

Zilma Silveira Nogueira Reis

**Uso de modelos de previsão para avaliação
do custo da assistência à saúde
na Fundação FIAT**

CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA

Instituto de Ciências Exatas da UFMG

Junho/2008

Uso de modelos de previsão para avaliação do custo da assistência à saúde na Fundação FIAT

Trabalho apresentado ao Curso de Estatística
do ICEX-UFMG para obtenção do título de
Especialista em Estatística.

Aluno: Zilma Silveira Nogueira Reis

Orientador: Prof. Ela Mercedes M de Toscano

CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA

Instituto de Ciências Exatas da UFMG

Junho/2008

RESUMO

A previsão de valores futuros a uma série de valores observados é uma questão importante em diversas áreas, especialmente em economia, administração e também em epidemiologia, pois permite ferramentas para controle e planejamento do processo analisado. Há uma variedade de metodologias de previsão em estatística e é preciso escolher o mais apropriado a um dado conjunto de condições.

A administração de operadoras de saúde, em todo mundo, tem como desafio maior o uso racional de seus recursos econômicos, tendo em vista os custos crescentes de despesas com atendimento médico, internações, emprego de novas tecnologias para diagnóstico e tratamento de doenças. A interface da Medicina com a Estatística pode propiciar novas ferramentas de análise e permitir melhor controle de gastos em saúde.

Neste estudo foram utilizadas as séries históricas do gasto mensal com a saúde de empregados pelas empresas do Grupo FIAT no Brasil, através do gerenciamento realizado pela Fundação FIAT. O período avaliado foi de janeiro de 2002 a dezembro de 2007. Os objetivos da análise foram ajustar um modelo estatístico para explicar o comportamento da série mensal de Custo-saúde e identificar fatores causais no fenômeno analisado. Do ponto de vista prático o modelo deve possibilitar controle mensal do gasto e fazer previsões anuais.

Os modelos utilizados foram do tipo SARIMA e regressão dinâmica com erros correlacionados. Os dois modelos mostraram-se bem ajustados e foram aceitos uma vez que o univariado é capaz de fazer previsões baseadas apenas no passado da série e o modelo dinâmico pode ser usado para fazer previsões em cenários simulados diversos, a partir das variáveis que o compõe. O modelo dinâmico mostrou que as variáveis: indicador de inflação (Fipe-saúde), salários dos profissionais de saúde (Salários) e consumo dos serviços de saúde (Sinistralidade) são capazes de explicar 90,2% da variabilidade da série, ajustado para valores atípicos e erros correlacionados.

Palavras-chave: previsão; modelos ARIMA; regressão dinâmica; gastos em saúde; administração de serviços de saúde

SUMÁRIO

	Página
Agradecimentos	6
Lista de Figuras	7
Lista de Tabelas	9
Lista de abreviaturas	10
1 INTRODUÇÃO	11
1.1 Impacto do custo da assistência médica nas empresas	11
1.2 A Fundação FIAT	11
1.3 Emprego de séries temporais na Medicina	12
1.4 Objetivos	13
2 METODOLOGIA	14
2.1 Conceituação e importância das séries temporais	14
2.2 Características de uma série temporal	15
2.3 Modelos de previsão em séries temporais	15
2.3.1 Modelos como processo estocástico	16
2.3.2 Função de autocorrelação	17
2.3.3 Modelos univariados compostos	20
2.3.4 Modelos causais	22
2.3.5 Validação dos modelos	23
2.4 Metodologia estatística do estudo	24
2.4.1 Estatística descritiva	24
2.4.2 Modelagem estatística	25
2.4.3 Recursos computacionais	26
3 APLICAÇÃO	28
3.1 Escolha das variáveis e coleta de dados	28
3.2 Análise descritiva da série Custo-saúde	29
3.3 Ajuste de modelo univariado para a série Custo-saúde	32
3.4 Ajuste de modelo causal – regressão dinâmica	40

4 DISCUSSÃO	52
5 CONCLUSÕES	54
6 BIBLIOGRAFIA	55
ANEXOS	57
Anexo 1: Função de autocorrelação e autocorrelação parcial da série Custo-saúde	57
Anexo 2: Saídas evIEWS para os modelos univariados	60
Anexo 3: Análise de resíduos dos modelos univariados	64
Anexo 4: Saídas computacionais do modelo dinâmico 1	72
Anexo 5: Saídas computacionais do modelo dinâmico 2	75
Anexo 6: Saídas computacionais dos modelos dinâmicos 3, 4, e 5	78
Anexo 7: Saídas computacionais do modelo dinâmico 6	81

AGRADECIMENTOS

Aos professores do Curso de Especialização em Estatística do Instituto de Ciências Exatas da UFMG, que vêm contribuindo com seriedade e disposição na formação de profissionais das mais diversas áreas, ministrando curso de especialização elevada qualidade.

Ao Hospital das Clínicas da UFMG, em especial ao Professor Henrique Vitor Leite e Antônio Carlos Vieira Cabral que viabilizaram minha participação no Curso.

À Fundação FIAT, em especial ao gestor do sistema de saúde Dr. Alexandre Veloso e à Jacqueline Zoglio e Julis Verne pela disponibilização das informações, confiança e contribuições para tornar o resultado final do trabalho de utilidade na rotina da administração do Sistema de Saúde desta empresa.

A todos os professores e colegas de curso e aos amigos Eura, Alexandre, Ricardo e Augusto pelos momentos compartilhados, incentivo e parcerias.

À mestra e orientadora Professora Ela Mercedes M. de Toscano pela disponibilidade e brilhantismo, agradeço especialmente e expresso minha grande admiração.

À minha família pelo incondicional incentivo.

LISTA DE FIGURAS

Figura 1	Série como processo estocástico como uma família de trajetórias	16
Figura 2	Processo estocástico como uma família de variáveis aleatórias	17
Figura 3	Função de autocorrelação de uma série no programa MINITAB	18
Figura 4	Função de autocorrelação de uma série com componente média móveis	19
Figura 5	Função de autocorrelação parcial de uma série no programa MINITAB	20
Figura 6	Função de autocorrelação parcial de uma série com componente média móveis	20
Figura 7	Gráfico da série Custo-saúde (janeiro 2002 a janeiro de 2007)	29
Figura 8	Descrição da variável Custo-saúde	30
Figura 9	<i>Boxplot</i> da série Custo-saúde, por ano	30
Figura 10	<i>Boxplot</i> da série Custo-saúde, por meses do ano	30
Figura 11	Correlograma fac da série Custo-saúde	30
Figura 12	Correlograma facp da série Custo-saúde	29
Figura 13	Correlograma fac da série Custo-saúde, diferenciada em um nível	31
Figura 14	Correlograma facp da série Custo-saúde, diferenciada em um nível	31
Figura 15	Correlograma fac da série Custo-saúde, diferenciada em dois níveis	32
Figura 16	Correlograma facp da série Custo-saúde, diferenciada em dois níveis	32
Figura 17	Metodologia de modelagem sugerida por Box & Jenkins	32
Figura 18	Série dos resíduos do modelo 1	35
Figura 19	Histograma dos resíduos do modelo 1	35
Figura 20	Série dos resíduos do modelo 2	35
Figura 21	Histograma dos resíduos do modelo 2	35
Figura 22	Série dos resíduos do modelo 3	36
Figura 23	Histograma dos resíduos do modelo 3	36
Figura 24	Série dos resíduos do modelo 4	36
Figura 25	Histograma dos resíduos do modelo 4	36
Figura 26	Zona de predição pelo modelo 1	37
Figura 27	Zona de predição pelo modelo 2	37
Figura 28	Zona de predição pelo modelo 3	37
Figura 29	Zona de predição pelo modelo 4	37
Figura 30	Zona de predição para o ano 2007, pelo modelo 1	38
Figura 31	Zona de predição para o ano 2007, pelo modelo 2	39

Figura 32	Zona de predição para o ano 2007, pelo modelo 3	39
Figura 33	Zona de predição para o ano 2007, pelo modelo 4	39
Figura 34	Gráfico das séries Custo-saúde e vidas, no período de janeiro de 2002 a dezembro de 2007	41
Figura 35	Gráfico das séries Custo-saúde e FIPE saúde, no período de janeiro de 2002 a dezembro de 2007	41
Figura 36	Gráfico das séries Custo-saúde e Sinistralidade, no período de janeiro de 2002 a dezembro de 2007	41
Figura 37	Gráfico das séries Custo-saúde e salários, no período de janeiro de 2002 a dezembro de 2007	42
Figura 38	Metodologia de modelagem para modelos dinâmicos	43
Figura 39	Série dos resíduos do modelo dinâmico 1	45
Figura 40	Histograma de resíduos do modelo dinâmico 1	45
Figura 41	Série dos resíduos do modelo dinâmico 5	47
Figura 42	Histograma dos resíduos do modelo dinâmico 5	47
Figura 43	Correlograma dos resíduos do modelo dinâmico 5	48
Figura 44	Série dos resíduos do modelo dinâmico 6	50
Figura 45	Histograma dos resíduos do modelo dinâmico 6	50
Figura 46	Zona de predição do modelo dinâmico 6, para o período amostral	50
Figura 47	Zona de predição do modelo dinâmico 6, para o ano 2007	51

LISTA DE TABELAS

Tabela 1	Análise dos coeficientes dos melhores modelos univariados para representação da série Custo-saúde	33
Tabela 2	Medidas estatísticas de ajuste dos melhores modelos univariados para representação da série Custo-saúde	34
Tabela 3	Análise de resíduos dos melhores modelos univariados para representação da série Custo-saúde	35
Tabela 4	Medidas estatísticas de avaliação da predição dos melhores modelo univariados para representação da série Custo-saúde	37
Tabela 5	Medidas estatísticas de avaliação da predição dos melhores modelos univariados para representação da série Custo-saúde, entre janeiro a dezembro de 2007	38
Tabela 6	Descrição das variáveis do modelo de regressão dinâmica	40
Tabela 7	Matriz de correlação entre as variáveis, com p-valor para $H_0: \rho=0$	42
Tabela 8	Análise de regressão univariada de cada variável resposta em relação ao Custo-saúde	43
Tabela 9	Modelo de regressão dinâmica 1 para série Custo-saúde	44
Tabela 10	Modelo de regressão dinâmica 2 para série Custo-saúde	45
Tabela 11	Medidas estatísticas de ajuste do modelo dinâmico com a introdução gradativa das <i>dummies</i> 1, 2 e 3	46
Tabela 12	Modelo dinâmico 5 para a série Custo-saúde (tratamento de valores atípicos)	47
Tabela 13	Modelo dinâmico 6 para a série Custo-saúde (tratamento de valores atípicos e erros correlacionados)	49
Tabela 14	Medidas de ajuste dos melhores modelos para a representação da série Custo-saúde	52
Tabela 15	Avaliação comparativa dos parâmetros de previsão dos modelos para o período de janeiro de 2007 a dezembro de 2007	53

LISTA DE ABREVIATURAS

AR	Autoregressivo
AIC	Critério de Informação de Akaike
ARCH	Teste de heterocedasticidade autoregressiva condicional
ARIMA	Autoregressivo integrado e médias móveis
ARMA	Autoregressivo e médias móveis
BIC	Critério de informação Bayesiano
fac	Função de autocorrelação
facp	Função de autocorrelação parcial
facv	Função de autocovariância
MA	Médias móveis
MAE	Erro médio absoluto
MAPE	Porcentagem do erro médio absoluto
R^2	Coeficiente de determinação
R^2_{aj}	Coeficiente de determinação ajustado
RMSE	Erro quadrático médio
RSS	Soma do erro quadrático
SARIMA	ARIMA sazonal
VIF	Fator de inflação da variável

1.1 Impacto do custo da assistência médica nas empresas

Países desenvolvidos investem 8 a 10% de seu Produto Interno Bruto em saúde. O Brasil investe apenas 4% em saúde pública, contra 16 % do investido pelos Estados Unidos da América. Esse valor, nitidamente insuficiente, é apontado como responsável pela falência do sistema público de saúde, sendo as empresas obrigadas a suprir essa carência, oferecendo aos seus empregados os benefícios da saúde privada. Como mais de 80% dos clientes dos planos de saúde são associados através de seus empregadores, o custo da medicina de grupo acaba por incidir sobre o faturamento das empresas em geral. Observa-se que o crescimento das despesas no Brasil supera a incidência mundial já que o excesso de consultas e pedidos de exames é um fenômeno nacional. (MARQUES, 2002). Neste contexto, as empresas precisam desenvolver instrumentos de avaliação e acompanhamento do sistema de saúde adotado para medir o valor efetivamente agregado.

Tendo em vista os custos crescentes de despesas com atendimento médico, internações, emprego de novas tecnologias para diagnóstico e tratamento de doenças, o controle de gastos é um dos grandes desafios das operadoras de saúde em todo mundo. Em especial, as montadoras de veículos automotores se voltam neste momento à questão do mau gerenciamento de recursos com benefícios de seus empregados, tendo como exemplo a situação atual General Motors, que acaba de perder liderança no mercado mundial por este motivo, levando ao fechamento de várias fábricas (LOOMIS, 2006; NEAL et al., 2007).

1.2 A Fundação FIAT e o modelo de assistência à saúde

A Fundação Fiat é entidade sem fins lucrativos que se destina à realização de atividades de assistência médica, social, cultural e esportiva, bem como atividades educacionais, visando ao interesse coletivo e o bem estar de seus beneficiários. Atende a cerca de 60mil pessoas, empregados e dependentes das 15 empresas mantenedoras que compõe o grupo FIAT no Brasil.

A assistência à saúde, um dos benefícios que estas empresas oferecem aos seus empregados de forma gratuita, que consiste em:

- Assistência médica e hospitalar – consultas, exames, internações, terapias.
- Assistência farmacêutica – cobertura de 50% dos gastos com medicamentos.

- Assistência odontológica – tratamento básico gratuito nos centros de saúde e especializado na rede conveniada, com tabela de descontos.

O modelo de assistência à saúde, adotado pela Fundação FIAT, é denominado autogestão patrocinada, plano este regulamentado na Agência Nacional de Saúde. O serviço de auditoria médica e regulação atua gerenciando o custo da assistência médica através da auditoria de contas referentes ao gasto em saúde e fazendo cumprir regras do plano de saúde. Através de sistema informatizado, o Benner Saúde, todos os valores a serem pagos são processados e permanecem armazenados em bancos de dados próprios. O acesso às informações de consumo de serviços de saúde, quantidade e custo, é então possível e corresponde ao total de recursos efetivamente dispendido na modalidade plano-padrão.

1.3 Emprego das séries temporais em Medicina

É cada vez mais freqüente o uso de técnicas estatísticas para auxiliar a tomada de decisão em saúde. E é justamente dentro de um ambiente de pesquisa multidisciplinar, onde as áreas biomédicas e a estatística trabalham juntas, que muitas questões de interesse comum poderão ser adequadamente trabalhadas.

Recentemente o Ministério da Saúde anunciou o lançamento do Manual de Análises de Séries Temporais que tem como objetivo de auxiliar profissional na área de atenção ambiental atmosférica de interesse para a Saúde (MINISTÉRIO DA SAÚDE, 2008). A idéia é estabelecer as bases do conhecimento da situação de saúde de sua população, no que diz respeito aos agravos respiratórios e cardiovasculares e outras doenças de interesse de acordo com aspectos regionais, frente à exposição das populações aos poluentes atmosféricos.

Também o processo de gestão estratégica da informação em saúde deve ser capaz de identificar e descrever as condições, necessidades e resultados de saúde para os clientes (UNIMED 2008). Neste tipo de abordagem a estatística e particularmente a análise de séries temporais são conhecimentos capazes de orientar, com evidências mais bem definidas, a tomada de decisão de gestores do setor público e privado.

- Estudar o comportamento da série temporal Custo-saúde, no período de janeiro de 2002 a dezembro de 2007, referente ao gasto mensal da assistência à saúde oferecida pela Fundação FIAT.
- Elaborar modelo de previsão para o Custo-saúde mensal, baseado em suas características históricas e variáveis causais
- Identificar relação de causa-efeito entre a série de Custo-saúde com outras séries, representativas de fenômenos provavelmente determinantes de sua variabilidade:
 - ✓ número de beneficiários (efeito diluição do risco de adoecer),
 - ✓ indicador de inflação (efeito preço da atividade assistencial),
 - ✓ utilização dos serviços de saúde através da sinistralidade (efeito consumo)
 - ✓ salários dos profissionais de saúde (efeito administração da rede própria de atendimento).

2.1 Conceituação e importância das séries temporais

Uma série temporal, ou série histórica, é um conjunto de observações tomadas regularmente em um período de tempo específico, que apresentam uma dependência serial entre elas. Em sua análise, o fenômeno estudado é interpretado em um modelo matemático que possibilita descrever seu comportamento, fazer estimativas e avaliar os fatores que influenciam sua variabilidade. Também é possível pela modelagem estatística definir relações de causa-efeito entre duas ou mais séries.

A análise de um conjunto de observações X_t , tomadas regularmente em um período de tempo específico t , conjunto esse denominado de série temporal, pode ser extremamente interessante no entendimento de fenômenos periódicos. Uma seqüência de dados obtidos em intervalos regulares de tempo por um período específico, pode ser modelada matematicamente, o que possibilita descrever seu comportamento, procurar periodicidades relevantes, fazer previsões de valores futuros e investigar fatores geradores da série e determinantes de seu comportamento.

Na realidade, o que chamamos de série temporal é uma função amostral, parte de uma trajetória dentre muitas outras que poderiam ter sido observadas, ou seja, é um conjunto de dados que pode ser entendido como um processo estocástico ou probabilístico (CHATFIELD, 2003a). Um modelo é uma descrição probabilística de uma série temporal e cabe ao proponente como elaborá-lo visando à utilização segundo seus objetivos (MORETIN & TOLOI, 2006a). A modelagem estatística é derivada da representação matemática das relações entre variáveis conjuntamente com uma medida inerente de incerteza. Modelos devem ser simples e parcimoniosos, sempre que possível, e de fácil manipulação aos interessados.

A previsão de valores futuros a uma série de valores observados é uma questão importante em diversas áreas, especialmente em economia e administração, pois permite controle e planejamento do processo analisado. Há uma variedade de metodologias de previsão em estatística e é preciso escolher o mais apropriado a um dado conjunto de condições (MORETIN & TOLOI, 2006a).

2.2 Características de uma série temporal

Uma série histórica pode ser composta por três elementos: tendência (T_t), sazonalidade (S_t) e a variação aleatória denominada de ruído branco (a_t).

A tendência se refere ao crescimento ou decrescimento dos valores da série ao longo do tempo. Para analisar a tendência de uma série os procedimentos mais usados são: o ajuste de uma função polinomial ou a análise do comportamento da série ao redor de um ponto, estimando a tendência naquele ponto, através de modelos auto-regressivos (LATORRE e CARDOSO, 2001).

Define-se um fenômeno sazonal como aquele que ocorre regularmente em períodos fixos de tempo, geralmente com periodicidade inferior a um ano. Se houver uma sazonalidade determinística podem ser utilizados modelos de regressão que incorporem funções do tipo seno ou cosseno à variável tempo (LATORRE e CARDOSO, 2001).

Antes de se conduzir uma análise é importante definir se a série é estacionária ou não, para estabelecer a estrutura do modelo probabilístico que estimará a série. Uma série é considerada estacionária quando suas observações ocorrem aleatoriamente em torno de uma média constante. Ou seja, não há tendência ou sazonalidade. Para isto, determina-se a função de auto-correlação da série, calculando-se o coeficiente de correlação entre as observações no instante t e $t+j$. Cada período da série é chamado j (*lag j*). Na série estacionária a correlação entre as observações é zero, se elas são realmente independentes (LATORRE e CARDOSO, 2001).

Um modelo mais simples, livre de tendência e sazonalidade, pode ser obtido nas séries estacionárias, uma vez que ela é consequência da variação aleatória do ruído branco ao redor de uma média.

2.3 Modelos de previsão em séries temporais

Uma parte importante da análise de uma série temporal é a seleção de um modelo probabilístico aceitável para os dados. Qualquer que seja a classificação que se faça para os modelos de séries temporais, podemos considerar um número muito grande de modelos diferentes para descrever o comportamento de uma série particular (MORETTIN e TOLOI, 2006a). De maneira prática, os modelos de previsão em série temporais podem ser classificados em univariados, multivariados e causais:

- Nos modelos univariados, a previsão dos valores futuros é explicada somente pelos valores do passado da série. As metodologias mais usadas são a de decomposição, alisamento exponencial e os baseados na média. Esses últimos são baseados na teoria geral dos sistemas lineares e são robustos e acurados para previsões. Englobam os modelos ARMA (p,q), ARIMA (p,d,q), SARIMA (p,d,q)(P,D,Q)s, entre outros (TOSCANO 2007).
- Muitas séries, na prática, são melhores consideradas como multivariadas, pois apresentam não apenas dependência serial dentro de cada componente da série, mas também interdependência entre diferentes séries componentes. Muito da teoria univariada se estende naturalmente à multivariada, mas novos problemas surgem. Nos modelos multivariados as previsões dos valores futuros k séries é explicada somente pelos valores do passado das séries (BROCKWELL e DAVIS, 2002b).
- Modelos causais trazem a previsão de uma variável a partir da interpretação de outras variáveis relacionadas. A relação causa-efeito permite antecipar mudanças significativas na série temporal. As metodologias mais usadas são: regressão dinâmica e modelos ARIMA (TOSCANO 2007).

2.3.1 Modelos como processo estocástico

Os modelos utilizados para descrever séries temporais são processos estocásticos, isto é, controlados por leis probabilísticas. O processo estocástico nada mais é do o conjunto de todas as possíveis trajetórias que poderíamos observar em um fenômeno, como representado na Figura 1 (MORETTIN e TOLOI, 2006b).

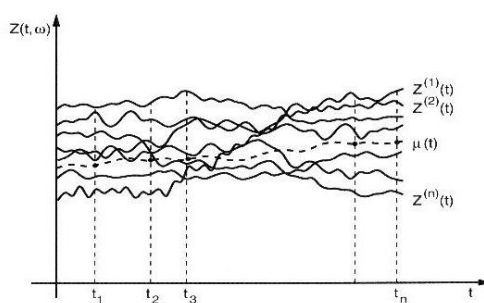


Figura 1 – Série como processo estocástico como uma família de trajetórias

Seja T um conjunto arbitrário. Um processo estocástico é uma família $Z = \{Z(t), t \in T\}$, tal que para cada $t \in T$, $Z(t)$ é uma variável aleatória. Sendo assim, na realidade $Z(t)$ é uma função de dois argumentos definida sobre Ω , $Z(t, \omega)$, $t \in T$, $\omega \in \Omega$, como ilustrado na Figura 2.

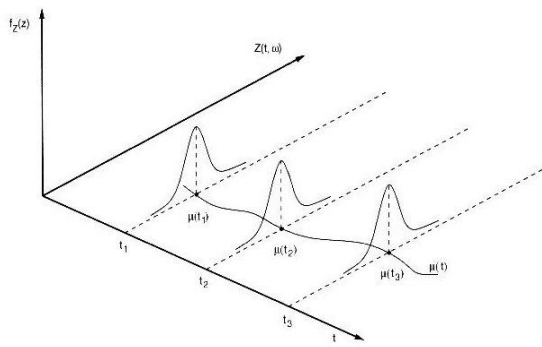


Figura 2 – Processo estocástico como uma família de variáveis aleatórias

Processos estocásticos são chamados de estacionários quando se a distribuição de probabilidade não muda com o tempo e portanto:

$$E(Z_t) = \mu(t) \text{ constante para todo } t$$

$$\text{Var}(Z_t) = \sigma^2 \text{ é finita e constante para todo } t$$

$$\text{Cov}(Z_t, Z_s) = \gamma_k = |t - s| \text{ função que só depende da diferença do tempo} \\ (\text{TOSCANO, 2007}).$$

Nas situações em que se pretende usar modelos para descrever séries temporais é necessário assumir suposições simplificadoras que conduzam à análise (MORETTIN e TOLOI, 2006a):

- Processos estacionários ou não, de acordo com a independência ou não relativamente à origem dos tempos;
- Processos normais (Gaussianos) ou não normais de acordo com as funções de distribuição que caracterizam os processos;
- Processos Markovianos ou não-Markovianos, de acordo com a independência dos valores do processo.

2.3.2 Função de autocorrelação

As relações entre as observações da mesma série entre diferentes instantes de tempo Z_t e Z_{t-k} são obtidas pela função de autocovariância e autocorrelação. Essas funções são guias importantes para o conhecimento das propriedades de uma série pois medem a correlação, se presente, entre observações de uma série a diferentes intervalos de tempo, provendo informações descritivas (CHATFIELD, 2004a). Dados N pares de observações Z_t e Z_{t-k} , a função de autocovariância (facv) e a função de autocorrelação (fac) são dadas por:

$$\sum_{j=1}^n \sum_{k=1}^n a_j a_k \gamma_{\vec{q}-\tau\kappa} \geq 0,$$

para quaisquer números reais a_1, \dots, a_n , e τ_1, \dots, τ_n de Z .

Enquanto a fac do processo é definida por:

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0}, \tau \in Z$$

(MORETTIN e TOLOI, 2006b).

Para melhor interpretação dos coeficientes de autocorrelação de uma série utiliza-se o gráfico denominado correlograma (Figura 1). Nesse gráfico, os coeficientes de autocorrelação amostral R_k são plotados contra o *lag* k para $k = 0, 1, \dots, M$, onde M é usualmente muito menor que N , geralmente $N/4$ (CHATFIELD, 2004a). No gráfico da Figura 3, o programa MINITAB apresenta as autocorrelações e estatísticas t e Ljung-Box Q , com bandas pontilhadas correspondentes a probabilidade de significância de 5% para rejeitar a hipótese de que as correlações são iguais à zero (MINITAB 2004).

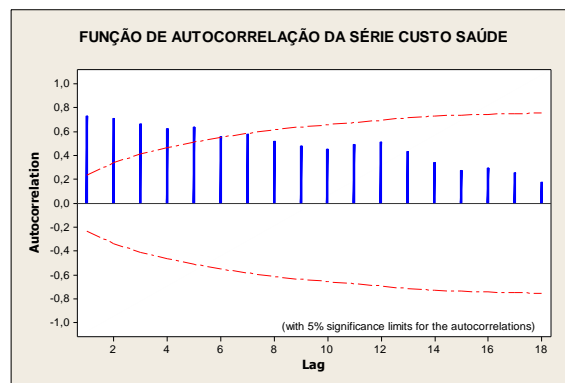


Figura 3 – Função de autocorrelação de uma série no programa MINITAB

Para avaliar o grau de dependência nos dados amostrais e selecionar um modelo que reflita esses dados, a fac é uma importante ferramenta. Se acreditarmos que os dados representam valores de uma série estacionária $Z(t)$, então a fac amostral será uma estimativa da série $Z(t)$. Por exemplo, quando a fac amostral é próxima de zero, para todas os *lags* não-zero, sugere-se que o modelo apropriado para os dados deve ser ruído branco (BROCKWELL e DAVIS, 2002a). Um processo autoregressivo $AR(p)$ tem fac que decai de acordo com exponenciais e ou senóides amortecidas (Figura 1), enquanto um processo médias móveis $MA(q)$ tem fac finita, no sentido que ela apresenta um corte abrupto após o *lag* q , ilustrado na Figura 2. Um processo $ARMA(p,q)$ tem fac infinita, a qual decai em extensão, de acordo com exponenciais e/ou senóides amortecidas após o *lag* $q-p$ (MORETTIN e TOLOI, 2006c). Se a série apresenta variação sazonal, então também o correlograma irá apresentar oscilações na mesma frequência (CHATFIELD, 2004a).

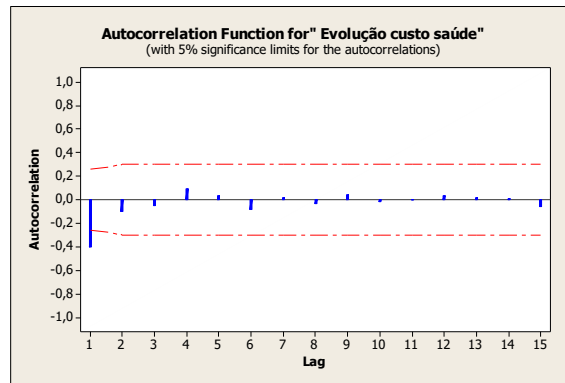


Figura 4 – Função de autocorrelação de uma série com componente médias móveis

No entanto, Box, Jenkins e Reinsel (BOX & JENKINS, 1976) propõem a utilização de um outro instrumento para facilitar a identificação do modelo, a função de autocorrelação parcial (facp). Esta função avalia a intensidade da relação entre dois valores de uma série, mantendo sob controle os demais. Seja ϕ_{kj} o j -ésimo coeficiente de um modelo $AR(k)$, de tal modo que $\phi_{\kappa\kappa}$ seja o último coeficiente e P_{κ} a matriz de autocorrelações e P_{κ}^* é a matriz P_{κ} com a última coluna substituída pelo vetor de autocorrelações:

$$\phi_{\kappa\kappa} = \frac{|P_{\kappa}^*|}{|P_{\kappa}|},$$

A quantidade $\phi_{\kappa\kappa}$, em função de κ , é chamada função de autocorrelação parcial (facp). Uma maneira de obtê-la consiste em estimar, sucessivamente, modelos auto-regressivos de ordens $p = 1, 2, 3, \dots$ por mínimos quadrados e tomar as estimativas do último coeficiente de cada ordem (MORETTIN e TOLOI, 2006c). O facp é útil para identificar modelos $AR(p)$, que tem $\text{facp} \neq 0$, para $\kappa \leq p$ e $\text{facp} = 0$, para $\kappa > p$, ilustrado na Figura 3. Um processo $MA(q)$ tem facp que se comporta de maneira similar à fac de um processo $AR(p)$, composta por exponenciais e/ou senóides amortecidas (Figura 4). Um processo $ARMA(p, q)$ tem facp que se comporta como a facp de um processo MA puro (MORETTIN e TOLOI, 2006c).

O MINITAB apresenta no correlograma do facp, as autocorrelações com estatísticas t e Ljung-Box Q, com bandas pontilhadas correspondentes a probabilidade de significância de 5% para rejeitar o teste de hipóteses de as correlações são iguais à zero (*Help* MINITAB 14.12.0).

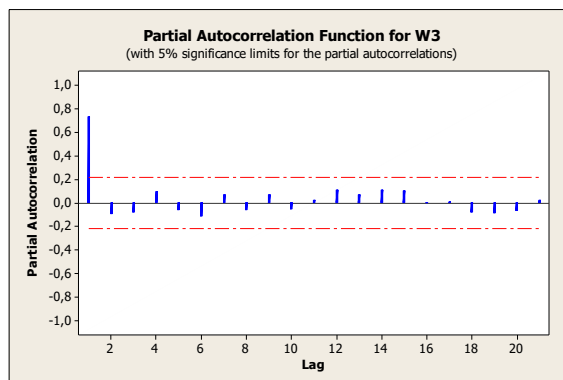


Figura 5 – Função de autocorrelação parcial de uma série no programa MINITAB

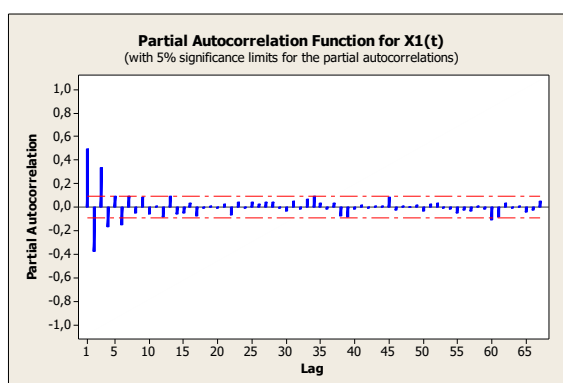


Figura 6 - Função de autocorrelação parcial de uma série com componente médias móveis

2.3.3 Modelos univariados compostos

Para muitas séries, a melhor solução para se encontrar a dependência serial está na combinação linear de valores próximos do passado da série, representação auto-regressiva de ordem p , com uma combinação linear dos ruídos brancos próximos ao valor da série, representação média móvel de ordem q (LATORRE e CARDOSO, 2001).

Tanto modelos auto-regressivos, como médias móveis ou ARMA são usados para séries estacionárias. No entanto, se o processo é não estacionário explosivo ou o chamado não-estacionário homogêneo, é incorporado um processo de diferenças ($D_d Z$) na série para aplicar os modelos ARMA. Este modelo é conhecido como auto-regressivo integrado de médias móveis, onde d é a ordem das diferenças necessárias para anular a tendência da série (LATORRE e CARDOSO, 2001).

Os modelos ARIMA podem também resolver a sazonalidade quando há lags de baixa ordem. No entanto, se a sazonalidade ocorre em múltiplos períodos é necessário considerar uma sazonalidade estocástica. Para isto, emprega-se os modelos SARIMA, que incorpora funções trigonométricas (geralmente seno e cosseno) ao modelo ARIMA e cuja ordem da sazonalidade pode ser estimada $(P,Q,Q)s$ (LATORRE e CARDOSO, 2001).

A Notação estatística a ser usada nos modelos univariados compostos é de grande importância para sua representação matemática. Para muitas séries, se quisermos um modelo com um número menor de parâmetros, a inclusão de termos auto-regressivos $AR(p)$ e de médias móveis $MA(q)$ é a solução adequada $ARMA(p,q)$. Um modelo $ARMA(1,1)$ é representado pela equação:

$$Y_t = \theta_0 + \phi Y_{t-1} + a_t + \theta a_{t-1}, \quad t = 1, 2, \dots, T$$

onde:

- $\theta_0 = \mu(1 - \phi)$, μ é a média do processo;
- O processo é considerado estacionário se $-1 < \phi < 1$;
- O processo é invertível se $-1 < \theta < 1$;
- $\{a_t\}$ é um ruído branco, ou seja, é uma série de observações com média zero ($E(a_t) = 0$) e variância constante $Var(a_t) = \sigma^2$.

Um modelo $ARIMA(p,d,q)$ é apropriado para representar séries cujo comportamento seja não-estacionário, mas não explosivo. As séries Z_t , tomadas em um número finito de diferenças, d , tornam-se estacionárias, sendo chamadas não-estacionárias homogêneas. A equação representativa de um modelo $ARIMA$ é dada por:

$$\Phi(B)(1-B)^d Y_t = \Theta(B)a_t, \quad t = 1, 2, \dots, T$$

sendo que p e q são as ordens dos operadores $\Phi(B)$ e $\Theta(B)$, respectivamente. O modelo pressupõe que a d -ésima diferença da série Z_t pode ser representada por um modelo $ARMA$, estacionário e invertível. Na maioria dos casos $d = 1$ ou $d = 2$.

Quando $\{Z_t\}$ exibe um comportamento sazonal determinístico com período definido, um modelo que pode ser útil é: $Z_t = \mu_t + \varepsilon_t$, $t = 1, 2, \dots, T$

onde μ_t é uma função determinística periódica, satisfazendo $\mu_t - \mu_{t-12} = 0$, ou

$$(1 - B^{12})\mu_t = 0 \text{ e } \varepsilon_t \text{ é um processo estacionário que pode ser modelado por um } ARMA(p,q).$$

Mas como μ_t é um processo estocástico, satisfaz as condições de um processo estacionário Y_t :

$$(1 - B^{12})\mu_t = Y_t,$$

Aplicando-se o operador $(1 - B^{12})$ à equação inicial,

$$(1 - B^{12})Z_t = (1 - B^{12})\mu_t + (1 - B^{12})\varepsilon_t$$

Pode-se demonstrar que essa expressão é equivalente à:

$$\Phi(B^{12})\Delta_{12}^D Z_t = \Theta(B^{12})a_t,$$

onde: $\Phi(B^{12})$ é o operador auto-regressivo sazonal de ordem P , estacionário; $\Theta(B^{12})$ é o operador de médias móveis sazonal de ordem Q , invertível; $\Delta_{12}^D Z_t = (1 - B^{12})Z_t^\Delta$ é o operador diferença sazonal

onde D indica o número de diferenças sazonais. Se o processo ε_t satisfaz um modelo ARIMA(p,d,q), o modelo é denominado SARIMA, ou ARIMA sazonal multiplicativo de ordem $(p,d,q)(P,D,Q)$ 12:

$$\Phi(B_p^{12})\Phi(B_p)Z_t = \theta(B_q)\Theta(B_Q^{12})a_t \quad t = 1, 2, \dots, T$$

$$\Phi(B_p^{12})\Phi = 1 - \Theta_1 B^{12} - \Theta_2 B^{24} - \dots - \Theta_{12} B^{112}$$

$$\Theta(B_Q^{12}) = 1 - \Theta B^{12} - \Theta_2 B^{24} - \dots - \Theta_{12} B^{112}$$

a_t é um processo de ruído branco

2.3.4 Modelos causais

Neste grupo de modelos a variável dependente ou resposta (Y) é tomada por contagem e as variáveis independentes são as candidatas a explicar o comportamento da série ao longo do tempo.

Os modelos de regressão dinâmica combinam a dinâmica de séries temporais e o efeito de variáveis explicativas. Atenta-se que o termo “regressão dinâmica” está relacionado à estrutura de defasagem da série temporal e das variáveis causais do modelo (TOSCANO, 2007). Isto não indica que os parâmetros do modelo evoluem no tempo, mas é um modelo de regressão no qual se inclui a estrutura de dependência de uma série temporal. Modelos de regressão dinâmica devem ser usados quando existe uma estrutura de dependência entre a variável de interesse e possíveis variáveis causais e, ao mesmo tempo, quando a estrutura de correlação da série dependente indicar que não se pode supor a independência dos erros. Nesses modelos, a variável dependente é explicada por seus valores defasados e pelos valores atuais e passados de variáveis causais ou exógenas (ZANINI, 2000).

A modelagem da sazonalidade pode ser feita através de variáveis dummies sazonais, defasagens sazonais nas variáveis dependentes ou erros estruturados sazonalmente ($\alpha_{t-s}, \alpha_{t-2s}, \dots$), (TOSCANO, 2007).

O processo de construção do modelo dinâmico é a partir de um modelo simples, verificação dos parâmetros de ajuste, incremento de defasagens, dummies, se necessário, até um mais elaborado em que seja possível verificar ruído branco.

A estrutura do modelo de regressão dinâmica é dada por:

$$\Phi(B)Y_t = \beta X_t + a_t, \quad t = 1, 2, \dots, T, \text{ onde:}$$

Y_t representa a variável (endógena) dependente;

X_t é um vetor de variáveis (exógenas) causais;

β é o vetor de coeficientes das variáveis exógenas;

$\Phi(B) = (Y_t - \Phi_1 Y_{t-1} - \Phi_2 Y_{t-2} - \dots - \Phi_p Y_{t-p})$ é um polinômio autoregressivo de ordem p ; a_t

são os erros i.i.d. Normal $(0, \sigma^2)$. Neste modelo é possível escrever uma série em função de seu passado

e outras variáveis explicativas, defasadas ou não, admite-se a existência de erros estruturados (TOSCANO, 2007).

2.3.5 Validação dos modelos

Após estimar o modelo temos que verificar se ele representa adequadamente os dados. Insuficiências reveladas nos testes podem sugerir modelos alternativos mais adequados (MORETTIN e TOLOI, 2006e). Para isto, empregam-se testes estatísticos que avaliam a adequação do modelo:

- Estatística de Durbin Watson: originalmente proposta para uso em modelos de regressão múltipla, tem sua aplicação também nas séries temporais (CHATFIELD, 2004b). O teste avalia a presença de autocorrelação nos resíduos do modelo. Se a_t é um resíduo associado a uma observação no tempo t , então o teste será:

$$d = \frac{\sum_{t=2}^T (a_t - a_{t-1})^2}{\sum_{t=1}^T a_t^2}, \text{ cujo valor está compreendido entre zero e quatro.}$$

Um valor 2, ou próximo, indica que aparentemente não há autocorrelação nos resíduos do modelo, enquanto um valor substancialmente menor que 2 indica a presença de autocorrelação (WIKIPEDIA, 2008).

- Teste de Portmanteau ou Ljung-Box(Q): baseado na função de autocorrelação de uma série, avalia a hipótese de que as k primeiras autocorrelações são nulas (TOSCANO 2007):

$H_0: \rho_1 = \rho_2 = \rho_3 = \dots = \rho_k = 0$, sendo a estatística do teste dada pela equação:

$$Q = \frac{T(T+2) \sum_{t=1}^T (r_t^2)}{(T-1)} \approx \chi^2_{(k-m)}, \text{ onde } m \text{ é o número de parâmetros livres do modelo.}$$

Rejeita-se H_0 de ruído branco para os resíduos, quando Q é grande quando comparado a um percentil apropriado. Em geral, basta utilizar as 10 ou 15 primeiras autocorrelações da série de resíduos (MORETTIN e TOLOI, 2006d).

Para se comparar modelos diferentes e escolher o mais adequado, são empregados critérios e testes estatísticos, sendo os mais usados:

- R^2 : o coeficiente de determinação é a proporção da variabilidade dos dados explicada pelo modelo:

$$R^2 = SS_{regressão} / SS_{total}$$

Esse coeficiente leva em conta o número de parâmetros do modelo e por isto é altamente influenciado por eles, sendo limitado na comparação de modelos com número de parâmetros diferentes. O R^2 ajustado para o número de termos explicativos aumenta somente se um novo termo melhora o

modelo, mais do que esperado aleatoriamente, sendo útil para comparar estágios na construção de um modelo (WIKIPEDIA 2008); é definido como:

$$R^2_{aj} = (1 - R^2) \frac{T-1}{T-1-p}, \text{ onde } T \text{ é o tamanho da série e } p \text{ o número de parâmetros da}$$

regressão.

Alternativas mais sofisticadas para seleção de modelos, que penalizam o excesso de parâmetros são o AIC, BIC e Schwartz (CHATFIELD, 2004c).

- Critério de Informação de Akaike (AIC): avalia a bondade de um modelo. Baseia-se no conceito da entropia, oferecendo uma medida relativa da informação perdida em um modelo usado para descrever uma realidade (WIKIPEDIA 2008). Na seleção de um modelo, o menor valor do AIC indica o mais adequado, pois se trata de uma função de probabilidade sujeita a um termo que penaliza o número de parâmetros, para prevenir o superajustamento. Por outro lado, AIC pode ser viesado para pequenas amostras (CHATFIELD, 2004c). É definido como:

$$AIC = T \ln(\hat{\sigma}_\varepsilon^2) + 2m, \text{ onde } m \text{ é o número de parâmetros estimados e } \hat{\sigma}_\varepsilon^2 \text{ é a variância do vetor de resíduos, } \ln(\cdot) \text{ é a função logaritmo neperiano.}$$

- Critério Bayesiano de Informação (BIC): trata-se de uma alternativa ao AIC que substitui seu termo final ($2m$) pela expressão $(m + m \ln T)$, onde T é o número de observações. Com isto, a adição de parâmetros extras é mais severamente penalizada (CHATFIELD, 2004c).

- Critério Bayesiano de Schwartz: nesse critério, outra alternativa ao AIC, similar a ele pela dependência em $\ln(N)$, mas substitui o termo por $m \ln(T)$ (CHATFIELD, 2004c), isto é:

$$AIC = T \ln(\hat{\sigma}_\varepsilon^2) + m \ln(T)$$

Em quaisquer dos três critérios, AIC, BIC e Schwartz, o modelo mais adequado é aquele que possui menores valores para os critérios.

2.4 Metodologia estatística do estudo

2.4.1 Estatística descritiva

Todas as séries mensais, Custo-saúde, Vidas, Sinistralidade, Salários e FIPE saúde, tomadas no período de janeiro de 2002 a dezembro de 2007, serão exploradas quanto ao seu comportamento (tendência, sazonalidade, ruído, variações cíclicas).

A série Custo-saúde será também avaliada quanto às suas funções de autocorrelação e autocorrelação parcial para orientar a escolha dos modelos mais adequados.

As correlações entre a série Custo-saúde (variável dependente) e as possíveis preditoras (variáveis independentes) serão obtidas individualmente, assim como a matriz de correlação entre as cinco variáveis.

2.4.2 Modelagem estatística

Para modelagem de modelos univariados as condições de estacionariedade do processo estocástico será inicialmente avaliada. Em seguida será empregado procedimento padrão para utilização da metodologia de previsão ARIMA pela metodologia Box & Jenkins (1976):

1. Realiza-se a diferenciação da série original tantas vezes quanto necessário para torná-la estacionária, não sendo recomendável ultrapassar duas vezes;

2. Identifica-se os valores p e q através da análise das funções de autocorrelação (ACF) e de autocorrelação parcial (PACF) estimadas e a estimação dos parâmetros (ϕ_i , $i = 1, \dots, p$ e θ_i , $i = 1, \dots, q$). A análise destas funções permite a tomada de decisão acerca do número p de parâmetros auto-regressivos e do número q de parâmetros de médias móveis que devem ser escolhidos de maneira a se obter uma estrutura parcimoniosa (ou seja, uma estrutura que tenha o menor número de parâmetros dentre todas as estruturas que se ajustem aos dados da série);

3. Avalia-se a adequação do modelo através de análise de resíduos: a diferença entre o valor observado e o ajustado pelo modelo será verificada quanto à normalidade e homocedasticidade, com objetivo de se obter ruído branco. O teste de Durbin Watson será empregado para detectar autocorrelações de primeira ordem nos resíduos e um valor próximo de 2 considerado satisfatório.

4. Realizam-se as previsões (obtenção dos novos valores da série) e os intervalos de confiança para as mesmas.

Avaliação comparativa entre os modelos: os modelos melhores ajustados serão apresentados e comparados entre si quanto ao ajuste: coeficiente de determinação, AIC, Schwartz e BIC e também quanto a suas qualidades preditivas pelo MAE, MAPE, RMSE e coeficiente de desigualdade.

Para os modelos causais, a relação de causa-efeito será avaliada através de regressão dinâmica com estratégia *bottom-up*, isto é, considera-se inicialmente um modelo simples para depois melhorá-lo e incluir novas variáveis até encontrar um modelo apropriado. Os mesmos testes estatísticos serão utilizados para mensurar o grau de adequação dos modelos.

2.3.4 Recursos computacionais

O desenvolvimento de programas estatísticos tem tornado a utilização do conhecimento estatístico mais expressiva e rápida. Nesta análise serão empregados: programa MINITAB® Release 14.12.0 e Eviews 4.1.

O *software* Econometric Views (Eviews) é um pacote estatístico desenvolvido com um conjunto de ferramentas para manipular dados de séries temporais (coletados ao longo do tempo). Foi desenvolvido por David M. Lilien, Richard Startz, Scott Ellsworth, Jaesun Noh e Robert Engle, da *Quantitative Micro Software*, localizada em Irvine, California. Embora o EViews tenha sido desenvolvido por economistas e com maioria de aplicações na Economia, não há nada que o impeça de ser usado em outras áreas. As mais importantes funções que o programa EViews (TOSCANO, 2007):

- Cálculo de uma série, baseada em fórmula de qualquer complexidade;
- Estatística descritiva, gráficos de linha, gráficos de barra e de torta; diagramas de dispersão;
- Produz Regressões através dos Métodos de Mínimos Quadrados Ordinários, Mínimos Quadrados de Dois e Três Estágios;
- Estimação de Funções Não-Lineares;
- Estimação dos modelos de escolha binária Probit, Logit e Tobit;
- Estimação linear e não linear de sistemas de equações;
- Combinação e estimação de dados séries temporal e cross-section;
- Estimação de Modelos ARCH-GARCH;
- Estimação e análise de sistemas de vetor autorregressivo;
- Distribuição Polinomial;
- Solução de modelos simultâneos; etc.

O Programa MINITAB (<http://minitabbrasil.com.br/minitab/vantagens.asp>, 2008) analisa séries temporais através dos procedimentos:

- Gráficos de séries temporais;
- Autocorrelação, autocorrelação parcial e autocorrelação cruzada;
- Análise de modelos ARIMA;
- Análise de tendência
- Decomposição
- Suavização exponencial
- Método de Winter

- Médias móveis

Em seus processos destacam-se:

Diferenças: calcula e armazena diferenças entre valores de uma série temporal. A diferenciação é geralmente usada para simplificar a estrutura de correlação ou revelar padrões, preparando as séries com tendência e sazonalidade para ajuste de modelos ARIMA.

Lag: calcula e armazena intervalos de tempo de uma série. Para isso, o MINITAB move valores originais para baixo da coluna e insere valores perdidos no alto da coluna. O número de valores perdidos inseridos depende do tamanho do *lag*.

Autocorrelação: calcula e faz gráficos de autocorrelações de uma série (FAC), revelando a correlação existente entre observações de uma série que estão separadas por k unidades de tempo. O procedimento é fundamental para escolha dos termos do modelo ARIMA.

Autocorrelação parcial: calcula e faz gráficos de autocorrelações parciais de uma série (FACP), revelando a correlação existente entre conjuntos de dados de uma série. A FACP a um *lag* k é a correlação entre resíduos em um tempo t de um modelo autoregressivo.

Correlação cruzada: processa e faz gráficos de correlação entre duas séries temporais.

ARIMA: ajusta modelos ARIMA de (BOX & JENKINS, 1976) filtrando passos até obter ruído aleatório.

3.1 Escolha das variáveis e coleta de dados

- **Custo-saúde:** se refere ao valor mensal total gasto pela Fundação Fiat com seus beneficiários (empregados e dependentes), registrado no livro razão da empresa. Os valores, tomados em Reais, somam o gasto mensal da Fundação FIAT com consultas médicas e procedimentos ambulatoriais em sua rede credenciada, internações, exames, terapias (fisioterapias, fonoaudiologia, quimioterapia), custo administrativos com sua rede própria de atendimento e serviço de auditoria médica. Não entram na composição do custo as despesas relativas a pagamento de reembolso médico ou odontológico e atendimento dos escritórios regionais através do plano Saúde-Bradesco.
- **Número de vidas (Vidas):** quantidade de beneficiários do sistema de saúde, empregados e dependentes, oferecido pelas empresas do grupo Fiat, computado no último dia do mês.
- **Eventos em saúde (Sinistralidade):** soma do número de ocorrências pagas pela utilização da rede credenciada do sistema de saúde, em um mês (consultas, exames, internações, atendimentos ambulatoriais), registradas no sistema de informação adotado pela Fundação FIAT. Representa a utilização dos serviços credenciados pela carteira de beneficiários.
- **Salários:** valor mensal gasto pela Fundação FIAT no pagamento de salários e outras despesas trabalhistas aos profissionais de saúde em sua rede própria de atendimento, registrado no livro razão da empresa.
- **Inflação (FIPE-saúde):** o indicador escolhido foi o IPC-FIPE SAÚDE da Fundação Getúlio Vargas e a série utilizada obtida no site: www.fgv.br.

A coleta das informações utilizadas no estudo foi realizada por técnico do Serviço de auditoria médica, a partir de sistema informatizado de gerenciamento em saúde adotado pela Fundação FIAT, o Benner Saúde.

3.2 Análise descritiva da série Custo-saúde

A série Custo-saúde, tomada no período de janeiro de 2002 a dezembro de 2007, com periodicidade mensal foi apresentada graficamente na Figura 5. A composição do Custo-saúde se refere aos valores gastos, tomados em Reais, com internações, exames complementares, terapias, consultas médicas e procedimentos ambulatoriais realizados na rede credenciada pela Fundação FIAT e com custos administrativos próprios.

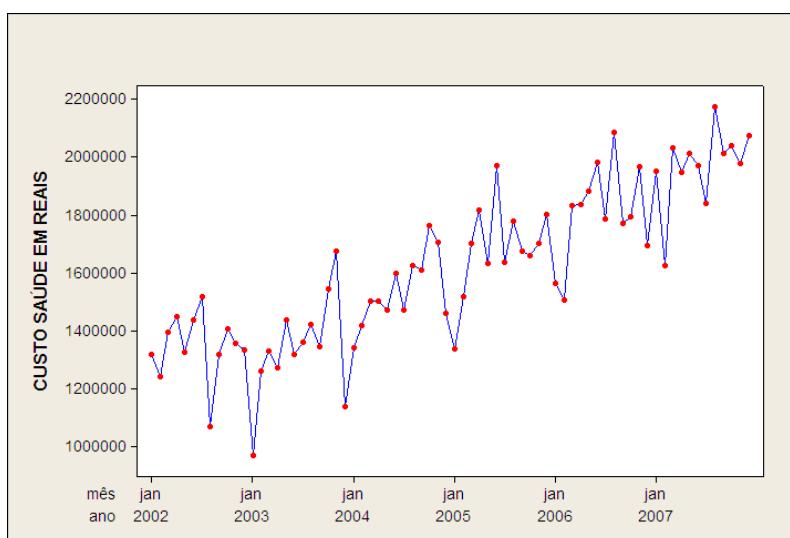


Figura 7 – Gráfico da série Custo-saúde (janeiro 2002 a janeiro de 2007)

A variável Custo-saúde apresentou valor médio mensal de R\$1.616.040,00, com desvio padrão de R\$ 273.974,00. Os demais parâmetros descritivos da variável se encontram na Figura 6. A variável apresenta evidência de distribuição normal pelo teste de Anderson Darling que não rejeita a hipótese nula de que a distribuição segue padrão gaussiano ($p\text{-valor} = 0,096$).

Padrões de comportamento da série Custo-saúde: ao longo dos 6 anos, a série apresenta tendência crescente (Figura 7). Observam-se também características de leve sazonalidade (Figura 8). Há um crescimento gradativo do custo de janeiro a março, estabilização dos valores até outubro, seguido de queda em novembro e dezembro. O mês que apresenta maior variabilidade é agosto. Aparentemente não há outras variações cíclicas no período estudado.

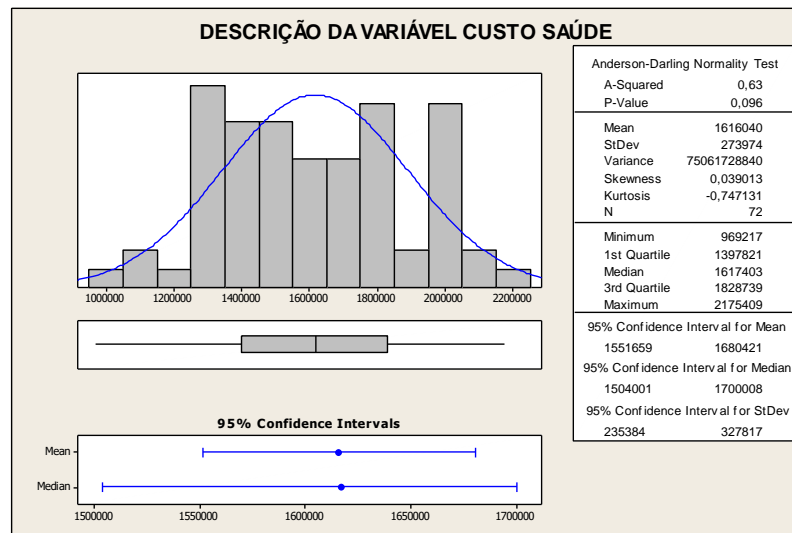


Figura 8 - Descrição da variável Custo-saúde

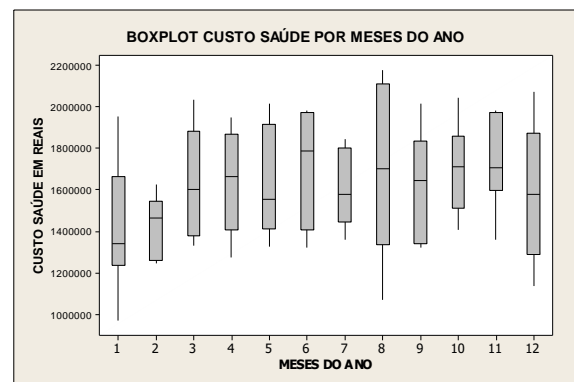
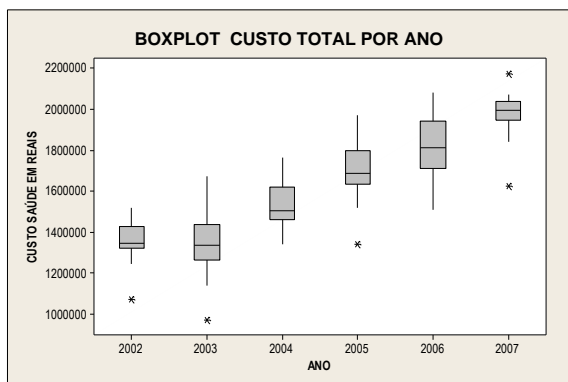


Figura 9 – *Boxplot* da série Custo-saúde, por ano

Figura 10 – *Boxplot* da série Custo-saúde, por meses do ano

A função de autocorrelação entre as observações da série nos instantes de tempo t e $t+k$ é apresentada no correlograma fac (Figura 9) que mostrou decaimento lento com autocorrelação até o quinto valor. A autocorrelação entre observações em diferentes instantes foi apresentada na facp, que evidenciou duas espículas iniciais, com decaimento alternado (Figura 10).

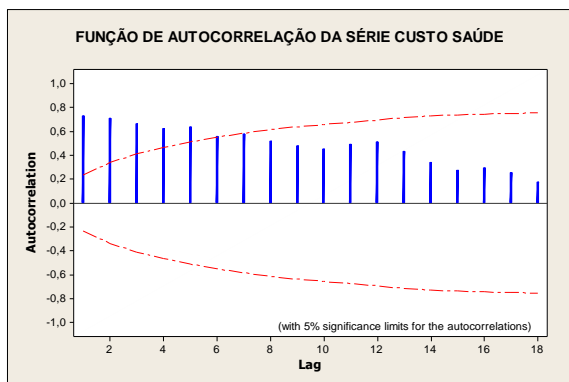


Figura 11 – Correlograma fac da série Custo-saúde

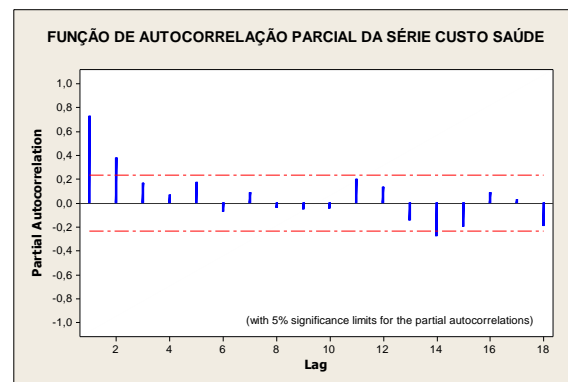


Figura 12 – Correlograma facp da série Custo-saúde

Nos gráficos acima (Figuras 9 e 10), que trazem os correlogramas até a *lag* 18, os intervalos de confiança das correlações estão representados em pontilhado. A hipótese nula, $H_0: \rho_1=\rho_2=\rho_3=\rho_4=\rho_5=0$, não pode ser rejeitada ao nível de confiança de 95%, ou seja, existe correlação entre os valores da série. A mesma situação é encontrada no gráfico facp. As saídas computacionais dos correlogramas, com os valores das correlações encontram-se no Anexo 1.

A partir das características da dependência serial observada nos correlogramas, o modelo para explicar esta série deve conter componente autoregressivo e médias móveis.

A mesma série, diferenciada em um nível, apresenta funções de autocorrelação: fac dadas nas figuras 11 e 12. A análise dos gráficos mostra que pode haver vantagens na diferenciação regular da série, pois seus correlogramas se aproximam mais de uma condição de estacionariedade do que em seus valores originais. Há menos correlações ultrapassando o intervalo de confiança representado pelo pontilhado no fac e facp, do que na série original (Figuras 9 e 10).

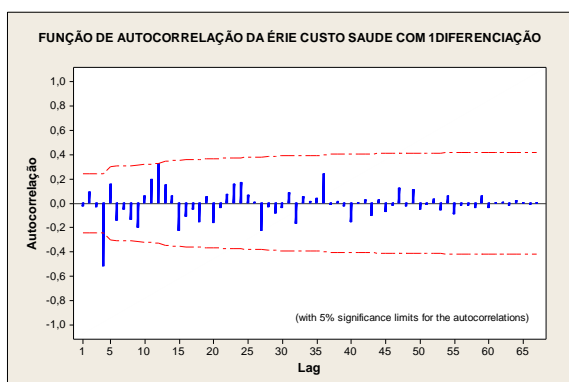


Figura 13 – Correlograma fac da série Custo-saúde, diferenciada em um nível

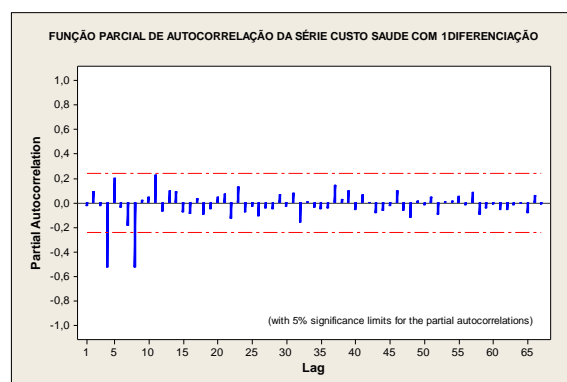


Figura 14 – Correlograma facp da série Custo-saúde, diferenciada em um nível

A diferenciação em dois níveis, apresentada nas Figuras 13 e 14, parece não trazer vantagens em relação à diferenciação em um nível. Os correlogramas são semelhantes. As saídas computacionais

dos correlogramas da série com uma e duas diferenciações, com os valores das correlações encontram-se no Anexo 1.

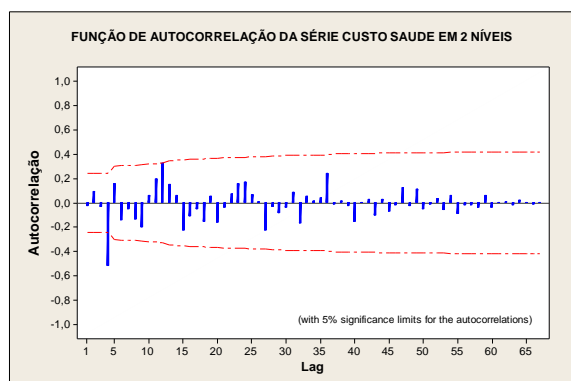


Figura 15 – Correlograma fac da série Custo-saúde, diferenciada em dois níveis

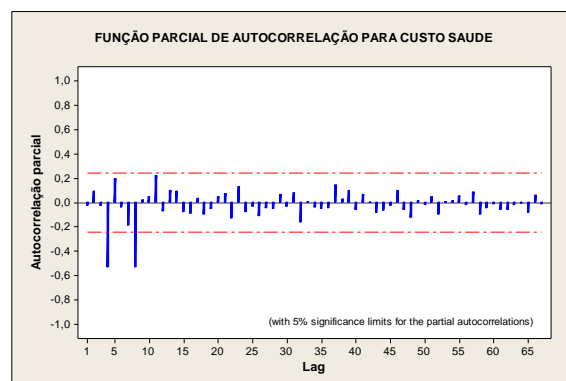


Figura 16 – Correlograma fac da série Custo-saúde, diferenciada em dois níveis

3.3 Ajuste de modelo univariado para a série Custo-saúde

Para ajuste de modelos, a série será dividida em 2 segmentos: as observações mensais do período de janeiro de 2002 a dezembro de 2006 serão consideradas como período amostral e empregadas no ajuste das equações propostas nos modelos. As observações de janeiro a dezembro de 2007 serão consideradas como período de validação e usadas para avaliar a capacidade preditiva dos modelos ajustados.

A partir da análise descritiva da série, análise das funções de autocorrelação, autocorrelação parcial da série original e diferenciada, tentativas foram feitas como sugerido por Box & Jenkins (1976) (Figura 15).

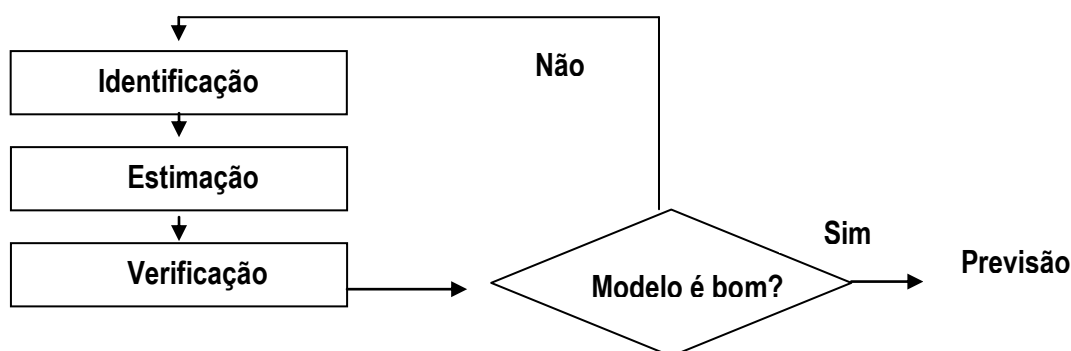


Figura 17 – Metodologia de modelagem sugerida por Box & Jenkins

Os melhores modelos univariados obtidos para a série Custo-saúde foram:

- Modelo 1: SARIMA (0,0,1) (0,1,1)¹², representada pela equação:

$$(1-B)^{12}Y_t = (1-\theta\beta)(1-\Theta B^{12})a_t$$
- Modelo 2: SARIMA (1,0,1) (1,0,0)¹², representada pela equação:

$$(1-\Theta B)(1-\Phi\beta^{12})Y_t = (1-\Theta B)a_t$$
- Modelo 3: SARIMA (2,1,1) (0,0,1)¹², representada pela equação:

$$(1-\Theta B)(1-\Theta B^2)(1-\Phi\beta^{12})Y_t = (1-\Theta B^{12})a_t$$
- Modelo 4: SARIMA (2,0,0) (0,0,1)¹², representada pela equação:

$$(1-\Theta B)(1-\Theta B^2)Y_t = (1-\Theta B^{12})a_t$$

A Tabela 1 resume os coeficientes dos modelos e seus respectivos testes de validação estatística:

Tabela 1 – Análise dos coeficientes dos melhores modelos univariados para representação da série
Custo-saúde, no período amostral

Modelos		Coeficiente	Erro padrão coef	Estatística t	p-valor
SARIMA (0,0,1) (0,1,1) ¹²	θ	-0,637976	0,097460	-6,546033	0,0000
	Θ	0,821949	0,052825	15,55988	0,0000
SARIMA (1,1,1) (1,1,0) ¹²	ϕ	-0,582765	0,133392	-4,368804	0,0001
	Φ	0,381158	0,136525	2,791848	0,0079
	θ	-0,959696	0,019273	-49,79506	0,0000
SARIMA (2,1,1) (0,1,1) ¹²	ϕ_1	-0,787224	0,132334	-5,948746	0,0000
	ϕ_2	-0,486068	0,088793	-5,474158	0,0000
	θ	-0,971597	0,032309	-30,07181	0,0000
	Θ	0,858557	0,043404	19,78037	0,0000
SARIMA (2,0,0) (0,0,1) ¹²	Θ	1620791	111521,3	14,53347	0,0000
	ϕ_1	0,358316	0,128948	2,778757	0,0075
	ϕ_2	0,369126	0,125472	2,941890	0,0048
	θ	0,862644	0,050271	17,15991	0,0000

Nesta tabela, a hipótese nula do teste de hipóteses que avalia a nulidade dos coeficientes foi rejeitada em nível de significância de 5% todos os casos, ou seja, todos os coeficientes

dos 4 modelos são diferentes de zero. Os quatro modelos apresentaram coeficientes com valores dentro do círculo unitário, entre -1 a 1, atendendo a condição de estacionariedade e invertibilidade. Apenas o modelo 4 apresenta constante enquanto os modelos 1, 2 e 3 se comportaram melhor sem ela.

A Tabela 2 traz os parâmetros de ajuste dos quatro modelos propostos. Comparando-se os resultados de cada modelo, nota-se que o modelo 3 tem a variabilidade da série Custo-saúde melhor explicada, pois apresenta maior valor de R^2 e R^2 ajustado (0,8656 e 0,85785, respectivamente). Pelo critério AIC, que oferece uma medida de ajuste baseada na diferença entre o valor estimado pelo modelo e o valor real, os quatro modelos apresentaram valores próximos, sendo o modelo 3 o menor deles. Já os modelos 2 e 4 apresentaram melhor comportamento quanto à estatística de Durbin Watson, pois trazem os maiores valores. Nota-se que nos modelos 1, 2 e 4 o valor está acima de 2 e apenas o modelo 3 abaixo deste limite indicando possível autocorrelação nos resíduos. O critério de informação Bayesiano (BIC), ou critério de Schwarz, por ser calculado a partir da soma do erro quadrático (RSS), deve ser o menor possível na seleção comparativa entre modelos, mas, na verdade, assumiu valores muito próximos nos quatro propostos. As saídas computacionais completas dos quatro modelos encontram-se no Anexo 2.

Tabela 2 – Medidas estatísticas de ajuste dos melhores modelos univariados para representação da série Custo-saúde

	R^2	R^2 ajustado	AIC	Schwarz	Durbin-Watson
Modelo1	0.564496	0.556855	26.35534	26.42576	2.258158
Modelo 2	0.751675	0.739850	26.85323	26.97368	2.383134
Modelo 3	0.865601	0.857847	26.33302	26.47768	1.818584
Modelo 4	0.741185	0.726807	26.34301	26.48511	2.287763

Melhor resultado em negrito

Quanto à análise de resíduos, apresentada na Tabela 3 e Figuras de 16 a 23, nota-se que os modelos 2 e 3 não violaram o teste de normalidade dos resíduos. O modelo 4 apresentou probabilidade de significância de 0,04, para o teste de Jarque-Bera, muito próxima da não rejeição da distribuição gaussiana dos resíduos, enquanto o modelo 1 a rejeita. Em relação à homocedasticidade dos resíduos dos modelos, avaliada pelo teste ARCH, cuja hipótese nula é a ausência de heterocedasticidade autoregressiva condicional (Anexo 3), a homocedasticidade foi aceita em todos os modelos.

Tabela 3 – Análise de resíduos dos melhores modelos univariados
para representação da série Custo-saúde

	Teste Jarque-Bera	p-valor	ARCH test Estatística F	p-valor
Modelo1	14,728	0,0006	0,042822	0,988072
Modelo 2	1,406	0,495	0,096211	0,961660
Modelo 3	4,65	0,098	0,171920	0,914864
Modelo 4	6,225	0,044	0,075055	0,973142

A possibilidade de autocorrelação nos resíduos, avaliada pelo fac e facp, para os quatro modelos se encontra no Anexo 3. Todos os modelos apresentaram condição próxima de ruído branco.

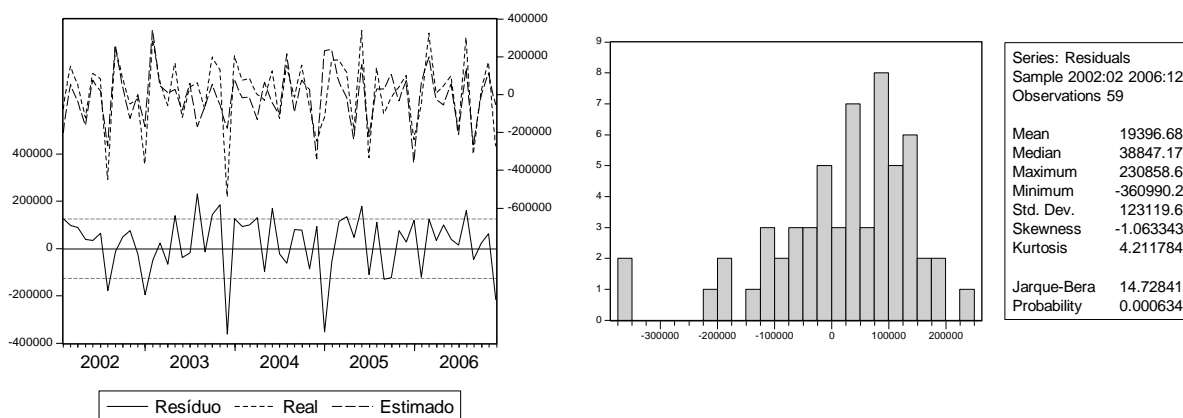


Figura18 – Série dos resíduos do modelo 1

Figura19 – Histograma de resíduos do modelo 1

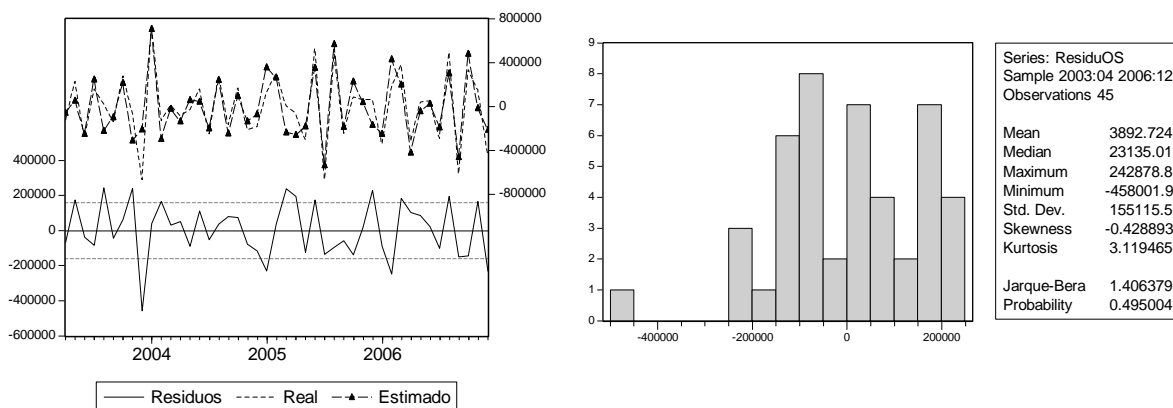


Figura 20– Série dos resíduos do modelo 2

Figura 21 – Histograma de resíduos do modelo 2

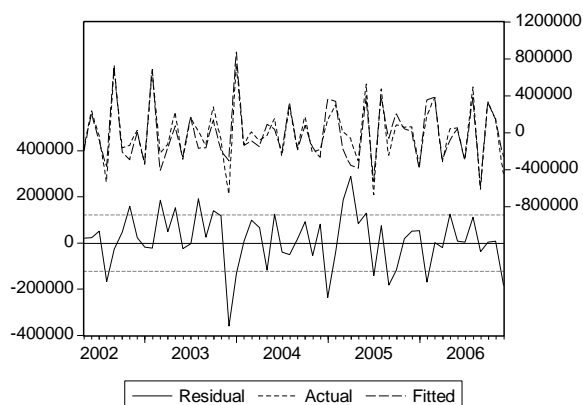


Figura 22 – Série dos resíduos do modelo 3

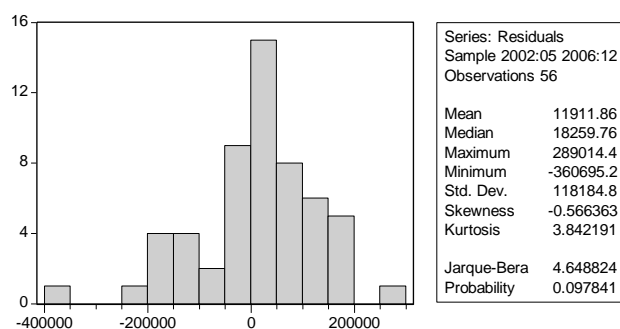


Figura 23 – Histograma de resíduos do modelo 3

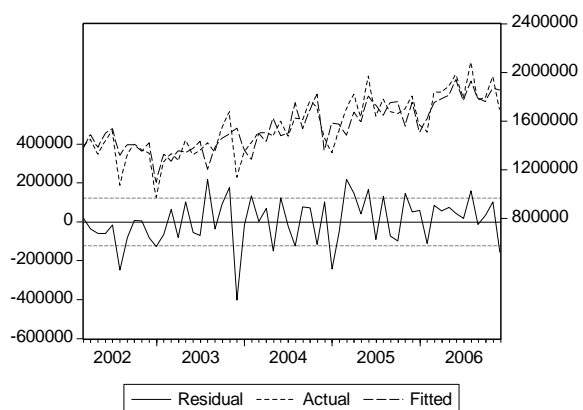


Figura 24 – Série dos resíduos do modelo 4

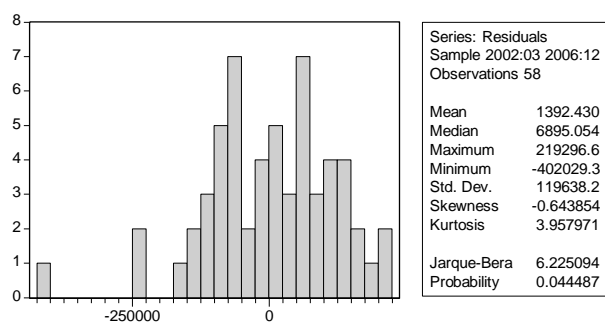


Figura 25 – Histograma de resíduos do modelo 4

Quanto aos parâmetros estatísticos de predição dos modelos, observa-se que, considerando-se o período amostral da série (janeiro de 2002 a dezembro de 2006), os modelos 1 e 4 se destacam por apresentar menores valores de erro: MAE, MAPE, RMSE, além de menor coeficiente de desigualdade.

Tabela 4 – Medidas estatísticas de avaliação da predição dos modelos univariados para representação da série Custo-saúde, no período amostral

	MAE	MAPE	RMSE	Coeficiente de desigualdade
Modelo1	261117	15,7	327347	0,113
Modelo 2	3319548	195,5	3941511	0,557
Modelo 3	1575355	98,1	1761456	0,363
Modelo 4	180081	12,4	223566	0,071

Melhor resultado em negrito

Os gráficos das Figuras 23 a 24 ilustram os parâmetros estatísticos de predição, no período amostral. Os modelos 2 e 3 (Figuras 25 e 26) mostram um aumento expressivo do intervalo superior e inferior da predição com o tempo, enquanto que esta divergência não é tão acentuada nos modelos 1 e 4 (Figuras 24 e 27).

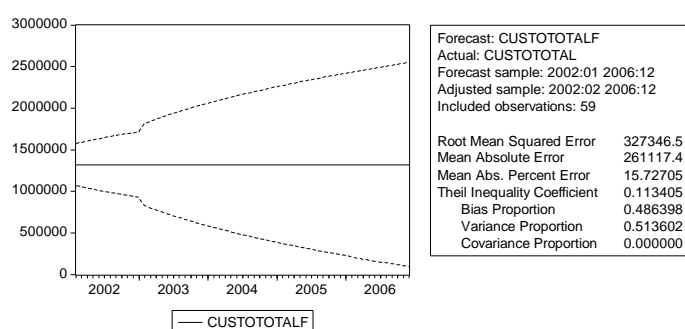


Figura 26 – Zona de predição pelo modelo 1

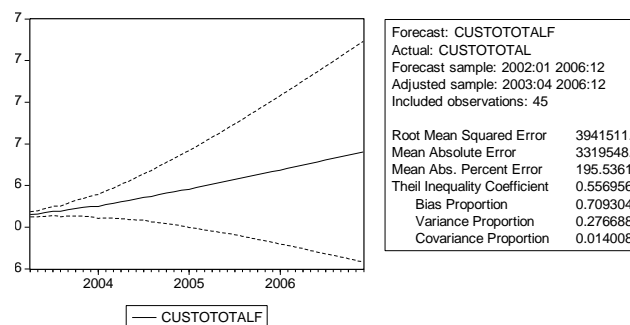


Figura 27 – Zona de predição pelo modelo 2

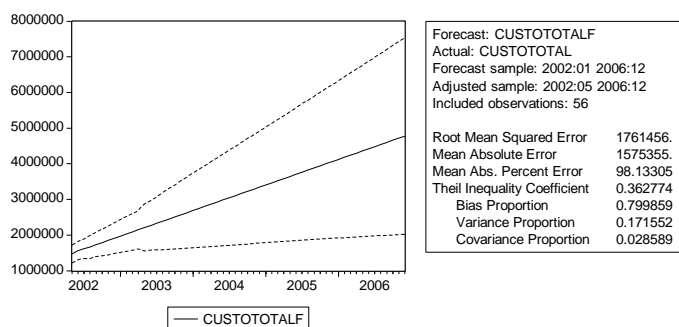


Figura 28– Zona de predição pelo modelo 3

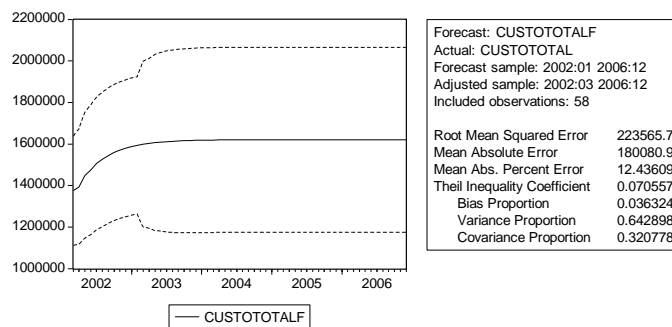


Figura 29 – Zona de predição pelo modelo 4

Considerando-se o período de validação (janeiro a dezembro de 2007), os parâmetros de qualidade preditiva dos modelos estão resumidos na Tabela 5. Nota-se que o modelo 1 e 3 possuem menor erro MAE, MAPE e RMSE, assim como menor coeficiente de desigualdade. Os modelos 2 e 3 apresentaram expressiva tendência à divergência na zona de predição (Figura 29 e 30), enquanto os demais mantêm estes intervalos relativamente constantes (Figuras 28 e 31).

Tabela 5 – Medidas estatísticas de predição dos melhores modelos univariados para representação da série Custo-saúde, no período de validação

	MAE	MAPE	RMSE	Coeficiente de desigualdade
Modelo1	95959	4.86	116378	0.030
Modelo 2	148662	7.36	167463	0.044
Modelo 3	153171	7.63	177780	0.047
Modelo 4	224632	11.09	251445	0.674

Melhor resultado em negrito

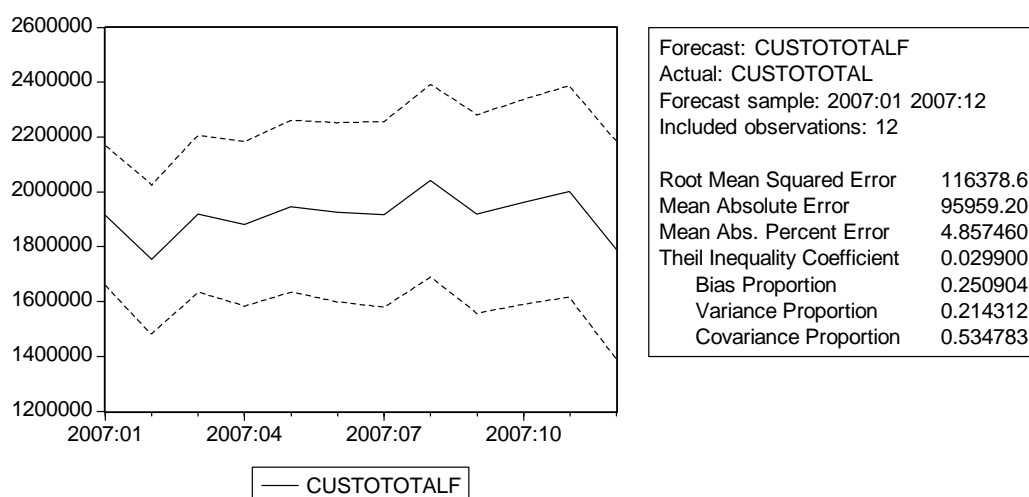


Figura 30 – Zona de predição para o ano 2007, pelo modelo 1

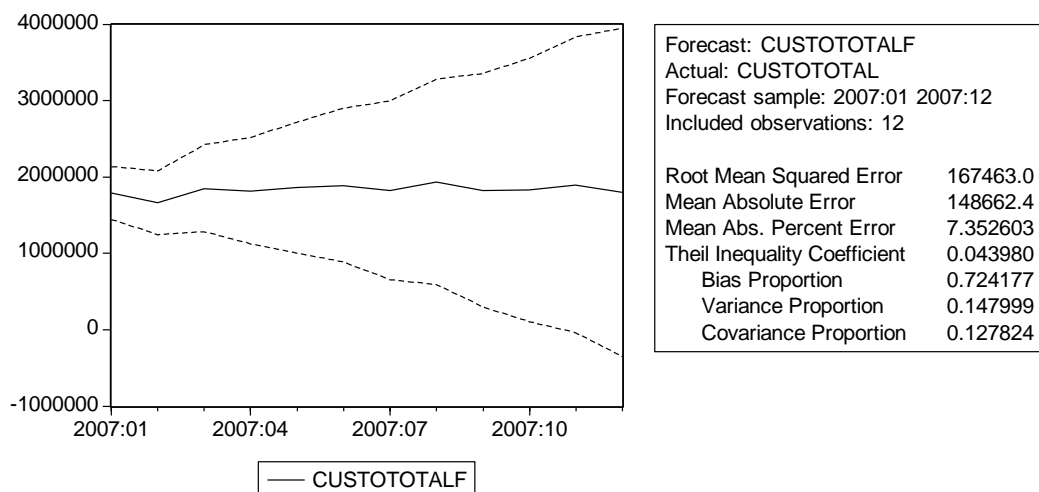


Figura 31 – Zona de predição para o ano 2007, pelo modelo 2

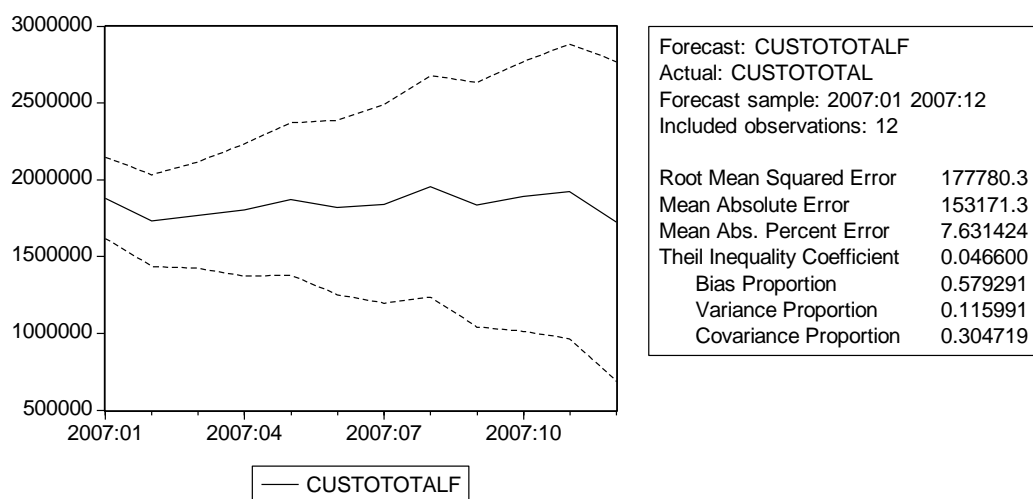


Figura 32 – Zona de predição para o ano 2007, pelo modelo 3

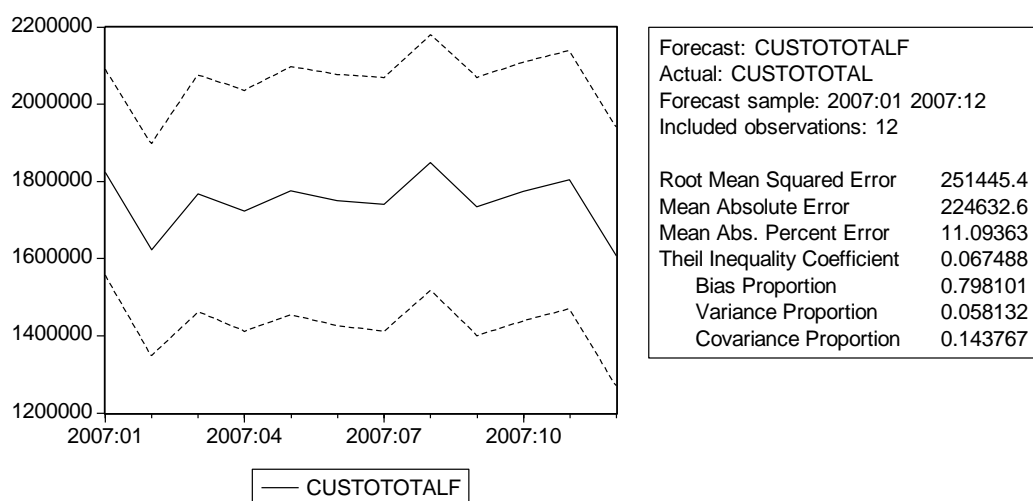


Figura 33 – Zona de predição para o ano 2007, pelo modelo 3

Considerando-se que o objetivo mais importante na elaboração do modelo para a análise de Custo-saúde é fazer previsões. Para isto, o modelo 1, SARIMA (0,0,1) (0,1,1)¹² parece ser o mais adequado entre os univariados. Neste modelo a maioria das condições de ajuste são atendidas: possui raízes invertíveis, coeficientes não são nulos estatisticamente, resíduos são homocedásticos, sem evidências de autocorrelação. Embora não apresente os melhores parâmetros de ajuste em relação aos demais, ele se aproxima muito dos demais (Tabelas 2 e 4), e os resíduos não apresentem distribuição normal (Tabela 3), foi o que mostrou a melhor condição para realizar previsões no período de validação (Tabela 5), com menor valor de MAE, MAPE, RMSE. A equação do modelo escolhido foi:

$$(1 - \beta)^{12} Y_t = (1 + 0.638\beta)(1 - 0.822\beta^{12})\alpha_t$$

3.4 Ajuste de modelo causal - regressão dinâmica

Nesta proposta de modelo, a variabilidade da série Custo-saúde tentará ser explicada pela metodologia de regressão dinâmica, cujas variáveis explicativas são as séries: Vidas, inflação (FIPE-saúde), utilização dos serviços (Sinistralidade) e Salários, tomadas no período amostral de janeiro de 2002 a dezembro de 2006. A Tabela 6 traz a descrição estatística das variáveis explicativas. A variável Custo-saúde, já apresentada na Figura 6 apresenta distribuição normal, assim como as variáveis Sinistralidade e Salários, que não tiveram a distribuição normal rejeitada pelo teste de Anderson Darling.

Tabela 6 – Descrição das variáveis do modelo de regressão dinâmica

Séries	Média	Desvio padrão	Valor mínimo	Valor máximo	p-valor (Teste Anderson Darling)
Vidas	39326	3503	35470	49010	<0.005
FIPE-saúde	316,25	44,21	242,93	386,18	0.018
Sinistralidade	54948	8420	40280	97976	0.476
Salários R\$	398056	70419	195908	541621	0.926

Cada série candidata a variável independente no modelo foi inicialmente plotada simultaneamente à série Custo-saúde, como mostrado nas Figuras de 32 a 35. A avaliação intuitiva desta comparação, apenas descritiva, aponta para o indexador escolhido da inflação no período, o FIPE-saúde, e os salários dos profissionais de saúde como capazes de acompanhar melhor a tendência crescente da série (Figuras 33 e 35).

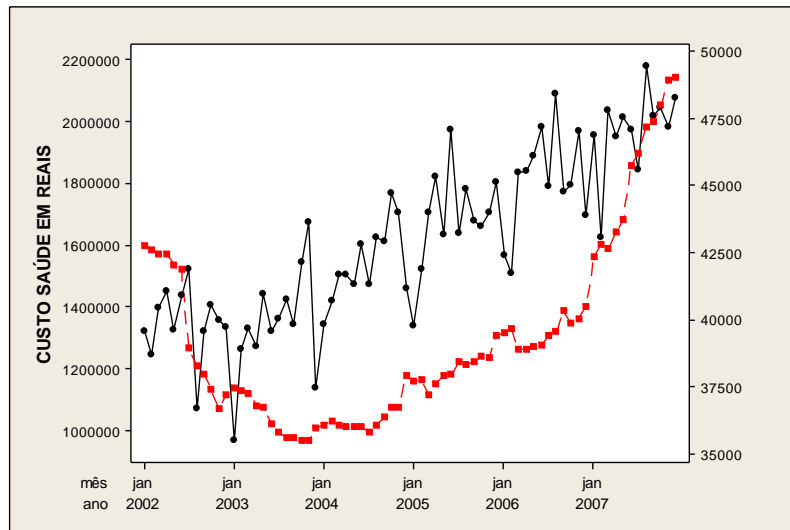


Figura 34 – Gráfico das séries Custo-saúde e Vidas, no período de janeiro de 2002 a dezembro de 2007

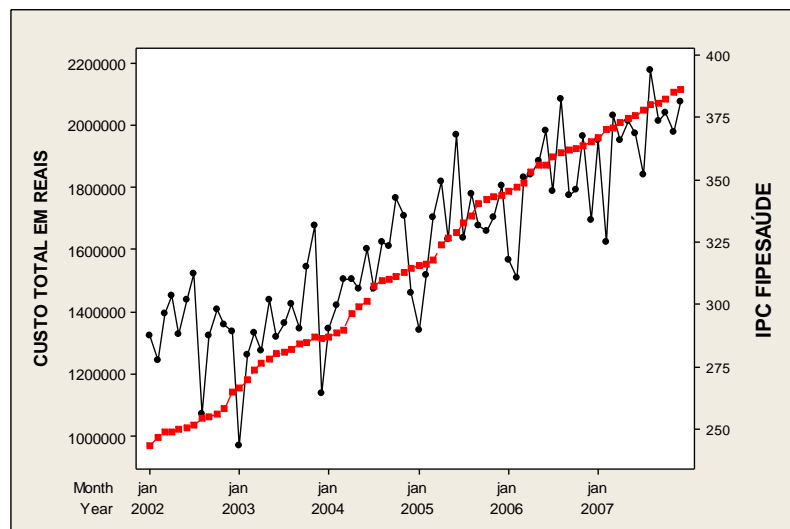


Figura 35 – Gráfico das séries Custo-saúde e FIPE-saúde, no período de janeiro de 2002 a dezembro de 2007

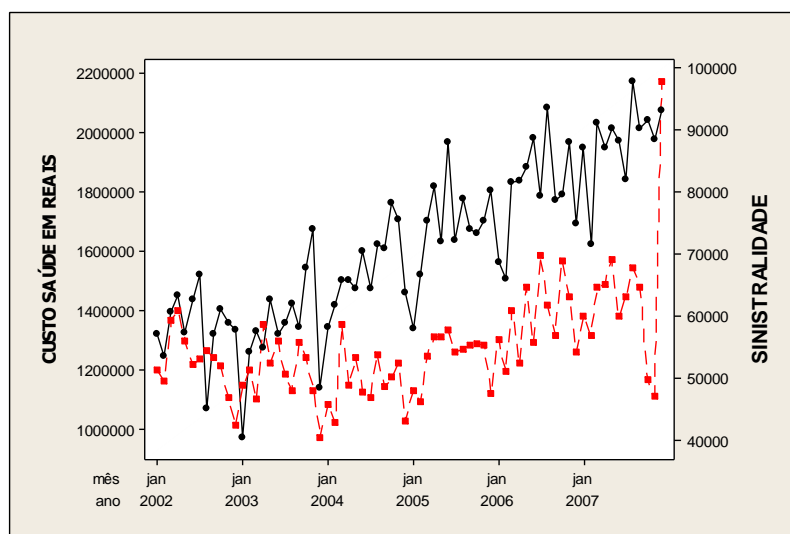


Figura 36 – Gráfico das séries Custo-saúde e sinistralidade, no período de janeiro de 2002 a dezembro de 2007

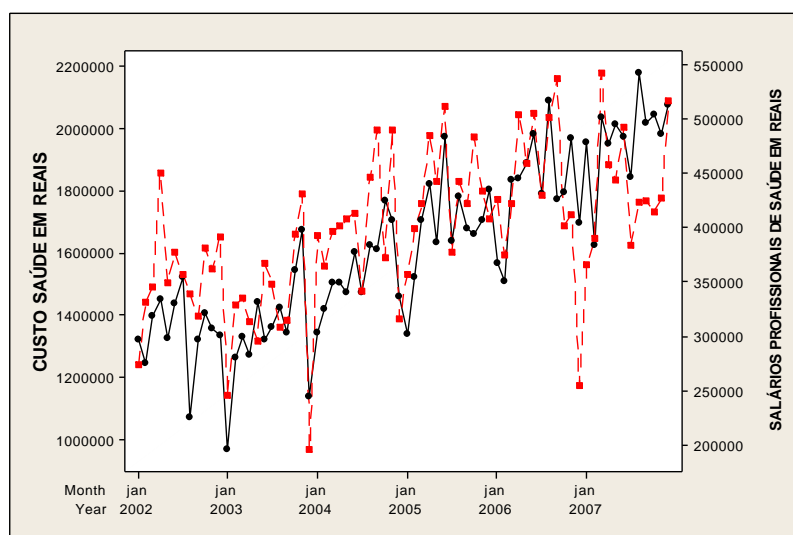


Figura 37 – Gráfico das séries Custo-saúde e Salários, no período de janeiro de 2002 a dezembro de 2007

A matriz de correlação entre os valores da série, apresentada na Tabela 7, confirma esta impressão inicial, pois justamente as variáveis FIPE-saúde e Salários possuem os maiores coeficiente de correlação de Pearson, em relação à variável resposta: Custo-saúde. Todas as variáveis escolhidas estão correlacionadas individualmente à variável resposta, com p-valores 0,000, que rejeitam a hipótese de que esta correlação seja zero. No entanto, o pressuposto de normalidade só pode ser confirmado para as variáveis Sinistralidade e Salários. Observa-se ainda que FIPE-saúde e Vidas estão correlacionadas entre si ($p=0,536$, $p=0,000$), mas são variáveis que descrevem fenômenos bem distintos, quais sejam, a inflação e o número de empregados contratados pela empresa. Também as variáveis FIPE-saúde e Salários, apresentam alguma correlação entre elas ($p=0,526$, $p=0,000$), embora os salários sejam reajustados por outro indexador.

Tabela 7 – Matriz de correlação entre as variáveis, com p-valor para $H_0: \rho=0$

	Custo-saúde	Vidas	FIPE-saúde	Sinistralidade
Vidas	0,559 (0,000)			
FIPE-saúde	0,864 (0,000)	0,536 (0,000)		
Sinistralidade	0,572 (0,000)	0,550 (0,000)	0,506 (0,000)	
Salários R\$	0,728 (0,000)	0,255 (0,030)	0,526 (0,000)	0,453 (0,000)

Avaliadas também individualmente pela regressão simples, na Tabela 8, a variável Custo-saúde foi mais bem explicada pela variável FIPE-saúde, com R^2 ajustado de 74,6%.

Tabela 8 – Análise de regressão univariada de cada variável resposta em relação ao Custo-saúde

	Coefficientes	R²	R² ajustado	p-valor (estatística F)
Vidas	43,69	31,2	30,2	0,000
FIPE-saúde	5354,8	74,7	74,6	0,000
Sinistralidade	18,60	32,7	31,7	0,000
Salários R\$	2,83	52,9	52,3	0,000

Melhor resultado em negrito

A diferenciação das séries Vidas, FIPE-saúde, Sinistralidade e Salários, até o segundo nível, não mostrou melhora dos coeficientes de correlação, tomadas individualmente, com a variável reposta. As matrizes de correlação encontram-se no Anexo 1.

A construção do modelo de regressão dinâmica seguirá a estratégia de estimação e verificação, partindo de um modelo simples, até chegar a um mais elaborado e ajustado aos dados, como ilustrado na Figura 36.

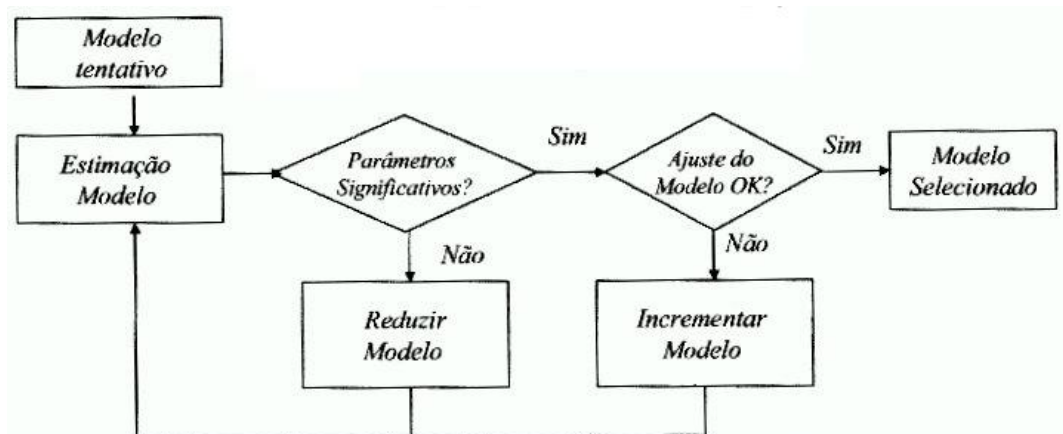


Figura 38 – Metodologia de modelagem para modelos dinâmicos (TOSCANO 2007)

O modelo dinâmico inicial, baseado no modelo mais simples com todas as variáveis foi resumido na Tabela 9 e as saídas computacionais apresentadas no Anexo 4.

Tabela 9 – Modelo de regressão dinâmica 1 para série Custo-saúde

	Coeficientes	Erro padrão coeficientes	p-valor Estatística t	VIF
Constante	-453630	288759	0.000	
Vidas	6.25	4.87	0.430	1.2
FIPE-saúde	3303.9	457.3	0.000	1.5
Sinistralidade	3.81	2.92	0.198	1.6
Salários R\$	1.42	0.24	0.000	1.5

$R^2 = 81,76\%$, R^2 ajustado = $80,07\%$; Estatística F $57,09$, p-valor $0,0000$; Durbin-Watson estatística = $2,252$; Critério de Schwarz = $26,270$; AIC = $26,066$

O teste de hipóteses para os coeficientes do modelo dinâmico inicial rejeitou a nulidade dos coeficientes das variáveis FIPE-saúde e Salários, ambas com probabilidade de significância 0,000. No entanto, a significância dos parâmetros relacionados às variáveis Vidas e Sinistralidade não foi verificada pois o p-valor está acima de 0,05. A estatística VIF apresentou valores baixos para todas as variáveis, indicando que provavelmente não há problemas de multicolinearidade entre as variáveis.

A análise de resíduos do modelo dinâmico inicial foi mostrada graficamente nas Figuras 37 e 38. Os resíduos apresentaram distribuição normal (teste de Jarque-Bera com p-valor = 0,813 não rejeita a hipótese de distribuição gaussiana). Os correlogramas fac e facp dos resíduos, saída computacional no Anexo 4, mostraram alguns indicativos de autocorrelação, enquanto a hipótese nula é rejeitada no teste de heterocedasticidade foi rejeitado (*White test*, estatística F com p-valor 0,246, Anexo 4).

Análise de resíduos

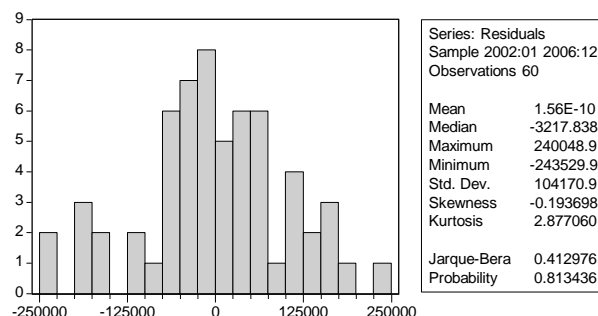
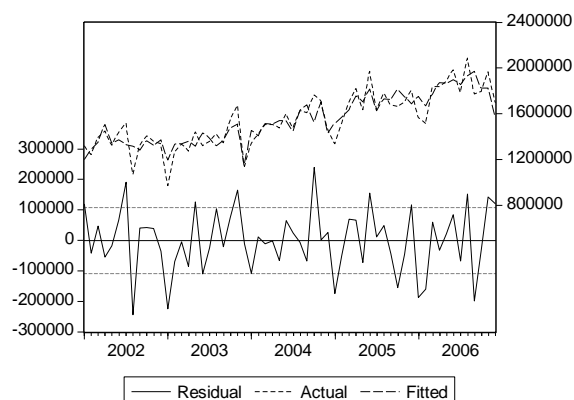


Figura 39 – Série dos resíduos do modelo dinâmico1

Figura 40 – Histograma de resíduos do modelo dinâmico1

Como o modelo não foi aprovado pelos testes de ajustes utilizados, modificações deverão ser efetuadas para sua adequação.

Incrementos e decréscimos ao modelo dinâmico inicial

A primeira medida para adequação do modelo inicial foi a reavaliação das variáveis preditoras. A variável Vidas foi retirada, por apresentar maiores valores de probabilidade de significância no teste de hipótese de sua constante (Tabela 9) e possibilidade de multicolinearidade com a variável Sinistralidade.

Tabela 10 – Modelo dinâmico 2 para série Custo-saúde

	Coefficientes	Erro padrão coeficientes	p-valor Estatística t	VIF
Constante	-254229	142164.2	0.0791	
FIPE-saúde	3268.8	453.7	0.0000	1,5
Sinistralidade	4.82	2.62	0.0714	1,3
Salários R\$	1.41	0.24	0.0000	1,5

$R^2 = 80,37\%$, R^2 ajustado = $79,31\%$; Estatística F 76,41, p-valor 0,0000; Durbin-Watson estatística = 2,246; Critério de Schwarz = 26,073; AIC = 26,066

Se considerarmos o nível de significância de 5% para o teste de hipóteses para os parâmetros do modelo dinâmico 2, a nulidade correspondente à variável Sinistralidade não será rejeitada, no entanto,

o p-valor está limítrofe (7%, Tabela 10). A análise de resíduos do modelo dinâmico 2 foi adequada quanto à normalidade e homocedasticidade (Anexo 5). Os correlogramas fac e facp dos resíduos deste modelo, no entanto, mostraram alguns indicativos de autocorrelação.

A intervenção seguinte no modelo de regressão dinâmica foi a introdução de variáveis de intervenção ou variáveis *dummy*. O objetivo deste tipo de procedimento é considerar e tratar as situações atípicas. Situações atípicas foram identificadas nos meses de outubro de 2004 (Atípico 1), agosto de 2002 (Atípico 2) e setembro de 2006 (Atípico 3), através da análise do *plot* de resíduos do modelo dinâmico 1 (Anexo 4) e em seguida foram individualmente tratadas com *dummies*.

A introdução gradativa de cada *dummy* resultou em novos modelos (dinâmico 3, 4 e 5), apresentados sumariamente na Tabela 11. As saídas computacionais, assim como análise de resíduos estão no Anexo 6. Observou-se uma melhoria gradativa do coeficiente de determinação da regressão.

Tabela 11 – Medidas estatísticas de ajuste dos modelos dinâmicos com a introdução gradativa das variáveis *dummies* 1, 2 e 3

	Atípico tratado	R ²	R ² ajustado	Coef. Durbin- Watson	p-valor (estatística F)
Modelo dinâmico 3	1	82.08	80.78	2.28	0.000
Modelo dinâmico 4	1 e 2	84.01	82.53	2.07	0.000
Modelo dinâmico 5	1, 2 e 3	85.05	83.36	1.97	0.000

O modelo dinâmico 5, contendo as variáveis explicativas FIPE-saúde, Sinistralidade e Salários, com tratamento para os três valores atípicos, apresentou resultado do teste de hipóteses, para seus parâmetros, sumarizado na Tabela 12.

Tabela 12 – Modelo dinâmico 5 para série Custo-saúde (tratamento de valores atípicos)

	Coefficientes	Erro padrão coeficientes	p-valor Estatística t	VIF
Constante	-275625,0	129533,5	0,0380	
FIPE-saúde	3082,56	414,85	0,0000	1,6
Sinistralidade	1,501552	0,223315	0,0000	1,3
Salários R\$	5,669441	2,374715	0,0206	1,6
Atípico 1	236789,3	97719,79	0,0188	1,0
Atípico 2	-254937,3	99385,59	0,0132	1,0
Atípico 3	-195044,0	101456,0	0,0599	1,1

Nota-se que o isolamento dos valores atípicos da série resultou na melhora das variáveis FIPE-saúde, Sinistralidade e Salários como explicativas da variabilidade da variável resposta, Custo-saúde. Apenas a *dummy* para o Atípico 3 não teve o teste de hipótese rejeitado para a nulidade de sua constante pela estatística t, em nível de 5% (Tabela 12). A análise de resíduos deste modelo representada graficamente nas Figuras 39 e 40, mostram distribuição normal (Teste de Jarque-Bera para normalidade dos resíduos não rejeitou a hipótese nula, p-valor: 0,83) e homocedasticidade (Anexo 6).

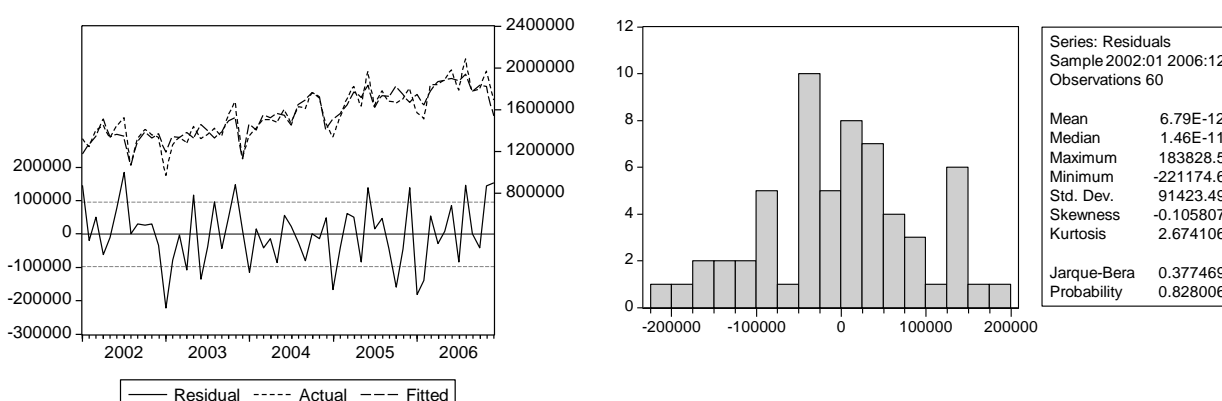


Figura 41 – Série dos resíduos do modelo dinâmico 5 Figura 42 – Histograma de resíduos do modelo dinâmico 5

O correlograma dos resíduos, no entanto, mostrou evidência de autocorrelação na *lag* 7, como apresentado na Figura 41.

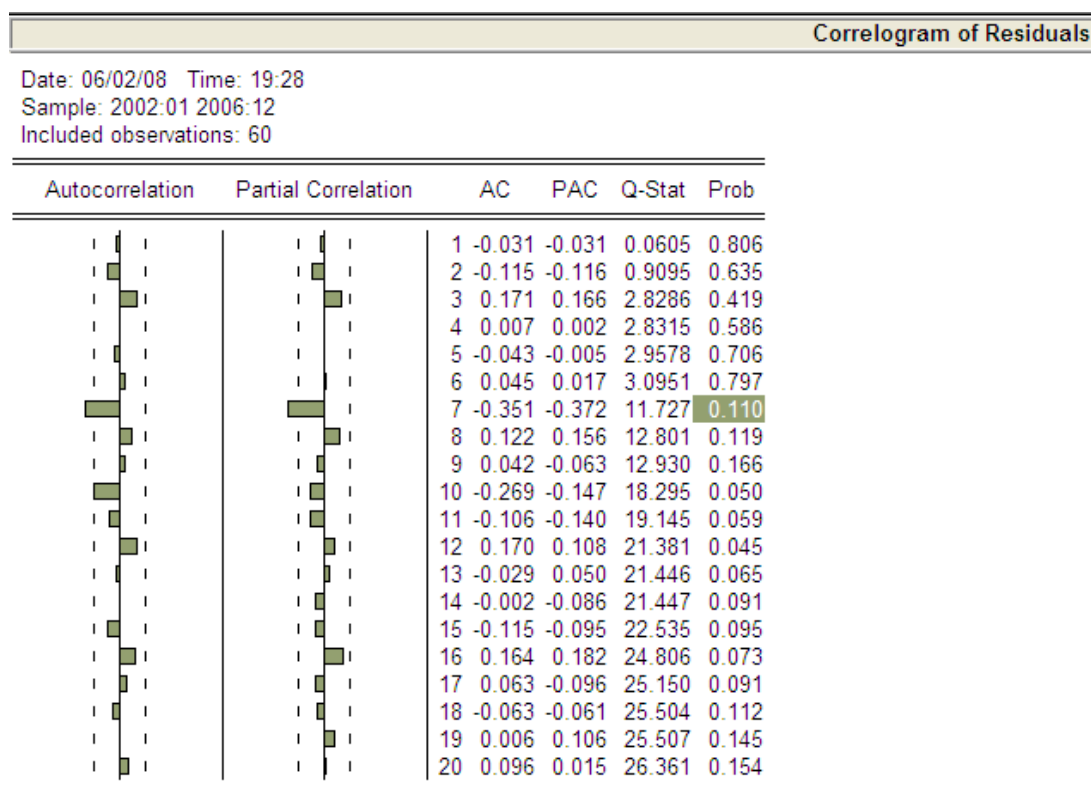


Figura 41 – Correlograma dos resíduos do modelo dinâmico 5

Como intervenção final, a autocorrelação do resíduo no termo 7 foi tratada através de defasagens, introduzindo-se no modelo o termo médias móveis de ordem 7 dos resíduos. O modelo resultante, denominado modelo dinâmico 6 foi apresentado na Tabela 10. A saída computacional encontra-se no Anexo 7. A equação do modelo final foi:

$$\text{Custo-saúde} = -355643,261 + 3781\text{FIPESAÚDE} + 1,33\text{SALARIOS} + 4,22\text{SINISTRALIDADE} + 209883,9\text{AT1} - 187586,7\text{AT2} - 340526,6\text{AT3} + [\text{MA}(7)=-0,93\text{BACKCAST}=2002:01]$$

Tabela 13 – Modelo dinâmico 6 para série Custo-saúde
(tratamento de valores atípicos e erros correlacionados)

	Coefficientes	Erro padrão coeficientes	Coefficientes padronizados	p-valor Estatística t	VIF
Constante	-355643,3	67866,31		0,0000	
FIPE-saúde	3781,13	273,58	13,8	0,0000	1,9
Salários R\$	1,326200	0,122400	10,8	0,0000	1,6
Sinistralidade	4,221077	1,732099	2,4	0,0183	1,6
Atípicos 1	209883,9	54170,42	---	0,0003	1,0
Atípicos 2	-187586,7	81742,39	---	0,0258	1,1
Atípicos 3	-340526,6	60982,02	---	0,0000	1,1
Erro - 7	-0,932531	0,021838	---	0,0000	1,1

$R^2 = 91,34\%$, R^2 ajustado = $90,17\%$; Estatística F 78,36, p-valor 0,0000; Durbin-Watson estatística = 2,009; Critério de Schwarz = 25,667; AIC = 25,388

O teste de hipóteses para os coeficientes do modelo dinâmico 6 rejeitou a nulidade para todas as variáveis, incluindo os valores atípicos tratados e o termo (Erro – 7). Portanto todos os coeficientes são estatisticamente significantes ao nível de 95% de confiança. O coeficiente de determinação da regressão foi elevado, mostrando que o modelo explica 91,34% da variabilidade da série Custo-saúde. A estatística de Durbin Watson acima do valor 2 indica que provavelmente não há autocorrelação nos resíduos.

A análise de resíduos do modelo dinâmico 6, Figuras 41 e 42, indica que os resíduos apresentaram distribuição normal (teste de Jarque-Bera com p-valor = 0,781 não rejeita a distribuição Gaussiana). Os correlogramas fac e facp dos resíduos, saída computacional no Anexo 7, confirmam graficamente esta observação, sem indicativos de autocorrelação. A hipótese nula do teste de heterocedasticidade foi rejeitada (*White test*, estatística F com p-valor 0,813), saída computacional no Anexo 7.

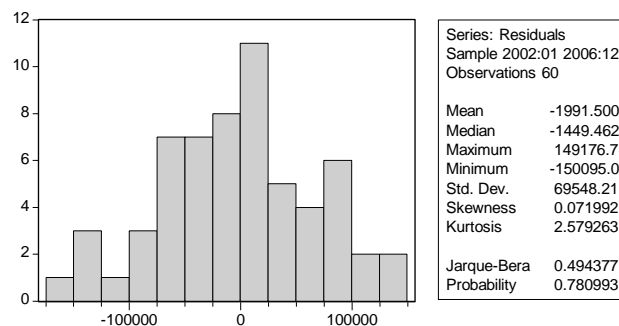
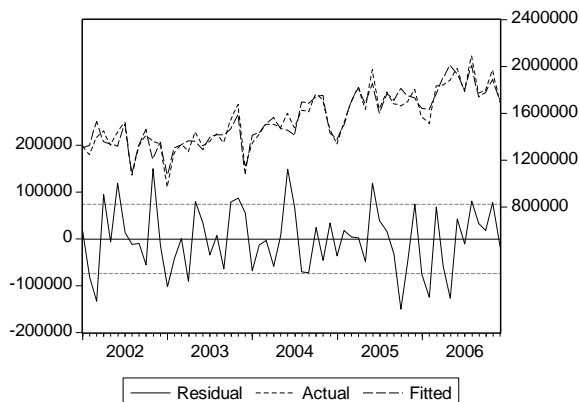


Figura 44– Série dos resíduos do modelo dinâmico 6 Figura 45– Histograma dos resíduos do modelo dinâmico 6

A partir de modelo completamente ajustado, o dinâmico 6, as propriedades preditivas foram estudadas e estão graficamente apresentadas nas Figuras 43 e 44. Considerando a previsão no período amostral (janeiro de 2002 a dezembro de 2006), o MAE, MAPE e RMSE foram respectivamente: 76033.3, 5.027, 96006.7 enquanto o coeficiente de desigualdade foi 0.0308.

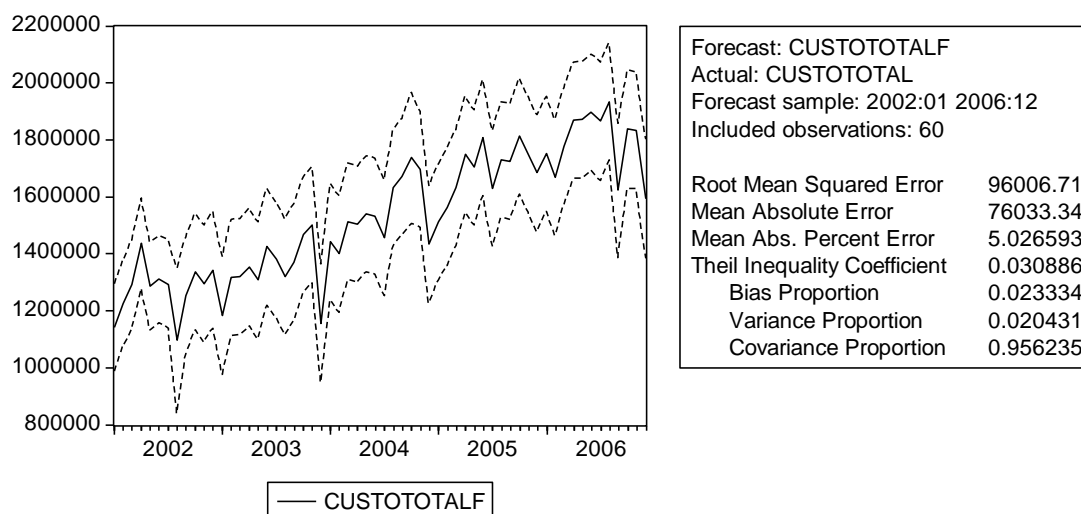


Figura 46 - Zona de predição pelo modelo dinâmico 6, para o período amostral

Para o período de validação, janeiro a dezembro de 2007, o MAE, MAE e RMSE foram respectivamente: 123197.3, 6.255 e 141160.6, enquanto o coeficiente de desigualdade foi 0.063 (Figura 42).

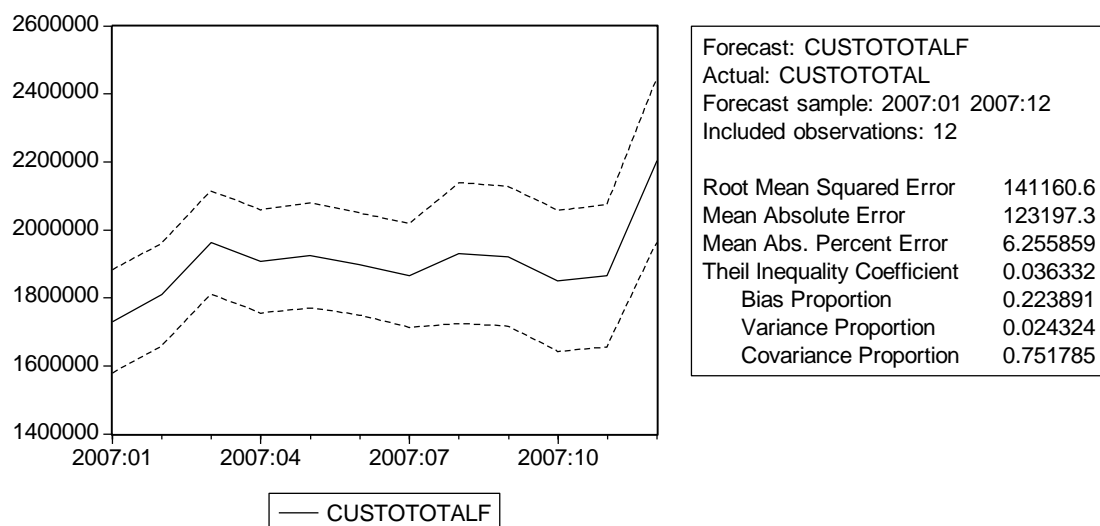


Figura 47 - Zona de predição pelo modelo dinâmico 6, para o ano 2007

Uma análise dos modelos melhores ajustados, obtidos por metodologias distintas, univariada e regressão dinâmica, foi sumarizada na Tabela 14:

Tabela 14 – Medidas de ajuste dos melhores modelos para representação da série Custo-saúde

	R ²	R ² ajustado	AIC	Schwarz	Durbin-Watson
Modelo univariado	0,5645	0,5569	26,355	26,426	2,258
Modelo dinâmico	0,9134	0,9017	25,388	25,667	2,009

É possível observar que os parâmetros de ajuste são muito próximos, embora, pelo coeficiente de determinação, o modelo dinâmico consiga explicar melhor a variabilidade da série Custo-saúde. A regressão dinâmica permite que se avalie a importância relativa de cada variável explicativa em relação ao custo: a inflação medida pelo FIPE-saúde, que apresenta coeficiente padronizado de 13,8 mostrou ser o principal determinante o custo, seguida do salário dos profissionais de saúde e da Sinistralidade (coeficientes 10,8 e 2,4 respectivamente).

Os resultados levam a crer que a inflação, um fenômeno externo ao sistema de saúde da Fundação FIAT, representativo da estabilidade econômica do país, atua nos dois determinantes principais do custo das empresas com a saúde: FIPE-saúde e Salários. Portanto, faz-se um alerta de que o principal foco de atenção dos gestores do benefício-saúde destas empresas mereça ser o gerenciamento dos contratos com prestadores de serviços de saúde e reajuste de salários dos funcionários, ambos norteados pela inflação. Neste sentido, o modelo causal permite simulação de cenários com avaliação do impacto de indicadores de inflação projetados para os meses do ano e também um estudo prévio de propostas de reajustes salariais, avaliando-se antecipadamente seu impacto no custo total. Também a sinistralidade, que representa nesta análise o consumo dos serviços de saúde pelos beneficiários do plano, faz sentido à realidade da oscilação de gastos. Há uma expectativa real de relação causa-efeito entre uma maior utilização da assistência à saúde e elevação no gasto mensal com este benefício para seus empregados.

Quanto ao modelo univariado, SARIMA (0,0,1) (0,1,1)¹², chama atenção sua simplicidade e capacidade de predição, com parâmetros apresentados na Tabela 15. Apresenta menores erros absoluto, quadrático e erro médio percentual e o coeficiente de desigualdade mais próximo de zero. Trata-se de um modelo parcimonioso, portanto de fácil utilização e bem ajustado aos dados, que possui uma ótima capacidade de predição, pelo menos em intervalo de um ano, com a vantagem de não

depende de outros parâmetros, apenas do passado da série. A questão da ausência de normalidade nos resíduos, não parece em princípio comprometer sua funcionalidade, mas ele deve ser testado e reavaliado quando um novo conjunto de dados puder ser incorporado à amostra.

Tabela 15 – Avaliação comparativa dos parâmetros de previsão dos modelos para o período de janeiro de 2007 a dezembro de 2007

	MAE	MAPE	RMSE	Coeficiente de desigualdade
Modelo univariado	95959	4,86	116378	0,03
Modelo dinâmico	123197,3	6,26	141160,6	0,22

Na definição do modelo adequado, é necessário levar em conta não só a significância dos parâmetros, mas também sua estrutura e sua aplicabilidade, verificando se os coeficientes estimados são coerentes com o contexto. O modelo univariado traz a vantagem de só depender do passado da série para ser usado como preditor de valores futuros. As previsões geradas por um modelo de regressão dinâmica dependem não só de valores passados da série, mas também dos valores previstos para as variáveis causais. Logo, para se obter as previsões da série para $t+1$, $t+2$, $t+3$, etc, é necessário fornecer ao modelo os valores futuros do vetor de variáveis causais X_t . Se as previsões destas variáveis exógenas não forem apropriadas, o modelo de regressão dinâmica irá também gerar previsões inadequadas.

Na verdade, os dois modelos são adequados e podem ser usados para solucionar as questões distintas propostas nos objetivos do estudo. O modelo univariado tem vantagens ao possibilitar a verificação mensal do valor gasto com o benefício-saúde dentro de limites previstos pela série, além de permitir previsão da expectativa de gastos anuais sem a necessidade de se informar quaisquer parâmetros. O modelo por regressão dinâmica, por outro lado, permite a previsão de valores futuros em cenários projetados em função das variáveis que o compõe: Sinistralidade, Salários dos profissionais de saúde e inflação medida pelo FIPE-saúde.

O presente estudo, possível pela fundamentação teórica do Curso de Especialização em Estatística e viabilizado pela integração entre áreas afins da Medicina e Estatística, superou as expectativas iniciais, pois além de um exercício de conceitos, fez transformar em boa prática algumas questões fundamentais da gestão de um importante plano de saúde do mercado brasileiro.

- A série Custo-saúde, referente ao gasto mensal da assistência à saúde oferecida pela Fundação FIAT no período de janeiro de 2002 a dezembro de 2007, apresentou características de tendência crescente e leve sazonalidade.
- O modelo univariado SARIMA (0,0,1) (0,1,1)¹² possibilitará prever valores futuros ao período estudado baseado apenas no passado da série. Após ampliação do período amostral acima de 60 observações, espera-se melhorar a condição de normalidade no vetor de resíduos.
- Existem evidências de relação de causa-efeito entre a série de Custo-saúde com outras séries, representativas de fenômenos determinantes de sua variabilidade, na seguinte ordem de importância:
 - ✓ FIPE-saúde: indicador de inflação que representa o efeito do preço da atividade assistencial,
 - ✓ Salários dos profissionais de saúde, que representa o efeito administração da rede própria de atendimento,
 - ✓ Sinistralidade que representa a utilização dos serviços de saúde e portanto o efeito consumo.
- O modelo dinâmico escolhido pode ser usado para fazer previsão em cenários simulados diversos, a partir das variáveis que o compõe.
- O presente estudo, além de cumprir os objetivos propostos, possibilitou o exercício de uma importante e complexa metodologia da Estatística, aplicável ao contexto atual da saúde no Brasil, abrindo um campo de trabalho em comum nestas duas áreas para estudos futuros.

Box, G. E. P.; Jenkins, G. M. *Time series analysis forecasting and control*. San Francisco: Holden-Day, 1976. Edição revisada.

Brockwell, P. J., Davis, R. A. Introduction. In: _____. *Introduction to Time Series and Forecasting*. 2ed., New York: 2002a, cap. 1, p. 1-40.

Brockwell, P. J., Davis, R. A. Introduction. In: _____. *Introduction to Time Series and Forecasting*. 2ed., New York: 2002b, cap. 7, p. 223-256.

Carol, J Loomis The tragedy of general motors. *Fortune*. New York: Feb 20, 2006. Vol. 153, Iss. 3; pg. 59, 11.

Chatfield, C. Simple Descriptive Techniques. In: _____. *The analysis of time series: an introduction*. 6ed., Boca Raton: CRC Press, 2004a, cap. 2, p. 33-50.

Chatfield, C. Fitting Time-Series Models in the Time Domain. In: _____. *The analysis of time series: an introduction*. 6ed., Boca Raton: CRC Press, 2004b, cap. 4, p. 55-70.

Chatfield, C. Some more advanced Topics. In: _____. *The analysis of time series: an introduction*. 6ed., Boca Raton: CRC Press, 2004c, cap. 13, p. 255-268.

Latorre, M. R. D. O.; Cardoso, M. R. A. Análise de séries temporais em epidemiologia: uma introdução sobre os aspectos metodológicos. *Rev. bras. epidemiol.* vol.4 no.3 São Paulo Nov. 2001.

MARQUES, D. O custo da medicina de grupo. *Notícias Fiesp/Ciesp* - 27/05/2002. In: *Crescimento das despesas no setor incide sobre faturamento das empresas*.

Ministério da Saúde 2008. Notícia em Portal da Saúde. Acedido em 9 de Maio de 2008, em: http://portal.saude.gov.br/portal/saude/visualizar_texto.cfm?idtxt=27788.

MINITAB Release 14.12.0 Help. 2004.

Morettin P. A., Toloi C. M. C. Preliminares. In: _____. *Análise de série temporais*. 2ed., São Paulo: Edgard Blücher, 2006a, cap. 1, p. 1-16.

Morettin P. A., Toloi C. M. C. Modelos para série temporais. In: _____. *Análise de série temporais*. 2ed., São Paulo: Edgard Blücher, 2006b, cap. 2, p. 19-41.

Morettin P. A., Toloi C. M. C. Modelos ARIMA. In: _____. *Análise de série temporais*. 2ed., São Paulo: Edgard Blücher, 2006c, cap. 5, p. 105-42.

Morettin P. A., Toloi C. M. C. Diagnóstico de modelos ARIMA. In: _____. *Análise de série temporais*. 2ed., São Paulo: Edgard Blücher, 2006d, cap. 8, p. 199-217.

Morettin P. A., Toloi C. M. C. Modelos sazonais. In: _____. *Análise de série temporais*. 2ed., São Paulo: Edgard Blücher, 2006e, cap. 10, p. 243-74.

Neal E. Boudette and Stephen Power. GM CEO Says Mergers Unlikely; Wagoner Sees No Easing Of Industry Overcapacity; Health-Care Issue Ripens. Wall Street Journal. (Eastern Edition). New York, N.Y.:Mar 8, 2007. p. A.3

Oliveira, P. J., Campos, E. F., Gontijo, M. C. F. Programa de desenvolvimento da saúde. In:____Campos, E. F. Desenvolver a saúde. Modelo Cuidador da Federação das UNIMEDS de Minas Gerais. Coleção Desenvolvimento da Saúde vol1. UNIMED Federação Minas, 2008, cap 1, p.13-35.

Toscano, E. M. M. Métodos estatísticos de previsão. 2007. Versão preliminar. Site:www.est.ufmg.br.
<ftp://ftp.est.ufmg.br/pub/mercedes/especializacao>.

Jussara Teixeira da Silva, Ela Mercedes M. de Toscano. Econometric views guia do usuário (Versão Preliminar)

Wikipedia Enciclopedia Livre. Akaike information criterion. Acedido em 9 de Maio de 2008, em: http://en.wikipedia.org/wiki/Akaike_information_criterion.

Wikipedia Enciclopedia Livre. Durbin Watson statistic. Acedido em 9 de Maio de 2008, em: http://en.wikipedia.org/wiki/Durbin-Watson_statistic.

Zanini, A. Redes neurais e regressão dinâmica: um modelo híbrido para previsão de curto prazo da demanda de gasolina automotiva no Brasil. Rio de Janeiro: PUC-Rio, 2000 (Dissertação de Mestrado).

ANEXO 1 – Funções de autocorrelação e autocorrelação parcial da série custo saúde

a) Série original

Date: 04/20/08 Time: 18:22

Sample: 2002:01 2007:12

Included observations: 72

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
. *****	. *****	1	0.729	0.729	39.879	0.000
. *****	. ***	2	0.709	0.380	78.191	0.000
. *****	. *	3	0.665	0.168	112.32	0.000
. *****	. *	4	0.624	0.072	142.86	0.000
. *****	. *	5	0.641	0.172	175.52	0.000
. ****	. *	6	0.559	-0.071	200.73	0.000
. ****	. *	7	0.577	0.092	228.01	0.000
. ****	. .	8	0.522	-0.034	250.68	0.000
. ****	. .	9	0.482	-0.052	270.31	0.000
. ***	. .	10	0.453	-0.042	287.93	0.000
. ****	. **	11	0.493	0.203	309.14	0.000
. ****	. *	12	0.513	0.134	332.49	0.000
. ***	. *	13	0.430	-0.140	349.18	0.000
. ***	** .	14	0.343	-0.276	360.00	0.000
. **	** .	15	0.276	-0.193	367.11	0.000
. **	. *	16	0.297	0.087	375.52	0.000
. **	. .	17	0.254	0.032	381.79	0.000
. *	. *	18	0.174	-0.188	384.77	0.000
. *	. .	19	0.181	-0.004	388.07	0.000
. *	. .	20	0.110	-0.033	389.30	0.000
. *	. *	21	0.106	0.101	390.47	0.000
. *	. .	22	0.075	0.039	391.07	0.000
. *	. .	23	0.103	0.042	392.23	0.000
. *	. *	24	0.090	-0.128	393.12	0.000
. .	. .	25	0.043	0.007	393.33	0.000
. .	. .	26	-0.010	0.026	393.34	0.000
* .	. *	27	-0.077	-0.098	394.04	0.000
. .	. *	28	-0.052	-0.066	394.37	0.000
* .	. .	29	-0.079	0.022	395.14	0.000
* .	. *	30	-0.099	0.066	396.38	0.000
* .	. .	31	-0.140	-0.005	398.94	0.000
* .	. *	32	-0.186	-0.137	403.53	0.000
* .	. .	33	-0.169	-0.010	407.42	0.000
** .	. .	34	-0.197	0.014	412.84	0.000
** .	. *	35	-0.200	-0.066	418.59	0.000
* .	. *	36	-0.163	0.087	422.54	0.000

b) Série com uma diferenciação

Date: 03/30/08 Time: 09:18

Sample: 2002:01 2007:12

Included observations: 71

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
**** .	**** .	1	-0.514	-0.514	19.554	0.000
. . .	** . .	2	0.055	-0.284	19.784	0.000
. . .	* . .	3	0.000	-0.162	19.784	0.000
* . .	*** . .	4	-0.164	-0.342	21.857	0.000
. **	. . .	5	0.273	-0.008	27.718	0.000
** . .	* . .	6	-0.222	-0.146	31.657	0.000
. * .	* . .	7	0.088	-0.121	32.285	0.000
. . .	* . .	8	0.006	-0.077	32.288	0.000
.	9	-0.031	-0.028	32.366	0.000
* . .	** . .	10	-0.077	-0.283	32.870	0.000
. . .	*** . .	11	-0.023	-0.358	32.914	0.001
. **	* . .	12	0.213	-0.090	36.890	0.000
. . .	. * .	13	-0.010	0.115	36.899	0.000
. . .	. **	14	-0.005	0.210	36.901	0.001
*	15	-0.188	-0.017	40.187	0.000
. *	16	0.093	-0.025	41.004	0.001
. * .	. * .	17	0.113	0.150	42.237	0.001
** . .	* . .	18	-0.213	-0.104	46.673	0.000
. *	19	0.183	-0.026	50.013	0.000
*	20	-0.143	-0.056	52.105	0.000
. *	21	0.078	-0.036	52.738	0.000
. . .	* . .	22	-0.049	-0.090	52.992	0.000
. . .	. * .	23	0.011	0.140	53.006	0.000
.	24	0.046	-0.025	53.244	0.001
. . .	* . .	25	0.037	-0.070	53.395	0.001
. . .	. * .	26	0.046	0.088	53.641	0.001
**	27	-0.195	0.010	58.142	0.000
. * .	* . .	28	0.082	-0.129	58.946	0.001
. . .	* . .	29	0.025	-0.067	59.025	0.001
. . .	* . .	30	-0.033	-0.068	59.158	0.001
. *	31	0.084	0.063	60.064	0.001
*	32	-0.166	-0.023	63.704	0.001
. *	33	0.130	-0.017	66.005	0.001
.	34	-0.027	0.055	66.104	0.001
* . .	* . .	35	-0.078	-0.088	66.982	0.001
. * .	* . .	36	0.139	-0.093	69.856	0.001

c) Série com duas diferenciações

Date: 03/30/08 Time: 09:23

Sample: 2002:01 2007:12

Included observations: 70

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
***** .	***** .	1	-0.685	-0.685	34.279	0.000
. **	***** .	2	0.205	-0.499	37.384	0.000
. .	** .	3	0.036	-0.240	37.481	0.000
** .	***** .	4	-0.201	-0.458	40.580	0.000
. **	** .	5	0.313	-0.196	48.189	0.000
** .	** .	6	-0.274	-0.197	54.086	0.000
. *	** .	7	0.137	-0.196	55.578	0.000
. .	. *	8	-0.015	-0.177	55.597	0.000
. .	. .	9	-0.004	0.055	55.598	0.000
. .	. .	10	-0.025	-0.001	55.652	0.000
. *	** .	11	-0.070	-0.298	56.072	0.000
. *.	** .	12	0.165	-0.314	58.432	0.000
. *	** .	13	-0.083	-0.206	59.039	0.000
. .	. *	14	0.063	0.109	59.394	0.000
. *	. *	15	-0.148	0.102	61.395	0.000
. *.	. .	16	0.077	-0.037	61.942	0.000
. *.	. **	17	0.122	0.208	63.352	0.000
** .	. *	18	-0.242	0.081	69.026	0.000
. **	. *	19	0.238	0.089	74.625	0.000
. *	. .	20	-0.178	0.037	77.813	0.000
. *.	. .	21	0.111	0.054	79.084	0.000
. *	. *	22	-0.066	-0.181	79.535	0.000
. .	. .	23	0.019	0.021	79.572	0.000
. .	. .	24	0.010	0.031	79.583	0.000
. .	. *	25	-0.006	-0.114	79.587	0.000
. *	. .	26	0.083	-0.010	80.372	0.000
. *	. *	27	-0.171	0.099	83.804	0.000
. *.	. .	28	0.109	-0.013	85.242	0.000
. .	. .	29	0.003	-0.021	85.243	0.000
. *	. *	30	-0.062	-0.113	85.723	0.000
. *.	. .	31	0.125	0.007	87.731	0.000
. *	. .	32	-0.182	0.005	92.104	0.000
. *.	. .	33	0.146	-0.034	95.021	0.000
. *	. *	34	-0.037	0.081	95.212	0.000
. *	. .	35	-0.084	0.009	96.238	0.000
. *	. *	36	0.145	-0.094	99.341	0.000

ANEXO 2 – Saídas computacionais dos modelos univariados

Modelo 1: SARIMA (0,0,1) (0,1,1)

Estimation Command:

=====

```
LS(DERIV=AA) D(CUSTOTOTAL,0,1) MA(1) SMA(12)
```

Estimation Equation:

=====

```
D(CUSTOTOTAL,0,1) = 0 + [MA(1)=C(1),SMA(12)=C(2),BACKCAST=2002M02]
```

Substituted Coefficients:

=====

```
D(CUSTOTOTAL,0,1) = 0 + [MA(1)=-0.6379761305,SMA(12)=0.8219492266,BACKCAST=2002M02]
```

Dependent Variable: D(CUSTOTOTAL,0,1)

Method: Least Squares

Date: 04/21/08 Time: 10:28

Sample(adjusted): 2002:02 2006:12

Included observations: 59 after adjusting endpoints

Convergence achieved after 8 iterations

Backcast: 2001:01 2002:01

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	-0.637976	0.097460	-6.546033	0.0000
SMA(12)	0.821949	0.052825	15.55988	0.0000
R-squared	0.564496	Mean dependent var		6325.271
Adjusted R-squared	0.556855	S.D. dependent var		188905.8
S.E. of regression	125752.9	Akaike info criterion		26.35534
Sum squared resid	9.01E+11	Schwarz criterion		26.42576
Log likelihood	-775.4824	Durbin-Watson stat		2.258158
Inverted MA Roots	.95+.25i	.95 -.25i	.70 -.70i	.70+.70i
	.64	.25+.95i	.25 -.95i	-.25+.95i
	-.25 -.95i	-.70+.70i	-.70+.70i	-.95 -.25i
	-.95+.25i			

Modelo 2: SARIMA (1,0,1) (1,0,0)

Estimation Command:

=====

```
LS(DERIV=AA) D(CUSTOTOTAL,1,1) AR(1) MA(1) SAR(12)
```

Estimation Equation:

=====

```
D(CUSTOTOTAL,1,1) = 0 + [AR(1)=C(1),SAR(12)=C(2),MA(1)=C(3),BACKCAST=2003M04]
```

Substituted Coefficients:

=====

```
D(CUSTOTOTAL,1,1) = 0 + [AR(1)=-0.5827651553,SAR(12)=0.3811575328,MA(1)=-0.9596962158,BACKCAST=2003M04]
```

Dependent Variable: D(CUSTOTOTAL,1,1)

Method: Least Squares

Date: 03/11/08 Time: 14:29

Sample (adjusted): 2003M04 2006M12

Included observations: 45 after adjustments

Convergence achieved after 12 iterations

Backcast: 2003M03

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.582765	0.133392	-4.368804	0.0001
SAR(12)	0.381158	0.136525	2.791848	0.0079
MA(1)	-0.959696	0.019273	-49.79506	0.0000
R-squared	0.751675	Mean dependent var		-7585.556
Adjusted R-squared	0.739850	S.D. dependent var		311376.0
S.E. of regression	158816.9	Akaike info criterion		26.85323
Sum squared resid	1.06E+12	Schwarz criterion		26.97368
Log likelihood	-601.1977	Durbin-Watson stat		2.383134
Inverted AR Roots	.92	.80+.46i	.80 -.46i	.46+.80i
	.46 -.80i	.00 -.92i	-.00+.92i	-.46+.80i
	-.46 -.80i	-.58	-.80 -.46i	-.80+.46i
	-.92			
Inverted MA Roots	.96			

Modelo 3: SARIMA (2,1,1) (0,0,1)

Estimation Command:

=====

```
LS D(CUSTOTOTAL, 1,1) AR(1) AR(2) MA(1) SMA(12)
```

Estimation Equation:

=====

```
D(CUSTOTOTAL,1,1) = 0 + [AR(1)=C(1),AR(2)=C(2),MA(1)=C(3),SMA(12)=C(4),BACKCAST=2002:05]
```

Substituted Coefficients:

=====

```
D(CUSTOTOTAL,1,1) = 0 + [AR(1)=-0.7872240506,AR(2)=-0.4860679761,MA(1)=-0.9715971791,SMA(12)=0.8585567461,BACKCAST=2002:05]
```

Dependent Variable: D(CUSTOTOTAL,1,1)

Method: Least Squares

Date: 04/27/08 Time: 18:41

Sample(adjusted): 2002:05 2006:12

Included observations: 56 after adjusting endpoints

Convergence achieved after 15 iterations

Backcast: 2001:04 2002:04

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	-0.787224	0.132334	-5.948746	0.0000
AR(2)	-0.486068	0.088793	-5.474158	0.0000
MA(1)	-0.971597	0.032309	-30.07181	0.0000
SMA(12)	0.858557	0.043404	19.78037	0.0000
R-squared	0.865601	Mean dependent var		-5833.607
Adjusted R-squared	0.857847	S.D. dependent var		324040.1
S.E. of regression	122173.2	Akaike info criterion		26.33302
Sum squared resid	7.76E+11	Schwarz criterion		26.47768
Log likelihood	-733.3245	Durbin-Watson stat		1.818584
Inverted AR Roots	-.39 -.58i	-.39+.58i		
Inverted MA Roots	.97	.95 -.26i	.95+.26i	.70 -.70i
	.70 -.70i	.26 -.95i	.26+.95i	-.26 -.95i
	-.26+.95i	-.70+.70i	-.70+.70i	-.95+.26i
	-.95 -.26i			

Modelo 4: SARIMA (2,0,0) (0,0,1)

Estimation Command:

=====
LS(DERIV=AA) CUSTOTOTAL C AR(1) AR(2) SMA(12)

Estimation Equation:

=====
CUSTOTOTAL = C(1) + [AR(1)=C(2),AR(2)=C(3),MA(12)=C(4)]

Substituted Coefficients:

=====
CUSTOTOTAL = 1620790.54 + [AR(1)=0.3583157616,AR(2)=0.3691262634,MA(12)=0.8626441818]

Dependent Variable: CUSTOTOTAL

Method: Least Squares

Date: 03/11/08 Time: 14:31

Sample (adjusted): 2002M03 2006M12

Included observations: 58 after adjustments

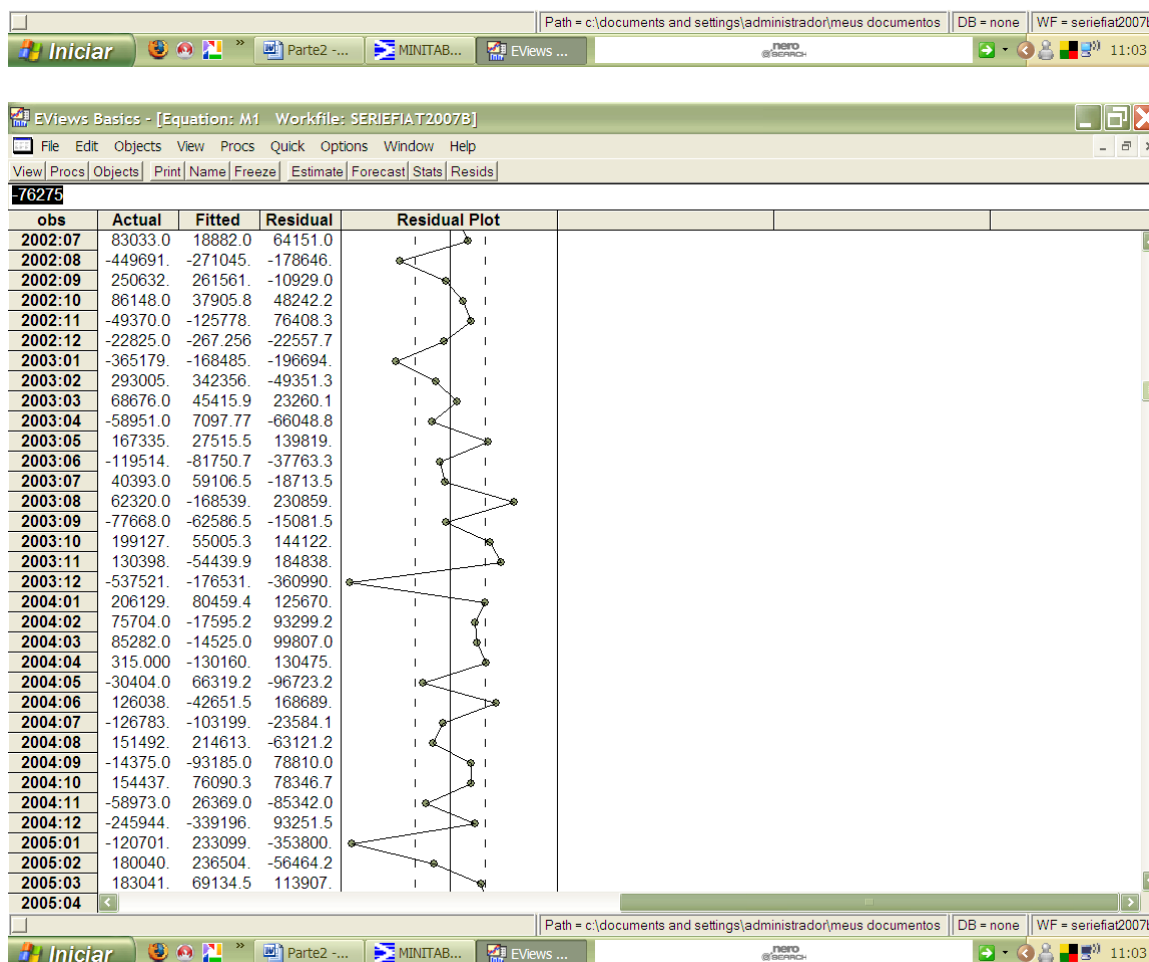
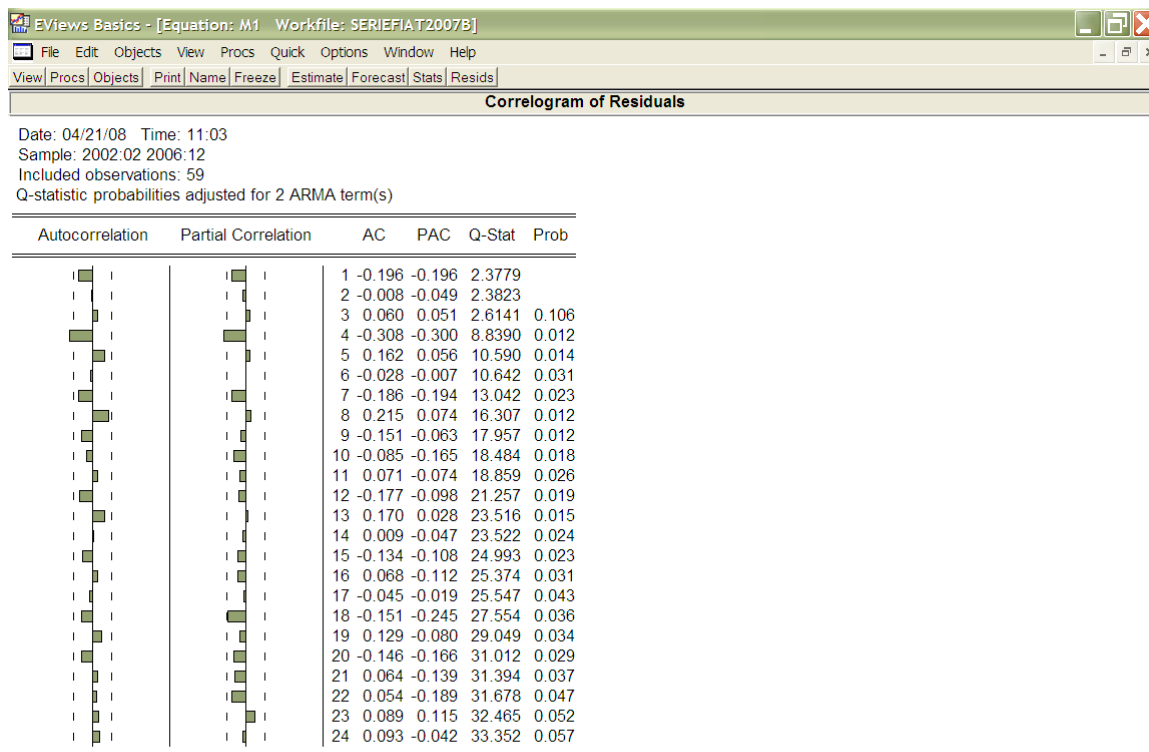
Convergence achieved after 17 iterations

Backcast: 2001M03 2002M02

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1620791.	111521.3	14.53347	0.0000
AR(1)	0.358316	0.128948	2.778757	0.0075
AR(2)	0.369126	0.125472	2.941890	0.0048
MA(12)	0.862644	0.050271	17.15991	0.0000
R-squared	0.741185	Mean dependent var		1553826.
Adjusted R-squared	0.726807	S.D. dependent var		235182.8
S.E. of regression	122925.1	Akaike info criterion		26.34301
Sum squared resid	8.16E+11	Schwarz criterion		26.48511
Log likelihood	-759.9473	F-statistic		51.54784
Durbin-Watson stat	2.287763	Prob(F-statistic)		0.000000
Inverted AR Roots	.81	-.45		
Inverted MA Roots	.95+.26i	.95 -.26i	.70 -.70i	.70+.70i
	.26+.95i	.26 -.95i	-.26+.95i	-.26 -.95i
	-.70 -.70i	-.70 -.70i	-.95+.26i	-.95 -.26i

ANEXO 3 – Análise de resíduos dos modelos univariados

Modelo 1 : Fac, facp, plot e teste de heterocedasticidade dos resíduos do modelo univariado 1



H0: resíduos não são heterocedásticos

ARCH Test:

F-statistic	0.042822	Probability	0.988072
Obs*R-squared	0.138007	Probability	0.986915

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

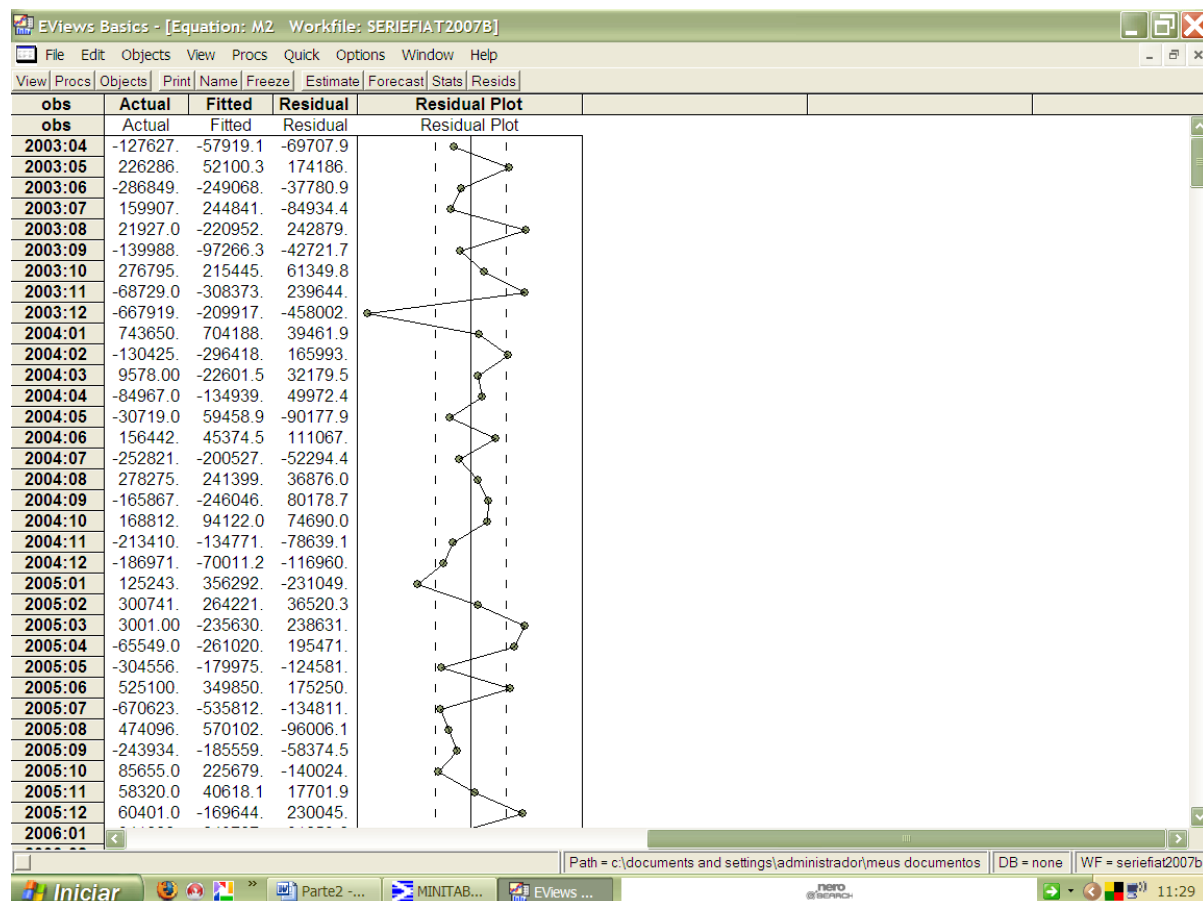
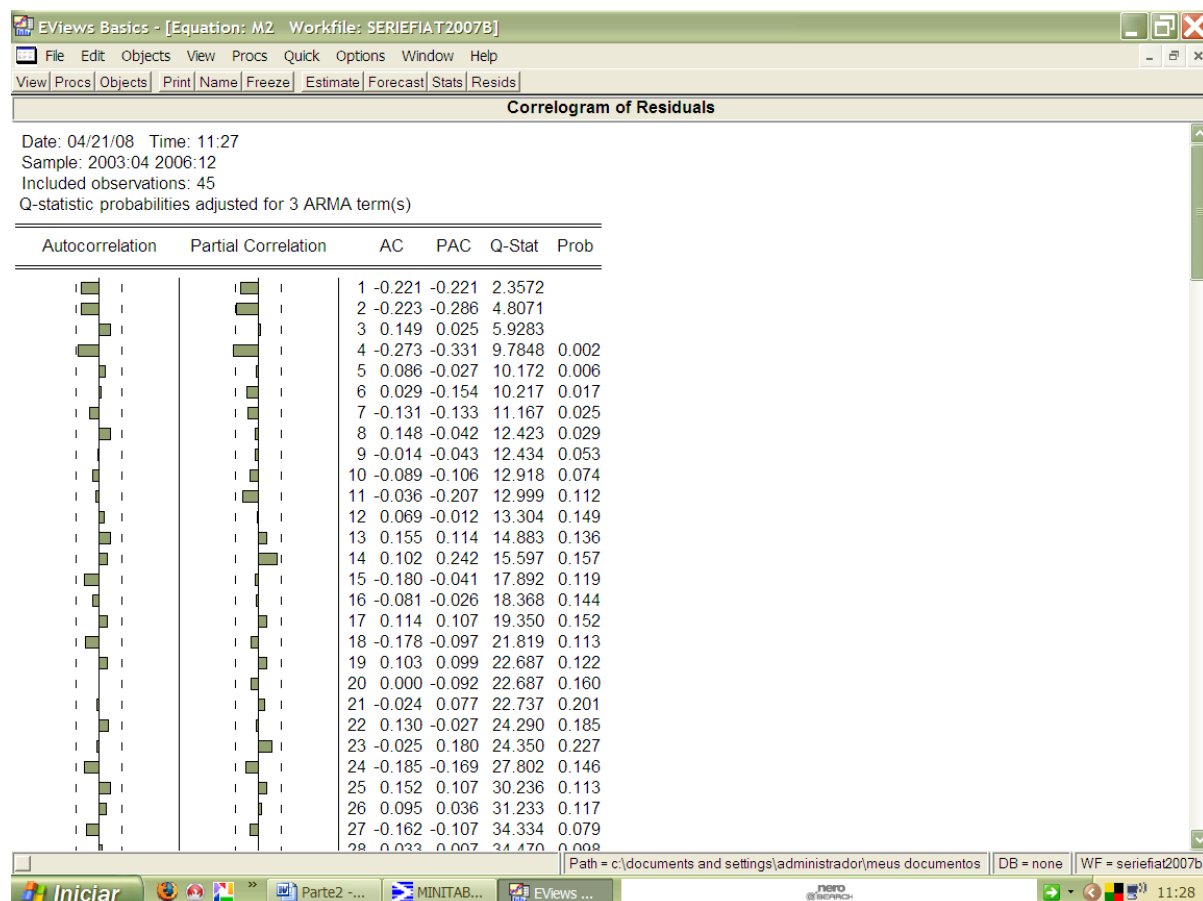
Date: 04/27/08 Time: 19:25

Sample(adjusted): 2002:05 2006:12

Included observations: 56 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.68E+10	5.04E+09	3.327637	0.0016
RESID^2(-1)	-0.015554	0.140511	-0.110697	0.9123
RESID^2(-2)	-0.033302	0.140680	-0.236721	0.8138
RESID^2(-3)	-0.035237	0.141148	-0.249648	0.8038
R-squared	0.002464	Mean dependent var		1.55E+10
Adjusted R-squared	-0.055086	S.D. dependent var		2.50E+10
S.E. of regression	2.57E+10	Akaike info criterion		50.84684
Sum squared resid	3.44E+22	Schwarz criterion		50.99151
Log likelihood	-1419.712	F-statistic		0.042822
Durbin-Watson stat	1.967782	Prob(F-statistic)		0.988072

Modelo 2: Fac, facp, plot e teste de heterocedasticidade dos resíduos do modelo univariado 2



H0: resíduos não são heterocedásticos

ARCH Test:

F-statistic	0.096211	Probability	0.961660
Obs*R-squared	0.316609	Probability	0.956875

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

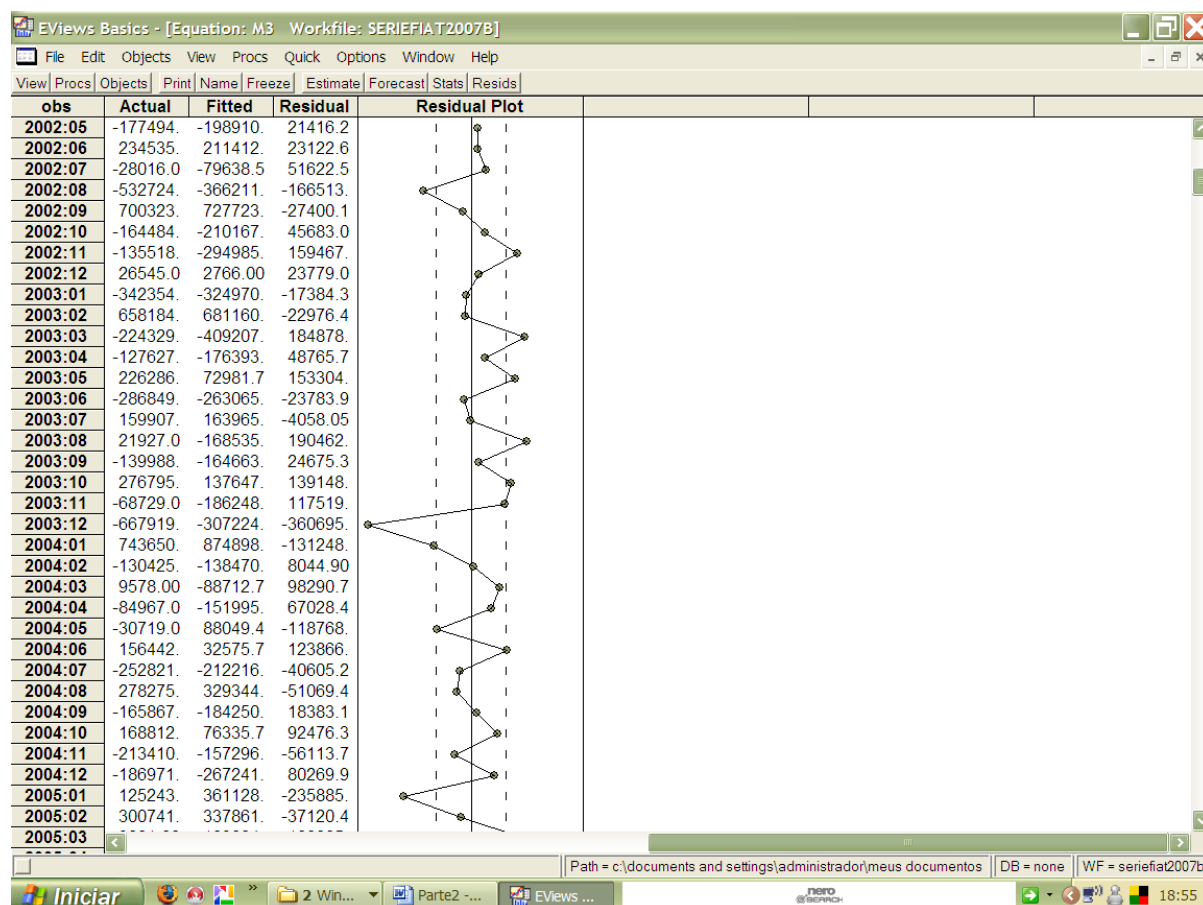
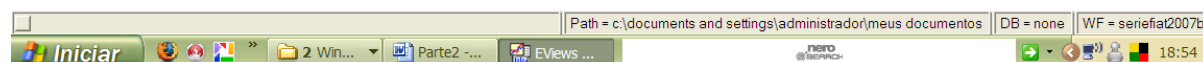
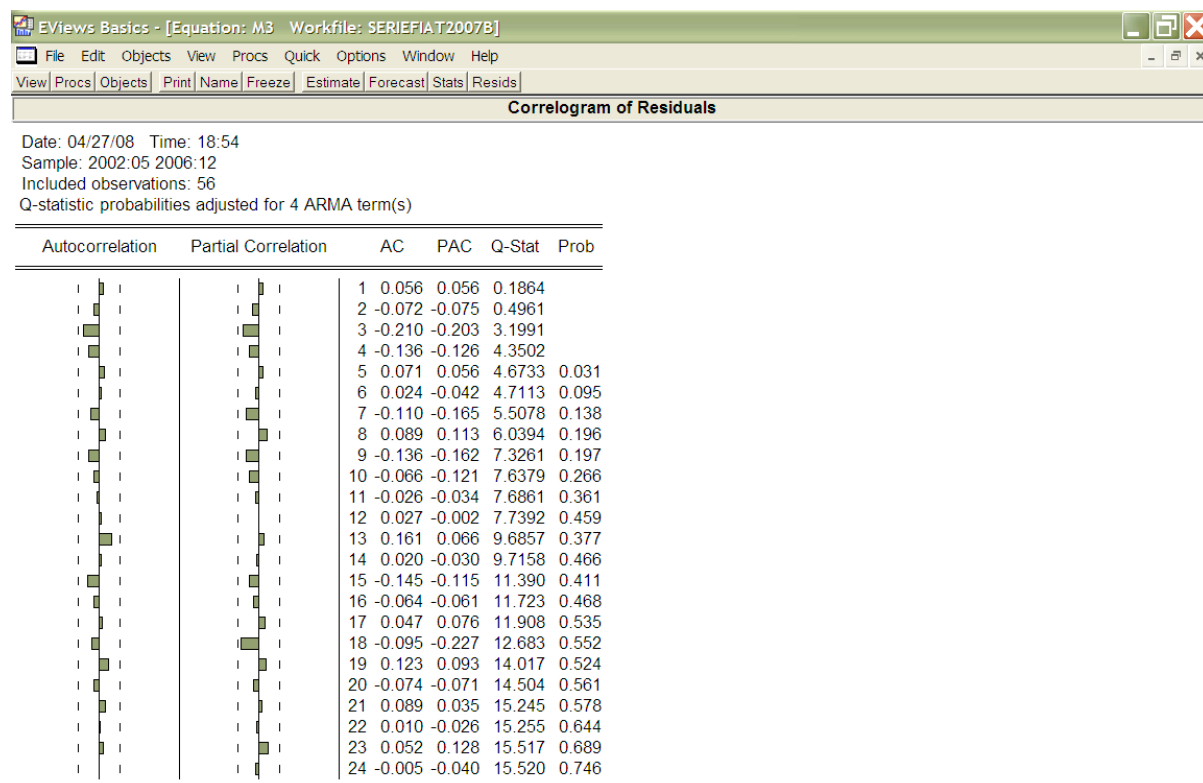
Date: 04/27/08 Time: 19:27

Sample(adjusted): 2003:07 2006:12

Included observations: 42 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.56E+10	8.49E+09	3.020089	0.0045
RESID^2(-1)	0.041618	0.162459	0.256176	0.7992
RESID^2(-2)	-0.027957	0.162469	-0.172075	0.8643
RESID^2(-3)	-0.069960	0.161877	-0.432181	0.6681
R-squared	0.007538	Mean dependent var		2.44E+10
Adjusted R-squared	-0.070814	S.D. dependent var		3.52E+10
S.E. of regression	3.64E+10	Akaike info criterion		51.56641
Sum squared resid	5.05E+22	Schwarz criterion		51.73190
Log likelihood	-1078.895	F-statistic		0.096211
Durbin-Watson stat	1.989278	Prob(F-statistic)		0.961660

Modelo 3: Fac, facp, plot e teste de heterocedasticidade dos resíduos do modelo univariado 3



H0: resíduos não são heterocedásticos

ARCH Test:

F-statistic	0.171920	Probability	0.914864
Obs*R-squared	0.552051	Probability	0.907316

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

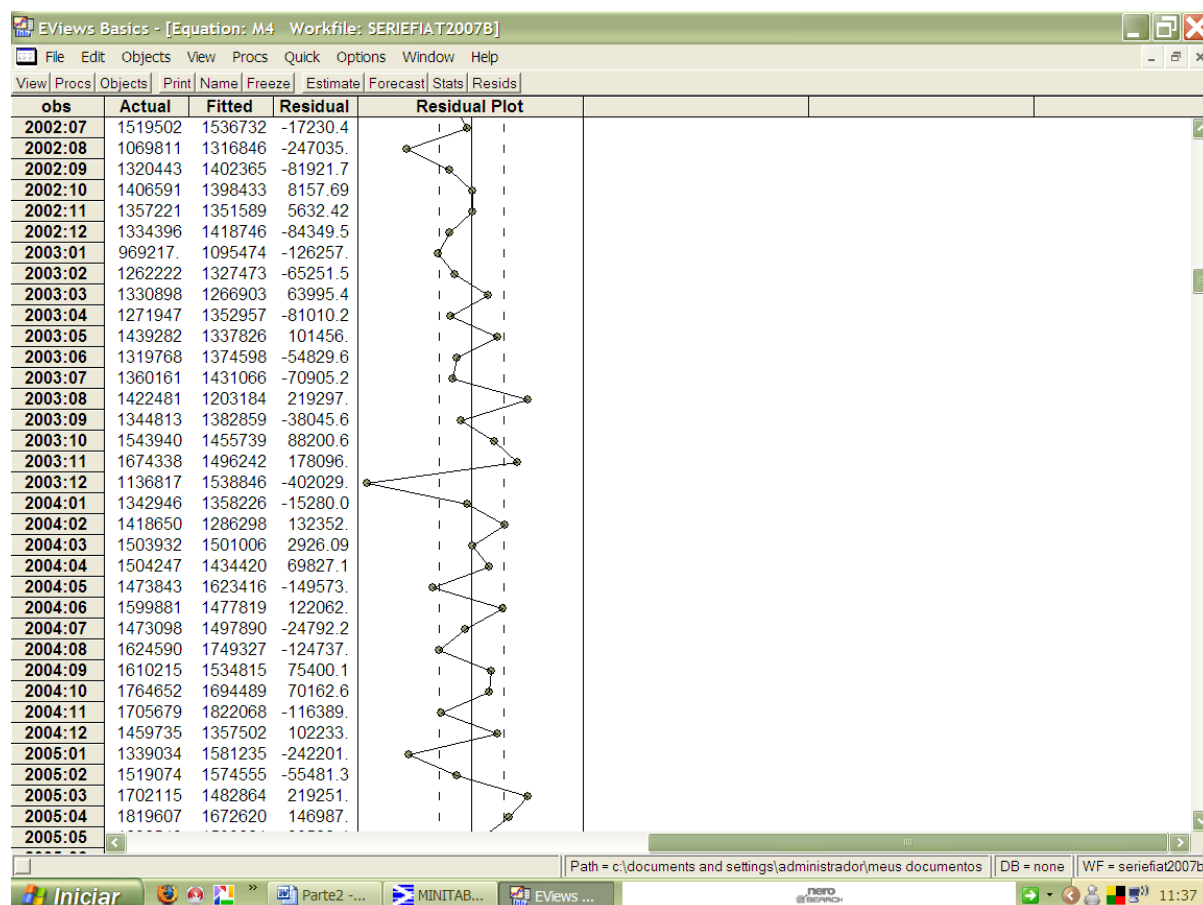
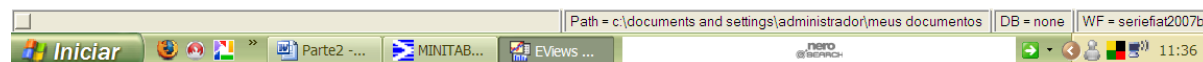
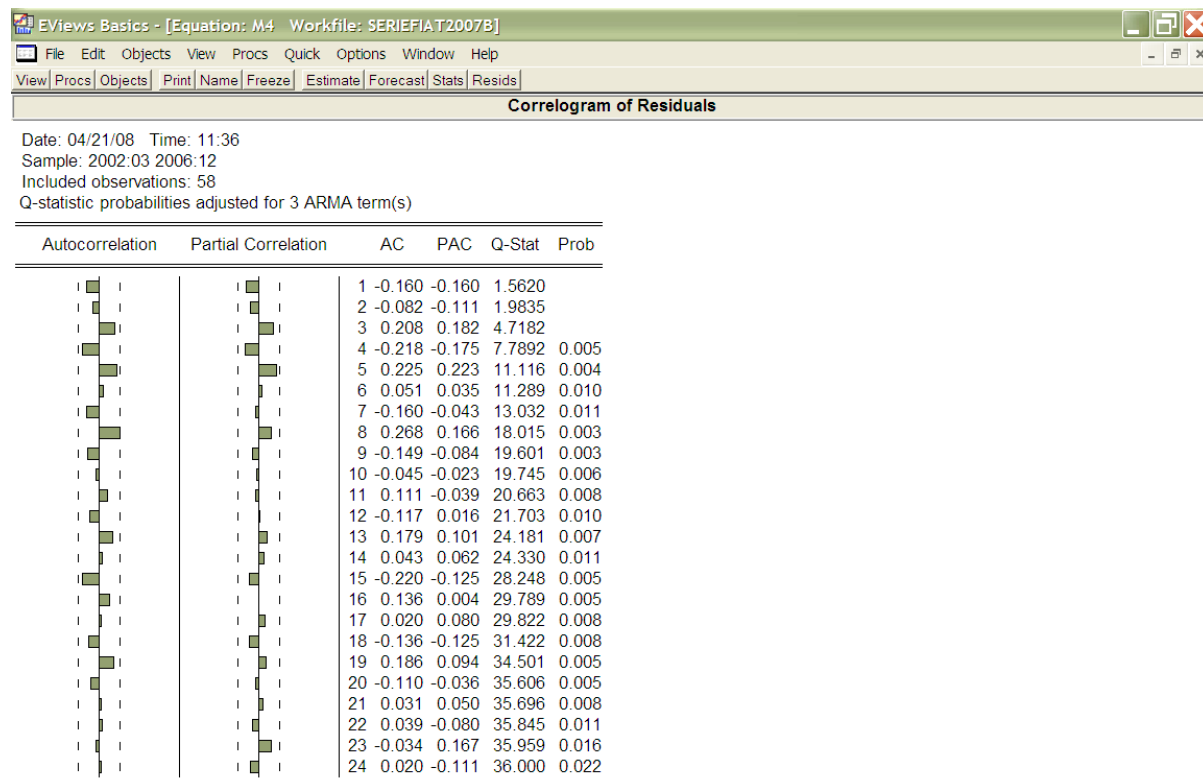
Date: 04/27/08 Time: 19:21

Sample(adjusted): 2002:08 2006:12

Included observations: 53 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.39E+10	4.75E+09	2.919281	0.0053
RESID^2(-1)	-0.013230	0.142789	-0.092655	0.9266
RESID^2(-2)	-0.032383	0.142805	-0.226765	0.8216
RESID^2(-3)	0.096223	0.142856	0.673568	0.5038
R-squared	0.010416	Mean dependent var		1.46E+10
Adjusted R-squared	-0.050171	S.D. dependent var		2.30E+10
S.E. of regression	2.35E+10	Akaike info criterion		50.67475
Sum squared resid	2.72E+22	Schwarz criterion		50.82345
Log likelihood	-1338.881	F-statistic		0.171920
Durbin-Watson stat	1.966227	Prob(F-statistic)		0.914864

Modelo 4: Fac, facp, plot e teste de heterocedasticidade dos resíduos do modelo univariado 4



H0: resíduos não são heterocedásticos

ARCH Test:

F-statistic	0.075055	Probability	0.973142
Obs*R-squared	0.241758	Probability	0.970582

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

Date: 04/27/08 Time: 19:28

Sample(adjusted): 2002:06 2006:12

Included observations: 55 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.61E+10	4.89E+09	3.295362	0.0018
RESID^2(-1)	-0.030667	0.139748	-0.219443	0.8272
RESID^2(-2)	-0.007974	0.139462	-0.057174	0.9546
RESID^2(-3)	-0.058262	0.139376	-0.418020	0.6777
R-squared	0.004396	Mean dependent var		1.47E+10
Adjusted R-squared	-0.054169	S.D. dependent var		2.48E+10
S.E. of regression	2.54E+10	Akaike info criterion		50.82595
Sum squared resid	3.30E+22	Schwarz criterion		50.97193
Log likelihood	-1393.714	F-statistic		0.075055
Durbin-Watson stat	1.991499	Prob(F-statistic)		0.973142

ANEXO 4 – Saídas computacionais do modelo dinâmico 1

MODELO DINÂMICO 1

Dependent Variable: CUSTOTOTAL

Method: Least Squares

Date: 06/04/08 Time: 09:16

Sample: 2002:01 2006:12

Included observations: 60

Variable	Coefficient	Std. Error	t-Statistic	Prob.
VIDAS	6.247701	7.866773	0.794189	0.4305
FIPESAÚDE	3303.949	457.3053	7.224820	0.0000
SINISTRALIDADE	3.807727	2.920356	1.303857	0.1977
SALARIOS	1.420090	0.244707	5.803232	0.0000
C	-453629.8	288758.6	-1.570965	0.1219
R-squared	0.805889	Mean dependent var	1544770.	
Adjusted R-squared	0.791772	S.D. dependent var	236440.4	
S.E. of regression	107892.5	Akaike info criterion	26.09531	
Sum squared resid	6.40E+11	Schwarz criterion	26.26984	
Log likelihood	-777.8594	F-statistic	57.08586	
Durbin-Watson stat	2.251664	Prob(F-statistic)	0.000000	

Cálculo do VIF para o modelo dinâmico 1

Regression Analysis: CUSTO TOTAL versus VIDAS; IPC FIPESAÚDE; ...

The regression equation is

$$\text{CUSTO TOTAL} = -453630 + 6,25 \text{ VIDAS} + 3304 \text{ IPC FIPESAÚDE} + 3,81 \text{ TOTAL EVENTOS} + 1,42 \text{ SALÁRIOS}$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	-453630	288759	-1,57	0,122	
VIDAS	6,248	7,867	0,79	0,430	1,2
IPC FIPESAÚDE	3303,9	457,3	7,22	0,000	1,5
TOTAL EVENTOS	3,808	2,920	1,30	0,198	1,6
SALÁRIOS	1,4201	0,2447	5,80	0,000	1,5

S = 107892 R-Sq = 80,6% R-Sq(adj) = 79,2%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	2,65810E+12	6,64524E+11	57,09	0,000
Residual Error	55	6,40243E+11	11640777545		
Total	59	3,29834E+12			

Source	DF	Seq SS
VIDAS	1	74132304141
IPC FIPESAÚDE	1	2,10469E+12
TOTAL EVENTOS	1	87247019675
SALÁRIOS	1	3,92030E+11

Unusual Observations

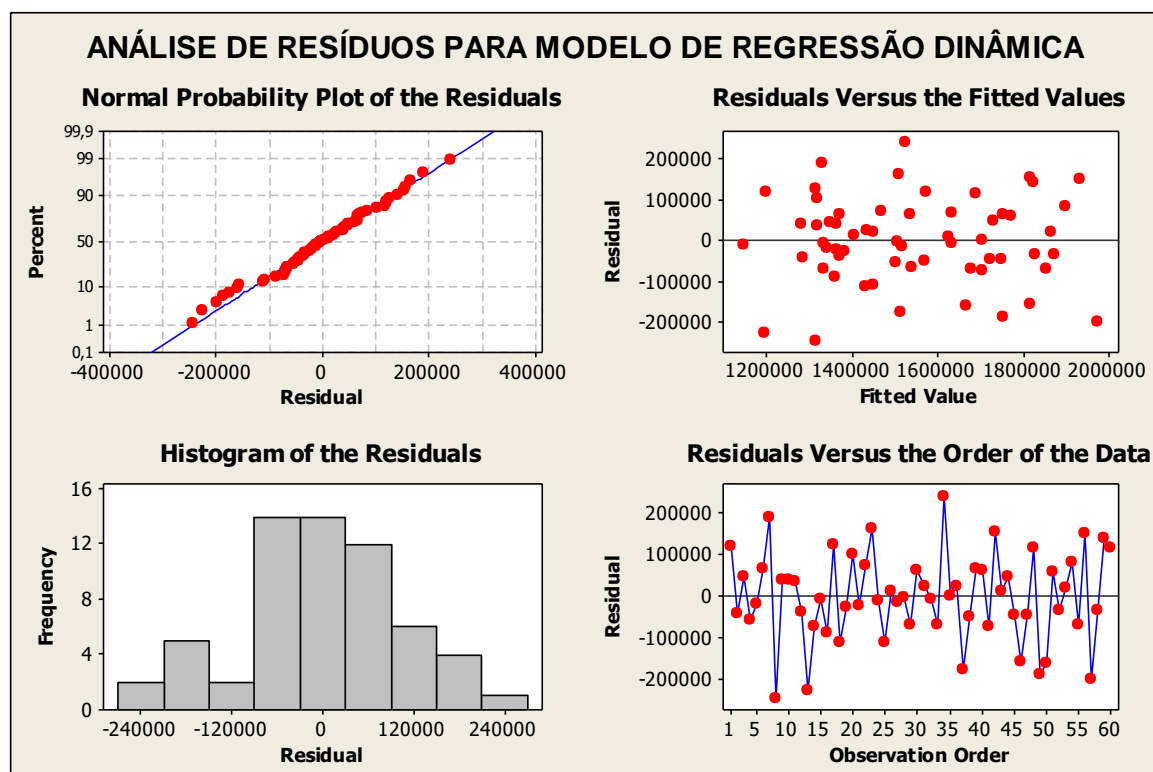
Obs	VIDAS	CUSTO TOTAL	Fit	SE Fit	Residual	St Resid
8	38286	1069811	1313341	26011	-243530	-2,33R
13	37440	969217	1194586	32221	-225368	-2,19R
34	36738	1764652	1524603	18407	240049	2,26R
60	40498	1693467	1574421	57138	119046	1,30 X

R denotes an observation with a large standardized residual.

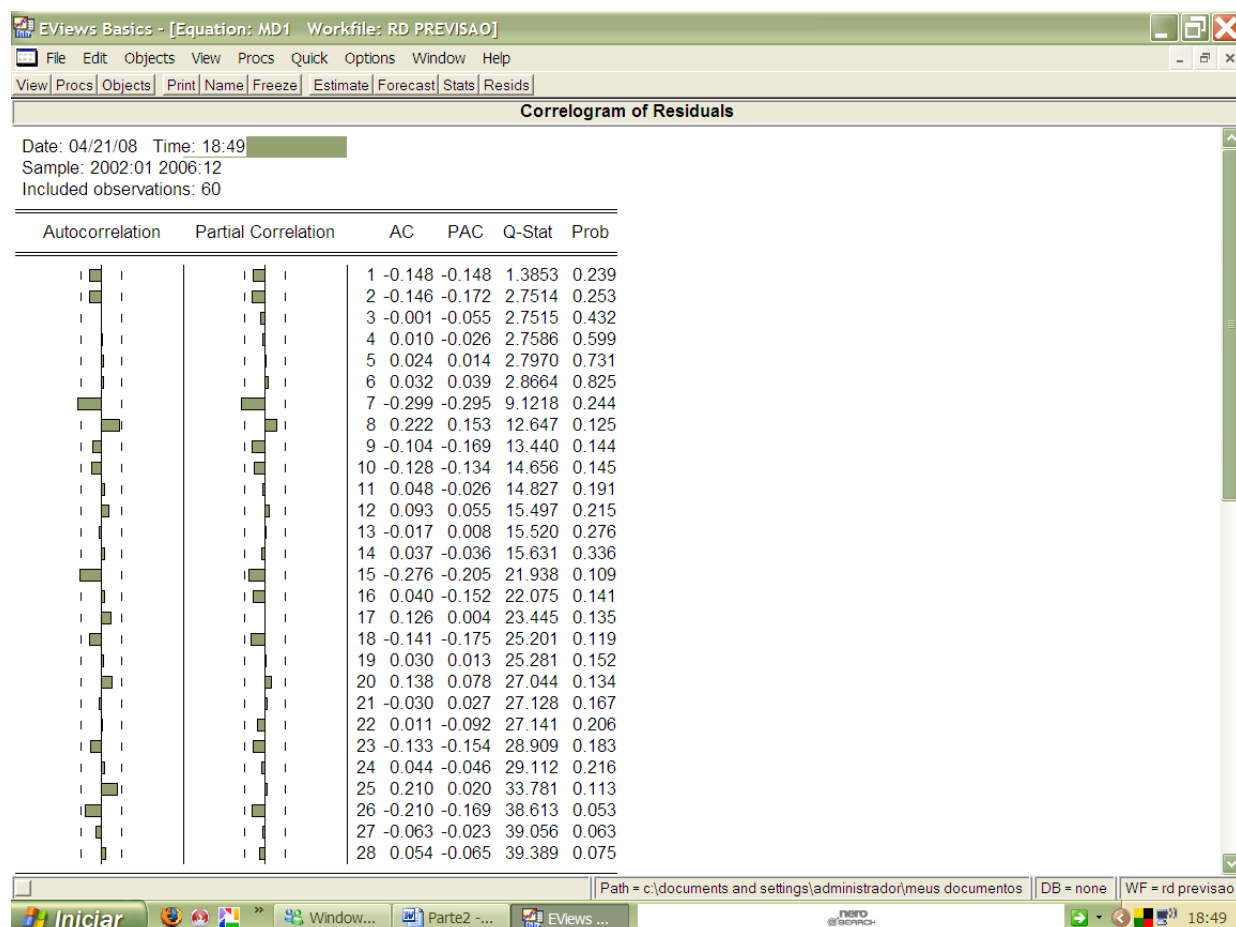
X denotes an observation whose X value gives it large influence.

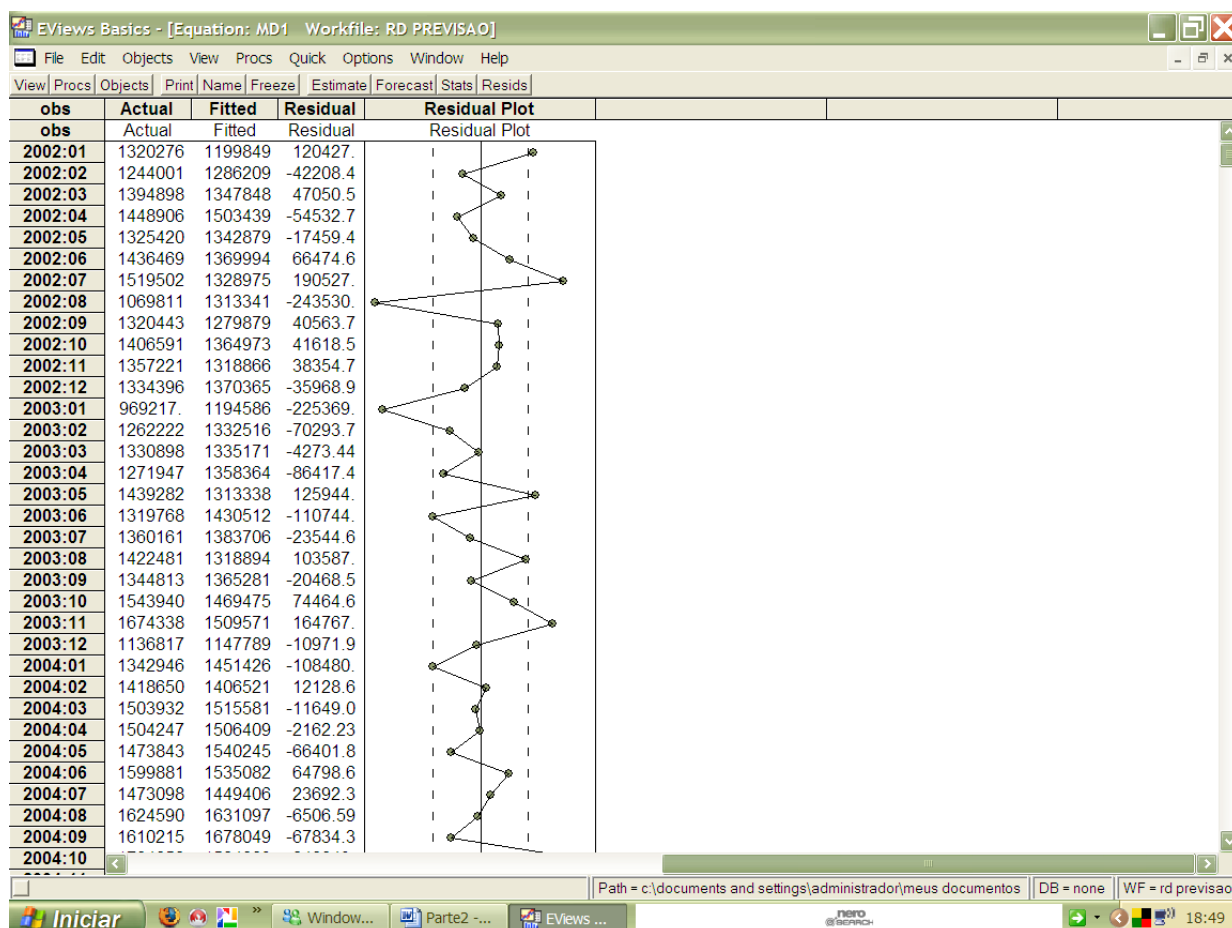
Durbin-Watson statistic = 2,25167

Análise de resíduos do modelo dinâmico 1:



Análise de resíduos do modelo dinâmico 1: fac, facp e plot dos resíduos





Teste de heterocedasticidade dos resíduos do modelo dinâmico 1

H0: resíduos não são heterocedásticos

White Heteroskedasticity Test:

F-statistic	1.338802	Probability	0.246190
Obs*R-squared	10.41356	Probability	0.237191

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

Date: 06/04/08 Time: 09:20

Sample: 2002:01 2006:12

Included observations: 60

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.49E+12	9.63E+11	-1.547202	0.1280
VIDAS	80944109	43907168	1.843528	0.0711
VIDAS^2	-1048.508	559.9984	-1.872341	0.0669
FIPESAUDE	-1.35E+09	1.95E+09	-0.693517	0.4911
FIPESAUDE^2	2188459.	3235004.	0.676493	0.5018
SINISTRALIDADE	7727458.	4086948.	1.890765	0.0643
SINISTRALIDADE^2	-71.98904	37.62379	-1.913392	0.0613
SALARIOS	-272497.3	226715.8	-1.201933	0.2349
SALARIOS^2	0.322987	0.295576	1.092736	0.2796
R-squared	0.173559	Mean dependent var	1.07E+10	
Adjusted R-squared	0.043921	S.D. dependent var	1.47E+10	
S.E. of regression	1.44E+10	Akaike info criterion	49.75849	
Sum squared resid	1.06E+22	Schwarz criterion	50.07264	
Log likelihood	-1483.755	F-statistic	1.338802	
Durbin-Watson stat	2.067104	Prob(F-statistic)	0.246190	

ANEXO 5 - Saídas computacionais do modelo dinâmico 2

Dependent Variable: CUSTOTOTAL

Method: Least Squares

Date: 06/04/08 Time: 09:26

Sample: 2002:01 2006:12

Included observations: 60

Variable	Coefficient	Std. Error	t-Statistic	Prob.
FIPESAÚDE	3268.784	453.6536	7.205463	0.0000
SINISTRALIDADE	4.816290	2.621098	1.837508	0.0714
SALÁRIOS	1.408607	0.243472	5.785487	0.0000
C	-254232.0	142164.3	-1.788297	0.0791
R-squared	0.803663	Mean dependent var	1544770.	
Adjusted R-squared	0.793145	S.D. dependent var	236440.4	
S.E. of regression	107536.1	Akaike info criterion	26.07338	
Sum squared resid	6.48E+11	Schwarz criterion	26.21301	
Log likelihood	-778.2015	F-statistic	76.40807	
Durbin-Watson stat	2.246211	Prob(F-statistic)	0.000000	

Cálculo do VIF para o modelo dinâmico 2

Regression Analysis: CUSTO-SAUDE versus FIPESAÚDE; SALÁRIOS; ...

The regression equation is

CUSTO-SAUDE = - 254229 + 0,327 FIPESAÚDE + 1,41 SALÁRIOS + 4,82 SINISTRALIDADE

Predictor	Coef	SE Coef	T	P	VIF
Constant	-254232	142164	-1,79	0,079	
IPC FIPESAÚDE	3268,8	453,7	7,21	0,000	1,5
SINISTRALIDADE	4,816	2,621	1,84	0,071	1,3
SALÁRIOS	1,4086	0,2435	5,79	0,000	1,5

S = 107536 R-Sq = 80,4% R-Sq(adj) = 79,3%

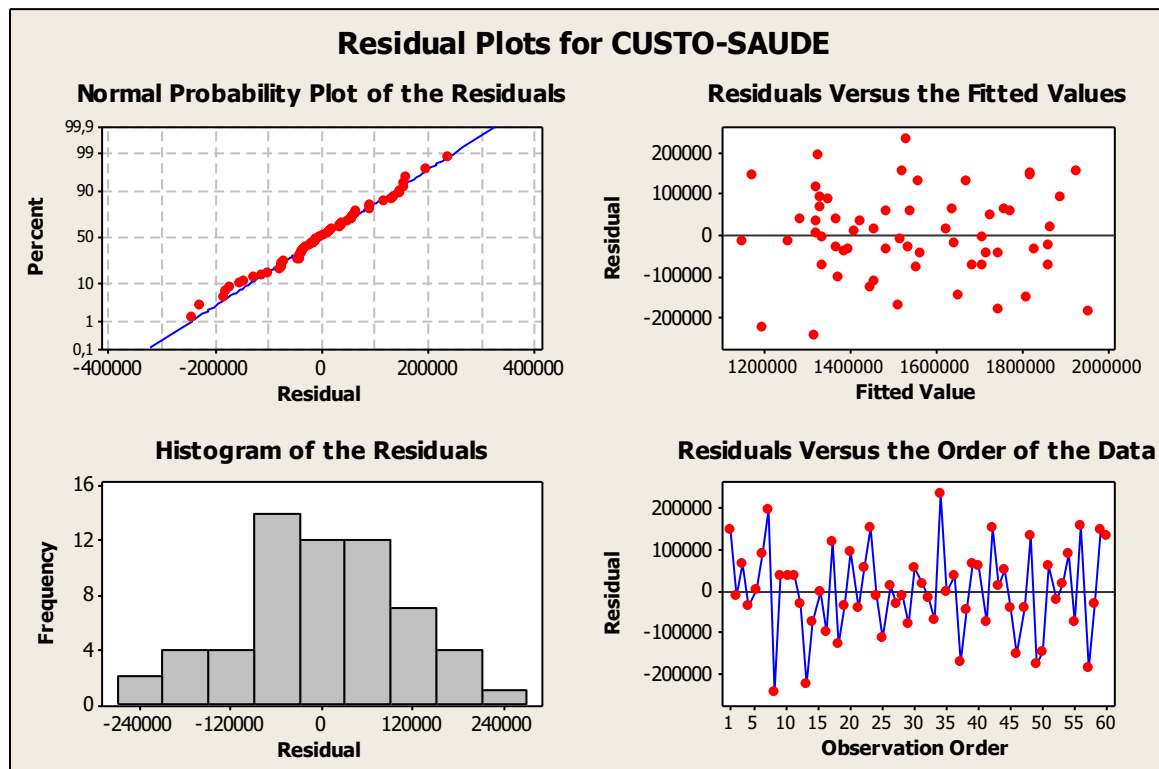
Analysis of Variance

Source	DF	SS	MS	F	P
Regression	3	2,65076E+12	8,83585E+11	76,41	0,000
Residual Error	56	6,47585E+11	11564022094		
Total	59	3,29834E+12			

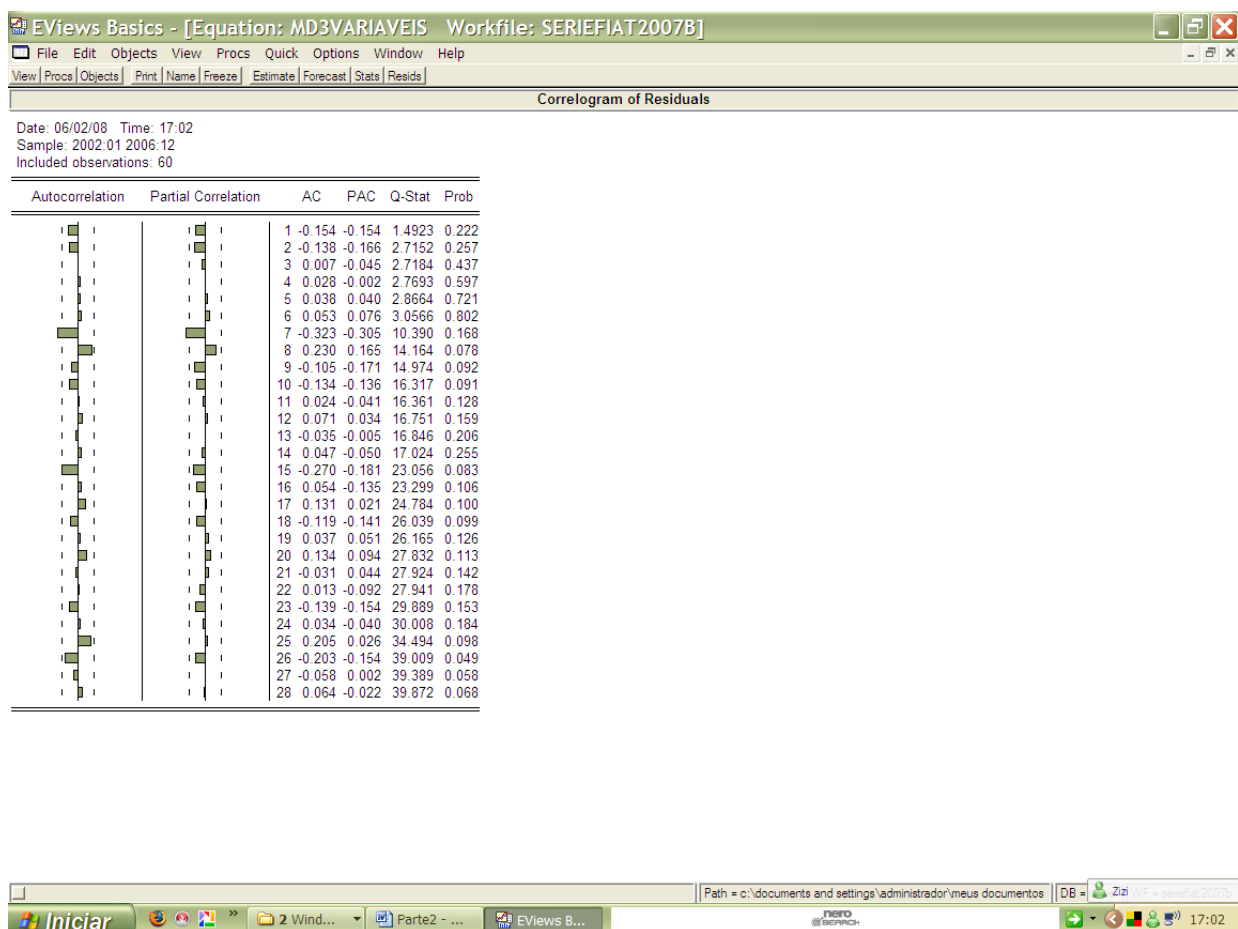
Source	DF	Seq SS
FIPESAÚDE	1	2,14027E+12
SALÁRIOS	1	4,71444E+11
SINISTRALIDADE	1	39044687358

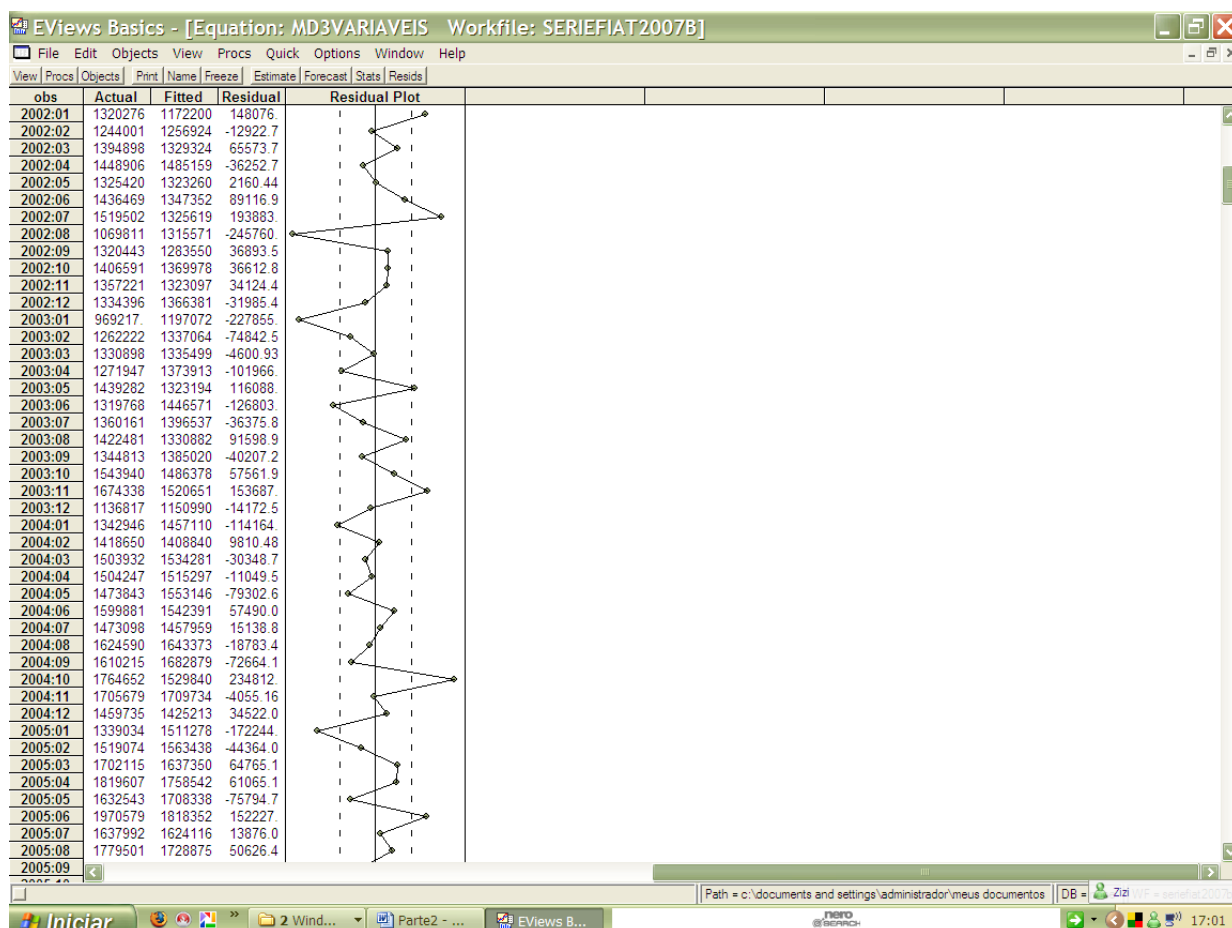
Durbin-Watson statistic = 2,24621

Análise de resíduos do modelo dinâmico 2:



Análise de resíduos do modelo dinâmico 2: fac, facp e plot dos resíduos





Teste de heterocedasticidade dos resíduos do modelo dinâmico 2

H0: resíduos não são heterocedásticos

White Heteroskedasticity Test:

F-statistic	1.270214	Probability	0.286857
Obs*R-squared	7.543174	Probability	0.273517

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

Date: 06/04/08 Time: 09:27

Sample: 2002:01 2006:12

Included observations: 60

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2.87E+10	1.98E+11	0.145239	0.8851
FIPESAUDE	-1.26E+09	1.12E+09	-1.129799	0.2636
FIPESAUDE^2	2173358.	1852707.	1.173072	0.2460
SINISTRALIDADE	8104935.	3986835.	2.032925	0.0471
SINISTRALIDADE^2	-75.58863	36.61612	-2.064354	0.0439
SALARIOS	-243230.4	222075.5	-1.095260	0.2784
SALARIOS^2	0.269896	0.289823	0.931243	0.3560
R-squared	0.125720	Mean dependent var	1.08E+10	
Adjusted R-squared	0.026744	S.D. dependent var	1.45E+10	
S.E. of regression	1.43E+10	Akaike info criterion	49.71327	
Sum squared resid	1.08E+22	Schwarz criterion	49.95761	
Log likelihood	-1484.398	F-statistic	1.270214	
Durbin-Watson stat	1.947244	Prob(F-statistic)	0.286857	

ANEXO 6 – Saídas computacionais dos modelos dinâmicos modelos dinâmicos 3, 4 e 5 (com tratamento de dados atípicos):

Primeiro passo

Modelo 3: tratamento para atípico 1

Dependent Variable: CUSTOTOTAL

Method: Least Squares

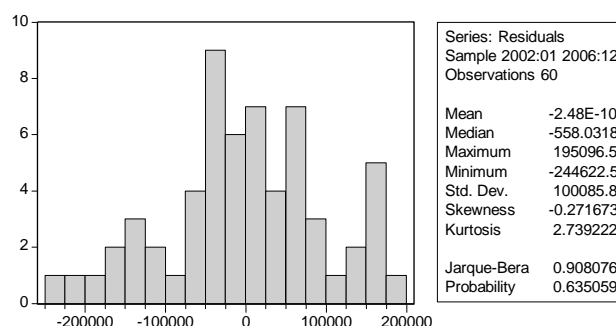
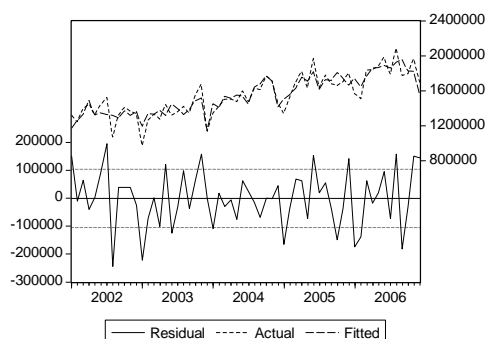
Date: 06/04/08 Time: 09:30

Sample: 2002:01 2006:12

Included observations: 60

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-267765.2	137168.8	-1.952086	0.0560
FIPESAUDE	3201.230	438.2977	7.303779	0.0000
SALARIOS	1.428523	0.234860	6.082438	0.0000
SINISTRALIDADE	5.235750	2.533261	2.066802	0.0435
AT1	240924.6	105002.0	2.294476	0.0256
R-squared	0.820815	Mean dependent var	1544770.	
Adjusted R-squared	0.807783	S.D. dependent var	236440.4	
S.E. of regression	103661.4	Akaike info criterion	26.01530	
Sum squared resid	5.91E+11	Schwarz criterion	26.18983	
Log likelihood	-775.4591	F-statistic	62.98629	
Durbin-Watson stat	2.227852	Prob(F-statistic)	0.000000	

Análise de resíduos do modelo dinâmico 3: plot dos resíduos e histograma com teste de normalidade



Segundo passo: tratamento para atípico 1 e atípico 2

Modelo dinâmico 4:

Dependent Variable: CUSTOTOTAL

Method: Least Squares

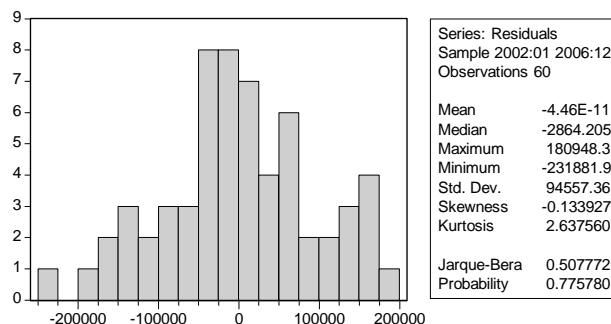
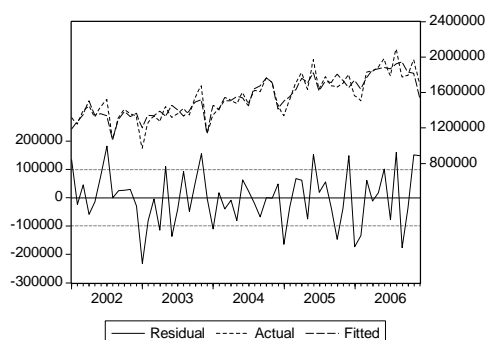
Date: 06/04/08 Time: 09:31

Sample: 2002:01 2006:12

Included observations: 60

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-238895.2	131275.8	-1.819796	0.0743
FIPESAUDE	3021.243	423.8261	7.128496	0.0000
SALARIOS	1.413943	0.224005	6.312094	0.0000
SINISTRALIDADE	5.910868	2.429865	2.432591	0.0183
AT1	239667.4	100117.6	2.393859	0.0202
AT2	-259536.2	101806.7	-2.549303	0.0137
R-squared	0.840063	Mean dependent var	1544770.	
Adjusted R-squared	0.825254	S.D. dependent var	236440.4	
S.E. of regression	98838.17	Akaike info criterion	25.93499	
Sum squared resid	5.28E+11	Schwarz criterion	26.14443	
Log likelihood	-772.0498	F-statistic	56.72676	
Durbin-Watson stat	2.073667	Prob(F-statistic)	0.000000	

Análise de resíduos do modelo dinâmico 4: plot dos resíduos e histograma com teste de normalidade



Terceiro passo: tratamento para atípico 1, atípico 2 e atípico 3
Modelo dinâmico 5:

Dependent Variable: CUSTOTOTAL
Method: Least Squares
Date: 06/04/08 Time: 09:32
Sample: 2002:01 2006:12
Included observations: 60

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-275627.5	129533.8	-2.127842	0.0380
FIPESAUDE	3082.561	414.8553	7.430449	0.0000
SALARIOS	1.501558	0.223315	6.723947	0.0000
SINISTRALIDADE	5.669461	2.374716	2.387427	0.0206
AT1	236789.1	97719.86	2.423142	0.0188
AT2	-254936.3	99385.68	-2.565121	0.0132
AT3	-195043.5	101456.0	-1.922444	0.0599
R-squared	0.850489	Mean dependent var	1544770.	
Adjusted R-squared	0.833563	S.D. dependent var	236440.4	
S.E. of regression	96459.76	Akaike info criterion	25.90092	
Sum squared resid	4.93E+11	Schwarz criterion	26.14526	
Log likelihood	-770.0276	F-statistic	50.24819	
Durbin-Watson stat	1.971378	Prob(F-statistic)	0.000000	

Teste de heterocedasticidade dos resíduos do modelo dinâmico 5
H0: resíduos não são heterocedásticos

White Heteroskedasticity Test:

F-statistic	1.354861	Probability	0.233864
Obs*R-squared	11.76364	Probability	0.226964

Test Equation:

Dependent Variable: RESID^2
Method: Least Squares
Date: 06/04/08 Time: 09:32
Sample: 2002:01 2006:12
Included observations: 60

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	3.58E+10	1.49E+11	0.239254	0.8119
FIPESAUDE	-9.58E+08	8.47E+08	-1.131259	0.2633
FIPESAUDE^2	1700675.	1401407.	1.213548	0.2306
SALARIOS	-239832.3	170051.3	-1.410352	0.1646
SALARIOS^2	0.260231	0.224905	1.157070	0.2527
SINISTRALIDADE	5934956.	2934231.	2.022661	0.0485
SINISTRALIDADE^2	-55.27401	26.97073	-2.049407	0.0457
AT1	-7.75E+09	1.08E+10	-0.721011	0.4743
AT2	-1.00E+10	1.08E+10	-0.926603	0.3586
AT3	-1.68E+10	1.21E+10	-1.387829	0.1713
R-squared	0.196061	Mean dependent var	8.22E+09	
Adjusted R-squared	0.051352	S.D. dependent var	1.07E+10	
S.E. of regression	1.04E+10	Akaike info criterion	49.12768	
Sum squared resid	5.45E+21	Schwarz criterion	49.47674	
Log likelihood	-1463.830	F-statistic	1.354861	
Durbin-Watson stat	2.112982	Prob(F-statistic)	0.233864	

ANEXO 7 – Saídas computacionais do modelo dinâmico 6 (com tratamento de dados atípicos e utilização dos resíduos no modelo MA(7))

Dependent Variable: CUSTOTOTAL
Method: Least Squares
Date: 06/04/08 Time: 09:36
Sample: 2002:01 2006:12
Included observations: 60
Convergence achieved after 14 iterations
Backcast: 2001:06 2001:12

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-355648.8	67865.87	-5.240467	0.0000
FIPESAÚDE	3781.136	273.5811	13.82090	0.0000
SALÁRIOS	1.326211	0.122399	10.83515	0.0000
SINISTRALIDADE	4.221093	1.732084	2.437003	0.0183
AT1	209879.6	54170.04	3.874459	0.0003
AT2	-187586.7	81742.15	-2.294859	0.0258
AT3	-340530.3	60981.09	-5.584195	0.0000
MA(7)	-0.932533	0.021839	-42.70130	0.0000
R-squared	0.913406	Mean dependent var	1544770.	
Adjusted R-squared	0.901749	S.D. dependent var	236440.4	
S.E. of regression	74112.15	Akaike info criterion	25.38811	
Sum squared resid	2.86E+11	Schwarz criterion	25.66736	
Log likelihood	-753.6434	F-statistic	78.35783	
Durbin-Watson stat	2.009940	Prob(F-statistic)	0.000000	
Inverted MA Roots	.99	.62 -.77i	.62+.77i	-.22 -.97i
	-.22+.97i	-.89+.43i	-.89 -.43i	

Cálculo do VIF para modelo dinâmico 6

Regression Analysis: CUSTO-SAUDE versus SINISTRALIDADE; SALÁRIOS; ...

The regression equation is

$$\text{CUSTO-SAUDE} = -369185 + 7,24 \text{ SINISTRALIDADE} + 1,46 \text{ SALÁRIOS} + 0,319 \text{ FIPESAÚDE} \\ + 204078 \text{ AT1} - 273483 \text{ AT2} - 217997 \text{ AT3} + 1,62 \text{ AVER(7)}$$

54 cases used, 6 cases contain missing values

Predictor	Coef	SE Coef	T	P	VIF
Constant	-369187	123333	-2,99	0,004	
IPC FIPESAÚDE	3185,6	469,0	6,79	0,000	1,9
SALÁRIOS	1,4650	0,2099	6,98	0,000	1,6
SINISTRALIDADE	7,240	2,445	2,96	0,005	1,6
AT1	204077	88333	2,31	0,025	1,0
AT2	-273482	91948	-2,97	0,005	1,1
AT3	-217997	91534	-2,38	0,021	1,1
AVER(7)	1,6169	0,4242	3,81	0,000	1,1

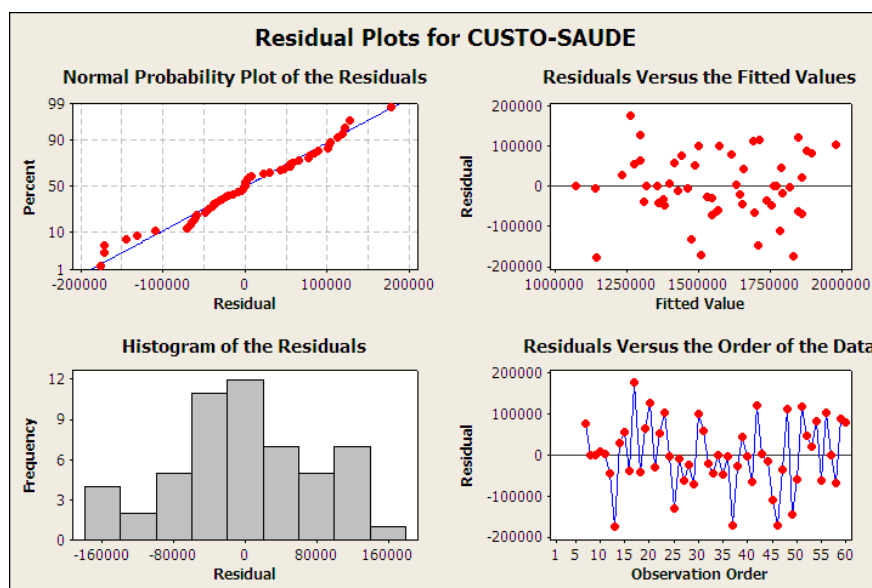
S = 86679,9 R-Sq = 88,6% R-Sq(adj) = 86,9%

Analysis of Variance

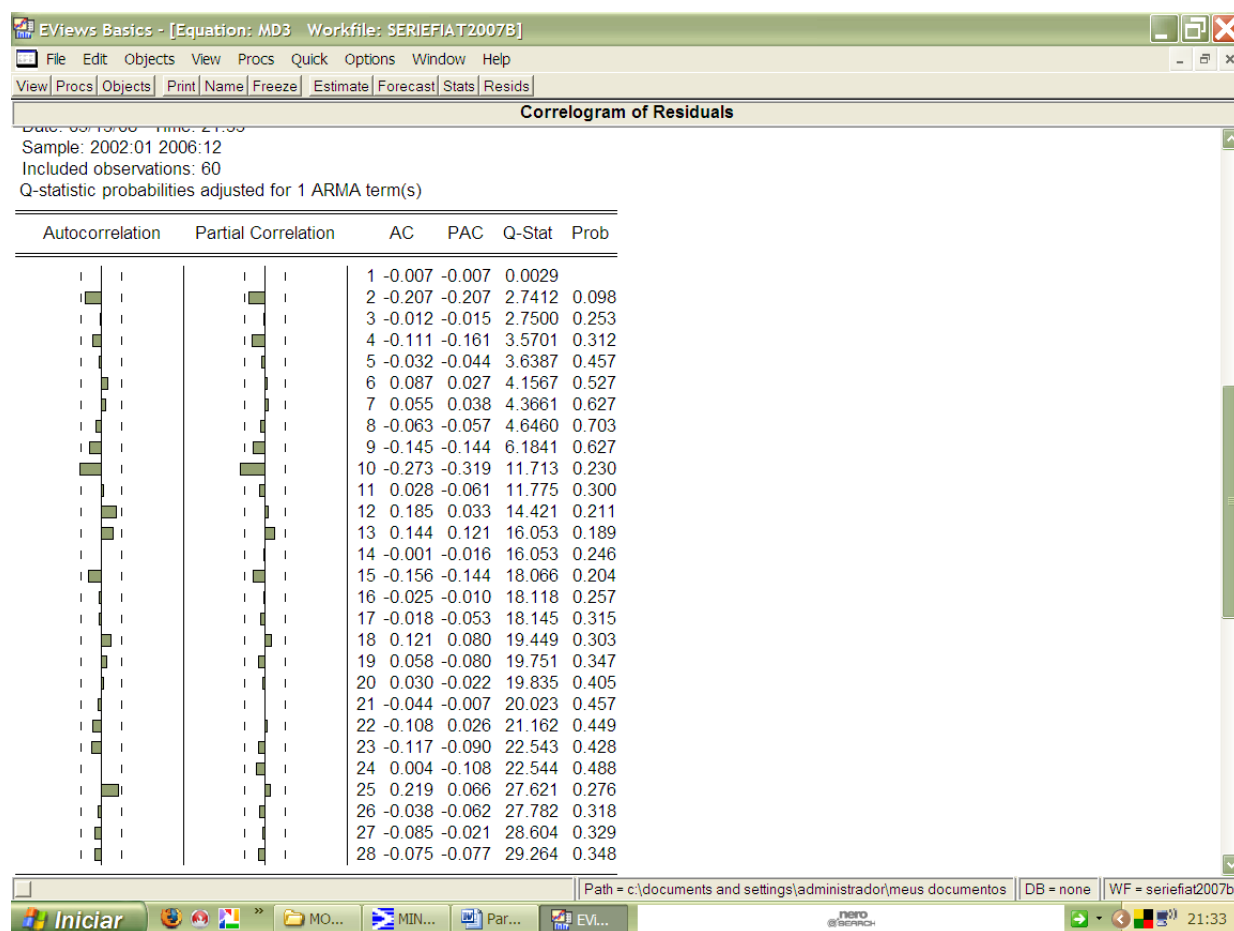
Source	DF	SS	MS	F	P
Regression	7	2,69802E+12	3,85431E+11	51,30	0,000
Residual Error	46	3,45616E+11	7513400489		
Total	53	3,04363E+12			
Source	DF	Seq SS			
SINISTRALIDADE	1	9,15837E+11			
SALÁRIOS	1	1,04784E+12			
FIPESAÚDE	1	4,75200E+11			
AT1	1	57719758269			
AT2	1	53122955847			
AT3	1	39142083924			
AVER(7)	1	1,09153E+11			

Durbin-Watson statistic = 1,96413

Análise de resíduos do modelo dinâmico 6:



FAC de resíduos do modelo dinâmico 6



Teste de heterocedasticidade dos resíduos do modelo dinâmico 6
H0: resíduos não são heterocedásticos

White Heteroskedasticity Test:

F-statistic	0.572924	Probability	0.812702
Obs*R-squared	5.609135	Probability	0.778310

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

Date: 06/04/08 Time: 09:40

Sample: 2002:01 2006:12

Included observations: 60

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8.13E+10	8.90E+10	0.912848	0.3657
FIPESAUDE	-5.37E+08	5.05E+08	-1.063183	0.2928
FIPESAUDE^2	837937.2	835110.2	1.003385	0.3205
SALARIOS	-37143.86	101335.0	-0.366545	0.7155
SALARIOS^2	0.078492	0.134023	0.585664	0.5607
SINISTRALIDADE	436770.1	1748533.	0.249792	0.8038
SINISTRALIDADE^2	-4.476295	16.07208	-0.278514	0.7818
AT1	-2.60E+09	6.41E+09	-0.405062	0.6872
AT2	-5.85E+09	6.45E+09	-0.906509	0.3690
AT3	-8.86E+09	7.22E+09	-1.226631	0.2257
R-squared	0.093486	Mean dependent var	4.76E+09	
Adjusted R-squared	-0.069687	S.D. dependent var	6.02E+09	
S.E. of regression	6.22E+09	Akaike info criterion	48.09234	
Sum squared resid	1.94E+21	Schwarz criterion	48.44140	
Log likelihood	-1432.770	F-statistic	0.572924	
Durbin-Watson stat	2.388818	Prob(F-statistic)	0.812702	