

UNIVERSIDADE FEDERAL DA BAHIA
DEPARTAMENTO ESTATÍSTICA
MAT185 - Análise das Séries Temporais
Prof(a): Gecynalda Gomes

ANÁLISE DA SÉRIE TEMPORAL DO FECHAMENTO DAS AÇÕES DA GOOGLE

Gerson Rodrigues Primo Junior

12 de dezembro de 2011

Sumário

1	Introdução	2
2	Metodologia	2
2.1	Modelo de Suavização de Holt-Winters (HW)	2
2.1.1	Holt-Winters Aditivo	2
2.1.2	Holt-Winters Multiplicativo	3
2.2	Método de Box & Jenkins	3
2.2.1	Modelo SARIMA	4
2.2.2	Identificação do Modelo	4
2.2.3	Diagnóstico do Modelo	5
3	Resultados e Discussões	5
3.1	Análise Descritiva	5
3.2	Ajuste pelo método de Holt-Winters (HW)	6
3.3	Ajuste método de Box & Jenkins	7
3.4	Comparação dos ajustes	9
4	Conclusões	10
	Referências	11
	Anexos	12
	Anexo - A	12
	Anexo - B	13
	Anexo - C	16

1 Introdução

A Google é uma empresa voltada para serviços online, no qual seu faturamento se baseia em publicidade pelo *AdWords*¹. Idealizada em 1996, como projeto de pesquisa de Larry Page e Sergey Brin, até então estudantes de doutorado na Universidade Stanford a google teve um crescimento em paralelo com o da internet e sem dúvida é hoje o principal meio de busca online. Desde seu surgimento a google foi uma das empresas que mais cresceu nos últimos anos e segundo a Brandz em 2010 a google ficou em 1ª lugar no ranking das maiores empresas do mundo.

A primeira venda das ações da Google no mercado de ações ocorreu em agosto de 2004, marco importante para o crescimento da empresa, pois a mesma começa a atrair investidores do mundo todo tendo assim uma grande demanda das suas ações e por consequência sua valorização. Através dos métodos de suavização de Holt-Winters e o de Box & Jenkins o presente trabalho tem como objetivo avaliar o crescimento das ações da Google desde início de sua comercialização, como também fazer previsões futuras.

2 Metodologia

Analizou-se a série referente ao fechamento das ações mensal da empresa GOOGLE durante o período de agosto de 2004 a setembro de 2011 e a base de dados foi obtida no site *yahoo finanças*².

As principais técnicas utilizadas no relatório são os modelos de série temporais descritos a seguir, devido a série estudada apresentar tendência e sazonalidade identificados a partir da análise gráfica e pelos o teste de Cox-Stuart para avaliar tendência e o teste de Kruskal-Wallis para avaliar a sazonalidade. Para os testes utilizados adotou-se o nível de significância de 5%.

2.1 Modelo de Suavização de Holt-Winters (HW)

Quando a série apresenta tendência e sazonalidade um modelo proposto para trabalhar com a série é o modelo de suavização de Holt-Winters, pois ele consegue incorporar as flutuações da série, no qual o modelo é escrito por três componentes (nível, tendência e sazonalidade) e dividido em dois modelos (Aditivo e Multiplicativo).

2.1.1 Holt-Winters Aditivo

Para o modelo Aditivo temos a seguinte expressão

$$Z_t = \mu_t + T_t + F_t + a_t$$

Onde;

Fator sazonal: $\hat{F}_t = D(Z_t - \bar{Z}_t) + (1 - D)\hat{F}_{t-s}, 0 < D < 1$,

Fator de nível: $\bar{Z}_t = A(Z_t - \hat{F}_{t-s}) + (1 - A)(\bar{Z}_{t-1} - \hat{T}_{t-1}), 0 < A < 1$

Fator de tendência: $\hat{T}_t = C(\bar{Z}_t - \bar{Z}_{t-1}) + (1 - C)\hat{T}_{t-1}, 0 < C < 1$

Sendo A, B e C constantes de suavização e a_t o resíduo aleatório com média zero e variância constante (σ_a^2).

Previsão da série sazonal aditiva

Considerando os resultados acima temos que:

$$\hat{Z}_t(h) = \bar{Z}_t + h\hat{T}_t + \hat{F}_{t+h-s}, h = 1, 2, \dots, 2;$$

$$\hat{Z}_t(h) = \bar{Z}_t + h\hat{T}_t + \hat{F}_{t+h-2s}, h = s + 1, \dots, 2s;$$

etc

Dado que \bar{Z}_t , \hat{T}_t e \hat{F}_t são obtido a partir dos fatores sazonal, nível e tendência é feita as atualizações, considerando a série sazonal aditiva.

¹São anúncios em forma de links encontrados, principalmente, nos sites de busca relacionados às palavras-chave que o internauta está procurando no motor de busca da página.

²<http://br.finance.yahoo.com/>

$$\begin{aligned}\hat{F}_{t+1} &= D(Z_{t+1} - \bar{Z}_{t+1}) + (1 - D)\hat{F}_{t+1-s}, \\ \bar{Z}_{t+1} &= A(Z_{t+1} - \hat{F}_{t+1-s}) + (1 - A)(\bar{Z}_{t-1} - \hat{T}_t) \\ \hat{T}_{t+1} &= C(\bar{Z}_{t+1} - \bar{Z}_t) + (1 - C)\hat{T}_t\end{aligned}$$

Diante as atualizações a previsão para o valor Z_{t+h} efeito por:

$$\begin{aligned}\hat{Z}_t(h-1) &= \bar{Z}_{t+1} + (h-1)\hat{T}_{t+1} + \hat{F}_{t+1+h-s}, \quad h = 1, 2, \dots, s+1; \\ \hat{Z}_t(h-1) &= \bar{Z}_{t+1} + (h-1)\hat{T}_{t+1} + \hat{F}_{t+1+h-2s}, \quad h = s+2, \dots, 2s+1;\end{aligned}$$

etc

2.1.2 Holt-Winters Multiplicativo

Para o modelo Multiplicativo temos a seguinte expressão

$$Z_t = \mu_t F_t + T_t + a_t; t = 1, \dots, N$$

Onde;

$$\text{Fator sazonal: } \hat{F}_t = D\left(\frac{Z_t}{\bar{Z}_t}\right) + (1 - D)\hat{F}_{t-s}, 0 < D < 1 \text{ e } t = s+1, \dots, N$$

$$\text{Fator de nível: } \bar{Z}_t = A\left(\frac{Z_t}{\hat{F}_{t-s}}\right) + (1 - A)(\bar{Z}_{t-1} - \hat{T}_{t-1}), 0 < A < 1 \text{ e } t = s+1, \dots, N$$

$$\text{Fator de tendência: } \hat{T}_t = C(\bar{Z}_t - \bar{Z}_{t-1}) + (1 - C)\hat{T}_{t-1}, 0 < C < 1 \text{ e } t = s+1, \dots, N$$

Sendo A, B e C constantes de suavização e a_t o resíduo aleatório com média zero e variância constante (σ_a^2).

Previsão da série sazonal multiplicativa

Considerando os resultados acima temos que:

$$\hat{Z}_t(h) = (\bar{Z}_t + h\hat{T}_t)\hat{F}_{t+h-s}, \quad h = 1, 2, \dots, 2;$$

$$\hat{Z}_t(h) = (\bar{Z}_t + h\hat{T}_t)\hat{F}_{t+h-2s}, \quad h = s+1, \dots, 2s;$$

etc

Dado que \bar{Z}_t , \hat{T}_t e \hat{F}_t são obtido parti dos fatores sazonal, nível e tendência é feita as atualizações, considerando a série sazonal multiplicativa.

$$\hat{F}_{t+1} = D\left(\frac{Z_{t+1}}{\bar{Z}_{t+1}}\right) + (1 - D)\hat{F}_{t+1-s},$$

$$\bar{Z}_{t+1} = A\left(\frac{Z_{t+1}}{\hat{F}_{t+1-s}}\right) + (1 - A)(\bar{Z}_{t-1} - \hat{T}_t)$$

$$\hat{T}_{t+1} = C(\bar{Z}_{t+1} - \bar{Z}_t) + (1 - C)\hat{T}_t$$

Diante as atualizações a previsão para o valor Z_{t+h} efeito por:

$$\hat{Z}_t(h-1) = (\bar{Z}_{t+1} + (h-1)\hat{T}_{t+1})\hat{F}_{t+1+h-s}, \quad h = 1, 2, \dots, s+1;$$

$$\hat{Z}_t(h-1) = (\bar{Z}_{t+1} + (h-1)\hat{T}_{t+1})\hat{F}_{t+1+h-2s}, \quad h = s+2, \dots, 2s+1;$$

Segundo (MORETTIN & TOLOI, 1981) as previsões obtidas por meio da expressões acima é ótima se Z_t for gerado por um processo ARIMA sazonal. Porém uma dificuldade do ajustamento pela suavização de Holt-Winters é encontrar os valores mais adequados para as constantes A, C e D . Ressaltando que as constantes devem ser obtidas de forma que minimize a soma dos quadrados dos erros de ajustamento técnica conhecida dentro da estatística com *Erro Quadrático Médio (EQM)*.

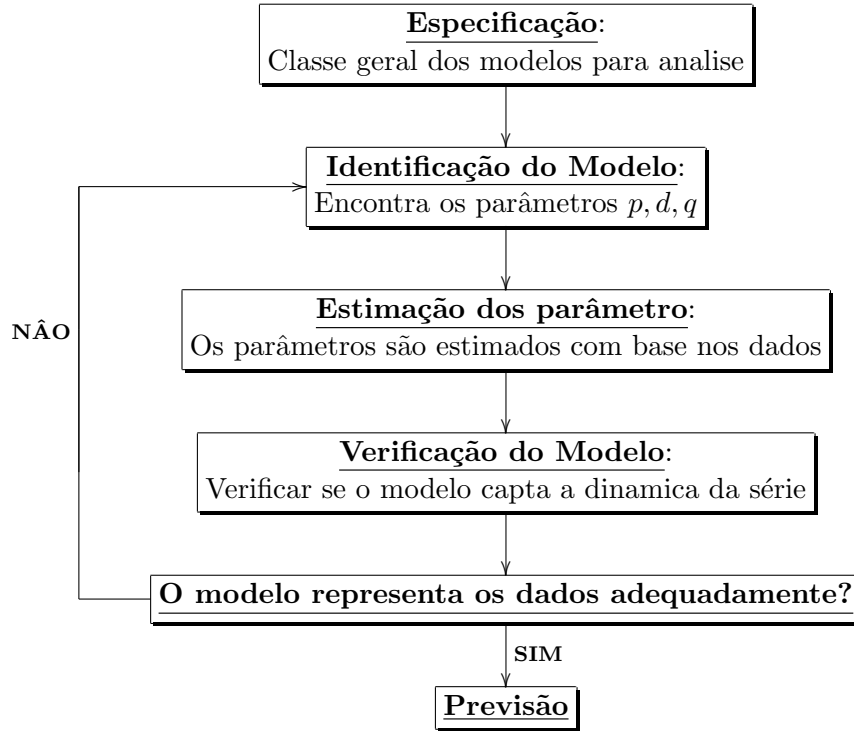
2.2 Método de Box & Jenkins

Se tratando de análise paramétrica de série temporal é bastante utilizado o os modelos ARIMA (Modelos auto-regressivos integrados de médias móveis) proposto por George Box e Gwilym Jenkins. De forma generalizada refere-se, respectivamente, às ordens de auto-regressão (p), de integração (d) e de média móvel (q), em que o modelo é expresso por: $ARIMA(p, d, q)$ ³.

Para diferentes valores de p, d, q se tem casos particulares do modelo, como no caso de $d = 0$, tem o modelo ARMA(p, q), e no caso de também $q = 0$, temos os modelo AR(p). O modelo ARIMA(0, 1, 0) é o passeio aleatório. Já no estudo em questão será considerado o modelo SARIMA por levar em consideração a sazonalidade encontrada na série em estudo.

A identificação do modelo com os parâmetros apropriado pode ser conduzida por um circulo interativo de seis estágios, como mostra o seguinte fluxograma.

³ p é o numero de termos auto-regressivo, d é o numero de diferença, q é o numero de termos da média móvel



Segundo (Morettin, 2006) a fase crítica do procedimento acima é a identificação. É possível que vários pesquisadores identifiquem modelos diferentes para a mesma série temporal.

2.2.1 Modelo SARIMA

Considerando que a série possui sazonalidade deve acrescentar ao método de Box & Jenhins o componente sazonal, logo considere a série Z_t e sejam:

$\phi(B^s) = 1 - \phi_1(B^1) - \dots - \phi_P(B^{Ps})$ o operador autorregressivo de ordem P;

$\theta(B^s) = 1 - \theta_1(B^1) - \dots - \theta_Q(B^{Qs})$ o operador autorregressivo sazonal de ordem Q e

$\Delta_s^D((1 - B)^s)^D$ o numero de diferenças sazonais.

Logo o modelo SARIMA (p,d,q)x(P,D,Q) é expresso por;

$$1 - \phi_1(B^1) - \dots - \phi_P(B^{Ps}) D_s((1 - B)^s)^D Z_t = 1 - \theta_1(B^1) - \dots - \theta_Q(B^{Qs}) Y_t$$

Onde, $Y_t = Z_t - Z_{(t-s)}$

2.2.2 Identificação do Modelo

Uma das etapas mais crítica do método de Box & Jenkis é a identificação do modelo. Esta escolha é feita com base nas autocorrelações parciais estimadas, no qual espera-se representar as respectivas quantidades teoria (desconhecidas). Considerando que exista uma grande correlação entre o valor do evento no tempo t e valores passados do evento (nos pontos $t-1, t-2, \dots$). O que acontece é que a variável Y na equação acima será substituída pela própria variável Z, atrasada k passos (lags). Fazendo isso, será possível estudar a correlação entre uma série Z_t com uma série Z_{t-1}, Z_{t-2} , etc. Geralmente este estudo é feito através das funções de autocovariância (ACVF) e autocorrelação (ACF). Onde ACF ρ_j é estimado por;

$$r_j = \frac{c_j}{c_0}, j = 0, 1, \dots, N-1,$$

Em que c_j é estimativa da ACVF γ ,

$$c_j = \frac{1}{N} \sum_{t=1}^{N-j} \{(Z_t - \bar{Z})(Z_{t+j} - \bar{Z})\}, j = 0, 1, \dots, N-1,$$

Sendo \bar{Z} a média amostral. Desta forma espera-se que: $r_j \pm t_{N-1}\hat{\sigma}_t$, onde t_{N-1} é o quantil da distribuição *t-student* e $\hat{\sigma}^2(r_j) \simeq \frac{1}{N} [1 + 2 \sum_{v=1}^q r_v^2]$, $j > q$.

2.2.3 Diagnóstico do Modelo

O diagnostico realizado neste trabalho para o modelo SARIMA foi feito considerando resíduo padronizado, que consiste na diferença do valor real com o predito dividido pelo seu desvio padrão, em que se espera que essas diferenças estejam aleatório em torno do 0 (zero), pela medida do ACF dos resíduos e pela estatística de Ljung & Box.

Estatística e Ljung & Box

O teste de Ljung-Box (nome de Greta M. Ljung e George EP Box) é um tipo de teste estatístico para saber se qualquer grupo de autocorrelações de uma serie temporal são diferentes de zero. Em vez dos testes de aleatoriedade em cada “lag” e as distintas, ele testa a aleatoriedade “global”, baseada num número de defasagens.

Sob a hipótese nula (H_0): Os dados são aleatórios

A estatística de teste é dada por:

$$Q = n(n+2) \sum_{k=1}^h \frac{\hat{\rho}_k^2}{n-k}$$

Onde n é o tamanho da amostra, $\hat{\rho}_k^2$ é a autocorrelação da amostra, K é o “lag”, e h é o numero de defasagens a ser testado. Para nível de significância α , a região crítica para rejeição da hipótese de aleatoriedade ocorre quando:

$$Q > \chi_{1-\alpha, h}^2$$

Sendo $\chi_{1-\alpha, h}^2$ o quantil da distribuição *Qui-quadrado* com h graus de liberdades.

3 Resultados e Discussões

3.1 Análise Descritiva

Durante os anos de 2004 a 2011 a cotação das ações de fechamento da GOOGLE, em média, foi de 325,78 milhões de dolares, com o coeficiente de variação em torno de 27%. No qual teve a cotação mínima (84,75 milhões) no início do estudo e a cotação máxima (484,63 milhões) em outubro de 2007 (ver Figura 1).

Tabela 1: Medidas descritivas do fechamento das ações da GOOGLE (2004-2011).

Mínimo	Máximo	Média	Desvio-Padrão	C.V(%)
84,75	484,63	325,78	88.80	27,26

Considerando o teste de Cox-Stuart, temos evidencias estatísticas, ao nível de significância de 5%, para rejeitarmos que o fechamento das ações da Google não tem tendência ($p = 0,001$). Já para avaliar a sazonalidade pelo do teste Kruskal-Wallis foi feita a primeira diferenciação para retirar a tendência e pelo teste de Cox-Stuart verificou que a primeira diferenciação não apresenta tendência ($p = 0,1215$) e então com o teste Kruskal-Wallis note-se que a mesma apresenta sazonalidade ($p = 0,02947$). Assim como pela Figura 1 pode ser observado o comportamento da tendência e sazonalidade indicado pelos testes, que de forma clara a Figura 1(b) mostra a tendência da série como pela Figura 1(c) a presença da sazonalidade.

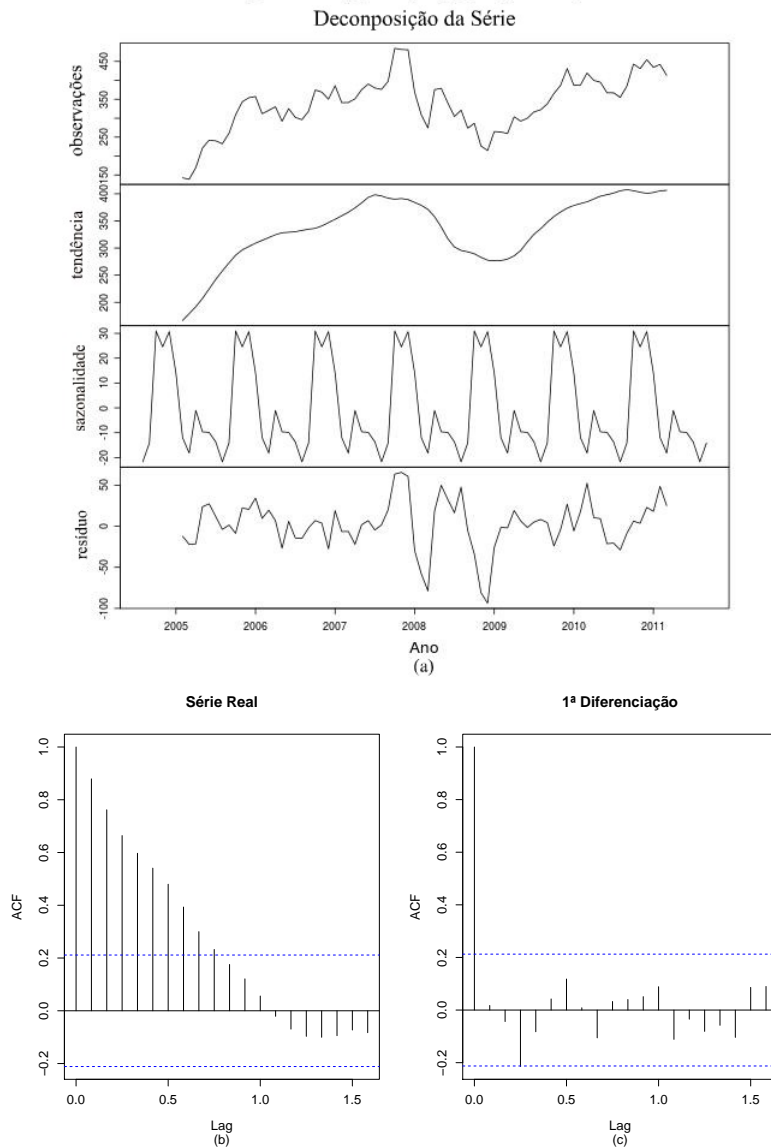


Figura 1: Decomposição e correlograma das ações de fechamento da Google.

Afim de verificar a bondade do ajuste na predição de valores futuros retirou-se as observações referentes aos 12 últimos meses da série e então refeito as análises descritivas, na qual apresentou as mesmas características da série completa, apresentando tendência e sazonalidade (em Anexo). Adiante disso utilizou a série dos últimos 12 meses para verificar a bondade do ajuste pelos métodos abortados no trabalho.

3.2 Ajuste pelo método de Holt-Winters (HW)

Aplicando o método de HW aditivo no fechamento das ações da Google com periodicidade 12 obteve, pelo erro quadrático médio, os valores $\alpha = 0,884$, $\beta = 0,0374$ e $\gamma = 1,000$ para o vetor das constantes de suavização e pelo método multiplicativo obteve os seguinte valores $\alpha = 0,765$, $\beta = 0,049$ e $\gamma = 1,000$, o qual não mostra um diferença gritante entre as estimativas dos parâmetros. A tabela 2 mostra os valores preditos pelo método de HW aditivo e HW multiplicativo para os 12 meses a frente a parti de outubro de 2010 e comparando com os valores reais deste período nota-se o método de HW-Aditivo houve uma maior subestimação do verdadeiro valor que o método HW-Multiplicativo.

A nível de esclarecimento o viés abordado na Tabela 2 é oriundo da subtração do valor predito do valor real e considerando o viés médio, o método de HW-multiplicativo demonstra captar melhor a dinâmica da série. No entanto quando analisado o viés em termos do seu desvio-padrão (SD)

Tabela 2: Previsões obtidas para o fechamento das ações da Google pelos métodos de Holt-Winters aditivo e multiplicativo com 12 passos à frente.

Valor Real		HW-Aditivo		HW-Multiplicativo	
Mês (2010-11)	Valor real	Predito	Viés	Predito	Viés
Outubro	442,910	422,430	20,480	422,481	20,430
Novembro	431,160	434,424	-3,264	422,354	8,806
Dezembro	454,500	451,742	2,758	444,934	9,566
Janeiro	434,450	447,763	-13,313	440,416	-5,966
Fevereiro	442,050	406,344	35,706	396,866	45,184
Março	413,270	387,333	25,937	358,986	54,284
Abril	362,470	403,345	-40,875	367,100	-4,630
Mai	364,840	419,786	-54,946	375,727	-10,887
Junho	347,000	434,014	-87,014	390,005	-43,005
Julho	422,550	432,165	-9,614	404,173	18,377
Agosto	376,580	426,657	-50,077	408,788	-32,208
Setembro	384,820	431,901	-47,081	417,353	-32,533
		Viés médio(SD)	-18,442(37,610)		2,285(30,110)

nota-se que apesar de ser, em média, menos vizado apresenta uma dispersão muito grande. A figura 3 ajudar visualizar melhor o comportamento das previsões para ambos os métodos.

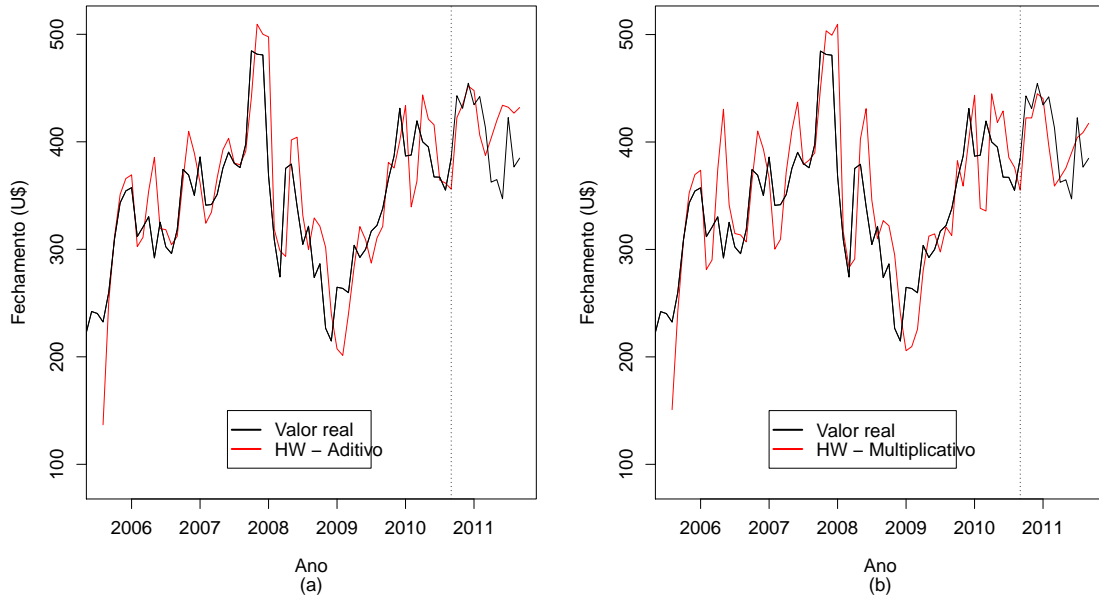


Figura 2: Serie temporal do fechamento das ações da Google, segundo o método de HW.

3.3 Ajuste método de Box & Jenkins

A seleção do modelo pelo método de Box & Jenkins foi utilizado o critério de AIC⁴, onde foram ajustado diferentes modelos, variando os valores de p e q entre 0 e 3, os valores de d , P , D e Q entre 0 e 1. A Tabela 3 mostra os modelos que apresentaram os menores AIC no qual os modelos selecionados para fazer as previsões foram SARIMA(1,1,1)x(1,1,0) e SARIMA(2,1,1)x(1,1,0).

⁴AIC(Akaike Information Criterion) é um valor dado por: $AIC = -2\log L + 2(p + q)$, onde L é função de verossimilhança, $p + q$ o numero de parâmetros e n o numero de observações e quanto menor o AIC melhor o modelo.

Tabela 3: **Modelo encontrados pelo método de Box & Jenkins.**

Modelo	(p,d,q)x(P,D,Q)	AIC
SARIMA1	(2,1,1)x(1,1,0)	643,65
SARIMA2	(1,1,1)x(1,1,0)	631,46
SARIMA3	(1,1,1)x(1,0,1)	728,15
SARIMA4	(2,1,1)x(1,1,0)	632,87
SARIMA5	(2,1,1)x(1,0,1)	728,88

Já a tabela 4 apresenta os resultados das previsões dos modelos SARIMA 2 e SARIMA 4, onde vemos que o viés médio do SARIMA 4 foi de 2,506, menor que o SARIMA 2, no entanto ambos os modelos apresentam variabilidade grande para o viés e a parti do diagnostico mostrado na Figura 3 tanto o modelo SARIMA 2, quanto o SARIMA 4 se ajustaram bem aos dados, pois apresentaram uma certa aleatoriedade dos resíduos padronizados, o ACF dos resíduos mantiveram dentro do intervalo especificado e com relação a estatística de Ljung & Box em todos os caso, ao nível de 5%, não rejeita a aleatoriedade dos dados.

Tabela 4: **Previsões obtidas para o fechamento das ações da Google pelo de método de Box & Jenkins com 12 passos à frente.**

Valor Real		SARIMA(1,1,1)x(1,1,0)		SARIMA(2,1,1)x(1,1,0)	
Mês (2010-11)	Valor real	Predito	Viés	Predito	Viés
Outubro	442,910	405,665	37,245	407,469	35,441
Novembro	431,160	380,428	50,732	379,542	51,618
Dezembro	454,500	392,345	62,156	389,315	65,185
Janeiro	434,450	402,315	32,135	398,267	36,183
Fevereiro	442,050	402,069	39,981	397,466	44,584
Março	413,270	413,254	0,016	408,349	4,921
Abril	362,470	430,406	-67,936	425,462	-62,992
Maio	364,840	421,833	-56,993	416,828	-51,988
Junho	347,000	414,351	-67,351	409,365	-62,365
Julho	422,550	423,985	-1,435	419,010	3,540
Agosto	376,580	421,956	-45,376	416,998	-40,418
Setembro	384,820	443,583	-58,763	438,605	-53,785
		Viés médio(SD)	-6,299(50,360)		2,506(49,100)

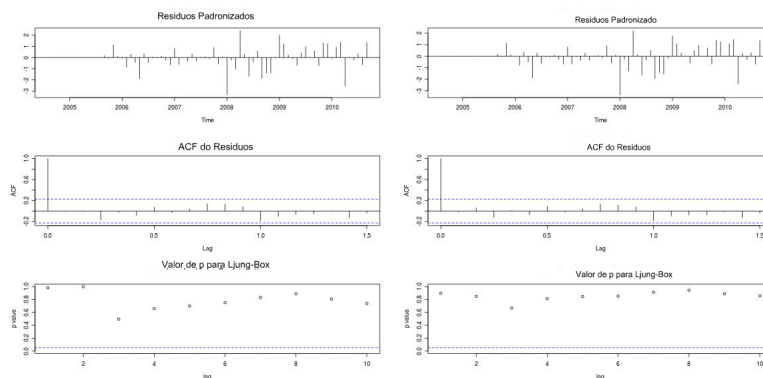


Figura 3: Diagnostico dos modelos: SARIMA 2 (a) e SARIMA 4 (b)

Com relação as estimativas pelo método de Box & Jenkins nota-se que para os dois modelos selecionados acompanham bem a dinâmica da série nos primeiros anos e a parti do ano de 2006 as séries ajustadas tiveram uma oscilação menor que a série real, como mostra a Figura 4. E quanto as previsões a Figura 4 mostra que as previsões para 12 passos a frente os modelos não mostram satisfatórios.

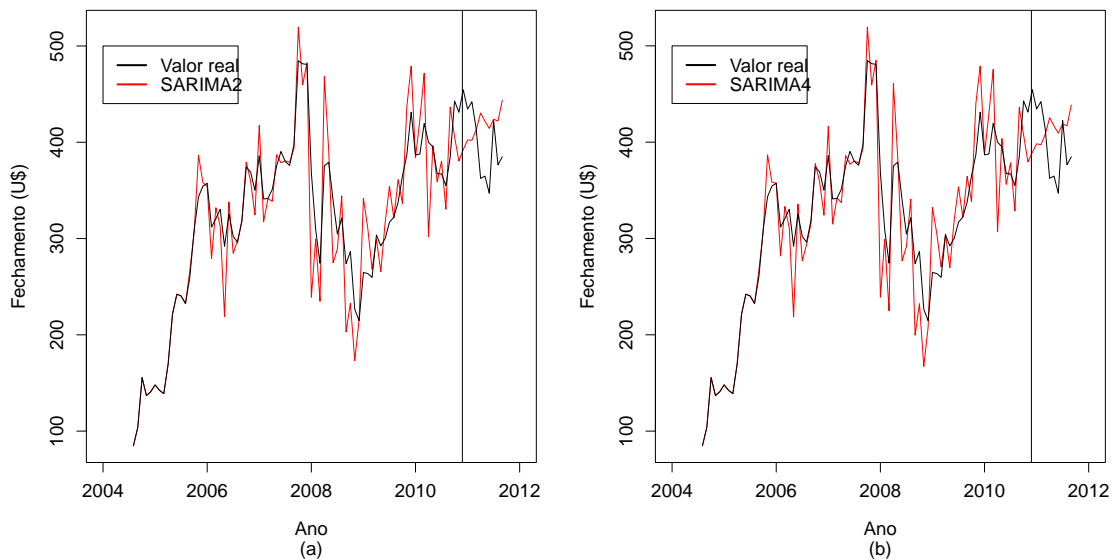


Figura 4: Série Temporal do Fechamento das ações da Google, segundo o método de Box & Jenkins

3.4 Comparação dos ajustes

Entre os métodos de suavização de Holt-Winters e o de Box & Jenkins pode observar que a suavização de Holt-Winters captou melhor a dinâmica da série, onde considerando a predição dos 12 passos a frente, constata uma melhor aproximação das suavizações de Holt-Winters com o valor real do período, como mostra a Figura 5. E ainda tomando com base o desvio médio das Tabelas 2 e 4, nota-se que o modelo de suavização de HW multiplicativo teve o menor viés médio (2,285) com a menor variabilidade (30,110), Ressaltando ainda que mesmo o viés médio do modelo de suavização de HW aditivo ter sido maior, em termos absoluto, que os do método de Box & Jenkins, a suavização de HW aditivo apresentou uma melhor predição (ver Figura 5).

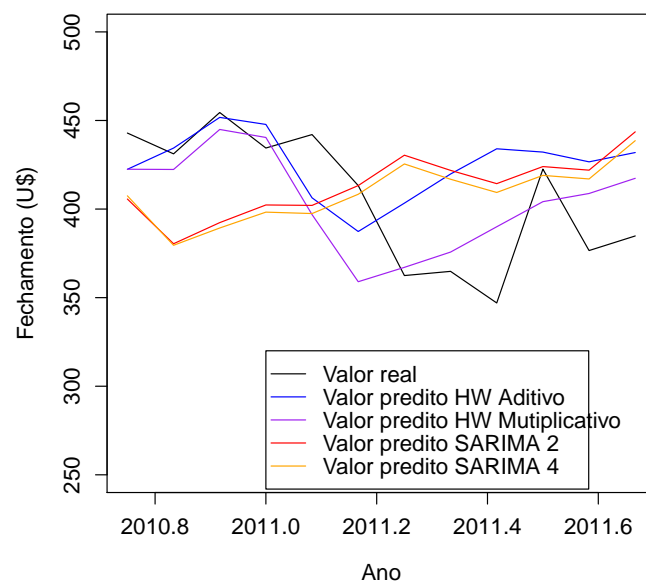


Figura 5: Previsão para 12 passos a frente do fechamento das ações da Google

E considerando o erro padrão (SSE) e o erro padrão médio (MSE) nota-se que, entre os modelos selecionados o que apresenta mais apropriado para o comportamento do fechamento das ações da

google são os modelos de suavização de Holt-Winters, como nota-se na Tabela 5 foram os ajustes que obtiveram os menores valores para o SSE e o MSE.

Tabela 5: **SSE e MSE dos modelos selecionados.**

Modelo	SSE	MSE
HW-Aditivo	19640,55	1636,71
HW-Multiplicativo	10033,26	836,11
SARIMA 2	28369,12	2364,09
SARIMA 4	26589,89	2215,82

4 Conclusões

O trabalho buscou identificar um modelo que mais se apropriava a série de fechamento das ações da Google e com base na teoria estuda nota-se que os modelos obtidos pelos os métodos de suavização de Holt-Winters foram os modelos que apresentaram os melhores ajustes, sobre tudo o HW-Multiplicativo. Onde pode ser visto tanto pelos Figuras 2 (b) quanto pela Figura 5 e Tabela 5 a maior eficiência do ajuste.

Referências

- [1] Gomes, A.S (2003), *Modelagem e previsão da arrecadação do imposto de renda no Brasil*, dissertação de mestrado, UFPE, Recife-PE.
- [2] Morettin, P.A & Toloi, C.M.C (2006), *Análise de Séries Temporais*, 2.ed São Paulo, Edgard Blucher.
- [3] Pacheco, A.G.F (200x), *Estudo de influências de variáveis meteorológicas no aparecimento de casos graves de leptospirose em Salvador-Ba via modelos de séries temporais*, Fundação Oswaldo Cruz, Salvador-BA.

.

Anexo - A

- Descritivas dos dados sem os últimos meses

Tabela 6: Medidas descritivas do fechamento das ações da GOOGLE (08/2004 - 09/2010).

Mínimo	Máximo	Média	Desvio-Padrão	C.V(%)
84,75	484.63	312,70	87.93	28,12

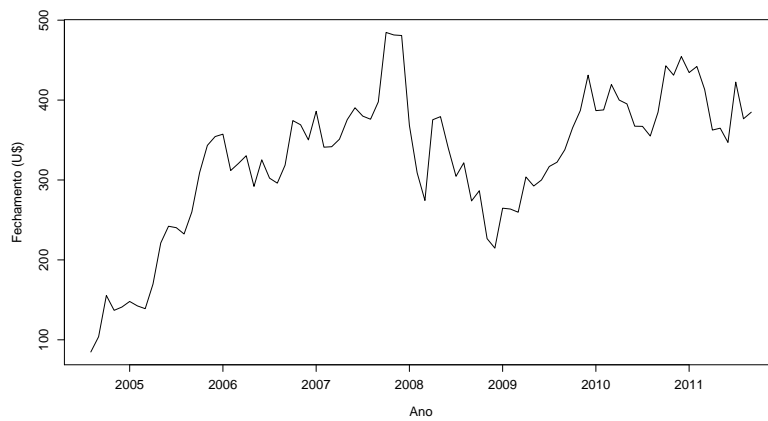


Figura 6: Série temporal das ações de fechamento das ações da Google (2004-2011)

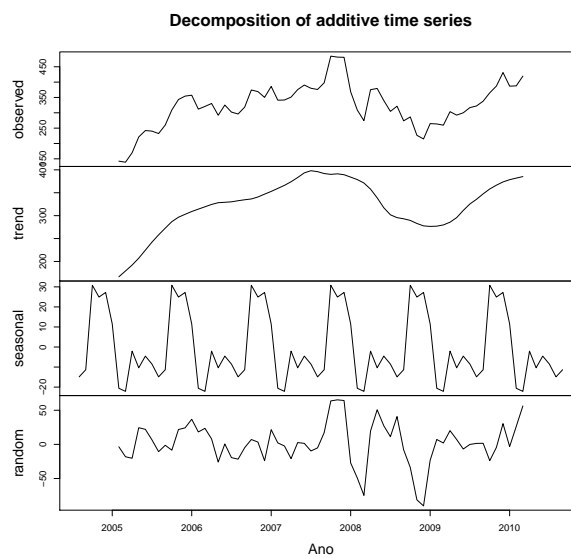


Figura 7: Decompose das ações de fechamento das ações da Google, retirado os 12 últimos meses.

Anexo - B

- Roteiro do R - Análise da série temporal das ações de fechamento da Google

```
##### Analise de Série Temporal - GOOGLE #####
### Banco de dados
dados=read.table("table_mensal.csv",head=T)
google = ts(dados,start=c(2004,8),frequency=12)
google
library(tseries)

##### Descritiva #####
plot(google, type="l",xlab="Ano",ylab="Fechamento (U$)")
summary(google);sd(google)
CV=(sd(google)/mean(google))*100;CV;sd(google)

# Decompondo a serie
plot(decompose(google),xlab="Ano",sub="(a)")

# Testes para tendência e sazonalidade
cox.stuart.test(google) # Utilizou implementação fornecida em sala
grupo=c(rep(seq(1,12),7),c(1,2))
kruskal.test(google,grupo) # Teste de Sazonalidade

# Primeira Diferenciação
cox.stuart.test(diff(google,difference=1))
grupo=c(rep(seq(1,12),7),1)
kruskal.test(diff(google,difference=1),grupo) # Teste de Sazonalidade 1o diferenciação

# Correlograma
par(mfrow=c(1,2))
acf(google,main="Série Real",sub="(b)")
acf(diff(google,difference=1),main="1ª Diferenciação",sub="(c)")

#### Retirando os ultimos 12 meses
base=google[-c(75:86)]
base_comp=google[75:86]
google2=ts(base,start=c(2004,8),frequency=12); google2
google_comp=ts(base_comp,start=c(2010,10),frequency=12)

# Descritiva
plot(google2, type="l",xlab="Ano",ylab="Fechamento (U$)",
      main="Serie Temporal do fechamento das ações da Google(2004-2010).")
summary(google2);sd(google2)
CV=(sd(google2)/mean(google2))*100;CV;sd(google2)

# Decompondo a serie
plot(decompose(google2),xlab="Ano",ylab=c("obe","sas","sss","sss"))

# Testes para tendência e sazonalidade
cox.stuart.test(google2) # Utilizou implementação fornecida em sala
grupo=c(rep(seq(1,12),6),c(1,2))
kruskal.test(google2,grupo) # Teste de Sazonalidade
```

```

# Primeira Diferenciação
cox.stuart.test(diff(google2,difference=1))
grupo=c(rep(seq(1,12),6),1)
kruskal.test(diff(google2,difference=1),grupo) # Teste de Sazonalidade 1o diferenciação
plot(diff(google2,difference=1))

##### Ajuste por HW #####
ajuste1=HoltWinters(google2) #previsão considerando sazonalidade aditiva
VF1=predict(ajuste1,12)
ajuste2=HoltWinters(google2, seasonal="multiplicative") # previsão considerando
# sazonalidade multiplicativa
VF2=predict(ajuste2,12)

# Gráfico do ajuste pelo HW - comparar os ajustamentos
par(mfrow=c(1,2))
plot(ajuste1,VF1,xlab="Ano",ylab="Fechamento (U$)",
     main="",sub="(a)");lines(google,type="l",lwd=1,col="black")
leg.txt=c("Valor real", "HW-Aditivo"); cores=c("black","red")
legend(list(x=2007.4,y=150), legend = leg.txt, col=cores, lwd=2, merge=FALSE)
plot(ajuste2,VF2,xlab="Ano",ylab="Fechamento (U$)",
     main="",sub="(b)");lines(google,type="l",lwd=1,col="black")
leg.txt=c("Valor real", "HW-Multiplicativo"); cores=c("black","red")
legend(list(x=2007,y=150), legend = leg.txt, col =cores, lwd=2, merge=TRUE)

# Matriz com os valores predito e real
predA= ts(VF1,start=c(2010,10),frequency=12)
predB= ts(VF2,start=c(2010,10),frequency=12)
testHW=data.frame(cbind(google_comp,predA,predB))

##### Ajuste do Modelo SARIMA #####
### Ajuste do modelo ARIMA(p,d,q)x(P,D,Q)

arima.bic=arima
fix(arima)

y=google2 # Série utilizada
lag.entrada=c(1,2) # Variáveis de Entrada
prev=12
l=max(lag.entrada) # Numero máximo de defasagens
n=length(y) # Numero de observações

#### Construindo a matriz de entrada Y
# Constroi a matriz de entrada para os lags especificos antes
Y=matrix(nrow=length(y),ncol=length(lag.entrada))
for(m in 1:length(lag.entrada)){
  e=lag.entrada[m]
  for(i in 1:length(y)){
    if(i<=e)
      for(a in 1:e){Y[a,m]=0}
    else Y[i,m]=y[i-e]
  }
}

```

```

### Ajuste do modelo ARIMA
y.teste=y[(n-prev+1):n] # Dados de teste
X.teste=Y[(n-prev+1):n,]
y=y[(1+1):(n-prev)]
X=Y[(1+1):(n-prev),] ## Matriz do modelo apos eliminaç o dos zeros

##### Seleç o do modelo ARIMA(p,1,q)x(0,1,1)-original #####
#d=1, Com a primeira diferenciaç o obteve a ss rie estacionaria
selecao.de.modelos <- function(serie=y, p.max=2,q.max=2,d=1,P=0,D=0,Q=0){
M<-matrix(0,p.max+1,q.max+1) ## Matriz para armazenar os resultados
if(P==0 && Q==0){
  for(i in 0:p.max){
    for(j in 0:q.max){
      if(i==0 && j==0) M[1,1]<-NA
      else M[i+1,j+1]<-arima(serie, order=c(i,d,j), seasonal=list(order=c(P,D,Q)))$aic #=BIC
    }
  }
} else {
  for(i in 0:p.max) {
    for(j in 0:q.max){
      M[i+1,j+1]<-arima(serie,order=c(i,d,j),seasonal=list(order=c(P,D,Q)))$aic
    }
  }
}
return(M)
}
M=selecao.de.modelos(google2)
p.selec=which(M==min(M),arr.ind=TRUE)[1];p.selec
q.selec=which(M==min(M),arr.ind=TRUE)[2];q.selec

# Modelos Selecionados
# a obtenç o do modelos se deu da seguinte forma:
# atribuiu valores 0,1 para os parametros d,P,Q,D
# e vez todas as possiveis compinaç es para os
# os valores de p,q variando entre 1 e 3.
# e seleciondos os modelos com menor BIC

mod1=arima(google2,order=c(2,0,1),seasonal=list(order=c(1,1,0)))
mod4=arima(google2,order=c(2,1,1),seasonal=list(order=c(1,1,0)));mod4
mod2=arima(google2,order=c(1,1,1),seasonal=list(order=c(1,1,0)));mod2
mod3=arima(google2,order=c(1,1,1),seasonal=list(order=c(1,0,1)));mod3
mod5=arima(google2,order=c(2,1,1),seasonal=list(order=c(1,0,1)));mod5
names(mod2)

## Previs o

y.prev1=predict(mod2,se.fit=FALSE,n.ahead=prev)
y.prev2=predict(mod4,se.fit=FALSE,n.ahead=prev)

###Gr fico da s rie e das previs es

par(mfrow=c(1,2))
# SARIMA 2
a1=mod2[["residuals"]]+google2

```



```

predA1=c(a1,y.prev1s)
predA1s=ts(predA1,start = c(2004,8), frequency = 12)
plot(predA1s,col="2",xlim=c(2004,2012),ylab="Fechamento (U$)",xlab="Ano",sub="(a)")
lines(google); abline(v=2010.9,lty=2)
leg.txt=c("Valor real", "SARIMA2"); cores=c("black","red")
legend(list(x=2004,y=500), legend = leg.txt, col =cores, lwd=2, merge=FALSE)
# SARIMA 4
a2=mod4[["residuals"]]+google2
predA2=c(a2,y.prev2s)
predA2s=ts(predA2,start = c(2004,8), frequency = 12)
plot(predA2s,col="2",xlim=c(2004,2012),ylab="Fechamento (U$)",xlab="Ano",sub="(b)")
lines(google); abline(v=2010.9,lty=2)
leg.txt=c("Valor real ", "SARIMA4 "); cores=c("black","red")
legend(list(x=2004,y=500), legend = leg.txt, col =cores, lwd=2, merge=FALSE)

# Previsão - 12 passos
plot(y.test1,xlab="Ano",lwd=1,ylab="Fechamento (U$)",ylim=c(250,500))
lines(VF1,lwd=1, col="blue")
lines(VF2,lwd=1,col="purple")
lines(y.prev1s,lwd=1,col="red")
lines(y.prev2s,lwd=1,col="orange")
leg.txt=c("Valor real", "Valor predito HW Aditivo",
          "Valor predito HW Mutiplicativo",
          "Valor predito SARIMA 2", "Valor predito SARIMA 4" )
cores=c("black", "blue", "purple","red","orange")
legend(list(x=2011,y=320), legend = leg.txt, col =cores, lwd=1, merge=TRUE)

## Diagnóstico
tsdiag(mod2)
tsdiag(mod4)

# SARIMA
sse1=sum((google_comp-y.prev1s)^2)
mse1=mean((google_comp-y.prev1s)^2)
sse2=sum((google_comp-y.prev2s)^2)
mse2=mean((google_comp-y.prev2s)^2)

#Holt-Winteres
sse.h=sum((google_comp-VF1)^2)
mse.h=mean((google_comp-VF1)^2)
sse.h2=sum((google_comp-VF2)^2)
mse.h2=mean((google_comp-VF2)^2)
aba=matrix(c(sse1,mse1,sse2,mse2,sse.h,
            mse.h,sse.h2,mse.h2),ncol=2,byrow=T)

##### FIM #####

```

Anexo - C

Tabela 7: Base de dados - Fechamento das ações da Google

Ano / Mês	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2004								84,75	103,85	155,47	136,86	141
2005	147,98	142,41	138,95	169,71	221,17	242,13	240,30	232,43	259,98	309,21	343,28	354,35
2006	357,34	311,85	320,71	330,36	291,98	325,25	302,30	296,05	318,55	374,34	369,05	350,18
2007	385,95	341,03	341,65	351,05	375,52	390,39	379,93	376,01	397,78	484,63	481,48	480,81
2008	368,40	308,92	274,35	375,38	379,28	339,99	304,60	321,50	273,80	286,55	226,60	214,70
2009	264,73	263,64	259,65	303,77	292,44	300,00	316,88	322,25	337,84	365,17	386,96	431,20
2010	386,80	387,70	419,48	400,03	395,20	367,30	367,10	355,03	384,85	442,91	431,16	454,50
2011	434,45	442,05	413,27	362,47	364,84	347,00	422,55	376,58	384,82			