsão *in-sample* para o período temporal que cobre os dados do ano 2016, a partir do modelo estudado.

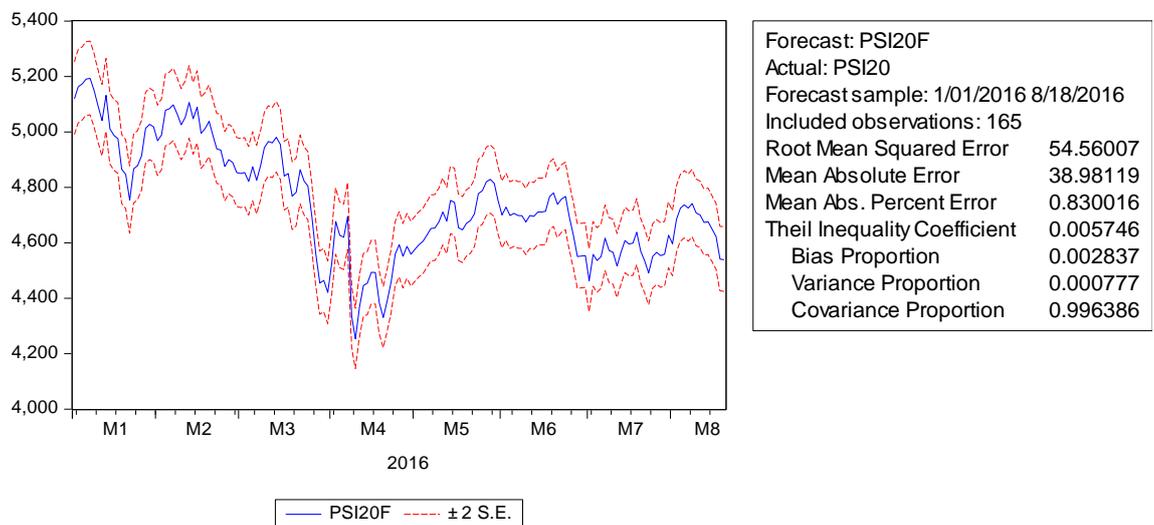


Figura 13. Representação gráfica da previsão da série psi20 do ano 2016

Como a previsão *in-sample* tem bastante qualidade, como se pode observar da Figura 13, procedemos com a previsão *out-of-sample* para a janela temporal de 1 dia (pois o modelo MA (1) só pode prever um valor no futuro). A previsão para o dia 19/08/2016 é 4772.1357.

A estatística U de Theil's é importante na previsão de séries temporais, indicando a qualidade da previsão nos modelos em estudo. Quanto mais próximo de zero for a estatística U de Theil's e o Bias e mais próximo de 1 for a proporção de covariância melhor será a previsão da série. Podemos então verificar que o valor de Theil's é muito próximo de zero, assim como a covariância é próximo de 1. Deste modo, a previsão deste modelo é relativamente assertiva e as linhas correspondentes ao atual e ao predito estão bastante próximas.

Vamos calcular o valor do erro de previsão para o dia 19 de agosto de 2016, isto é,

$$err = |psi20 - psi20f| = |4701,14 - 4772,13| = 70,99$$

sendo $psi20$ o valor original e $psi20f$ o valor predito. O erro de previsão é aproximadamente 7%.

Em baixo apresentamos o gráfico dos erros de previsão para a janela temporal considerada.

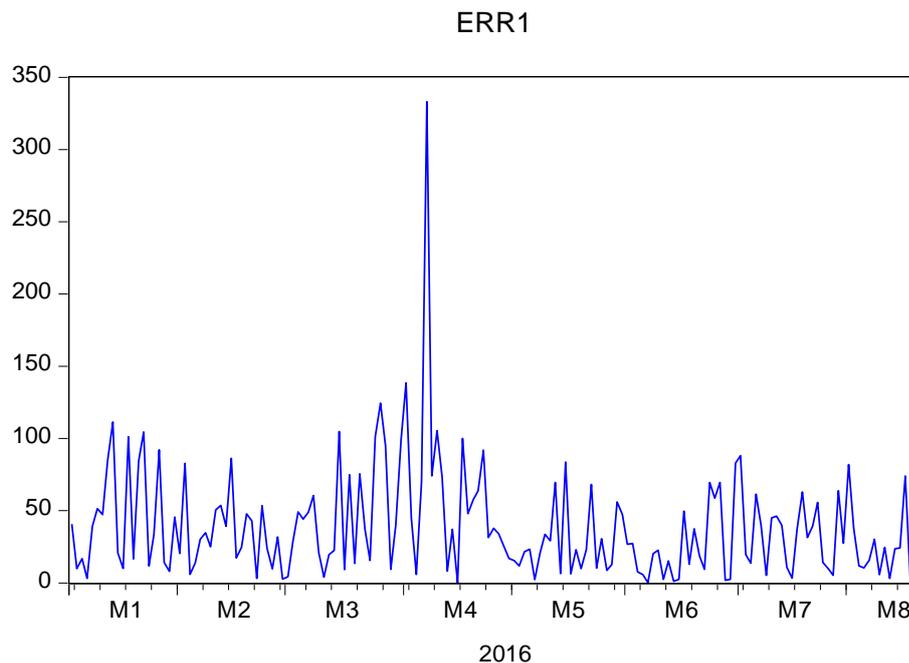


Figura 14. Representação gráfica dos erros do modelo MA (1)

Como os resíduos do modelo MA (1) apresentavam heterocedasticidade, procedemos com o estudo de um modelo GARCH.

4.1.2 ESTIMAÇÃO DO MODELO GARCH

Geralmente um modelo $GARCH(1,1)$ é suficiente para estimar uma série temporal. No quadro seguinte temos o output para o modelo $GARCH(1,1)$ com distribuição normal, onde utilizamos a constante para a equação da média.

| Dependent Variable: D(LOG(PSI20)) | | | | |
|---|-------------|-----------------------|-------------|--------|
| Method: ML - ARCH (Marquardt) - Normal distribution | | | | |
| Date: 09/11/17 Time: 23:06 | | | | |
| Sample (adjusted): 11/02/2004 8/18/2016 | | | | |
| Included observations: 3078 after adjustments | | | | |
| Convergence achieved after 13 iterations | | | | |
| Presample variance: backcast (parameter = 0.7) | | | | |
| GARCH = C(2) + C(3)*RESID(-1)^2 + C(4)*GARCH(-1) | | | | |
| Variable | Coefficient | Std. Error | z-Statistic | Prob. |
| C | 0.000557 | 0.000168 | 3.308759 | 0.0009 |
| Variance Equation | | | | |
| C | 1.28E-06 | 2.23E-07 | 5.732458 | 0.0000 |
| RESID(-1)^2 | 0.129357 | 0.009645 | 13.41135 | 0.0000 |
| GARCH(-1) | 0.873118 | 0.008913 | 97.95460 | 0.0000 |
| R-squared | -0.003201 | Mean dependent var | -0.000169 | |
| Adjusted R-squared | -0.003201 | S.D. dependent var | 0.012833 | |
| S.E. of regression | 0.012854 | Akaike info criterion | -6.214536 | |
| Sum squared resid | 0.508400 | Schwarz criterion | -6.206697 | |
| Log likelihood | 9568.171 | Hannan-Quinn criter. | -6.211720 | |
| Durbin-Watson stat | 1.826354 | | | |

Quadro 5. Estimação do modelo GARCH (1,1)

À semelhança do modelo anterior, para validar o modelo vamos recorrer à análise dos resíduos, verificando a correlação dos resíduos de estimação, normalidade e a existência de elementos ARCH (variância constante). Vamos visualizar o histograma dos resíduos.

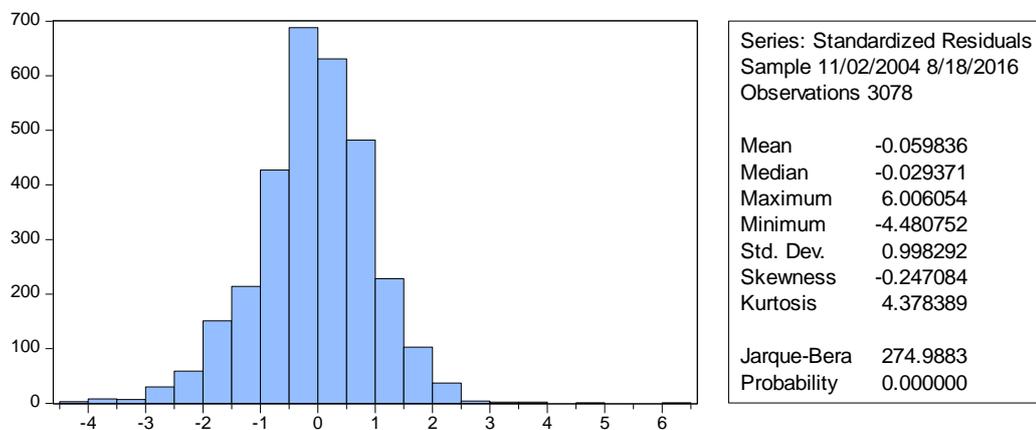


Figura 155. Gráfico dos resíduos do modelo GARCH (1,1).

Analisando o histograma, verificamos que a média é muito próximo do zero, o coeficiente de curtose é superior a 3 (leptocúrtica) e o coeficiente de assimetria é negativa. O correlograma dos resíduos é representado na Figura 16:

| Autocorrelation | Partial Correlation | AC | PAC | Q-Stat | Prob | |
|-----------------|---------------------|----|--------|--------|--------|-------|
| | | 1 | 0.009 | 0.009 | 0.2488 | 0.618 |
| | | 2 | -0.015 | -0.015 | 0.9580 | 0.619 |
| | | 3 | -0.005 | -0.004 | 1.0210 | 0.796 |
| | | 4 | 0.015 | 0.015 | 1.7166 | 0.788 |
| | | 5 | 0.010 | 0.010 | 2.0318 | 0.845 |
| | | 6 | -0.025 | -0.025 | 3.9327 | 0.686 |
| | | 7 | -0.014 | -0.013 | 4.4980 | 0.721 |
| | | 8 | -0.006 | -0.006 | 4.6009 | 0.799 |
| | | 9 | 0.003 | 0.003 | 4.6348 | 0.865 |
| | | 10 | -0.018 | -0.018 | 5.6678 | 0.842 |
| | | 11 | 0.023 | 0.024 | 7.2890 | 0.775 |
| | | 12 | 0.006 | 0.005 | 7.4025 | 0.830 |
| | | 13 | -0.016 | -0.017 | 8.2431 | 0.827 |
| | | 14 | -0.040 | -0.040 | 13.228 | 0.509 |
| | | 15 | -0.002 | -0.002 | 13.240 | 0.584 |
| | | 16 | -0.010 | -0.012 | 13.523 | 0.634 |
| | | 17 | 0.006 | 0.007 | 13.646 | 0.692 |
| | | 18 | -0.016 | -0.014 | 14.423 | 0.701 |
| | | 19 | -0.009 | -0.008 | 14.677 | 0.743 |
| | | 20 | -0.010 | -0.012 | 14.963 | 0.779 |
| | | 21 | 0.008 | 0.008 | 15.175 | 0.814 |
| | | 22 | -0.039 | -0.041 | 19.998 | 0.583 |
| | | 23 | -0.003 | -0.002 | 20.021 | 0.641 |
| | | 24 | 0.023 | 0.021 | 21.721 | 0.596 |
| | | 25 | 0.002 | 0.002 | 21.732 | 0.651 |
| | | 26 | -0.018 | -0.018 | 22.756 | 0.647 |
| | | 27 | -0.017 | -0.017 | 23.707 | 0.647 |
| | | 28 | 0.011 | 0.006 | 24.079 | 0.677 |
| | | 29 | -0.007 | -0.010 | 24.241 | 0.717 |
| | | 30 | -0.016 | -0.016 | 25.041 | 0.723 |
| | | 31 | -0.033 | -0.031 | 28.469 | 0.597 |
| | | 32 | -0.017 | -0.020 | 29.328 | 0.602 |
| | | 33 | 0.051 | 0.050 | 37.360 | 0.276 |
| | | 34 | -0.013 | -0.014 | 37.853 | 0.298 |
| | | 35 | -0.009 | -0.008 | 38.088 | 0.331 |
| | | 36 | -0.001 | -0.005 | 38.090 | 0.375 |

Figura 166. Correlograma dos resíduos aplicado ao modelo GARCH

Analisando o correlograma, podemos verificar que para 36 graus de liberdade o *p-value* obtido é 0.375, não se rejeita a hipótese nula, isto é, os resíduos são não correlacionados. Nele podemos verificar a existência de um padrão definido, evidenciando que não existe informação nos resíduos, apresentando-se como um “ruído branco”.

Vamos verificar a existência de efeitos ARCH para que o modelo seja validado, aplicando o teste de Heterocedasticidade.

| Heteroskedasticity Test: ARCH | | | |
|-------------------------------|----------|---------------------|--------|
| F-statistic | 0.248442 | Prob. F(1,3075) | 0.6182 |
| Obs*R-squared | 0.248583 | Prob. Chi-Square(1) | 0.6181 |

Quadro 6. Teste de Heterocedasticidade aplicado aos resíduos aplicado ao modelo GARCH

Verifica-se que, para os níveis de significância 5%, o *p-value* é bastante elevado, o que nos leva a não rejeitar a hipótese nula, concluindo-se que não existem mais evidências de termos ARCH na série dos resíduos. Concluindo, os resíduos têm média nula, são independentes e tem variância constante. Logo, podemos validar o modelo GARCH (1, 1).

No que segue, estudamos a previsão in-sample para toda a amostra e para a amostra que contém os dados do ano 2016. Como a última teve um melhor desempenho, vai ser utilizada para obter o valor futuro do índice PSI20 (isto é, para o dia 19 de agosto de 2016).

Procedemos com a previsão (*out-of-sample*) para a amostra do ano 2016, para determinar a qualidade do desempenho deste modelo.

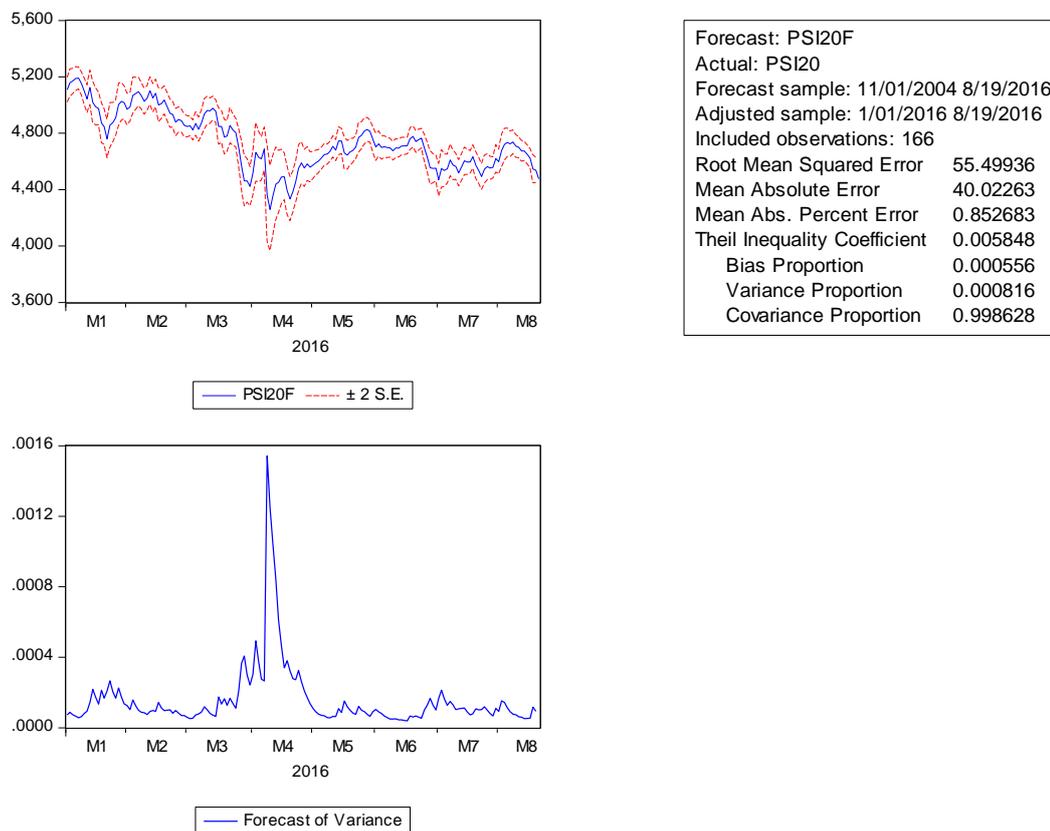


Figura 177. Previsão do ano 2016 utilizando o modelo GARCH (1, 1)

O Valor predito para o dia 19/08/2016 é 4776,256. De igual modo, temos que os valores da estatística de Theil's e do viés são próximos de zero e a covariância é próximo de 1, o que nos indica que a previsão utilizando o modelo *GARCH*(1,1) também é assertiva e as linhas correspondente ao atual e ao predito estão próximos.

À semelhança do modelo anterior, calculamos o erro de previsão para o dia 19 de agosto de 2016, que vem representado em baixo:

$$err = |psi20 - psi20f| = |4701,14 - 4776,25| = 75,11$$

isto é, aproximadamente 7,5%.

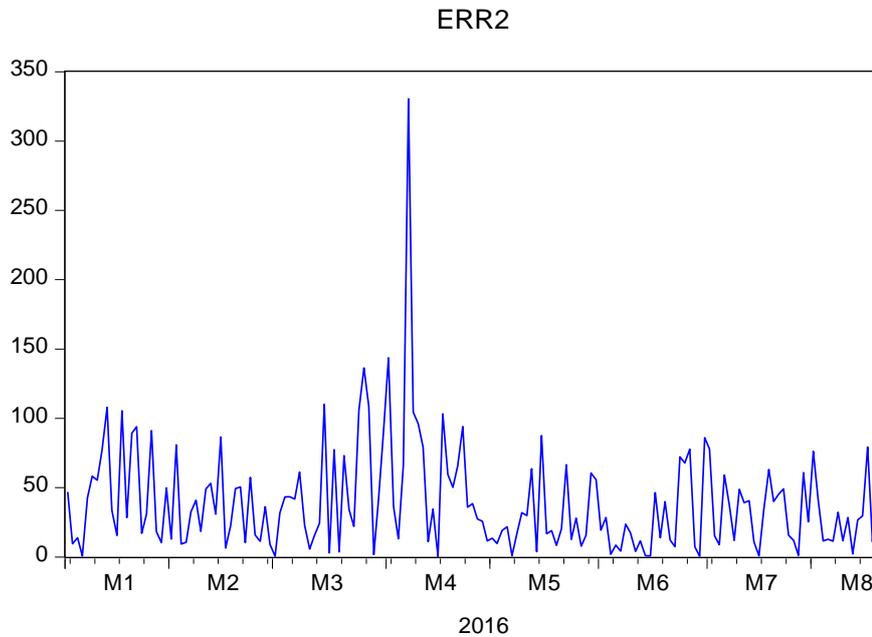


Figura 188. Representação gráfica dos erros do modelo GARCH (1, 1)

Finalmente, a Figura 18 representa os erros de previsão para o ano 2016. Em comparação com os erros do modelo MA (1), a diferença é pouca, observando-se o mesmo padrão e a mesma magnitude nos erros. Concluimos assim, que o desempenho dos dois modelos bastante semelhante.

COMPARAÇÃO DOS MODELOS ARCH e GARCH

Analisando os erros de previsão, as medidas do erro absoluto médio e da raiz do erro quadrático médio, obtidas para cada uma das previsões (Figuras 13 e 17), podemos verificar que estes valores são muito próximos, o que nos permite concluir que ambos os modelos têm um desempenho razoável na previsão. Como vimos anteriormente, o teste de Heterocedasticidade indica que existem evidências de efeitos ARCH na série, ou seja, a variância não é constante, o que permite concluir que o melhor modelo que se

ajusta aos dados é o modelo GARCH. Apesar deste modelo ser o mais apropriado para a série, temos que o modelo ARMA (em particular MA (1)) apresenta melhor resultados em relação aos erros e aos valores preditos.

5. CONCLUSÃO

O objetivo deste trabalho é fazer um estudo comparativo entre modelos lineares univariados e modelos heterocedásticos de modo a verificar qual dos modelos é o mais eficiente quando aplicados numa série não bem-comportada como as séries financeiras. Para ambos os modelos mostrou-se as suas características assim como as suas limitações.

Foi feita a análise dos dados, utilizando para o efeito a cotação de fecho do PSI20, com o objetivo principal de verificar se os modelos são eficientes quando aplicados, e conseqüentemente, verificar qual dos modelos melhor se ajusta e produz melhor resultado na previsão da série PSI20.

Depois de ter sido feita a análise dos dados e dos resultados, conclui-se que ambos os modelos produzem resultados semelhantes na previsão, no entanto os modelos lineares apresentaram maior desempenho. Contudo os modelos heterocedásticos, modelos GARCH, são modelos capazes de detetar e modelar manifestações de heteroscedasticidade condicionada nos resíduos das séries temporais, ou seja, pretendem responder a evidência de que a variância de dados não é constante ao longo do tempo.

BIBLIOGRAFIA

Terence, C. Mills, Raphael N. Markellos (2008), *The Econometric Modelling of Financial Time Series*.

Brockwell, J. Peter, Richard A. Davis (2001), *Introduction to Time Series and Forecasting*.

Hamilton, D. James (1954), *Time Series Analysis*.

Shumway, H. Robert, David S. Stoffer (2010), *Time Series Analysis and Its Applications*.

Svetlozar T. Rachev, Stefan Mittnik, Frank J. Fabozzi, Sergio M. Focardi, Teo Jasic, (2007), *Financial Econometrics - From Basics to Advanced Modeling Techniques*, John Wiley & Sons, Inc.

Mendes, D. (2013). *Econometria dos Mercados Financeiros, Apontamentos das aulas*. Mestrado de Matemática Financeira, ISCTE.

Furriel, A. (2011). *Modelos Heterocedásticos – ARCH e GARCH*. Dissertação de Mestrado em Métodos Quantitativos para a Economia e Gestão, Faculdade de Economia da Universidade do Porto.

Jerónimo, S. (2015), *Assimetria na volatilidade dos principais índices bolsistas de Portugal, Irlanda, Itália, Grécia e Espanha: um estudo comparativo*. Dissertação de Mestrado em Controlo de Gestão e dos Negócios, ISCAL.

Paulino, T. (2016), *Efeito dos dias da semana na volatilidade S&P 500: em estudo empírico*. Dissertação de Mestrado em Contabilidade e Análise Financeira, ISCAL.

Amaral, C. (2015), *Análise de modelos de previsão do value-at-risk aplicados ao principal índice de ações do mercado português*. Dissertação de Mestrado em Contabilidade e Análise Financeira, ISCTE.

Vieira, S. (2013), *A contribuição dos Modelos de Avaliação para a previsão do retorno dos títulos financeiros de empresa do PSI20 – Análise Comparativa entre Capital Asset Pricing Model e Arbitrage Pricing Theory*. Dissertação de Mestrado em Gestão, Instituto Superior de Gestão.

Sites na internet

Valores do fecho, <https://www.euronext.com/pt-pt/products/indices/PTING0200002-XLIS>.

Index rule book PSI20 Index, de <https://www.euronext.com/pt-pt>, consultado em 08/11/2016.

Relatório Anual Sobre a Atividade da CMVM e Sobre os Mercados de Valores Mobiliários (2015), de <http://www.cmvm.pt/> consultado em 08/11/2016.

A bolsa Portuguesa, Factos e Números 2014, de <https://www.euronext.com/pt-pt>, consultado em 12/11/2016.

<https://www.economias.pt/os-principais-indices-bolsistas-do-mundo/>, consultado em 13/07/2017.

ANEXOS

| Dependent Variable: D(LOG(PSI20)) | | | | |
|---|-------------|-----------------------|-------------|-----------|
| Method: Least Squares | | | | |
| Date: 09/11/17 Time: 06:07 | | | | |
| Sample (adjusted): 11/03/2004 8/18/2016 | | | | |
| Included observations: 3077 after adjustments | | | | |
| Convergence achieved after 3 iterations | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | -0.000171 | 0.000252 | -0.678562 | 0.4975 |
| AR(1) | 0.083732 | 0.017973 | 4.658743 | 0.0000 |
| R-squared | 0.007009 | Mean dependent var | | -0.000170 |
| Adjusted R-squared | 0.006686 | S.D. dependent var | | 0.012835 |
| S.E. of regression | 0.012792 | Akaike info criterion | | -5.879287 |
| Sum squared resid | 0.503207 | Schwarz criterion | | -5.875367 |
| Log likelihood | 9047.283 | Hannan-Quinn criter. | | -5.877879 |
| F-statistic | 21.70389 | Durbin-Watson stat | | 1.998174 |
| Prob(F-statistic) | 0.000003 | | | |
| Inverted AR Roots | .08 | | | |

Quadro 7. Estimação do modelo AR (1)

| Dependent Variable: D(LOG(PSI20)) | | | | |
|---|-------------|-----------------------|-------------|-----------|
| Method: Least Squares | | | | |
| Date: 09/11/17 Time: 05:11 | | | | |
| Sample (adjusted): 11/03/2004 8/18/2016 | | | | |
| Included observations: 3077 after adjustments | | | | |
| Convergence achieved after 7 iterations | | | | |
| MA Backcast: 11/02/2004 | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | -0.000171 | 0.000250 | -0.682898 | 0.4947 |
| AR(1) | -0.009730 | 0.212855 | -0.045710 | 0.9635 |
| MA(1) | 0.094135 | 0.211925 | 0.444192 | 0.6569 |
| R-squared | 0.007076 | Mean dependent var | | -0.000170 |
| Adjusted R-squared | 0.006430 | S.D. dependent var | | 0.012835 |
| S.E. of regression | 0.012794 | Akaike info criterion | | -5.878705 |
| Sum squared resid | 0.503173 | Schwarz criterion | | -5.872824 |
| Log likelihood | 9047.387 | Hannan-Quinn criter. | | -5.876592 |
| F-statistic | 10.95287 | Durbin-Watson stat | | 1.999591 |
| Prob(F-statistic) | 0.000018 | | | |
| Inverted AR Roots | -.01 | | | |
| Inverted MA Roots | -.09 | | | |

Quadro 8. Estimação do modelo ARMA (1, 1)

| Dependent Variable: D(LOG(PSI20)) | | | | |
|---|-------------|-----------------------|-------------|--------|
| Method: Least Squares | | | | |
| Date: 09/11/17 Time: 05:12 | | | | |
| Sample (adjusted): 11/03/2004 8/18/2016 | | | | |
| Included observations: 3077 after adjustments | | | | |
| Convergence achieved after 3 iterations | | | | |
| MA Backcast: 11/01/2004 11/02/2004 | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | -0.000171 | 0.000250 | -0.683457 | 0.4944 |
| AR(1) | 0.084357 | 0.018039 | 4.676511 | 0.0000 |
| MA(2) | -0.008258 | 0.018088 | -0.456582 | 0.6480 |
| R-squared | 0.007072 | Mean dependent var | -0.000170 | |
| Adjusted R-squared | 0.006426 | S.D. dependent var | 0.012835 | |
| S.E. of regression | 0.012794 | Akaike info criterion | -5.878701 | |
| Sum squared resid | 0.503175 | Schwarz criterion | -5.872821 | |
| Log likelihood | 9047.382 | Hannan-Quinn criter. | -5.876589 | |
| F-statistic | 10.94764 | Durbin-Watson stat | 1.999495 | |
| Prob(F-statistic) | 0.000018 | | | |
| Inverted AR Roots | .08 | | | |
| Inverted MA Roots | .09 | -.09 | | |

Quadro 9. Estimação do modelo ARMA (1, 2)

| Dependent Variable: D(LOG(PSI20)) | | | | |
|---|-------------|-----------------------|-------------|--------|
| Method: Least Squares | | | | |
| Date: 09/11/17 Time: 05:12 | | | | |
| Sample (adjusted): 11/04/2004 8/18/2016 | | | | |
| Included observations: 3076 after adjustments | | | | |
| Convergence achieved after 4 iterations | | | | |
| MA Backcast: 11/03/2004 | | | | |
| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
| C | -0.000170 | 0.000250 | -0.679656 | 0.4968 |
| AR(2) | -0.000900 | 0.018106 | -0.049688 | 0.9604 |
| MA(1) | 0.084394 | 0.018042 | 4.677634 | 0.0000 |
| R-squared | 0.007075 | Mean dependent var | -0.000169 | |
| Adjusted R-squared | 0.006428 | S.D. dependent var | 0.012837 | |
| S.E. of regression | 0.012796 | Akaike info criterion | -5.878389 | |
| Sum squared resid | 0.503168 | Schwarz criterion | -5.872507 | |
| Log likelihood | 9043.963 | Hannan-Quinn criter. | -5.876276 | |
| F-statistic | 10.94753 | Durbin-Watson stat | 1.999590 | |
| Prob(F-statistic) | 0.000018 | | | |
| Inverted MA Roots | -.08 | | | |

Quadro 10. Estimação do modelo ARMA (2, 1)

Previsão de séries temporais financeiras: o caso PSI 20

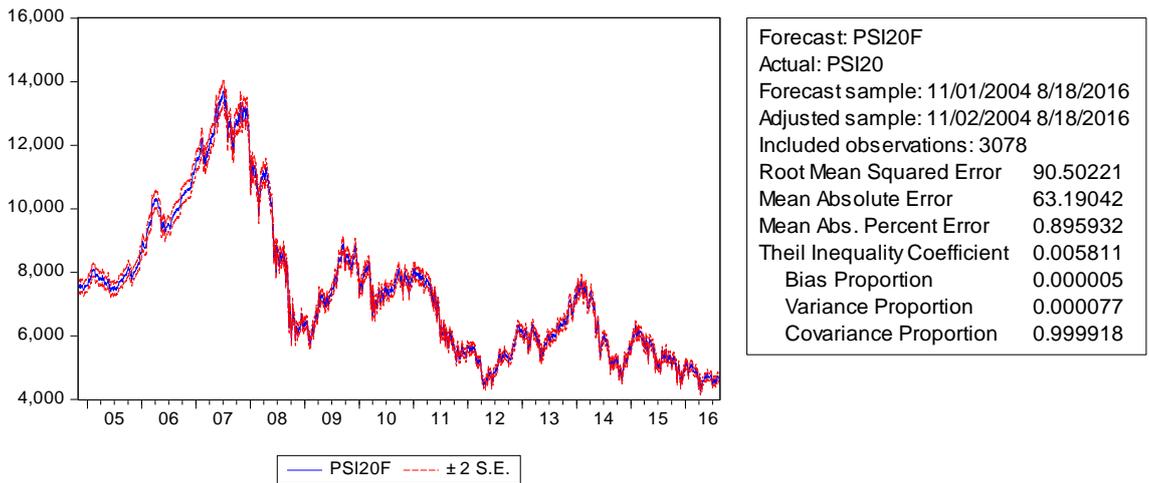


Figura 19. Representação gráfica da previsão da série psi20 aplicando o modelo MA (1)

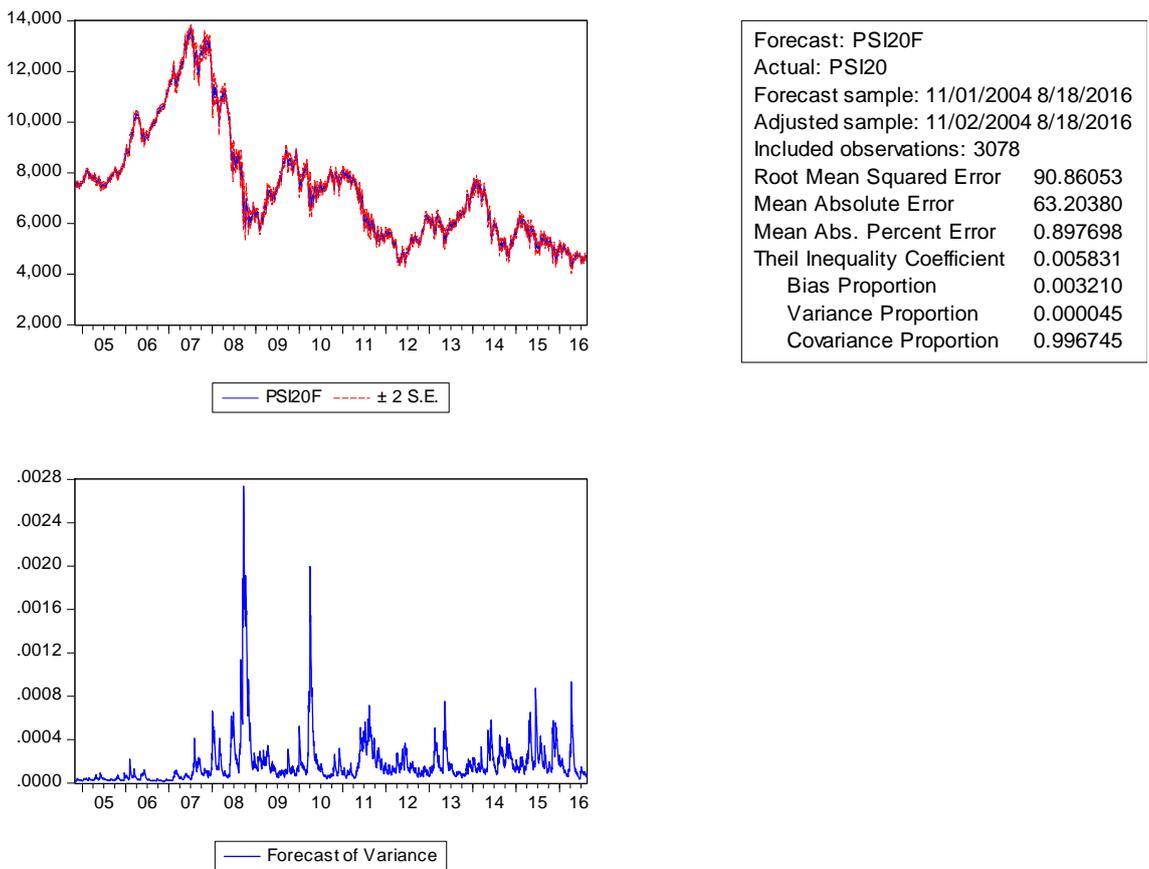


Figura 20. Representação gráfica da previsão da série psi20 aplicando o modelo GARCH (1, 1).