

UNIVERSIDADE ABERTA



**ANÁLISE E PREVISÃO DE SÉRIES TEMPORAIS COM MODELOS ARIMA E
ANÁLISE ESPECTRAL SINGULAR**

Jorge Manuel Nunes Xavier

Mestrado em Bioestatística e Biometria

Dissertação orientada Prof^ª Doutora Teresa Costa Azinheira Oliveira

Co-orientador: Professor Doutor Paulo Jorge Canas Rodrigues

2016

Resumo

Na nossa sociedade cada vez é mais importante a criação de modelos que nos ajudem a compreender os fenómenos que nos rodeiam, assim como fazer os mais diversos tipos de previsões.

Dos diversos tipos de dados disponíveis e recolhidos para posterior análise, serão neste trabalho tratados aqueles que são recolhidos em intervalos de tempo iguais (horas, dias, semanas, etc), ao longo de um período mais ou menos longo, ou seja as Séries Temporais.

Escolher o modelo mais fiável, o que dá as previsões mais fidedignas, sempre foi um propósito e uma preocupação para a Estatística ao longo dos tempos.

Neste trabalho, são focados em particular os modelos do tipo ARIMA e a Análise Espectral Singular, como ferramentas de trabalho na análise e previsão de séries temporais.

Pretendeu-se contribuir um pouco para melhorar a abordagem da análise e previsão das séries temporais, ilustrando com exemplos e recurso ao software R e ao *Gretl*.

Palavras-chave: Séries Temporais, modelos ARIMA, Análise Espectral Singular, software R e Gretl

Abstract

Abstract

In our society it is increasingly important to create models that help us understand the phenomena that surround us as well as make the most varied types of forecasts.

Of the various types of data available and collected for further analysis, in this study we process data collected at equal time intervals (hours, days, weeks, etc.) over a shorter or longer period, that is to say, time series.

Choosing the most reliable model, which provides the most reliable forecasts, was always a purpose and a concern for statistics over time.

In this work, we focus on particular models of ARIMA type and Spectral Simple Analysis as work tools in the analysis and forecasting of time series.

The aim is to compare these two techniques, contributing a little to improve the approach to the analysis and forecasting of time series.

Key Words: Time Series, ARIMA models, Singular Spectrum Analysis, software R and Gretl

Dedicatória

Para a minha mãe pelo seu amor e apoio constante.

À Fátima pela companhia nesta viagem.

Ao Luís e ao Nuno.

Agradecimentos

Os meus primeiros agradecimentos vão para a minha mãe, a mulher da minha vida, que ao longo de todo este tempo esteve e estará incondicionalmente ao meu lado.

Um agradecimento muito profundo à Fátima por me ter acompanhado desde o início nesta caminhada, que não acabará aqui.

Para o Luís e o Nuno, os meus filhos do coração, os meus agradecimentos pelo seu apoio, pela sua paciência nos momentos mais difíceis e nas minhas ausências, esperando que tenha sido um bom exemplo para eles.

A todos os restantes familiares e amigos, o meu mais sincero obrigado, pelo apoio, força e carinho me deram.

Por fim, os mais sinceros agradecimentos aos orientadores, Professora Doutora Teresa Costa Azinheira Oliveira e Professor Doutor Paulo Jorge Canas Rodrigues, pela sua paciência e toda a força que me deram nos momentos mais difíceis e na disponibilidade para resolver todos os meus problemas.

Não queria deixar de agradecer a toda a comunidade Universidade Aberta e em especial todos os colegas que trabalharam comigo durante esta etapa da minha vida, premindo-me chegar até aqui.

Índice

1. Introdução	1
2. Séries Temporais	2
2.1. História.....	2
2.2 Objectivo das Séries Temporais	3
2.3 Tipo de Séries Temporais.....	5
2.3.1 Séries Temporais estocásticas.	5
2.3.2 Processo Estocásticos não estacionários	7
2.3.2 Séries temporais - Contínuas e discretas, univariadas e multivariadas.	8
2.4 Tendência e sazonalidade das séries temporais.....	9
2.5 Métodos matemáticos de previsão de séries temporais.....	11
2.5.1 Métodos simples de previsão de séries.....	12
2.5.1.1 Modelos de Média Móvel.....	13
2.5.1.2 - Alisamento exponencial.....	14
2.5.1.2.1 - Modelos de alisamento exponencial simples.	14
2.5.1.2.2 - Modelo de Alisamento Exponencial Linear.....	14
2.5.1.2.3 – Método de Alisamento Exponencial Holt-Winter com sazonalidade.	15
2.6 Análise de séries temporais	15
2.6.1 Teste de normalidade	15
2.6.2 Teste da raiz unitária ou teste de estacionaridade.	16
3. Modelos ARIMA	21
3.1 Modelos autoregressivos (AR).....	21
3.2 Modelos de média móvel (MA).	22
3.3 Modelos autorregressivos e de médias móveis (ARMA).....	22
3.4 Modelos autorregressivos integrados e de médias móveis (ARIMA).....	23
3.5 Identificação de modelos ARIMA.	24
3.6 Estimação de modelos ARIMA.....	27
3.7 Diagnóstico de modelos ARIMA	27
3.7.1 Testes de Box-Pierce e Ljung-Box.....	28
3.7.2 Critério de escolha dos modelos.....	29
3.8 Previsão com modelos ARIMA.	30
3.9. Modelos SARIMA	31
4. - Análise Espectral Singular (AES).....	33

4.1 - Primeira fase – Decomposição.....	33
4.1.1 Embutimento (<i>Embedding</i>)	33
4.1.2. - Decomposição em Valores Singulares.....	34
4.2 - Segunda fase – Reconstrução.....	35
4.2.1 Agrupamento.....	35
4.2.2 - Diagonalização por Médias.....	36
5. Análise dos dados.....	38
5.1 Análise descritiva das séries temporais	39
5.1.1 Testes de normalidade	40
5.1.2 Testes de raiz unitária ou testes de estacionaridade.	41
5.2 Modelos ARIMA	41
5.2.1 Testes de Ljung-Box	41
5.2.2 Critério de escolha dos modelos.....	42
5.3 Análise Espectral Singular	44
5.4 Previsões	49
5.5 Comparação dos modelos.....	54
6. Considerações finais.....	56
7. Bibliografia	58
Anexo A - Resultados dos testes de estacionaridade	63
Anexo B - Resultados dos modelos ARIMA	67
Anexo C - Resultados da análise dos critérios de informação AIC, BIC e HQC.....	78
Anexo D – Raiz do Erro Quadrado Médio (RMSE).	82
Anexo E – Programação utilizada para o estudo dos modelos AES (scripts).	85
Anexo F – Dados.....	90

Índice de tabelas.

Tabela 5.1. Resultados dos testes de normalidade – Milk.....	40
Tabela 5.2. Resultados dos testes de normalidade – Pigs.....	40
Tabela 5.3. Resultados teste Ljung-Box para Milk.....	41
Tabela 5.4. Resultados do teste Ljung-Box para Pigs.....	42
Tabela 5.5 Resultados critério AIC – Milk	43
Tabela 5.6. Resultados critério AIC – Pigs.....	43
Tabela 5.7. Resultados critério BIC – Milk.....	43
Tabela 5.8. Resultados critério BIC - Pigs.....	43
Tabela 5.9. Resultados critério HQC – Milk.....	43
Tabela 5.10. Resultados critério HQC – Pigs.....	43
Tabela 5.11. – Milk – Previsão e intervalos de confiança (ARIMA).....	50
Tabela 5.12. – Pigs – Previsão e intervalos de confiança (ARIMA).....	50
Tabela 5.13. – Milk – Previsão e intervalos de confiança (AES).....	51
Tabela 5.14. – Pigs – Previsão e intervalos de confiança (AES).....	51
Tabela 5.15. Milk – Raiz do erro quadrado médio.....	54
Tabela 5.16. Pigs – Raiz do erro quadrado médio.....	54

Índice de figuras.

Figura 5.1. Gráfico dos dados da série temporal - Milk.....	39
Figura 5.2. Gráfico dos dados da série temporal - Pigs.	39
Figura 5.3. Milk Matriz correlação.....	45
Figura 5.4. Pigs Matriz correlação.	45
Figura 5.5. Milk - Decomposição em valores singulares.	45
Figura 5.6. Pigs – Decomposição em valores singulares.....	46
Figura 5.7. Reconstrução da série temporal Milk.....	46
Figura 5.8. Reconstrução da série temporal Pigs.....	47
Figura 5.9. Reconstrução com sazonalidade da série temporal - Milk.....	47
Figura 5.10. - Reconstrução com sazonalidade da série temporal - Pigs.	48
Figura 5.11 – Correlação ponderada para os agrupamentos finais.....	48
Figura 5.12. – Gráfico da previsão e intervalos de confiança modelo ARIMA (Milk)...	52
Figura 5.13. – Gráfico da previsão e intervalos de confiança modelo AES (Milk).....	52
Figura 5.14. – Gráfico da previsão e intervalos de confiança modelo ARIMA (Pigs)...	53
Figura 5.15. – Gráfico da previsão e intervalos de confiança modelo AES (Pigs).	53

Lista de abreviaturas, siglas e acrónimos

AR – *Auto-Regressive*.

AES - Análise Espectral Singular

ARIMA – *Auto-Regressive Integrated Moving Average*.

ARMA - *Auto-Regressive Moving Average* .

f.a.c - função autocorrelação

f.a.c.p.- função autocorrelação parcial

MA – *Moving Average*

SARIMA - *Seasonal Auto-Regressive Integrated Moving Average*

ST – Série Temporal

Var – Variância

SSA – Singular Spectrum Analysis Singular

Introdução

1. Introdução

Previsão ; Pre.vi.são; Sf (lat praevisione)

Acto ou efeito de prever; conjuntura

Léxico: Dicionário de português online

Ao longo da história da Humanidade, a necessidade de compreender os fenómenos e de os prever, sempre foram uma preocupação.

Para tomada de decisões na agricultura (plantação, aplicação de fertilizantes, etc), meteorologia (previsão do tempo)

Compreender e prever fenómenos, em que a informação (dados) é recolhida em intervalos de tempo iguais (horas, dias, semanas, etc), ao longo de um período mais ou menos longo é uma faceta particular da estatística, a análise de séries temporais.

No capítulo 2 será efectuada uma descrição breve das séries temporais.

Para a análise de séries temporais e posterior previsão de resultados, vários métodos e modelos podem ser utilizados de acordo com o seu tipo. Este trabalho centraliza-se sobre os chamados modelos ARIMA (capítulo 3) e um método mais recente a análise espectral singular (capítulo 4).

Posteriormente, no capítulo 5 será feita uma análise de dois conjuntos de dados, número de suínos abatidos mensalmente em Victoria (Pigs) e produção mensal de leite (Milk), que correspondem a duas séries temporais de características completamente diferentes, de modo a ter uma visão mais ampla das diferenças na aplicação das duas metodologias.

Para a análise e tratamentos dos dados foram utilizados os softwares Gretl e R de modo a verificar quais os que mais se adequavam a cada um dos diferentes métodos.

Pretendeu-se ao analisar conjuntamente estas duas metodologias, contribuir para um melhor esclarecimento relativamente à abordagem na análise e previsão das séries temporais.

Será efectuada uma análise à literatura especializada, de modo a permitir aferir uma análise conjunta dos métodos ARIMA e da análise espectral singular.

2. Séries Temporais

2.1. História

Na sua generalidade, os métodos de previsão analisam valores passados para prever valores futuros. Até ao início dos anos 20 do século passado, as previsões eram realizadas com a extrapolação simples de um valor global, ajustado em função do tempo.

Uma importante contribuição para o desenvolvimento de ferramentas e métodos para a análise de séries temporais deve-se ao trabalho de Yule (1927), em que o autor refere que uma série temporal deveria ser vista simplesmente como a realização de um processo estocástico. O autor criou o modelo auto-regressivo (AR), em que o valor previsto dependia dos valores passados.

Nos anos seguintes os peritos da área da estatística, foram realizando os seus trabalhos e estudos, com base em que o comportamento de uma série dependia dos modelos lineares e do ruído.

Contudo, os modelos lineares mostram-se insuficientes para a análise de séries temporais, em virtude de a maioria das séries reais apresentarem fortes tendências de não linearidade (CHAVES, 1991).

Para combater estas dificuldades, a partir de 1950, iniciaram-se novos estudos na área da previsão de séries temporais, tendo sido criada a técnica de alisamento exponencial. Foi também nesta época, com o aparecimento do computador que se iniciou uma evolução acentuada nos métodos de previsão.

Um trabalho que é considerado uma referência na história da previsão com séries temporais é o livro de Box and Jenkins (1976). Este livro integrou todo o conhecimento existente até à data relativamente às séries temporais, no que respeita à estimação de parâmetros, identificação e previsão. Estes autores abordaram a construção de modelos estocásticos no domínio do tempo para séries temporais discretas, que eram o mais simples possível, com número mínimo de parâmetros e satisfazendo portanto um critério de parcimónia. Os autores consideraram que a escolha deste critério se devia ao

facto de estes modelos serem capazes de revelar algo sobre a natureza dos processos que geram as séries temporais e à possibilidade de esses modelos serem utilizados para a obtenção de previsões óptimas de valores futuros das séries.

Recentemente Gooijer and Hyndman (2006), apresentaram uma análise dos trabalhos mais importantes realizados na área de previsões de séries temporais nos últimos 25 anos em todo o mundo.

2.2 Objectivo das Séries Temporais

Segundo Granger (1977) o propósito inicial da análise de séries temporais é tirar inferências sobre as propriedades ou características básicas do mecanismo gerador do processo estocástico das observações da série.

Ao analisar uma série temporal, espera-se que nesta exista uma causa relacionada com o tempo, que influenciou os dados e que possam continuar a influenciá-los futuramente.

Para Barbancho (1970) é através da abstracção de regularidades contidas nos fenómenos observáveis de uma série temporal que existe a possibilidade de se construir um modelo matemático como representação simplificada da realidade.

Os objectivos de se analisar uma série temporal são segundo MORETTIN e TOLOI (1981) e CHAFTFIELD (2000), os seguintes:

1. Descrição: Determinar as propriedades de uma série como, por exemplo, o padrão de tendência, a existência de alterações estruturais e sazonalidades, etc.
Antes de se tentar modelar e prever uma determinada série de tempo, é desejável ter uma ideia preliminar dos dados, para obter descrever algumas das suas principais propriedades. Isto será uma ajuda muito importante na realização do processo de modelação.
2. Explicação/Modelação: Encontrar um modelo estatístico adequado que permita explicar o comportamento da série no período observado.

3. Previsão: Estimar valores futuros de uma Série Temporal, com base em valores passados.
4. Controlo de processos: Boas previsões permitem ao estatístico, tomar medidas de forma a controlar um determinado processo.

As séries temporais e modelos dinâmicos, podem ser usadas em quatro áreas de aplicação distintas Box et al (1994):

1. A previsão de valores futuros de uma série temporal dados valores anteriores e correntes.
2. A determinação da função de transferência (*transfer function*) de um sistema sujeito a inércia – a determinação de um modelo entrada-saída (*input-output*) que permite mostrar no resultado de um sistema o efeito de qualquer série de valores da entrada dada.
3. O uso de variáveis de entrada, indicadoras nos modelos de funções de transferência para a representação e avaliação dos efeitos de eventos de intervenção invulgares no comportamento da série temporal.
4. O desenho de esquemas de controlo por meio dos quais, os potenciais desvios dos valores de saída do sistema obtidos a partir de um objectivo pretendido poderão, na medida do possível, ser compensados pelo ajustamento dos valores de entrada da série.

Em biologia, uma parte que pode ser isolada para se realizar um estudo sobre ela, em que se conhecem as influências internas e externas actuantes, dá-se o nome de *sistema*. Os sistemas que evoluem no tempo são especialmente classificados como *sistemas dinâmicos* (Torres, 2001). Um sistema dinâmico, de uma forma geral, pode estar sob a acção de p sinais de entrada e revelar o seu comportamento por meio de q sinais de saída, em que a variável independente t representa o tempo. Assim sendo estes sinais de entrada e saída de um sistema dinâmico poderão ser vistos como séries temporais, ou seja, valores de uma determinada grandeza em função do tempo (TONG, 1996)

2.3 Tipo de Séries Temporais

2.3.1 Séries Temporais estocásticas.

Para CHAFTFIELD (2000) a maioria das séries temporais são estocásticas, ou aleatórias, em que o futuro é apenas parcialmente determinado por valores passados, sendo o modelo para estas séries muitas vezes chamado de um processo estocástico.

Um processo estocástico pode ser descrito como uma família de variáveis aleatórias $\{x_1, x_2, \dots, x_t, t \in T\}$ indexadas no tempo, onde T é um conjunto ordenado de índices. Esta notação é usada quando as observações não são igualmente espaçadas no tempo

MORETTIN e TOLOI (1981) afirmam que uma série temporal estocástica é estacionária se a sua média, variância e autocovariância (em diferentes fases) permanecem as mesmas. Estas características verificam-se em qualquer período em que se faça a medição ou seja, não variam com o tempo.

As propriedades estatísticas de um conjunto finito de dados da série $\{x_1, x_2, \dots, x_n\} \in X_t$ serão sempre semelhantes às de um conjunto $\{x_{1+h}, x_{2+h}, \dots, x_{n+h}\} \in X_t$ para qualquer valor inteiro de h .

Sentido estrito - quando a distribuição conjunta de qualquer conjunto de variáveis, não se modifica quando transferimos as variáveis no tempo.

$$F(x_i, x_j, \dots, x_k) = F(x_{i+h}, x_{j+h}, \dots, x_{k+h})$$

Sentido débil – quando o primeiro e segundo dados da série são constantes e a Covariância entre duas variáveis depende somente da sua separação temporal.

$$E(X_t) = \mu$$

$$Var(X_t) = \sigma^2$$

$$\gamma(t, t - k) = \gamma(t, t + k) = \gamma(k) \text{ com } k = 0, \pm 1, \pm 2, \dots$$

Na literatura encontra-se também uma forma alternativa de descrever um processo estocástico estacionário, por meio de sua função de densidade espectral ou espectro, que é discreta e em que a transformada de Fourier de $\{\gamma_k\}$ será:

$$f(w) = \frac{1}{\pi}(\gamma_0 + 2 \sum_{k=1}^{\infty} \gamma_k \cos wk), \text{ para } 0 \leq w \leq \pi$$

A função descreve como a variância global do processo é distribuída ao longo de diferentes frequências de zero até à frequência *Nyquist*, π .

No trabalho de CHAFTFIELD (2000) o autor refere ainda que as duas funções $\{\gamma_k\}$ e $f(\omega)$ são equivalentes e complementares. Uma análise baseada principalmente em estimativas de uma função de autocorrelação (ou autocovariância) é chamada de análise no domínio do tempo. Uma análise que se baseia principalmente no espectro é chamada de análise no domínio da frequência ou análise espectral. Por vezes é necessário fazer a análise baseada em ambas as funções, mas em séries temporais de previsão é bastante raro usar métodos no domínio da frequência.

Alguns exemplos de processos estocásticos estacionários:

- Processo Ruído Branco. ARIMA (0,0,0)
- $X_t = a_t$
- Processo Homogéneo de ordem 1. ARIMA (1,0,0)
- $\nabla X_t = a_t$
- Processo Autoregressivo de ordem (p). ARIMA (p,0,0)
- $X_t = \phi_1 x_t + \dots + \phi_p x_{t-p} + a_t$
- Processo Média Móvel de ordem (q). ARIMA (0,0,q)
- $X_t = a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q}$
- Processo ARIMA (p,0,q)
- $X_t = \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q}$

2.3.2 Processo Estocásticos não estacionários

Uma série temporal estocástica é não-estacionária quando a sua média varia com o tempo e/ou é a variância que varia com o tempo. Nestas séries só é possível determinar uma parte dos dados, pelo que geram piores previsões que as séries estacionárias.

As propriedades estatísticas de um conjunto finito de dados da série $\{x_1, x_2, \dots, x_n\} \in X_t$ serão diferentes das de um conjunto $\{x_{1+h}, x_{2+h}, \dots, x_{n+h}\} \in X_t$ para pelo menos um valor inteiro de h . A série pode ser considerada:

- Série temporal não estacionária em média - quando o nível da série não é estável no tempo, podendo ter tendência crescente ou decrescente:

$$E_{(x_t)} \neq \mu$$

- Série temporal não estacionária nas covariâncias – quando a variância ou as covariâncias variam com o tempo:

$$Var(x_t) \neq \sigma^2$$

$$\gamma(t, t - k) \neq \gamma(t, t + k)$$

- Passamos a citar alguns dos exemplos de processos estocásticos não estacionários:
- Passeio Aleatório, ARIMA (0,1,0):

$$(1 - B)x_t = a_t$$

- Processo Alisamento Exponencial Simples:

$$(1 - B)x_t = (1 - \theta_1 B)a_t$$

- Processos Integrados ARIMA (p,d,q):

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - B)^d x_t = (1 - \theta_1 B - \dots - \theta_q B^q)a_t$$

- Processos de Memória Longa:

$$(1 - B)^d x_t = a_t \quad -0.5 < d < 0.5$$

- Processo ARFIMA (p,d,q):

$$\theta_p(B)\nabla^d z_t = \phi_q(B)a_t \quad -0.5 < d < 0.5$$

2.3.2 Séries temporais - Contínuas e discretas, univariadas e multivariadas.

Sendo uma série temporal um conjunto de observações, medidas sequencialmente através do tempo, estas podem ser referentes a séries temporais de tempo contínuo ou discreto. Serão séries temporais de tempo contínuo se os valores forem recolhidos de uma forma contínua ao longo do tempo, caso contrário serão de tempo discreto.

Numa série temporal de tempo contínuo, a variável observada é tipicamente um variável contínua, em que os valores são registrados continuamente, como por exemplo os dados da actividade cerebral registrada por uma máquina de EEG ou o registo da maré de um rio durante um ano.

$$T = \{t: t_1 < t < t_2\}$$

$$T = [0, 24]$$

Um exemplo de uma série temporal discreta no tempo será o registo da produção de leite ente entre 2012 e 2013:

$$T = \{t_1, t_2, \dots, t_n\}$$

$$T = \{01/2012, \dots, 12/2013\}$$

O método usual de análise de uma série temporal contínua no tempo consiste em amostrar os dados em intervalos de tempo iguais, de modo a convertê-la numa série temporal de tempo discreto. Segundo CHAFTFIEL (2000) pouca ou nenhuma informação é perdida por este processo, desde que o intervalo de agrupamento seja suficientemente pequeno. Mais, este autor refere que as séries de temporais discretas podem surgir em três formas distintas, a saber:

1. Sendo amostrados a partir de uma série contínua (por exemplo, temperatura medida em intervalos de tempo de uma hora). Tais dados podem surgir tanto por amostragem de uma série temporal contínua, ou porque as medições são tomadas apenas uma vez uma hora);
2. Ao serem agregadas ao longo de um período de tempo (por exemplo, as vendas totais em sucessivas meses);

3. Como uma série inerentemente discreta (por exemplo, o dividendo pago por uma empresa em anos sucessivos).

As séries temporais podem também ainda ser classificadas como univariadas se as observações são únicas e registadas sequencialmente durante intervalos de tempo iguais (dias, mês, anos, etc.) ou multivariadas se as observações são múltiplas. Estas observações serão do tipo $Y_1(t), \dots, Y_k(t), t \in T$, como por exemplo as vendas semanais de uma determinada firma $Y_1(t)$ e valor investido em publicidade $Y_2(t)$

2.4 Tendência e sazonalidade das séries temporais

As duas principais fontes de variação na maioria das séries temporais são a tendência e a variação sazonal, existindo contudo algumas variações cíclicas e flutuações irregulares que não se devem desprezar, de acordo com os especialistas desta área. Um comportamento muito importante é a tendência, que poderá ser definida como um comportamento de longo prazo da série temporal. Este tipo de variação está presente quando uma série exhibe constantes crescimentos ou declínios, em sucessivos períodos de tempo.

Se considerarmos que a série temporal não apresenta componente sazonal podemos considerar o modelo:

$$Z_t = T_t + a_t$$

sendo a_t uma variável aleatória com média zero, variância σ_a^2 e T_t o estimador da tendência.

Depois de estimada a tendência \hat{T}_t podemos considerar uma série ajustada:

$$Y_t = Z_t - \hat{T}_t$$

Um outro processo que se utiliza para retirar a tendência consiste em fazer sucessivas diferenças da série original até encontrar uma série estacionária.

$$\Delta Z_t = Z_t - Z_{t-1}$$

Um dos problemas na verificação da tendência de uma série temporal, prende-se com a necessidade de saber se os dados são independentes, dado que a maioria dos testes se baseia nessa hipótese. Os testes mais utilizados para detectar tendências são os testes de Wald-Wolfowitz, o teste de Cox-Stuart, o teste de Mann-Kendall e o teste de Spearman.

A percepção da tendência depende, em parte, da duração da observação da série temporal. Além disso, a decomposição da variação sazonal e das componentes de tendência poderão não ser independentes.

A variação sazonal é um tipo de variação em que padrões semelhantes de comportamento se repetem em períodos de tempo iguais, geralmente durante o ano, mas também podem ser períodos semanais, mensais ou trimestrais. São exemplos desta variação sazonal o padrão de vendas para gelados que é sempre mais elevada no verão, os ciclos de acasalamento e reprodução dos animais, as variações de temperatura numa determinada região registadas anual ou diariamente. Para se reconhecer variações anuais, a série temporal deverá ter no mínimo quatro anos completos (16 trimestres ou 48 meses por exemplo).

A sazonalidade pode ser dividida em dois tipos, determinista quando o padrão sazonal é regular e estável no tempo ou estocástica quando o padrão sazonal varia com o tempo.

Os testes mais utilizados para estimar a sazonalidade são o teste de Kruskal-Wallis (para várias amostras independentes), teste de Friedman (indicado para várias amostras dependentes/ emparelhadas) e o teste F, tendo sempre a preocupação de retirar, caso exista, a tendência da série.

Segundo CHAFTFIELD (2000) os métodos clássicos funcionam bastante bem quando a variação é dominado pela uma tendência linear regular e/ou sazonalidade regular. No entanto, eles não funcionam muito bem quando a tendência e/ou efeitos sazonais vão mudando ao longo do tempo ou quando os valores sucessivos das flutuações irregulares são correlacionados.

Se a série temporal tiver mais do que um componente não aleatório, MORETTIN E TOLOI (2006) propõem que se teste a existência de um deles (tendência ou sazonalidade) depois da eliminação do outro componente.

2.5 Métodos matemáticos de previsão de séries temporais.

Um método de previsão tem associado um conjunto de procedimentos que com base num histórico de dados permite prever qual será o comportamento esperado dos dados, a médio ou longo prazo. Para BARBACHO (1970) uma previsão reflete uma manifestação em relação aos sucessos desconhecidos num futuro determinado.

A previsão não constitui um fim em si, mas um meio de fornecer informações e subsídios para uma consequente tomada de decisão visando atingir determinados objectivos (MORETTIN, 1981).

CHAFTFIELD (2000) considera lamentável que o termo "modelo de previsão" seja usado de maneira pouco rigorosa por muitos outros autores verificando que é, por vezes, erradamente usado para descrever um método de previsão.

Ainda de acordo com WHEELWRIGTH (1985), os métodos de previsão de séries temporais, que se classificam como quantitativos, baseiam as suas previsões na extrapolação de características de observações passadas e no inter-relacionamento entre essas, fornecendo previsões fidedignas se o futuro apresentar comportamento similar ao passado.

WHEELWRIGTH (1985) afirma que a maioria dos métodos de previsão de séries temporais baseia-se na suposição de que as observações passadas contêm todas as informações sobre o padrão de comportamento da série temporal e esse padrão é recorrente no tempo.

Segundo CHAFTFIELD (2000), os métodos de previsão podem ser classificados em três tipos :

(a) As previsões *Judgemental* – que têm com base o julgamento subjectivo, a intuição, 'dentro' do conhecimento comercial e sem qualquer outra informação relevante.

(b) os métodos univariados – em que as previsões dependem apenas no presente dos valores passados, de uma única série que se está a estudar, podendo ser auxiliada por uma função do tempo ou por uma tendência linear.

(c) métodos multivariados – em que as previsões de uma dada variável dependem, pelo menos em parte, dos valores de uma ou mais variáveis adicionais, chamados preditores ou variáveis explicativas. As previsões utilizando métodos multivariados podem

dependem de um modelo multivariado, envolvendo mais de uma equação se as variáveis forem dependentes.

Dado que existem inúmeros métodos para previsões de séries temporais e que, segundo WHEELRIGHT et al. (1994), nem sempre com os métodos mais complexos se obtém os melhores resultados, será necessário avaliar primeiro os benefícios dos vários métodos, antes de iniciar uma previsão de uma determinada série temporal.

Na escolha do método de previsão que melhores resultados nos possa dar, é importante ter em atenção algumas destas medidas de erro:

- Comparação gráfica;
- Diagrama de dispersão;
- Coeficiente de correlação;
- Erro quadrático acumulado;
- Raiz do erro médio quadrático (RMSE);
- Erro percentual absoluto (MAPE);
- Etc.

Para CHAFTFIELD (2000) é também possível classificar os métodos de previsão de mais duas maneiras diferentes:

- Uma forma alternativa importante de classificar métodos de previsão é entre os métodos automáticos, que não requerem a intervenção humana, e métodos não-automáticos, os que requerem a intervenção humana.
- Uma outra classificação é entre métodos simples e complicados. Os métodos univariados são geralmente mais simples do que os métodos multivariados, mas é sempre uma questão de *juízo* relativamente à dificuldade de cada método para fazer uma previsão.

2.5.1 Métodos simples de previsão de séries.

Muitos estudos que envolvem séries temporais utilizam métodos de previsão simples, para efectuarem a previsão de valores futuros da série temporal através de observação de valores passados da série em estudo.

Segundo MORETTIN e TOLOI (1981) o objectivo destes métodos é identificar um padrão básico presente nos dados históricos da série e conseqüentemente prever futuros valores. Dado que estes métodos estatísticos univariados, são fáceis de programar, o seu custo informático é pequeno e conseguem obter uma razoável previsão, estes métodos são bastante utilizados.

2.5.1.1 Modelos de Média Móvel.

Os modelos de média móvel utilizam a média das observações registadas, como previsão para um determinado período no futuro. Nestes modelos tem-se que as médias móveis podem ser simples, centradas ou ponderadas.

Segundo MORETTIN e TOLOI (1981) o termo média móvel é utilizado porque à medida que uma observação está disponível, a média das observações é recalculada com a utilização deste valor e desprezando o mais antigo.

Para os modelos que utilizam a média móvel simples pode-se definir uma equação do tipo:

$$X_t = \frac{x_{t-1} + x_{t-2} + \dots + x_{t-n}}{n}$$

em que: n (janela de observações) – número de observações.

X_t – média.

Nestes modelos o número de observações é proporcional ao efeito do alisamento na previsão.

Dado que as séries temporais têm comportamentos diferentes deve-se:

- Utilizar uma janela de observações maior, para séries que apresentam muita aleatoriedade ou pequenas mudanças nos seus padrões, de modo a que o cálculo da média móvel fique imune a ruídos e movimentos curtos.
- Utilizar uma janela de observações menor, para séries que apresentam pouca flutuação aleatória nos dados ou mudanças significativas, de modo a reagir mais rapidamente a essas mudanças.

2.5.1.2 - Alisamento exponencial.

Os modelos de alisamento exponencial dividem-se em: modelos de alisamento exponencial simples, modelos de alisamento exponencial simples, modelos de Holt-Winters com e sem sazonalidade.

2.5.1.2.1 - Modelos de alisamento exponencial simples.

Tal como os modelos de média móvel os modelos de alisamento simples utilizam as observações registadas, para determinar o comportamento das séries. Mas enquanto que nos modelos de média móvel simples, todas as observações usadas têm um peso igual no cálculo da previsão, nos modelos de alisamento simples cada observação da série tem um peso diferente. Neste modelo o valor das últimas observações são evidenciadas pela aplicação de uma constante de alisamento α , na previsão no momento anterior.

A fórmula para o cálculo do alisamento exponencial é dado pela equação:

$$F_{t+1} = \alpha x_t + (1 - \alpha)F_t$$

Em que: F_{t+1} – representa a previsão no tempo t+1;
 α ($0 < \alpha < 1$) (constante de alisamento) – peso atribuído à observação;
 x_t – observação mais recente;
 F_t – valor da previsão anterior.

Para valores de α pequenos, o peso das observações iniciais é maior na previsão do comportamento da série enquanto que valores de α maiores (perto de 1), serão as observações recentes que mais afectaram a previsão.

2.5.1.2.2 - Modelo de Alisamento Exponencial Linear

Segundo MORETTIN e TOLOI (1981) se o modelo de alisamento exponencial simples é aplicado para previsão, em que a série temporal apresenta tendência para as observações mais antigas então os valores reais poderão ser superestimados ou subestimados, pelo que o valor da previsão poderá ser posto em causa.

Este método foi desenvolvido para reconhecer a presença de uma tendência na série temporal, melhorando assim o valor da previsão, que será dada pela seguinte equação:

$$F_{t+n} = S_t + mT_t$$

onde: S_t designa a previsão no tempo t e é dada por:

$$S_t = \alpha x_t + (1 - \alpha)(S_{t-1} + T_{t-1});$$

T_t representa a componente de tendência e é dada por:

$$T_t = \beta(S_t - S_{t-1}) + (1 - \beta)(T_{t-1});$$

α e β são coeficientes de alisamento.

2.5.1.2.3 - Método de Alisamento Exponencial Holt-Winter com sazonalidade.

Este método de previsão é similar ao método de alisamento exponencial linear, contudo é capaz de estudar Séries Temporais que apresentam tendências sazonais e o valor da previsão será dado pelas equações:

$$S_t = \alpha \frac{x_t}{I_{t-l}} + (1 - \alpha)(S_{t-1} + T_{t-1})$$

$$T_t = \beta(S_t - S_{t-1}) + (1 - \beta)(T_{t-1})$$

$$I_t = \gamma \frac{x_t}{S_t} + (1 - \gamma)I_{t-l}$$

$$F_{t+m} = (S_t + mT_t)I_{t-l+m}$$

onde: I_t - alisamento do factor de sazonalidade $\frac{x_t}{S_t}$;

l - intervalo de sazonalidade ;

γ - peso atribuído ao factor de sazonalidade.

2.6 Análise de séries temporais

2.6.1 Teste de normalidade

Existem vários testes utilizados para averiguar se os dados de uma série temporal são normalmente distribuídos. Para os dados em estudo aplicaram-se o teste de Doornik-Hansen, o teste de Shapiro-Wilk, o teste de Lilliefors e o teste de Jarque-Bera, que estão disponíveis no Gretl.

Para estes testes a hipótese nula é que os dados provêm de uma distribuição normal, e a hipótese alternativa é de que os dados não provêm de uma distribuição normal.

2.6.2 Teste da raiz unitária ou teste de estacionaridade.

Uma série temporal será estacionária se as suas propriedades não se alterarem ao longo do tempo. A análise do gráfico da série temporal poderá dar uma ideia do comportamento dos dados, mas é insuficiente para verificar se será estacionária. Existem vários testes para fazer este estudo, baseando-se a maioria em encontrar uma raiz unitária. Segundo a literatura deve-se utilizar mais do que um teste de raiz unitária de modo a certificar a estacionaridade da série. Alguns dos testes mais utilizados são os testes de Dickley-Fuller Aumentado, teste de Phillips-Perron e teste de KPSS.

Teste de Dickley-Fuller Aumentado.

Este teste também é conhecido como teste ADF (*Augmented Dickley – Fuller*) e baseia-se na seguinte regressão:

$$\Delta y_t = \beta_1 + \beta_2 t + \delta y_{t-1} + \sum_{i=1}^m \alpha_i \Delta y_{t-1} + \varepsilon_t$$

com: β_1 – intercepto da série, também conhecido como *drit* da série;

β_2 – coeficiente de tendência ;

δ – coeficiente de presença de raiz unitária ;

m – número de defasagens tomada na série;

Neste teste considera as seguintes hipóteses:

$H_0 : \delta = 0$ - Existe pelo menos uma raiz unitária.

$H_1 : \delta \neq 0$ – A série não é estacionaria.

Será feita uma regressão de Δy_t em $y_{t-1}, \Delta y_{t-1}, \dots, \Delta y_{t+p-1}$ e de seguida calculada a estatística T, que será dada por:

$$T = \frac{\hat{\delta}}{se(\hat{\delta})}$$

com: $\hat{\delta}$ - estimador para δ ;

$se(\hat{\delta})$ – estimador para o desvio padrão do erro de δ .

Os valores críticos da estatística T são calculados por Dickley e Fuller através de simulação Monte Carlo.

Teste de Phillips-Perron

Este teste é conhecido na literatura pela sigla PP e é uma generalização do teste de Dickley-Fuller para os casos em que os erros $\{\varepsilon_t\}, t \in \mathbb{Z}$ são correlacionados e, possivelmente, heterocedásticos, considerando portanto as mesmas hipóteses, baseando-se no estudo da seguinte regressão:

$$\Delta y_t = \beta_1 + \beta_2 t + \delta y_{t-1} + \sum_{i=1}^m \alpha_i \Delta y_{t-1} + \varepsilon_t$$

Particularmente neste teste, a estatística Z é calculada por:

$$Z = n\hat{\delta}_n - \frac{n^2\hat{\sigma}^2}{2s_n^2} (\hat{\lambda}_n^2 - \hat{\gamma}_{0,n})$$

sendo:

$$\hat{\gamma}_{j,n} = \frac{1}{n} \sum_{i=1+j}^n r_i r_{i-j}$$

$$\hat{\lambda}_n^2 = \hat{\gamma}_{0,n} + 2 \sum_{j=1}^q \left(1 - \frac{j}{q+1}\right) \hat{\gamma}_{j,n}$$

$$s_n^2 = \frac{1}{n-k} \sum_{i=1}^n r_i^2$$

com: r_i - representado o resíduo em y_i utilizando estimadores de mínimos quadrados

k - número de covariáveis na regressão

q - número de defasagens utilizadas para calcular $\hat{\lambda}_n^2$

De referir que Z é um ajuste da estatística de Dickley-Fuller e que se o processo for não correlacionado terá covariâncias nulas e portanto $\hat{\lambda}_n^2 = \hat{\gamma}_{0,n}$. Por outro lado se o processo não for heterocedástico ficará $se(\delta) = \frac{1}{n}$ e Z será dada por:

$$Z = n\hat{\delta} = \frac{\hat{\delta}}{se(\hat{\delta})}$$

Assim sendo Z transforma-se na estatística de Dickley-Fuller e portanto, terá a mesma distribuição da estatística do teste ADF.

Teste KPSS

Este teste tem por finalidade determinar a estacionariedade numa série temporal e foi criado por Denis Kwiatkowski, Peter C. B. Phillips, Peter Schmidt e Yongcheol Shin.

Comparando com os testes anteriores (ADF e PP), as hipóteses para determinar a estacionariedade são trocadas, ou seja:

H_0 = A série é não estacionária;

H_1 = A série apresenta raiz unitária.

Considerando $Y_t : t \in \mathbb{N}$, as observações de uma série temporal, esta será decomposta em componentes de tendência determinista (ξ_t), passeio aleatório (r_t) e erro estacionário (ε_t) sendo:

$$Y_t = \xi_t + r_t + \varepsilon_t$$

$$r_t = r_{t-1} + \mu_t$$

onde: μ_t - representa as flutuações do passeio aleatório.

Considerando $e_t, t \in \mathbb{N}$ os resíduos de uma regressão em y explicado pelas componentes de tendência, passeio aleatório e intercepto. O estimador para a variância dos erros da regressão é dado por:

$$\hat{\sigma}_\varepsilon^2 = \frac{SQE}{E}$$

Define-se a soma parcial dos resíduos por:

$$S_t = \sum_{i=1}^t e_i, \quad t = 1, 2, \dots, T$$

a estatística deste teste é definida por:

$$LM = \sum_{t=1}^N \frac{S_t^2}{N^2 \hat{\sigma}_\varepsilon^2}$$

Para a análise da estacionaridade das séries temporais a estudar foram utilizados os testes que o programa *GRET*L tem à disposição, teste de Dickey-Fuller aumentado, teste ADF-GLS e o teste KPSS.

3. Modelos ARIMA

Neste capítulo será feita uma análise à metodologia de Box-Jenkins, para posterior comparação com a modelação feita com Análise Espectral Singular.

3.1 Modelos autoregressivos (AR).

Quando se pretende utilizar a modelação ARIMA, na análise de séries temporais é necessário, em primeiro lugar, identificar o modelo ARIMA que melhor se ajusta aos dados em questão.

Recorre-se principalmente às autocorrelações e autocorrelações parciais estimadas para a escolha do modelo mais adequado.

a_t é o erro aleatório do modelo, com $E(a_t) = 0$ e $Var(a_t) = \sigma^2$.

Descrição de um modelo AR.

O modelo AR(1) é um modelo autorregressivo de ordem $p = 1$ que assume grande importância, em que a série temporal X_t definido por:

$$X_t = \phi_1 X_{t-1} + a_t$$

onde:

ϕ_1 – parâmetro auto-regressivo de ordem 1;

X_{t-1} – série de tempo desfasado um período;

a_t – termo do erro do modelo;

Generalizando e considerando X_t , $t \in \mathbb{Z}$, pode-se afirmar que se trata dum modelo autorregressivo de ordem p ($X_t, \sim AR(p)$) se:

$$X_t = \phi_0 + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + a_t$$

sendo: $\phi_0, \phi_1, \dots, \phi_p$ parâmetros reais.

3.2 Modelos de média móvel (MA).

Diz-se que um processo linear $\{X_t, t \in \mathbb{Z}\}$ é um processo de média móvel de ordem q , e designa-se por MA(q), se satisfaz a seguinte equação de diferenças:

$$X_t = \alpha_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

onde $\mu, \theta_1, \dots, \theta_q$ são constantes reais. e $a_t \sim N(0, \sigma^2)$.

3.3 Modelos autorregressivos e de médias móveis (ARMA).

Os modelos autorregressivos e de médias móveis designam-se por ARMA(q,p) e são a junção dos modelos AR e MA e podem representar-se por:

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \alpha_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

O modelo ARMA(1,1) é um dos que mais se utiliza, pelo que:

$$X_t = \phi_1 X_{t-1} + \alpha_t - \theta_1 a_{t-1}$$

A f.a.c.v. poderá ser calculada por:

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + \dots + \phi_p \gamma_{j-p} + \gamma_{x_a}(j) - \theta_1 \gamma_{x_a}(j-1) - \dots - \theta_q \gamma_{x_a}(j-p)$$

sendo $\gamma_{x_a}(j)$ a covariância cruzada ente X_t e α_t e definida por:

$$\gamma_{x_a}(j) = \begin{cases} = 0, & j > 0, \\ \neq 0, & j \leq 0. \end{cases}$$

a f.a.c.v. será definida por:

$$\gamma_j = \phi_1 \gamma_{j-1} + \phi_2 \gamma_{j-2} + \dots + \phi_p \gamma_{j-p}, \quad j > q$$

e a f.a.c. será definida por:

$$\rho_j = \frac{\gamma_j}{\gamma_0} = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2} + \dots + \phi_p \rho_{j-p}, \quad j > q$$

Se esta f.a.c. for infinita, a decrescer de acordo com exponenciais e/ou senóides amortecidas após “lag” p-q é um bom indicativo para a escolha do modelo ARMA(p,q) para a análise dos dados.

3.4 Modelos autorregressivos integrados e de médias móveis (ARIMA).

Os modelos ARIMA são modelos estatísticos lineares para análise de séries temporais. A abreviação ARIMA em inglês significa “Auto-Regressive Integrated Moving Average Model”, ou seja, auto-regressivo, integrado e médias móveis. Os termos auto-regressivos correspondem a defasagens da série e as médias móveis são as defasagens dos erros aleatórios.

Os modelos ARIMA (p , d, q) são em teoria, a classe mais geral de modelos para a previsão de uma série temporal que não sendo estacionárias podem ser estacionadas por diferenciações.

Os modelos ARIMA são uma generalização dos modelos ARMA que visam basicamente tornar o processo estacionário através de operações diferenças.

Os atrasos da série diferenciada que aparecem na equação de previsão são chamados de termos "auto-regressivos "

Os defasamentos dos erros de previsão são chamados de termos de "média móvel", e uma série de temporal que necessita ser diferenciada para ser estacionária é dita ser uma versão "integrada" de uma série estacionária.

Ao verificar o gráfico de uma série temporal se os dados indicarem uma não estacionaridade, então procedesse à sua diferenciação, para transformar a série não estacionária numa série estacionária. Neste processo efectua-se a subtracção de um dado ao dando anterior.

Far-se-á uma diferenciação de 1ª ordem se a série esta a crescer a uma taxa constante e uma diferenciação de 2ª ordem se a serie estiver a crescer com uma taxa de crescente.

$$W_T = \nabla^d X_t = (1 - B)^d X_t$$

$$W_t = \alpha_1 W_{t-1} + \dots + \alpha_p W_{t-p} + a_t + \beta_1 a_{t-q}$$

$$\phi(B)(1 - B)^d X_t = \theta(B)\epsilon_t$$

Um modelo ARIMA não sazonal é classificado como um modelo " ARIMA (p , d, q) ", em que:

p é o número de termos auto-regressivos;

d é o número de diferenciações para que a série se torne estacionária;

q é o número termos de médias móveis.

Os termos p, d e q são todos inteiros maiores ou iguais a zero.

Assim, passeios aleatórios, modelos aleatórios de tendências, modelos auto-regressivos e os modelos de suavização exponencial (isto é, de médias móveis exponenciais ponderadas) são todos casos especiais de modelos ARIMA.

Considerando Z_t uma série temporal não estacionária, $W_t = \Delta Z_t = Z_t - Z_{t-1}$ será a Série diferenciada uma vez de Z_t .

Assim sendo $W_t^d = \Delta^d Z_t$ Será a série temporal diferenciada d vezes de Z_t .

3.5 Identificação de modelos ARIMA.

A identificação de um modelo ARIMA é efectuada recorrendo às funções de autocorrelação (f.a.c.) e às funções de autocorrelação parcial (f.a.c.p.).

Função autocorrelação (f.a.c.)

Nalguns modelos, nomeadamente os AR(p), MA(q) e ARMA(p,q), pode-se encontrar características especiais das f.a.c. que permitem optar pela sua escolha:

1. se a f.a.c. for infinita em extensão e a decrescer como uma função exponencial e/ou função sinusoidal amortecida, possível a opção pelo modelo AR(p);
2. se a f.a.c. for finita e apresenta um corte após a defasagem q, possível a opção pelo modelo MA(q);
3. se a f.a.c. for infinita e decresce com características exponenciais e/ou sinusoidais amortecidas após a defasagem (q,p), possível a opção pelo modelo ARMA(p,q).

Com a estimação das f.a.c. pode-se encontrar um padrão que se identifique com algum modelo em particular, sendo mais eficaz para identificar em particular os modelos MA(q).

Uma f.a.c. ρ_j é estimada por:

$$r_j = \frac{g_j}{g_0}, \quad j = 0, 1, \dots, N - 1$$

Sendo g_j é a estimativa da f.a.c.v. γ_j definida por:

$$g_j = \frac{1}{N} \sum_{t=1}^{N-j} [(X_t - \hat{X})(X_{t+j} - \hat{X})], \quad j = 0, 1, \dots, N - 1$$

Em que \hat{X} é a média amostra definida por:

$$\hat{X} = \frac{1}{N} \sum_{j=1}^N X_t$$

Sendo funções pares as funções de auto correlação, então $g_{-j} = g_j$ e $r_{-j} = r_j$.

Função autocorrelação parcial (f.a.c.p.)

Esta função corresponde à correlação de X_t e X_{t-k} em que se retira o efeito das observações $X_{t-1}, X_{t-2}, \dots, X_{t-k+1}$ e é denotada por ϕ_{kk} e tem a seguinte expressão:

$$\phi_{kk} = \text{Corr}((X_t, X_{t-k}) / (X_{t-1}, X_{t-2}, \dots, X_{t-k+1}))$$

A f.a.c.p. é útil para identificar modelos AR puros ou da parte AR de um processo ARMA (p,q), não sendo útil para identificar modelos MA e ARMA.

Um processo simples para estimar as f.a.c.p. consiste em substituir nas equações de Yullie-Walker as f.a.c. pelas suas estimativas.

Equações de Yule-Walker

Processo $AR(p)$: $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \delta + \varepsilon_t$

Covs. de lag k : $\gamma_k = E[(Y_{t-k} - \mu)(\phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \dots + \phi_p(Y_{t-p} - \mu) + \varepsilon_t)]$

Sistema de equações para as covariâncias

$$\begin{aligned} \gamma_0 &= \phi_1 \gamma_1 + \phi_2 \gamma_2 + \dots + \sigma^2 \\ \gamma_1 &= \phi_1 \gamma_0 + \phi_2 \gamma_1 + \dots + \phi_p \gamma_{p-1} \\ &\dots\dots\dots \\ \gamma_p &= \phi_1 \gamma_{p-1} + \phi_2 \gamma_{p-2} + \dots + \phi_p \gamma_0 \end{aligned}$$

Para $k > p$: $\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \dots + \phi_p \gamma_{k-p}$

Se derivarmos as equações anteriores.

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1 + \dots + \phi_p \rho_{p-1} \\ &\dots\dots\dots \\ \rho_p &= \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \dots + \phi_p \end{aligned}$$

Para $k > p$: $\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \dots + \phi_p\rho_{k-p}$

Se conhecermos ρ_1, \dots, ρ_p , será possível resolver as equações de Yule-Walker para ϕ_1, \dots, ϕ_p .

3.6 Estimação de modelos ARIMA.

Terminado o processo de identificação da ordem do modelo ARIMA para a análise de uma série temporal, pode-se estimar os parâmetros desse modelo utilizando:

- método dos momentos;
- estimadores de mínimos quadrados;
- estimadores de máxima verossimilhança.

3.7 Diagnóstico de modelos ARIMA

Depois de realizados os dois passos anteriores (identificação e estimação do modelo) é necessário verificar se este está adequado aos dados. É a análise dos resíduos que nos permite fazer o diagnóstico do modelo.

Considerando o modelo ajustado

$$\phi(B)W_t = \theta(B)a_t$$

com

$W_t = \nabla^d X_t$ com d diferenças nas séries

$\phi(B)$ e $\theta(B)$ são polinómios do operador retardado B

os resíduos serão descritos por

$$a_t = \theta^{-1}(B)\phi(B)W_t$$

Considera-se que o modelo é adequado aos dados se os resíduos exactos a_t se se identifica. com uma distribuição $N(0, 1)$.

Para verificar se os resíduos estimados são não correlacionados podemos utilizar os testes de Box-Pierce ou o teste de Ljung Box.

Para a escolha mais acertada do modelo ARIMA serão utilizados os critérios AIC, BIC e HQC.

3.7.1 Testes de Box-Pierce e Ljung-Box

Estes testes são ferramentas para testar o ajuste de um modelo de série temporal. Este teste é aplicado aos resíduos de uma série temporal, após o ajustamento de um modelo. São analisadas m autocorrelações dos resíduos e quanto menor for esse valor melhor o ajustamento do modelo escolhido.

A estatística de Box-Pierce pode ser definida da seguinte forma:

$$Q = N \sum_{k=1}^m \hat{r}_k^2$$

onde: N – é o tamanho da amostra;

m – a duração da defasagem;

\hat{r}_k - autocorrelação estimada da série.

Para grandes amostras, a estatística Q distribui-se segundo uma distribuição qui-quadrado com m graus de liberdade (gl).

O teste de Ljung-Box é uma alteração do teste de Box-Pierce, tornando-o mais generalizado e apresenta, segundo a literatura, melhores resultados.

A estatística de Ljung-Box pode se expressa como:

$$LB = N(N + 2) \sum_{k=1}^m \left(\frac{\hat{r}_k^2}{N - k} \right)$$

A distribuição assintótica da estatística LB também é qui-quadrado com m graus de liberdade (gl)

A hipótese nula para os dois testes refere a não existência de auto correlação conjunta dos resíduos. Esta hipótese será rejeitada se o valor da estatística do teste for superior ao valor crítico da distribuição qui-quadrado para um determinado nível de significância α e grau de liberdade $gl=m$, isto quer dizer que as autocorrelações para um determinado número de *lags* poderão ser significativamente diferentes de zero, mostrando que os valores não são aleatórios e independentes ao longo do tempo.

3.7.2 Critério de escolha dos modelos.

Critério de Informação de Akaike (AIC)

O Critério de Informação de Akaike (AIC) baseia-se na existência de um modelo “real” que é desconhecido mas permite descrever os dados. Este critério minimiza a divergência de Kullback-Leibler (K-L).

A estimativa do AIC para um determinado modelo é dada por $AIC = -2L + 2K$, onde L é a log-verossimilhança máxima e K é o número de parâmetros do modelo.

O modelo com menor valor de AIC é considerado o modelo de melhor ajuste.

Critério de informação Bayesiano (BIC)

Este critério também denominado de Schwarz é definido como a estatística que maximiza a probabilidade de se identificar o melhor modelo dentre os que se estão a estudar.

A estimativa do BIC para um determinado modelo é dada por

$$AIC = -2L + 2K \ln(n)$$

sendo: L é a log-verossimilhança máxima

K é o número de parâmetros do modelo

n o número de observações.

Tal como no critério anterior, o modelo que apresentar menor valor da estimativa BIC é considerado o de melhor ajuste.

Critério Informações Hannan-Quinn (HQC)

O critério informações Hannan-Quinn é outro instrumento de selecção de modelos. É uma alternativa ao critério de informação de Akaike (AIC) e critério de informação Bayesiano (BIC).

A estimativa do HQC para um determinado modelo é dada por:

$$HQC = -2L + 2K \ln(\ln(n))$$

sendo: L é a log-verossimilhança máxima;

K é o número de parâmetros do modelo;

n o número de observações.

Burnham & Anderson (2002, p. 287) dizem que HQC ", enquanto muitas vezes citado, parece ter visto pouco uso na prática." Eles também observam que HQC, como BIC, mas ao contrário de AIC, não é um estimador de Kullback-Leibler divergência. Claeskens & Hjort (2008, cap. 4), note que HQC, como BIC, mas ao contrário de AIC, não é assintoticamente eficiente, e ainda de salientar que qualquer método que está sendo usado para afinar o critério será mais importante na prática do que o termo $\ln \ln n$, uma vez que este último número é pequeno, mesmo para muito grande n .

3.8 Previsão com modelos ARIMA.

Para fazer a previsão de um determinado valor X_{t+h} , $h \geq 1$, partindo do pressuposto que são conhecidos todos os valores até ao instante t , ou seja, $X_{t-2}X_{t-1}, X_t$, considera-se t como a origem das previsões.

A previsão de h passos a partir da origem t será representada por $\hat{X}_t(h)$.

Conhecido o modelo ARIMA (p,d,q), estacionário, invertível e com parâmetros conhecidos são conhecidas 3 fórmulas de previsão:

- i. fórmula de equações de diferenças:

$$X_{t+h} = \phi_1 X_{t+h-1} + \dots + \phi_{p+d} X_{t+h-p-d} - \theta_1 a_{t+h-1} - \theta_2 a_{t+h-2} - \dots - \theta_q a_{t+h-q} + a_{t+h}$$

- ii. fórmula de choque aleatórios:

$$X_{t+h} = \sum_{j=-\infty}^{t+h} \psi_{t+h-j} a_j = \sum_{j=0}^{\infty} \psi_j a_{t+h-j}$$

com $\psi_0 = 1$ e os outros pesos obtido pela resolução do sistema de operadores de retardos dado por $\theta(B) = \phi(B)\psi(B)$.

- iii. fórmula invertida:

$$X_{t+h} = \sum_{j=1}^{\infty} \pi_j X_{t+h-j} + a_{t+h}$$

com os π_j obtidos pela resolução do sistema de operadores de retardos dado por $\phi(B) = \theta(B)\pi(B)$.

3.9. Modelos SARIMA

A fim de englobar o comportamento da sazonalidade de algumas séries temporais, aplicam-se os modelos ARIMA sazonais multiplicativos (SARIMA).

Os modelos SARIMA apresentam uma parte não sazonal a que corresponde os parâmetros (p,d,q) e uma parte sazonal a que corresponde os parâmetros (P,D,Q). Assim temos uma componente sazonal autorregressiva SAR(P) e média móvel SMA(Q), sendo apresentado como SARIMA (p,d,q)(P,D,Q), em que D se refere à ordem de diferenciação sazonal.

A definição do modelo SARIMA segundo BOX e JENKINS (1974) e dado por:

$$\phi(B)\Phi(B^s)W_t = \theta(B)\Theta(B^s)a_t$$

com: $W_t = \nabla_s^d \nabla^d X_t$

e com os seguintes operadores:

- auto-regressivo não sazonal - $\phi(B) = (1 - \alpha_1 B - \dots - \alpha_p B^p)$;

- auto-regressivo sazonal - $\Phi(B^s) = (1 - \phi_s B^s - \dots - \phi_p B^{ps})$;

- média móvel não sazonal.- $\theta(B) = (1 + \beta_1 B + \dots + \beta_q B^q)$;

$\Theta(B^s) = (1 + \theta_s B^s - \dots - \theta_p B^{qs})$ – operador média móvel sazonal;

∇^d – operador diferença não sazonal de ordem d;

∇_s^d – operador diferença sazonal de ordem D.

Neste trabalho, nos capítulos seguintes, sempre que são referidos os modelos ARIMA, considera-se que estes englobarão os modelos SARIMA.

4. - Análise Espectral Singular (AES).

A análise espectral singular também conhecido por SSA (*Singular Spectrum Analysis*) é uma nova e poderosa ferramenta para a análise de séries temporais.

Segundo a literatura consultada, a AES tem a sua génese nos trabalhos de BROOMHEAD e KING (1986a, 1986b) e BROOMHEAD e colaboradores (1987).

Só em 1996, foram reunidos e compilados os vários aspectos teóricos e metodológicos desta técnica, por ELSNER e TSONIS, num livro intitulado *Singular Spectral Analysis. A New Tool in Time Series Analysis*. São ainda autores importantes e de referência, com trabalhos publicados sobre a AES, NINA E. GOLYANDINA e HOSSEIN HASSANI.

Em Portugal é de salientar o trabalho de PAULO RODRIGUES e MIGUEL CARVALHO intitulado “Método de Imputação Recorrente: Análise Espectral Singular com Valores Omissos”, prémio SPE (Sociedade Portuguesa de Estatística) em 2009.

Renata Esquível (2012) refere que “a AES possui basicamente duas finalidades:

- *Modelação ou ajuste dos dados: destina-se a modelar a dependência entre as observações da série temporal. Para este objectivo da AES, o propósito principal é o entendimento do comportamento da série, ou seja, é a parte responsável pela identificação ou extracção de padrões geradores da série temporal.*
- *Previsão: Objectiva realizar previsões de valores futuros da série temporal; após o processo de modelação dos dados, assim como ocorre nos métodos clássicos de previsão”.*

4.1 - Primeira fase – Decomposição

Esta primeira fase é dividida em embutimento e decomposição em valores singulares.

4.1.1 Embutimento (*Embedding*)

O primeiro passo do algoritmo básico da AES é a construção da matriz de trajectórias dos dados iniciais da ST

Matriz trajectória $\mathbf{X}_{l \times k}$ é definida por:

$$\mathbf{X} = \begin{bmatrix} y_1 & y_2 & \cdots & y_k \\ y_2 & y_3 & \cdots & y_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_l & y_{l+1} & \cdots & y_{l+(k+1)} \end{bmatrix}$$

onde k é escolhido de modo a incluir todas as observações da Série Temporal na matriz \mathbf{X} , ou seja, $k = n - l + 1$.

No presente trabalho cada vector será representado na forma $x_i = [x_1 \dots x_{l+(i-1)}]'$ que se denominam $L - lagged\ vectors$ (vectors desfasados), com uma janela de comprimento l .

O parâmetro l será escolhido pelo utilizador (RODRIGUES e CARVALHO, 2008; HASSANI et al. 2009).

A matriz trajectória \mathbf{X} é simétrica e as suas entradas são constantes ao longo das diagonais paralelas à diagonal secundária, pelo que se pode considerar que é uma matriz Hankel.

Matriz trajectória $\mathbf{X}_{k \times l}^T$ é definida por:

$$\mathbf{X}^T = \begin{bmatrix} y_1 & y_2 & \cdots & y_l \\ y_2 & y_3 & \cdots & y_{k+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_k & y_{l+1} & \cdots & y_{l+(k+1)} \end{bmatrix}$$

esta matriz é também uma matriz trajectória da Série.

4.1.2. - Decomposição em Valores Singulares.

Neste segundo passo é realizada uma decomposição em valores singulares (DVS), da matriz trajectória \mathbf{X} na soma de matrizes singulares.

Vamos considerar $\mathbf{X}\mathbf{X}'$, como o produto da matriz trajectória pela sua transposta. Quando se realizar a decomposição em valores singulares, obtemos os valores

singulares ($\lambda_1 \geq \dots \geq \lambda_L \geq 0$) que são ordenados por ordem decrescente e os respectivos vectores singulares (U_1, \dots, U_L) ortogonais e normalizados.

Considerando $d = \max\{i, \text{tal que } \lambda_i > 0\} = \text{rank } \mathbf{X}$ em que d significa o número de valores singulares diferentes de zero da matriz $\mathbf{X}\mathbf{X}'$, e a transformação $V_i = \frac{\mathbf{X}'U_i}{\sqrt{\lambda_i}}$ então a DVS da matriz trajectória poderá ser expressa pela expressão:

$$\mathbf{X} = \mathbf{X}_1 + \dots + \mathbf{X}_d$$

em que $\mathbf{X}_i = \sqrt{\lambda_i}U_iV_i'$ ($i = 1, \dots, d$).

Na literatura especializada e consultada, $(\sqrt{\lambda_i}U_iV_i')$ denomina-se *i*-ésimo autotriple (*eigen-triple*) da matriz trajectória, $\sqrt{\lambda_i}$ é chamado de valor singular da matriz trajectória e o conjunto $\{\sqrt{\lambda_i}\}$ representa o espectro de \mathbf{X} (GOLYANDINA, NEKRUTKIN E ZHINGLAVSKY, 2001).

4.2 - Segunda fase – Reconstrução

Nesta fase, analisam-se as matrizes de modo a formar grupos semelhantes com a finalidade de formar os componentes e finalmente reconstruir a série temporal.

4.2.1 Agrupamento.

Este primeiro passo da reconstrução consiste na selecção dos componentes aditivos da série temporal em termos de matrizes separáveis. Vamos organizar grupos de componentes que estão mais correlacionadas entre si.

Encontrando componentes correlacionadas forma-se um grupo $I = \{i_1, \dots, i_p\}$ de índices i_1, \dots, i_p , em que a matriz \mathbf{X}_I correspondente a esse grupo fica definida por

$$\mathbf{X}_I = \mathbf{X}_{i_1} + \dots + \mathbf{X}_{i_p}$$

O derrame do conjunto de índices $J = 1, \dots, d$ no subconjunto disjunto I_1, \dots, I_m faz corresponder a representação.

$$\mathbf{X} = \mathbf{X}_{I_1} + \dots + \mathbf{X}_{I_m}$$

O processo de escolha dos conjuntos I_1, \dots, I_m é denominado *eigentriple grouping*.

4.2.2 - Diagonalização por Médias

Este passo é conhecido por *diagonal averaging* e consiste na transformação de cada matriz resultante num componente aditiva da Série Temporal original Y_t .

Considere-se uma matriz $Z_{L \times K} = (z_{ij}); 1 \leq i \leq L \wedge 1 \leq j \leq k$, em que $L \leq K$. O termo de ordem k da série resultante é calculado determinando a média z_{ij} para todo o $i, j : i + j = k + 2$. A este processo denomina-se média diagonal ou Hankelização da matriz Z .

Ao fazer a hankelização da matriz Z obtém-se a matriz Hankel. O resultado da hankelização de uma matriz $\mathcal{K}Z$, que não é mais do que a matriz trajectória correspondente à série temporal inicial, obtida com o resultado da média diagonal.

O principal interesse na utilização das matrizes de Hankel, baseia-se no facto de que qualquer matriz de Hankel é uma matriz trajectória muito próxima das séries temporais Y_N com $N = K + L - 1$ e vice versa (NINA GOLYANDINA, VLADIMIR NEKRUTKON, ANATOLY ZHIGLJAVSKY, 2001).

O resultado da aplicação do operador \mathcal{K} a uma matriz de Hankel se os seus elementos z_{ij} coincidem nas diagonais $i + j = s$ para qualquer $2 \leq s \leq L + K$.

Considerando $L^* = \min(L, K), K^* = \max(L, K), N = L + K - 1$ e $y_{ij}^* = y_{ij}$ se $L < K$, então cada termo da série “resultante” (pelo processo de Hankelização) pode-se determinar, segundo Buchstaber (1994) com a seguinte condição:

$$G_t = \begin{cases} \frac{1}{s-1} \sum_{q=1}^{s-1} Z_{(q,s-q)}, & 2 \leq s \leq L-1 \\ \frac{1}{L} \sum_{q=1}^L Z_{(q,s-q)}, & L-1 \leq s \leq L-1 \\ \frac{1}{K+L-s+1} \sum_{q=s-k}^L Z_{(q,s-q)}, & 2 \leq s \leq L-1 \end{cases}$$

em que s é a soma dos índices da matriz $Z_{L \times K}$ ($s = i + j$) e $N = L + K - 1$.

5. Análise dos dados.

Neste capítulo vão ser analisadas duas séries temporais, que foram seleccionadas por terem características distintas, de modo a que no final se possa atingir uma análise mais abrangente. Uma das séries é denominada MILK e refere-se à produção mensal de leite (libras por vaca), dados recolhidos entre Janeiro de 1962 e Dezembro de 1975. A outra série é denominada PIGS e refere-se ao número total mensal de suínos abatidos na província de Victoria, com dados recolhidos entre Janeiro 1980 e Agosto 1995.

No sub-capítulo 5.1 será feita uma análise descritiva das duas séries temporais, onde serão apresentados os resultados da realização de testes de normalidade e testes de raiz unitária ou de estacionaridade.

No sub-capítulo 5.2 será feita a análise de vários modelos auto-regressivos. A cada modelo em estudo foi aplicado os testes de Ljung-Box e avaliados os valores dos critérios de informação Akaike (AIC), Bayesiano (BIC) e de Hannan-Quinn (HQC). No final será proposto um modelo que melhor reflecte os dados em estudo.

No sub-capítulo 5.3 será utilizada a AES no tratamento da série temporal e escolhido o modelo que melhor traduz o desenvolvimento de cada um dos conjuntos de dados.

As previsões serão analisadas no sub-capítulo 5.4. Para os modelos escolhidos nos sub-capítulos 5.2 e 5.3 será feita a previsão com base nos últimos 12 meses de dados conhecidos.

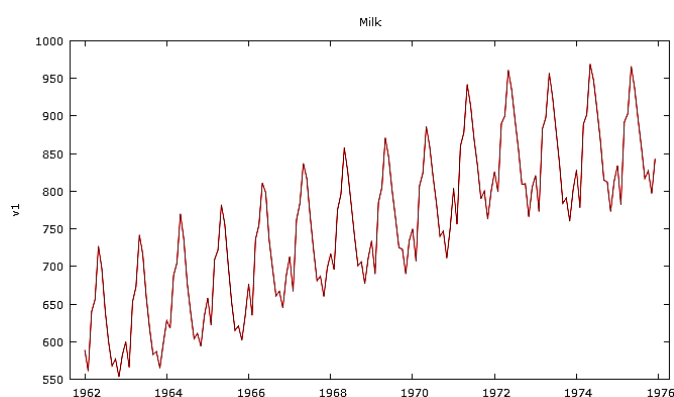
Finalmente no sub-capítulo 5.5, será feita uma análise comparativa das previsões para cada série temporal, de modo a verificar qual a técnica que melhores resultados apresenta.

5.1 Análise descritiva das séries temporais

Como se pretende a comparação das previsões feitas com os modelos ARIMA e os modelos utilizando AES, foram escolhidos dois conjuntos de dados com comportamentos diferentes, como se pode verificar graficamente nas figuras 5.1 e 5.2.

A série temporal Milk apresenta uma tendência crescente e um comportamento bem definido quanto à sazonalidade dos seus dados.

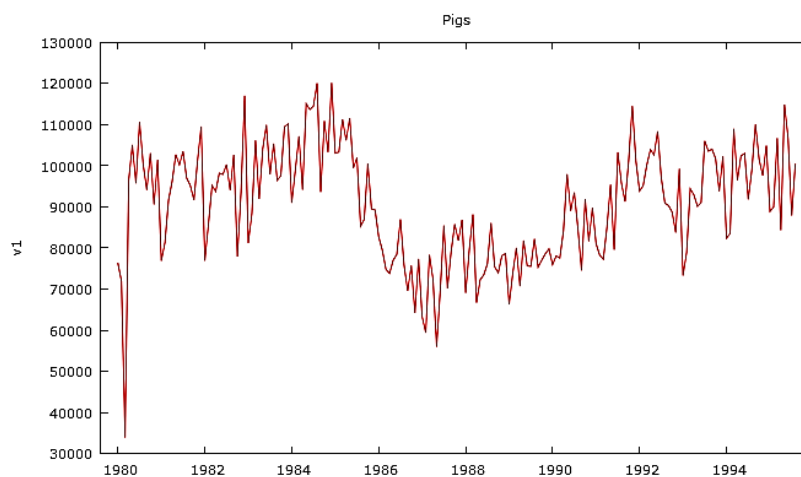
Figura 5.1. Gráfico dos dados da Série Temporal - Milk.



Fonte: Autor

Por seu lado a série temporal Pigs, figura 5.2, não apresenta nenhum padrão particular quanto à sua monotonia, podendo contudo verificar-se graficamente uma ligeira sazonalidade, não tão vincada como a anterior, mas que, como se verificará, será importante na escolha dos modelos.

Figura 5.2. Gráfico dos dados da Série Temporal - Pigs.



Fonte: Autor

5.1.1 Testes de normalidade

Existem vários testes utilizados para analisar se os dados de uma série temporal são normalmente distribuídos. Para os dados em estudo aplicaram-se, disponíveis no *Gretl*, o teste de Doornik-Hansen, o teste de Shapiro-Wilk, o teste de Lilliefors e o teste de Jarque-Bera.

Para estes testes a hipótese nula é que os dados provêm de uma distribuição normal, contra a alternativa de que os dados não provêm de uma distribuição normal.

Os resultados obtidos foram sintetizados nas tabelas 5.1. e 5.2.

Tabela 5.1. Resultados dos testes de normalidade - Milk

Dados - Milk	Valor de teste	Valores de p
Teste de Doornik-Hansen	4,89443	0,0865344
Teste de Shapiro-Wilk	0,982782	0,0353962
Teste de Lilliefors	0,0457531	0,52
Teste de Jarque-Bera	4,2782	0,117761

Fonte: Autor

Tabela 5.2. Resultados dos testes de normalidade - Pigs

Dados - Pigs	Valor de teste	Valores de p
Teste de Doornik-Hansen	4,80086	0,0906791
Teste de Shapiro-Wilk	0,978654	0,00564931
Teste de Lilliefors	0,0590931	0,11
Teste de Jarque-Bera	5,79785	0,0550824

Fonte: Autor

Da análise dos resultados pode concluir-se que, as duas séries temporais e considerando um nível de significância de 0,05, os dados provêm duma distribuição normal, dado que só o teste de Shapiro-Wilk tem um valor de $p < 0,05$ e este é mais indicado para um número de dados pequenos.

De seguida as séries temporais serão analisadas quanto à sua estacionaridade.

5.1.2 Testes de raiz unitária ou testes de estacionaridade.

Os resultados dos testes disponíveis no *Gretl*, encontram-se no anexo A. Pode-se verificar a análise produzida para cada série temporal, relativamente ao teste aumentado de Dickey-Fuller, Dickey-Fuller (GLS) e teste de KPSS.

Tendo-se concluído que as séries temporais não eram estacionárias, o passo seguinte será verificar qual o modelo ARIMA que melhor representa dos dados, para isso serão efectuados o teste de Ljung-Box e estudados os critérios de informação AIC, BIC e HQC.

5.2 Modelos ARIMA

Nesta secção serão efectuados os testes de Ljung-Box e os Critério de informação AIC, BIC e HQC.

5.2.1 Testes de Ljung-Box

No software utilizado, *Gretl*, é possível determinar o valor da estatística de Ljung-Box para os todos os modelos estudados.

São apresentados nas tabelas 5.3. e 5.4. seguintes os cinco valores mais baixos encontrados para o teste de Ljung-Box, referentes às duas séries, para um nível de significância de 0,05.

Tabela 5.3. Resultados do teste Ljung-Box para Milk

Milk		
Modelo	Ljung-Box Q'	valor p
(1,1,1)(1,1,1)	8,95597	P = 0,3460
(2,1,1)(1,1,1)	6,87398	P = 0,4421
(2,1,1)(2,1,1)	6,75871	P = 0,3437
(2,2,2)(1,1,1)	5,95156	P = 0,4286
(3,1,1)(1,1,1)	5,63673	P = 0,4651

Fonte: autor

Tabela 5.4. Resultados do teste Ljung-Box para Pigs

Pigs		
Modelo	Ljung-Box Q'	valor p
(1,1,1)(1,1,0)	32,2877	p = 0,0001776
(1,1,1)(1,1,1)	27,6701	p = 0,0005412
(2,0,1)(2,0,1)	15,3153	p = 0,01794
(2,1,1)(1,1,1)	14,9244	p = 0,03698
(2,1,1)(1,1,0)	13,0733	p = 0,1094

Fonte: autor

Os resultados deste teste não sendo decisivos na escolha do modelo mais apropriado, são uma referência para a continuidade do estudo. Tal como se constatou graficamente a série temporal Milk apresenta uma significativa tendência sazonal, facto constatado pelos valores mais baixos da estatística do teste, terem componente sazonal. Também se consegue observar que neste aspecto os dados Pigs não têm as mesmas características, apesar de a presença de sazonalidade ser um factor importante, que irá influenciar a escolha do modelo mais adequado.

Começa assim a ser interessante o estudo destas duas séries temporais, quer ao nível do comportamento dos modelos ARIMA, como posteriormente nos modelos AES.

5.2.2 Critério de escolha dos modelos.

Os critérios mais utilizados na escolha do modelo ARIMA, que melhor representa uma determinada série temporal são os critérios de informação de Akaike (AIC), Bayesiano (BIC) e de Hannan-Quinn (HQC). Todos os valores destes critérios foram obtidos através do programa *Gretl*.

Dado que o modelo com menor valor em cada critério é considerado o modelo de melhor ajuste, serão apresentados nesta secção os modelos que registaram o 5 menores valores para cada critério (tabelas 5.5 a 5.10).

Contudo podem ser consultados no anexo C, tabelas C1 a C6, os resultados dos testes efectuados para os vários modelos estudados.

Critérios de informação de Akaike (AIC).

Tabela 5.5 Resultados critério AIC - Milk

Milk	
Modelo	Critério de Akaike
(2,2,2)(1,1,1)	1073,726
(3,1,1)(1,1,1)	1072,562
(2,1,1)(2,1,1)	1072,153
(1,1,1)(1,1,1)	1071,773
(2,1,1)(1,1,1)	1070,47

Fonte: Autor

Tabela 5.6. Resultados critério AIC - Pigs

Pigs	
Modelo	Critério de Akaike
(2,0,1)(2,0,1)	3740,748
(1,1,1)(1,1,0)	3739,478
(1,1,1)(1,1,1)	3723,499
(2,1,1)(1,1,0)	3721,998
(2,1,1)(1,1,1)	3708,799

Fonte: Autor

Critério de informação Bayesiano (BIC).

Tabela 5.7. Resultados critério BIC - Milk

Milk	
Modelo	Critério de Schwarz
(3,1,1)(1,1,1)	1096,909
(2,1,1)(2,1,1)	1096,501
(2,2,1)(1,1,1)	1096,199
(2,1,1)(1,1,1)	1091,774
(1,1,1)(1,1,1)	1090,033

Fonte: Autor

Tabela 5.8. Resultados critério BIC - Pigs

Pigs	
Modelo	Critério de Schwarz
(2,0,1)(2,0,1)	3762,942
(1,1,1)(1,1,0)	3755,301
(1,1,1)(1,1,1)	3742,488
(2,1,1)(1,1,0)	3740,987
(2,1,1)(1,1,1)	3730,952

Fonte: Autor

Critério Informações Hannan-Quinn (HQC).

Tabela 5.9. Resultados critério HQC - Milk

Milk	
Modelo	Critério Hannan-Quinn
(2,2,1)(1,1,1)	1083,576
(3,1,1)(1,1,1)	1082,451
(2,1,1)(2,1,1)	1082,043
(1,1,1)(1,1,1)	1079,19
(2,1,1)(1,1,1)	1079,123

Fonte: Autor

Tabela 5.10. Resultados critério HQC - Pigs

Pigs	
Modelo	Critério Hannan-Quinn
(2,0,1)(2,0,1)	3749,75
(1,1,1)(1,1,0)	3745,896
(1,1,1)(1,1,1)	3731,202
(2,1,1)(1,1,0)	3729,701
(2,1,1)(1,1,1)	3717,785

Fonte: Autor

Os valores encontrados nestes critérios, estão em sintonia com os obtidos nos testes de Ljung-Box e com as observações gráficas efectuadas. Continuam a mostrar que os modelos com valores mais baixos para os dados Milk são os que têm uma componente sazonal. Esta situação não é tão evidente para os modelos com valores mais baixos nos dados Pigs. Pela análise dos três critérios verificou-se que para os dois conjuntos de dados os cinco modelos com valores mais baixos eram sempre os mesmos.

Assim considerou-se que o modelo ARIMA (2,1,1)(1,1,1) apresenta o melhor ajuste para as duas séries. Serão estes os modelos escolhidos para fazer a previsão e posterior comparação com os modelos AES.

5.3 Análise Espectral Singular

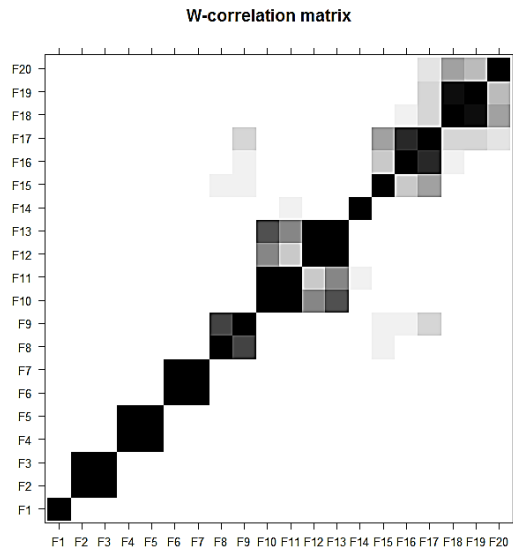
Para a análise das duas séries temporais utilizei o software R, dado ser de utilização livre e aquele com que mais me identificava.

A programação utilizada para o estudo dos modelos AES (scripts) será apresentada na totalidade no anexo E.

Primeiramente foram construídas as matrizes correlação (matrizes trajectórias), para os dados Milk e Pigs, que permitem determinar quantas componentes devem ser retidas pelo modelo da AES. Estas matrizes são apresentadas nas figuras 5.3. e 5.4.

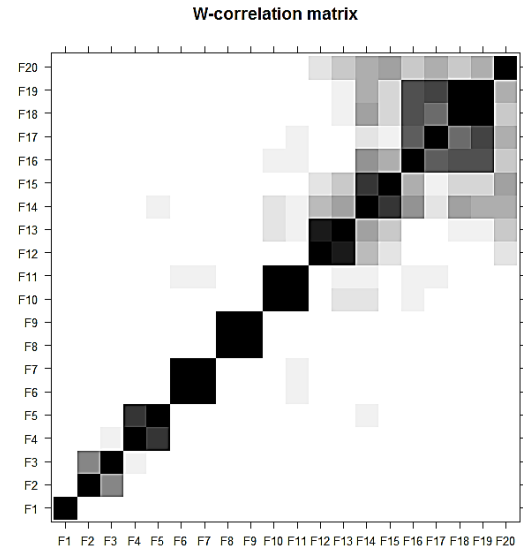
No canto inferior esquerdo verificamos as componentes associadas ao sinal, enquanto no canto superior direito as componentes associadas ao ruído.

Figura 5.3. Milk Matriz correlação



Fonte: Autor

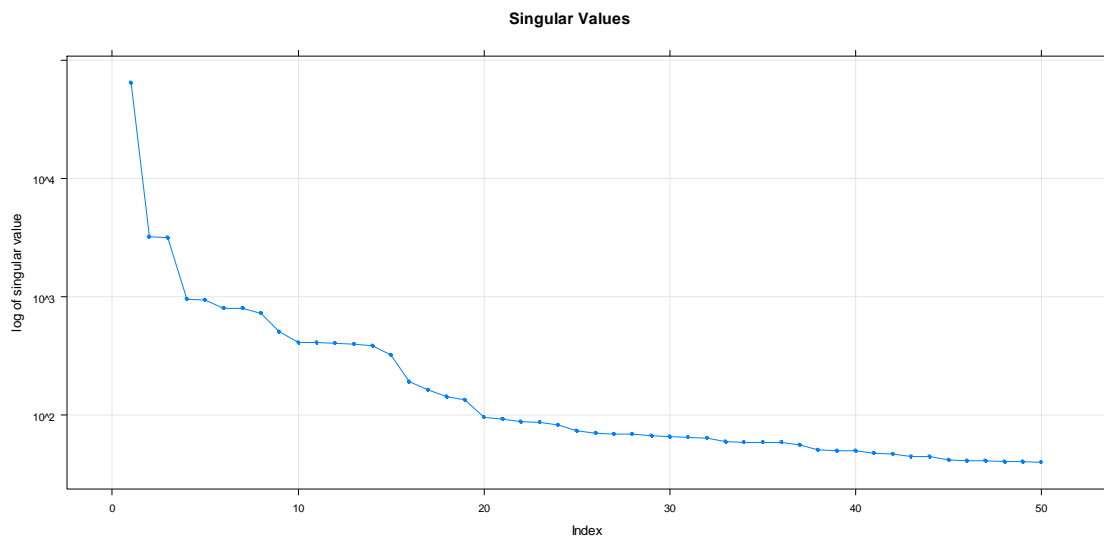
Figura 5.4. Pigs Matriz Correlação



Fonte: Autor

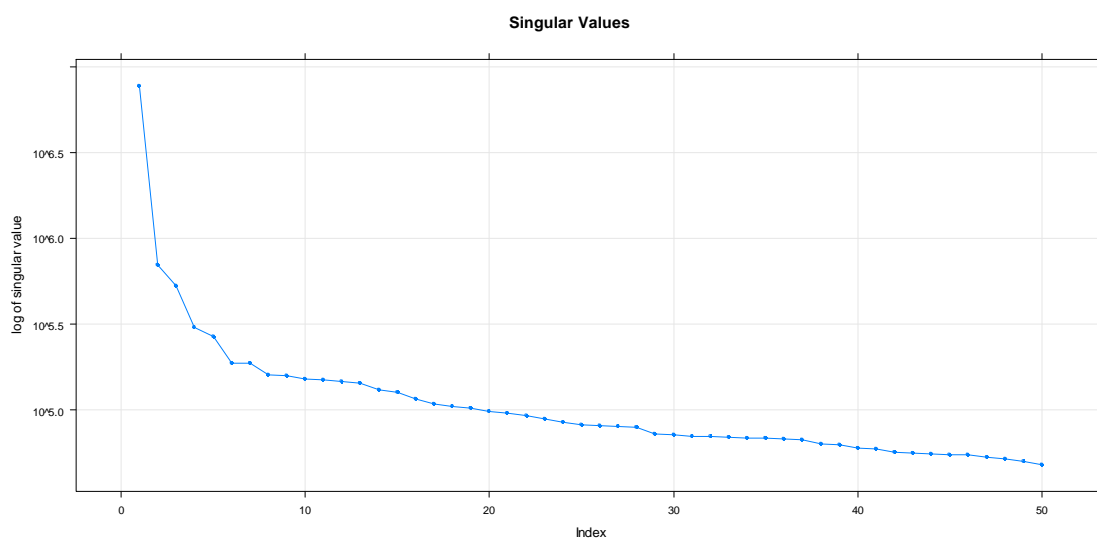
Um dos passos importantes na AES é a decomposição em valores singulares. Para os dados em estudo foi efectuado no R a decomposição em valores singulares, as séries temporais, e obtidos os gráficos das figuras 5.5. e 5.6.

Figura 5.5. Milk - Decomposição em valores singulares.



Fonte: Autor

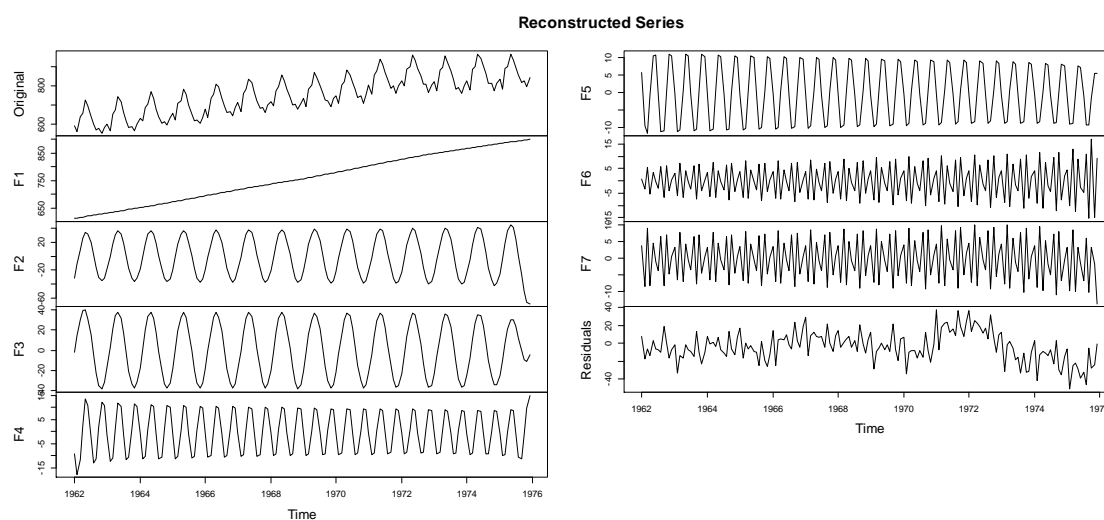
Figura 5.6. Pigs – Decomposição em valores singulares.



Fonte: Autor

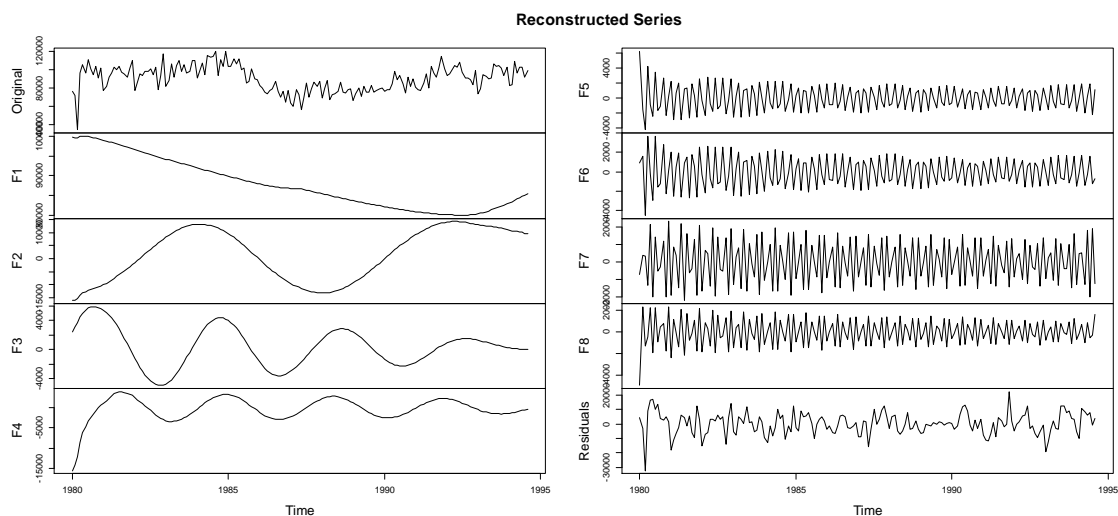
O passo seguinte foi a análise das matrizes correlação de modo a estudar as componentes a reter para cada série temporal. Assim para os dados Milk foram retidas sete componentes (F1 - F7) e para os dados Pigs retidas nove componentes (F1 – F9), tendo-se obtido a reconstrução das séries, como mostra as figuras 5.7. e 5.8.

Figura 5.7. Reconstrução da Série Temporal Milk.



Fonte: Autor

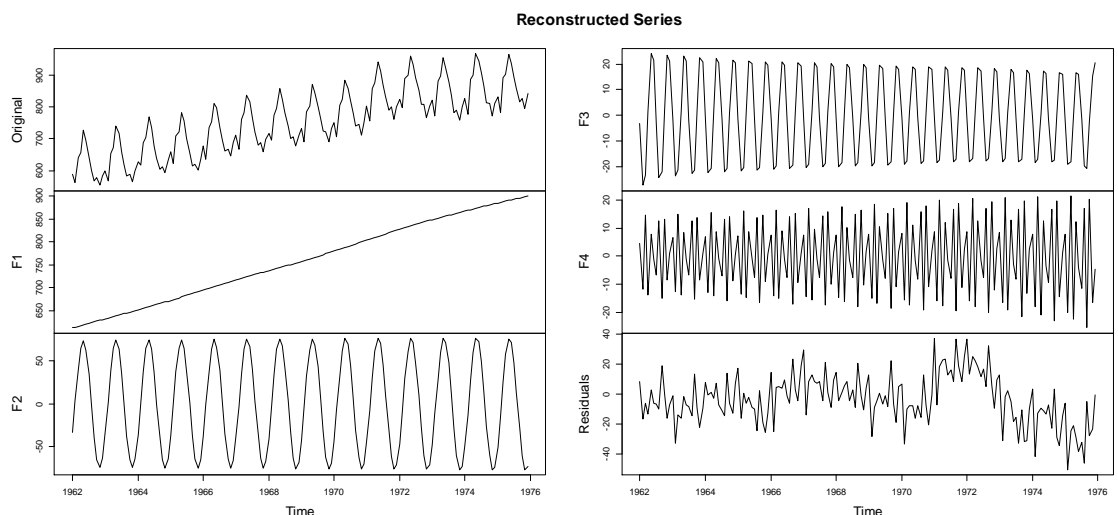
Figura 5.8. Reconstrução da série temporal Pigs.



Fonte: Autor

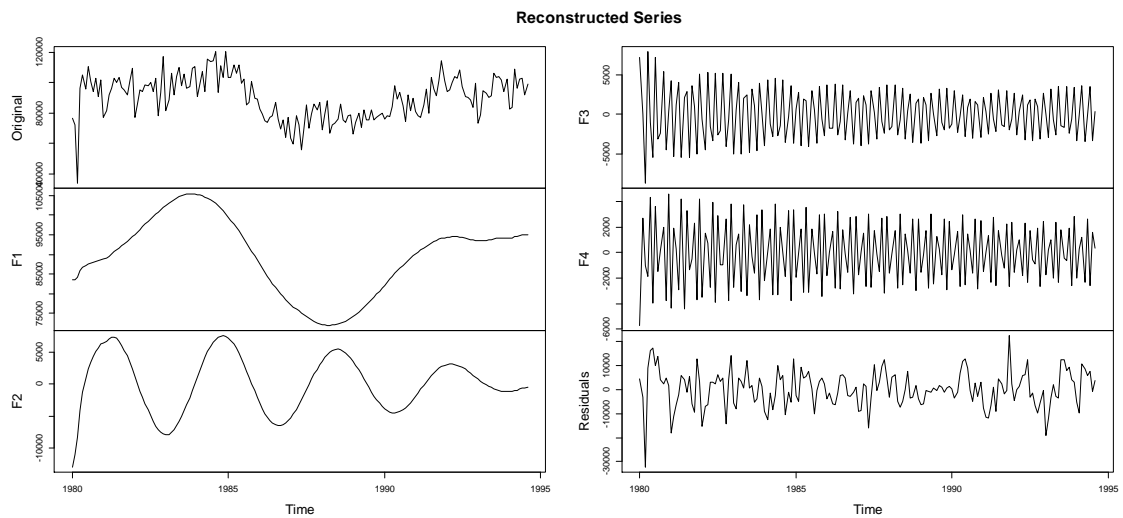
As componentes também foram agrupadas de acordo com as que tinham sazonalidades que apresentavam durações semelhantes, obtendo-se as reconstruções como mostram as figuras 5.9. e 5.10.

Figura 5.9. Reconstrução com sazonalidade da série temporal Milk.



Fonte: Autor

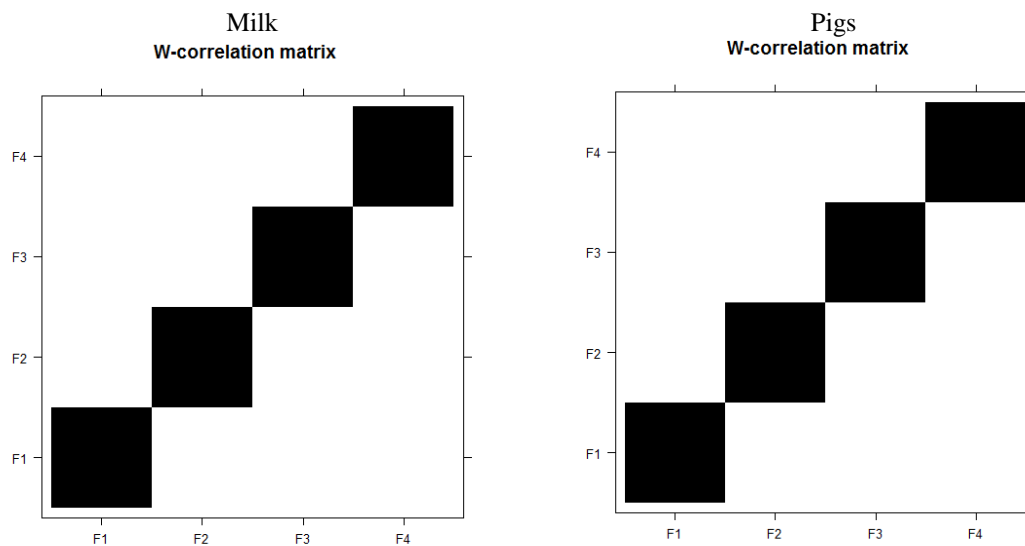
Figura 5.10. - Reconstrução com sazonalidade da série temporal Pigs.



Fonte: Autor

A figura 5.11 mostra a correlação ponderada para os agrupamentos finais, para cada série analisada.

Figura 5.11. Correlação ponderada para os agrupamentos finais.



Fonte: Autor

Utilizando o mesmo software, foi feita uma previsão para os últimos 12 meses de cada uma das séries temporais. Contudo estas previsões serão tradas no capítulo 5.4. em conjunto com as previsões com os modelos ARIMA.

5.4 Previsões

Nos capítulos anteriores, foram escolhidos os modelos ARIMA e AES que melhor se ajustam aos dados em estudo.

Será necessário agora fazer as previsões utilizando todos os modelos e criar condições que permitam fazer uma análise comparativa.

Renata Esquivel (2012) refere que, para os casos em que os valores futuros não são conhecidos, “*costuma-se avaliar o melhor método de previsão ao retirar as n últimas observações da Série Temporal e compará-las, através de uma medida de erro, com os correspondentes valores preditos a partir do valor estimado*”.

Não existindo uma recolha de dados posterior, que nos permita fazer uma previsão e consequente comparação, a opção neste trabalho foi utilizar as potencialidades do *Gretl* e para cada série criar uma nova amostra em que foi retirado os últimos 12 dados, servindo estes valores retirados, como elementos de comparação.

Assim para os dados Milk a amostra passou a estar compreendida entre de 1962:01 e 1974:12, enquanto para os dados Pigs a amostra ficou compreendida entre 1980:01 a 1994:08. Os dados retirados n igual a 12 (meses), foram utilizados como comparação para previsão.

Com estas amostras aplicou-se o modelo ARIMA, que mostrou ser o melhor ajuste e foi feita a respectiva previsão para mais n dados.

A Tabela 5.9. mostra a previsão feita para os meses de Janeiro de 1995 até Dezembro de 1995 e os respectivos intervalos de confiança. Pode-se verificar que todas as previsões se encontram dentro desses intervalos.

Tabela 5.11. – Milk – Previsão e intervalos de confiança (ARIMA).

Observações	V. Original	V. Previsão	erro padrão	intervalo a 95%
1975:01	834,000	836,355	7,21829	822,208 - 850,503
1975:02	782,000	793,385	9,09964	775,550 - 811,220
1975:03	892,000	896,577	10,5876	875,826 - 917,329
1975:04	903,000	911,995	12,1987	888,086 - 935,904
1975:05	966,000	973,683	13,3112	947,593 - 999,772
1975:06	937,000	947,861	14,6042	919,237 - 976,484
1975:07	896,000	905,331	15,5751	874,804 - 935,858
1975:08	858,000	865,181	16,6625	832,524 - 897,839
1975:09	817,000	815,102	17,5442	780,716 - 849,488
1975:10	827,000	817,671	18,4946	781,423 - 853,920
1975:11	797,000	780,409	19,3092	742,563 - 818,254
1975:12	843,000	820,632	20,1627	781,114 - 860,150

Fonte: Autor

A tabela 5.12. apresenta os resultados obtidos pelo *Gretl* da previsão para o intervalo compreendido entre Setembro de 1994 e Agosto de 1995, último valor conhecido dos dados Pigs, assim como os respectivos intervalos de confiança. Para estes dados, todas as previsões também se encontram dentro dos intervalos de confiança.

Tabela 5.12. – Pigs – Previsão e intervalos de confiança (ARIMA).

Observações	Pigs	Previsão	erro padrão	intervalo a 95%
1994:09	110.067,00	96.663,51	8.950,00	79.121,83 - 114.205,19
1994:10	101.599,00	91.403,35	9.632,41	72.524,17 - 110.282,53
1994:11	97.646,00	94.892,68	10.005,12	75.283,01 - 114.502,35
1994:12	104.930,00	100.110,92	12.077,63	76.439,20 - 123.782,65
1995:01	88.905,00	80.426,62	12.656,22	55.620,90 - 105.232,35
1995:02	89.936,00	85.934,88	13.231,65	60.001,33 - 111.868,43
1995:03	106.723,00	95.829,77	14.379,87	67.645,19 - 124.013,22
1995:04	84.307,00	90.981,77	14.939,98	616.99,94 - 120.263,60
1995:05	114.896,00	97.581,59	15.553,57	67.097,14 - 128.066,04
1995:06	106.749,00	95.271,43	16.356,51	63.213,27 - 12.7329,60
1995:07	87.892,00	96.021,41	16.901,44	62.895,21 - 129.147,62
1995:08	100.506,00	95.471,50	17.490,02	61.191,70 - 129.751,30

Fonte: Autor

Quer pelas tabelas de resultados como pelos gráficos é fácil verificar que todos os valores previstos se encontram dentro dos intervalos de confiança, pelo que temos uma taxa de cobertura de 100%.

A última fase da análise espectral singular é a previsão. Para poder comparar estas previsões com as previsões feitas com o modelos ARIMA, foram executadas 12 previsões. Este intervalo corresponde exactamente ao intervalo estudado para os modelos ARIMA, para cada série temporal.

Os valores das previsões obtidas pelo R e os respectivos intervalos de confiança, são apresentados nas tabelas 5.13. e 5.14.

Tabela 5.13. – Milk – Previsão e intervalos de confiança (AES).

Milk - AES	V. Original	V. Previsto	intervalo a 95%
1975:01	834,00	857,7002	845,1731 - 864,2349
1975:02	782,00	856,7572	843,1549 - 871,1940
1975:03	892,00	945,0376	936,8910 - 952,6171
1975:04	903,00	949,6276	940,6024 - 957,8707
1975:05	966,00	1.020,7950	1.010,1514 – 1.032,4916
1975:06	937,00	997,5950	987,6564 – 1.009,0082
1975:07	896,00	957,4729	944,1794 - 966,5169
1975:08	858,00	937,9110	918,2905 - 950,5440
1975:09	817,00	858,8647	849,9378 - 898,6826
1975:10	827,00	889,7951	876,0217 - 898,6826
1975:11	797,00	846,8274	835,5440 - 861,0298
1975:12	843,00	869,2761	858,3710 - 878,4215

Fonte: Autor

Tabela 5.14. – Pigs – Previsão e intervalos de confiança (AES).

Pigs - AES	V. Original	V. Previsto	intervalo a 95%
1994:09	110.067,00	90.927,71	84.196,78 - 94.598,63
1994:10	101.599,00	90.951,81	84.989,98 - 99.135,41
1994:11	97.646,00	91.581,37	88.402,17 - 98.807,06
1994:12	104.930,00	93.045,57	86.038 - 97.971,60
1995:01	88.905,00	88.153,65	83.207,40 - 93.014,74
1995:02	89.936,00	93.622,46	88.465,30 - 98.666,37
1995:03	106.723,00	90.760,66	83.027,01 - 97.994,40
1995:04	84.307,00	87.120,46	81.090,20 - 93.570,85
1995:05	114.896,00	95.983,09	88.620,61 - 102.505,53
1995:06	106.749,00	86.123,78	78.578,49 - 91.146,72
1995:07	87.892,00	89.713,58	82.436,98 - 97.463,71
1995:08	100.506,00	93.265,18	84.457,09 - 100.591,31

Fonte: Autor

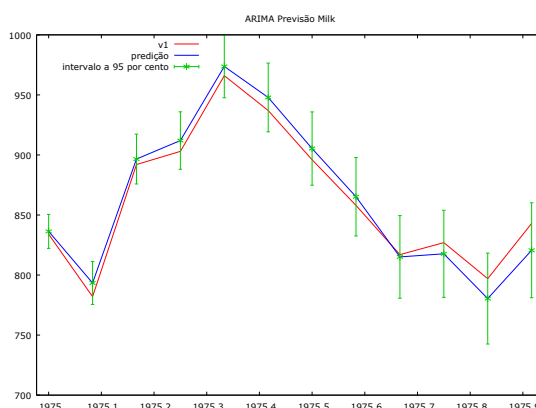
Também para as duas séries temporais, nestes modelos obteve-se uma taxa de cobertura de 100%, dado que todos os valores de previsão se encontram dentro dos intervalos de confiança.

Para todos os valores de previsão, quer para os modelos ARIMA como os modelos AES, também foi calculado a raiz do erro quadrado médio, para cada uma das série temporais, que se apresenta nas tabelas D1 a D4 do anexo D.

Apesar de a comparação os modelos ARIMA e AES ser feita no próximo capítulo, será interessante mostrar os gráficos correspondentes aos dados das tabelas anteriores. Poder-se-á assim visualizar os valores reais, os valores previstos e os respectivos intervalos de confiança.

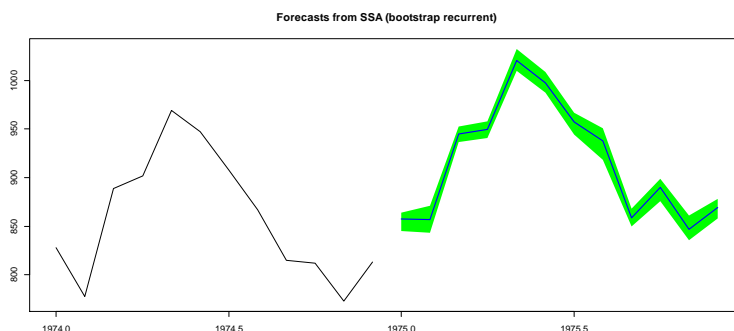
Para os dados Milk, pode verificar-se que os gráficos são muito idênticos, quer para os valores revistos quer para os intervalos de confiança.

Figura 5.12. – Gráfico da previsão e intervalos de confiança modelo ARIMA (Milk).



Fonte: Autor

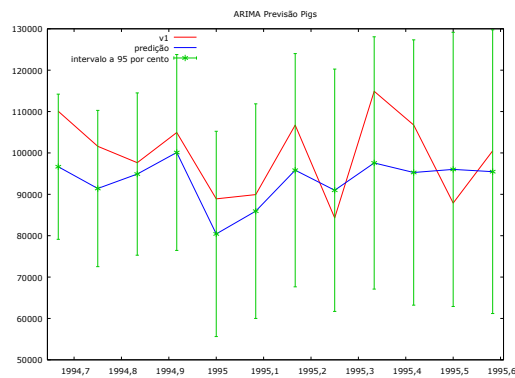
Figura 5.13. – Gráfico da previsão e intervalos de confiança modelo AES (Milk).



Fonte: Autor

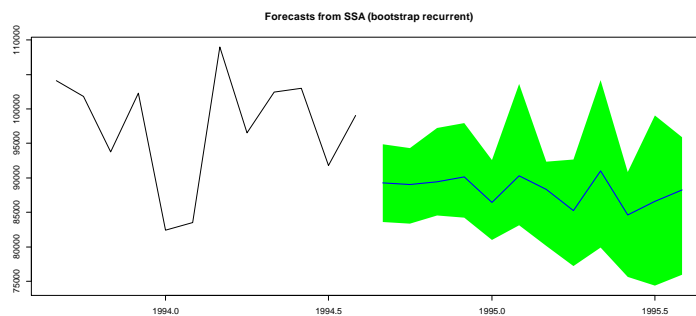
Para a série temporal Pigs, pode-se verificar uma maior diferença nos gráficos, quer na curva da previsão, quer como nos intervalos de confiança.

Figura 5.14. – Gráfico da previsão e intervalos de confiança modelo ARIMA (Pigs).



Fonte: Autor

Figura 5.15. – Gráfico da previsão e intervalos de confiança modelo AES (Pigs).



Fonte: Autor

5.5 Comparação dos modelos

Com as análises e cálculos executados anteriormente, podemos agora fazer a comparação dos modelos, qual será a técnica mais eficaz, modelos ARIMA ou modelos AES?

Nesse sentido vamos comparar a raiz do erro quadrado médio (RMSE em inglês) de todas as previsões. Estes erros como referido anteriormente podem ser consultados no Anexo D nas tabelas D1 a D4.

Na tabela 5.15. é feita a comparação dos erros quadrados médios para a série temporal Milk, com as previsões para os últimos 12 meses dos dados recolhidos. O modelo ARIMA (2,1,1)(1,11) foi aquele em que o valor RMSE foi o mais baixo. O modelo AES apresentou um valor bastante mais elevado, pelo que se pode concluir que não será um bom modelo para representar os dados desta série.

Tabela 5.15. Milk – Raiz do erro quadrado médio.

Modelo	RMSE
ARIMA	10,77
AES	55,41

Fonte: Autor

A Tabela 5.16 apresenta a comparação dos RMSE das previsões para a série temporal Pigs. O modelo ARIMA (2,1,1)(1,1,1) apresenta o valor mais baixo. Contudo para este conjunto de dados a diferença entre o modelo ARIMA e AES não é tão significativo que na série Milk.

Tabela 5.16. Pigs – Raiz do erro quadrado médio

Modelo	RMSE
ARIMA	9.525,97
AES	11.156,64

Fonte: Autor

Quanto à percentagem de cobertura de cada modelo, ela é de 100 % pelo que não permite fazer nenhuma distinção quanto ao melhor modelo.

Pela análise efectuada, podemos concluir com segurança, que para estas séries temporais, os modelos ARIMA são aqueles que melhor se ajustam aos dados estudados.

No seu trabalho, Renata Esquível (2012) refere nas conclusões que “*o algoritmo recorrente de previsão SSA consegue representar melhor as variações existentes nos dados, como flutuações sazonais e picos, características tradicionais encontradas na prática isto é, em muitas séries temporais.*”. Neste trabalho concluímos exactamente o contrário. Sendo que na série temporal em que as flutuações sazonais e os picos são mais acentuados (Milk), o modelo ARIMA fez melhores previsões.

Poderá colocar-se desde já duas questões:

- Os modelos ARIMA representam melhor a totalidade das séries temporais?
- Os modelos ARIMA representam melhor as séries temporais em que o seu comportamento quanto à monotonia e sazonalidade estão bem definidos?

6. Considerações finais.

O presente trabalho científico teve como principal objectivo avaliar e comparar o processo de modelação e a qualidade de previsão do modelo de Box-Jankins ARIMA e da técnica AES, com aplicação a duas séries temporais reais.

No capítulo cinco procurou-se qual o melhor modelo ARIMA que se ajustasse aos valores reais de cada uma das séries. A escolha para ambas as séries, recaiu sobre o modelo ARIMA (2,1,1)(1,1,1), ou seja um caso particular também denominado modelo SARIMA.

Tendo escolhido o software *Gretl*, utilizou-se os testes disponíveis para escolher o melhor modelo, o teste de Ljung e os testes de informação AIC, BIC e HQC, para a escolha do modelo com melhor ajuste de previsão.

Neste programa, foram analisados vários modelos ARIMA, tendo-se alterado sucessivamente um valor das características do modelo, de modo a verificar como se comportavam os resultados dos testes de Ljung-Box.

Neste capítulo foi ainda escolhido os modelos AES, que melhor representam as séries temporais.

Pelos estudos efectuados às duas séries temporais, verificou-se que a metodologia de Box e Jenkins, que neste trabalho foi referenciada como modelos ARIMA, através do modelo SARIMA, apresentou melhor capacidade de previsão relativamente aos modelos AES.

Neste trabalho constatou-se no entanto que, para a série temporal em que a componente sazonal era bem definida (MILK), o valor da RMSE obtido para o modelo AES foi cinco vezes superior ao modelo ARIMA. Na série temporal Pigs a componente sazonal não está tão bem definida e o valor da RMSE obtido pelo modelo AES é só duas vezes superior ao modelo ARIMA.

Poderemos assim questionar, se os modelos ARIMA serão mais indicados, para prever valores futuros de séries temporais com sazonalidade bem definida, em detrimento da AES.

Considerações finais

Há que concordar com Renata Esquível (2012), no sentido em que a simplificação nos entendimentos das séries temporais na técnica, vem traduzida numa simplificação nos cálculos, uma vez que diminui a intervenção do analista.

O presente trabalho almeja constituir um auxiliar de informação nas abordagens aos modelos ARIMA e AES, tanto para os principiantes como para investigadores menos versados nas lides computacionais, já que mostramos em detalhe o recurso ao R e ao *Gretl* para a parte computacional. Será certamente um contributo valioso na comparação destas duas importantes técnicas de previsão, sendo por isso uma mais valia para análise e previsão de séries temporais.

Nesse sentido ao testar os vários modelos ARIMA e dada a facilidade demonstrada com a utilização do *Gretl*, apresentam-se em anexo os resultados mais interessantes dos testes efectuados. Será possível assim analisar o comportamento dos modelos consoante várias alterações que podem vir a ser introduzidas, permitindo a simulação e análise de modelos estocásticos.

Nota: Neste trabalho, por opção do autor, não foi utilizado o novo Acordo Ortográfico

7. Bibliografia.

BARROS, Mónica, Processos Estocásticos, 202-229. Site http://www.mbarros.com/documentos/upload/Livro_Processos_Cap_5_parte.pdf consultado Novembro 2015.

BELL, W. R., and HILLMER, S. C. (1984). Issues involved with the seasonal adjustment of time series. J. Bus. Econ.

BERTOLO, Luíz António, Técnicas de Previsão de Box-Jenkins no Excel. Site: <http://www.bertolo.pro.br/MetodosQuantitativos/Simulacao/TecnicasPrevisaoBoxJenkinsNoExcel.pdf> consultado em Dezembro 2015

BOX, G.E., JENKINS, G.M. and REINSEL, G. C. (1994). Time Series: Forecasting and Control (3rd edition). Prentice Hall.

BOZDONGAN. H. (1987). Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. **Psychometrika**. v.52

BROCKWELL, P.J. and DAVIS, R.A. (1991). Time Series. Theory and Methods (2nd edition). Springer.

BROCKWELL, P. J. and DAVIS, R. A. (2002). Introduction to Time Series and Forecasting, 2nd. ed., Springer-Verlag

CHAFTFIELD, Cris, (2000). Time Series Forecasting, Chapman & Hall/CRC.

CHATFIELD, C. (2004): The Analysis of Time Series: An Introduction (6th Edition). Chapman and Hall.

Bibliografia

CHUJAI, P., KERDPRASOP, N. and KERDPRASOP, K.(2013). Time Series Analysis of Household Electric Consumption with ARIMA and ARMA Models. Proceedings of the International MultiConference of Engineers and Computer Scientists. Vol I, IMECS 2013, March 13 - 15, Hong Kong.

DANILOV, D.(1997). Principal components in time series forecast. Journal of computational and graphical statistics

DAVILA, Vitor Hugo Lacho, Introdução às séries temporais. Site: <http://www.ime.unicamp.br/~hlauchos/MaterialSeries.pdf> consultado Novembro 2015.

DOUGLAS C. Montgomery, CHERYL L. Jennings, MURAT Kulahci (2015). Introduction to Time Series Analysis and Forecasting. John Wiley & Sons.

ESQUÍVEL, R. M. (2012). Análise Espectral Singular: Modelação de séries temporais através de estudos comparativos usando diferentes estratégias de previsão. SENAI CIMATEC

FILIFE, José dos Santos (2012). Aplicação de modelos ARIMA em série de preços de soja no norte do Paraná. Site: <http://www.fatecbt.edu.br/seer/index.php/tl/article/viewFile/150/151> consultado Outubro de 2015

GARDNER, W.H. (1988). - Statistical spectral analysis - A nonprobabilistic theory. Prentice-Hall Inc. Englewood Cliffs, NJ, USA.

GEORGE E. P. Box, GWILYM M. Jenkins, GREGORY C. Reinsel, GRETA M. Ljung (2015). Time Series Analysis: Forecasting and Control. John Wiley & Sons.

GOLYANDINA, Nina, NEKRUTKIN, Vladimir, e ZHINGLJAVSKY, Anatoly (2001). Analysis of Time Series Structure – SSA and Related Techniques.

Bibliografia

HAMILTON, J.D., (1994), *Time Series Analysis*, Princeton: Princeton University Press.

HASSANI, Hossein, *A Brief Introduction to Singular Spectrum Analysis*. (2014). Artigo publicado por *ResearchGate* site: https://www.researchgate.net/profile/Hossein_Hassani7/publication/267723014_A_Brief_Introduction_to_Singular_Spectrum_Analysis/links/548036910cf250f1edbf826.pdf consultado em Outubro 2015.

HYNDMAN, R. J. (2014). *Forecast: Forecasting Functions for Time Series and Linear Models*. R Package Version 5.4. Available online at: <http://CRAN.R-project.org/package=forecast>

HYNDMAN, R. J., and ATHANASOPOULOS, G. (2014). *Forecasting: Principles and Practice*. OTexts. Available online at: <http://otexts.org/fpp/>

HYNDMAN, R. J., and KHANDAKAR, Y. (2008). *Automatic time series forecasting: the forecast package for R*. J. Stat. Softw.

JONATHAN, D. Cryer and CHAN, Kung-Sik (2008). *Time Series Analysis: With Applications in R*. Springer Science & Business Media.

KLEIN, J.L.(1997). *Statistical visions in time: a history of time series analysis, 1662-1938*. Cambridge: Cambridge University Press

LITTELLI, R. C.; Milliken, G. A. Stroup, W. W & Wolfinger, R. D. (2002). *SAS System for Mixed Models*. Cary: Statistical Analysis System Institute.

MADSEN, Henrik (2007). *Time Series Analysis*. CRC Press

MITCHELL, T. R., and JAMES, L. R. (2001). *Building better theory: time and the specification of when things happen*.

MORETTIN, Pedro A. e Toloi, Clécia M.C. Instituto de Matemática e Estatística da Universidade de São Paulo (Janeiro 2006). Site:

Bibliografia

https://issuu.com/editorablucher/docs/issuu_analise_temporais_isbn9788521203896/5?e=1099747/5170484 consultado Dezembro 2015.

MORETTIN, Pedro A. e TOLOI, Clélia M (1981). Modelos para Previsão de Séries Temporais, Edgard Blucher,.

MYUNG, N.K.(2009). Singular Spectrum Analysis. Thesis submitted for the degree Master of Science in Statistics, University of California, Los Angeles.

NAIDU, P.S. (1996). Modern spectrum analysis of time series. CRC Press Inc., Boca Raton, FL, USA

NETTO, Maria Aparecida Cavalcanti, A previsão com a Metodologia de Box-Jenkins, publicado pela Universidade Federal do Rio de Janeiro no site: <http://www.ie.ufrj.br/download/APrevisaoComMetodologiadeBox-Jenkins.pdf> consultado em Outubro 2015.

PORTAL ACTION, Séries Temporais. Site: <http://www.portalaction.com.br/series-temporais/modelos-arima> consultado em Outubro 2015.

PREZ, M. (2014). Time series analysis with matlab: arima and arimax models, createspace independent pub

QUEIROZ, Vívian dos Santos, Estimação de Modelos ARMA e ARIMA. Site: <http://www.ppge.ufrgs.br/sabino/eco02007/ARMA-ARIMA.pdf> consultado em Novembro 2015.

R Development Core Team. (2011). *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. Available online at: <http://www.R-project.org/>

Rodrigues, Paulo Canas e Carvalho, Miguel. Método de Imputação Recorrente: Análise Espectral Singular com Valores Omissos. (2009)

Bibilografia

SHUMWAY, R.H., STOFFER, D.S. (2010). Time Series Analysis and Its Applications: With R Examples. Springer Science & Business Media.

THIRION, Michel e COLLIS, Robert, ARIMA: Como evitar a mentalidade de rebanho ao analisar os dados de séries temporais. Artigo publicado por Minitab no site <http://www.minitab.com/pt-BR/Published-Articles/ARIMA--Como-evitar-a-mentalidade-de-rebanho-ao-analisar-os-dados-de-s%C3%A9ries-temporais/> consultado Março de 2016.

TONG, Howell (2011) Rejoinder from Howell Tong to the discussions on 'threshold models in time series analysis - 30 years on'. Statistics and Its Interface.

TONG, Howell (1999) Some recent nonparametric tools for time series data analysis. Bulletin of the International Statistical Institute, 53rd s (Tome 5).

WARNER, R. M. (1998). Spectral analysis of time-series data. Guilford Press, New York, NY, USA.

WAYNE, A. Woodward, HENRY L. Gray, ALAN C. Elliot (2011). Applied Time Series Analysis. CRC Press.

WHEELWRIGHT, S.C., MAKRIDAKIS, S. e McGee, V.E. (1983). Forecasting: Methods and Applications, 2nd Ed. New York: John Wiley and Sons

WOLFINGER, R. D. (1993).Covariance estrutura selection in general mixed models. Communications in Statistics. V.22.

ZOU H. & YANG, Y. (2004). Combining time series models for forecasting. International Journal of forecasting.

Anexo A - Resultados dos testes de estacionaridade

Anexos

O software *Gretl* tem à disposição os seguintes testes da raiz unitária, teste de Dickey-Fuller aumentado, teste ADF-GLS e por fim o teste KPSS.

Para o estudo de cada uma das Séries Temporais foram realizados os três testes.

Nas tabelas A1 e A2 apresentam-se uma síntese dos resultados do teste aumentado de Dickey-Fuller, com constante sem e com tendência variáveis auxiliares de sazonalidade aditiva, para as duas Séries Temporais. Neste teste foi utilizado o critério AIC.

Tabela A.1. - Resultados Teste Aumentado de Dickey-Fuller – Milk

Teste Aumentado de Dickey-Fuller - Milk		
Teste com constante variáveis auxiliares de sazonalidade aditiva		
	sem tendência	com tendência
modelo:	$(1-L)y = b_0 + (a-1)*y(-1) + \dots + e$	$(1-L)y = b_0 + b_1*t + (a-1)*y(-1) + \dots + e$
valor estimado de (a - 1):	-0,00784784	-0,107054
estatística de teste:	$\tau_c(1) = -1,1156$	$\tau_{ct}(1) = -2,13049$
valor p assintótico	0,712	0,5281
coef. de 1ª-ordem para e:	-0,003	-0,007
diferenças desfasadas:	$F(6, 142) = 3,586 [0,0024]$	$F(13, 127) = 2,476 [0,0049]$

Fonte: autor

Tabela A.2. - Resultados Teste Aumentado de Dickey-Fuller – Pigs

Teste Aumentado de Dickey-Fuller - Pigs		
Teste com constante variáveis auxiliares de sazonalidade aditiva		
	sem tendência	com tendência
modelo:	$(1-L)y = b_0 + (a-1)*y(-1) + \dots + e$	$(1-L)y = b_0 + b_1*t + (a-1)*y(-1) + \dots + e$
valor estimado de (a - 1):	-0,0897483	-0,0885113
estatística de teste:	$\tau_c(1) = -1,71073$	$\tau_{ct}(1) = -1,66851$
valor p assintótico	0,4258	0,7653
coef. de 1ª-ordem para e:	0,003	0,003
diferenças desfasadas:	$F(10, 154) = 8,663 [0,0000]$	$F(10, 153) = 8,605 [0,0000]$

Fonte: autor

Anexos

As tabelas A3 e A4 apresentam uma descrição do teste aumentado de Dickey-Fuller (GLS), realizado a cada uma das duas Séries. Foram incluídos 11 desfasamentos e utilizados os critérios AIC modificado e Perron-Qu modificado. Hipótese nula de raiz unitária: $a = 1$.

Tabela A.3. - Resultados Teste Aumentado de Dickey-Fuller (GLS) – Milk

Teste Aumentado de Dickey-Fuller (GLS) - Milk				
Teste com constante e tendência variáveis auxiliares de sazonalidade aditiva				
modelo:	$(1-L)y = b_0 + b_1*t + (a-1)*y(-1) + \dots + e$			
valor estimado de (a - 1):	-0,00334388			
estatística de teste:	tau = -0,0434797			
valores críticos	10%	5%	2,5%	1%
valor p assintótico	-2,64	-2,93	-3,18	-3,46
coef. de 1ª-ordem para e:	0,664			
diferenças desfasadas:	F(11, 144) = 84,711 [0,0000]			

Fonte: autor

Tabela A.4. - Resultados do teste Aumentado de Dickey-Fuller (GLS) – Pigs

Teste Aumentado de Dickey-Fuller (GLS) - Pigs				
Teste com constante e tendência variáveis auxiliares de sazonalidade aditiva				
modelo:	$(1-L)y = b_0 + b_1*t + (a-1)*y(-1) + \dots + e$			
valor estimado de (a - 1):	-0,0733287			
estatística de teste:	tau = -1,46203			
valores críticos	10%	5%	2,5%	1%
valor p assintótico	-2,64	-2,93	-3,18	-3,46
coef. de 1ª-ordem para e:	-0,013			
diferenças desfasadas:	F(12, 162) = 10,918 [0,0000]			

A tabela A5 apresenta a descrição dos resultados do teste KPSS, onde a ordem de defasamento para o teste foi 4 e inclui-se a tendência e auxiliares de sazonalidade

Tabela A.5. - Resultados do teste KPSS (incluindo tendência e sazonais)

	Milk		Pigs
Estatística do teste	0,353462		0,469705
Valores referência	10%	5%	1%
Valores críticos	0,120	0,148	0,216
Valor de p	< .01		< .01

Anexo B - Resultados dos modelos ARIMA

Com a finalidade de fornecer ao leitor uma informação rápida do estudo completo dos modelos ARIMA escolhidos para cada Série Temporal, apresenta-se neste anexo toda a análise ao modelo ARIMA (2,1,1)(1,1,1) quer para Milk e como para Pigs.

- .Análises ao modelo ARIMA
- Resultados do teste para autocorrelação
- Resultados da análise dos resíduos
- Gráfico dos resíduos
- . Correlograma dos resíduos
- Periodograma dos resíduos
- Gráfico Quanti-Quanti dos resíduos
- Resultados da análise da predição

Tabela B. 1. Análises ao modelo ARIMA (2,1,1)(1,1,1) - Milk

ARIMA (2,1,1)(1,1,1), usando as observações 1963:02-1975:12 (T = 155)					
Variável dependente: (1-L)(1-Ls) Milk					
Erros padrão baseados na Hessiana					
	Coefficiente	Erro Padrão	z	valor p	
const	-0,0202106	0,208052	-0,0971	0,9226	
phi_1	-1,15886	0,0919283	-12,6061	<0,0001	***
phi_2	-0,26424	0,0779521	-3,3898	0,0007	***
Phi_1	-0,0574459	0,120487	-0,4768	0,6335	
theta_1	0,943855	0,0599505	15,7439	<0,0001	***
Theta_1	-0,581981	0,0987288	-5,8947	<0,0001	***

Média var. dependente	0,122581	D.P. var. dependente	8,940159
Média de inovações	0,127307	D.P. das inovações	7,153559
Log. da verosimilhança	-528,2349	Critério de Akaike	1070,470
Critério de Schwarz	1091,774	Critério Hannan-Quinn	1079,123

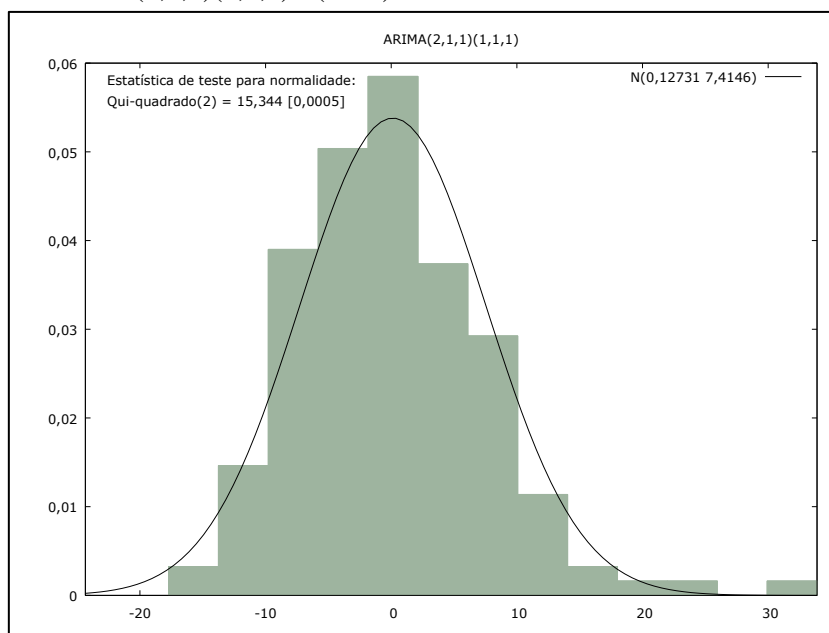
		Real	Imaginária	Módulo	Frequência
AR	Raiz 1	-1,1809	0,0000	1,1809	0,5000
	Raiz 2	-3,2047	0,0000	3,2047	0,5000
AR (sazonal)	Raiz 1	-17,4077	0,0000	17,4077	0,5000
MA	Raiz 1	-1,0595	0,0000	1,0595	0,5000
MA (sazonal)	Raiz 1	1,7183	0,0000	1,7183	0,0000

Tabela B. 2. Resultados do teste para autocorrelação - ARIMA (2,1,1)(1,1,1) - Milk

ordem	Ljung-Box Q'	valor p
12	6,87398	P(Qui-quadrado(7) > 6,87398) = 0,4421
24	13,9492,	P(Qui-quadrado(19) > 13,9492) = 0,7866
36	17,2093	P(Qui-quadrado(31) > 17,2093) = 0,9784

Fonte: autor

Figura B 3. Gráfico da análise da normalidade dos resíduos ARIMA (2,1,1)(1,1,1) - (Milk)



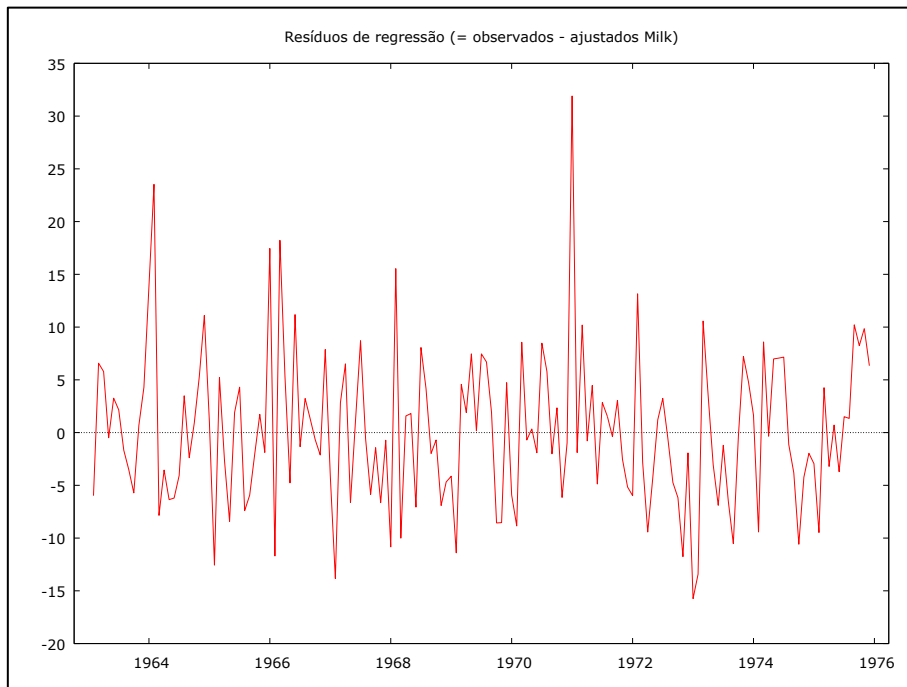
Fonte: autor

Tabela B 4. Resultados da análise dos resíduos - ARIMA (2,1,1)(1,1,1) - Milk

intervalo	ponto médio	freqüência	rel.	rel.	acum.
< -13,753	-15,738	2	1,29%	1,29%	
-13,753 - -9,7825	-11,768	9	5,81%	7,10%	**
-9,7825 - -5,8124	-7,7975	24	15,48%	22,58%	*****
-5,8124 - -1,8424	-3,8274	31	20,00%	42,58%	*****
-1,8424 - 2,1277	0,14267	36	23,23%	65,81%	*****
2,1277 - 6,0978	4,1127	23	14,84%	80,65%	*****
6,0978 - 10,068	8,0828	18	11,61%	92,26%	****
10,068 - 14,038	12,053	7	4,52%	96,77%	*
14,038 - 18,008	16,023	2	1,29%	98,06%	
18,008 - 21,978	19,993	1	0,65%	98,71%	
21,978 - 25,948	23,963	1	0,65%	99,35%	
25,948 - 29,918	27,933	0	0,00%	99,35%	
>= 29,918	31,903	1	0,65%	100,00%	

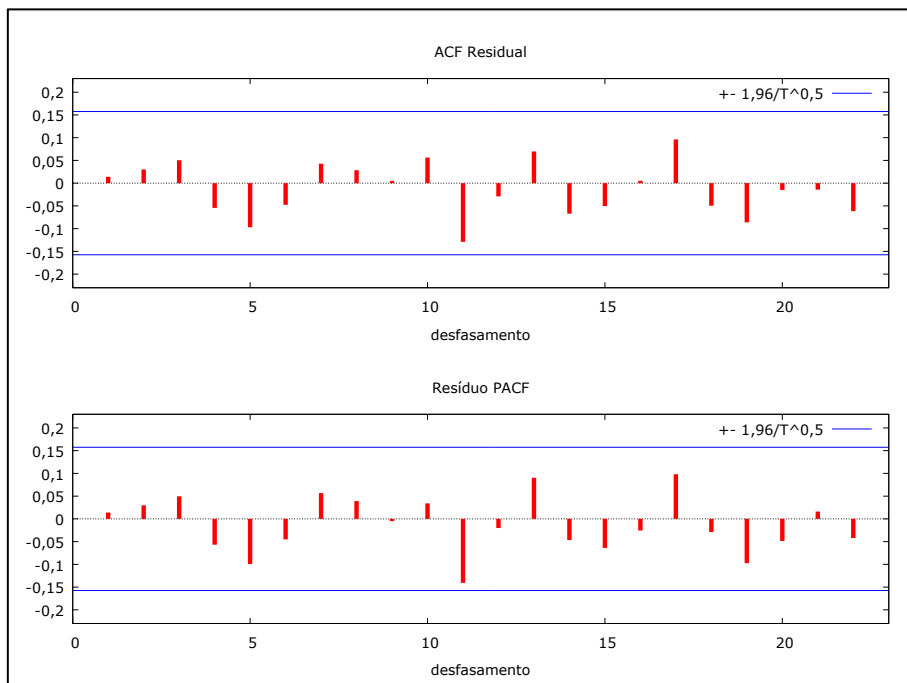
Fonte: autor

Figura B 5. Gráfico dos resíduos - ARIMA (2,1,1)(1,1,1) - Milk



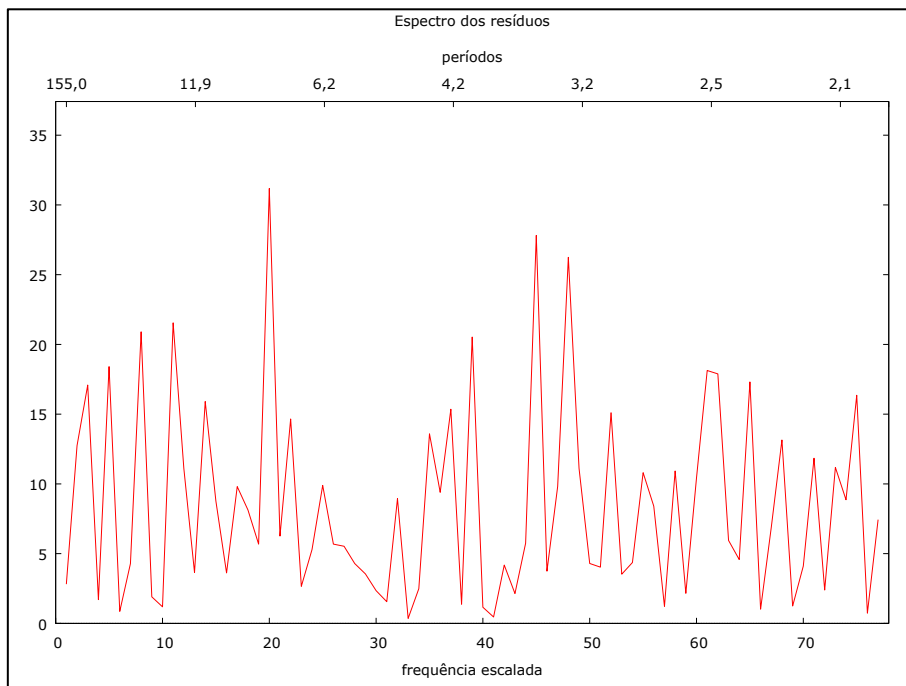
Fonte: autor

Figura B.6. Correlograma dos resíduos - ARIMA (2,1,1)(1,1,1) - Milk



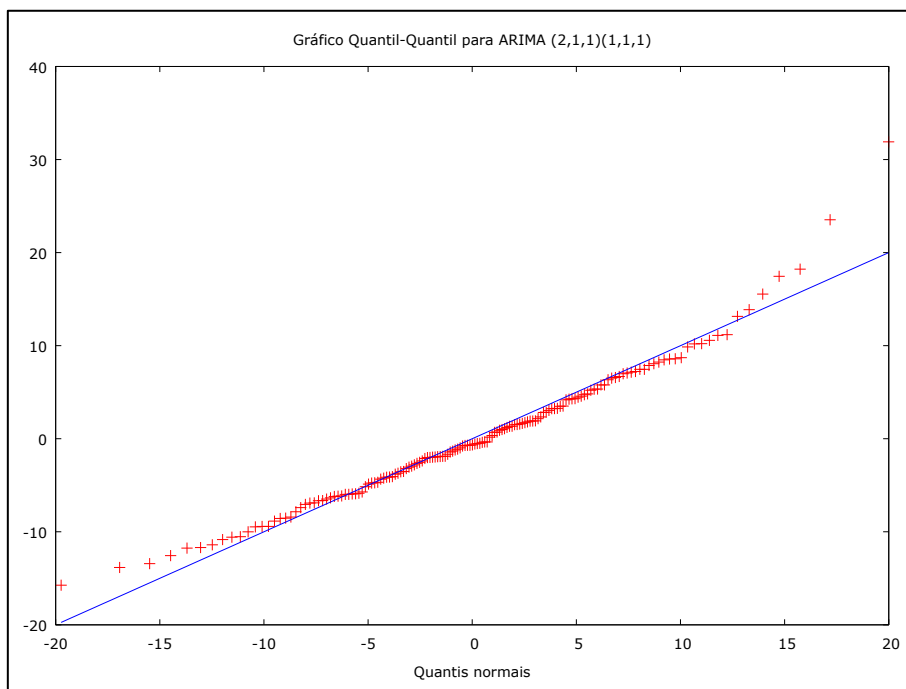
Fonte: autor

Figura B.7. Periodograma dos resíduos - modelo ARIMA (2,1,1)(1,1,1) - Milk



Fonte: autor

Figura B.8. Gráfico Quanti-Quanti dos resíduos - modelo ARIMA (2,1,1)(1,1,1) - Milk



Fonte: autor

Tabela B.9. Resultados da análise da predição - ARIMA
(2,1,1)(1,1,1) – Milk

Estatísticas da avaliação da Predição	
Erro Médio	0,12731
Erro Médio Quadrado	52,865
Erro Unitário Médio Quadrado	7,2708
Erro Médio Absoluto	5,5756
Erro Médio Percentual	0,0090156
Erro Médio Percentual Absoluto	0,74548
U de Theil	0,1628
Proporção do enviesamento, UM	0,00030658
Proporção da regressão, UR	7,6209e-005
Proporção da perturbação, UD	0,99962

Fonte: autor

Tabela B. 10. Análises ao modelo ARIMA (2,1,1)(1,1,1) - Pigs

ARIMA (2,1,1)(1,1,1), usando as observações 1981:02-1995:08 (T = 175)
 Variável dependente: (1-L)(1-Ls) Pigs
 Erros padrão baseados na Hessiana

	Coefficiente	Erro Padrão	z	valor p	
const	-42,508	102,08	-0,4164	0,6771	
phi_1	-0,825035	0,130455	-6,3243	<0,0001	***
phi_2	-0,606573	0,0848403	-7,1496	<0,0001	***
Phi_1	0,164655	0,121871	1,3511	0,1767	
theta_1	0,203233	0,153714	1,3222	0,1861	
Theta_1	-0,839236	0,108417	-7,7408	<0,0001	***

Média var. dependente	5,302857	D.P. var. dependente	12836,45
Média de inovações	-70,48440	D.P. das inovações	8980,916
Log. da verosimilhança	-1847,399	Critério de Akaike	3708,799
Critério de Schwarz	3730,952	Critério Hannan-Quinn	3717,785

		Real	Imaginária	Módulo	Frequência
AR	Raiz 1	-0,6801	-1,0891	1,2840	-0,3388
	Raiz 2	-0,6801	1,0891	1,2840	0,3388
AR (sazonal)	Raiz 1	6,0733	0,0000	6,0733	0,0000
MA	Raiz 1	-4,9205	0,0000	4,9205	0,5000
MA (sazonal)	Raiz 1	1,1916	0,0000	1,1916	0,0000

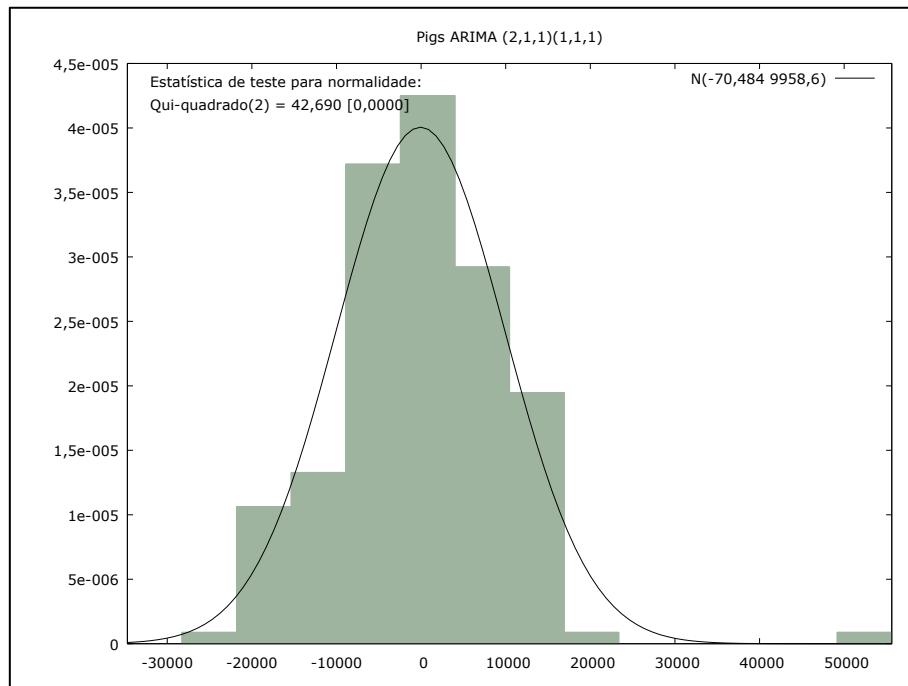
Fonte: autor

Tabela B 11. Resultados do teste para autocorrelação - ARIMA (2,1,1)(1,1,1) - Pigs

ordem	Ljung-Box Q'	valor p
12	14,9244	P(Qui-quadrado(7) > 14,9244) = 0,03698
24	29,4252	P(Qui-quadrado(19) > 29,4252) = 0,05959
36	35,5345	P(Qui-quadrado(31) > 35,5345) = 0,2631

Fonte: autor

Figura B 12. Gráfico da análise da normalidade dos resíduos ARIMA (2,1,1)(1,1,1) - (Pigs)



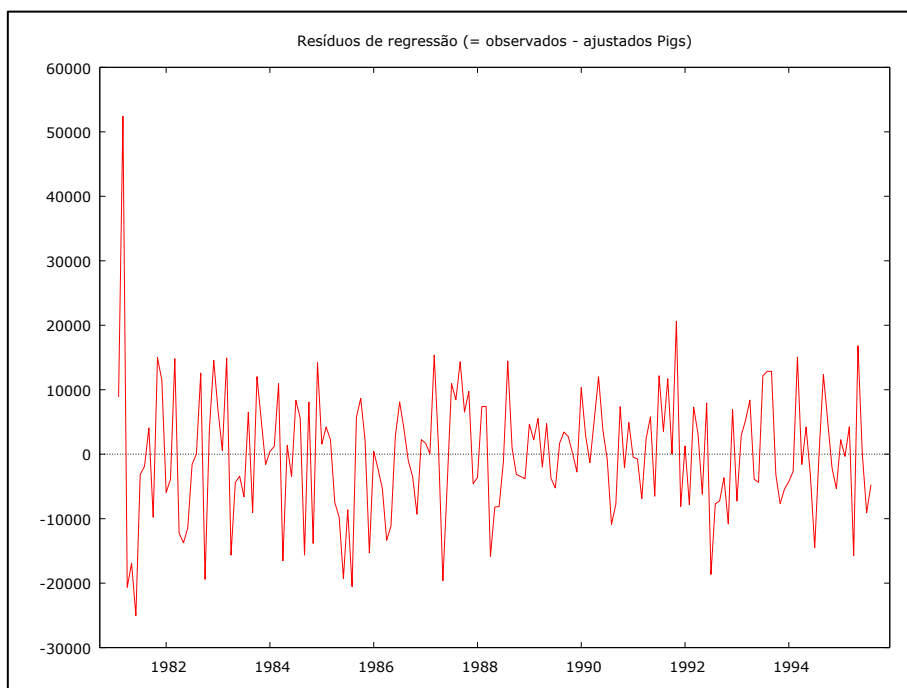
Fonte: autor

Tabela B 13. Resultados da análise dos resíduos - ARIMA (2,1,1)(1,1,1) - Pigs

intervalo	ponto médio	frequência	rel.	acum.
< -21809,	-25035,	1	0,57%	0,57%
-21809, -	-15356,	12	6,86%	7,43% **
-15356, -	-8902,8	15	8,57%	16,00% ***
-8902,8 -	-2449,8	42	24,00%	40,00% *****
-2449,8 -	4003,1	48	27,43%	67,43% *****
4003,1 -	10456,	33	18,86%	86,29% *****
10456, -	16909,	22	12,57%	98,86% ****
16909, -	23362,	1	0,57%	99,43%
23362, -	29815,	0	0,00%	99,43%
29815, -	36268,	0	0,00%	99,43%
36268, -	42721,	0	0,00%	99,43%
42721, -	49174,	0	0,00%	99,43%
>= 49174,	52400,	1	0,57%	100,00%

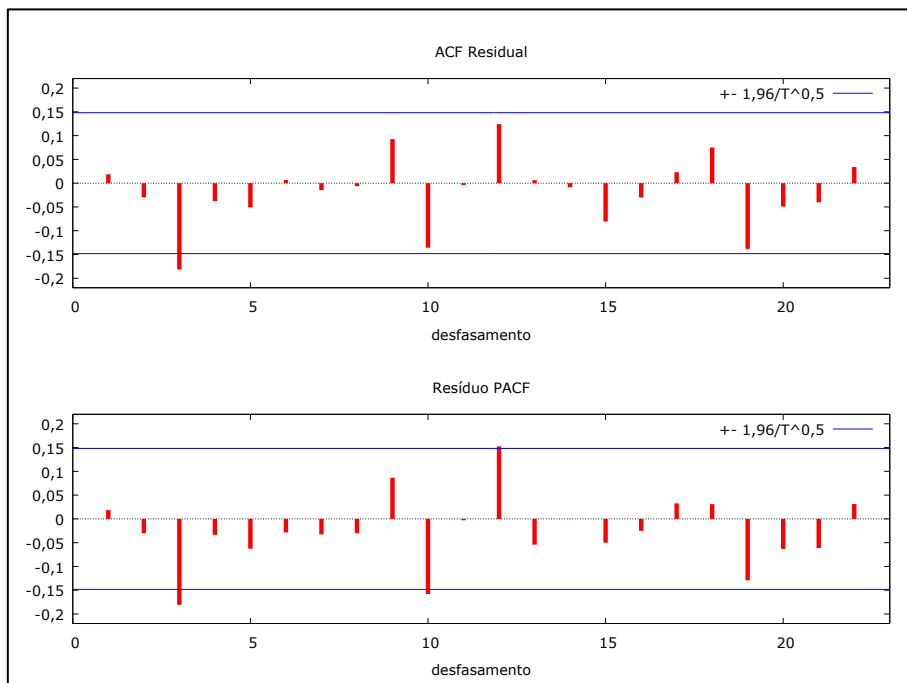
Fonte: autor

Figura B 14. Gráfico dos resíduos - ARIMA (2,1,1)(1,1,1) - Pigs



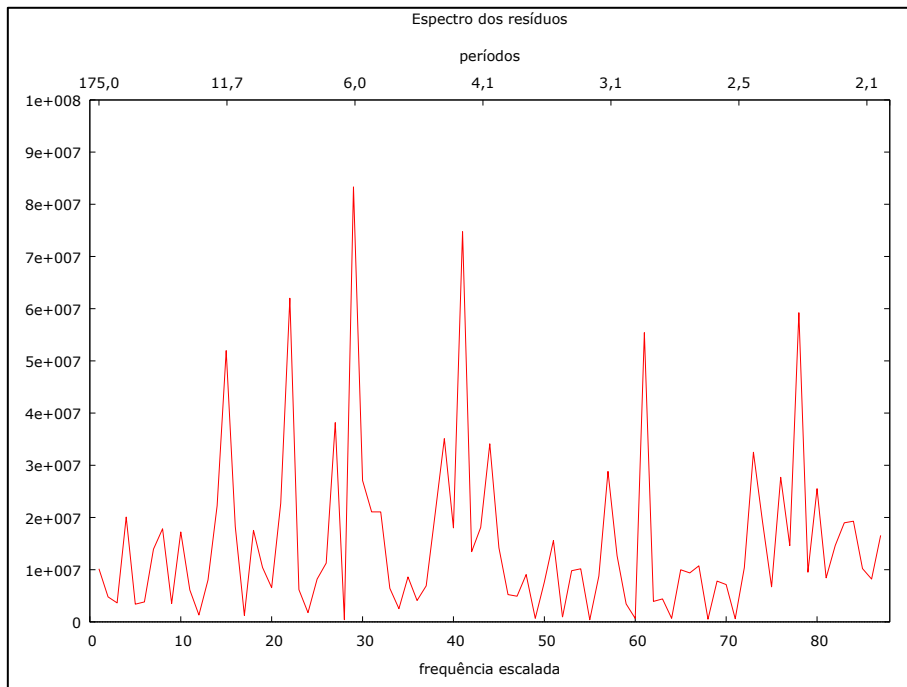
Fonte: autor

Figura B 15. Correlograma dos resíduos - ARIMA (2,1,1)(1,1,0) - Pigs



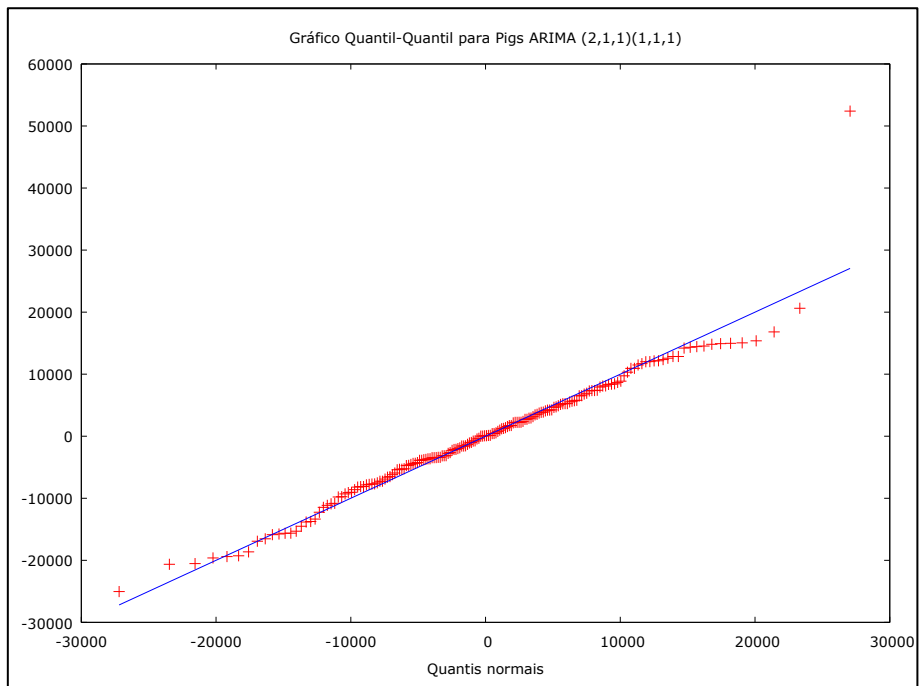
Fonte: autor

Figura B 16. Periodograma dos resíduos - modelo ARIMA (2,1,1)(1,1,1)
- Pigs



Fonte: autor

Figura B 17. Gráfico Quanti-Quanti dos resíduos - modelo ARIMA (2,1,1)(1,1,1) - Pigs



Fonte: autor

Tabela B 18. Resultados da análise da predição - ARIMA
(2,1,1)(1,1,1) – Pigs

Estatísticas da avaliação da Predição	
Erro Médio	-70,484
Erro Médio Quadrado	9,5778e+007
Erro Unitário Médio Quadrado	9786,6
Erro Médio Absoluto	7421,3
Erro Médio Percentual	-0,44091
Erro Médio Percentual Absoluto	8,1977
U de Theil	0,88009
Proporção do enviesamento, UM	5,187e-005
Proporção da regressão, UR	0,21546
Proporção da perturbação, UD	0,78449

Fonte: autor

Anexo C - Resultados da análise dos critérios de informação AIC, BIC e HQC

Os critérios disponíveis no *Gretl* para aplicar a séries temporais são o critério ACI, BCI e HQC.

Para cada uma das séries temporais foram estudados vários modelos ARIMA, optou-se. Sendo esta informação muito importante para a escolha do melhor método, optou-se por colocar neste anexo todos os resultados dos critérios de informação. Assim o leitor terá a possibilidade de verificar para cada série e dada as suas características como variam os resultados dos testes, relativamente às várias variações que se pode fazer nos modelos.

O modelo que apresentar menor valor num critério de informação é o modelo com melhor ajuste, assim as tabelas foram organizadas por ordem decrescente dos valores dos critérios.

Tabela C1. Resultados do critério ACI para Milk

Milk	
Modelo	Critério de Akaike
(1,0,1)	1759,402
(1,1,1)	1746,55
(1,0,1)(1,0,0)	1252,919
(1,1,1)(1,0,0)	1245,308
(4,2,1)(1,1,0)	1088,183
(4,1,1)(1,1,0)	1087,748
(3,2,1)(1,1,0)	1086,42
(1,1,1)(1,1,0)	1085,516
(4,2,2)(1,1,1)	1076,721
(4,2,1)(1,1,1)	1076,149
(2,2,1)(1,1,1)	1074,941
(4,2,2)(3,1,1)	1074,884
(4,2,1)(3,1,1)	1074,473
(2,2,2)(1,1,1)	1073,726
(3,1,1)(1,1,1)	1072,562
(2,1,1)(2,1,1)	1072,153
(1,1,1)(1,1,1)	1071,773
(2,1,1)(1,1,1)	1070,47

Fonte: autor

Tabela C2. Resultados do critério ACI para Pigs

Pigs	
Modelo	Critério de Akaike
(1,0,1)	4011,807
(1,1,1)	3994,043
(2,0,1)(1,0,0)	3984,631
(1,0,1)(1,0,0)	3982,809
(1,1,1)(1,0,0)	3967,951
(2,1,1)(1,0,0)	3959,023
(4,2,2)(2,0,1)	3945,012
(3,2,1)(2,0,1)	3943,989
(2,2,2)(2,0,1)	3943,746
(3,2,2)(2,0,1)	3943,067
(4,2,1)(2,0,1)	3940,374
(2,0,1)(2,0,1)	3740,748
(1,1,1)(1,1,0)	3739,478
(1,1,1)(1,1,1)	3723,499
(2,1,1)(1,1,0)	3721,998
(2,1,1)(1,1,1)	3708,799

Critério de informação Bayesiano (BIC)

Tal como no critério anterior, o modelo que apresentar menor valor da estimativa BIC é considerado o de melhor ajuste.

Tabela C3. Resultados do critério BIC para Milk

Milk	
Modelo	Critério de Schwarz
(1,0,1)	1771,898
(1,1,1)	1734,078
(1,0,1)(1,0,0)	1268,539
(1,1,1)(1,0,0)	1260,898
(4,2,1)(1,1,0)	1112,479
(4,1,1)(1,1,0)	1112,096
(4,2,2)(3,1,1)	1111,328
(4,2,1)(3,1,1)	1107,88
(3,2,1)(1,1,0)	1107,679
(4,2,2)(1,1,1)	1107,091
(4,2,1)(1,1,1)	1103,481
(1,1,1)(1,1,0)	1100,733
(2,2,2)(1,1,1)	1098,021
(3,1,1)(1,1,1)	1096,909
(2,1,1)(2,1,1)	1096,501
(2,2,1)(1,1,1)	1096,199
(2,1,1)(1,1,1)	1091,774
(1,1,1)(1,1,1)	1090,033

Fonte: autor

Tabela C4. Resultados do critério BIC para Pigs

Pigs	
Modelo	Critério de Schwarz
(1,0,1)	4024,753
(1,1,1)	4006,967
(2,0,1)(1,0,0)	4004,049
(1,0,1)(1,0,0)	3998,992
(1,1,1)(1,0,0)	3984,107
(4,2,2)(2,0,1)	3980,495
(2,1,1)(1,0,0)	3978,409
(3,2,2)(2,0,1)	3975,325
(3,2,1)(2,0,1)	3973,02
(2,2,2)(2,0,1)	3972,778
(4,2,1)(2,0,1)	3972,631
(2,0,1)(2,0,1)	3762,942
(1,1,1)(1,1,0)	3755,301
(1,1,1)(1,1,1)	3742,488
(2,1,1)(1,1,0)	3740,987
(2,1,1)(1,1,1)	3730,952

Em conclusão, pode-se verificar que o modelo (2,1,1)(1,1,1) é, no conjunto dos três testes, o que mostra menores valores e será uma boa escolha, para representar cada uma das Séries Temporais.

Critério Informações Hannan-Quinn (HQC)

Tabela C5. Resultados do critério HQC para Milk

Milk	
Modelo	Critério Hannan-Quinn
(1,0,1)	1764,474
(1,1,1)	1739,14
(1,0,1)(1,0,0)	1259,259
(1,1,1)(1,0,0)	1251,636
(4,2,1)(1,1,0)	1098,052
(4,1,1)(1,1,0)	1097,638
(3,2,1)(1,1,0)	1095,055
(1,1,1)(1,1,0)	1091,697
(4,2,2)(3,1,1)	1089,688
(4,2,2)(1,1,1)	1089,057
(4,2,1)(3,1,1)	1088,043
(4,2,1)(1,1,1)	1087,251
(2,2,2)(1,1,1)	1083,595
(2,2,1)(1,1,1)	1083,576
(3,1,1)(1,1,1)	1082,451
(2,1,1)(2,1,1)	1082,043
(1,1,1)(1,1,1)	1079,19
(2,1,1)(1,1,1)	1079,123

Fonte: autor

Tabela C6. Resultados do critério HQC para Pigs

Pigs	
Modelo	Critério Hannan-Quinn
(1,0,1)	4017,052
(1,1,1)	3999,28
(2,0,1)(1,0,0)	3992,498
(1,0,1)(1,0,0)	3989,366
(1,1,1)(1,0,0)	3974,497
(2,1,1)(1,0,0)	3966,878
(4,2,2)(2,0,1)	3959,391
(3,2,2)(2,0,1)	3956,139
(3,2,1)(2,0,1)	3955,753
(2,2,2)(2,0,1)	3955,511
(4,2,1)(2,0,1)	3953,446
(2,0,1)(2,0,1)	3749,75
(1,1,1)(1,1,0)	3745,896
(1,1,1)(1,1,1)	3731,202
(2,1,1)(1,1,0)	3729,701
(2,1,1)(1,1,1)	3717,785

Tal como nos critérios anteriores, o modelo que apresentar menor valor da estimativa BIC é considerado o de melhor ajuste.

Anexo D – Raiz do Erro Quadrado Médio (RMSE).

Tabela D 1. – Cálculo da raiz do erro quadrado médio do modelo ARIMA para MILK

Milk	V.O.	V.P	(VO-VP)	(VO-VP) ²
1975:01	834	836,355	-2,355	5,546025
1975:02	782	793,385	-11,385	129,618225
1975:03	892	896,577	-4,577	20,948929
1975:04	903	911,995	-8,995	80,910025
1975:05	966	973,683	-7,683	59,028489
1975:06	937	947,861	-10,861	117,961321
1975:07	896	905,331	-9,331	87,067561
1975:08	858	865,181	-7,181	51,566761
1975:09	817	815,102	1,898	3,602404
1975:10	827	817,671	9,329	87,030241
1975:11	797	780,409	16,591	275,261281
1975:12	843	820,632	22,368	500,327424
			Somatório	1.418,87
RMSE				10.87

Tabela D 2. – Cálculo da raiz do erro quadrado médio do modelo AES para MILK

Milk	V.O.	V.P	(VO-VP)	(VO-VP) ²
1975:01	834,00	857,70	-23,70	561,70
1975:02	782,00	856,76	-74,76	5.588,64
1975:03	892,00	945,04	-53,04	2.812,99
1975:04	903,00	949,63	-46,63	2.174,13
1975:05	966,00	1.020,80	-54,80	3.002,49
1975:06	937,00	997,60	-60,60	3.671,75
1975:07	896,00	957,47	-61,47	3.778,92
1975:08	858,00	937,91	-79,91	6.385,77
1975:09	817,00	858,86	-41,86	1.752,65
1975:10	827,00	889,80	-62,80	3.943,22
1975:11	797,00	846,83	-49,83	2.482,77
1975:12	843,00	869,28	-26,28	690,43
			Somatório	36.845,47
RMSE				55.416

Tabela D 3. – Cálculo da raiz do erro quadrado médio do modelo ARIMA para PIGS

Pigs	V.O.	V.P	(VO-VP)	(VO-VP)²
1994:09	110.067,00	96.663,51	13.403,49	179.653.544,18
1994:10	101.599,00	91.403,35	10.195,65	103.951.278,92
1994:11	97.646,00	94.892,68	2.753,32	7.580.771,02
1994:12	104.930,00	100.110,92	4.819,08	23.223.532,05
1995:01	88.905,00	80.426,62	8.478,38	71.882.927,42
1995:02	89.936,00	85.934,88	4.001,12	16.008.961,25
1995:03	106.723,00	95.829,77	10.893,23	118.662.459,83
1995:04	84.307,00	90.981,77	-6.674,77	44.552.554,55
1995:05	114.896,00	97.581,59	17.314,41	299.788.793,65
1995:06	106.749,00	95.271,43	11.477,57	131.734.613,10
1995:07	87.892,00	96.021,41	-8.129,41	66.087.306,95
1995:08	100.506,00	95.471,50	5.034,50	25.346.190,25
			Somatório	1.088.472.933,19
			RMSE	9.525,97

Tabela D 4. – Cálculo da raiz do erro quadrado médio do modelo ARIMA para Pigs

Pigs	V.O.	V.P	(VO-VP)	(VO-VP)²
1994:09	110.067,00	90.927,71	19.139,29	366.312.421,70
1994:10	101.599,00	90.951,81	10.647,19	113.362.654,90
1994:11	97.646,00	91.581,37	6.064,63	36.779.737,04
1994:12	104.930,00	93.045,57	11.884,43	141.239.676,42
1995:01	88.905,00	88.153,65	751,35	564.526,82
1995:02	89.936,00	93.622,46	-3.686,46	13.589.987,33
1995:03	106.723,00	90.760,66	15.962,34	254.796.298,28
1995:04	84.307,00	87.120,46	-2.813,46	7.915.557,17
1995:05	114.896,00	95.983,09	18.912,91	357.698.164,67
1995:06	106.749,00	86.123,78	20.625,22	425.399.700,05
1995:07	87.892,00	89.713,58	-1.821,58	3.318.153,70
1995:08	100.506,00	93.265,18	7.240,82	52.429.474,27
			Somatório	1.773.406.352,35
			RMSE	12.156,64

Anexo E – Programação utilizada para o estudo dos modelos AES (scripts).

AES

```
#####
```

```
# MILK DATA
```

```
#####
```

```
require(Rssa)
```

```
milk1<-read.table("C:/Users/Jorge Xavier/Desktop/milkssa.csv",sep=";",header=TRUE)
```

```
milk<- ts(milk1[,2], start=c(1962,1), frequency = 12)
```

```
#####
```

```
# transformar os dados numa serie temporal
```

```
#####
```

```
plot(milk)
```

```
m<-ssa(milk)
```

```
#summary(m)
```

```
plot(m)
```

```
plot(m,type="vector")
```

```
plot(m,type="paired")
```

```
plot(m,type="wcor", groups = 1:20)
```

```
#####
```

```
# utilização de 7 componentes (separação entre sinal e ruído)
```

```
#####
```

```
milkrecon<- reconstruct(m,groups=list(c(1),c(2),c(3),c(4),c(5),c(6),c(7)))
```

```
plot(milkrecon)
```

```
#####
```

Anexos

```
# Agrupamento das componentes de acordo com as que têm sazonalidade com duração semelhante
```

```
#####
```

```
milkrecon1<- reconstruct(m,groups=list(c(1),c(2,3),c(4,5),c(6,7)))
```

```
plot(milkrecon1)
```

```
plot(wcor(m,groups=list(c(1),c(2,3),c(4,5),c(6,7))))
```

```
#####
```

```
# Forecasting
```

```
#####
```

```
m<-ssa(milk)
```

```
f.m <-forecast(m, groups = list(1:7), method = "bootstrap-recurrent", len= 12, R = 10)
```

```
# Plot the result including the last 12 points of the series
```

```
plot(f.m, include = 12, shadecols= "green", type = "l")
```

```
# previsões com intervalos de previsão
```

```
f.m
```

```
#####
```

```
# PIGS DATA
```

```
#####
```

```
require(Rssa)
```

```
pigs1<-read.table("C:/Users/Jorge Xavier/Desktop/pigsssa.csv",sep=";",header=TRUE)
```

```
pigs<- ts(pigs1[,2], start=c(1980,1), frequency = 12)
```

```
#####
```

Anexos

```
# transformação dos dados numa serie temporal

#####

plot(pigs)

p<-ssa(pigs)

#summary(m)

plot(p)

plot(p,type="vector")

plot(p,type="paired")

plot(p,type="wcor", groups = 1:20)

#####

# Utilização de 9 componentes (separação entre sinal e ruido)

#####

pigsrecon<- reconstruct(p,groups=list(c(1),c(2,3),c(4),c(5),c(6), c(7), c(8), c(9)))

plot(pigsrecon)

# Agrupamento das 9 componentes de acordo com as que têm sazonalidade com
duração semelhante

pigsrecon1<- reconstruct(p,groups=list(c(1,2,3),c(4,5),c(6,7),c(8,9)))

plot(pigsrecon1)

plot(wcor(p,groups=list(c(1,2,3),c(4,5),c(6,7),c(8,9))))

#####

# Forecasting

#####

p<-ssa(pigs)
```

Anexos

```
f.p <-forecast(p, groups = list(1:9), method = "bootstrap-recurrent", len= 12, R = 10)
```

```
# Plot the result including the last 12 points of the series
```

```
plot(f.p, include = 12, shadecols= "green", type = "l")
```

```
# previsões com intervalos de previsão
```

```
f.p
```

Anexo F – Dados.

Série Temporal - Milk

1962-01	589	1965-07	702	1968-12	711	1972-06	935
1962-02	561	1965-08	653	1969-01	734	1972-07	894
1962-03	640	1965-09	615	1969-02	690	1972-08	855
1962-04	656	1965-10	621	1969-03	785	1972-09	809
1962-05	727	1965-11	602	1969-04	805	1972-10	810
1962-06	697	1965-12	635	1969-05	871	1972-11	766
1962-07	640	1966-01	677	1969-06	845	1972-12	805
1962-08	599	1966-02	635	1969-07	801	1973-01	821
1962-09	568	1966-03	736	1969-08	764	1973-02	773
1962-10	577	1966-04	755	1969-09	725	1973-03	883
1962-11	553	1966-05	811	1969-10	723	1973-04	898
1962-12	582	1966-06	798	1969-11	690	1973-05	957
1963-01	600	1966-07	735	1969-12	734	1973-06	924
1963-02	566	1966-08	697	1970-01	750	1973-07	881
1963-03	653	1966-09	661	1970-02	707	1973-08	837
1963-04	673	1966-10	667	1970-03	807	1973-09	784
1963-05	742	1966-11	645	1970-04	824	1973-10	791
1963-06	716	1966-12	688	1970-05	886	1973-11	760
1963-07	660	1967-01	713	1970-06	859	1973-12	802
1963-08	617	1967-02	667	1970-07	819	1974-01	828
1963-09	583	1967-03	762	1970-08	783	1974-02	778
1963-10	587	1967-04	784	1970-09	740	1974-03	889
1963-11	565	1967-05	837	1970-10	747	1974-04	902
1963-12	598	1967-06	817	1970-11	711	1974-05	969
1964-01	628	1967-07	767	1970-12	751	1974-06	947
1964-02	618	1967-08	722	1971-01	804	1974-07	908
1964-03	688	1967-09	681	1971-02	756	1974-08	867
1964-04	705	1967-10	687	1971-03	860	1974-09	815
1964-05	770	1967-11	660	1971-04	878	1974-10	812
1964-06	736	1967-12	698	1971-05	942	1974-11	773
1964-07	678	1968-01	717	1971-06	913	1974-12	813
1964-08	639	1968-02	696	1971-07	869	1975-01	834
1964-09	604	1968-03	775	1971-08	834	1975-02	782
1964-10	611	1968-04	796	1971-09	790	1975-03	892
1964-11	594	1968-05	858	1971-10	800	1975-04	903
1964-12	634	1968-06	826	1971-11	763	1975-05	966
1965-01	658	1968-07	783	1971-12	800	1975-06	937
1965-02	622	1968-08	740	1972-01	826	1975-07	896
1965-03	709	1968-09	701	1972-02	799	1975-08	858
1965-04	722	1968-10	706	1972-03	890	1975-09	817
1965-05	782	1968-11	677	1972-04	900	1975-10	827
1965-06	756	1968-11	677	1972-05	961	1975-11	797
						1975-12	843

Série Temporal -Pigs

1980-01	76378	1983-12	110191	1987-11	81778	1991-10	101244
1980-02	71947	1984-01	90974	1987-12	86852	1991-11	114525
1980-03	33873	1984-02	98981	1988-01	69069	1991-12	101139
1980-04	96428	1984-03	107188	1988-02	79556	1992-01	93866
1980-05	105084	1984-04	94177	1988-03	88174	1992-02	95171
1980-06	95741	1984-05	115097	1988-04	66698	1992-03	100183
1980-07	110647	1984-06	113696	1988-05	72258	1992-04	103926
1980-08	100331	1984-07	114532	1988-06	73445	1992-05	102643
1980-09	94133	1984-08	120110	1988-07	76131	1992-06	108387
1980-10	103055	1984-09	93607	1988-08	86082	1992-07	97077
1980-11	90595	1984-10	110925	1988-09	75443	1992-08	90901
1980-12	101457	1984-11	103312	1988-10	73969	1992-09	90336
1981-01	76889	1984-12	120184	1988-11	78139	1992-10	88732
1981-02	81291	1985-01	103069	1988-12	78646	1992-11	83759
1981-03	91643	1985-02	103351	1989-01	66269	1992-12	99267
1981-04	96228	1985-03	111331	1989-02	73776	1993-01	73292
1981-05	102736	1985-04	106161	1989-03	80034	1993-02	78943
1981-06	100264	1985-05	111590	1989-04	70694	1993-03	94399
1981-07	103491	1985-06	99447	1989-05	81823	1993-04	92937
1981-08	97027	1985-07	101987	1989-06	75640	1993-05	90130
1981-09	95240	1985-08	85333	1989-07	75540	1993-06	91055
1981-10	91680	1985-09	86970	1989-08	82229	1993-07	106062
1981-11	101259	1985-10	100561	1989-09	75345	1993-08	103560
1981-12	109564	1985-11	89543	1989-10	77034	1993-09	104075
1982-01	76892	1985-12	89265	1989-11	78589	1993-10	101783
1982-02	85773	1986-01	82719	1989-12	79769	1993-11	93791
1982-03	95210	1986-02	79498	1990-01	75982	1993-12	102313
1982-04	93771	1986-03	74846	1990-02	78074	1994-01	82413
1982-05	98202	1986-04	73819	1990-03	77588	1994-02	83534
1982-06	97906	1986-05	77029	1990-04	84100	1994-03	109011
1982-07	100306	1986-06	78446	1990-05	97966	1994-04	96499
1982-08	94089	1986-07	86978	1990-06	89051	1994-05	102430
1982-09	102680	1986-08	75878	1990-07	93503	1994-06	103002
1982-10	77919	1986-09	69571	1990-08	84747	1994-07	91815
1982-11	93561	1986-10	75722	1990-09	74531	1994-08	99067
1982-12	117062	1986-11	64182	1990-10	91900	1994-09	110067
1983-01	81225	1986-12	77357	1990-11	81635	1994-10	101599
1983-02	88357	1987-01	63292	1990-12	89797	1994-11	97646
1983-03	106175	1987-02	59380	1991-01	81022	1994-12	104930
1983-04	91922	1987-03	78332	1991-02	78265	1995-01	88905
1983-05	104114	1987-04	72381	1991-03	77271	1995-02	89936
1983-06	109959	1987-05	55971	1991-04	85043	1995-03	106723
1983-07	97880	1987-06	69750	1991-05	95418	1995-04	84307
1983-08	105386	1987-07	85472	1991-06	79568	1995-05	114896
1983-09	96479	1987-08	70133	1991-07	103283	1995-06	106749
1983-10	97580	1987-09	79125	1991-08	95770	1995-07	87892
1983-11	109490	1987-10	85805	1991-09	91297	1995-08	100506