

AJUSTE SAZONAL DE SÉRIES TEMPORAIS:
O MÉTODO X-11 E SUA APLICAÇÃO ÀS
SÉRIES BRASILEIRAS

Irene Mauricio Cazorla



UNICAMP

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E CIÊNCIA DA COMPUTAÇÃO

C318a

7653/BC

CAMPINAS - SÃO PAULO
BRASIL

ALUOTE SAZONAL DE SÉRIES TEMPORAIS: O MÉTODO

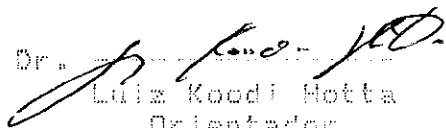
X-11 ARIMA E SUA APLICAÇÃO AS SÉRIES

BRASILEIRAS.

Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pelo Sr. Irene Mauricio Gazorla e aprovada pela Comissão Julgadora.

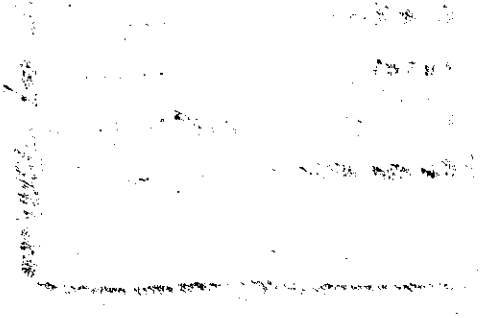
Campinas, 09 de dezembro de 1986

Prof. Dr.


Luiz Koodi Hotta
Orientador

Dissertação apresentada ao Instituto de Matemática, Estatística e Ciência da Computação da Universidade Estadual de Campinas, como requisito parcial para a obtenção do título de Mestre em Estatística.

UNICAMP
BIBLIOTECA CENTRAL



"A mis padres"

Viente

cinco

1911

AGRADECIMENTOS

Tantas pessoas contribuíram para que este trabalho pudesse ser realizado, que me seria impossível citá-las nominalmente. Mas quero registrar aqui meus agradecimentos sinceros a todos aqueles que, de alguma forma me auxiliaram ao longo do caminho percorrido.

Sou especialmente grata ao Professor Luiz Koodi Hotta pela orientação segura, aliada à amizade e paciência que me dispensou ao longo destes quase três anos e à Professora Nazira Gait, da USP, pela constante ajuda, sem a qual não seria possível a execução do presente trabalho.

Todavia, se foi possível meu contato com as pessoas acima mencionadas, devo-o à Universidad Nacional de San Cristóbal de Huamanga, do Peru, que me liberou dos trabalhos didáticos para que eu pudesse me dedicar ao programa de Pós-Graduação, aqui no Brasil, e à Universidade Estadual de Campinas que me recebeu e me propiciou todas as condições que possibilitaram a realização deste trabalho.

E por último, mas não menos importante, sou profundamente grata às amigas que se fizeram minhas irmãs no Brasil: Gloria, Inês, Odete e Olga, e aos amigos que, no Peru, em nome da amizade, garantiram minha estadia no Brasil. Destre eles, quero destacar Ronald, Fermín, Juan, Walter e, em especial, Fernando Arce.

RESUMO

O trabalho analisa o desempenho do método X-11 desenvolvido pelo Bureau de Censo dos E.U.A. e do X-11 ARIMA desenvolvido no Statistics Canada, frente a séries que apresentam fortes mudanças estruturais na tendência e sazonalidade, características estas encontradas nas séries brasileiras. É analisada a série da taxa de mortalidade infantil do Estado de São Paulo, no período de 1933 a 1985. A análise mostra o perigo do uso indiscriminado da opção automática, devido as deficiências mostradas na qualidade do ajuste e a demora para detectar mudanças estruturais. Isto no entanto, é atenuado pela utilização adequada das opções existentes no próprio programa. O exemplo mostra a importância do ajuste sazonal, e das interpretações que se podem conseguir através da análise dos componentes.

ABSTRACT

This work studies the performance of the U.S. Bureau of Census X-11, and the Statistics Canada X-11 ARIMA Methods, in the analysis of time series presenting strong structural changes in both trend and sazonality, common in some brazilian series. It analyses the infant mortality rate monthly series in the State of São Paulo, Brazil, from January 1933 to December 1985. The analysis reveals the dangers of the indiscriminate use of the automatic option due to the deficiencies shown in the fitting quality and the long time needed to detect structural changes. However this problem is attenuated by the appropriate use of the existing options in the program. The example illustrates the importance of the seasonal adjustment and of the interpretations which may be gotten through the analysis of the components.

INDICE

Capítulo I : Introdução	1
Capítulo II : O ajuste sazonal de séries temporais	4
2.1 Introdução	4
2.2 Razões para o ajuste sazonal	5
2.3 Histórico	5
2.4 Métodos de ajuste sazonal	7
2.4.1 Procedimentos de médias móveis ou de suavização linear ..	7
2.4.2 Os métodos de regressão e o enfoque bayesiano ..	8
2.4.3 Ajuste sazonal "Model-based"	9
2.4.4 Discussão dos procedimentos de ajuste sazonal	10
2.5 Avaliação do ajuste sazonal	11
2.5.1 Critérios para avaliar o ajuste sazonal	11
2.5.2 Revisões	11
Capítulo III: O método de ajuste sazonal X-11 e o X-11 ARIMA	13
3.1 Introdução	13
3.2 O método de ajuste sazonal X-11	13
3.2.1 Histórico	13
3.2.2 Definição e notação dos componentes	14
3.3 Fundamentos do método de decomposição sazonal X-11 ARIMA	15
3.3.1 Introdução	15
3.3.2 Os modelos ARIMA, sua extrapolação e sua seleção	15
3.3.2.1 A opção ARIMA automática	16
3.3.2.2 A identificação do modelo ARIMA pelo usuário	18
3.3.3 Principais passos para o ajuste sazonal X-11 ARIMA	19
3.3.4 Propriedades básicas dos filtros do X-11 e do X-11-ARIMA ..	21
3.3.5 Rotina variável para a tendência-ciclo	25
3.3.6 Rotina variável para o fator sazonal	26
3.3.7 Tratamento dos valores extremos	26
3.3.8 Vantagens do X-11 ARIMA sobre o X-11	27
3.3.9 Testes para sazonalidade	28
3.3.9.1 Testes para sazonalidade estável	28
3.3.9.2 O teste Fm para sazonalidade móvel	29
3.3.9.3 O teste combinado para sazonalidade	
identificável	30
3.3.9.4 Teste para sazonalidade residual	31
3.3.10 Testes para aleatoriedade dos resíduos	31
3.3.10.1 O teste da duração média de sequência (ADR) ..	31
3.3.10.2 A função de autocorrelação dos resíduos (FAC) ..	31
3.3.10.3 O teste para periodicidade nos resíduos	32
3.3.11 As estatísticas M e G de controle de qualidade	32
3.4 Descrição do pacote do X-11 ARIMA	40

Capítulo IV : Metodologia	42
4.1 Introdução	42
4.2 A metodologia	42
4.3 A opção automática	42
4.4 Critérios para avaliar a qualidade do ajuste	42
Capítulo V : Aplicação	49
5.1 Introdução	49
5.2 Definição da série	49
5.3 Descrição da série	49
5.3.1 Descrição da taxa anual	49
5.3.2 Descrição da taxa mensal	51
5.4 Escolha dos pontos de estudo	55
5.5 Estudo do desempenho do método no período de 1933 a 1948	55
5.5.1 Análise dos resultados do ajuste da opção automática	56
5.5.1.1 Análise do ajuste da opção ARIMA	56
5.5.1.2 Análise da qualidade do ajuste	59
5.5.1.3 Discussão dos resultados	60
5.5.2 Desempenho das opções alternativas para a série no período de 1933 a 1948: TX3348, X-11	63
5.5.3 Discussão e escolha da melhor opção	66
5.5.4 Estudo das revisões da opção automática versus a melhor opção	67
5.6 Estudo do desempenho do método no período de 1950 a 1985	77
5.6.1 Análise dos resultados do ajuste da opção automática	77
5.6.1.1 Análise do ajuste da opção ARIMA	77
5.6.1.2 Análise da qualidade do ajuste	77
5.6.2 Desempenho das opções alternativas	78
5.6.2.1 Discussão e escolha da melhor opção para a série no período de 1950 a 1975: TX5075, X-11 ARIMA ...	79
5.6.2.2 Discussão e escolha da melhor opção para a série no período de 1950 a 1965: TX5065, X-11	81
5.6.2.3 Discussão e escolha da melhor opção para a série no período de 1960 a 1975: TX6075, X-11 ARIMA ...	83
5.6.2.4 Discussão e escolha da melhor opção para a série no período de 1970 a 1985: TX7085, X-11 ARIMA ...	83
5.6.2.5 Discussão e escolha da melhor opção para a série no período de 1960 a 1985: TX6085, X-11 ARIMA ...	85
5.6.3 Estudo das revisões no período de 1968 a 1969	87
5.7 Análise dos componentes	90
5.7.1 Análise dos componentes no período de 1933 a 1948	90
5.7.2 Análise dos componentes no período de 1950 a 1985	91
5.8 Conclusões	92
Apêndices	95
Lista de gráficos	114
Lista de tabelas	115
Referências bibliográficas	118

Capítulo I : INTRODUÇÃO.

O estudo da evolução sócio-econômica de um país geralmente utiliza indicadores ou índices para medir o nível da atividade econômica e das condições sociais. Exemplos destes índices são: a taxa de desemprego, índice de custo de vida, taxa de mortalidade, etc. A maioria destes indicadores são fortemente influenciados pela sazonalidade (condições físico-climáticas, as quatro estações, costumes da população, festas religiosas) que interferem no comportamento das séries ao longo do ano. Por exemplo, a taxa de desemprego normalmente aumenta no final do ano acadêmico, quando estudantes procuram empregos, ou no período de entressafra em regiões agrícolas; as atividades relacionadas com turismo aumentam no período de férias, em épocas festivas ou em feriados longos, etc.

Dada esta influência sazonal, a simples comparação dos índices mês a mês, e mesmo com o mesmo mês em anos anteriores pode levar o analista a interpretações incorretas. Tome-se o caso da taxa de desemprego como exemplo. Não se pode concluir apressadamente que está havendo uma diminuição do desemprego se houver aumento da oferta de emprego num mês, se este mês for caracterizado por baixas taxas de desemprego. A caracterização do que é tendência e o que é simplesmente efeito sazonal pode ajudar bastante neste caso.

Desta forma, se o objetivo é a análise da tendência do nível da atividade econômica, dados não ajustados (sem remover o efeito da sazonalidade) comprometem seriamente a validade das conclusões e das decisões tomadas, fazendo com que os usuários destas informações e público em geral percam a credibilidade nos responsáveis pelas análises. Os tomadores de decisões, que assumem a responsabilidade de controlar o nível da economia, necessitam contar com dados sazonalmente ajustados de forma precisa, pois são importantes para a tomada de decisões corretas (Dagum, 1978).

Há pouco tempo o Brasil passou por uma severa crise econômica. Muitos analistas (principalmente da área de economia), jornalistas e o próprio governo, tentaram mostrar através de indicadores econômicos, que o país estava retomando o crescimento econômico, saindo da época de recessão. Porém, muitos dos indicadores por eles apresentados estavam baseados em dados não ajustados, o que poderia levar a erros, conforme foi chamada a atenção em Hotta(1984).

O conhecimento do componente sazonal também é importante, como é o caso de controle de estoques, ou na orientação do governo, na formulação de política de abastecimento. Pode também auxiliar no entendimento de sistemas, como será visto nesta tese, ao relacionar a magnitude da sazonalidade às condições de saúde no Estado de São Paulo.

O uso de dados sazonalmente ajustados leva à questão dos métodos a serem usados para conseguir este objetivo. O método mais conhecido neste contexto é o X-11, devido aos bons resultados obtidos em outros países. Atualmente, no Brasil, este método não é usado para o ajuste de séries em grande escala, embora se tenha grupos pesquisando nesta área. Portanto, surge a necessidade de avaliar o desempenho do

método em séries brasileiras.

Uma das vantagens do método X-11, citada frequentemente pelos empregadores do método (Pierce, 1980; Hillmer e Tiao, 1982; e outros) é o seu aspecto automático, ou quase automático. Esta característica é considerada fundamental ao se trabalhar com um número bastante grande de séries. E, como todo método automático, ele está "calibrado" para os usuários em potencial, ou para a tarefa a que ele se dispõe. O método X-11 e o método X-11 ARIMA, que é uma versão mais ampla do X-11, por serem desenvolvidos nos E.U.A e no Canada, respectivamente, certamente consideraram as séries americanas e canadenses. Desta forma, os modelos e suas opções padrões não devem ser adotados automaticamente para o ajuste das séries brasileiras, sem uma análise prévia.

Dagum e Morry (1983), ao analisar a série de emprego e desemprego de adultos do Canada na época de recessão em relação à época normal (mudança brusca na composição da sazonalidade e no nível) mostraram-se preocupados com o desempenho do X-11 e X-11 ARIMA devido às fortes mudanças estruturais de tendência e sazonalidade. Estas fortes mudanças estruturais de tendência e sazonalidade pode ser considerada uma característica típica das séries brasileiras, devido às grandes mudanças que tem ocorrido no Brasil nas últimas décadas. Este fato motivou a presente tese, que é a de verificar o comportamento destes programas numa série brasileira. Como uma das grandes vantagens do método é o fato dele ser quase automático, podendo ser aplicado rapidamente para um grande número de séries, a preocupação maior do trabalho foi a de estudar o desempenho da opção automática do X-11, e ao mesmo tempo estudar como as opções disponíveis poderiam ser utilizadas para melhorar o ajuste.

Foi escolhida a série da taxa de mortalidade infantil do Estado de São Paulo como a série representativa das séries brasileiras por vários motivos: i) a taxa de mortalidade infantil é um dos indicadores mais completos para avaliar a evolução da qualidade de vida de uma população, e, portanto, seu nível de desenvolvimento, ii) é altamente sensível às mudanças sócio-econômicas e, como consequência, esta série possui na sua estrutura sazonal e sua tendência, diversas características que fazem dela uma série "representativa" das séries brasileiras e iii) os dados são mais confiáveis e completos.

A taxa de mortalidade infantil do Estado de São Paulo, no período de 1933 a 1985, apresenta na década de 30 e parte da década de 40, um padrão sazonal fortemente estável, com altas taxas nos meses de verão e baixas nos meses de inverno, tendo o nível quase constante e muito elevado, consequência da crise desta época (Ortiz, 1982a). Em parte da década de 40 e na década de 50, a influência sazonal e a tendência decrescem rapidamente e permanecem quase constantes na década de 60. Isto é devido à grande expansão econômica do Estado e principalmente ao crescimento do salário mínimo real (Monteiro, 1982). No fim da década de 60, observa-se um aumento no nível da taxa. Este aumento parece ser explicado pelo fato do país ter entrado num processo de concentração da renda com diminuição do salário mínimo real, além do aumento do fluxo migratório (Ortiz, 1982a). A partir de 1975, esta taxa novamente começa a decrescer. Segundo Monteiro, isto se deve, principalmente, à ampliação da rede de água e esgoto. Neste último per-

ríodo não há um padrão predominante e em alguns anos as taxas dos meses de frio possuem maior importância relativa. Este fenômeno parece ser explicado pela mudança na estrutura de causalidade da mortalidade infantil (Ortiz, 1982a).

Para atingir o objetivo, a série foi dividida em períodos mais curtos. Os cortes são feitos nos períodos onde ocorrem mudanças bruscas, a fim de analisar como o método acompanha estas mudanças. A estratégia consiste em aplicar, em princípio, a opção automática (aquela onde o programa possui a mais ampla liberdade de escolha entre as opções que ele proporciona) na série, nos períodos definidos anteriormente, fazendo-se a análise da qualidade do ajuste. Em seguida são aplicadas algumas opções alternativas e escolhida a que proporciona o "melhor ajuste". Para finalizar é realizado um estudo comparativo das revisões entre a opção do "melhor ajuste" e a opção automática. Desta forma, é possível verificar a eficiência da opção automática, ao mesmo tempo em que é estudada a possibilidade da utilização das opções disponíveis no próprio programa X-11 ARIMA para sanar as falhas da opção automática do programa.

Os critérios para avaliar a qualidade do ajuste estão baseados nos trabalhos de Shiskin e Plewes (1978), de Dagum e Morry (1983) e naqueles que o próprio X-11 ARIMA proporciona (Dagum, 1980). Os critérios são: sazonalidade, adequabilidade do modelo, aleatoriedade dos resíduos, participação do componente irregular, suavidade da série sazonalmente ajustada e as revisões.

Na maioria dos casos, a opção automática não fornece um ajuste adequado, as deficiências estão estreitamente relacionados com as mudanças de estrutura da sazonalidade e da tendência. O uso apropriado das opções alternativas, principalmente dos filtros, aumenta a qualidade do ajuste, atingindo, em alguns casos, bons ajustes. A análise dos componentes mostra resultados compatíveis com as análises já existentes sobre a série, indicando que o método consegue seu objetivo. A análise da sazonalidade indica como este fator auxilia a compreensão do processo gerador de dados.

O trabalho mostra que o método X-11 ARIMA, através da utilização adequada de suas opções pode apresentar um bom desempenho frente a séries que apresentam fortes mudanças estruturais de sazonalidade e tendência. Estes resultados indicam a viabilidade de aplicar o método às séries brasileiras, embora, certamente, não se possa afirmar que se consiga um bom desempenho em relação a qualquer série. Cleveland e Tiao (1976), inclusive mostram que para certas séries, o método X-11 não apresenta um bom desempenho.

Na organização geral da tese, o Capítulo II apresenta uma revisão geral do conceito de ajuste sazonal, os métodos disponíveis, e um breve comentário sobre estes métodos. O Capítulo III apresenta o método X-11 e o X-11 ARIMA, enfatizando o segundo, pois este abrange o primeiro. O capítulo IV apresenta a metodologia utilizada no trabalho e o capítulo V apresenta a aplicação da metodologia e as conclusões do estudo.

Capítulo II. O AJUSTE SAZONAL DE SÉRIES TEMPORAIS.

2.1 Introdução.

O ajuste sazonal de uma série temporal consiste em decompor a série $Z(t)$, basicamente em dois componentes não observáveis: sazonal $S(t)$ e não sazonal $N(t)$. Geralmente o componente não sazonal é decomposto em outros componentes, tais como: tendência, ciclo, variação do calendário, componente irregular.

A tendência reflete os movimentos a longo prazo. O ciclo é composto por oscilações quase periódicas, caracterizadas por épocas de expansão e contração. A sazonalidade representa o efeito composto de eventos climáticos e institucionais que se repetem com certa regularidade a cada ano. Supõe-se que estes três tipos de flutuações seguem um padrão sistemático ao longo do tempo e que podem ser medidos diretamente na série. Por outro lado, o componente irregular constitui-se de movimentos imprevisíveis, que, em geral têm uma aparência aleatória estável. O componente irregular é distinto de outros tipos de irregularidade; os outliers, os quais têm causas bem identificadas, tais como, enchentes, greves, variações climáticas atípicas, etc. (Dagum, 1978a).

As três características importantes da sazonalidade, são:

- 1) o fenômeno se repete a cada ano com certa regularidade, embora possa se modificar ao longo dos anos;
- 2) o fenômeno pode ser medido e separado das outras forças que influenciam o movimento da série; e,
- 3) o fenômeno é principalmente causado por forças não econômicas, exógenas ao sistema econômico, que não podem ser controladas ou modificadas pelos tomadores de decisões num intervalo a curto prazo.

Para Dagum, esta última característica constitui a principal razão para se remover a sazonalidade, já que as variações da série sazonalmente ajustada são devidas à tendência-ciclo e às flutuações aleatórias, ocasionadas principalmente por desequilíbrios introduzidos pelos ciclos econômicos, que podem ser corrigidos num período relativamente curto.

Embora haja uma crescente tendência em utilizar dados ajustados sazonalmente, deve-se evitar a utilização indiscriminada destas técnicas pois elas podem levar a resultados absurdos, principalmente quando não se têm claros, os objetivos e as razões para se fazer o ajuste sazonal, a definição dos componentes, os métodos a serem usados e suas suposições implícitas, etc. (Bell e Hillmer, 1984). A série sazonalmente ajustada não deve substituir a série original, mas complementá-la.

O ajuste sazonal tem sido oficialmente aplicado pelas agências oficiais da maioria dos países desenvolvidos. Entre os métodos mais usados, o método X-11 (Shiskin, Musgrave e Young, 1967) é o mais popular, graças aos bons resultados obtidos e à sua facilidade de aplicação (Pierce, 1980a; Dagum, 1978a; Kenny e Durbin, 1982; etc.). A

maioria dos autores consideram o X-11 como o ajuste padrão, servindo como referência para avaliar novos métodos.

Neste capítulo é apresentada uma breve revisão sobre o ajuste sazonal, histórico e métodos utilizados.

2.2 Razões para o ajuste sazonal.

Como foi visto anteriormente, para Dagum, a principal razão para se retirar a sazonalidade é que esta não pode ser controlada pelo tomadores de decisões. Segundo Bell e Hillmer (1984) as razões para o ajuste sazonal tem sido geralmente vagas, mas estas giram em torno de três temas:

- 1) ajudar na previsão a curto prazo;
- 2) ajudar a relacionar uma série temporal com outras séries, com eventos externos ou com variáveis de política; e,
- 3) conseguir comparabilidade nos valores da série mês a mês.

Para Bell e Hillmer, uma possível justificativa do ajuste sazonal é que os dados ajustados sazonalmente são mais fáceis de serem interpretados por usuários leigos em estatística, sempre e quando a perda de informação não for significativa. Esta questão é importante ser levada em consideração, pois, para se estabelecer se a justificativa acima é válida, é necessário conhecer a quantidade de informação perdida no ajuste sazonal, mesmo quando o método for adequado. Esta situação será pior quando o ajuste sazonal for feito sob suposições incorretas.

Há alguns trabalhos que estudam a consequência de se usar dados ajustados, concentrando-se na maneira pela qual o ajuste sazonal afeta a precisão nas previsões e no relacionamento de uma série com outras (Wallis, 1974). Destes trabalhos se conclui que é difícil dizer o efeito que tem o uso de dados ajustados na precisão das previsões. Além disso, o uso de dados ajustados tem uma severa desvantagem que é a impossibilidade do cálculo da variância do erro de previsão, impedindo o uso de intervalos de confiança para as previsões, algo que pode ser feito para dados não ajustados. Deve-se considerar que o uso de dados ajustados implica em correr algum risco de vício e ineficiência na estimação das relações entre séries. Às vezes, a simplicidade de dados ajustados é ilusória.

2.3 Histórico

O ajuste sazonal se inicia na década de 20 como uma ferramenta para análise de séries econômicas sem ajustar modelos estatísticos. Foi desenvolvido dentro da tradição de se considerar componentes não observáveis de uma série temporal. Os primeiros a analisarem esta técnica, perceberam que as séries temporais continham tendência não estacionária e sazonalidade móvel, e que este comportamento não poderia ser descrito por uma função do tempo matematicamente explícita. Eles desenvolveram métodos empíricos de ajuste sazonal, usando ferramentas como as médias móveis para lidar com estes problemas. Este enfoque foi e é fortemente criticado pela sua natureza empírica, mas naquela época não havia modelos disponíveis para séries temporais e não se possuía o

auxílio do computador.

Muitos conceitos considerados importantes foram, não entanto, definidos desde àquela época:

- 1) a idéia de sazonalidade móvel;
- 2) a necessidade de avaliar a tendência e ciclo ao se estimar o componente sazonal;
- 3) a impossibilidade de descobrir tendências e ciclos através de fórmulas; e,
- 4) a necessidade de se lidar com valores extremos.

Os primeiros a utilizarem estas técnicas perceberam que era necessário ajustar primeiro os dados para efeitos da tendência, ou estimar a tendência ao mesmo tempo em que se estimava o componente sazonal, o que leva à iteração entre a estimação da tendência e a da sazonalidade, como é feito no X-11. Enquanto alguns autores preferiram ajustar a tendência via transformações ou diferenças outros preferiram estimar primeiro a sazonalidade e removê-la através da razão entre a série original e a estimativa da sazonalidade. Paralelamente, desenvolveram-se métodos de regressão, cujo princípio básico é a suposição de que os componentes são funções determinísticas do tempo. Tais métodos tiveram pouco sucesso, uma vez que este enfoque requer formas explícitas para os componentes.

Em 1950, Julius Shiskin começou a fazer ajustes sazonais já com o auxílio do computador do Bureau do Censo dos E.U.A, em um grande número de séries com bastante êxito. Ele e sua equipe desenvolveram o método chamado Census Method I e lideraram um programa de pesquisa intenso para melhorá-lo. À medida que conseguiram modificações substanciais, eles introduziram novas versões, que foram desde a versão X-3 até à versão final, o X-11. O método X-11 é baseado em filtros de médias móveis e não explicita o modelo subjacente, nem da série nem dos componentes.

No contexto do ajuste sazonal, o X-11 se transformou no método mais popular, devido, principalmente, a sua facilidade de aplicação e aos bons resultados obtidos em muitos países. Além disto, o X-11 fornece ajustes com qualidade superior à qualidade de outros procedimentos (Burridge e Wallis, 1984), tanto assim, que o X-11 é considerado como o método padrão para a avaliação do desempenho de novos procedimentos (Akaike, 1980).

Contudo, seu caráter empírico (Hillmer e Tiao, 1982), a sensibilidade dos fatores sazonais frente às revisões (Dagum, 1978a, 1982), e a autocorrelação dos resíduos (Cleveland e Tiao, 1976), fazem com que o método seja frequentemente questionado. Outra crítica feita ao X-11 é a dificuldade de se verificar como os resultados são alterados frente a uma mudança em alguma das opções (Akaike, 1980). Muitas discussões tem sido feitas a respeito, e as tentativas para superar estas deficiências continuam.

Em 1970, Box e Jenkins apresentaram os modelos ARIMA e o software apropriado para estes modelos. Isto contribuiu para um rápido desenvolvimento de técnicas alternativas ao X-11 a fim de superar as

deficiências mostradas por este.

Em 1975, Dagum apresentou o método X-11 ARIMA que ajusta a série segundo um modelo ARIMA, estende-a com valores extrapolados pelo modelo ajustado e aplica o X-11, permitindo o uso de filtros próximos aos simétricos e com isto reduzindo as revisões das estimativas do fator sazonal em relação ao X-11.

Outra técnica alternativa que vem recebendo grande atenção nos últimos anos é o enfoque "model-based", cuja principal vantagem reside na especificação dos modelos para a série observada e seus componentes, geralmente usando modelos ARIMA e impondo restrições aos modelos. Para a estimação dos componentes, usa-se a teoria da extração do sinal (Cleveland e Tiao, 1976; Box, Hillmer e Tiao, 1978; Burman, 1980; Hillmer e Tiao, 1982; Burrridge e Wallis, 1984).

Em fins da década de 70 e início da década de 80 aparece o enfoque Bayesiano (catalogado dentro dos procedimentos de regressão) desenvolvido por Harrison e Stevens (1976), por Akaike e Ishiguro (1983), entre outros. Segundo Akaike (1980), este enfoque pode ser visto como um procedimento de modelagem de séries temporais, o qual permite o uso efetivo de informação a priori de cada série particular, e a seleção do modelo básico é feita minimizando um critério objetivamente definido (Bulach, 1983).

2.4 Métodos de ajuste sazonal.

Os métodos de ajuste sazonal são determinados pelas suposições feitas, implicitamente ou não, acerca dos componentes (Bell e Hillmer, 1984). Portanto, é essencial definir rigorosamente os componentes a serem estimados. Segundo Akaike (1980), os procedimentos podem ser classificados como:

- 1) procedimentos de médias móveis ou de suavização linear, para a extração dos componentes de tendência e sazonalidade, que são assumidas como sendo funções suaves do tempo, isto é, estocásticas.
- 2) procedimentos de regressão, que assumem que os componentes são funções determinísticas do tempo. Aqui se encaixa o enfoque Bayesiano;
- 3) modelagem de séries temporais, onde a série e seus componentes são modelados, principalmente, usando modelos ARIMA. Este enfoque é conhecido como model-based, incluindo nesta classificação os estruturais (Harvey e Todd, 1983).

2.4.1 Procedimentos de médias móveis ou de suavização linear.

Para decompor a série, estes procedimentos utilizam uma sequência de médias móveis, chamadas também de filtros. O X-11 é o mais conhecido dentro desta classe, sendo também o X-11 ARIMA incluído nesta categoria (Dagum, 1978a). Entre outros métodos tem-se o modelo de alisamento exponencial de Holt-Winters, e o modelo geral de Brown (Mortimer e Toloi, 1981).

Estudos empíricos envolvendo séries temporais de diversos países mostram que as revisões do fator sazonal produzidas pelo X-11 são menores que as outras produzidas por outros procedimentos de médias móveis, Dagum (1983a). O X-11 e o X-11 ARIMA serão amplamente discutidos no capítulo III.

2.4.2 Procedimentos de regressão e o enfoque Bayesiano.

O método de regressão assume que a parte sistemática de uma série temporal pode ser aproximada por simples funções do tempo. Para a tendência, geralmente são usados polinômios de grau baixo, o que implica supor que o fenômeno possui um movimento suave, lento e progressivo ao longo do tempo. Entretanto, para sazonalidade e ciclos, são usadas combinações lineares de funções seno e cosseno de diferentes amplitudes e frequências, representando as oscilações, estritamente periódicas ou não (Dagum, 1978a).

Estas funções são estimadas por métodos de mínimos quadrados. Para obter estimadores eficientes, supõe-se que o componente aleatório seja independente e uma versão apropriada de mínimos quadrados generalizados poderá ser aplicada. O método assume sazonalidade estável ou linear móvel.

O enfoque Bayesiano, ao contrário dos procedimentos de médias móveis, usa a técnica de regressão como uma estrutura simples que permite, facilmente, a incorporação de várias opções a fim de incrementar o número de variáveis independentes. Por outro lado, dado que este enfoque está baseado num modelo estatístico explicitamente definido, um critério útil e objetivo para a seleção do melhor modelo está disponível. Assim, este procedimento satisfaz dois requisitos importantes, a disponibilidade de opções e de um critério de decisão objetivo (Akaike, 1980; Akaike e Ishiguro, 1983).

O modelo básico assume que a série possui uma estrutura aditiva, onde se espera que: i) a tendência $T(t)$, pelo menos localmente, mostre um comportamento suave, ii) o componente sazonal $S(t)$ tenha um padrão anual estável. De tal modo que a parte sistemática $S(t)+T(t)$ não se desvie das observações da série original. Embora a estrutura básica seja simples, o BAYSEA (Bayesian Seasonal Adjustment, pacote desenvolvido por Akaike em 1980) contém muitos parâmetros que devem ser especificados antes de se utilizar o pacote. Este problema é resolvido como um problema estatístico de seleção do modelo.

2.4.3 Ajuste sazonal "Model-based".

A principal crítica feita aos métodos de ajuste sazonal através de médias móveis, principalmente ao X-11, é o enfoque empírico, isto é, à falta de modelos explícitos subjacentes usados no ajuste. Esta insatisfação levou muitos autores a pesquisarem o uso de modelos de séries temporais para fazer ajuste sazonal, chamados de métodos de ajuste sazonal "model-based".

Os métodos de ajuste sazonal "model-based" geralmente usam a decomposição aditiva, $Z(t) = S(t) + N(t)$, ou uma decomposição aditiva para alguma transformação de $Z(t)$, tal como logarítmo de $Z(t)$, e usam modelos estatísticos explícitos (ou densidades espectrais) para $Z(t)$, $S(t)$ e $N(t)$. O modelo para $Z(t)$ pode ser estimado a partir dos dados observados, mas $S(t)$ e $N(t)$ não são observáveis e seus modelos dependem de suposições arbitrárias. Os vários métodos diferem no tipo de modelo ajustado aos valores observados $Z(t)$'s e nas suposições usadas na especificação dos modelos para $S(t)$ e $N(t)$. $S(t)$ e $N(t)$ são estimados diretamente quando $Z(t)$ é ajustado (como nos métodos de regressão) ou depois de ajustar o modelo para $Z(t)$, usando a teoria de extração do sinal (Whittle, 1963).

Os métodos de regressão fornecem o primeiro enfoque "model-based" para o ajuste sazonal. O enfoque básico consiste em especificar as formas funcionais da tendência e do componente sazonal que dependem linearmente de alguns parâmetros. Estes parâmetros são estimados por mínimos quadrados, subtraindo a sazonalidade da série. As especificações mais populares usadas são polinômios no tempo para a tendência e médias sazonais para o componente sazonal (com modificações para manipular a sazonalidade móvel). O erro é geralmente assumido como um ruído branco.

Alguns autores sugerem métodos baseados na estimação espectral. Outros autores sugerem métodos que envolvem ajuste de modelos ARIMA para a série original e para seus componentes. Para estimar os componentes é usada a extração do sinal. Para isto é necessário fazer certas suposições, incluindo a de que a variância do componente irregular deve ser maximizada, o que torna mínima a variância da sazonalidade e da tendência. O método é chamado de decomposição canônica.

Box, Hillmer e Tiao (1978), começaram com um modelo ARIMA $(0,1,1) \times (0,1,1)$. Este enfoque foi mais tarde estendido para um modelo ARIMA mais geral, por Burman (1980) e por Hillmer e Tiao (1982), os quais discutem algumas propriedades do enfoque (v. também Hillmer, Bell e Tiao, 1983; Burrigde e Wallis, 1984; entre outros). Outros enfoques são referidos por Bell e Hillmer (1984).

O X-11 pode ser representado por um conjunto de filtros lineares (Wallis, 1982). Alguns dos trabalhos propõem modelos para estes filtros através dos modelos de componentes não observáveis. Este enfoque foi iniciado por Cleveland e Tiao em 1976. Eles formularam um modelo para os filtros simétricos do X-11; mais tarde, Burrigde e Wallis (1984) ampliaram este enfoque para os filtros assimétricos. Os resultados destes trabalhos mostram que, sob certas condições, a decomposição feita pelo X-11 pode ser feita através dos modelos propostos, e, portanto, nestes casos, este procedimento supera o do X-11.

2.4.4 Discussão dos procedimentos de ajuste sazonal.

Dagum (1978a) discute as suposições básicas dos métodos de regressão e das técnicas de suavização linear, concluindo que, as principais críticas aos procedimentos de regressão são: a rigidez para captar os movimentos subjacentes da série, a influência das observações mais distantes nas estimativas dos valores mais recentes, e, a tendência das revisões serem altas e não possuírem um comportamento monótono. Uma limitação dos procedimentos de médias móveis é o caráter empírico destes procedimentos; isto é, estes métodos não estão baseados num modelo paramétrico explícito. Portanto, as propriedades não estão bem definidas e a inferência estatística é limitada. Outra dificuldade são os pesos dos filtros para as primeiras e últimas observações.

Mesmo que o desempenho dos métodos de regressão possa ser melhorado através de maiores especificações para casos particulares, é difícil derivar modelos cujo desempenho forneça ajustes de maior qualidade que os dos outros procedimentos. Além disto, o enfoque de regressão é muito sensível, particularmente, no início e no final da série, para a escolha de variáveis independentes (Akaike, 1980). Um resumo dos procedimentos e problemas de ajuste sazonal pode ser encontrado em Pierce (1980a).

Como foi visto, no contexto do ajuste sazonal, o X-11 é o método que proporciona, na maioria das vezes, uma qualidade superior aos dos outros procedimentos, sendo considerado atualmente como o método padrão para comparação com novos procedimentos de ajuste sazonal. Contudo, o X-11 mostra sérias deficiências.

O X-11 ARIMA supera em parte estas deficiências. Pierce (1980b) mostra, que teóricamente sob certas condições, o X-11 ARIMA fornece o erro quadrático médio mínimo das estimativas do fator sazonal e deste modo minimiza o erro quadrático médio das revisões. Embora na prática isto não seja verdade necessariamente, tem sido demonstrado que este método reduz significativamente as revisões quando comparado com outros procedimentos de médias móveis (Dagum, 1979b).

A maioria dos novos procedimentos procura superar as deficiências mostradas pelo X-11. Assim, o enfoque "model-based" tenta suprir a natureza empírica deste, fornecendo modelos para as estimativas dos componentes. Para alguns casos tem-se obtido bons resultados. As principais limitações deste enfoque são as dificuldades práticas na implementação destes métodos e a especificação dos modelos para os componentes não observáveis. Isto é, a cobertura dos métodos deste enfoque é menor que a do X-11 e, além disso, tais métodos requerem um conhecimento estatístico maior, que os usuários geralmente não possuem.

Conclui-se, portanto, que na atualidade não existe um método alternativo ao X-11 e/ou X-11 ARIMA que o supere no contexto geral do ajuste sazonal.

2.5 Avaliação do ajuste sazonal.

Há dois pontos diferentes, às vezes conflitantes, que devem ser levados em consideração na avaliação da adequabilidade de procedimentos de ajuste sazonal: uma é a técnica e o outro a aceitabilidade pública. Isto é importante, principalmente quando se tratar de séries altamente sensíveis, envolvendo dados polêmicos, como por exemplo, a série do desemprego (Shiskin e Plewes, 1978).

Segundo Bell e Hillmer (1984), vários critérios têm sido propostos para avaliar a adequabilidade do ajuste sazonal e/ou para decidir quando um método faz um melhor trabalho de ajuste que outro, mas estas tentativas têm falhado. Assim, hoje não há padrões aceitos pelos quais os ajustes possam ser julgados. Estudos empíricos, comparando o desempenho de diferentes métodos sobre vários conjuntos de dados, não produzem um método único para determinar se um método é melhor que outro ou não.

2.5.1 Critérios para avaliar o ajuste sazonal.

No presente trabalho, não se pretende comparar o X-11 com outros métodos de ajuste sazonal. Portanto a avaliação será restrita ao X-11 e ao X-11 ARIMA. O trabalho de Dagum e Morry (1983) sobre a avaliação do desempenho do X-11 e do X-11 ARIMA na série de emprego e desemprego do Canadá na épocas de recessão e não recessão, e o trabalho de Shiskin e Plewes (1978) na série da taxa de desemprego dos E.U.A., guiam o presente enfoque.

Para Dagum e Morry o principal critério para a escolha do método que proporciona o melhor ajuste são as revisões, embora o trabalho esteja restrito à comparação do X-11 e o X-11 ARIMA. Entretanto para Shiskin e Plewes (além das revisões) os critérios são: sazonalidade, suavidade, aleatoriedade dos resíduos, participação do componente irregular em relação aos outros componentes e adequabilidade do modelo. O próprio X-11 possui uma série de estatísticas que avaliam a qualidade do ajuste global. Na presente seção é apresentado, de modo mais detalhado, somente o critério das revisões; os outros critérios serão apresentados com maiores detalhes nas seções 3.3.9, 3.3.10 e 3.3.11 e um resumo destas na seção 4.4.

2.5.2 Revisões

As revisões são as mudanças nos valores ajustados à medida que se incorporam novas observações. Para Shiskin e Plewes (1978), as revisões são, em essência, a medida do erro de especificação da sazonalidade prevista. Portanto, revisões grandes implicam erros grandes. Uma qualidade desejável é que as revisões sejam pequenas e que mostrem uma mudança de direção consistente, isto é, um comportamento monótono de revisão a revisão. Revisões grandes evidenciam deficiências na qualidade do ajuste do método e podem comprometer as decisões baseadas nestas estimativas, além da perda de credibilidade pública.

Muitos pesquisadores têm conduzido estudos empíricos de métodos de ajuste sazonal usando as revisões como critério para avaliar diversos métodos. Isto faz sentido quando se compara métodos de ajuste

que proporcionam o mesmo ajuste final, como o X-11 e o X-11 ARIMA, ou o X-11 um ano à frente e "concurrents models". Neste caso, os diferentes métodos tentam atingir o mesmo objetivo: o ajuste final do X-11. Comparações das magnitudes do total das revisões refletem quão perto o ajuste inicial está do ajuste final. Desde que o ajuste final seja, presumivelmente, melhor que o ajuste inicial, revisões totais menores são melhores (Bell e Hillmer, 1984).

Bell e Hillmer ilustram a impropriedade de usar medidas de revisão para julgar méritos relativos a métodos de ajuste sazonal que fornecem diferentes ajustes finais. A decisão sobre qual ajuste final é apropriado deveria ser baseada na informação dos dados, nas crenças sobre sazonalidade, e, se for possível, nos objetivos do ajuste sazonal. Portanto, na escolha do método de ajuste sazonal é importante concentrar a atenção no que se está estimando ao invés de considerar somente as revisões. Avaliar métodos de ajuste sazonal que produzem diferentes ajustes finais, usando revisões, é como julgar quão rápido um estimador de um parâmetro converge ao se incrementar o tamanho da amostra, mesmo que convirja para um valor errado. Eles concluem também que os filtros mais compridos levam a revisões anuais (mudança nos valores ajustados, conforme um ano de dados é adicionado) menores, e que estes atingem o ajuste final de forma mais lenta que os filtros mais curtos.

Para calcular o tamanho das revisões, é utilizado o Erro Absoluto Médio (MAE) dos fatores sazonais atuais, definido por Dagum e Morry (1983) como:

$$MAE = \sum_{t=1}^N |\hat{S}_t^c - \hat{S}_t^f| / N,$$

onde $\hat{S}_t^c$ denota o valor do fator sazonal atual ou um fator sazonal projetado um ano à frente. $\hat{S}_t^f$ denota o fator sazonal "final" no sentido de que ele não mudará significativamente quando a série for aumentada com novos dados. Para o X-11 e o X-11 ARIMA, um fator sazonal se transforma em final quando pelo menos três anos e meio de dados são adicionados (Young, 1968 e Wallis, 1974).

O X-11 usa filtros simétricos para observações centrais e assimétricos para observações no início e no final da série. Entretanto, o X-11 ARIMA amplia a série, podendo-se usar filtros próximos aos simétricos. Dagum e Morry mostram que as revisões do X-11 ARIMA são menores que as do X-11, mesmo em séries que possuem mudanças bruscas, como a série de desemprego em época de recessão.

A natureza e o tamanho das revisões dos métodos de ajuste sazonal X-11 e do X-11 ARIMA têm sido extensivamente analisados para indicadores econômicos de muitos países, abrangendo diversos períodos (Kenny e Durbin, 1982; Pierce, 1980b; Young, 1968, e outros). As conclusões dependem das séries e do período de tempo escolhido.

Capítulo III. O MÉTODO DE AJUSTE SAZONAL X-11 E O X-11 ARIMA.

3.1 Introdução.

O método X-11 desenvolvido por Shiskin, Musgrave e Young em 1967 no Bureau do Censo dos E.U.A., é sem dúvida o método de ajuste sazonal mais conhecido. Sua popularidade se deve aos bons resultados obtidos e à sua flexibilidade de aplicação em uma ampla gama de séries. Ele pertence à categoria dos procedimentos de médias móveis e é baseado na premissa de que o componente sazonal pode ser medido na série original e que pode ser separado dos outros componentes não sazonais.

Como foi discutido no capítulo anterior, este método é fortemente criticado, principalmente, pela sua natureza empírica e pela sensibilidade das estimativas dos fatores sazonais frente às revisões. Estas deficiências são, em parte, superadas pelo X-11 ARIMA, que ajusta um modelo ARIMA à série, estende-a com dados extrapolados pelo modelo e aplica o X-11 à série estendida. Assim podem ser usados filtros quase simétricos no final da série (o uso de filtros assimétricos no final da série pelo X-11 constitui a principal causa das altas revisões). Estes filtros assimétricos são uma combinação dos filtros do X-11 e dos filtros do modelo ARIMA que dependem da série e do modelo particular usado. Tais filtros conseguem acompanhar os movimentos mais recentes da série.

Como o método X-11 ARIMA engloba o X-11, pois além de conter os principais passos do X-11, fornece estatísticas adicionais, e, como no presente trabalho é utilizado o pacote do X-11 ARIMA, neste capítulo será apresentada a descrição do pacote do X-11 ARIMA, enfatizando-se as diferenças entre os dois programas.

3.2 O método de ajuste sazonal X-11.

3.2.1 Histórico.

Em 1954, o Bureau do Censo dos E.U.A. introduz o primeiro programa em computador eletrônico para ajustar sazonalmente séries temporais econômicas, aplicando o método da razão de média móvel em grande escala pela primeira vez. Em 1955, o programa original, o Census Method I foi substituído por um procedimento revisado chamado Census Method II. A partir daí o Bureau do Censo tem conduzido um programa extenso de pesquisa para melhorar o método. Estes esforços têm se concentrado no melhoramento das médias móveis, nos pesos dos filtros usados para o início e o final da série, na estimação da variação do calendário e na manipulação de valores extremos.

Variantes experimentais do programa Census Method II que incorporam os resultados destas pesquisas, estão identificadas com a letra X e um número subsequente. A primeira variante disponível ao público (1960) foi o X-3. Este difere do original na forma de substituir os valores extremos e no método de calcular os fatores sazonais para os anos mais recentes. As últimas variantes introduzidas nos anos recentes foram o X-9 em 1961, o qual foi usado como o programa padrão, e o X-10 para ajustar séries mais irregulares. O X-10 contém uma rotina

variável para o fator sazonal, que seleciona a média móvel para este. A largura do filtro depende das amplitudes relativas das flutuações do fator sazonal e do componente irregular num mês determinado, ampliando a gama das séries a serem ajustadas.

Em outubro de 1965, a versão X-11 substituiu as versões X-9 e X-10 como o programa padrão no Bureau do Censo. Este incluiu diversas melhorias sobre as primeiras versões. Muitas das novas características do X-11 contêm ferramentas adicionais para a análise de séries temporais. Embora o programa padrão seja suficiente para a maioria das aplicações, o usuário pode selecionar opções oferecidas pelo programa a fim de captar melhor as características peculiares de sua série.

Como resultado da disponibilidade destas opções, o X-11 se transformou num instrumento, não só para ajustes sazonais em grande escala de séries temporais, mas também para séries não usuais e ainda, como referência para pesquisar novas técnicas de ajuste de séries temporais.

3.2.2 Definição e notação dos componentes.

Existem particularidades nas definições e notações do método X-11 ARIMA. As principais são:

- O série original;
- S componente sazonal, definido como o padrão de variação entre anos, que se repete de forma quase constante e/ou envolve alguma particularidade de um ano para outro;
- C componente tendência-ciclo, que inclui o termo a longo prazo e os ciclos econômicos;
- D ou TD, variação pelo efeito calendário, que consiste nas variações atribuídas à composição do calendário;
- I componente irregular, que é composto pelas variações residuais, tais como: impactos repentinos de eventos políticos, efeito de greves, condições climáticas fora de estação, erros experimentais, etc.;
- CI é a série sazonalmente ajustada, consiste do componente tendência e irregular (C+I no caso aditivo);
- SI é a razão composta pela sazonalidade e pelo componente irregular (S+I no caso aditivo);
- M_c filtro para a tendência; e,
- M_s filtro para o fator sazonal.

3.3 Fundamentos do método de decomposição sazonal X-11 ARIMA.

3.3.1 Introdução.

Como foi visto anteriormente, o método X-11 é criticado principalmente pela falta de um modelo explícito para a série e seus componentes, e pela sensibilidade das estimativas do fator sazonal frente às revisões. Estas duas razões levaram Dagum e equipe a pesquisarem e proporem, em 1975, um método alternativo que superasse estas deficiências: o método X-11 ARIMA.

O X-11 ARIMA utiliza os modelos $ARIMA(p,d,q)(P,D,Q)_s$ (Box e Jenkins, 1976) para ajustar a série original, ampliando-a com um ou dois anos de dados extrapolados segundo o modelo ajustado e, finalmente, aplica o X-11 para decompor a série. A combinação do modelo ARIMA, sua extrapolação e as médias móveis do X-11 reduzem significativamente o tamanho das revisões das previsões e dos valores correntes das estimativas dos fatores sazonais (Dagum, 1978b, 1980).

O fato de se utilizar uma série estendida faz com que os filtros aplicados às observações para gerar previsões sejam quase simétricos e próximos aos filtros aplicados às observações centrais. Isto dá maior confiabilidade às estimativas (Dagum, 1980). Por outro lado, pode ser usado o filtro 24-termos para a estimativa inicial da tendência, o qual fornece melhores resultados quando a série é fortemente afetada por ciclos curtos ou mudanças repentinas na tendência.

Os pesos finais do X-11 ARIMA para estimar a tendência são uma combinação de filtros simétricos H-Henderson e de pesos assimétricos do modelo ARIMA usados na extrapolação dos dados. Tendo em vista que estes pesos finais mudam conforme o modelo ARIMA ajustado, eles refletem os movimentos mais recentes da série, e, como resultado, raramente perdem um ponto de inflexão (Dagum, 1979b). Deste modo são obtidas melhores estimativas da razão (ou diferença) SI.

Uma vantagem importante deste método é que oferece um modelo ARIMA para a série global, o qual serve de controle para analisar se a série satisfaz os princípios "subjacentes" da decomposição. Quando a série não segue um modelo ARIMA, a decomposição é duvidosa, porque isto pode significar que ela tenha sido gerada por um processo aleatório puro, ou que ela esteja contaminada por irregulares sistematicamente não identificáveis.

3.3.2 Os modelos ARIMA, sua extrapolação e sua seleção.

As razões que levaram Dagum a escolher os modelos ARIMA foram: robustez, parcimônia, simplicidade, eficiência para fazer extrapolações a curto prazo, extrapolações ótimas no sentido de possuir o erro quadrático médio mínimo (EQM), e previsões que acompanhem adequadamente o movimento entre anos, mesmo quando este altera seu nível (Dagum, 1978a, 1978b).

O modelo $ARIMA(p,d,q)(P,D,Q)_s$:
$$\phi_p(B)\bar{\phi}_p(B)\Delta^d\bar{\Delta}_s\Delta_t^D = \theta_q(B)\bar{\theta}_q(B)a_t$$

Δ^d, Δ^D São os operadores diferença simples e sazonal de ordem d e D respectivamente, onde $\Delta = 1-B$ e $\Delta_s = 1-B^s$.

$\phi(B), \Phi(B)$ São os polinômios autorregressivos (AR), ordinário e sazonal de ordem p e P respectivamente, com raízes fora do círculo unitário. Medem o impacto do passado.

$\theta(B), \Theta(B)$ São os polinômios de médias móveis (MA) ordinária e sazonal de ordem q e Q respectivamente, com raízes fora do círculo unitário. Medem o impacto dos resíduos sobre Z_t .

s é o período sazonal.

Assim um modelo ARIMA é um modelo dinâmico com resíduos autocorrelacionados.

O programa oferece duas opções: uma é a opção automática, onde o programa ajusta e escolhe um modelo ARIMA e a outra é a opção na qual o usuário ajusta seu próprio modelo.

Quando se usa a opção automática, pode acontecer que os modelos sejam rejeitados por não cumprirem certos critérios (a serem descritos posteriormente). Entretanto, a opção do usuário ajusta e extrapola o modelo, mesmo que o modelo possa não cumprir com os critérios.

Se for possível, a identificação do modelo deve ser feita com dados que já sofreram tratamento para eliminação da influência de valores extremos. A substituição de valores extremos é importante, principalmente, quando estes acontecem nos anos mais recentes, pois estes podem levar à rejeição de modelos adequados.

3.3.2.1 A opção ARIMA automática.

O programa ajusta e verifica automaticamente três modelos para a opção aditiva e multiplicativa: $(0,1,1)(0,1,1)_s$, $(0,2,2)(0,1,1)_s$ e $(2,1,2)(0,1,1)_s$. Para a opção multiplicativa os dados sofrem transformação logarítmica nos dois primeiros modelos. Estes modelos foram escolhidos por Dagum e equipe, baseados na experiência com 12 modelos aplicados em 174 séries econômicas (Dagum, 1978b, 1979a, 1980, 1981; Lothian e Morry, 1978c). Trabalhos recentes da equipe propõem outros modelos e modificações nos critérios de escolha (Huot et al. 1986).

Os modelos ajustados devem cumprir certos critérios de "bom ajuste" e o programa seleciona o melhor dentro dos modelos que satisfazem estes critérios.

Os critérios de adequabilidade dos modelos são:

- 1) aleatoriedade dos resíduos;
- 2) evidência de superdiferença; e,
- 3) geração de previsões razoáveis.

Estes critérios são avaliados da seguinte maneira:

- 1) Para examinar a aleatoriedade dos resíduos é usado o teste de Portmanteau, desenvolvido por Box e Pierce (1970). Este teste verifica a adequabilidade do modelo ajustado ARIMA, tomando as autocorrelações dos resíduos como um todo. Esta estatística X^2 é incorporada no X-11 ARIMA. Dado o risco e a gravidade de se cometer o erro de tipo II (aceitação da hipótese nula quando falsa), ao se testar a hipótese nula de não autocorrelação, o valor crítico do teste é normalmente considerado igual a 10%.

É importante verificar se o modelo ARIMA foi ajustado para dados sem outliers, porque estes podem facilmente distorcer o valor da estatística, e assim, incrementar ou diminuir falsamente o p-valor do teste.

- 2) A excessiva aplicação do operador diferença para gerar uma série estacionária pode induzir a um processo de média móvel não invertível para os resíduos. Quando há presença de superdiferença, a autocorrelação de primeira ordem dos resíduos é muito alta, mas frequentemente isto não é detectado pelo teste de Portmanteau. O enfoque correto para detectar superdiferença é testar se as raízes associadas com o polinômio de médias móveis, são significativamente diferentes de 1. Para um processo de média móvel de ordem maior que 1, existe evidência de superdiferença, quando a soma dos parâmetros de média móvel ordinários ou a soma dos parâmetros de média móvel sazonal (denotado por R^2) for maior que 0.9 (Dagum, 1981).
- 3) O critério aplicado pelo X-11 ARIMA para avaliar o desempenho das previsões (para frente ou para o passado) dos modelos é o erro percentual absoluto médio (MAPE), estimado por:

$$MAPE = \frac{100}{N} * \sum_{l=1}^N \left| \frac{Z_t - \hat{Z}_t(l)}{Z_t} \right|$$

onde $\hat{Z}_t(l)$ é o valor extrapolado pelo modelo l passos à frente do origem t .

Este erro é expresso em percentagem do nível da série não ajustada para permitir sua inclusão na opção automática de seleção do modelo. Este erro é calculado para cada um dos três primeiros e três últimos anos da série, assim como para todos os meses dos três primeiros e três últimos anos. As extrapolações são produzidas em diferentes origens para um ano e são comparadas com o valor observado correspondente, para determinar a magnitude do erro. Analisando só os três últimos anos, o desempenho do modelo é, principalmente, avaliado pela sua habilidade para acompanhar os movimentos mais recentes da série original.

O limite de aceitação foi fixado para o máximo erro de previsão para frente em 12% e para trás em 18%. Estes limites estão baseados na experiência com 92 séries americanas e canadenses (Dagum, 1981).

O programa escolherá o modelo com menor MAPE, mas se a diferença com respeito aos outros modelos for pequena, é recomendável selecionar o modelo com menor número de parâmetros.

A opção automática somente pode ser usada para séries com mais de 5 anos. Para séries com mais de 15 anos, só os últimos 15 anos são usados para o ajuste e extrapolação. Quando a série possui menos de 15 anos o programa estende automaticamente a série original com um ano de dados no início e no final da série e aplica o X-11. Neste caso, se algum modelo for aceito, o programa apresentará todas as estatísticas relativas à previsão para trás.

Caso nenhum dos três modelos seja aceito, uma mensagem é dada, indicando que os valores extrapolados não foram incorporados. Isso ocorre mais frequentemente com séries de fluxo, tais como importações, comércio e outros que podem ser afetadas fortemente por greves ou pelo calendário. É necessário que estas fontes de variação sejam removidas antes de se usar a opção automática.

O programa oferece uma opção onde os valores extremos são substituídos pelos valores ajustados pelo modelo ARIMA que foi escolhido na primeira execução da opção e, então, o mesmo modelo é resubmetido para modificar a série a extrapolar. Neste caso o programa apresenta todas as estatísticas recalculadas com os valores extremos substituídos, chamando-o de modelo final (mf).

Quando os 3 modelos da opção automática forem rejeitados, o usuário deve determinar se a rejeição é devida a um erro médio de extrapolação extremamente grande para um ano particular. Se este for o caso, os modelos ainda poderão ser considerados bons se o ano em questão tiver sido um ano incomum, por exemplo, devido a uma forte recessão. O melhor modelo será resubmetido usando a opção correspondente ao modelo ARIMA identificado pelo usuário.

3.3.2.2 A identificação do modelo ARIMA pelo usuário.

Mesmo que a opção automática rejeite os 3 modelos, ainda assim estes podem ser colocados numa segunda opção, se for considerada satisfatória uma probabilidade X^2 entre 5% e 10% e/ou MAPE entre 12% e 15%, para séries altamente irregulares.

Se nenhum dos 3 modelos for aceito, o usuário deverá identificar um novo modelo. Na maioria dos casos a identificação de um bom modelo, requer mudanças pequenas em relação aos modelos da opção automática. As seguintes regras têm sido úteis para melhorar o ajuste e extrapolação de um grande número de séries (Dagum, 1980).

- 1) Correção da baixa probabilidade X^2 . Isto acontece quando os resíduos do modelo ajustado são autocorrelacionados, o que ocorre pelo uso indevido da transformação logarítmica, ou pelo não uso quando necessário. Logo, deve-se tentar a transformação (ou não) para corrigir o valor X^2 . Em outros casos um valor pequeno de X^2 é resultado de superdiferença.

- 2) Correção da superdiferença. Evidência de superdiferença leva a cancelamento de parâmetros, sugerindo um modelo mais parcimonioso. É necessário ter cuidado ao se eliminar parâmetros pois a sazonalidade fica perturbada. Para evitar isto é preciso que se analise cuidadosamente os testes relacionados com a presença de sazonalidade.

Geralmente, ao se fazer as correções, tem-se um ganho na diminuição dos erros de extrapolação. Entretanto, se este não for o caso, o usuário deverá identificar seu próprio modelo.

3.3.3 Principais passos do ajuste sazonal X-11 ARIMA.

Os passos são os mesmos que os do X-11, as principais diferenças são:

- 1) a série é estendida em um ano no início e no final, se a opção ARIMA é usada e aceita;
- 2) pode ser usada a opção 24-terms para a estimação inicial da tendência; e
- 3) séries curtas de 3 ou 4 anos são sazonalmente ajustadas com a opção de sazonalidade estável (média simples).

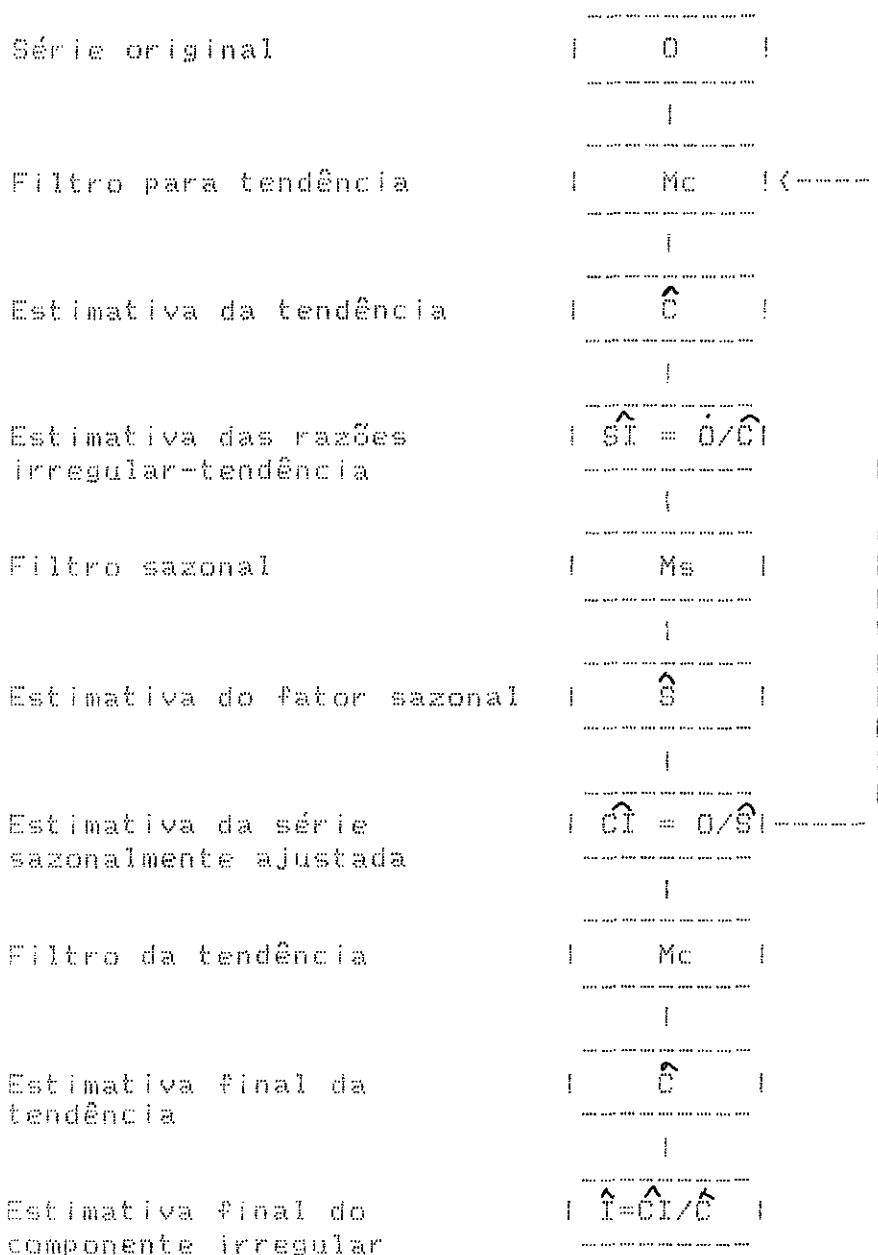
Os principais passos do X-11 ARIMA para o ajuste sazonal, usando a opção padrão (default option) multiplicativa são:

- Passo 1. Estimativa inicial da razão $\hat{SI}$, como quociente entre a série original e a estimativa da tendência, que é a série ajustada pela média móvel centrada 12-terms. $\hat{SI} = O/\hat{C}$
- Passo 2. Estimativa inicial do fator sazonal $\hat{S}$, aplicando a média móvel 5-terms (3x3) a cada mês, separadamente da razão $\hat{SI}$, estimada no passo anterior. $\hat{S} = Ms(\hat{SI})$
- Passo 3. Ajuste da soma a 12, fazendo o quociente entre a estimativa do fator sazonal $\hat{S}$ e do fator sazonal $\hat{S}^*$ ajustado através da aplicação do 12-terms. Para se recuperar as 6 observações do início e final da série, repete-se o primeiro e o último valores da móvel disponível seis vezes. $\hat{S} = \hat{S}/\hat{S}^*$
- Passo 4. Estimativa do componente irregular fazendo o quociente entre a razão $\hat{SI}$ e fator sazonal. $\hat{I} = \hat{SI}/\hat{S}$
- Passo 5. Identificação e eliminação de valores extremos. Um esquema gradual é usado para detectar e remover valores extremos (ver seção 3.3.7).
- Passo 6. Obtenção da estimativa preliminar dos fatores sazonais $\hat{S}$, aplicando a média móvel 5-terms a cada mês das razões $\hat{SI}$ com os valores extremos substituídos.
- Passo 7. Ajuste da soma a 12, idem passo 3.
- Passo 8. Estimativa preliminar da série sazonalmente ajustada, fazendo o quociente entre a série original e a estimativa dos fatores sazonais. $\hat{CI} = O/\hat{S}$
- Passo 9. Estimativa da tendência, aplicando o filtro 9-H, 13-H ou 23-H segundo a razão $\hat{I}/\hat{C}$ à série sazonalmente ajustada. Na primeira estimativa só se usa o 13-H. $\hat{C} = Mc(\hat{CI})$
- Passo 10. Nova estimativa da razão $\hat{SI}$, fazendo o quociente entre a série original e a estimativa da tendência. $\hat{SI} = O/\hat{C}$

- Passo 11. Estimativa final dos fatores sazonais aplicando a cada mês separadamente da razão SI o filtro 7-termos (3x5). $\hat{S} = M_s(SI)$
- Passo 12. Ajuste da soma a 12, idem passo 3.
- Passo 13. estimativa final da série sazonalmente ajustada. Dividir a série original entre o fator sazonal. $\hat{CI} = O/\hat{S}$

O X-11 é um procedimento iterativo. Repete alguns destes passos mais de uma vez e neste processo se obtém resultados mais suaves. O método ganha flexibilidade dado que o programa possui várias opções as quais permitem ao usuário se aproximar do mecanismo gerador de suas séries, conseguindo resultados mais precisos.

Estrutura do X-11.



3.3.4 Propriedades básicas dos filtros do X-11 e do X-11 ARIMA.

A estimação dos componentes é feita através de diferentes tipos de combinações lineares das observações, chamadas de médias móveis ou filtros (Dagum, 1976). Um filtro transforma uma série em outra, segundo a seguinte relação.

$$O_t^* = \sum_{j=-m}^m \delta_j O_{t-j}, \text{ onde } \sum_{j=-m}^m \delta_j = 1$$

Os pesos δ_j são constantes. Neste caso o comprimento do filtro é de $2m+1$. Segundo os valores dos pesos, os filtros são classificados em simétricos e assimétricos. Os primeiros são usados para observações centrais e os outros para as m primeiras e m últimas observações. A soma destes pesos é 1, e portanto, a média da série original não é alterada no processo de filtragem.

As propriedades básicas dos filtros, segundo Dagum (1978a) são:

- 1) preservação de escala;
- 2) princípio de superposição; e,
- 3) invariância no tempo.

A preservação de escala significa que, se a série original for ampliada por uma constante, então a série suavizada deverá ser ampliada pelo mesmo fator. O princípio de superposição significa que, se um filtro for aplicado à soma de duas séries o resultado será a soma das duas séries ajustadas separadamente pelo mesmo filtro. Estas duas propriedades são consequências do fato do filtro ser uma transformação linear. A invariância no tempo significa que se duas entradas forem as mesmas, exceto para uma defasagem no tempo; então as saídas também serão as mesmas, exceto para a mesma defasagem no tempo; isto é, o comportamento deste tipo de transformação ou filtro não muda no tempo, os filtros sempre responderão da mesma maneira.

As funções que não podem ser eliminadas nem afetadas pelos filtros são: as linhas retas, polinômios de qualquer grau, exponenciais, funções trigonométricas e produtos de qualquer destas por potências de t .

Por outro lado, é importante que os filtros não introduzam defasagens na fase. Os filtros simétricos introduzem defasagens em alguns componentes e em outros não. Em termos práticos, estas defasagens são consideradas nulas. Entretanto, os filtros assimétricos introduzem defasagens em todos os componentes. Isto é indesejável pois a série ajustada pode ter uma inflexão antes ou depois da série original. No X-11, os filtros assimétricos para o fator sazonal introduzem defasagem; maiores detalhes podem ser encontrados em Dagum (1978a).

Nesta seção são apresentados cada um dos filtros e suas características principais.

O conjunto de pesos dos filtros pode ser gerado segundo duas técnicas, ajustando um polinômio ou através do operador soma. O primeiro método requer que o comprimento do filtro (n), assim como o grau

do polinômio (p), sejam escolhidos antecipadamente. Fixados n e p , os coeficientes são estimados por mínimos quadrados e então o polinômio é usado para determinar o valor do comprimento de n médias móveis de um polinômio de grau p , de tal modo que não dependa das observações. O valor suavizado é uma média ponderada das observações. Os pesos independem de qualquer parte da série que seja tomada. Para um polinômio de grau fixo, a variância da série suavizada decresce quando o comprimento do filtro cresce (Dagum, 1978a).

O X-11 usa os filtros gerados através de fórmulas de soma. O princípio básico desta técnica é a combinação de operações de diferença e soma, de tal forma que a diferença acima de certa ordem é desprezível, pois eles reproduzem estas funções. O mérito deste procedimento é que os valores assim obtidos são funções de um grande número de valores observados, cujos erros para uma extensão considerável se cancelam.

Henderson (1916), salientou que para suavizar dados reais, deve existir uma compatibilidade entre o ajuste e a suavidade da série ajustada. A falta de ajuste é medida pela soma dos quadrados dos desvios entre os valores observados e os valores ajustados. A falta de suavidade é medida pela soma de quadrados da terceira diferença da curva suavizada. Assim, ele desenvolveu uma fórmula para a qual a soma da terceira diferença da série suavizada é mínima para qualquer número de termos.

Estes filtros têm a propriedade de que, quando são ajustados a um polinômio de segundo ou terceiro grau, os resultados caem exatamente sobre essas curvas. Quando são ajustados a dados estocásticos, os filtros fornecem resultados mais suaves do que os obtidos através dos pesos de um ponto central de uma parábola de segundo grau ajustada pelo método de mínimos quadrados. Estes pesos são dados pelo produto da minimização da soma dos quadrados da terceira diferença da série. Os filtros de Henderson são aplicados para melhorar as estimativas finais da tendência-ciclo (Dagum, 1978a).

Os filtros H-Henderson e os filtros para o fator sazonal, segundo o comprimento $2m+1$, possuem m filtros assimétricos, e, além disto, o filtro sazonal possui um filtro adicional para extrapolar os fatores. A tabela A apresenta os pesos do filtro 3×3 para o fator sazonal.

Tabela A. Pesos do filtro sazonal 3×3 .

Fator por ano	pesos dados as razões SI no ano				
	N-4	N-3	N-2	N-1	N
N+1	0.000	-0.056	0.148	0.426	0.481
N	0.000	0.000	0.185	0.407	0.407
N-1	0.000	0.111	0.259	0.370	0.259
N-2	0.111	0.222	0.333	0.222	0.111

- 1) O 12-termos centrado. é uma média de 13 termos, onde o primeiro e o último têm pesos igual a $1/24$ e os restantes $1/12$. Para um modelo aditivo tendo as características de tendência linear e sazonalidade estável, este filtro reproduz exatamente a primeira e elimina a segunda. Se a relação entre os componentes é multiplicativa, então só uma tendência constante multiplicada por uma sazonalidade estável será perfeitamente reproduzida.

A principal limitação é que o filtro perde os picos e baixos dos pequenos ciclos (2 ou 3 anos), e, a menos que a variação irregular seja pequena, os resultados não serão satisfatórios. Contudo, este filtro é geralmente bom para a estimativa inicial da tendência, dado que a maioria das séries econômicas possuem grandes ciclos, de 40 anos ou mais (Dagum, 1980).

- 2) O 24-termos centrado. Este filtro é incorporado para lidar, principalmente com séries dominadas por flutuações cíclicas curtas (2 ou 3 anos) ou afetadas por mudanças repentinas no nível da tendência. Este filtro reduz a variação irregular mais que o filtro 12-term, e usa filtros assimétricos, pois a qualidade das estimativas das 12 primeiras e 12 últimas observações se deteriora gradualmente conforme as observações se afastem do centro (Dagum, 1980).

- 3) Os filtros de Henderson. Estes filtros são aplicados nas séries sazonalmente ajustadas a fim de estimar a tendência e são: 9-H, 13-H e o 23-H, cujos comprimentos são 9, 13 e 23 respectivamente. O filtro 13-H é aplicado para a estimativa preliminar da tendência e geralmente para estimativa final da maioria das séries. O filtro 9-H é usado para séries mais suaves e o filtro 23-H para séries altamente irregulares. Estes dois últimos filtros são aplicados na estimativa final da tendência a fim de melhorar estas estimativas (Dagum, 1978a, 1980).

A escolha do filtro adequado pode ser feita através da opção padrão ou o usuário escolhe o filtro segundo suas necessidades. No caso da opção padrão, a escolha é feita com base na razão I/C (ver rotina variável para a tendência, seção 3.3.5).

- 4) Os filtros para o fator sazonal. Estes filtros são aplicados às razões (diferenças) SI para cada mês separadamente sobre todos os anos, a fim de estimar o componente sazonal. Seus pesos são todos positivos, e, em consequência, eles reproduzem o valor central de uma linha reta no período de seus comprimentos (Dagum, 1980). Portanto, estes filtros podem aproximar muito bem mudanças sazonais graduais que seguem um padrão não linear ao longo de toda a série.

Os filtros mais usados são o 3x3 e o 3x5, que são aplicados à maioria das séries. Para séries altamente irregulares, o programa proporciona o filtro 3x9 e o filtro estável n-termos; este último é uma média simples de n termos. Para séries curtas, tem-se modificações de todos estes filtros. A seleção do filtro é feita através do próprio programa ou o usuário escolhe o filtro segundo as características de sua série (v. seção 3.3.6).

- a) O filtro 5-termos (3x3). Este filtro possui 5 termos, sendo o filtro sazonal de menor comprimento, e portanto muito flexível. Este filtro permite seguir satisfatoriamente mudanças rápidas em direção, mas o componente irregular deve ser pequeno para que as razões SI possam ser suavizados de forma satisfatória. Em geral este filtro é aplicado para a estimativa preliminar do fator sazonal, mas pode ser utilizado na estimativa final pela opção do usuário (ver seção 3.3.6).
 - b) O filtro 7-termos (3x5). O comprimento deste é de 7 termos, este filtro é menos flexível que o filtro anterior, e é aplicado para a estimativa final do fator sazonal, ou em séries cujo componente irregular é grande. Para séries altamente irregulares o programa dispõe de outros filtros adicionais mais adequados, os quais são aplicados para períodos maiores e produzem estimativas dos fatores sazonais mais suaves.
- 5) Os filtros assimétricos. Em todo processo de suavização através de médias móveis, as m primeiras e as m últimas observações não podem ser suavizadas pelos filtros simétricos usados para as observações centrais. Isto se constitui na principal limitação do X-11, pois as estimativas dos fatores sazonais no início e no final da série são menos confiáveis que as das observações centrais, e como consequência disto tem-se a sensibilidade das estimativas dos fatores sazonais frente às revisões (Dagum, 1982).

Diferentemente do X-11, os filtros usados pelo X-11 ARIMA na estimação das m primeiras e m últimas observações da série são obtidos de uma combinação linear de dois tipos de filtros: (Dagum 1978a, 1979b, 1980)

- a) Filtros assimétricos do modelo ARIMA usados para extrapolar dados originais.
- b) Filtros do X-11 usados para ajuste sazonal.

Os filtros do modelo ARIMA mudam com a série e portanto são mais flexíveis. Estes filtros refletem a maioria dos movimentos recentes da série, particularmente para mudanças sazonais rápidas.

Os filtros do X-11 aplicados à série original estendida para a estimação da tendência-ciclo são simétricos. Em consequência eles não perdem pontos de inflexão e não introduzem defasagens, o que permite estimar a variação cíclica muito bem.

Os filtros para estimar os fatores sazonais são ainda assimétricos, mas quase próximos aos filtros simétricos usados nas observações centrais. Assim, com um ano de dados extrapolados, os fatores sazonais previstos são obtidos de dados extrapolados com os filtros do X-11 usados para produzir o ajuste sazonal atual (Dagum, 1980).

Esta combinação dos dois tipos de filtros, fixo e flexível do modelo ARIMA faz do X-11-ARIMA um método de ajuste sazonal melhor do que o X-11 (Dagum 1978a, 1980, 1983a).

3.3.5 Rotina variável para tendência-ciclo.

Esta rotina é um subprograma construído com o objetivo de selecionar o filtro para o ajuste final da tendência. Esta rotina oferece duas opções. Uma opção padrão na qual o programa seleciona automaticamente o filtro segundo a razão $I/\bar{C}$ e uma opção na qual o usuário escolhe o filtro de acordo com as características peculiares de sua série (Dagum, 1980).

A razão $I/\bar{C}$ consiste no quociente entre as médias das variações absolutas mensais do componente irregular e da tendência. Esta razão mede a variabilidade do componente irregular versus a tendência. Um valor alto da razão indica que a série é altamente irregular e alerta contra o perigo de uma separação insatisfatória destes componentes (v. estatística M3 na seção 3.3.11).

A opção automática seleciona uma média móvel que compatibiliza a necessidade de suavizar o irregular com uma média móvel inflexível de muitos termos e a necessidade de reproduzir de forma precisa o elemento sistemático com uma média flexível de poucos termos. Para séries altamente irregulares, o filtro 23-H é escolhido e, para séries suaves, o programa escolhe o filtro 9-H; entretanto, o filtro 13-H é usado na maioria das séries. O filtro é selecionado segundo os valores de $I/\bar{C}$ mostrados na tabela B.

Tabela B. Escolha do filtro para a tendência.

razão $I/\bar{C}$	filtro para tendência selecionado	grau de irre- gularidade
0.00 - 0.99	9-termos de Henderson	baixa
1.00 - 3.49	13-termos de Henderson	padrão
3.50 a mais	23-termos de Henderson	alta

As três médias móveis substituem a média 15-termos de Spencer usada nas primeiras versões do X-11. Os três filtros foram desenvolvidos por Henderson, sob o mesmo critério que o 15-termos, isto é, eles minimizam a soma de quadrados da terceira diferença da curva. A diferença com o 15-termos é que o filtro 9-H é usado para séries suaves e o filtro 23-H para séries altamente irregulares (Shiskin et al. 1967). Nota-se que esta rotina serve para a escolha do filtro adequado para a estimativa final da tendência. Nas outras estimativas o programa usa o filtro 13-H.

3.3.6 Rotina variável para o fator sazonal.

No X-11 esta rotina oferece duas opções, uma onde o programa escolhe automaticamente os filtros para a estimativa preliminar (3x3) e final (3x5) e outra onde o usuário escolhe o filtro para cada mês, segundo suas necessidades. Os filtros disponíveis são: 3x1, 3x3, 3x5, 3x9. Na versão X-10 os filtros eram selecionados automaticamente, segundo a razão $\bar{I}/\bar{S}$ (Shiskin et al. 1967) dados na tabela C.

Tabela C. Escolha do filtro sazonal.

$\bar{I}/\bar{S}$	média móvel
0.00 - 1.49	3 - termos
1.50 - 2.49	3x3 - termos
2.50 - 4.49	3x5 - termos
4.50 - 6.49	3x9 - termos
6.50 - 8.49	3x15 - termos
8.50 a mais	n - termos

Do mesmo modo que o X-11, a opção padrão do X-11 ARIMA ajusta as razões SI com o filtro 3x3 para a estimativa preliminar e o filtro 3x5 para a estimativa final. A diferença, consiste nas opções que oferece ao usuário. Estas opções são: 3x3, 3x5, 3x9, n-termos e a opção de "livre escolha", onde n-termos é uma média simples de n termos. A opção de "livre escolha" consiste em que o programa escolhe o filtro adequado para cada mês, segundo o valor da razão $\bar{I}/\bar{S}$, para a estimativa final. Nesta opção o usuário não tem controle na escolha dos filtros (Dagum, 1980).

3.3.7 Tratamento dos valores extremos.

Um esquema gradual é usado para detectar e remover valores extremos. Este esquema consiste em: a) calcular um desvio padrão móvel (σ) de 5 anos; b) testar se o valor do ano central é superior a $2,5*\sigma$; e, c) remover valores acima deste limite. Recalcula-se o σ e testa-se novamente, mas nesta vez, para remover os valores aplica-se o seguinte esquema:

```

Ot < x*sigma      peso 1
x*sigma < Ot < y*sigma      peso linear entre 0 e 1
Ot > y*sigma      peso 0

```

onde x e y são escolhidos pelo usuário. Na opção automática os valores são: $x=1,5$ e $y=2,5$. Maiores detalhes podem ser encontrados em Shiskin et al. (1967).

3.3.8 Vantagens do X-11-ARIMA sobre o X-11.

As principais vantagens segundo Dagum (1978a, 1980), são:

- 1) A existência de um modelo de ajuste indica que os dados cumprem com o princípio fundamental do ajuste sazonal, isto é, que a série é decomponível. A inexistência de um modelo questiona a validade do ajuste sazonal, pois indica que a série é uma combinação de componentes irregulares, sistematicamente não identificáveis. Ou pode indicar deficiências nos dados, por exemplo, no tipo de intervalo amostral.

Quando a série segue um modelo ARIMA, o valor esperado e a variância da série podem ser calculados, e portanto podem ser construídos intervalos de confiança para as observações, que permitirão a identificação de valores extremos, particularmente no final da série.

- 2) A extrapolação do modelo ARIMA é a extrapolação do erro quadrático médio mínimo e pode ser usada como valor projetado.
- 3) Quando os fatores sazonais atuais são aplicados para obter dados atuais sazonalmente ajustados, não é necessário revisar a série mais de duas vezes. Para muitas séries uma única revisão será suficiente para fornecer estimativas do fator sazonal que são finais no sentido estatístico (Dagum, 1983b).
- 4) O erro total nas previsões do fator sazonal é significativamente reduzido para todos os meses. Geralmente a redução de 30% para o vício e de 20% no valor absoluto do erro total foi encontrada em séries americanas e canadenses (Dagum, 1978a, 1980).

Isto ocorre porque o X-11-ARIMA usa a combinação de dois tipos de filtros; desta forma as previsões do fator sazonal são obtidas de dados extrapolados na base de filtros ainda assimétricos mas que são mais próximos dos filtros aplicados às observações centrais comparados com os do X-11.

- 5) Outra vantagem é que a estimativa da última observação da tendência é feita com pesos simétricos do filtro de Henderson (que pode reproduzir uma cúbica num espaço de tempo) combinado com os pesos do modelo ARIMA usados para extrapolar dados. Dado que os pesos do modelo ARIMA mudam com o modelo para uma série particular, eles refletem os movimentos mais recentes, e uma melhor estimativa é obtida destes pesos combinados. Isto é particularmente verdadeiro para anos com pontos de inflexão, porque o X-11 aplica pesos assimétricos de Henderson, os quais só podem ser adequadamente estimados se a tendência for linear.
- 6) Finalmente, adicionando-se 1 ou 2 anos de dados extrapolados (sem extremos, dado que são meras projeções) pode-se obter uma estimativa melhor da variância, o que permite uma melhora significativa na identificação de outliers, que podem distorcer seriamente as estimativas obtidas com uma suavização linear.

Para fatores sazonais atuais, as mesmas observações são válidas, exceto que os filtros sazonais estão mais próximos dos filtros centrais que os correspondentes às previsões dos fatores sazonais extrapolados. Por esta razão, o número de revisões na estimativa do fator sazonal é também significativamente reduzido. Tem-se que um ano de previsão para frente e para trás é o melhor compromisso para a maioria das séries quando se usa a opção padrão.

3.3.9 Testes para sazonalidade.

Existem basicamente dois tipos de sazonalidade: a estável e a móvel. A variação é dita estável, quando pode ser representada por uma função estritamente periódica. Quando a periodicidade for de um ano e o modelo for aditivo, os valores dos fatores sazonais são 12 constantes que somam zero de qualquer origem no tempo, (no modelo multiplicativo, estes fatores somam 1200). Para a maioria das séries, a amplitude e/ou a forma da sazonalidade mudam ao longo do tempo. Em tais casos, diz-se que a série tem sazonalidade móvel (Dagum, 1974).

Segundo Shiskin e Plewes (1978), uma propriedade altamente desejável para qualquer procedimento de ajuste sazonal é a capacidade de identificar um padrão sazonal em uma série que mostre alguma estabilidade ao longo do tempo. A presença de sazonalidade móvel pode dificultar a identificação do fator sazonal pelo método, distorcendo as estimativas, especialmente quando se está interessado em fazer previsão da sazonalidade de futuras observações.

Os testes para sazonalidade estável e móvel são aplicados às estimativas da razão sazonal-irregular SI, isto é, à série livre da tendência.

3.3.9.1 Testes para sazonalidade estável.

O teste Fs_{B1} é o teste F baseado na análise de variância de um fator (mês) sobre as razões SI (ou diferenças), similar ao usado na versão do X-11 para detectar a existência de sazonalidade estável. A hipótese nula é que não há diferença entre as médias mensais, isto é, não há sazonalidade estável. A diferença reside na estimativa da tendência, que é feita diretamente na série original pela aplicação do filtro 12-termos centrado. Esta estimativa é removida fazendo o quociente (ou diferença) com a série original. Maiores detalhes podem ser encontrados em Lothian e Morry (1978b).

O valor da razão Fs_{B1} é o quociente de duas variâncias, uma que é a variância entre meses e que é devida principalmente à variação sazonal e a outra que é a variância residual devida principalmente às flutuações aleatórias do componente irregular. Portanto, valores altos de Fs_{B1} indicam a presença de sazonalidade estável.

Dado que muitas das hipóteses básicas do teste Fs_{B1} são provavelmente violadas, o valor da razão Fs_{B1} a ser usado para rejeitar a hipótese nula (presença de sazonalidade estável não significativa) é testado ao nível de 0,1%. Na versão X-11, o valor crítico de F é 2,41 e segundo Lothian e Morry, valores de F menores que 10, indicam que

não há suficiente sazonalidade para usar o X-11.

Este mesmo teste (denotado por Fs_{D8}) é aplicado aos valores da estimativa final das razões SI sem modificação pelos valores extremos. A diferença entre estes dois testes reside no tipo da estimativa para a tendência. No primeiro caso a estimativa é preliminar e no segundo a estimativa é final. Espera-se que esta segunda estimativa renova com mais precisão a tendência, tornando mais estáveis as razões SI, melhorando a qualidade da estatística F. Em geral espera-se que este valor seja superior ao correspondente a Fs_{B1} .

O teste $X2_{K_W}$ qui-quadrado de Kruskal-Wallis (Morettin e Toloi, 1981) é um teste não paramétrico usado para testar a diferença entre meses, baseado no posto médio de cada mês dos postos assinalados a cada estimativa final da razão SI. Calcula uma estatística e a compara com um valor qui-quadrado com $s-1=11$ graus de liberdade. Valores altos desta estatística evidenciam a existência de diferença significativa entre meses (Lehmann, 1974).

3.3.9.2 O teste F_m para sazonalidade móvel.

A sazonalidade móvel pode ser descrita pela evolução ou pelo movimento no padrão sazonal. Sazonalidade, para a maioria das séries, não é uma função periódica, que se repete com periodicidade de 12 meses. Há condições de mudanças constantes, as quais resultam em mudanças graduais do padrão sazonal e sua amplitude ao longo do tempo, por exemplo, mudanças em tecnologia que alteram a influência do clima, mudanças nos costumes ou hábitos, mudanças na importância relativa de grupos demográficos ou fenômenos econômicos que compõem as estimativas agregadas, mudanças institucionais, etc.

A fim de medir a proporção do movimento sazonal, um teste F, para sazonalidade móvel foi desenvolvido por Higginson (1975), baseado na análise de variância de dois fatores sobre as razões SI. Este teste, denotado por F_m , tem por função detectar a sazonalidade móvel caracterizada pelas mudanças graduais na amplitude sazonal, mas não na fase.

A variância total é a soma de:

- i) σ^2_m é a variância entre meses, mede a magnitude da sazonalidade e é igual à soma dos quadrados das diferenças entre as médias de cada mês e a média geral, corrigida pelos graus de liberdade correspondentes.
- ii) σ^2_y é a variância entre anos, mede o movimento ano a ano da sazonalidade, e é igual à soma de quadrados das diferenças entre a média anual e o total, corrigida pelos graus de liberdade correspondentes.
- iii) σ^2_r é a variância residual, obtida pela diferença entre a variância total e a soma de σ^2_m e σ^2_y .

A razão F_m é o quociente entre a variância entre anos e a variância residual. Valores altos de F_m implicam a rejeição da hipótese

nula, reduzindo a possibilidade de um estimador digno de confiança para os fatores sazonais. A hipótese nula consiste na não existência de diferença significativa entre as médias anuais (sazonalidade móvel não significativa).

3.3.9.3 O teste combinado para sazonalidade identificável.

A sazonalidade é chamada de identificável se o erro associado à estimativa do fator sazonal não for suficientemente grande para camuflar o padrão sazonal subjacente. Por exemplo, se houver sazonalidade estável (entre meses) pequena e a maioria do processo for dominada por uma sazonalidade móvel (entre anos) rápida, estas mudanças dificultarão a identificação do fator sazonal pelo X-11 e, portanto, as estimativas não serão precisas.

Este teste combina os dois testes prévios Fs_{D8} e Fm e o teste não paramétrico $X2_{K_W}$ de Kruskal-Wallis.

O objetivo principal do teste é determinar se a sazonalidade da série é identificável ou não. Segundo Lothian e Morry (1978b), o teste consiste basicamente em combinar os valores dos 3 testes anteriores da seguinte forma:

- 1) se o p-valor associado a Fs_{D8} for menor que 0,1%, então H_0 (a sazonalidade não é identificável) será aceita. Se for maior que 0,1, a série passará no teste para sazonalidade estável;
- 2) se a série passar no teste Fs_{D8} , mas o teste Fm para presença de sazonalidade móvel mostrar significância ao nível de 5%, então o valor de Fm será combinado com Fs_{D8} da seguinte forma:

$$T1 = \frac{Z}{Fs_{D8}} ; \quad T2 = \frac{3Em}{Fs_{D8}} ; \quad T = \frac{I1 + I2}{2}$$

segundo Dagum (1978a), $T1 = 7/(Fs_{D8} - Fm)$

Se T for maior ou igual a 1, então H_0 (sazonalidade não identificável) será aceita;

- 3) se Fm mostrar sazonalidade móvel não significativa, ou se T for maior ou igual a 1, então serão verificados os valores de $T1$ e $T2$. Se nenhum deles exceder 1, o programa imprimirá a seguinte mensagem: "sazonalidade identificável provavelmente presente";
- 4) se (iii) passar, mas o teste de $X2_{K_W}$ der um p-valor menor que 0,1%, o programa imprimirá a seguinte mensagem: "sazonalidade identificável provavelmente não presente"; e,
- 5) se houver sazonalidade presente, (iii) passar e o teste de $X2_{K_W}$ indicar a presença de sazonalidade estável, o programa imprimirá: "sazonalidade identificável presente".

3.3.9.4 Teste para sazonalidade residual.

A presença de sazonalidade estável na série ajustada, significa que o procedimento não desempenhou satisfatoriamente sua função de remover a variação sazonal da série. Qualquer padrão anual repetitivo presente na série deve ser eliminado, e não deve aparecer na série sazonalmente ajustada, na tendência, ou no componente irregular.

Este é um teste F aplicado aos valores da série sazonalmente ajustada, e é calculado para todos os valores da série, assim como para os últimos 3 anos. O efeito da tendência é removido pela diferença de ordem 1 de lag 3 ($C=CI(t) - CI(t-3)$). Este teste é de muita importância, pois a presença de sazonalidade residual invalida todo o ajuste sazonal.

3.3.10 Teste para aleatoriedade dos resíduos.

3.3.10.1 O teste da Duração Média de Sequências (ADR).

O X-11 usa o ADR (Average Duration of Run) para testar a aleatoriedade das estimativas finais dos resíduos. Este é um teste não paramétrico desenvolvido por Wallis e Moore (1941), e é construído na base do número de pontos de inflexões. Um ponto de inflexão ocorre quando o sinal da mudança mensal se inverte (Gait, 1975). É eficiente para testar a aleatoriedade dos resíduos versus a hipótese alternativa de que os resíduos seguem um processo AR(1), isto é, $I(t) = \rho I(t-1) + e(t)$, onde ρ é o coeficiente de autocorrelação e $e(t)$ é um processo totalmente aleatório.

Se um processo for puramente aleatório com tamanho infinito, então, a estatística ADR será igual a 1,5. Para séries de 120 observações, o intervalo de confiança de 99% se estende de 1,36 a 1,75 o que implica que não existe correlação. Valores menores que 1,36 indicam autocorrelação negativa e valores maiores que 1,75 indicam autocorrelação positiva.

O teste ADR também é aplicado às estimativas da série sazonalmente ajustada e à tendência. Espera-se valores altos para a tendência.

3.3.10.2 A Função de autocorrelação dos resíduos (FAC).

A função de autocorrelação dos resíduos de um processo estocástico é formada pelas autocorrelações de lag k , e no X-11 ARIMA, k é igual a 14. A função de autocorrelação é a principal ferramenta para a identificação de modelos e para verificação de diagnóstico, isto é, para verificar se os resíduos de um ajuste são aleatórios ou não, a fim de medir a qualidade do ajuste. Uma condição necessária para que um ajuste seja considerado bom é a aleatoriedade dos resíduos, isto é, que os resíduos provenham de um ruído branco, de média igual a zero, variância constante e autocorrelação zero (Box e Jenkins, 1976). O ponto de referência para avaliar a proporção da autocorrelação é o erro padrão, que no caso, supõe-se seja $1/\sqrt{n}$.

3.3.10.3 Teste para periodicidade nos resíduos.

O teste ADR não é eficiente para detectar a existência de componentes periódicos nos resíduos. Para testar independência dos resíduos versus a hipótese alternativa que inclui um processo periódico, o teste do periodograma acumulado normalizado, (Box e Jenkins, 1976), foi incluído no programa.

Se os resíduos forem um ruído branco, o gráfico do periodograma acumulado normalizado $C(f)$ versus a frequência f deverá coincidir, ou estaria espalhado ao redor da reta que passa pelos pontos $(0;0)$ e $(0,5;1)$. Desvios sistemáticos desta reta, indicam presença de periodicidade de período igual ao inverso da frequência f onde acontecem estes desvios. Os desvios são avaliados através do teste de Kolmogorov-Smirnov (Morettin e Toloi, 1981).

Na tabela apresentada para a análise deste teste, mostra-se a estrutura percentual dos desvios que caem nas faixas de confiança de menos de 75%, de 75% a 95% e mais de 95%. Isto implica que se, por exemplo, 10% dos desvios caírem na faixa de mais de 95%, a hipótese de que os resíduos são ruído branco é rejeitada ao nível de significância for 5%. Para afirmar que existe algum componente periódico deve-se verificar que estes desvios sejam sistemáticos.

Em todos os experimentos realizados no presente trabalho, onde a percentagem dos desvios que caem na faixa de mais de 95% supera 5%, estes desvios apresentam um comportamento monótono, fato que permite afirmar que existe evidência da presença significativa de componentes periódicos nos resíduos.

3.3.11 As Estatísticas M e Q de Controle de Qualidade

A tabela F2 da saída do programa do X-11 ARIMA mostra estatísticas com respeito às estimativas dos componentes sazonal, tendência-ciclo e irregular. Estas estatísticas fornecem informações sobre a mudança percentual média de cada componente em diferentes períodos, sobre a contribuição relativa dos componentes às mudanças percentuais na série original, sobre duração média das corridas, etc.

Além disto, o programa fornece onze estatísticas de controle de qualidade (Lothian e Morry, 1978a; Dagum, 1980), que avaliam diversos aspectos da qualidade do ajuste. Estas onze estatísticas denominadas estatísticas M, são combinadas linearmente proporcionando uma estatística chamada estatística Q, que avalia a qualidade global do ajuste.

O valor das estatísticas M e Q flutua de 0 a 3, e só valores menores que 1 são considerados aceitáveis. Para as estatísticas M, valores maiores que 1 indicam falha nos testes, o que não implica necessariamente que o ajuste esteja comprometido. Já no caso da estatística Q, valores maiores que 1 são inaceitáveis, pois comprometem a qualidade do ajuste.

M1: A contribuição relativa do componente irregular no período de 3 meses.

Do ponto de vista do ajuste sazonal, é importante conhecer a proporção da contribuição relativa do componente irregular na variância da mudança percentual da série original em relação à contribuição relativa do fator sazonal. Se a variação do componente irregular for maior quando comparada com a variação do sazonal, os dois componentes não poderão ser separados satisfatoriamente.

Para se obter estacionaridade (sem o que não se terá variância) remove-se a tendência linear da série original, que é estimada através da diferença com períodos de três meses. Lothian e Morry (1978a), chegaram a este resultado após estudos da função de transferência das diferenças de períodos de 1 a 12. A máxima contribuição aceitável do componente irregular à variância global foi estabelecida em 10%. Assim, se o valor de M1 for maior que 1, a contribuição do componente irregular será considerada demasiado alta.

$$M1 = 100 \times \frac{\bar{I}(3)^2}{\bar{O}(3)^2} / 10 ,$$

onde a contribuição do componente irregular no período de 3 meses, $\bar{I}(3)^2$ é a média do quadrado das diferenças $I(t) - I(t-3)$ e $\bar{O}(3)^2$ é a soma das contribuições dos três componentes. A contribuição do fator sazonal e da tendência é calculada da mesma forma que a do componente irregular.

M2: A contribuição relativa do componente irregular na variância da porção estacionária da série.

A medida é similar a M1, mas a tendência linear é estimada e removida através do ajuste de uma linha (ou, se o modelo for multiplicativo, o crescimento exponencial é ajustado e todos os componentes são logaritmicamente transformados). Neste caso, esta estimativa também é removida da tendência. A contribuição de cada um dos componentes é calculada como o quociente entre a variância das estimativas finais dos componentes e a série "estacionarizada". O critério é o mesmo que para M1, portanto a estatística toma a forma de:

$$M2 = 100 \times \text{contribuição do componente irregular} / 10$$

Se M2 for maior que 1 a variação do componente irregular contribuirá demasiado na variação global da série.

M3: A proporção da mudança mensal do irregular comparado com o da tendência.

Para que o X-11 cumpra seu objetivo de extrair o fator sazonal é importante e necessário que a tendência e o componente irregular sejam bem identificados. Se o movimento mensal do componente irregular for dominante na série sazonalmente ajustada (CI) será difícil separar esses dois componentes, e a qualidade global do ajuste sazonal será prejudicada.

A estatística que mede a relação entre o componente irregular e a tendência é a razão $\bar{I}/\bar{C}$, onde $\bar{I}$ e $\bar{C}$ são as médias absolutas das mudanças das estimativas finais da tendência e do componente irregular, respectivamente. Quando a razão $\bar{I}/\bar{C}$ for superior a 3, a proporção do movimento irregular será considerada demasiado alta. A estatística correspondente é a seguinte:

$$M3 = (\bar{I}/\bar{C} - 1)/2$$

Logo, se M3 superar o valor de 1, a série falhará neste teste.

M4: A proporção da autocorrelação do componente irregular descrita como a duração média de sequências (ADR).

Uma das hipóteses básicas da estatística F no método X-11 é que o componente irregular é um processo puramente aleatório com variância constante e autocovariância zero. O programa usa a estatística ADR (ver seção 3.3.10.1) para testar a aleatoriedade das estimativas finais dos resíduos. A estatística M4 usa a fórmula de Bradley baseada na aproximação normal Lothian e Morry (1978a).

$$M4 = \frac{\left| \frac{N-1}{ADR} - \frac{2(N-1)}{3} \right|}{\sqrt{\frac{16 \times N - 22}{90}}} \times \frac{1}{2,58}$$

onde o valor de 2,58 é o valor do limite de 1% da distribuição normal no teste bilateral. Quando M4 for maior que 1 implicará na presença de autocorrelação significativa e a série falhará neste teste.

M5: O número de meses para dominação cíclica (MCD).

Esta estatística mede o número de meses (trimestres) que uma mudança média absoluta da tendência leva para dominar a mudança média do componente irregular (Dagum, 1974). Esta medida é similar a M3, e examina o tamanho relativo da mudança dos componentes irregular e tendência.

Se o MCD for igual a k, então a mudança absoluta na tendência levará k meses para ser maior que a do componente irregular. Esta estatística toma só valores inteiros e para evitar complicações uma nova estatística MCD' é calculada.

A estatística MCD de 6 meses ou mais, tem sido tradicionalmente considerada inaceitável. Assim, a estatística final M5 toma a forma:

$$M5 = (MCD' - 0,5) / 5,0$$

Portanto valores maiores que 1 são rejeitados.

M6: A proporção da mudança anual do componente irregular em relação à mudança anual do fator sazonal.

O X-11 aplica o filtro 3x5 à razão SI (diferença) a fim de separar o fator sazonal do componente irregular. A experiência tem mostrado que quando o irregular é demasiado pequeno comparado com a mudança anual do sazonal, descrito por um valor baixo de $\bar{I}/\bar{S}$, a média móvel 3x5 não é suficientemente flexível para acompanhar o movimento sazonal. De outra maneira, quando a razão é alta o filtro evidencia demasiada flexibilidade e, como resultado, os fatores sazonais são contaminados pelo movimento irregular (Dagum, 1978a).

Estudos com 421 séries com sazonalidade presente ajustadas pelo X-11, indicaram que quando $\bar{I}/\bar{S}$ cai entre 1,5 e 6,5, a média móvel (3x5) trabalha relativamente bem. Fora deste intervalo, para separar os dois componentes corretamente, é necessário o uso de um filtro sazonal curto (para valores de $\bar{I}/\bar{S}$ demasiado pequenos) ou do uso de um filtro comprido (para razões altas), (Lothian e Morry, 1978a).

Das 421 séries, 2% tiveram valores abaixo de 1,5 e 2% acima de 6,5. A medida M6 é baseada nos pontos de corte 1,5 e 6,5 e é formulada da seguinte forma:

$$M6 = 1 (\bar{I}/\bar{S} - 4,0) / 2,5$$

Se este valor exceder 1 a série falhará para este teste, mas o problema poderá ser remediado ajustando a série com uma média móvel 3x1, se a razão $\bar{I}/\bar{S}$ for menor que 1,5 ou usando sazonalidade estável se a razão for maior que 6,5.

M7: A proporção da sazonalidade estável em relação a sazonalidade móvel.

Como se viu na seção 3.3.9 o teste Fs e o teste Fm são combinados para detectar presença de sazonalidade identificável. Aqui, a sazonalidade será chamada de identificável se o valor absoluto do erro (ou distorção) introduzido na estimativa final do fator sazonal não for demasiado alto.

Tem-se que a distorção depende de ambos os valores: Fs e Fm. Valores pequenos de Fs sugerem alta distorção, enquanto valores altos de Fm indicam que distorção adicional foi introduzida devido ao movimento. O ponto de corte foi baseado nas séries de 10 anos e na correspondente combinação dos valores Fs e Fm que indica 50% de distorção na estimativa final do fator sazonal. Assim a estatística toma a seguinte forma:

$$M7 = \sqrt{\frac{1}{2} \left[\frac{7}{Fs} - \frac{3Fm}{Fs} \right]}$$

Maiores detalhes podem ser encontrados em Lothian e Morry (1978a). Valores maiores que 1 indicam que a sazonalidade da série não é identificável.

As estatísticas de controle de qualidade que descrevem o movimento ano a ano do componente sazonal.

Os filtros para o fator sazonal do X-11 são eficientes somente para sazonalidade constante nos primeiros e últimos 3 anos da série, enquanto que nos anos intermediários estes filtros podem reproduzir uma linha ou uma constante. Deste modo, somente um componente sazonal constante pode ser estimado adequadamente para a série completa. Se a sazonalidade contiver movimentos ano a ano, a estimativa do fator sazonal terá erros consideráveis.

Distinguem-se dois tipos de movimentos: um que exhibe flutuações quase aleatórias e outro onde as mudanças aparecem na mesma direção ao longo dos anos. O tamanho do primeiro tipo de movimento pode ser medido através da média dos valores absolutos das mudanças ano a ano do fator sazonal, enquanto a média aritmética das mudanças proporciona uma indicação do tamanho do movimento sistemático. Flutuações aleatórias são medidas pelas estatísticas M8 e M10. As estatísticas M9 e M11 descrevem o tamanho do movimento linear. M8 e M9 são calculadas usando todos os valores da estimativa final do componente sazonal e M10 e M11 são calculadas para os últimos anos. Estas duas últimas estatísticas foram introduzidas porque geralmente os usuários estão interessados nos movimentos mais recentes da série.

Para evitar problemas de unidades, as estimativas do fator sazonal (S) são padronizadas (S').

M8: O tamanho das flutuações do componente sazonal ao longo da série inteira.

Como mencionado antes, as flutuações são medidas através da mudança média absoluta.

$$|\overline{\Delta S'}| = \sum_{j=1}^J \sum_{i=2}^N |S'_{ji+j} - S'_{j(i-1)+j}| / J*(N-1),$$

onde N é o número de anos e J é 12 ou 4 (para séries mensais ou trimestrais). A mudança máxima aceitável foi colocada em 10%. Assim a estatística M8 toma a seguinte forma:

$$M8 = 100 * |\overline{\Delta S'}| / 10$$

M9: A média do movimento linear no componente sazonal através da série inteira.

Medidas da quantidade do movimento sistemático são obtidas calculando a média das mudanças ano a ano para cada mês. Se houver flutuações somente aleatórias de ano a ano esta média estará perto de zero. Se a maioria das mudanças for na mesma direção por mês a média estará perto da média sistemática da mudança.

Usando a fórmula :

$$\sum_{i=1}^{N-1} (\Delta S'_{J(i-1)+J}) = S'_{J(i-1)+J} - S'_{J(i-1)+J}$$

Colocando o limite de aceitação de 10% obtém-se:

$$M9 = 100 * \frac{\sum_{j=1}^J |S'_{J(i-1)+J} - S'_{J(i-1)+J}|}{J * (N-1)} * \frac{1}{10}$$

M10: O tamanho das flutuações do componente sazonal nos anos recentes.

Esta estatística é equivalente a M8 com a exceção de que só os anos N-2, N-3, N-4 e N-5 são envolvidos nos cálculos.

M11: A média do movimento linear no componente sazonal nos anos recentes.

Esta medida corresponde a M9 usando os dados dos anos N-2, N-3, N-4 e N-5.

Quando M11 excede 1, há forte indicação de que os fatores sazonais para os anos recentes são altamente distorcidos devido ao efeito de achatamento dos pesos finais sobre os movimentos lineares.

Q: A estatística de controle de qualidade global do ajuste.

A estatística Q avalia a qualidade global do ajuste e é uma soma ponderada das 11 estatísticas estudadas anteriormente. Os pesos assinalados a cada uma delas dependem de sua importância e sua participação na qualidade global do ajuste.

É possível que a série falhe nas estatísticas M1 ou M2, mas o ajuste não é necessariamente mal. Estas duas estatísticas medem a variação do componente irregular em relação à variação sazonal. A série média ajustada tem um ciclo que contribui com cerca de 5% a 10% na variância da porção estacionária da série. O nível de entrada para as estatísticas M1 e M2 é baseado nesta suposição. Se a série não contiver ciclo, o componente irregular contribuirá com 13% a 14% na variação total (resultando que M1 e M2 excedem 1) e ainda será aceitável. Similarmente, se o ciclo contribuir com mais que 10% o nível de entrada deverá ser abaixado.

Se a série tiver uma tendência achatada será possível que exista uma razão $\bar{I}/\bar{C}$ excedendo 3, e deste modo o teste M3 falhará sem prejudicar a qualidade do ajuste. Atualmente o programa X-11 compensa a falta de ajuste da tendência aplicando o filtro 23-H para estimar a tendência. Se o principal objetivo for a análise dos ciclos econômicos, valores altos de M3 constituirão sinal de sérios problemas. Isto

indica que a série ajustada sazonalmente contém uma proporção muito alta do componente irregular, o que alerta ao usuário sobre a necessidade de uma identificação apropriada da tendência.

Uma autocorrelação significativa no componente irregular indicado por um valor de M4 maior que 1, pode significar, por exemplo, que o usuário terá que aplicar regressão de "trading days" e assim o ajuste não será válido. Ao mesmo tempo, é possível que o componente irregular original seja autocorrelacionado devido ao desenho amostral. Isto não afetará o ajuste sazonal X-11 que é baseado nas características reconhecidas do comportamento sazonal e da tendência na série e obtém o componente irregular como resíduo do procedimento. Assim o fator sazonal pode ser bem identificado. Foi observado que o valor de M4 se comporta independentemente das outras estatísticas. Frequentemente é a única que falha ou, contrariamente, a única que não falha e, portanto, lhe foi assinalado peso mínimo.

No caso de M5, o que foi dito relativo a M3 se aplica aqui. É possível que a irregularidade seja demasiadamente alta, mas é também possível que a série contenha uma tendência constante, o que não impede o X-11 de isolar o movimento sazonal corretamente.

Como foi visto, M6 é a única estatística onde a falha pode ser corrigida. O usuário é aconselhado a usar de novo o X-11 e aplicar a média móvel sazonal apropriada à série SI a fim de aumentar a qualidade do ajuste sazonal.

Se uma série qualquer falhar na estatística M7 é porque não tem sazonalidade, ou as estimativas são distorcidas, ou o componente sazonal não é identificável, como é indicado na mensagem na saída do programa. Esta medida é uma das mais importantes no conjunto das estatísticas de controle de qualidade e lhe é, portanto, assinalado o maior peso. Quando M7 falha o usuário é aconselhado a não ajustar a série, embora possa haver exceções. É possível que devido ao uso da opção aditiva no ajuste de uma série onde os componentes são relacionados multiplicativamente, caso em que a tendência cresce rapidamente, o valor Fm do teste de sazonalidade móvel seja muito alto. Isto pode resultar num valor de M7 que excede 1.

Se o ajuste for feito novamente com a opção multiplicativa, o valor de Fm será reduzido significativamente e M7 provavelmente passará pelas especificações.

A falha das estatísticas M8 e M10 pode não ser crucial, como por exemplo, no caso em que M9 e M11 passam pelas especificações (sendo seus valores menores que 1) e o usuário está preocupado somente com o vício nas estimativas do componente sazonal. Similarmente, se M9 e M11 excederem 1, mas o usuário estiver somente interessado nos fatores sazonais históricos, estas estimativas poderão ainda ser necessárias porque os pesos centrais da média móvel sazonal podem seguir qualquer movimento linear. Não obstante, se alguém estiver interessado nos fatores sazonais atuais, valores altos de M9 e M11 indicarão a presença de distorção significativa nas estimativas.

Da discussão anterior percebe-se que nenhuma estatística isoladamente pode avaliar a qualidade do ajuste. Se as 11 estatísticas M falharem o ajuste será inaceitável. Mas o que acontecerá se alguma falhar e as outras não? Uma estatística de controle de qualidade global foi desenvolvida para este caso. É a estatística Q, que consiste na soma ponderada das 11 estatísticas M. A cada estatística M foi conferido um peso de acordo com a sua importância relativa na qualidade global do ajuste. Os pesos associados às 11 estatísticas aparecem na seguinte tabela. Maiores detalhes podem ser encontrados em Lothian e Morry (1978a).

Estatística M e peso assinalado

M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11
13	13	10	5	11	10	16	7	7	4	4

Esta combinação das 11 estatísticas foi bem sucedida na avaliação da qualidade do ajuste das 421 séries (cujos comprimentos variam de 5 a 30 anos), testadas pelos autores. Estas séries foram ajustadas no Statistics Canada com o programa X-11 ARIMA. Os valores médios para as 11 estatísticas e a estatística Q para as 421 séries são apresentadas na tabela seguinte.

Valores médios para as estatísticas

M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	Q
.719	.605	.485	.424	.593	.380	.403	.640	.394	.794	.684	.529

Se a estatística Q for maior que 1, o ajuste da série será declarado inaceitável. O ajuste também será rejeitado se falhar o teste de sazonalidade identificável. Para as séries mensais, 8,4% tem valores maiores que 1. Outras 3,7% foram rejeitadas porque não passaram pelo teste de sazonalidade identificável. 12,1% dos 421 ajustes sazonais foram rejeitados. As 51 séries que falharam foram examinadas em detalhe pelos autores e para todas elas, os ajustes foram julgados inaceitáveis.

As estatísticas de controle de qualidade apresentadas aqui permitem avaliar a qualidade de um grande número de séries. Mais do que isto, permitem ainda que esta avaliação possa ser feita por pessoas que não dominem profundamente os métodos de ajuste sazonal. A estatística Q proporciona a avaliação geral da qualidade do ajuste, mas os usuários devem ser cautelosos em atribuir o nível de significância a pequenas mudanças na estatística.

3.4 Descrição do pacote X-11 ARIMA.

Opções disponíveis no programa.

- 1) dispõe do programa X-11-Q para séries trimestrais;
- 2) tipo de ajuste: aditivo, multiplicativo e logarítmico;
- 3) tipo de programa: sazonal ou resumo de medidas;
- 4) ajuste prévio da tendência por greves;
- 5) rotina para estimar a variação do calendário;
- 6) tratamento de valores extremos;
- 7) rotina variável para a tendência-ciclo; e,
- 8) rotina variável para o fator sazonal.

Parte A. Ajustes prévios (opcional)

1. O ajuste prévio é feito para corrigir mudanças no nível da série, efeito de certos feriados, etc.
2. Ajuste prévio por variação do calendário. O usuário pode substituir os pesos para os 7 dias.

Parte B. Estimação preliminar dos pesos e a variação do calendário.

Estimativas preliminares dos pesos e fatores para reduzir o efeito dos extremos ou de valores quase-extremos são desenvolvidos a partir dos dados. Estas estimativas são redefinidas na parte C.

1. Estimativa preliminar da tendência e da razão SI.
2. Substituição de valores extremos.
3. Estimativa do fator sazonal S, da série sazonalmente ajustada e da tendência.
4. Ajuste da tendência-ciclo por greves(opcional).
 - a) Substituição de valores extremos.
 - b) Estimação do fator sazonal, da série sazonalmente ajustada, do componente irregular.
5. Ajuste por variação do calendário (opcional).
 - a) Substituição de valores extremos pela regressão da variação do calendário.
 - b) Regressão preliminar da variação do calendário.
 - c) Fatores de variação do calendário derivados dos coeficientes de regressão.
 - d) Pesos preliminares para o componente irregular.
 - e) fatores de variação do calendário combinados com os pesos diários.
6. Série ajustada por variações prévias e por variação do calendário.

Parte C. Estimação final dos pesos do componente irregular da variação do calendário.

1. Série original modificada pelos pesos preliminares e ajustada pela variação do calendário e variação prévia.
2. Estimativa da tendência-ciclo, da razão SI, do fator sazonal e da série sazonalmente ajustada.
3. Nova estimativa da tendência-ciclo, razão SI, do fator sazonal, da série sazonalmente ajustada e da série do componente irregular.
4. Ajuste pela variação do calendário, a fim de obter melhores estimativas.
 - a) Substituição de valores extremos pela regressão da va-

riação do calendário.

b) Regressão final da variação do calendário.

c) Fatores finais da variação do calendário da regressão de variação do calendário.

d) Pesos finais para o componente irregular.

e) Fatores finais da variação do calendário dos pesos diários combinados.

5. Série original ajustada pela variação do calendário e pela variação prévia.

Parte D. Estimação final do fator sazonal, da tendência-ciclo, do componente irregular e da série sazonalmente ajustada.

1. Estimativa da tendência-ciclo, da razões SI, do fator sazonal e da série sazonalmente ajustada.

2. Nova estimativa da tendência.

3. Estimativa final das razões SI, substituição final dos valores extremos, estimativa final dos fatores sazonais, série final sazonalmente ajustada, estimativa final da tendência e do componente irregular.

Parte E. Série original modificada, sazonalmente ajustada e do componente irregular.

A série original, sazonalmente ajustada e a do componente irregular são modificadas por extremos (além de 2.5-sigma). Tabelas de controle de qualidade, razões do total anual, mudanças percentuais na série original e na série sazonalmente ajustada, são apresentadas aqui.

Parte F. Resumo de medidas e a média MCD.

1. Média móvel MCD.

2. Resumo das medidas: médias das mudanças percentuais.

Capítulo IV. METODOLOGIA.

4.1 Introdução

O objetivo principal do trabalho diz respeito à verificação do comportamento dos métodos do X-11 e do X-11 ARIMA para séries sócio-econômicas brasileiras, que costumam apresentar fortes mudanças estruturais de tendência e sazonalidade. Para tanto tomou-se como exemplo a série de mortalidade infantil do Estado de São Paulo, que apresenta estas características de forma marcante. Como subproduto da análise é mostrado também como o componente sazonal pode trazer informações e interpretações interessantes.

Esta dissertação trata especificamente da utilização da opção automática do método X-11 ARIMA. Existem vários aspectos que devem ser levados em consideração na análise da qualidade do ajuste conseguido, mas três pontos merecem destaque na tese: revisões, velocidade de detecção de pontos de inflexão e as próprias estatísticas dadas no programa X-11 ARIMA.

A metodologia propriamente dita utilizada para se conseguir este objetivo é apresentada na seção 4.2, a opção automática definida na seção 4.3, e a discussão da avaliação da qualidade do ajuste é realizada na seção 4.4.

4.2. A Metodologia

Tendo-se a série histórica da Mortalidade Infantil no Estado de São Paulo no período de 1933 a 1985 é possível verificar com elevada precisão os pontos onde existem fortes mudanças de tendência e sazonalidade. O objetivo da dissertação é verificar o comportamento do método exatamente nestes pontos.

Para exemplificar a metodologia suponha que tenha havido fortes mudanças no período T . Serão então realizados ajustes utilizando-se dados até o ano $(T+t1)$, $(T+t2)$, ..., $(T+ti)$. Desta forma poder-se-á avaliar a qualidade do ajuste automático em relação a revisões, rapidez na detecção de pontos de inflexão e em relação às próprias estatísticas de controle do método X-11, nos períodos onde ocorrem fortes mudanças.

Uma vez definida a opção automática, ela é aplicada para avaliar o seu desempenho. Através da análise da qualidade do ajuste desta opção são propostas mudanças, principalmente nos filtros a fim de se conseguir uma "melhor opção". A introdução das modificações na opção automática é realizada gradualmente, a fim de detectar as consequências de cada modificação. Após a avaliação do desempenho da opção automática e das opções alternativas, é escolhida a "melhor opção". Esta melhor opção é definida como o conjunto de escolha de opções do programa que fornece o ajuste com melhor desempenho, com respeito as estatísticas de controle de qualidade. A seguir é realizado um estudo comparativo entre a opção automática e a "melhor opção". Desta forma é possível estudar a opção automática, ao mesmo tempo em que é verificada a potencialidade das opções existentes no programa X-11 ARIMA.

Os cortes da série são feitos basicamente nos anos 1941, 1944, e 1968. Isto porque os dois primeiros anos contém quedas bruscas e no último porque é um ano controverso. Além disto, são realizados ajustes em outros períodos envolvendo características contrastantes, por exemplo, o período de 1950 a 1965 que inclui a década de 50 caracterizada com uma queda rápida na tendência e na amplitude sazonal e parte da década de 60 caracterizada pela constância em ambos componentes. A aplicação é mostrada no Capítulo V.

Para o estudo de revisões é necessário definir as estimativas da tendência e da sazonalidade que são consideradas históricas ou finais, a fim de estudar as revisões. Dada à falta de informação para 1949, à limitação do X-11 que permite ajustar séries de, no máximo 30 anos, e ao próprio fato das revisões caírem praticamente para zero depois da inclusão de 3 anos, considerou-se como ajuste final todo ajuste realizado em pontos que estivessem no mínimo 3 anos e meio distantes das extremidades.

Para avaliar a qualidade do ajuste é necessário definir critérios, estatísticas e testes, assim como definir o que é um bom ajuste. A definição dos critérios é baseada principalmente nos trabalhos de Shiskin e Plewes (1978) e Dagum e Morry (1983). A análise destes critérios é realizada através das estatísticas de testes implementados no próprio programa (v. Dagum, 1978a, 1980; Lothian e Morry, 1978a). O estudo das revisões é baseado no artigo de Dagum e Morry (1983). O problema de revisão tem sido citado em todo o trabalho, principalmente na seção 2.5.2. Os testes estatísticos dados no programa X-11 ARIMA para avaliar a qualidade do ajuste são revistos novamente na seção 4.4.

4.3. A Opção Automática

A maioria dos trabalhos envolvendo análise do desempenho do X-11, tomam como referência a opção padrão (aditiva ou multiplicativa). As principais características da opção padrão do X-11 e do X-11 ARIMA são: escolha automática do filtro para a estimativa final da tendência; uso do filtro 3x3 para a estimativa preliminar e uso do filtro 3x5 para a estimativa final do fator sazonal; limites inferior e superior respectivamente iguais a 1,5 e 2,5 para o desvio padrão móvel para a eliminação de valores extremos. Deve-se notar que a opção padrão do X-11 ARIMA não inclui a opção automática ARIMA (ajuste automático dos modelos ARIMA).

Neste trabalho a opção automática é aquela em que o programa tem a mais ampla liberdade de escolha entre as opções proporcionadas por este. Diferentemente da opção padrão do X-11 é dado ao programa a liberdade de escolha do filtro para a estimativa final do fator sazonal em cada mês separadamente. Além disto, é utilizada a opção ARIMA com substituição de valores extremos.

Estas modificações foram introduzidas por duas razões: a primeira porque se houver um mês com sazonalidade muito variável ou muito suave o programa automaticamente escolherá um filtro mais comprido ou mais curto, o que não acontece com a opção padrão. Este fato torna a opção automática mais flexível que a opção padrão. A segunda

razão refere-se à opção ARIMA com substituição de valores extremos, pelas vantagens já mencionadas (v. seção 3.3.2). Quando a opção ARIMA é rejeitada é denominada "opção automática X-11", quando é aceita, "opção automática X-11 ARIMA".

Deve-se notar que a opção automática assume um modelo multiplicativo. Isto porque existe uma grande relação entre a tendência e a sazonalidade da série da taxa de mortalidade infantil, como será visto na seção 5.3.

As principais características da opção automática são :

Tipo de modelo	: multiplicativo
Valores extremos	: limites sigma entre 1,5 e 2,5
Filtro sazonal	: usa o filtro 3x3 para estimativa inicial do fator sazonal, e para a estimativa final escolhe o filtro sazonal para cada mês separadamente, segundo a razão $\bar{I}/\bar{S}$.
Filtro para tendência:	escolhe o filtro para estimativa final da tendência, segundo a razão $\bar{I}/\bar{C}$.
Opção ARIMA	: com substituição de valores extremos pelos valores ajustados pelo modelo.

Notação e descrição das diversas opções.

1) Opção auto	: opção automática
2) Opção 3x3	: similar à opção automática. A diferença consiste em que o filtro para a estimativa do fator sazonal é o 3x3, para todos os meses em todas as iterações. Igualmente poderiam ser usados os filtros 3x5, ou 3x9; neste caso a opção seria chamada de 3x5 ou 3x9 respectivamente.
3) Opção 9-H	: Diferentemente da opção automática, esta opção usa o filtro 9-H para a estimativa final da tendência. De forma similar poderia ser usado o filtro 23-H; neste caso, a opção seria chamada de opção 23-H. Não foi usado o filtro 13-H porque este é escolhido automaticamente pelo programa.
4) Opção ARIMA (p,d,q)(P,D,Q)s	: similar à opção automática, mas o usuário escolhe o modelo ARIMA com substituição de valores extremos. Esta opção é usada quando a opção automática rejeita os modelos ARIMA.
5) Opção s/out	: inclui a manipulação de valores extremos, os sigma limites são ± 1 e ± 2 .
6) Opção aditiva	: similar à opção automática mas o tipo de modelo é aditivo.
7) Opção X-11	: similar à opção automática sem a opção ARIMA. Esta opção é utilizada quando a opção automática aceita algum modelo ARIMA. Isto é realizado para avaliar o desempenho das duas versões.
8) Opção 24-termos	: A estimativa inicial da tendência é feita através do filtro 24-termos. Neste caso a série tem que ser ajustada por um modelo ARIMA a fim de se contar com a série estendida.

As outras opções são combinações das opções mais simples antes descritas. Por exemplo, a opção 3x3,9-H, significa que o filtro para o fator sazonal é o 3x3 para todas as iterações e todos os meses, e, o filtro para a estimativa final da tendência é o 9-H. Esta opção, por usar os filtros mais curtos, é chamada também de opção dos filtros curtos. As opções mais usadas são a opção 3x3,9-H, a opção 3x5,9-H, a opção 3x9,9-H. Algumas das opções mais utilizadas usam a manipulação de valores extremos; nesse caso a opção é chamada, por exemplo de opção 3x9,9-H,s/out.

Outra notação utilizada no presente trabalho é a da caracterização da série num certo período, por exemplo, a série no período de 1933 a 1948 é chamada de TX3348, onde os dois primeiros dígitos indicam o ano inicial do período e os dois últimos o ano final.

4.4 Critérios para avaliar a qualidade do ajuste.

Os critérios para avaliar a qualidade do ajuste estão baseados naqueles usados por Shiskin e Plewes (1978), Dagum e Morry (1983), e naqueles que o próprio X-11 ARIMA oferece (Dagum, 1980; Lothian e Morry, 1978a). São eles:

- 1) Presença de sazonalidade: Estável, móvel, identificável e residual, ver seção 3.3.9.
- 2) Adequabilidade do modelo: multiplicativo ou aditivo.
- 3) Aleatoriedade dos resíduos: presença de autocorrelação e periodicidade, ver seção 3.3.10.
- 4) Participação do componente irregular em relação aos outros componentes, ver seção 3.3.10.
- 5) Efeito das revisões, ver seção 2.5.2
- 6) Grau de suavidade da série sazonalmente ajustada, ver seção 3.3.11.

Será apresentado aqui apenas um resumo destes critérios, maiores detalhes podem ser encontrados nas seções mencionadas.

1) Sazonalidade.

- a) Sazonalidade estável é uma característica desejável, medida através de três testes: O teste F de um fator, aplicado às estimativas preliminar e final das razões "sazonal-irregular" SI (ou diferenças), e o teste não paramétrico de Kruskal-Wallis aplicado a estas últimas.
- b) Sazonalidade móvel, característica a ser evitada, cuja presença é medida através do teste F de dois fatores, aplicado às estimativas finais das razões SI.
- c) Sazonalidade identificável, que mede a relação entre a sazonalidade estável e móvel, Lothian e Morry (1978b).
- d) Sazonalidade residual, mede se o método cumpre sua tarefa principal, a de retirar o componente sazonal da série. É avaliado através de um teste F de um fator aplicado à série sazonalmente ajustada.

UNICAMP
BIBLIOTECA CENTRAL

- 2) Para avaliar a adequabilidade do modelo aplica-se o teste Bell Canada Model (BCM), que é um programa desenvolvido por Statistics Canada (Higginson, 1977). Este teste consiste na estimação da tendência-ciclo através da média centrada 12-termos e na construção de uma parte sistemática da série por adição ou multiplicação. As duas partes sistemáticas são então removidas da série original, e, qualquer modelo para o qual a soma de quadrados dos resíduos for significativamente menor será considerado melhor (Shiskin e Plewes, 1978).

O problema da identificação do modelo é mais importante quando é focado do ponto de vista sazonal dos anos mais recentes, particularmente de séries com tendências apresentando um rápido aumento e/ou diminuição.

Segundo Dagum (1976), no modelo aditivo, os componentes da série são considerados como sendo independentes e, portanto, o efeito sazonal não é afetado pelo nível da atividade econômica condicionada pelos estágios dos ciclos econômicos. Pelo contrário, no modelo multiplicativo estes fatores estão altamente relacionados. No caso da mortalidade infantil a influência da sazonalidade, determinada principalmente pelas estações, está altamente relacionada com o nível da qualidade de vida, que por sua vez depende da evolução sócio-econômica, do avanço tecnológico, entre outros fatores. Portanto a opção multiplicativa parece ser empiricamente mais razoável. Isto é confirmado para todos os períodos de estudo da série, pela aplicação do teste BCM, embora a opção aditiva também possa ser usada. Lothian (1978) apresenta as relações existentes entre os modelos e suas restrições.

$$\begin{array}{ll} \text{Multiplicativo:} & Ot = St * Ct * It \\ \text{Aditivo} & : Ot = St + Ct + It \\ \text{Logarítmico} & : \ln Ot = \ln St + \ln Ct + \ln It \end{array}$$

- 3) Aleatoriedade dos resíduos.
- a) A presença de autocorrelação é medida por: o teste ADR, a estatística de controle de qualidade M4 que está baseada neste e a função de autocorrelação dos resíduos.
 - b) A presença de periodicidade é medida através do teste de Kolmogorov-Smirnov.
- 4) A participação do componente irregular em relação aos outros componentes é medida através de: i) participação do componente irregular na variância da porção estacionária da série original; ii) estatísticas de controle de qualidade M1, M2 e M3, que medem a participação do componente irregular em relação à participação dos componentes sazonal e tendência.

A participação dos componentes é apresentados numa tabela chamada de "Estrutura da variância da porção estacionária da série original", e é calculada como o quociente entre a variância das mudanças percentuais do componente e a variância da série estacionarizada; esta última é a soma das variâncias dos componentes (v. estatísticas M1 e M2 na seção 3.3.11).

- 5) As revisões são medidas através do erro médio absoluto MAE (Dagum e Morry, 1983), que consiste na média dos valores absolutos das diferenças entre as estimativas correntes e as estimativas históricas dos componentes.
- 6) O grau de suavidade da série ajustada é medida pelo número de meses gastos para que a tendência domine o componente irregular (MCD) (Dagum, 1974) e através da estatística M5, que é função do MCD.

Uma restrição prática é a necessidade de que a série ajustada exiba um movimento suave sobre o tempo. A falta de suavidade implica em mudanças frequentes de direção da série sazonalmente ajustada, devidas principalmente às mudanças do componente irregular. Essas mudanças inquietam e diminuem a confiança e a credibilidade do público (Shiskin e Plewes, 1978).

A suavidade também pode ser medida na base do número de pontos de inflexão na série sazonalmente ajustada e no seus componentes. O X-11 aplica o teste ADR para a série sazonalmente ajustada, para o componente irregular, para a tendência e apresenta os valores destes, além de apresentar os valores do MCD e o número de meses para a dominação cíclica.

Além destas estatísticas, tem-se as estatísticas de controle de qualidade M6-M11, que avaliam outros aspectos da qualidade do ajuste (seção 3.3.11).

"Bom ajuste"

Após a discussão apresentada até aqui sobre as qualidades desejáveis que todo ajuste deve possuir, quais sejam, os critérios para avaliar estas qualidades, as estatísticas, os limites de especificação, etc., decidiu-se que, para o presente trabalho, um bom ajuste deve cumprir, no mínimo, as seguintes especificações:

- i) Aleatoriedade dos resíduos: ausência de periodicidade e autocorrelação.
- ii) Suavidade da série: MCD menor ou igual a 4.
- iii) Nenhuma estatística M deve falhar.
- iv) A participação do componente irregular deve ser menor que 10%.

"Melhor opção"

é aquela que fornece o melhor ajuste, mesmo que este ajuste às vezes, não seja bom, pois, como será visto no capítulo V, nem sempre existe uma opção que proporcione um ajuste livre de deficiências.

Capítulo V. APLICAÇÃO.

5.1 Introdução.

A taxa de mortalidade infantil é um dos principais índices de qualidade de vida de uma população. Esta taxa reflete, através de sua sazonalidade, o grau de vulnerabilidade da saúde da população às condições climáticas, e, através de sua tendência a longo prazo, a evolução geral das condições sócio-econômicas.

Esta série foi escolhida porque sua estrutura, ao longo de todo o período em estudo, apresenta características contrastantes na sazonalidade e na tendência. Dependendo do período a sazonalidade, pode se apresentar fortemente estável, ou completamente móvel (amplitude) ou defasada, ou sem predominância de um padrão sazonal, ou ainda com padrão invertido em alguns anos. A tendência possui um comportamento menos variável, mas mesmo assim, apresenta quedas bruscas. Estas características tornam esta série adequada para avaliar o desempenho do X-11.

5.2 Definição da série.

A taxa de mortalidade num dado período é calculada como o quociente entre o número de óbitos de menores de um ano e o número de nascidos vivos multiplicado por mil. Os dados foram obtidos da Fundação Sistema Estadual de Análise de Dados (SEADE), que possui as séries mensais do número de nascidos vivos e do número de óbitos de menores de um ano do Estado de São Paulo.

Segundo Ferreira e Ortiz (1982) a série de nascidos vivos, até antes de 1960 incluía sistematicamente os nascidos fora de prazo (o nascido fora de prazo é aquele que é registrado num ano posterior ao do nascimento). A partir de 1961, o SEADE retira estes nascimentos e apresenta duas séries de nascidos vivos: uma que inclui os fora de prazo e a outra não, embora, só a partir de 1975 estes dados sejam apresentados mensalmente. O ideal seria que os fora de prazo fossem incluídos na época certa, pois esta série é a real e embora o SEADE possua fatores de correção para uniformizar a série, não foi possível usá-la, pois esta informação não é disponível mensalmente.

A série de óbitos de menores de um ano, no presente trabalho, é aquela que registra os óbitos por ocorrência (os que ocorreram no Estado de São Paulo), isto porque é a única apresentada de forma mensal completa. A outra série é por residência, mas esta série não é completa, existindo apenas a partir de 1971. As diferenças entre estas duas séries, segundo o caso, podem ser ignoradas (Giraldelli, 1982). A série por ocorrência é susceptível ao efeito de evasão e, principalmente de invasão de óbitos, devido ao fato do Estado de São Paulo possuir a infraestrutura hospitalar mais completa do país (Margone Camargo e Cunha, 1982).

Muitos problemas foram encontrados nos dados. Para a série de óbitos de menores de um ano, não existe informação (por ocorrência) de 1974, embora o SEADE conte com estimativas preliminares, as quais foram usadas. Isto, infelizmente, não acontece com a série de nascidos

vivos, que não tem informação para o ano de 1949, motivo pelo qual a série fica interrompida neste ano. Também se deve levar em consideração que, ao longo do período, as definições mudam, o sistema de captação dos dados melhora sensivelmente, sobretudo na série de nascidos vivos, o que pode afetar a taxa (Altmann, 1982). Estes fatores devem ser levados em consideração ao se fazer a análise.

São Paulo é o Estado que possui as estatísticas vitais mais confiáveis e completas no país. O grau de integridade das estatísticas de óbitos infantis do Estado de São Paulo até meados deste século, encontrava-se por volta de 80%. A partir de 1950 o registro é praticamente completo atingindo 99,5% em 1976 (Ortiz, 1982b). O mesmo não acontece com o registro dos nascidos vivos, cujo grau de cobertura atinge a 91% em 1973 (Altman e Ferreira, 1982).

5.3 Descrição da série.

5.3.1 Descrição da taxa anual.

A série formada pela taxa anual de mortalidade infantil (v. gráfico 1 e anexo 7) apresenta uma tendência quase constante na década de 30, situando-se em níveis elevados, acima de 166 por mil. A partir de 1942, quando possui uma taxa de 144,90 por mil, observa-se uma tendência decrescente bem acentuada até 1945, quando atinge o valor de 72,58 por mil, reduzindo a taxa anual em quase 50%, em relação ao início deste período. A partir de 1966 a tendência começa a registrar um incremento quase constante até 1975, (neste período a taxa média é de 82,68 por mil) para logo retomar a tendência decrescente, atingindo o mínimo em 1985 com uma taxa de 36,53 por mil, isto é, reduzindo novamente em quase 50% com respeito ao período anterior.

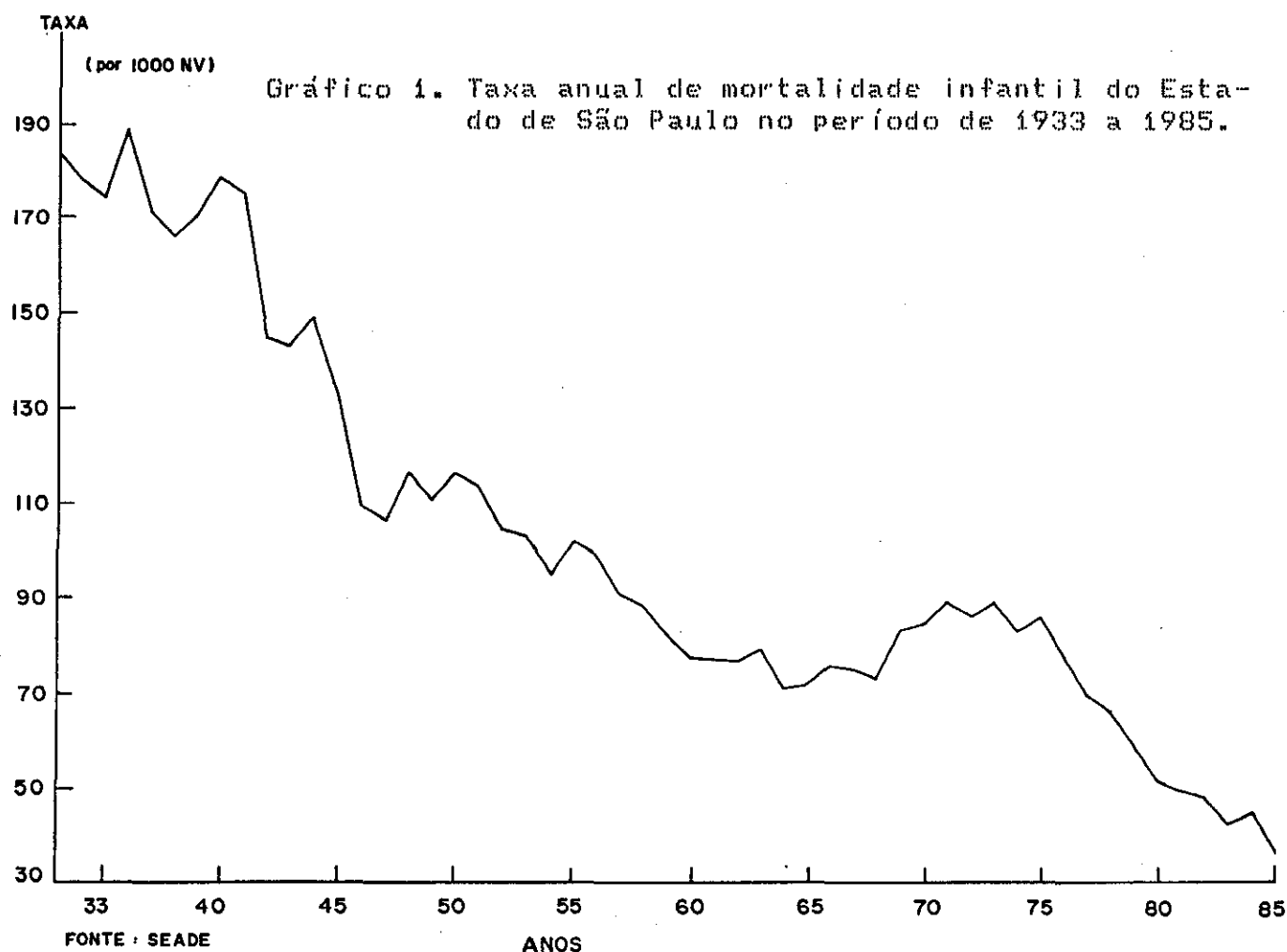
Segundo Ortiz (1982a) as altas taxas dos primeiros anos do período se devem a doenças infecciosas e parasitárias, comumente associadas à desnutrição. Esta situação, segundo ele, provavelmente foi causada por problemas sanitários e acesso restrito à assistência médica, além das profundas consequências da grande crise de 1929/1930, que envolveu todas as áreas econômicas, especialmente as atividades relacionadas com o café.

Nos anos seguintes, que vão de 1940 até 1964, começam algumas mudanças sócio-econômicas, derivadas da evolução industrial que se registra no Estado, sendo as mais significativas nas áreas de saneamento, nutrição, habitação, medicina preventiva e curativa, além dos avanços técnicos na medicina (quimioterapia de sulfa e antibióticos, imunização mais eficaz, etc.) que conseguem atenuar as oscilações da mortalidade devidas a epidemias. Como consequência destas transformações, a mortalidade infantil apresenta uma queda bem acentuada. O trabalho de Monteiro (1982) mostra que o principal fator para esta queda é o crescimento do salário mínimo real.

Na fase seguinte, de 1965 a 1975, a mortalidade infantil aumenta. Este fenômeno já foi motivo de muitas controvérsias. Diversas hipóteses têm sido formuladas, sendo a maioria delas ligada à conjuntura sócio-econômica do Estado; a concentração da renda, a redução do salário mínimo real e os fluxos migratórios parecem ser os principais

causadores (Ortiz, 1982a). Durante os anos do milagre, a migração para o Estado de São Paulo contribuiu significativamente para a elevação das taxas de mortalidade infantil, principalmente nas áreas urbanas. Isto teria ocorrido por dois motivos. Em primeiro lugar, os migrantes provinham de áreas onde a mortalidade era mais alta, trazendo o padrão dos seus lugares de origem. Em segundo lugar, o rápido crescimento da periferia urbana, para onde os migrantes se dirigiam, causou problemas na expansão da infra-estrutura sanitária que, via de regra, não tem conseguido acompanhar o ritmo do crescimento da demanda por água e esgoto.

A partir de 1975, a taxa retoma seu processo decrescente. Possíveis explicações seriam dadas pela ampliação da rede de água e esgoto, pela crescente cobertura de vacinas, além dos programas de saúde e de atendimento pré-natal (Monteiro, 1982). Além disto, pela diminuição da migração em função do desemprego, que implicaria na diminuição da importação de padrões de mortalidade mais elevada e no alívio da pressão sobre certos serviços escassos.



5.3.2 Descrição da taxa mensal.

Infelizmente, não se tem encontrado trabalhos sobre as consequências das estações, isto é, sobre a influência do tempo e temperatura nas frequências mensais dos óbitos e suas relações no campo da saúde. Todos os trabalhos são feitos sobre a taxa anual de mortalidade, embora se possa deduzir, por exemplo, que as doenças infecciosas e parasitárias (principalmente diarreicas) ocorram principalmente nos meses mais quentes, e que doenças do aparelho respiratório (principalmente pneumonia) aconteçam nos meses frios.

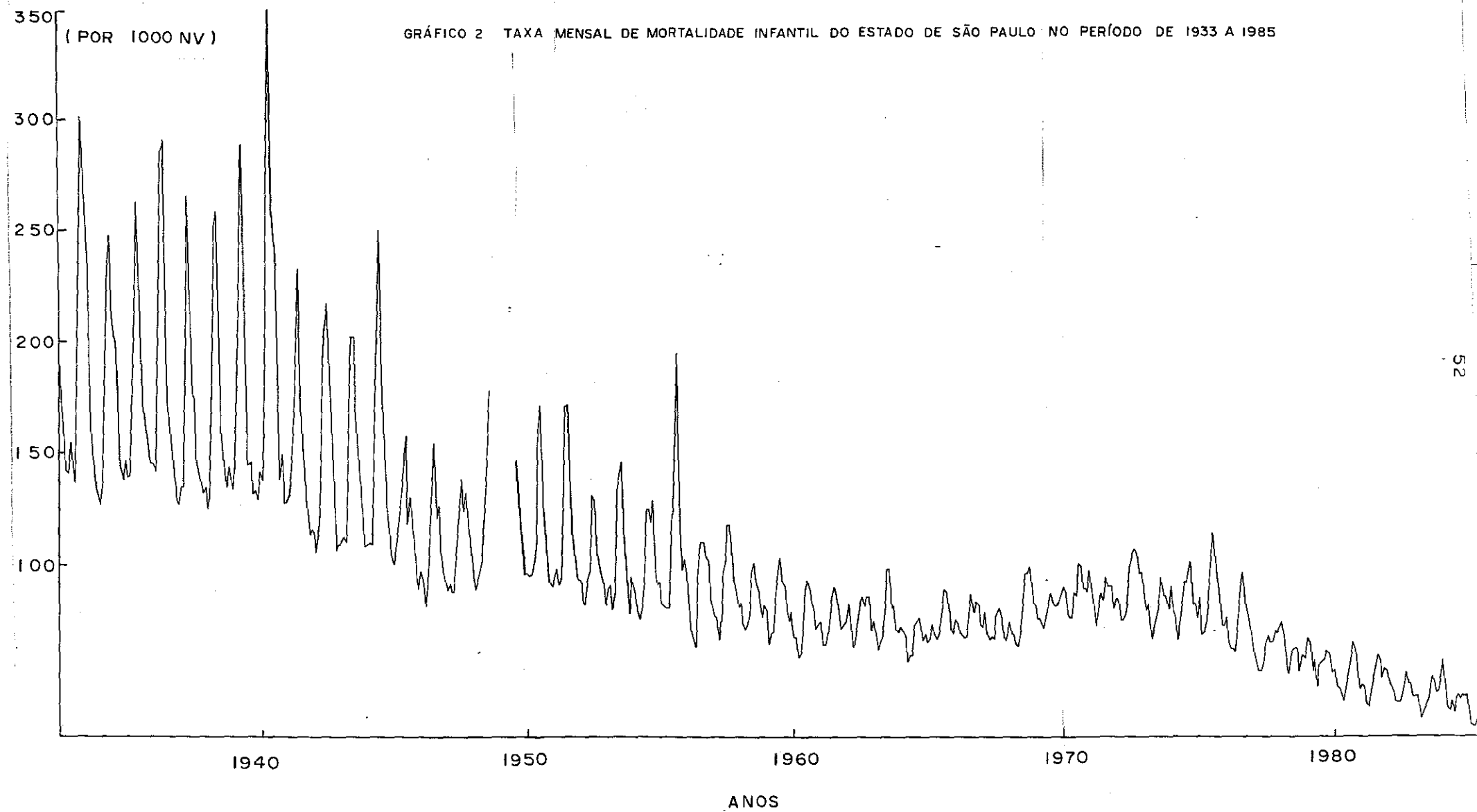
Giraldelli e Saad (1982) ao estudar a sazonalidade dos eventos vitais no Estado de São Paulo no período de 1930 a 1980 concluem que na série de nascimentos (apesar de apresentar um comportamento diferencial entre os meses), não se consegue distinguir com clareza um padrão de comportamento mensal. Entretanto, na série de óbitos, pode-se observar um comportamento regular, verificando-se um decréscimo até à metade do ano, passando a crescer até o final. Os autores afirmam que o padrão sazonal se modifica a longo do período em estudo e que o modelo encontrado no final do período é quase inverso ao padrão do início do período.

No gráfico 2 pode ser apreciada a série em forma mensal. A tendência mostra ligeira diferença em relação à tendência mostrada pela taxa anual, em particular nos pontos de mudança de estrutura. Pode ser apreciada também a estrutura sazonal ao longo do período.

A tendência mostra várias etapas, a primeira que vai até 1940, onde esta permanece quase constante. A partir de 1941 a tendência começa a decrescer até 1960, onde a queda é mais acentuada na década de 50. Na década de 60 a tendência permanece quase constante, onde, nos últimos anos se observa um processo crescente na queda da taxa, permanecendo neste estado durante parte da década de 70, atingindo o máximo em 1975. A partir de então, a tendência retoma seu processo decrescente até o final do período, 1985.

A estrutura de sazonalidade vai mudando ao longo do período, de uma estrutura sazonal bem acentuada nos primeiros anos, com altas taxas nos meses de verão e baixas nos meses de inverno, passando a ter uma sazonalidade mais suave nos últimos anos, mostrando a diminuição da vulnerabilidade da saúde da população às condições climáticas, como consequência do melhoramento das condições de vida. A estrutura sazonal pode ser dividida em cinco períodos, de 1933 a 1945, de 1946 a 1958, de 1959 a 1965, de 1966 a 1974, e, de 1975 a 1985.

O primeiro período que vai de 1933 até 1945, se caracteriza pela forte sazonalidade estável que apresenta. O padrão sazonal se assemelha a uma parábola com concavidade positiva, onde os meses de calor possuem taxas acima da média, atingindo o máximo no mês de dezembro, e os meses de frio possuem taxas abaixo da média atingindo o mínimo, principalmente, nos meses de agosto ou junho. Os valores da taxa nos meses de calor são tão elevados que a presença de máximos e mínimos relativos neste período é irrelevante.



Isto fica explicado porque neste período a população sofre fortemente as consequências do clima, produto do baixo nível da qualidade de vida. As altas taxas atingidas nos meses de calor são devidas principalmente às doenças infecciosas e parasitárias, (35% do total de óbitos, onde as doenças diarréicas são responsáveis por mais de 70% destes). As taxas baixas nos meses de frio se devem às doenças do aparelho respiratório, responsáveis por 10% do total, sendo que a pneumonia é responsável por 80% destas (Ortiz, 1982a).

O padrão sazonal de 1946 até 1958 mantém as características do período anterior, mas este é mais suave; a amplitude tende a ser menor. A concavidade da parábola fica interrompida no mês de fevereiro de 1946 a 1948, onde se observa um ligeiro aumento relativo na taxa. Em outros anos, observa-se a presença de um mínimo relativo, principalmente, no mês de maio e um máximo relativo, principalmente, no mês de junho. A importância relativa destes pontos ainda é irrelevante quando comparada com as altas taxas dos meses de calor. O mínimo absoluto é atingido com maior frequência no mês de agosto, às vezes em julho, entretanto o máximo absoluto é atingido principalmente no mês de dezembro. Neste período a amplitude sazonal decresce rapidamente.

A redução das doenças diarréicas, provavelmente a causa de morte mais sensível à influência do meio social e cultural e às medidas de saúde pública (Ortiz, 1982a), poderia explicar a diminuição considerável da amplitude sazonal.

De 1959 a 1965, observa-se que o padrão é aquele que começa em janeiro, descendo até março ou abril, e logo sobe até atingir um máximo relativo em maio ou junho, descendo de novo até atingir o mínimo absoluto em agosto ou julho, para logo continuar sua subida, até atingir o máximo, principalmente, no mês de dezembro. A diferença em relação ao período anterior é que estes pontos relativos adquirem maior importância relativa, devido ao fato da queda nos valores dos pontos máximos. A amplitude sazonal continua decrescendo, embora com menor rapidez em relação ao período anterior.

De 1966 até 1974, observa-se uma sazonalidade diferente da dos anos anteriores. Dezembro deixa de ser o mês mais importante da série. O máximo é atingido, com maior frequência, no mês de fevereiro, duas vezes em dezembro, duas em janeiro, e uma vez em junho (1970). O mínimo é atingido em agosto ou setembro. Quanto aos pontos relativos, o mínimo é atingido, principalmente, em maio e o máximo, principalmente em junho; alguns destes pontos tem maior importância que no período anterior. Pode-se apreciar uma defasagem na sazonalidade. Por outro lado, pode-se observar que a amplitude sazonal cresce ligeiramente. O padrão sazonal de 1970 é um outlier, o mínimo relativo é atingido no mês de fevereiro, logo sobe até atingir o máximo absoluto no mês de junho, descendo novamente até atingir o mínimo no mês de outubro, para logo subir novamente até dezembro.

A partir de 1975, o máximo absoluto passa a ser atingido principalmente em fevereiro, e o mínimo, principalmente em outubro. Observa-se uma defasagem maior com respeito aos períodos anteriores. Se for comparada com os anos iniciais a defasagem é de dois meses, e em relação ao período anterior a defasagem é de um mês. A estrutura de mínimos e máximos relativos, é similar a dos anos anteriores, mas às vezes se confunde com os máximos absolutos. A amplitude sazonal deste período é decrescente.

Ainda neste período, a estrutura de sazonalidade muda substancialmente. Não há padrão sazonal predominante. Para alguns anos, o padrão é similar ao dos anos anteriores, mas para outros anos as taxas mais altas acontecem nos meses de frio e as mais baixas nos meses de calor. Este é o caso dos anos 1978, 1984, 1979 e 1985, onde o máximo é atingido no mês de julho, para os dois primeiros anos, no mês de junho para 1979 e no mês de abril para o ano de 1985. Na verdade, não se pode falar de aumento da taxa nos meses de frio, o que se pode afirmar é que, relativamente a outros anos, esta taxa aumenta, e o aumento relativo se deve também à diminuição da taxa nos meses de calor.

A evolução do padrão sazonal ao longo de todo o período pode ser explicada pela mudança na estrutura da causa de mortalidade infantil. Ortiz (1982a), ao estudar a evolução da mortalidade infantil no Estado de São Paulo, no período de 1930 a 1979, mostra que as doenças "não sazonais" (endógenas) evoluem de 10% em 1935 a 42% em 1979, e que as doenças afetadas pela sazonalidade, tais como, as doenças infecciosas e parasitárias, evoluem no mesmo período, de 38% a 25%, enquanto que as doenças do aparelho respiratório, de 10% a 17%. Embora a estrutura se modifique, a redução da mortalidade neste período é bem acentuada, como se viu no estudo da tendência, devido ao fato de que foram eliminadas, quase que totalmente, algumas causas ligadas às condições de saneamento e saúde pública do Estado. Contudo, observa-se um aumento relativo das mortes devidas à doenças diarréicas, o que se constitui num motivo de preocupação a respeito da qualidade da assistência médica que vem tendo o recém-nascido nos hospitais, especialmente tendo em conta o desenvolvimento econômico do Estado.

A amplitude sazonal está fortemente ligada à tendência ao longo de todo o período. Quer dizer, há uma estreita relação entre a sazonalidade e a tendência, significando que, à medida que se reduz a mortalidade infantil, reduz-se a influência das estações nesta.

5.4 Escolha dos pontos de estudo.

A falta de informação de 1949 fez com que se dividisse a série em dois períodos de estudo: o primeiro que vai desde 1933 até 1948 e, o segundo, que vai desde 1950 a 1985.

Dadas as características peculiares dos períodos definidos previamente, são feitos estudos diferentes em cada um deles. Para o período de 1933 a 1948, é concentrada a atenção nos pontos mais importantes que são: o maior ponto de inflexão (40/41), e nos períodos 42/44 e 46/48 onde acontecem quedas no nível da tendência e ligeira mudança na estrutura sazonal no último período. Estes estudos são feitos para se analisar como e com que velocidade o método acompanha estas mudanças.

O objetivo da análise do segundo período é diferente do primeiro. Neste período pretende-se avaliar o desempenho do método quando a série possui diversas características. Assim, a série é dividida em períodos que possuem características contrastantes:

-1950 a 1965, que inclui a década de 50, caracterizada, principalmente, pela amplitude sazonal e tendência decrescentes e parte da década de 60 onde estes fatores permanecem quase constantes (decrecem mais lentamente).

-1960 a 1975, que inclui a década de 60, caracterizada, principalmente, pelo nível da tendência e amplitude sazonal quase constante e parte da década de 70 onde ambos fatores crescem ligeiramente. Neste período o padrão sazonal muda e apresenta ligeira defasagem.

-1970 a 1985, que inclui os últimos 10 anos, cuja principal característica, é a falta de um padrão dominante, onde alguns anos a taxa máxima é atingida nos meses de frio.

Entretanto, os períodos de 1950 a 1975 e de 1960 a 1985 tem o objetivo de ajuste histórico, e os períodos que incluem o ano de 1968 tem como objetivo à análise deste período tão controvertido.

Nas duas seções seguintes são detalhadas as características da série nos pontos de estudo e descritos os resultados dos experimentos.

5.5 Estudo do desempenho do método no período de 1933 a 1948.

Na série original (v. gráfico 3, pag. 68), observa-se que a tendência possui três níveis. O primeiro, que vai de 1933 até 1940, atingindo uma taxa anual média de 176,59 por mil. No final de 1940 e começo de 1941 a tendência atinge um pico, logo decresce no final do ano 1941 onde atinge um novo patamar. Este segundo nível vai de 1942 a 1944, e é inferior ao anterior, atingindo uma taxa anual média de 145,46 por mil. Em 1945 observa-se uma queda mais forte que a anterior (109,78), e, o último nível, que abrange os anos de 1946 a 1948, possui uma taxa média de 111,02 por mil. O maior valor observado acontece em dezembro de 1940.

A estrutura sazonal que vai de 1933 a 1945 é estável, o padrão vigente é aquele que atinge o máximo em dezembro, onde começa uma queda constante, até atingir o mínimo com mais frequência nos meses de junho ou agosto, raramente em julho e uma vez só em setembro. Logo começa a subir novamente até atingir o máximo no mês de dezembro com maior frequência, duas vezes em novembro e duas em janeiro. Este padrão se modifica ligeiramente nos últimos 3 anos, onde se observa que a taxa do mês de fevereiro aumenta relativamente com respeito à taxa dos anos anteriores. A amplitude sazonal é maior nos anos intermediários e decresce substancialmente nos últimos anos.

Dado o objetivo de avaliar o desempenho do programa, foram realizados cortes na série onde acontecem mudanças bruscas no nível da tendência e na sazonalidade, a fim de estudar o grau de velocidade que o método possui para acompanhar estas mudanças. Os cortes foram feitos de tal maneira que incorporassem gradualmente estas mudanças. O primeiro corte foi feito em 1940. Neste período (1933-1940) a série é fortemente sazonal, estável e bem comportada. Logo se incorporou 1941, onde acontece a primeira queda no nível, depois se fez o ajuste até 1944 tendo-se incorporado nesta série a primeira mudança de nível. Em 1945, acontece a segunda queda, mais forte que a anterior; e, finalmente, os anos 46, 47 e 48 onde acontece o segundo nível e a mudança na sazonalidade. Foram realizados então os seguintes experimentos:

1. Ajuste da série no período 1933-1940: TX3340
2. Ajuste da série no período 1933-1941: TX3341 (inclusão da primeira queda)
3. Ajuste da série no período 1933-1944: TX3344 (inclusão de mudança brusca de nível)
4. Ajuste da série no período 1933-1945: TX3345 (inclusão da segunda queda)
5. Ajuste da série no período 1933-1946: TX3346 (inclusão da segunda mudança de nível e da sazonalidade).
6. Ajuste da série no período 1933-1947: TX3347 (idem)
7. Ajuste da série no período 1933-1948: TX3348 (idem)

Depois foi ajustada a opção automática para a série em todos estes períodos, a fim de avaliar o seu desempenho ao se introduzir mudanças na tendência e na sazonalidade.

5.5.1 Análise dos resultados do ajuste da opção automática.

A seguir são apresentados os resultados da aplicação da opção automática na série nos diversos períodos.

5.5.1.1 Análise do ajuste da opção ARIMA.

A tabela 1 mostra a qualidade do ajuste dos modelos ARIMA da opção automática, (ver seção 3.3.2.1). Apresenta os valores do erro absoluto percentual médio (MAPE), que mede o erro de previsão do modelo em cada um e para todos os meses dos três últimos anos. O MAPE serve para medir o grau de ajuste destes modelos nos anos mais recentes, a fim de captar as mudanças substanciais neste período. Valores acima de 15% provocam a rejeição do modelo. Além disto, esta tabela também

apresenta o valor do p-valor associado ao teste X^2 de Portmanteau para testar a hipótese nula de aleatoriedade dos resíduos. Se o p-valor for menor que 10%, rejeita-se a hipótese nula e, portanto, o modelo será descartado. Inclui-se ainda nesta tabela o valor de R^2 que serve para avaliar a presença de superdiferença, o qual deve ser menor que 0,9.

Dado que se usa a opção ARIMA com substituição de valores extremos, estas estatísticas são apresentadas para o modelo final modificado (mf), se algum dos três modelos for aceito. Por outro lado se algum modelo for aceito ou for escolhido pelo usuário e a série tiver menos de 15 anos de dados, estas estatísticas também são calculadas para a previsão para trás. No caso, o MAPE é calculado para cada um dos três primeiros anos e para os três anos juntos; neste caso, o limite de aceitação é de 18%.

Tabela 1. Ajuste dos modelos ARIMA da opção automática

série	modelos	M A P E				2 IX (%)	2 R
		média	1	1-1	1-2		
TX3340	*lg(0,1,1)(0,1,1)	7,41	12,73	5,95	3,55	29,75	0,87
	lg(0,2,2)(0,1,1)	12,02	23,60	0,91	3,36	44,85	0,85
mf	*lg(0,1,1)(0,1,1)	8,06	13,22	6,18	4,78	19,54	0,88
	backf *lg(0,1,1)(0,1,1)	13,14	18,62	6,95	11,86	19,34	0,87
TX3341	lg(0,1,1)(0,1,1)	15,35	26,86	13,10	6,10	15,34	0,86
	lg(0,2,2)(0,1,1)	29,07	54,76	22,98	9,46	26,71	0,84
	** --(2,1,2)(0,1,1)	7,46	8,13	7,29	6,95	82,75	0,85
TX3344	lg(0,1,1)(0,1,1)	7,31	11,56	6,89	3,49	1,13	0,89
	lg(0,2,2)(0,1,1)	11,27	15,54	9,19	9,09	2,23	0,87
	** --(2,1,2)(0,1,1)	5,80	3,90	8,43	5,05	50,29	0,88
TX3345	lg(0,1,1)(0,1,1)	13,87	22,71	11,39	7,02	1,29	0,89
	lg(0,2,2)(0,1,1)	20,37	36,31	15,66	9,14	1,66	0,87
	** --(2,1,2)(0,1,1)	8,32	11,65	4,54	8,76	83,80	0,87
TX3346	lg(0,1,1)(0,1,1)	14,90	10,46	22,62	11,62	2,17	0,90
	lg(0,2,2)(0,1,1)	24,57	22,26	36,23	15,24	3,76	0,89
	--(2,1,2)(0,1,1)	11,29	9,31	14,24	10,29	65,08	0,88
TX3347	lg(0,1,1)(0,1,1)	14,61	10,42	10,95	22,45	5,03	0,91
	lg(0,2,2)(0,1,1)	23,36	12,56	22,24	35,27	7,03	0,90
	** --(2,1,2)(0,1,1)	10,60	9,08	11,72	11,01	76,55	0,89
TX3348	lg(0,1,1)(0,1,1)	15,80	24,58	11,18	11,64	10,80	0,92
	lg(0,2,2)(0,1,1)	15,67	23,92	10,73	12,35	12,32	0,91
	--(2,1,2)(0,1,1)	20,86	27,62	21,25	13,71	44,87	0,89

* escolhido pelo programa.

** aparentemente o modelo ARIMA(2,1,2)(0,1,1) cumpre com as especificações requeridas para sua escolha.

A série TX3340 aceita o modelo $ARIMA(1g(0,1,1)(0,1,1))$, ampliando a série com um ano no início (1932) e um ano no final (1941). Isto acontece porque esta série é bem comportada.

Observa-se que os modelos com transformação logarítmica, com exceção das séries TX3340, TX3341 e TX3348, possuem valores de X^2 abaixo de 10%, o que leva à rejeição da hipótese de aleatoriedade dos resíduos. Isto implica que, o fato de usar transformação log, introduz autocorrelação nos resíduos.

Quanto ao modelo $ARIMA(2,1,2)(0,1,1)$, observa-se que, com exceção das séries TX3346 e TX3348 (onde os valores do MAPE superam os limites permissíveis), aparentemente cumpre com todas as especificações requeridas. Porém, a opção automática o rejeita em todos os casos. Isto pode ser explicado, ou pelos erros de previsão muito grandes, ou porque, ao ampliar a série, se introduz autocorrelação. Para se ter uma idéia clara do que acontece, este modelo foi ajustado às séries TX3344 e TX3348, mesmo considerando que este último possui altos valores do MAPE.

Tabela 2. Ajuste do modelo $ARIMA(2,1,2)(0,1,1)$ às séries TX3344 e TX3348.

série	M A P E					
	média	1	1-1	1-2	X^2 (%)	R
TX3344						
Ajuste	5,80	3,90	8,43	5,05	50,27	0,88
forecast	5,78	4,32	7,10	5,85	3,32	0,89
backcast	8,26	10,38	7,53	6,86	1,61	0,90
TX3348						
Ajuste	20,86	27,62	21,25	13,71	44,87	0,89
forecast	17,96	25,13	19,48	9,27	4,84	0,90

A tabela 2 mostra que os resíduos do modelo $ARIMA(2,1,2) \times (0,1,1)$ não são aleatórios ao se ampliar a série com dados ajustados por este modelo. Esta é a razão pela qual o programa o rejeita.

Tabela 3. Ajuste do modelo ARIMA opção aditiva à série TX3348.

modelo	M A P E					
	média	1	1-1	1-2	X^2 (%)	R
(011)(011)	43,34	47,32	36,30	46,41	10,71	0,86
(022)(011)	49,35	50,17	40,61	57,20	18,98	0,87
(212)(011)	20,86	27,62	21,25	13,71	44,87	0,89

Os modelos da opção aditiva são rejeitados pelos altos valores do MAPE (v. tabela 3).

Da observação das tabelas anteriores conclue-se que nenhum modelo ARIMA possui qualidade suficiente para ajustar os dados de forma adequada. Em parte isto era esperado, pelo menos para as séries posteriores a 1940, porque dificilmente existiria um modelo ARIMA que ajustasse adequadamente uma série que é dominada por uma estrutura de sazonalidade e tendência quase constantes e que, nos últimos anos, sofre quedas fortes na tendência. Estes modelos não captam estas mudanças, fazendo previsões segundo a estrutura dominante, aumentando os erros de previsão, e por isso são rejeitados.

5.5.1.2 Análise da qualidade do ajuste.

A contribuição relativa de cada componente na variância da mudança percentual da porção estacionária da série é calculada como o quociente entre a variância do componente e da série estacionarizada. Esta estatística serve para medir a participação da cada componente, especialmente do componente irregular, do qual se espera que contribua com, o máximo, 10% da variação.

Tal como era esperado, estas séries possuem um componente irregular pequeno (menor que 6%), e a maior parte da variação fica explicada pela sazonalidade (ao redor de 85%). À medida que se incorporam dados adicionais, a participação da tendência aumenta, e, em compensação, a participação do fator sazonal diminui (tabela 4).

Tabela 4. Estrutura da variância da porção estacionária da série original (%).

compon	TX3340	TX3341	TX3344	TX3345	TX3346	TX3347	TX3348
I	5	4	4	4	4	4	4
C	4	6	8	9	13	14	14
S	91	90	88	87	83	82	82

O teste ADR serve para testar a aleatoriedade dos resíduos e está baseado no número de pontos de inflexão (quando a sequência troca de sinal). O intervalo de aceitação da hipótese nula ao nível de 1% é de 1,36 a 1,75 (v. seção 3.3.10.1).

O ADR do componente irregular fica quase no limite permissível de aceitação da hipótese de aleatoriedade dos resíduos (1,75), acusando presença de autocorrelação positiva em alguns casos. O ADR da tendência apresenta forte autocorrelação positiva, o que era esperado. O número de meses para a dominação das mudanças da tendência com respeito às mudanças do componente irregular, é alto. Isto implica que a opção automática não está fornecendo uma série sazonalmente ajustada suave (tabela 5).

Tabela 5. Duração média de sequências (ADR) e o número de meses para dominação cíclica (MCD).

Estat.	TX3340	TX3341	TX3344	TX3345	TX3346	TX3347	TX3348
CI	1,76	1,88	1,93	1,85	1,86	1,86	1,95
I	1,58	1,75	1,79	1,76	1,74	1,63	1,77
C	8,64	7,64	7,53	7,75	7,59	7,46	6,59
MCD	3,64	3,68	3,56	3,77	3,54	3,50	3,40
#meses	5	5	5	5	5	5	4

As razões $\bar{I}/\bar{C}$ e $\bar{I}/\bar{S}$ medem o grau de variabilidade entre os componentes. $\bar{I}$, $\bar{C}$, e $\bar{S}$ são as médias das mudanças absolutas do componente irregular, da tendência e da sazonalidade. Estas razões são sumamente importantes, pois, valores altos indicam que a série é altamente irregular e que a separação dos componentes pode não ser satisfatória, comprometendo a qualidade do ajuste. (seção 3.3.11).

Na tabela 6 não se observa uma tendência clara na razão dos componentes. As razões, em relação à tendência, caem no intervalo 1,0 a 3,0 indicando que o filtro adequado para todas as séries é o 13-M, o qual, é de fato escolhido pela opção automática. Para o fator sazonal, as razões caem no intervalo de 1,5 a 6,5, podendo-se considerar a série como "padrão" (nem muito suave, nem muito irregular) nestes períodos.

Tabela 6. Razão final entre os componentes.

razão	TX3340	TX3341	TX3344	TX3345	TX3346	TX3347	TX3348
$\bar{I}/\bar{C}$	1,95	2,43	2,57	2,47	3,34	2,43	2,37
$\bar{I}/\bar{S}$	3,97	4,66	5,01	4,67	5,59	4,65	4,08

A tabela 7 apresenta os testes para avaliar a presença de sazonalidade estável e móvel, aplicados às razões SI (v. seção 3.3.9). As três primeiras linhas contêm as estatísticas associadas à hipótese nula de presença de sazonalidade estável não significativa. Os dois primeiros são um teste F de um fator, Fs_B1 denota que é aplicado na estimativa preliminar de SI e Fs_D8 nas estimativas finais (sem substituição final de valores extremos). O teste $X2_K_W$ é um teste não paramétrico aplicado a Fs_D8 . Valores altos indicam a presença de sazonalidade estável. O teste Fm é um teste F de dois fatores que serve para testar a hipótese nula de presença de sazonalidade móvel não significativa. é apresentado só o p-valor associado a Fm , pois, para todos os casos, o p-valor associado aos teste de sazonalidade estável, é próximo de zero. O p-valor está dado em percentagem.

Os altos valores das estatísticas mostram que a presença da sazonalidade estável é forte. A nível de 0,1%, a hipótese de presença de sazonalidade móvel é aceita para a série TX3348, o que confirma as expectativas.

Tabela 7. Teste para sazonalidade.

Estat.	TX3340	TX3341	TX3344	TX3345	TX3346	TX3347	TX3348
Fs_B1	89,3	83,0	125,0	148,7	138,4	145,4	132,0
Fs_D8	136,3	120,6	191,3	192,9	218,5	199,8	181,6
X2_K_W	87,8	95,2	127,5	138,4	148,7	158,8	168,1
Fm	0,5	0,5	0,5	0,9	1,0	2,7	3,0
p-valor	82,5	83,4	90,8	56,0	45,8	1,6	0,03

A função de autocorrelação dos resíduos serve para analisar a aleatoriedade dos resíduos e está formada pelas autocorrelações de lag k . Foi tomado como erro padrão das estimativas das autocorrelações, o quociente $1/\sqrt{n}$. Valores que superam a faixa de confiança de 95%, isto é, $2/\sqrt{n}$, indicam possível presença de autocorrelação do lag correspondente (v. seção 3.3.10.2).

Observa-se na tabela 8 que a série TX3340 está livre da presença de autocorrelação, e, a partir de 1941 a autocorrelação de lag 2 e de lag 3 é alta. A autocorrelação de lag 9 aparece nas duas últimas séries.

Tabela 8. Função de autocorrelação dos resíduos.

lag k	TX3340	TX3341	TX3344	TX3345	TX3346	TX3347	TX3348
1	0,09	0,13	0,11	0,12	0,10	0,13	0,14
2	-0,19	-0,23*	-0,27*	-0,32*	-0,33*	-0,30*	-0,29*
3	-0,04	-0,20*	-0,16	-0,18*	-0,20*	-0,19*	-0,19*
4	0,07	-0,12	-0,06	-0,03	-0,08	-0,02	-0,03
5	-0,00	-0,04	-0,03	-0,00	0,03	-0,00	-0,00
6	0,03	0,10	0,03	0,00	0,02	-0,01	-0,01
7	-0,03	0,02	-0,02	-0,02	-0,05	-0,07	-0,04
8	-0,11	-0,07	-0,05	-0,01	-0,04	-0,02	-0,01
9	-0,17	0,01	0,07	0,13	0,15	0,17*	0,17*
10	0,04	-0,07	-0,04	-0,02	0,02	0,04	0,04
11	0,16	-0,04	-0,09	-0,14	-0,14	-0,12	-0,11
12	-0,03	-0,07	-0,05	-0,08	-0,06	-0,07	-0,09
13	-0,12	0,09	0,13	0,14	0,14	0,13	0,13
14	-0,02	-0,08	-0,08	-0,06	-0,07	-0,06	-0,04
$2/\sqrt{n}$	0,204	0,192	0,167	0,160	0,154	0,149	0,144

* supera os limites

A tabela 9 mostra a percentagem dos desvios do periodograma acumulado normalizado em relação à diagonal que une os pontos $(0,0;0,0)$ e $(0,5;1,0)$ que caem nas faixas de confiança menor que 75%, de 75% a 95%, e acima de 95%. Se a percentagem da última faixa for superior a 5%, a hipótese nula de componente periódico não significativo, será rejeitada ao nível de 5%. Os desvios são medidos pelo teste de Kolmogorov-Smirnov (v. seção 3.3.10.3).

A partir de 1945, o componente periódico torna-se cada vez mais evidente. Isto mostra que os filtros usados pela opção automática (especialmente o filtro da tendência) não estão conseguindo captar os movimentos subjacentes da série, particularmente nos últimos anos.

Tabela 9. Teste de Kolmogorov-Smirnov (%).

Interv.:	TX3340	TX3341	TX3344	TX3345	TX3346	TX3347	TX3348
-75%:	100	93	81	65	67	66	64
75%-95%:	0	7	19	19	24	17	11
+95%:	0	0	0	16	9	17	25

As estatísticas M's de controle de qualidade avaliam aspectos particulares da série. Elas tomam valores de 0 a 3, e, valores acima de 1 indicam problemas no ajuste no aspecto em questão, mas não necessariamente invalidam o ajuste e em muitos casos as deficiências podem ser corrigidas. M1 e M2 medem a participação do componente irregular em relação à sazonalidade, M3, em relação à tendência, M4, o grau de autocorrelação nos resíduos, M5, o grau de suavidade da série sazonalmente ajustada, M6, a relação entre o componente irregular e a sazonalidade anual, M7, a relação entre a sazonalidade estável e a móvel, M8 e M10, o tamanho das flutuações aleatórias da sazonalidade, onde M10 é calculado só para os últimos anos, M9 e M11 medem o movimento linear da sazonalidade, onde M11 é calculada só para os anos mais recentes. Maiores detalhes podem ser encontrados na seção 3.3.11. A estatística Q é uma média ponderada das 11 estatísticas anteriores e mede a qualidade global do ajuste. Valores acima de 1 invalidam o ajuste.

Tabela 10. As estatísticas de controle de qualidade.

Estat.:	TX3340	TX3341	TX3344	TX3345	TX3346	TX3347	TX3348
M1	0,265	0,332	0,297	0,321	0,350	0,343	0,353
M2	0,483	0,455	0,396	0,448	0,422	0,419	0,401
M3	0,475	0,713	0,767	0,737	0,668	0,717	0,683
M4	0,348	0,953	1,209*	1,161*	1,118*	0,666	1,312*
M5	0,844	0,726	0,785	0,756	0,701	0,742	0,772
M6	0,010	0,264	0,406	0,349	0,634	0,341	0,031
M7	0,177	0,189	0,149	0,158	0,151	0,182	0,210
M8	0,226	0,240	0,267	0,257	0,205	0,291	0,343
M9	0,195	0,182	0,050	0,048	0,042	0,093	0,168
M10	0,314	0,290	0,316	0,337	0,260	0,375	0,516
M11	0,312	0,288	0,285	0,313	0,253	0,361	0,516
Q	0,37	0,44	0,45	0,45	0,42	0,42	0,50
Nº falhas	--	--	1	1	1	--	1

* falha

Na tabela 10, observa-se que, em geral, a qualidade da maioria das estatísticas se deterioram à medida que se adiciona um ano, sendo a mais afetada a estatística M4, que a partir de 1944 começa a falhar. Esta estatística está relacionada com a aleatoriedade dos re-

síduos. Tal fato já tinha sido observado através do teste ADR, da FAC e do teste K-S. Note-se que, em 1945, os valores das estatísticas Q e M4 são superiores aos valores da série até 46 e 47, e que são superados só na série de 48. Este fato fica explicado porque, ao se incluir gradualmente um ano, a série se "estabiliza" e o que pode ser considerado como um outlier referente a uma série, deixa de o ser quando a série é mais longa.

5.5.1.3 Discussão dos resultados.

Conclui-se dos resultados anteriores, que a opção automática tem um bom desempenho em 1940, o que fica explicado porque a série, neste período, é bem comportada. A partir de 1941, o desempenho da opção automática deixa muito a desejar, principalmente pelas deficiências mostradas através dos resíduos, que indicam a presença de autocorrelação e periodicidade, desta forma, procurou-se outras opções que melhorassem a qualidade do ajuste.

5.5.2 Desempenho das opções alternativas para a série no período de 1933 a 1948: TX3348, X-11.

Dado que a série no período de 1933 a 1948 inclui todas as mudanças mencionadas anteriormente e dado que a opção automática não proporciona um ajuste desejável, pois apresenta resíduos com presença de autocorrelação e periodicidade, tenta-se aqui superar estas deficiências através da manipulação de opções, principalmente, dos filtros.

Descrição das opções:

1. Opção automática
2. Opção ARIMA (2,1,2)(0,1,1)
3. Opção padrão (é a opção padrão do X-11)
4. Opção 3x3
5. Opção 3x3, 9-H
6. Opção 3x5, 9-H
7. Opção 3x3, 9-H, s/out
8. Opção 3x3, 9-H, s/out, aditivo

A participação do componente irregular na estrutura da variância pode ser apreciada na tabela 11. As opções ARIMA e aditiva aumentam ligeiramente a participação do componente irregular, e as opções que interferem no tipo de filtro dos componentes a diminuem.

A opção que possui menor valor para o ADR do componente irregular é a dos filtros curtos: 3x3,9-H. Todas as opções alternativas à opção automática têm valores menores que 1,75, aceitando-se a hipótese de aleatoriedade dos resíduos (v. tabela 12). As opções que fornecem séries sazonalmente ajustadas mais suaves (3 meses) são aquelas que usam o filtro 9-H para a tendência.

Tabela 11. Estrutura da variância da porção estacionária da série original(%).

	!	!	!	!	!	!	!	!
	!	ARIMA	!	3x3	3x3	3x3	3x5	3x3,9H
compo-	auto	!(212)x	padrão!		9-H	9-H	9-H	aditivo
nentes!	!	(011)!	!	!	!	s/out	!	s/out
I	4	5	4	3	3	3	3	4
C	14	12	14	15	15	15	15	13
S	82	83	82	82	82	82	82	83

Tabela 12. Duração média de sequências (ADR) e número de meses para dominação cíclica (MCD).

	!	!	!	!	!	!	!	!
	!	ARIMA	!	3x3	3x3	3x3	3x5	3x3,9H
Estat.	auto	!(212)x	padrão!		9-H	9-H	9-H	aditivo
!	!	(011)!	!	!	!	s/out	!	s/out
CI	1,95	1,84	1,87	1,95	1,91	1,91	1,93	2,08
I	1,77	1,71	1,74	1,74	1,65	1,74	1,71	1,72
C	6,59	6,82	6,59	7,07	4,90	6,16	4,90	5,16
MCD	3,40	5,00	3,40	4,35	3,00	3,00	2,82	4,79
#meses	5	6	5	4	3	3	3	4

As opções que fornecem as menores razões são o 3x3,9-H,s/out e a opção 3x3,9-H, isto é, a dos filtros mais curtos, ver tabela 13.

Tabela 13. Razão final entre os componentes.

	!	!	!	!	!	!	!	!
	!	ARIMA	!	3x3	3x3	3x3	3x5	3x3,9H
razão	auto	!(212)x	padrão!		9-H	9-H	9-H	aditivo
!	!	(011)!	!	!	!	s/out	!	s/out
I/C	2,37	2,70	2,39	1,90	1,69	1,20	1,76	1,25
I/S	4,08	4,70	4,02	3,82	3,17	2,76	3,39	2,33

A hipótese de presença de sazonalidade estável é aceita, pois o p-valor associado a todos eles é quase zero. Ao nível de 0,1%, a hipótese de presença de sazonalidade móvel é aceita para quase todas as opções, com exceção das opções ARIMA e 3x5,9-H (v. tabela 14).

Nenhuma opção consegue eliminar totalmente a evidência de periodicidade nos resíduos. As opções que usam o filtro 9-H para a tendência, pelo menos controlam esta deficiência, e quando usadas com o filtro 3x3 para o fator sazonal, fornecem resultados ainda melhores. Isto evidencia que os filtros usados pelas outras opções (13-H para a tendência e para o fator sazonal segundo a razão I/S para cada mês) não conseguem captar os movimentos da tendência, deixando-os passar para o resíduo (tabela 15).

Tabela 14. Teste para sazonalidade.

		ARIMA		3x3	3x3	3x3	3x5	3x3,9H
Estat.	auto	(212)x	padrão		9-H	9-H	9-H	aditivo
		(011)				s/out		s/out
Fs_D1	132,0	147,1	132,0	132,0	132,0	132,0	132,0	82,6
Fs_D8	181,6	203,0	179,8	184,4	207,4	207,4	257,0	118,6
X2_K_W	168,1	169,8	168,2	168,0	168,9	168,9	170,9	163,6
Fm	3,0	2,1	3,0	3,1	3,1	3,1	2,6	7,3
p-valor	0,03	1,44	0,04	0,02	0,02	0,02	0,18	0,00

Tabela 15. teste de Kolmogorov-Smirnov.

		ARIMA		3x3	3x3	3x3	3x5	3x3,9H
Inter.	auto	(212)x	padrão		9-H	9-H	9-H	aditivo
		(011)				s/out		s/out
-75	64	55	65	71	80	75	73	69
75-95	11	10	11	20	18	24	20	26
+95	25	35	24	9	2	1	7	5

Observou-se que a opção automática estava fornecendo resíduos autocorrelacionados. Esta deficiência foi superada parcialmente pelo uso do filtro 9-H para a tendência, diminuindo ainda mais ao se combinar o filtro 9-H com o filtro 3x3 para a sazonalidade. Mesmo assim, nenhuma das opções consegue eliminar totalmente a autocorrelação dos resíduos (tabela 16).

Tabela 16. Função de autocorrelação dos resíduos.

		ARIMA		3x3	3x3	3x3	3x5	3x3,9H
lag	auto	(212)x	padrão		9-H	9-H	9-H	aditivo
k		(011)				s/out		s/out
1	0,14	0,24*	0,15*	0,10	-0,01	0,02	-0,01	0,03
2	-0,29*	-0,16*	-0,28*	-0,30*	-0,33*	-0,36*	-0,36*	-0,33*
3	-0,19*	-0,16*	-0,18*	-0,16*	-0,07	-0,14	-0,07	-0,19*
4	-0,03	-0,03	-0,03	-0,03	-0,13	0,11	0,18	0,01
5	-0,00	-0,01	-0,00	0,00	0,09	0,07	0,10	0,11
6	-0,01	-0,02	-0,01	-0,00	-0,05	-0,03	-0,06	-0,02
7	-0,04	-0,03	-0,05	-0,08	-0,08	-0,08	-0,06	-0,06
8	-0,01	0,02	-0,01	-0,04	-0,09	-0,06	-0,04	-0,04
9	0,17*	0,12	0,17*	0,20*	0,14	0,13	0,12	0,06
10	0,04	0,04	0,03	0,02	0,07	0,03	0,03	0,03
11	-0,11	-0,09	-0,12	-0,13	-0,10	-0,10	-0,08	0,03
12	-0,09	-0,04	-0,10	-0,14	-0,11	-0,07	-0,06	0,07
13	0,13	0,09	0,12	0,11	0,25*	0,26*	0,27*	0,13
14	-0,04	-0,05	-0,04	-0,05	-0,03	0,00	-0,02	-0,07

Na tabela 17, pode-se observar que com exceção da opção 3x0,9-H, as outras opções mostram evidência de autocorrelação nos resíduos (valor de M4 superior a 1), fato que faz descartá-las. A opção ARIMA, além disto apresenta problemas na estatística M5, que controla o grau de suavidade da série sazonalmente ajustada.

Tabela 17. As estatísticas de controle de qualidade.

Estat.	! auto	! ARIMA ! !(212)x! !(011)!	! padrão!	! 3x3 !	! 3x3 ! ! 9-H !	! 3x3 ! ! 9-H ! ! s/out !	! 3x5 ! ! 9-H !	! 3x3,9H ! ! aditivo ! ! s/out
M1	0,35	0,42	0,35	0,28	0,20	0,25	0,23	0,36
M2	0,40	0,48	0,41	0,34	0,26	0,30	0,29	0,38
M3	0,68	0,85	0,69	0,45	0,35	0,10	0,38	0,12
M4	1,31*	1,05*	1,18*	1,18*	0,78	1,18*	1,05*	1,11*
M5	0,77	1,02*	0,78	0,68	0,45	0,50	0,45	0,52
M6	0,03	0,28	0,01	0,07	0,33	0,50	0,24	0,67
M7	0,21	0,18	0,21	0,21	0,20	0,20	0,17	0,35
M8	0,34	0,29	0,36	0,52	0,52	0,45	0,32	0,55
M9	0,17	0,13	0,18	0,20	0,21	0,17	0,16	0,31
M10	0,52	0,40	0,53	0,67	0,72	0,61	0,51	0,75
M11	0,52	0,40	0,53	0,65	0,70	0,60	0,51	0,74
Q	0,50	0,52	0,45	0,47	0,38	0,38	0,36	0,47
#falhas	1	2	1	1	—	1	1	1

5.5.3 Discussão e escolha da melhor opção.

Embora os modelos ARIMA da opção automática tenham sido rejeitados, experimentou-se o modelo ARIMA(2,1,2)(0,1,1). Este modelo foi escolhido porque até 1947 vinha apresentando um desempenho razoável. Ao se expandir a série com valores extrapolados por este modelo, a qualidade do ajuste se deteriora sensivelmente. Os principais problemas são autocorrelação e periodicidade nos resíduos. Em termos gerais, a qualidade se deteriora em comparação à opção automática. Isto era esperado, porque se amplia a série com valores extrapolados por um modelo que apresenta sérios problemas de não aleatoriedade nos resíduos. Após esta experiência foi descartada toda possibilidade de uso da opção ARIMA.

A opção padrão difere da opção automática no tipo de filtro sazonal usado na estimativa final destes fatores. Enquanto a opção automática escolhe o tipo de filtro para a estimativa final do fator sazonal para cada mês segundo a razão $\bar{I}/\bar{S}$, a opção padrão aplica o filtro 3x5 para todos os meses. O ganho em termos de qualidade é insignificante. Deve-se notar que o filtro usado, pelas duas opções, para a tendência é o 13-H.

Dada a estrutura de sazonalidade bem comportada, pelo menos para os primeiros anos, com componente irregular pequeno, tentou-se o uso do filtro 3x3 que é recomendado nestes casos. Obtém-se uma melhora significativa nos resíduos, mas, mesmo assim o componente periódico ainda persiste.

A presença de componente periódico nos resíduos sugere o uso de um filtro mais curto para a tendência que possa captar estes movimentos. É o caso da opção 3x3,9-H. Esta opção consegue uma melhora significativa da qualidade do ajuste, em relação a todas as opções tentadas anteriormente. Elimina (pelo menos controla) a autocorrelação e a periodicidade dos resíduos, e há um ganho significativo nas estatísticas de controle de qualidade.

Dado que a opção 3x3,9-H é a que possui o melhor desempenho, tentou-se a manipulação de valores extremos, a fim de aprimorar o ajuste. Mas, isto tem um efeito contrário com relação à opção anterior, pois introduz autocorrelação nos resíduos.

A qualidade do ajuste melhorou fundamentalmente pelo uso do filtro 9-H para a tendência; portanto, fixou-se este e manipulou-se o filtro para o fator sazonal. Ao se usar o filtro 3x5 para a estimativa do fator sazonal, em todas as iterações, obtém-se mudanças quase insignificantes, mas se introduz autocorrelação. Isto leva à conclusão de que os filtros mais curtos para sazonalidade conseguem melhor ajuste que os filtros mais compridos.

Fixados os filtros 9-H e 3x3, usou-se a opção aditiva, juntamente com a manipulação dos valores extremos. A qualidade do ajuste comparado com a opção 3x3,9-H se deteriora significativamente. Portanto a opção escolhida como alternativa à opção automática, é a opção dos filtros mais curtos, isto é, a opção 3x3,9-H.

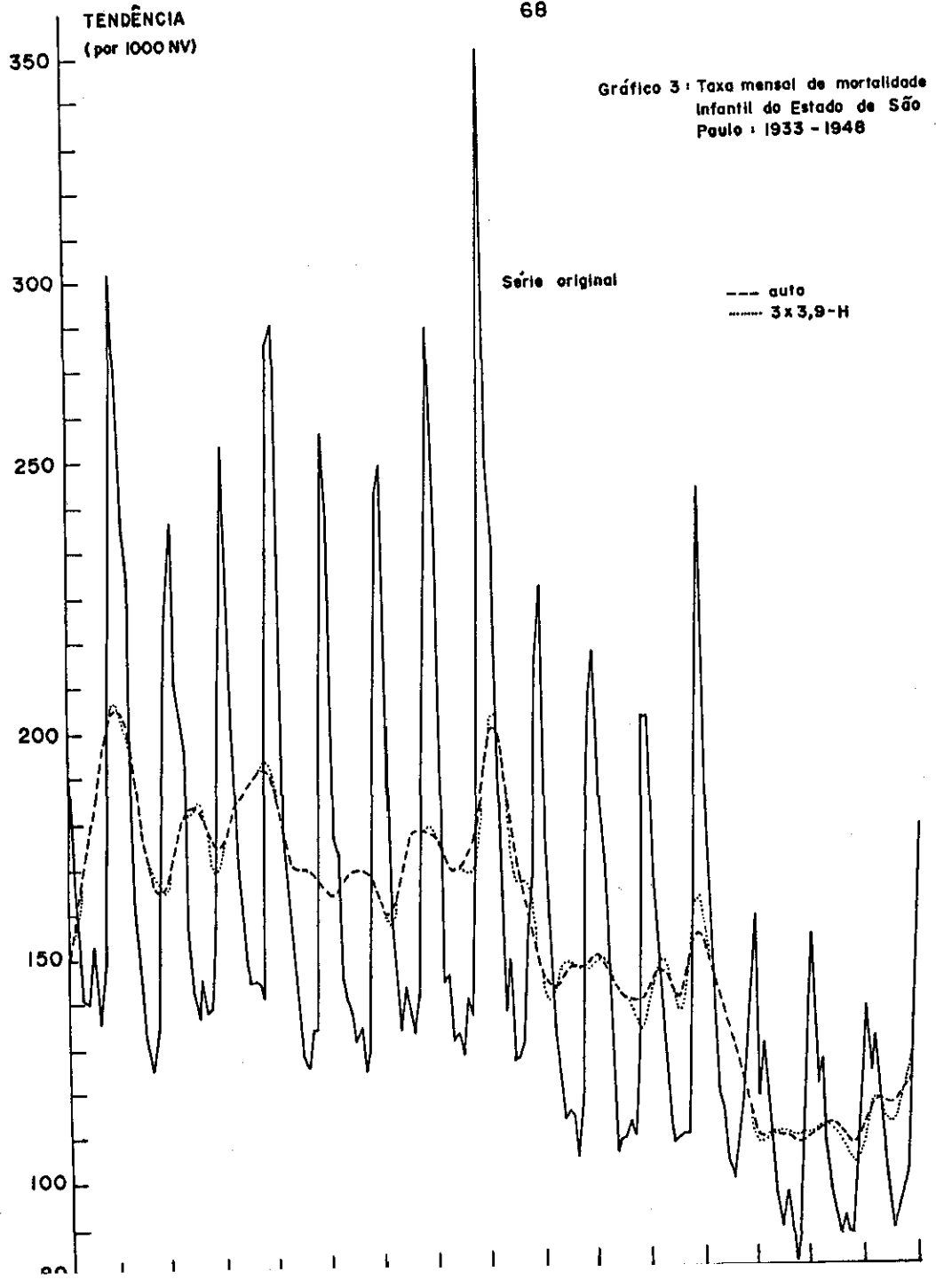
5.5.4 Estudo das revisões da opção automática versus a melhor opção.

As revisões são as mudanças nos valores ajustados ao se incorporar informação adicional e estão relacionadas com o erro de especificação da sazonalidade prevista. É desejável que estes valores sejam pequenos e que as mudanças de revisão a revisão sejam consistentes, isto é, possuam um comportamento monotônico. Para medir sua magnitude usa-se o MAE, que é a média dos valores absolutos das diferenças entre a estimativa atual e a estimativa histórica do componente (v. seção 2.5.2).

O estudo de revisão está restrito a duas opções, a automática e a opção alternativa a esta. A metodologia a seguir consiste em:

- 1) Considerar as estimativas dos componentes da série global, isto é, TX3348, como as estimativas históricas ou finais para cada opção separadamente.
- 2) Avaliar o comportamento das estimativas dos componentes atuais, isto é, antes de 1948, com respeito às estimativas finais de cada opção.

A não existência de informação para 1949, impede a avaliação das revisões nos últimos anos, pois uma estimativa corrente transforma-se em final só após três anos e meio de dados adicionais (v. seção 2.5.2). Mesmo assim se estuda o desempenho das estimativas das duas opções a fim de ter noção do comportamento delas neste período.



Entre todas as opções verificou-se que, aquela que fornece melhor qualidade em termos gerais, é a opção 3x3,9-H, que usa os filtros mais curtos, tanto para a tendência como para a sazonalidade. Em termos de ajuste, a opção 3x3,9-H, parece acompanhar melhor o movimento da tendência e da sazonalidade que a opção automática, embora ao se usar filtros curtos, corre-se o risco de contaminar as estimativas com o componente irregular.

São analisadas cuidadosamente as revisões nos períodos de 1940 a 1941, de 1942 a 1944 e de 1946 a 1948, isto porque no primeiro período ocorre o maior ponto de inflexão e nos dois últimos porque nelas acontecem mudanças bruscas.

No gráfico 3, pode-se apreciar as estimativas finais ou históricas da tendência, do fator sazonal e do componente irregular fornecida pelas duas opções. Para a tendência, as duas opções acompanham as duas defasagens na tendência e o maior ponto de inflexão. A principal discrepância entre as duas aparece nos pontos de inflexão. A opção automática é sempre mais conservadora (acompanha os movimentos de forma mais tardia) que a opção dos filtros mais curtos; este fato implica a presença de periodicidade nos resíduos. Entretanto a opção 3x3,9-H consegue acompanhar mais rapidamente os movimentos da série, mas nos pontos de inflexão tende a superestimar ou subestimar os componentes, incrementando (em geral) as revisões.

Em relação à estimativa do fator sazonal viu-se que a opção de filtros curtos tende a acompanhar melhor o movimento, isto é, no meio do período a amplitude do 3x3,9-H é maior que a da opção automática, e ao final do período a amplitude sazonal diminui mais rapidamente que a amplitude fornecida pela opção automática.

No gráfico 4, podemos observar o desempenho das duas opções no maior ponto de inflexão. Este ponto acontece em fevereiro de 1941. Deve-se notar que na série original o maior valor acontece no mês de dezembro de 1940, embora este esteja contaminado pela alta sazonalidade do mês de dezembro. As estimativas históricas das duas opções conseguem acompanhar este ponto, onde a opção dos filtros curtos fornece estimativas maiores que as da automática. As estimativas correntes de 1940, para ambas opções demoram quase dois meses para acompanhar o movimento histórico deste ano.

A opção automática para este ano, usa o modelo ARIMA (0,1,1) x(0,1,1) com transformação log, mesmo assim, apresenta as maiores revisões (tabela 10). Este fato pode ser explicado porque, ao se usar a opção ARIMA com substituição de valores extremos, o outlier de 1940 é substituído, reduzindo seu efeito. Deve-se notar que as estimativas da tendência histórica, não incorporam modelos ARIMA, por esta razão se ajustou a opção automática X-11, isto é, sem a opção ARIMA. Esta opção fornece revisões menores. A opção dos filtros curtos sempre superestima a tendência histórica nos pontos de inflexão positivos e subestima nos negativos; este fato faz com que as revisões sejam maiores. O gráfico 4a mostra o desempenho das opções automática e 3x3,9-H com manipulação de valores extremos. A primeira fornece estimativas mais suaves no maior ponto de inflexão, e a segunda, ao contrário, fornece estimativas mais elevadas.

Gráfico 4 : Desempenho das revisões da tendência : 1938-1941

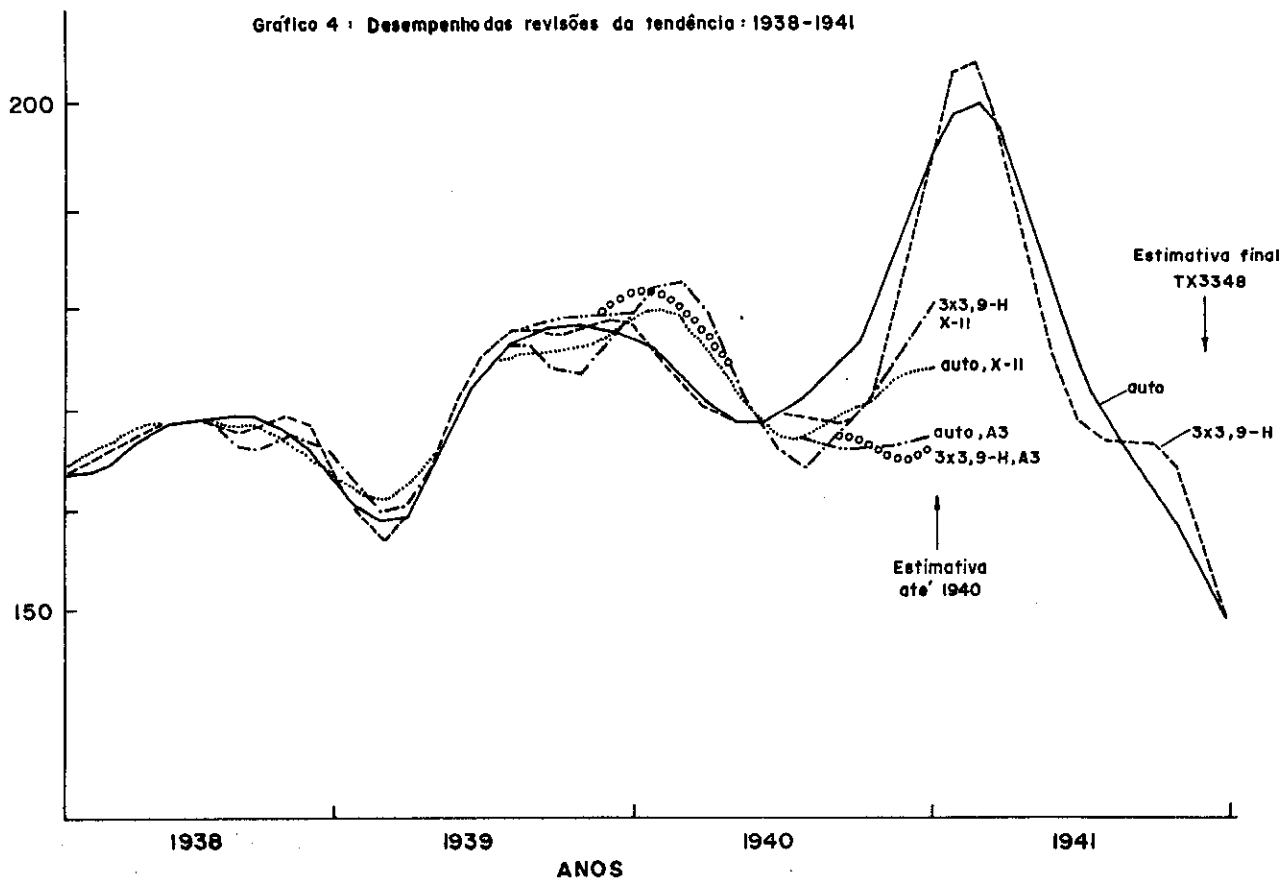
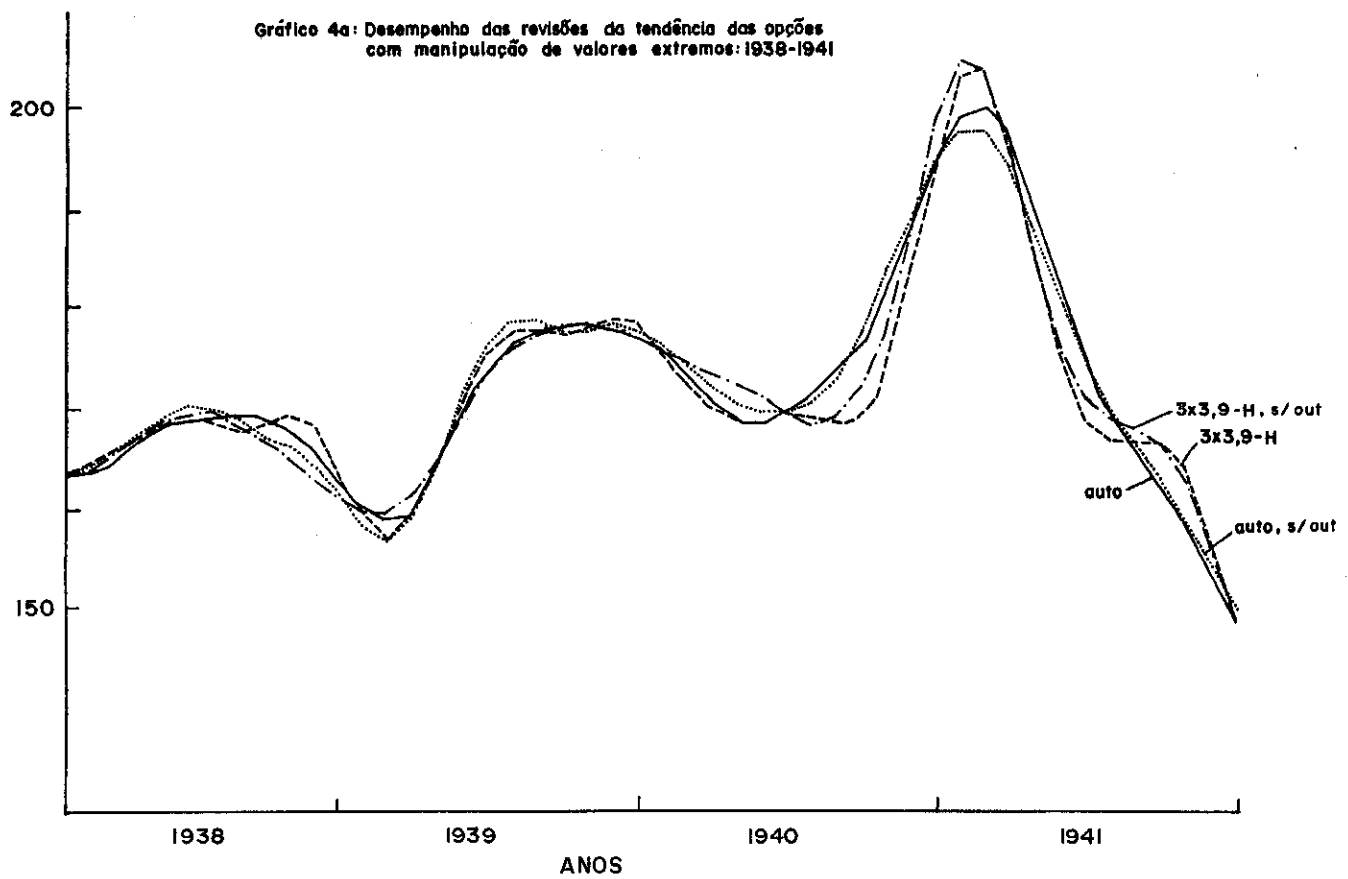
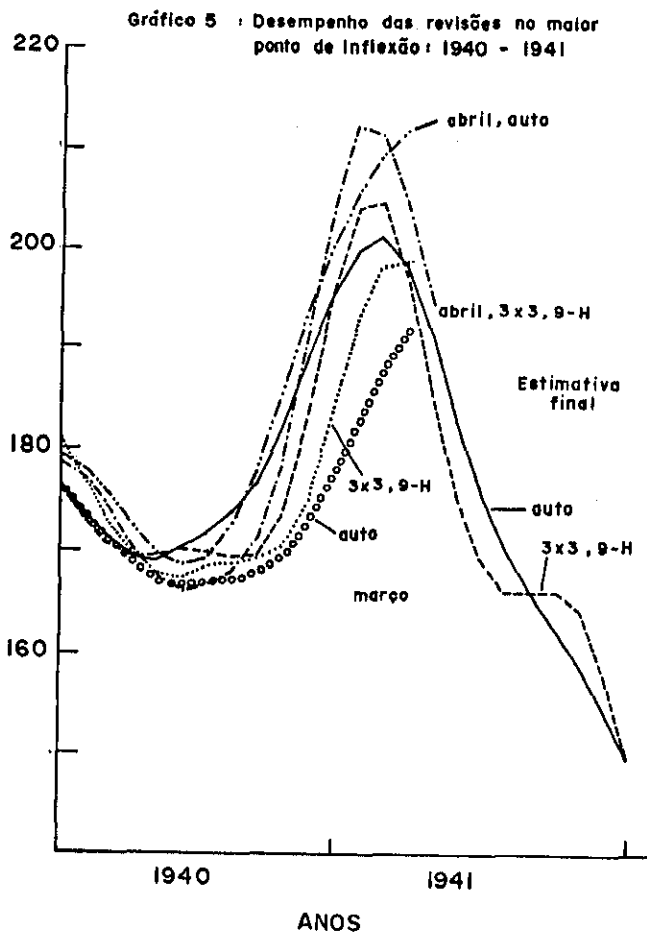


Gráfico 4a: Desempenho das revisões da tendência das opções com manipulação de valores extremos: 1938-1941



No gráfico 5, pode-se apreciar a evolução das duas opções nos meses de março e abril. A opção dos filtros curtos já em março detecta a proximidade de um ponto de inflexão e no mês de abril consegue acompanhá-lo; ao contrário, a opção automática não detecta o ponto de inflexão em nenhum dos dois meses. Isto é devido ao fato da opção dos filtros mais curtos reagir mais rapidamente a movimentos subjacentes da série, o que não acontece com a opção automática que usa filtros de comprimento médio.



No gráfico 6, pode-se observar as revisões da tendência no período de 1941 a 1944. As revisões das duas opções são relevantes nos dois últimos anos. As revisões da opção automática são sempre menores. As duas opções demoram quase um mês para acompanhar os movimentos da série. No gráfico 6a são mostradas as revisões nos anos 42 e 43, onde se observa que as duas opções acompanham o movimento histórico.

Gráfico 6 : Desempenho das revisões da tendência : 1941 - 1944

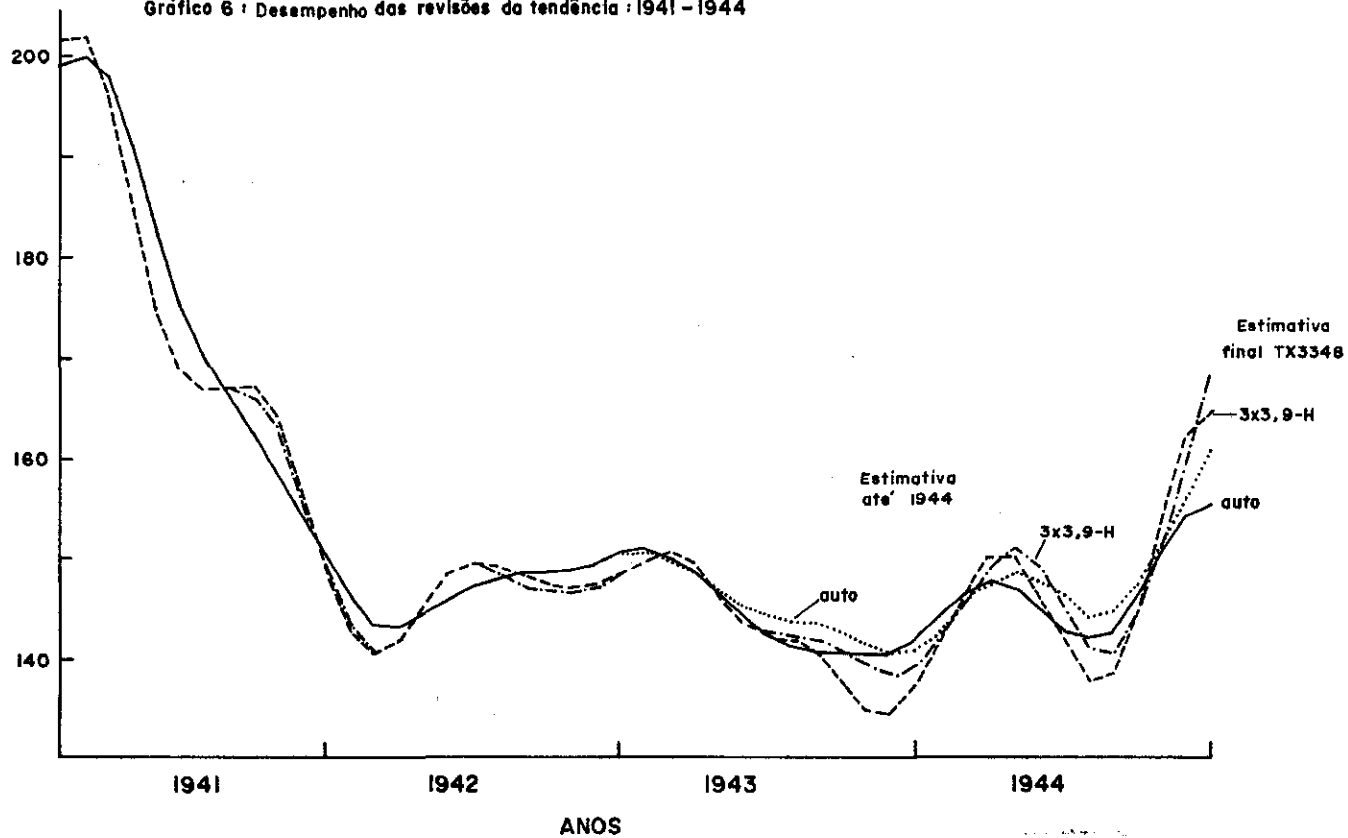


Gráfico 6.a : Desempenho das revisões da tendência nos anos : 1942 - 1943

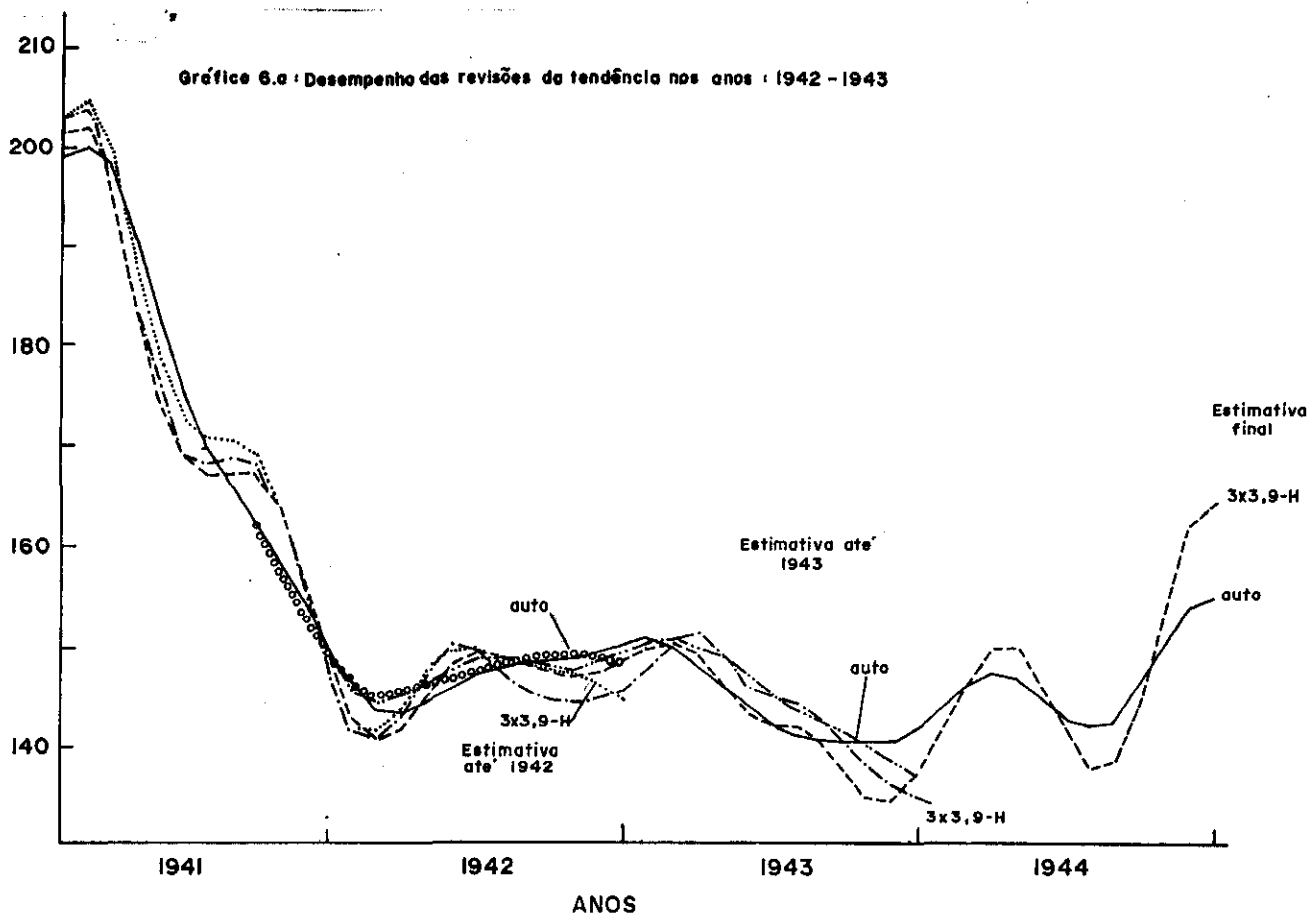


Gráfico 7: Revisão da sazonalidade : 1941-1944, opção automática

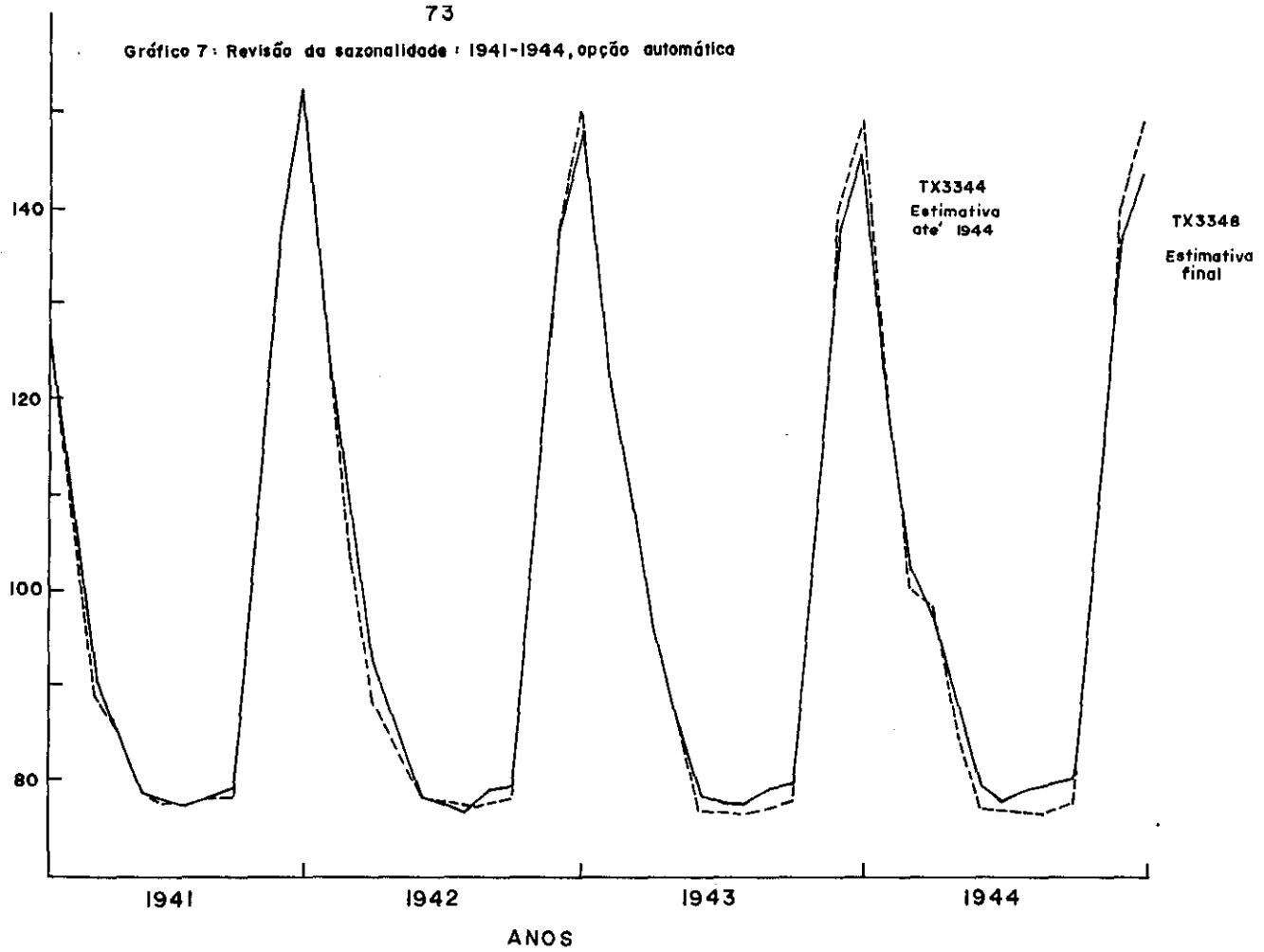
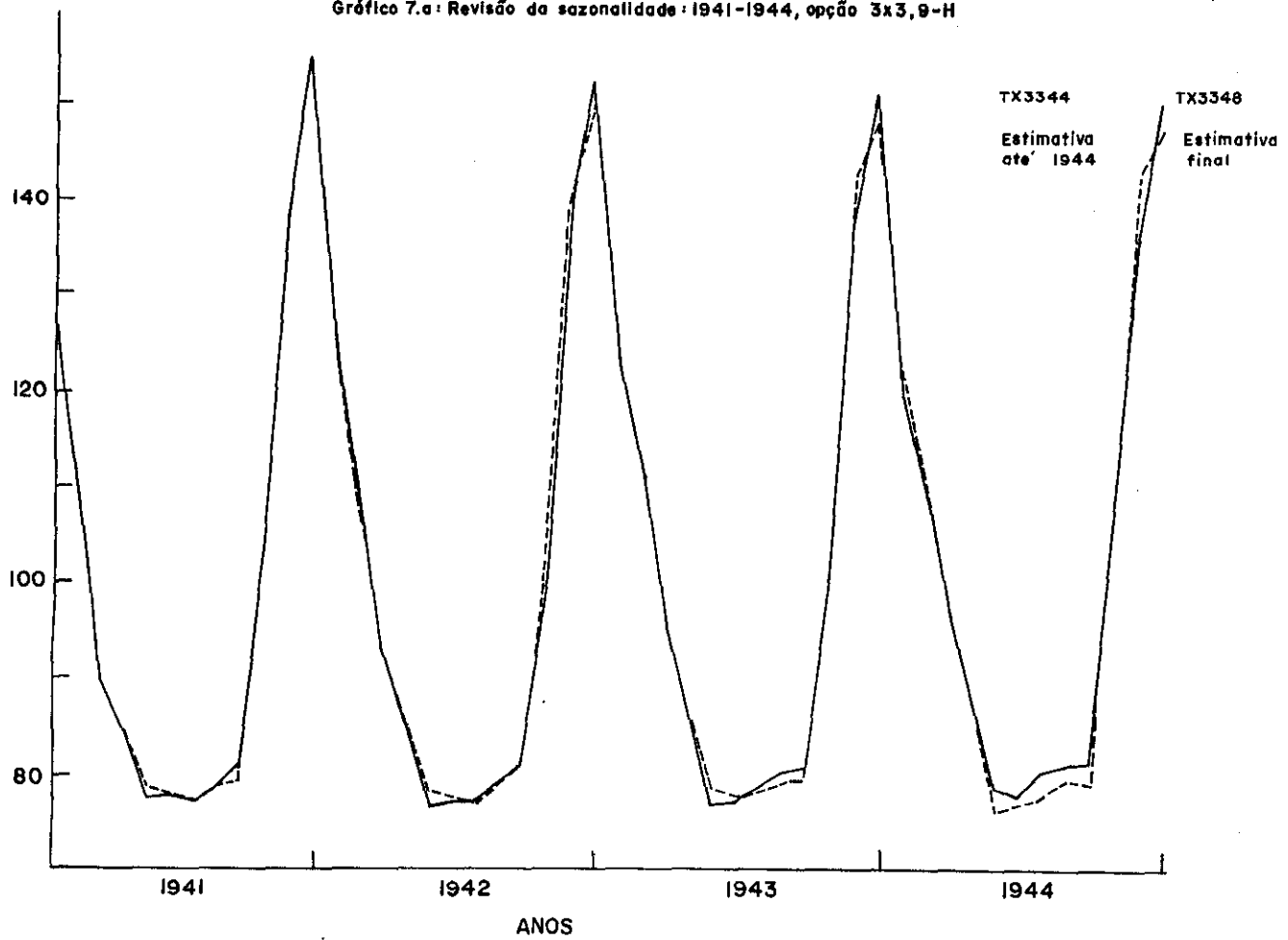
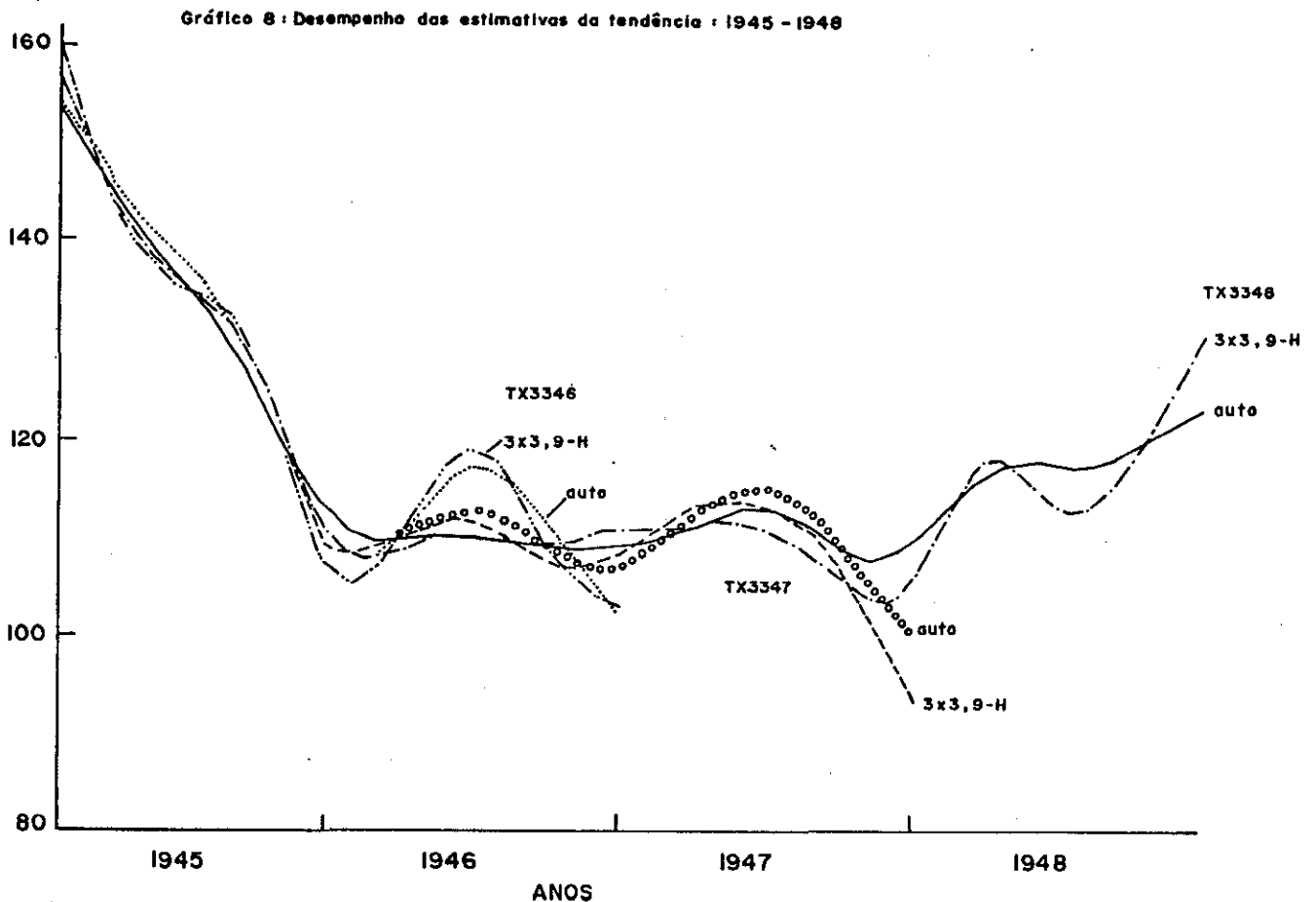


Gráfico 7.a: Revisão da sazonalidade : 1941-1944, opção 3x3,9-H



No gráfico 7, pode-se observar as revisões do fator sazonal para o período de 1941 a 1944. A opção automática, superestima ligeiramente a sazonalidade histórica nos picos e a subestima nas quedas do padrão sazonal; entretanto a opção 3x3,9-H (gráfico 7a), subestima ligeiramente o fator sazonal ao longo de todo o período. Neste caso as revisões da opção automática são sempre maiores que da outra opção.



O gráfico 8 mostra que as estimativas da tendência no período de 1945 a 1948, se afastam nos pontos de inflexão, mais do que no período anterior. Nenhuma das duas opções consegue acompanhar o movimento constante da tendência após a queda de 1945, sendo que a opção dos filtros curtos fornece estimativas mais distantes das estimativas históricas.

O desempenho das estimativas da opção automática do fator sazonal no período 1945-1948, pode ser apreciado no gráfico 9; as revisões para 1945 são pequenas. No gráfico 9a observa-se que as revisões para a opção 3x3,9-H são muito altas (ver tabela 18).

Tabela 18. O erro médio absoluto (MAE) dos dois últimos anos das diversas séries

Série	mês	opção	Tendência atual	Sazonalidade atual	previsão
TX3340		auto, X11	3,34	1,96	3,52
		*automática	4,56	2,06	3,24
		3x3,9-H,X11	3,73	2,29	3,96
		*3x3,9-H	3,88	2,22	3,47
TX3341	Março	automática	3,83	1,47	1,95
		3x3,9-H	2,08	1,72	2,84
	Abril	automática	2,83	2,17	3,78
		3x3,9-H	3,54	2,76	2,48
TX3342		automática	0,92	1,98	3,15
		3x3,9-H	1,71	1,75	3,05
TX3343		auto	1,53	1,79	3,32
		3x3,9-H	2,05	1,57	3,22
TX3344		automática	1,64	1,80	3,45
		3x3,9-H	1,76	1,80	3,27
TX3345		automática	2,35	2,24	3,08
		3x3,9-H	4,66	3,14	5,61

* extrapola o modelo ARIMA 1g(0,1,1)(0,1,1).

Vê-se que as estimativas correntes da tendência possuem um comportamento diferencial após as duas quedas do nível. Na primeira queda pode-se observar que as estimativas correntes de ambas opções conseguem acompanhar o movimento da série após a queda, o que não acontece com nenhuma das duas opções após a segunda queda, cujas estimativas demoram quase um ano para acompanhar este movimento. Dois fatores parecem explicar este comportamento: i) a primeira queda é menor que a segunda e ii) na segunda há uma ligeira mudança no padrão sazonal.

Gráfico 9 : Desempenho das estimativas da sazonalidade :
1945 - 1948, opção automática

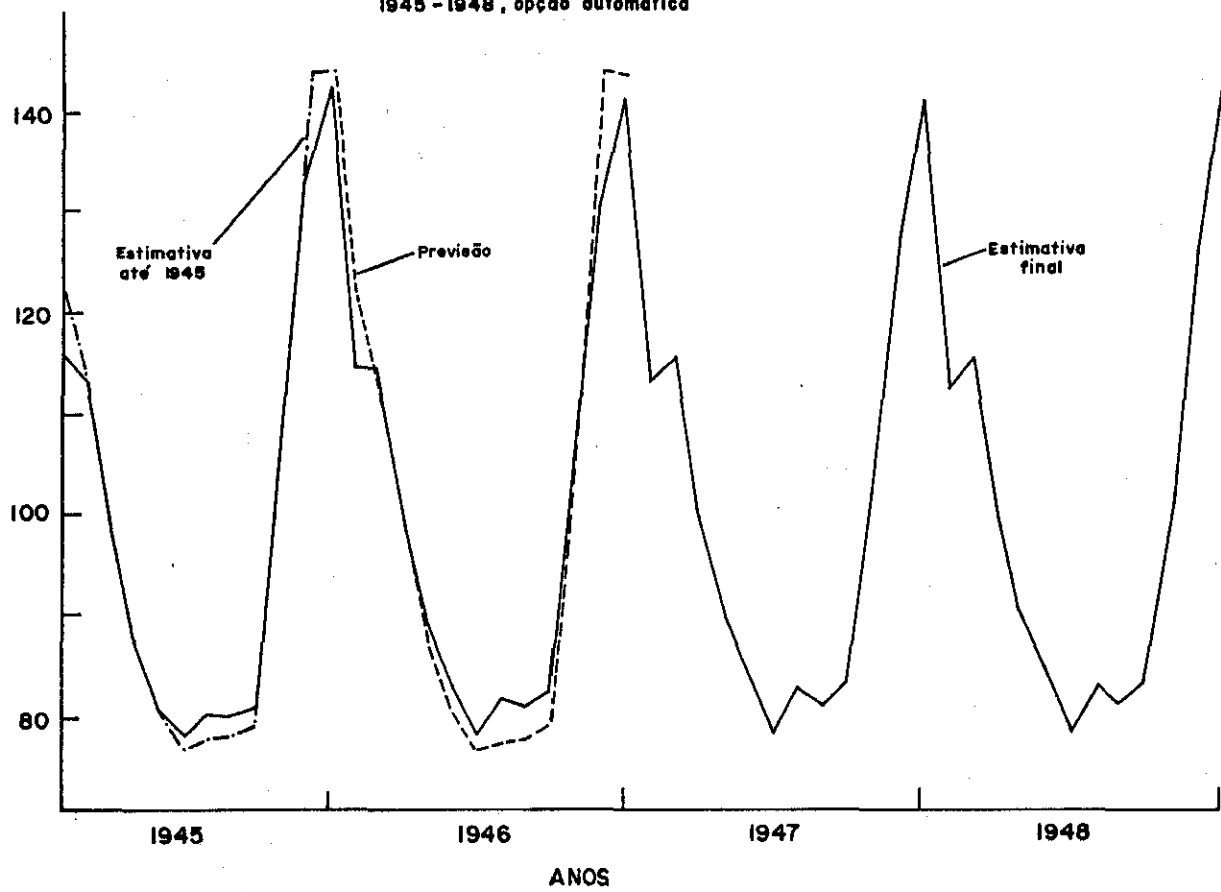
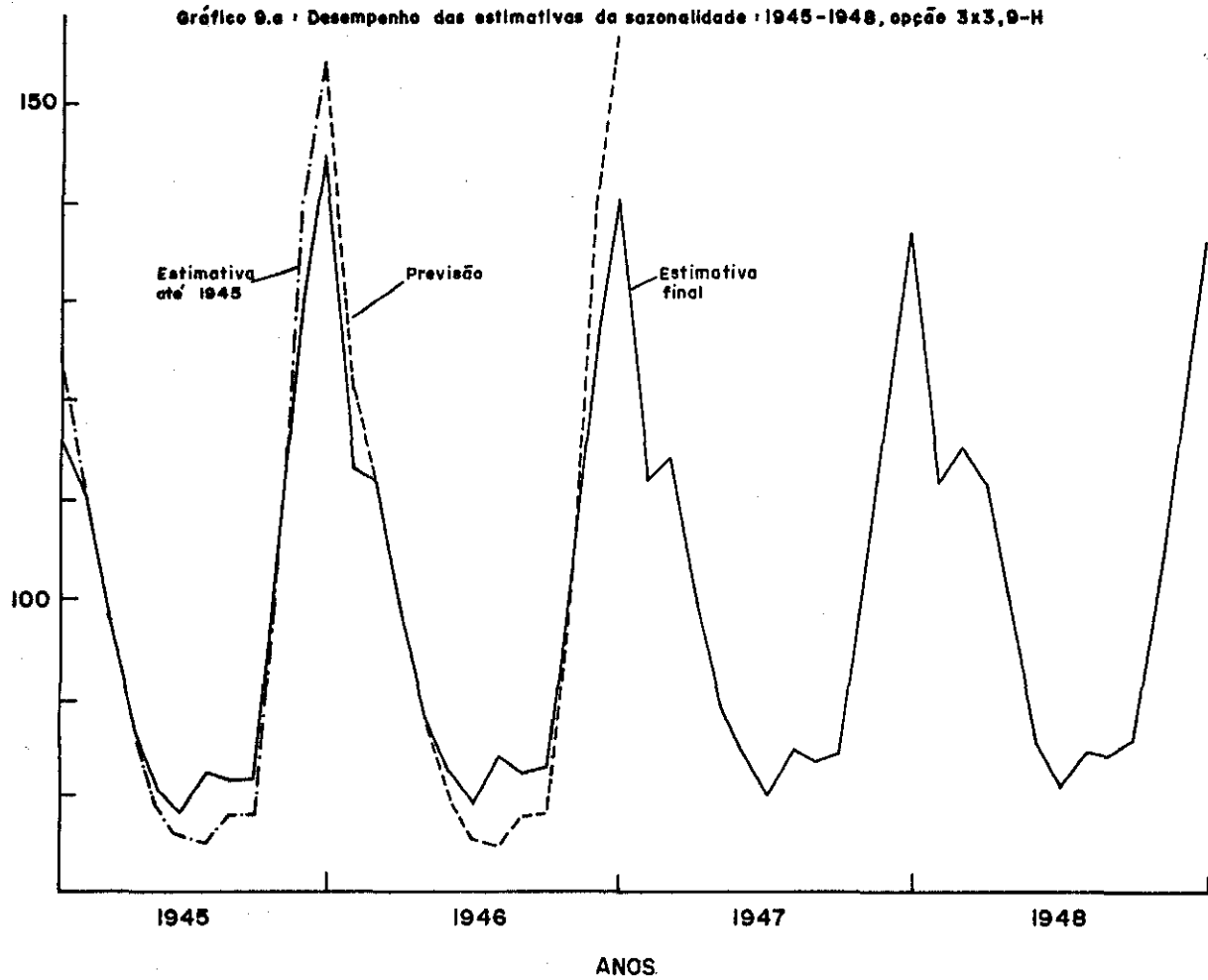


Gráfico 9.a : Desempenho das estimativas da sazonalidade : 1945-1948, opção 3x3,9-H



5.6 Estudo do desempenho do programa no período de 1950 a 1985.

Esta seção apresenta os resultados do desempenho da opção automática e das opções alternativas no período de 1950 a 1985. As tabelas que suportam estes resultados e os respectivos comentários são apresentados nos apêndices de 1 a 6.

Como foi visto anteriormente, a tendência é decrescente até 1960, quase constante de 1961 a 1968, cresce a partir de 1968/69 até 1975 e decrescente até 1985.

Entretanto o padrão sazonal mostra 4 períodos diferentes: i) de 1950 a 1958, caracterizado principalmente pela amplitude sazonal decrescente, ii) de 1959 a 1965, cuja diferença em relação ao padrão anterior reside nos máximos e mínimos relativos que são mais espaçados e mais acentuados, iii) de 1966 a 1974, cuja principal característica é a de passagem e amplitude sazonal ligeiramente crescente e, iv) de 1975 a 1985, caracterizado pela falta de um padrão sazonal dominante.

5.6.1 Análise dos resultados do ajuste da opção automática.

5.6.1.1 Análise do ajuste da opção ARIMA.

Os períodos onde a série inclui os 15 primeiros anos rejeitam os três modelos ARIMA. Isto deve-se a que a amplitude sazonal e a tendência dos 10 primeiros anos são fortemente decrescentes, não conseguindo se compatibilizar com a estrutura da década de 60, embora, ao se ampliar a série até 1975, o modelo $lg(0,1,1) \times (0,1,1)$ seja aceito. Isto ocorre porque o peso relativo destes 10 primeiros anos é menor. A partir de 1960 todas as séries aceitam (com exceção da série TX6068) o modelo $lg(0,1,1)(0,1,1)$ porque nestes períodos a tendência e a amplitude sazonal não apresentam mudanças tão bruscas como no caso anterior, embora haja uma diminuição da influência sazonal.

5.6.1.2 Análise da qualidade do ajuste.

A qualidade do ajuste da opção automática X-11 ARIMA, no período de 1950 a 1975, mostra deficiências na aleatoriedade dos resíduos, falta de suavidade da série sazonalmente ajustada e mostra ainda a existência de sazonalidade móvel (v. tabelas no apêndice 1).

No período de 1950 a 1965, onde se inclui a década de 50 e parte da década de 60, a opção automática X-11 não consegue controlar a participação do componente irregular em relação à sazonalidade, devido a que a variabilidade observada na década de 50 possui um peso maior neste período. Por outro lado, a série sazonalmente ajustada possui um MCD igual a seis, o que indica que as variações da série sazonalmente ajustada são devidas principalmente às variações do componente irregular. Há presença de sazonalidade móvel.

No período de 1950 a 1968, que inclui 3 anos a mais que o período anterior, a qualidade do ajuste da opção automática X-11 se deteriora em relação à qualidade do período anterior. É introduzida autocorrelação de lag 12, a participação do componente irregular aumenta, embora a presença de sazonalidade móvel desapareça.

A qualidade do ajuste X-11 no período de 1960 a 1968 (onde a tendência se mantém quase constante, com mudanças no padrão sazonal), é deficiente, pois mostra autocorrelação nos resíduos, série sazonalmente ajustada influenciada pela variação do componente irregular e alta participação do componente irregular em relação aos outros componentes. Neste período o desempenho é o mais deficiente.

No período de 1960 a 1975, onde se inclui a década de 60, caracterizada pela quase constância da amplitude sazonal e tendência, e parte da década de 70, caracterizada pelo crescimento relativo de ambos, mostrando mudanças no nível da tendência e no padrão sazonal, esta última com presença de defasagem, a opção automática mostra presença de autocorrelação de lag 12 nos resíduos e uma alta participação do componente irregular comparado com os outros componentes.

O período de 1970 a 1985 possui tendência e amplitude sazonal ligeiramente crescentes na primeira parte da década de 70, e a partir de 1975, uma tendência fortemente decrescente, sem possuir padrão sazonal predominante. Neste período, a qualidade do ajuste X-11 ARIMA não é satisfatória, pois apresenta problemas de autocorrelação nos resíduos, alta participação do irregular comparado com os outros componentes e falta de suavidade da série sazonalmente ajustada.

Finalmente, na série no período de 1960 a 1985, a qualidade do ajuste X-11 ARIMA mostra uma alta autocorrelação e ligeira periodicidade, alta participação do irregular comparado com a sazonalidade, e falta de suavidade na série sazonalmente ajustada.

Dos resultados anteriores pode-se concluir que a opção automática, em geral, não tem bom desempenho. A qualidade do ajuste depende fundamentalmente do padrão sazonal. Quando este é bem comportado, tem-se excelentes resultados, mas quando o padrão muda ao longo do período, a qualidade sofre sérios problemas. A tendência também tem um papel importante, mas esta é mais fácil de ser controlada. Tendo em vista tais resultados, procurou-se opções alternativas a fim de aumentar a qualidade do ajuste.

5.6.2 Desempenho das opções alternativas.

Dentre as opções alternativas disponíveis para melhorar a qualidade do ajuste da série na maioria dos períodos em estudo trabalhou-se, principalmente com os filtros, uma vez que na maioria dos casos, a opção automática escolhe filtros de comprimentos médios. As séries que incluem o ano de 1968 não foram contempladas com a análise acima mencionada, pois, neste ano, a análise foi feita sobretudo visando as especificidades do período.

5.6.2.1 Discussão e escolha da melhor opção para a série no período de 1950 a 1975: TX5075, X-11 ARIMA.

A qualidade do ajuste da opção automática X-11 ARIMA no período de 1950 a 1975 mostra deficiências na aleatoriedade dos resíduos e falta de suavidade na série sazonalmente ajustada. As tabelas e os respectivos comentários para as opções alternativas são apresentadas no apêndice 2.

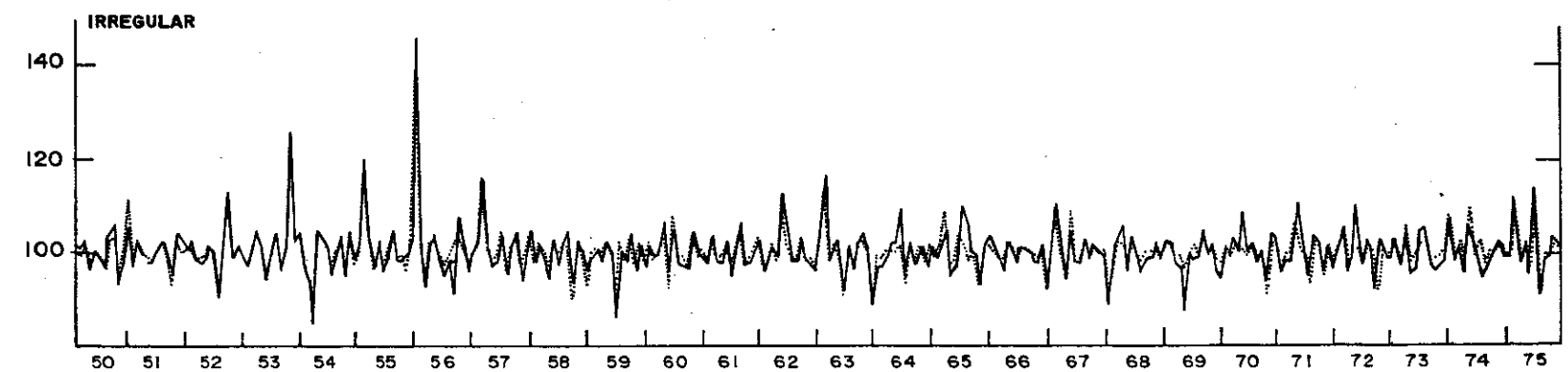
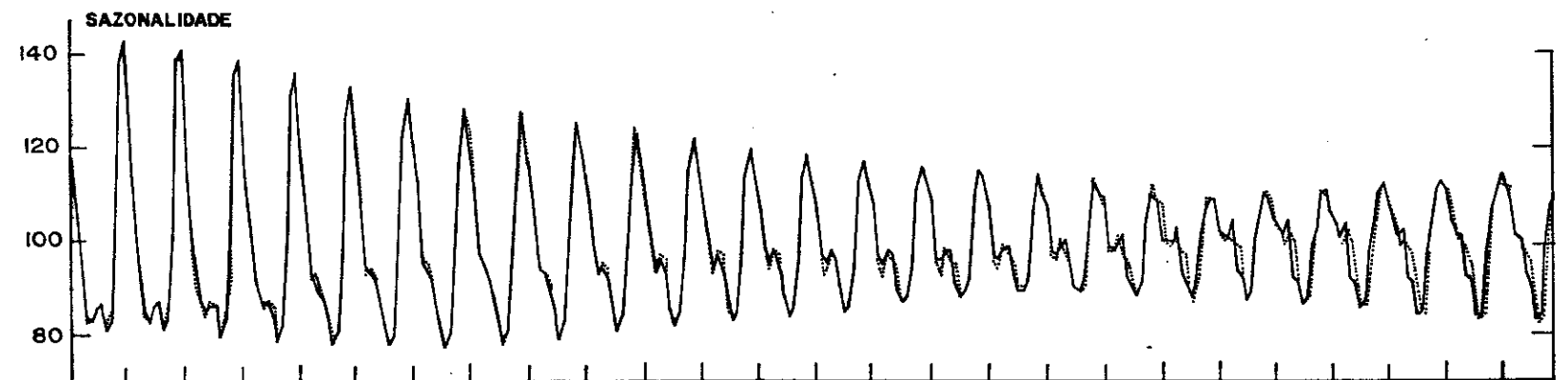
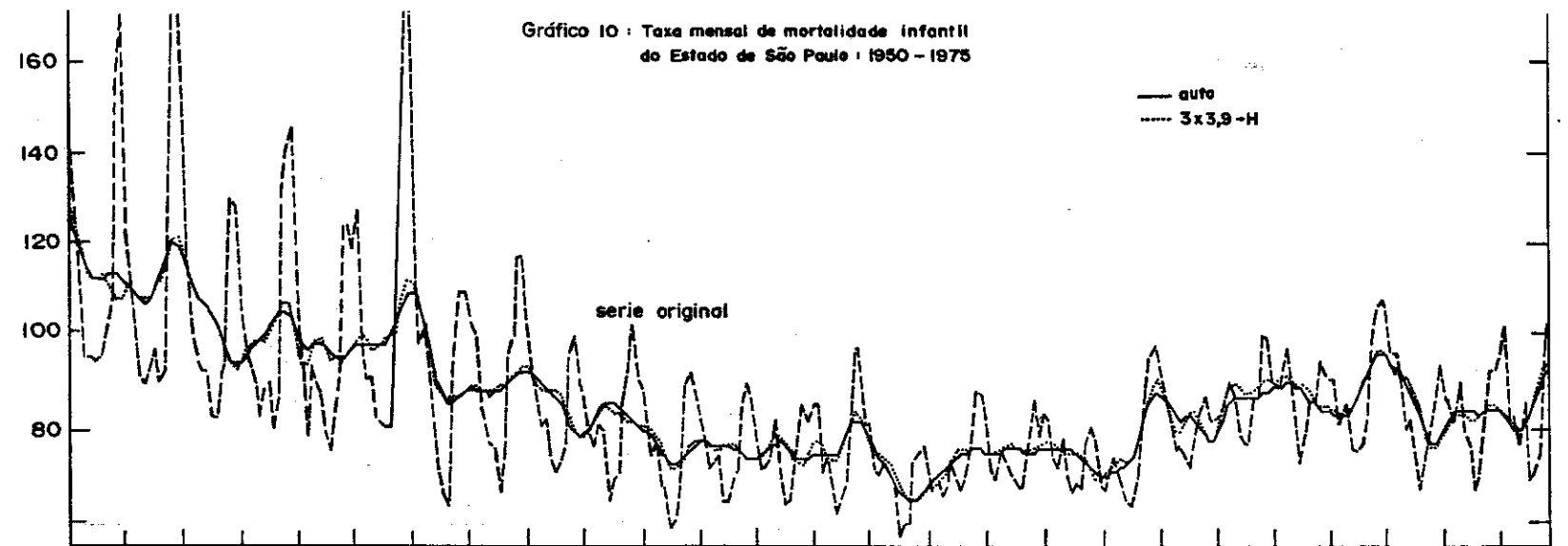
A autocorrelação nos resíduos não é eliminada ao se usar as opções alternativas. As opções que utilizam os filtros mais curtos conseguem suavizar melhor a série sazonalmente ajustada e melhorar a qualidade global do ajuste. Ao contrário, a manipulação dos valores extremos introduz periodicidade.

A melhoria na qualidade do ajuste deve-se fundamentalmente ao uso do filtro 9-H para a tendência, pois a opção automática usa o filtro 13-H e, ao se usar o filtro 23-H a qualidade se deteriora substancialmente. Quanto aos filtros do fator sazonal, a opção que utiliza o filtro 3x9 introduz periodicidade nos resíduos, entretanto, os filtros mais curtos conseguem melhorar a qualidade.

Embora nenhuma opção consiga eliminar a alta autocorrelação dos resíduos, as opções dos filtros mais curtos fornecem uma qualidade melhor que a da automática. No caso, tanto faz o uso das opções: 3x3,9-H ou 3x5,9-H. Estas duas opções pelo menos fornecem ajustes mais suaves que os da opção automática.

Estes resultados, em parte, eram esperados porque os filtros 3x3 e 3x5 são mais flexíveis para acompanhar as mudanças no padrão sazonal que ocorrem neste período (v. gráfico 10). Estas mudanças se explicam porque: i) a década de 50 é caracterizada por uma amplitude sazonal decrescente; ii) a década de 60 que tem um padrão sazonal diferente da década anterior e iii) parte da década de 70 é caracterizada pela presença de defasagem e amplitude sazonal ligeiramente crescente. O melhor desempenho mostrado pelos filtros mais curtos para o fator sazonal se deve, principalmente, ao fato destes acompanharem melhor as mudanças que ocorrem neste período, porém, o componente irregular é demasiado alto para que estes filtros possam ter o desempenho esperado.

No gráfico 10 pode-se observar o desempenho das estimativas da tendência, da sazonalidade e do componente irregular da opção automática e da opção 3x3,9-H. Quanto à tendência, pode-se dizer que a discrepância entre as duas opções reside, principalmente, nos pontos de inflexão, onde a opção do filtro 9-H acentua as estimativas nos picos positivos e diminui nos picos negativos. Quanto à sazonalidade, as opções mostram discrepâncias nas estimativas nos últimos anos. A opção que usa o filtro 3x3 mostra defasagem de quase um mês em relação às estimativas da opção automática.



ANOS

5.6.2.2 Discussão e escolha da melhor opção para a série no período de 1950 a 1965: TX5065, X-11.

No período de 1950 a 1965, a opção automática X-11 não consegue controlar a participação do componente irregular em relação a participação da sazonalidade e apresenta falta de suavidade na série sazonalmente ajustada. As tabelas e os respectivos comentários estão apresentados no apêndice 3.

A fim de controlar o componente irregular, tenta-se a opção de manipulação de valores extremos. Ao se fazer isto, M2 continua falhando, e além disto o M5, que está relacionado com o MCD, também falha. Isto mostra que há um efeito contrário ao desejado, portanto a opção é descartada.

A opção 3x3 melhora a qualidade geral do ajuste. Nota-se que a opção escolhe o filtro 13-H para a tendência, mas introduz autocorrelação de lag 12, indicando presença de sazonalidade residual; portanto esta opção fica descartada. Ao se combinar o filtro 3x3 com o filtro 9-H para a tendência e com a manipulação de valores extremos se consegue melhorar a qualidade, mas não se consegue eliminar a autocorrelação de lag 12. A opção 3x5,9-H, também não consegue este objetivo.

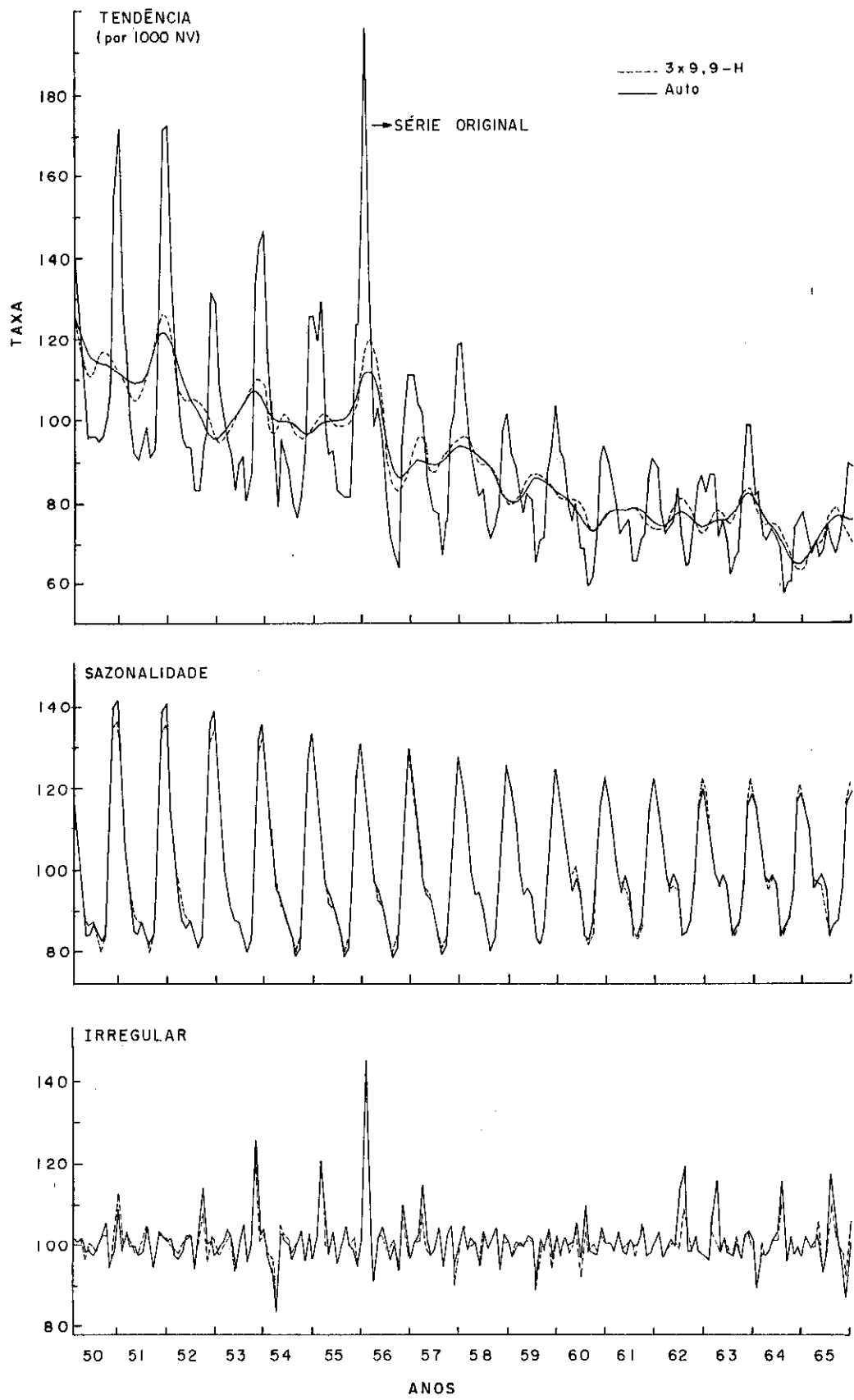
A opção, 3x9,9-H, consegue eliminar a autocorrelação dos resíduos, fornece uma série sazonalmente ajustada mais suave e melhora substancialmente a qualidade global do ajuste, candidatando-se como a melhor opção para a série neste período. Isto se deve fundamentalmente ao uso do filtro 3x9 (o de mais amplo comprimento), que consegue captar o movimento decrescente da amplitude sazonal que domina este período, o que pode ser confirmado pela diminuição dos valores das estatísticas M0-M11.

Para este período, a opção 3x9,9-H consegue uma substancial melhoria na qualidade do ajuste eliminando quase todas as deficiências mostradas pela opção automática, tornando-se a "melhor opção" para este período. A manipulação dos filtros mais curtos se traduz em forte autocorrelação.

O bom desempenho da opção 3x9,9-H se deve principalmente ao uso do filtro 3x9 para o fator sazonal. Este consegue captar o movimento do padrão sazonal deste período, que inclui a década de 50 e parte da década de 60, dominado pela amplitude sazonal decrescente (v. gráfico 11). Este resultado é totalmente oposto ao anterior, o que fica explicado pela maior abrangência do período anterior, incluindo outras mudanças no padrão sazonal.

O gráfico 11 mostra as estimativas dos componentes da série pela opção automática e pela opção 3x9,9-H. Quanto à tendência, as opções mostram discrepância significativas nos pontos de inflexão, principalmente, nos primeiros anos, devido à presença de fortes valores extremos. As estimativas da sazonalidade mostram diferenças na amplitude sazonal. A opção do filtro 3x9 fornece estimativas com amplitudes menores que as da opção automática nos primeiros anos, e maiores nos últimos anos. Isto deve-se ao fato do filtro 3x9 ser menos influenciado por valores extremos.

Gráfico 11 Taxa mensal de mortalidade infantil do Estado de São Paulo: 1950 - 1965



5.6.2.3 Discussão e escolha da melhor opção para a série no período de 1960 a 1975: TX6075, X-11 ARIMA.

No período de 1960 a 1975, onde se inclui a década de 60, caracterizada pela quase constância da amplitude sazonal e parte da década de 70, caracterizada pelo crescimento deste e por mudanças no padrão sazonal com presença de defasagem (v. gráfico 12), a opção automática mostra presença de autocorrelação nos resíduos, alta participação do componente irregular comparado com os outros componentes e falta de suavidade do ajuste. As tabelas podem ser observadas no anexo 4.

A opção $3 \times 3,9-H$ introduz periodicidade nos resíduos, não elimina a autocorrelação e não consegue captar satisfatoriamente o movimento sistemático da sazonalidade.

A opção $3 \times 5,9-H$ apresenta deficiências na aleatoriedade e periodicidade dos resíduos, embora melhore substancialmente a qualidade do ajuste geral. A opção $3 \times 9,9-H$ tem mais sucesso, pois consegue reduzir a autocorrelação dos resíduos e se transforma numa opção alternativa à opção automática.

Neste caso, novamente a opção $3 \times 9,9-H$ se constitui numa opção alternativa à opção automática, pois reduz significativamente a autocorrelação dos resíduos, elimina as outras deficiências e aumenta significativamente a qualidade do ajuste. Isto ocorre porque a série neste período apresenta um padrão sazonal irregular. Novamente, neste caso o uso dos filtros mais curtos para o fator sazonal tem um efeito contraproducente introduzindo periodicidade e autocorrelação nos resíduos, onde o filtro 3×3 não consegue acompanhar o movimento aleatório dos últimos anos.

O gráfico 12 mostra as estimativas dos componentes para as opções automática e $3 \times 9,9-H$. Os comentários sobre as estimativas da tendência nos períodos anteriores são também válidos neste período. As diferenças entre as estimativas da sazonalidade são muito pequenas.

5.6.2.4 Discussão e escolha da melhor opção para a série no período 1970 a 1985: TX7085, X-11 ARIMA.

No período de 1970 a 1985, a qualidade do ajuste X-11 ARIMA é insatisfatório, pois apresenta problemas de autocorrelação nos resíduos, alta participação do componente irregular comparado com os outros componentes e falta de suavidade da série sazonalmente ajustada. As tabelas que suportam esta análise são apresentadas no anexo 5.

A opção $3 \times 3,9-H$ consegue uma melhora substancial na qualidade do ajuste, tanto na estrutura dos resíduos como nas estatísticas de controle de qualidade, mas não consegue controlar a variabilidade aleatória da sazonalidade ao longo da série. A combinação desta opção com a manipulação dos valores extremos não consegue uma melhoria significativa.

Gráfico 12. Taxa mensal de mortalidade infantil do Estado de São Paulo : 1960-1975

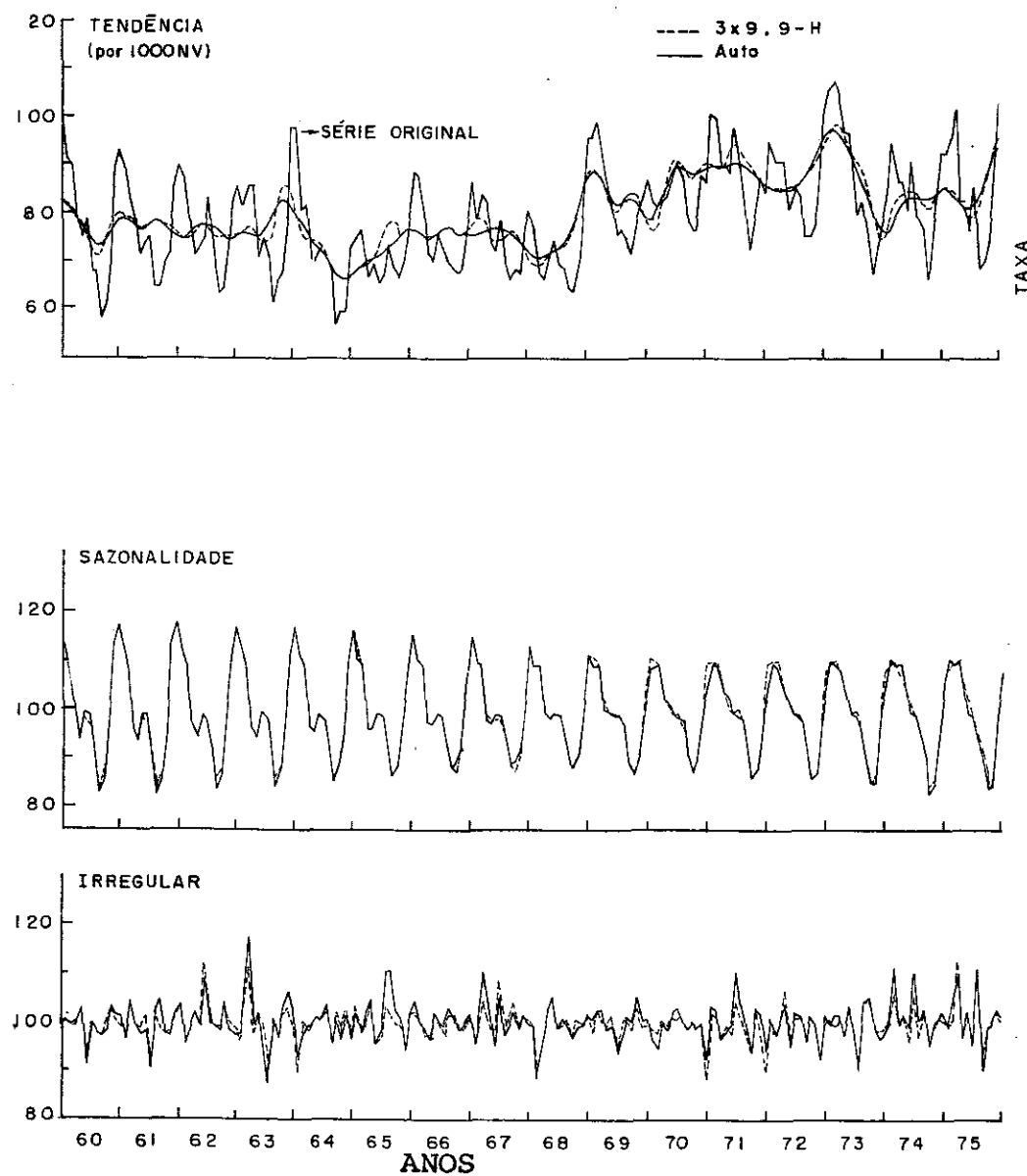
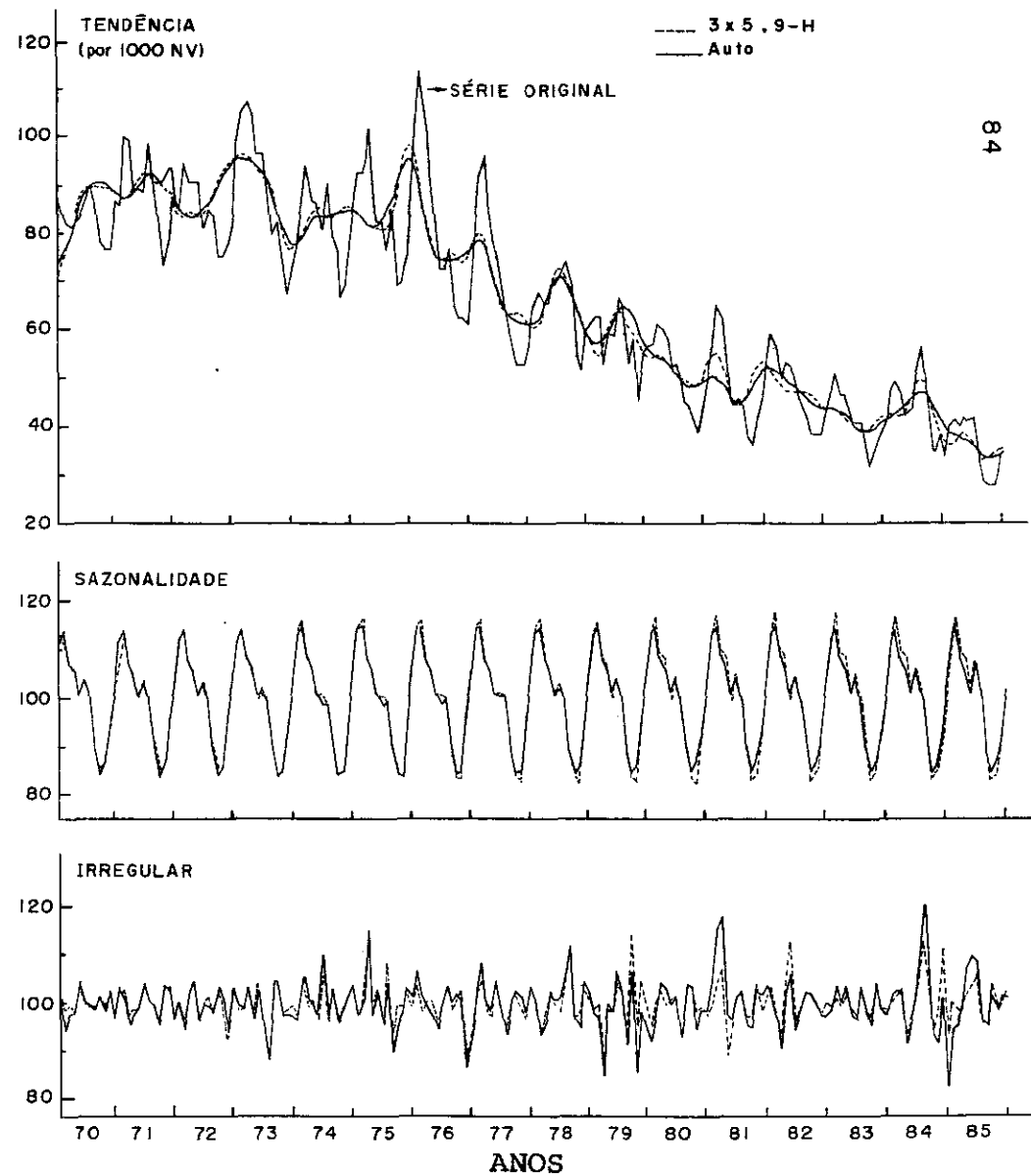


Gráfico 13 Taxa mensal de mortalidade infantil do Estado de São Paulo: 1970-1985



As opções 3x5,9-H e 3x9,9-H conseguem uma qualidade superior à opção automática, mas nenhuma consegue eliminar as deficiências mostradas nesta última, quais sejam autocorrelação e periodicidade nos resíduos. A combinação da opção 3x5,9-H com a manipulação de valores extremos (3x5,9-H,s/out) se constitui numa opção alternativa à opção automática; mesmo assim esta opção não consegue eliminar completamente a autocorrelação e periodicidade dos resíduos.

A falta de uma opção que forneça um bom ajuste, se deve fundamentalmente à variabilidade do padrão sazonal indicada pelo alto valor de I/S (o mais alto de todos os períodos). Este período se caracteriza, principalmente, pela falta de um padrão sazonal predominante.

O gráfico 13 apresenta as estimativas dos componentes pelas opções automática e 3x5,9-H. Neste gráfico pode-se observar que a opção que usa o filtro 3x5 fornece estimativas de maior amplitude para o fator sazonal, principalmente nos últimos anos. Também se pode observar que, neste período, as estimativas do componente irregular são superiores às estimativas em outros períodos.

5.6.2.5 Discussão e escolha da melhor opção para a série no período de 1960 a 1985: TX6085, X-11 ARIMA.

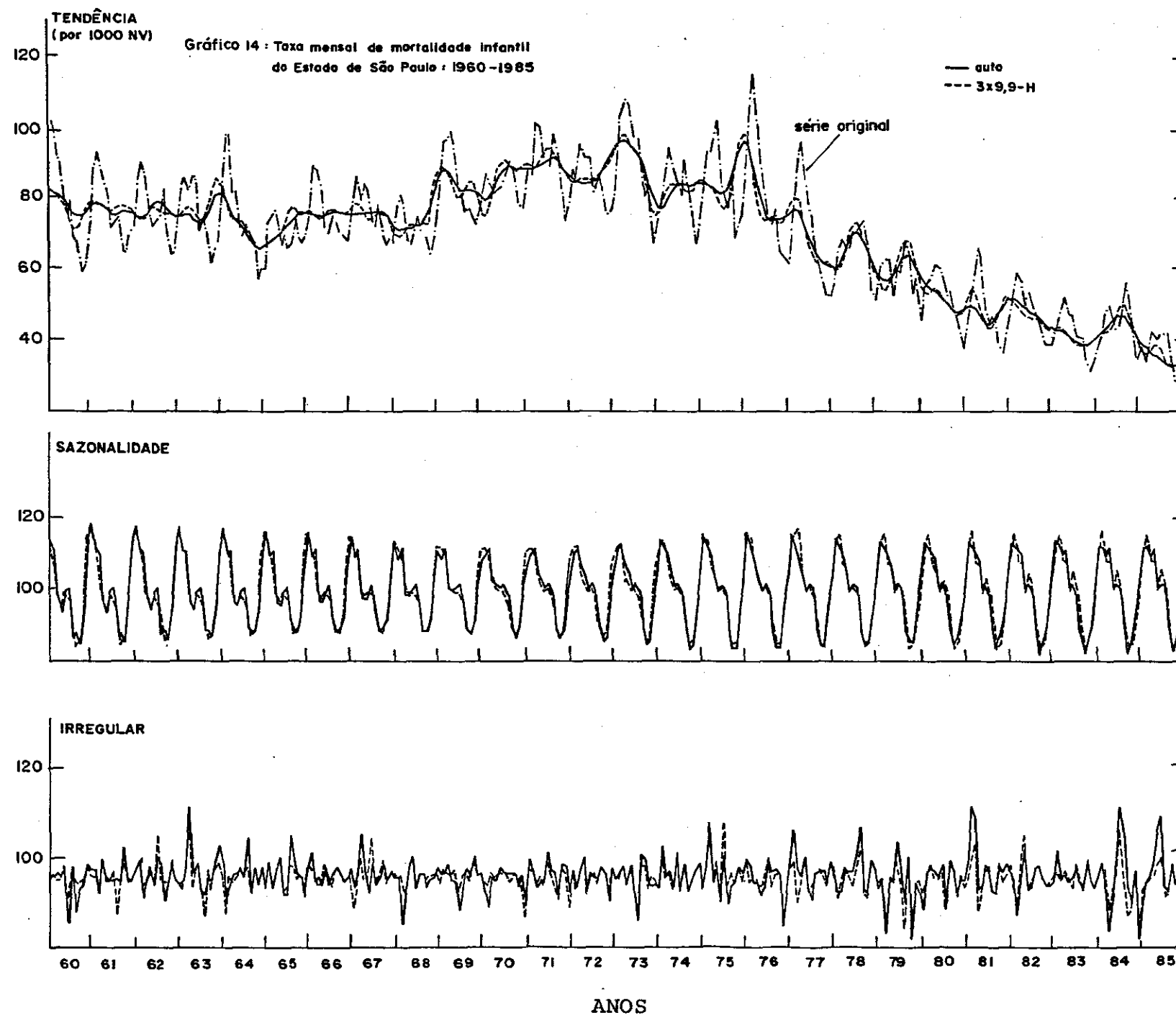
Na série no período de 1960 a 1985, a qualidade do ajuste X-11 ARIMA mostra uma alta autocorrelação e ligeira periodicidade nos resíduos, alta participação do componente irregular comparado com a sazonalidade, e falta de suavidade na série sazonalmente ajustada. As tabelas e os respectivos comentários das opções alternativas são apresentadas no anexo 6.

Novamente, o filtro mais adequado para a tendência é o filtro mais curto, isto é, o 9-H. O filtro 23-H fica descartado pela sua ineficiência para seguir os movimentos da série. Dado que o melhor filtro para a tendência é o 9-H, manipula-se o filtro para a sazonalidade.

A opção 3x3,9-H, apresenta problemas nas estatísticas de controle de qualidade M0 e M10, explicadas fundamentalmente pelo uso do filtro 3x3 para a sazonalidade, mostrando sua incapacidade de filtrar a variabilidade aleatória do movimento sazonal; portanto é uma opção descartada.

A opção 3x5,9-H consegue melhorar a qualidade do ajuste, mas não consegue eliminar a autocorrelação. Isto se consegue ao se utilizar a opção 3x9,9-H, mas esta introduz periodicidade. De qualquer maneira estas duas opções superam a qualidade da opção automática, tornando-se em opções alternativas a esta, embora não consigam controlar totalmente as deficiências mostradas através dos resíduos.

No gráfico 14 pode-se observar as estimativas dos componentes pelas opções automática e 3x9,9-H. As estimativas da sazonalidade mostram diferenças, principalmente, na amplitude sazonal dos últimos anos e nas estimativas do máximo relativo nos primeiros e últimos anos. A opção automática fornece estimativas superiores às estimativas da outra opção nos máximos relativos dos primeiros anos, e menores nos últimos anos.



5.6.3 Estudo das revisões para o período de 1968 a 1969.

Este período foi escolhido para uma análise mais detalhada, pela importância destes anos. Neste período o Brasil entra num processo de desenvolvimento acelerado, bastante controvertido, que leva a um aumento da taxa anual de mortalidade infantil. Dada a sensibilidade deste período, é analisada a existência de alguma diferença nos resultados do ajuste ao se utilizar diferentes opções.

Para se fazer o estudo das revisões, tomou-se como estimativas finais ou históricas as estimativas da série no período de 1950 a 1975. Como foi visto na seção 5.6.2.1, uma opção alternativa à opção automática é a opção 3x3,9-H.

O gráfico 15 mostra o desempenho das estimativas da tendência das duas opções no período de 1966 a 1969. As duas opções, a automática e a dos filtros curtos, fornecem estimativas históricas muito próximas ao longo de quase todo o período, com exceção dos meses de maio, junho e julho de 1969, onde a opção 3x3,9-H se afasta ligeiramente da opção automática. As estimativas correntes de ambas as opções conseguem acompanhar o movimento histórico muito bem, já que as revisões são insignificantes. As duas opções levam a uma mesma conclusão.

Neste gráfico pode-se ainda observar que a tendência permanece quase constante durante os anos de 1966 e 1967; em 1968 o nível decresce ligeiramente nos primeiros meses aumentando logo a seguir. Existe um aumento acelerado da mortalidade infantil a partir de setembro de 1968, atingindo o máximo em março de 1969. A partir deste mês, existe uma tendência de baixa, logo acompanhada de novas altas. De forma geral, a partir de 1968 a taxa de mortalidade infantil sobe para um novo patamar (v. também gráfico 16). O aumento da taxa de mortalidade infantil, a partir de 1968, tendo em vista a conjuntura econômica deste período, poderia sugerir que o processo de desenvolvimento adotado trouxe como consequência o incremento da mortalidade infantil.

O gráfico 16 mostra o desempenho das estimativas históricas do fator sazonal. As duas opções diferem na captação do máximo relativo. A opção automática demora um mês em relação à opção dos filtros curtos, embora nem sempre isso seja uma desvantagem, pois os máximos relativos dos anos 67 e 68 são captados por esta opção. Entretanto, os anos 66 e 69 são descritos de melhor forma pela opção dos filtros curtos. No caso, nenhuma das duas mostra superioridade em relação à outra.

As revisões do fator sazonal para a opção automática são apresentadas no gráfico 17. As maiores revisões ocorrem nos últimos anos, principalmente nos máximos relativos e absolutos. Nos picos máximos, as revisões superestimam as estimativas históricas, entretanto, nos mínimos relativos, as subestimam. Algo similar ocorre com as revisões da opção dos filtros curtos, (v. gráfico 18), embora as revisões desta opção, sejam menores. Quanto à sazonalidade, pode-se observar que a estrutura sazonal e a amplitude não mudam ao longo deste período, o que indica que o aumento da taxa de mortalidade infantil não foi motivado por fatores externos, como por exemplo, condições climáticas atípicas.

Gráfico 15: Desempenho das estimativas da tendência:
1968 - 1969

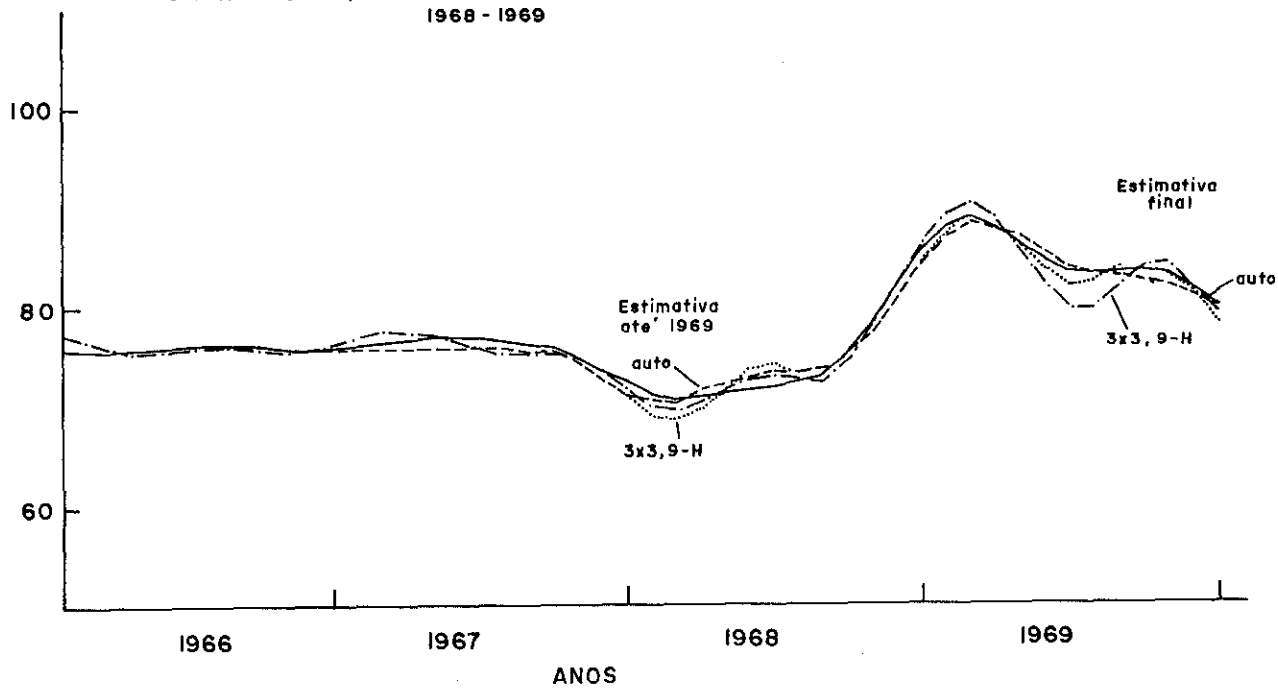


Gráfico 16: Desempenho das estimativas do fator sazonal:
1968 - 1969

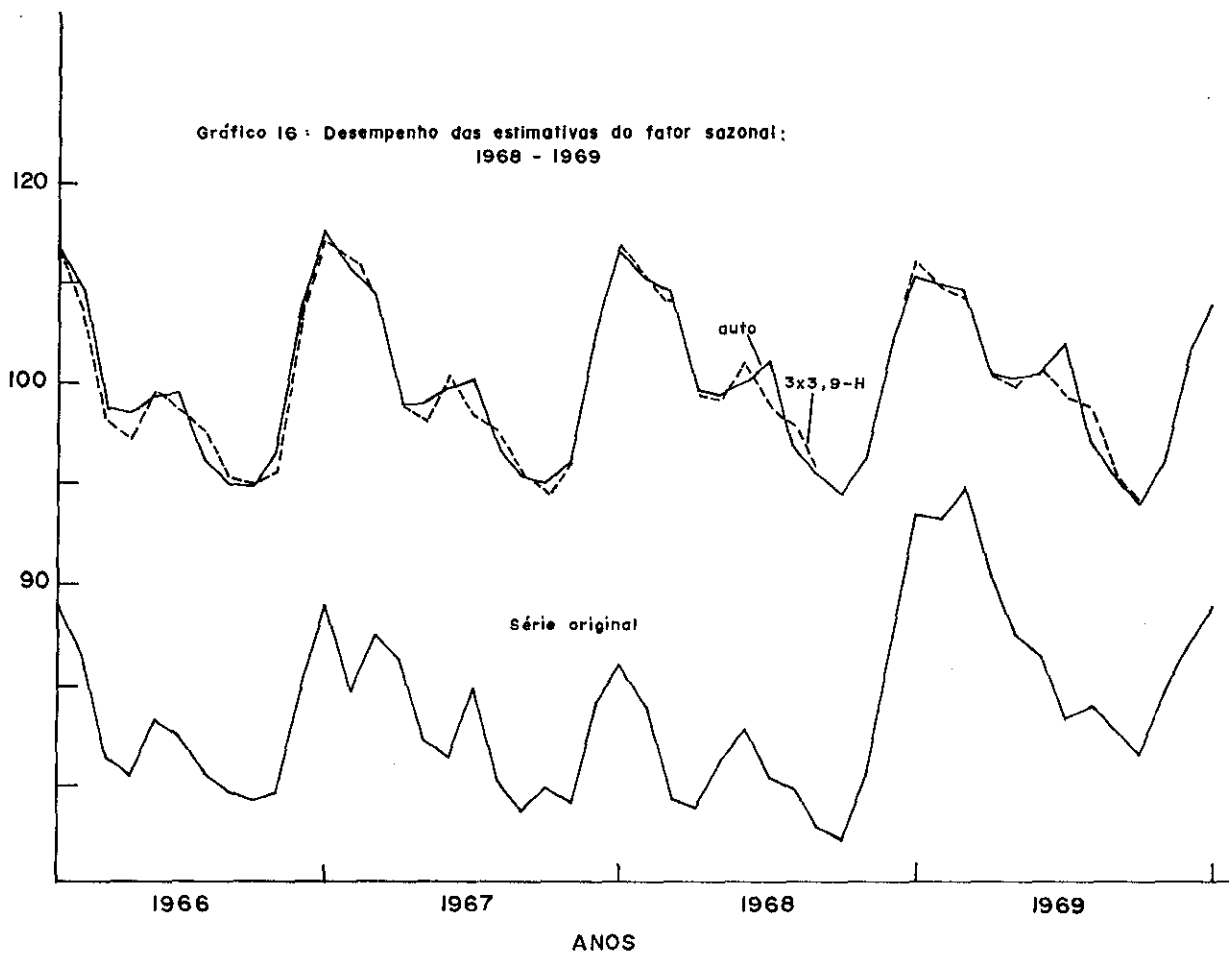


Gráfico 17: Desempenho das revisões do fator sazonal
para a opção automática

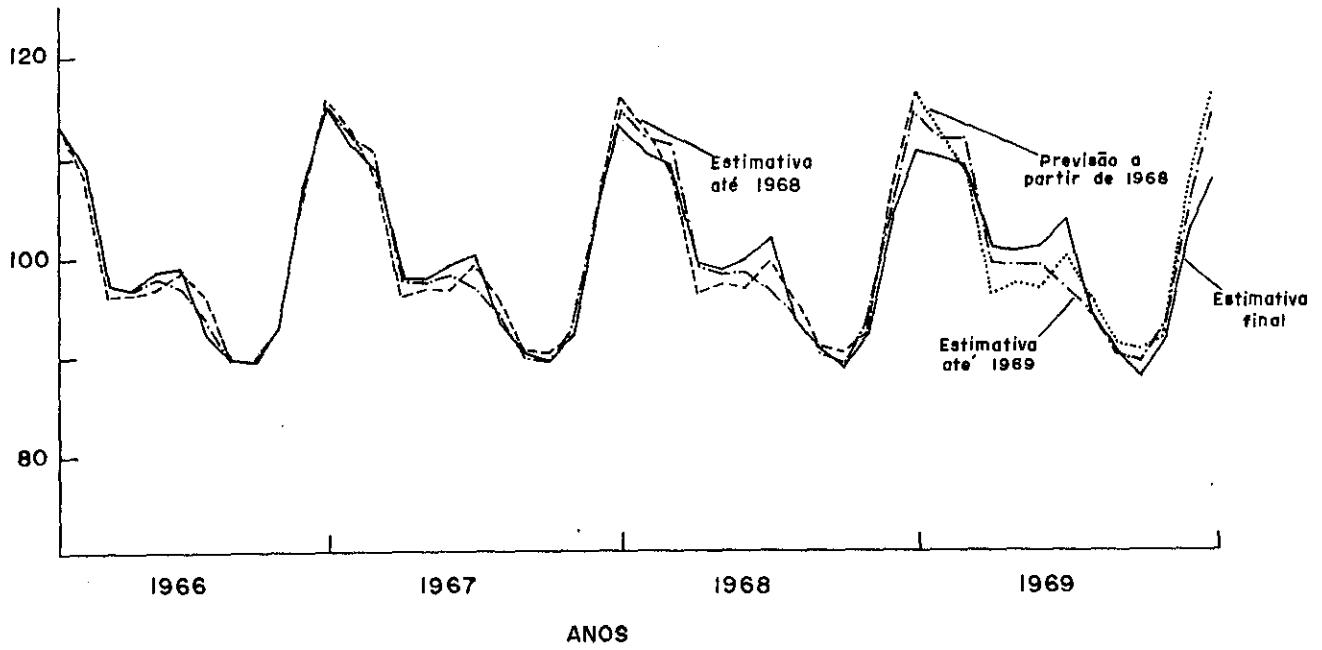
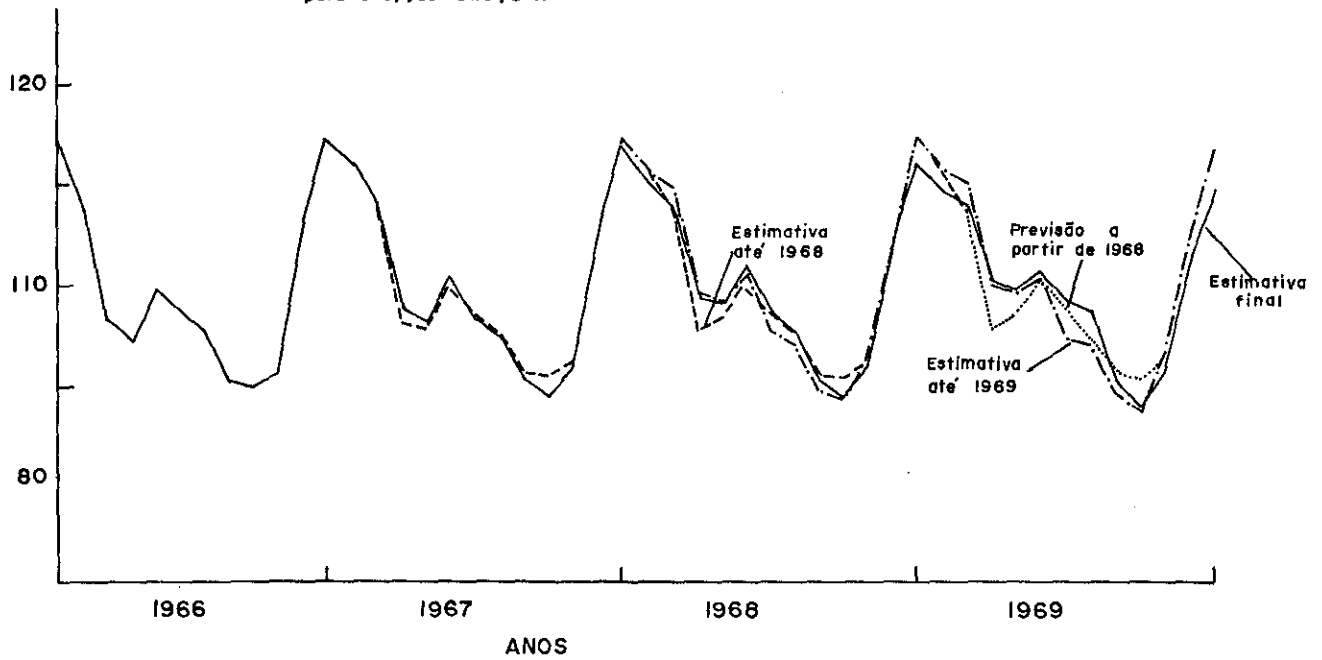


Gráfico 18: Desempenho das revisões do fator sazonal
para a opção 3x3, 9-H



5.7 Análise dos componentes.

Na análise dos componentes deve-se levar em consideração vários aspectos, já mencionados anteriormente e que de certo modo limitam a análise. Estes são: i) a limitação da série (falta de informação em 49), ii) a limitação do programa (ajusta no máximo 30 anos), e, iii) falta de referências bibliográficas sobre a influência das estações e do clima sobre a taxa de mortalidade infantil, ou estudos feitos sobre a taxa mensal.

A fim de uniformizar a análise, adotou-se a opção automática como a opção para decompor a série, e, os componentes são analisados nos períodos de 1933 a 1948 e de 1950 a 1985. Os componentes neste último período são analisados segundo dois ajustes. De 1950 a 1975 segundo o ajuste deste período e de 1975 a 1985 segundo o ajuste de 1960 a 1985.

5.7.1 Análise dos componentes no período de 1933 a 1948.

O gráfico 3 (pag. 68) mostra a tendência e a sazonalidade no período de 1933 a 1948.

A tendência mostra 3 níveis diferenciados, o primeiro nível abrange o período 33-40, onde a taxa mostra um nível médio mais alto de todo o período; o segundo nível vai de 1942 a 1944, e é intermediário entre os outros dois; o último nível vai de 1946 a 1948 e possui a taxa média mais baixa do período. Por outro lado pode-se observar também, a presença de pequenos ciclos com amplitude de mais ou menos 3 anos, onde o maior ponto de inflexão ocorre no início de 1941. Infelizmente não se tem encontrado referências bibliográficas sobre as possíveis causas destes ciclos.

Em relação à sazonalidade observa-se que o máximo é atingido em todos os anos em dezembro, enquanto o mínimo é atingido em agosto de 1933 até 1938, em julho de 1939 a 1940, e em junho de 1944 a 1948. Por outro lado, a partir de 1946 até 1948 observa-se um sutil máximo relativo no mês de fevereiro. A amplitude sazonal de 1933 é de 72,5, e, a partir deste ano, a amplitude cresce até 1939, atingindo o máximo valor de todo o período, 79,1. A partir de 1940, onde a amplitude é de 77,9, a amplitude começa a decrescer constantemente até atingir o mínimo do período em 1948, mínimo este de 68,1.

A alta taxa de mortalidade no meses de primavera e verão, principalmente no mês de dezembro, fica explicada pela incidência das doenças infecciosas, relacionadas com doenças diarreicas e com desidratação que ocorrem com maior frequência nos meses de calor. Os meses com menor mortalidade são os meses de inverno, junho, julho e agosto. A forte oscilação da amplitude sazonal indica que neste período a mortalidade infantil é fortemente influenciada pelas condições climáticas e pelas estações, mostrando a vulnerabilidade da saúde da população.

O elevado nível da tendência para a década de 30 e a amplitude crescente são produtos das condições precárias do Estado. A queda mostrada pela tendência e pela amplitude sazonal a partir de 1940 sugere o início do melhoramento das condições de vida como consequência da expansão industrial do Estado.

5.7.2 Análise dos componentes no período de 1950 a 1985.

Os gráficos 10 (pag. 80) e 14 (pag. 86) mostram a série original, a tendência e a sazonalidade respectivamente, nos períodos de 1950 a 1975 e de 1960 a 1985. A tendência é decrescente até 1960, quase constante de 1961 a 1963, sofre uma queda em 1964, novamente permanece constante até 1968, começa a crescer a partir de 1969 até 1975, quando retoma o decrescimento.

A sazonalidade pode ser dividida em 4 períodos, de 1950 a 1958, de 1959 a 1967, de 1968 até 1974, e de 1975 a 1985.

No primeiro período, que abrange os anos de 1950 a 1958, o padrão sazonal predominante começa em janeiro descendo até atingir o mínimo absoluto no mês de agosto; logo começa a subir até atingir o máximo absoluto no mês de dezembro. A amplitude sazonal decresce de 61,6 em 1950 até 45,8 em 1958. O segundo período que vai de 1959 a 1967, mostra uma estrutura similar à anterior, com a ligeira diferença de que, o mínimo e o máximo relativo possuem uma importância relativa maior que no período anterior. O mínimo relativo é atingido em abril e o máximo relativo em maio. O máximo absoluto é atingido no mês de dezembro e o mínimo no mês de agosto. A amplitude sazonal continua decrescendo, indo de 42,64 a 23,46 no final do período.

As principais características do padrão sazonal dos períodos observados anteriormente são a queda rápida da amplitude sazonal e a maior relevância dos máximos e mínimos relativos. A redução da amplitude sazonal indica a redução das oscilações da mortalidade infantil ao longo do ano, mostrando a menor vulnerabilidade da saúde da população frente às condições climáticas. Esta melhoria pode ser explicada pela evolução industrial do Estado e avanços da medicina, que conseguem atenuar as oscilações da mortalidade devida, principalmente, a epidemias.

No período de 1968 a 1974, o máximo absoluto é atingido no mês de janeiro ou fevereiro e o mínimo absoluto é atingido em setembro; o mínimo relativo também se translada para o mês de maio, acompanhando a defasagem ocorrida no período. A amplitude sazonal após atingir o mínimo de 21,6 em 1968, cresce até o final do período quando atinge o valor de 31,1, acompanhando a tendência. O crescimento da tendência e da amplitude sazonal reforça a hipótese da influência negativa do processo de desenvolvimento da economia adotado neste período sobre a mortalidade infantil.

A defasagem de um mês, mostrada no padrão sazonal em relação ao padrão dos anos anteriores, implica mudanças no tipo de influência da sazonalidade na mortalidade infantil, isto é, as taxas mais altas são atingidas, principalmente, nos meses de verão, e as taxas mais baixas nos meses de primavera. A importância dos máximos relativos se deve ao fato de que os máximos absolutos perdem seu papel predominante no padrão sazonal e, por outro lado, a taxa de mortalidade cresce relativamente, em relação aos anos anteriores, nos meses de inverno.

A partir de 1975, se observa que o máximo é atingido no mês de janeiro com exceção de 1985 quando é atingido em fevereiro. O mínimo absoluto é atingido em outubro nos anos de 1977, 1978 e 1979, entretanto, para os outros anos este é atingido no mês de setembro. O mínimo relativo ocorre no mês de maio e o máximo relativo no mês de junho. O ano de 1985 mostra um padrão muito particular. A amplitude sazonal permanece quase constante ao redor de 30,0.

A influência da sazonalidade na mortalidade infantil neste último período é similar à do período anterior; a diferença reside basicamente no fato de que, neste período, a amplitude sazonal permanece quase constante. Este fato deve chamar a atenção, especialmente quando se considera que a amplitude sazonal já foi menor.

A diminuição da amplitude sazonal é muito importante, pois ela mede o grau de vulnerabilidade da saúde da população às condições climáticas.

5.8 Conclusões.

Ao longo do trabalho foi observado que o desempenho do método depende das características particulares da série. Quando a série é bem comportada, com sazonalidade e tendência estável, a qualidade é ótima. À medida que se incorporam modificações estruturais, a qualidade vai se deteriorando tornando-se necessário o uso de opções alternativas.

A análise dos experimentos mostra resultados interessantes. Na maioria dos casos, a opção automática não fornece um ajuste adequado. Os problemas que normalmente surgem são: resíduos autocorrelacionados, periodicidade, alta variabilidade na série sazonalmente ajustada, alta participação do componente irregular em relação à sazonalidade e tendência. Estas deficiências estão estreitamente relacionadas aos períodos escolhidos, isto é, dependem da estrutura sazonal e da tendência. Por exemplo, a presença de sazonalidade móvel influencia negativamente na qualidade do ajuste; a alta variabilidade do padrão sazonal implica forte autocorrelação dos resíduos; a presença de defasagem se traduz em autocorrelação ou periodicidade, etc.

O desempenho da opção automática depende principalmente dos filtros usados por esta. Em todos os casos, o filtro para o ajuste final da tendência é o 13-H, e na maioria dos casos o filtro para o ajuste final do fator sazonal é o 3x5. Estes filtros são de comprimento médio, e às vezes, estes não conseguem acompanhar os movimentos da sazonalidade e da tendência de forma mais adequada.

Em todos os experimentos, o uso do filtro 9-H para a tendência significou uma melhora substancial na qualidade do ajuste e, em todas as situações, o uso do filtro 23-H implicou em autocorrelação e/ou periodicidade nos resíduos, além das outras deficiências mostradas através das estatísticas de controle de qualidade.

Quando a série apresenta quedas bruscas no nível da tendência, a opção automática fica contaminada por autocorrelação e periodicidade nos resíduos. Esta deficiência, em parte, é superada pelo uso do filtro 9-H (o de menor comprimento) pela sua flexibilidade para acompanhar este tipo de mudanças.

No caso dos filtros para o fator sazonal, estes dependem das características da série no período em estudo. Em alguns casos o filtro 3x3 (o de menor comprimento) consegue acompanhar os movimentos do fator sazonal, sobretudo quando o componente irregular é pequeno, como no caso da série no período de 1933 a 1948. Este filtro também mostra certas vantagens para a série no período de 1950 a 1975, embora seu desempenho não seja tão bom quanto ao anterior, pois, nesta série, o componente irregular é maior que na série anterior. Porém, quando a série possui um padrão sazonal altamente variável, este filtro tem um efeito contrário, pois, ele é muito sensível a movimentos irregulares.

Quando a série possui um padrão altamente variável, o filtro 3x9 (o de maior comprimento), mostra um bom desempenho acompanhando muito bem os movimentos sistemáticos do padrão sazonal, como no caso da série no período de 1950 a 1965, caracterizada pela amplitude sazonal decrescente.

Quando a participação do componente irregular é alta ou quando a variabilidade do padrão sazonal é forte, mostrada através de um alto valor de I/S , ou quando há presença de defasagem, o desempenho da opção automática é ruim. Este é o caso da série nos períodos 1960-1968 e 1970-1985. No primeiro período não há presença de defasagem. O uso de filtros adequados, especialmente de filtros para a sazonalidade consegue notáveis melhoras na qualidade do ajuste. Entretanto, eles não conseguem eliminar totalmente as deficiências mostradas pela opção automática.

Os períodos onde a série apresenta defasagem são, 1960-1975, 1970-1985 e 1960-1985. Nos dois primeiros períodos a defasagem é de um mês, com o último período incorporando as duas defasagens dos períodos anteriores. Nestes casos, a qualidade do ajuste da opção automática apresenta autocorrelação nos resíduos, sendo a forte autocorrelação explicada pelo fato de que este último período abrange duas defasagens. Nenhuma opção consegue eliminar completamente as deficiências mostradas pela opção automática; consegue-se apenas opções alternativas à esta última.

Na maioria dos casos, o uso da manipulação dos valores extremos não melhora o ajuste. Só em um caso esta opção consegue uma melhoria relativa na qualidade do ajuste. O uso de outras opções, como o 24-termos para a estimativa inicial da tendência, não consegue uma substancial melhora, e, em alguns casos, aumenta a participação do componente irregular. A consequência de se retirar a opção ARIMA é a introdução de periodicidade. O uso da opção ARIMA, quando esta é rejeitada, introduz autocorrelação e periodicidade nos resíduos.

Infelizmente, quase nenhuma opção consegue eliminar totalmente as deficiências mostradas pela opção automática, embora em todos os casos tenham sido encontradas opções alternativas que melhoraram a qualidade substancialmente.

O estudo das revisões mostra que a opção automática é sempre mais conservadora que a opção alternativa, o que é explicado pelo fato da opção automática usar filtros de comprimento médio, enquanto, as opções alternativas usam o filtro mais curto para a tendência e diversos filtros para o fator sazonal. O fato de usar o filtro mais curto para a tendência implica num acompanhamento mais rápido dos movimentos da série, pois este reage com maior rapidez frente à mudanças próximas, mas, corre-se o risco de se contaminar as estimativas dos componentes com parte do componente irregular, e além disto, as revisões são geralmente maiores. Como vantagem tem-se, em alguns casos, detecções antecipadas de pontos de inflexão.

Como conclusão, pode-se dizer que nem sempre a opção automática dá bons resultados, e que, em alguns casos a manipulação adequada das opções existentes pode ajudar bastante na melhora da qualidade do ajuste, mesmo que as séries apresentem grandes mudanças estruturais. A análise dos componentes mostra resultados compatíveis com as análises já existentes sobre a série indicando que o método consegue descrever razoavelmente bem, o comportamento da série. A análise do componente sazonal indica como este componente pode ajudar a entender o processo gerador de dados.

Finalmente, duas sugestões se impõem na conclusão deste trabalho. A primeira se refere à existência de outros métodos de ajuste que poderiam, talvez, obter melhores resultados no ajuste de séries brasileiras. Como o objetivo deste trabalho foi o estudo do X-11 ARIMA, não foram analisados estes métodos. Todavia, seria interessante que se desenvolvessem estudos comparativos neste sentido. Outro ponto seria deslocar a atenção para a série, visando um melhor entendimento das causas que a afetam e como o ajuste sazonal pode auxiliar nesta tarefa.

Apêndices.

Apêndice 1. Desempenho da opção automática para a série no período de 1950 a 1985	96
Apêndice 2. Desempenho das opções alternativas para a série no período de 1950 a 1975: TX5075, X-11 ARIMA	100
Apêndice 3. Desempenho das opções alternativas para a série no período de 1950 a 1965: TX5075, X-11 ARIMA	102
Apêndice 4. Desempenho das opções alternativas para a série no período de 1960 a 1975: TX6075, X-11 ARIMA	105
Apêndice 5. Desempenho das opções alternativas para a série no período de 1970 a 1985: TX7085, X-11 ARIMA	107
Apêndice 6. Desempenho das opções alternativas para a série no período de 1960 a 1985: TX6085, X-11 ARIMA	110
Apêndice 7. Dados originais da taxa mensal e anual de mortalidade infantil do Estado de São Paulo no período de 1933 a 1985	113

APÊNDICE 1. Desempenho da opção automática no período de 1950 a 1985.

Na tabela 19 observa-se que as séries que incluem os 15 primeiros anos rejeitam os três modelos ARIMA. Isto deve-se ao fato da amplitude sazonal e tendência dos 10 primeiros anos serem fortemente decrescentes, não conseguindo se compatibilizar com a estrutura da década de 60, embora, ao se ampliar a série até 1975, o modelo $lg(0,1,1) \times (0,1,1)$ seja aceito. Isto acontece porque o peso relativo destes 10 primeiros anos é menor. A partir de 1960 todas as séries aceitam (com exceção da série TX6068) o modelo $lg(0,1,1)(0,1,1)$ devido a tendência e a amplitude sazonal não apresentarem mudanças tão bruscas nestes períodos como no caso anterior, embora o padrão sazonal seja altamente variável.

Tabela 19. Ajuste dos modelos ARIMA da opção automática.

série	modelos	M A P E					
		média	1				R
			1	1-1	1-2	IX (%)	
TX5075	$lg(0,1,1)(0,1,1)$	9,56	6,71	10,52	11,44	18,63	0,70
	$lg(0,2,2)(0,1,1)$	13,91	7,94	21,01	12,77	18,67	0,65
	* $lg(0,1,1)(0,1,1)$	9,53	6,68	10,55	11,36	4,08	0,70
TX5065	$lg(0,1,1)(0,1,1)$	14,74	16,70	17,03	10,51	1,89	0,79
	$lg(0,2,2)(0,1,1)$	23,58	28,00	24,90	17,86	1,37	0,73
	--(2,1,2)(0,1,1)	7,56	7,36	7,86	7,46	8,11	0,74
TX5068	$lg(0,1,1)(0,1,1)$	5,10	6,61	5,01	3,67	5,16	0,77
	$lg(0,2,2)(0,1,1)$	8,76	15,88	5,31	5,09	1,73	0,62
	--(2,1,2)(0,1,1)	4,83	5,71	5,29	3,49	2,16	0,71
TX6075	$lg(0,1,1)(0,1,1)$	9,56	6,71	10,52	11,44	18,63	0,66
	$lg(0,2,2)(0,1,1)$	13,91	7,94	21,01	12,77	18,67	0,65
	* $lg(0,1,1)(0,1,1)$	9,53	6,68	10,55	11,36	4,08	0,70
TX6068	$lg(0,1,1)(0,1,1)$	5,25	6,96	5,12	3,68	44,14	0,55
	$lg(0,2,2)(0,1,1)$	7,88	13,81	6,64	3,19	23,86	0,42
	--(2,1,2)(0,1,1)	4,29	6,09	2,87	3,89	37,51	0,61
TX7085	$lg(0,1,1)(0,1,1)$	8,34	7,28	13,86	3,87	71,41	0,93
	$lg(0,2,2)(0,1,1)$	15,13	19,09	16,79	9,50	80,26	0,92
	* $lg(0,1,1)(0,1,1)$	8,52	8,07	13,77	3,73	58,82	0,93
TX6085	$lg(0,1,1)(0,1,1)$	8,53	7,19	14,49	3,90	78,76	0,93
	$lg(0,2,2)(0,1,1)$	15,02	18,67	17,25	9,15	87,63	0,92
	* $lg(0,1,1)(0,1,1)$	10,29	12,87	14,24	3,74	53,32	0,93

0 modelo escolhido

A série nos períodos mais curtos apresenta forte presença do componente irregular (acima de 10%), e a variação fica explicada principalmente pelas mudanças da sazonalidade. Entretanto, a tendência da série nos períodos mais compridos possui um papel mais importante, porque ao se ampliar o período de estudo, a tendência varia muito mais que nos períodos mais curtos. As séries que incluem o período de 1975 a 1985 mostram maior participação da tendência porque neste período a série decresce rapidamente (v. tabela 20).

Tabela 20. Estrutura da variância da porção estacionária da série original (%).

compon	TX5075	TX5065	TX5060	TX6060	TX6075	TX7085	TX6085
I	8	11	10	13	12	11	6
C	42	10	16	14	31	45	75
S	50	79	74	73	57	44	19

O ADR do componente irregular toma valores dentro dos limites permissíveis de aceitação da hipótese de aleatoriedade dos resíduos, que são 1,36 e 1,75 (tabela 21). O ADR da tendência apresenta forte correlação positiva, o que era esperado. O número de meses para a dominação das mudanças da tendência com respeito às mudanças do componente irregular é alto, sobretudo para os períodos que incluem os 10 primeiros anos; o que implica que a opção automática não está fornecendo uma série suave.

Tabela 21. Duração média de sequências (ADR) e o número de meses para dominação cíclica (MCD).

Estat.	TX5075	TX5065	TX5060	TX6060	TX6075	TX7085	TX6085
CI	1,65	1,65	1,72	1,60	1,72	1,77	1,81
I	1,56	1,57	1,58	1,65	1,54	1,65	1,57
C	6,48	6,59	6,68	5,35	5,97	7,64	6,76
MCD	2,95	5,70	5,53	4,59	2,92	5,34	4,59
#meses	5	6	6	6	5	4	5

A razão $\bar{I}/\bar{C}$ da série no período 50-60 supera o valor de 3, o que indica que a separação da tendência e do componente irregular pode não ser satisfatória. Para a série em outros períodos, o filtro adequado é o 13-H, o qual é de fato escolhido pela opção automática. Entretanto, para o fator sazonal, as razões (com execução da série no período 70-85) caem no intervalo que vai desde 1,5 até 6,5, o que implica que a série pode ser considerada "padrão". A razão para a série no período 70-85 indica que é altamente irregular, sugerindo o uso de filtros mais compridos (v. tabela 22).

Tabela 22. Razão final entre os componentes.

Razão	TX5075	TX5065	TX5068	TX6068	TX6075	TX7085	TX6085
$\frac{I}{S}$	2.99	2.91	3.15	2.95	2.68	2.25	2.42
$\frac{I}{\bar{S}}$	3.93	4.05	3.78	4.65	4.49	6.84	5.51

O p-valor associado aos testes de sazonalidade estável é quase zero para todos os casos (v. tabela 23). Os altos valores das estatísticas mostram que a presença da sazonalidade estável é forte. Quando comparados estes valores com os do período anterior observa-se que para Fs_{B1} , os valores alcançam só a metade e os valores de Fs_{D8} a terceira parte, indicando que a sazonalidade estável do período anterior é maior que neste período, confirmando as suposições anteriores. Ao nível de 0,1%, a hipótese de presença de sazonalidade móvel é aceita para a série nos três primeiros períodos; isto se deve à presença da década de 50, onde a amplitude sazonal decresce rapidamente. Para as outras séries, esta hipótese é rejeitada, fato que é confirmado pela amplitude sazonal, que a partir de 1960 não possui mudanças fortes.

Tabela 23. Teste para sazonalidade.

Teste	TX5075	TX5065	TX5068	TX6068	TX6075	TX7085	TX6085
Fs_{B1}	34,9	42,1	40,8	22,5	25,0	24,1	29,5
Fs_{D8}	40,1	49,3	48,0	39,3	36,4	45,1	42,4
$X2_{K-W}$	207,4	153,0	173,6	90,4	135,9	144,3	190,7
Fm	4,4	3,5	4,6	3,6	2,2	1,9	1,7
p-valor	0,0	0,0	0,0	0,5	0,9	2,8	0,5

O teste de Kolmogorov-Smirnov (v. tabela 24) indica a não presença do componente periódico, embora na última série exista ligeira evidência de periodicidade. Isto mostra que os filtros usados pela opção automática (especialmente o filtro da tendência) conseguem captar os movimentos subjacentes da série.

Tabela 24. Teste de Kolmogorov-Smirnov (%).

Interv.:	TX5075	TX5065	TX5068	TX6068	TX6075	TX7085	TX6085
-75%	100	98	95	100	98	82	64
75%-95%	0	2	5	0	2	18	35
+95%	0	0	0	0	0	0	1

A tabela 25 mostra que a série nos períodos 60-68 e 60-85 possui uma autocorrelação muito forte. A série nos períodos: 50-68, 60-68, 60-75 mostra autocorrelação de lag 12, o que pode indicar presença de sazonalidade nos resíduos. A série no período 50-55 apresenta menor índice de autocorrelação.

Tabela 25. Função de autocorrelação dos resíduos.

Lag k	TX5075	TX5065	TX5068	TX6068	TX6075	TX7085	TX6085
1	0,01	0,07	0,04	0,01	-0,03	0,18*	0,12*
2	-0,17*	-0,18*	-0,20*	-0,20*	-0,13	-0,15*	-0,12*
3	-0,06	-0,05	-0,04	-0,11	-0,15*	-0,14	-0,16*
4	-0,04	-0,06	-0,08	-0,20	-0,10	-0,10	-0,13*
5	-0,03	-0,03	-0,04	-0,07	-0,01	-0,10	-0,12*
6	0,01	-0,01	0,05	0,11	0,00	-0,07	-0,04
7	-0,05	-0,03	-0,02	0,11	0,00	-0,01	-0,00
8	-0,03	-0,04	-0,07	0,04	0,09	-0,03	0,07
9	0,12*	0,10	0,11	0,24*	0,21*	0,17*	0,14*
10	0,02	0,03	0,04	-0,08	-0,08	0,14	0,10
11	0,07	0,10	0,07	-0,22*	-0,03	0,05	0,01
12	-0,14*	-0,11	-0,16*	-0,25*	-0,15*	-0,18*	-0,09
13	0,05	0,05	0,06	0,12	0,08	-0,11	-0,10
14	0,18*	-0,21*	0,22*	0,12	0,02	-0,02	-0,02
$2/\sqrt{n}$	0,113	0,144	0,132	0,192	0,144	0,144	0,113

Com exceção da série TX5075 (v. tabela 26), as estatísticas de controle de qualidade, mostram que a opção automática não consegue controlar o movimento do componente irregular de forma efetiva. Para o caso da série no período 70-85, esta situação se torna mais crítica.

Tabela 26. As estatísticas de controle de qualidade.

Estat.	TX5075	TX5065	TX5068	TX6068	TX6075	TX7085	TX6085
M1	0,886	0,704	0,761	1,061*	1,270*	1,458*	1,586*
M2	0,737	1,039*	0,939	1,235*	1,171*	1,099*	0,557
M3	0,993	0,958	1,075*	0,973	0,838	0,624	0,709
M4	0,453	0,378	0,469	0,595	0,245	0,779	0,505
M5	0,872	0,936	0,975	1,042*	0,851	0,624	0,818
M6	0,029	0,020	0,036	0,261	0,196	1,106*	0,602
M7	0,501	0,420	0,465	0,451	0,455	0,375	0,386
M8	0,737	0,571	0,656	0,469	0,700	0,393	0,565
M9	0,496	0,389	0,464	0,410	0,503	0,893	0,356
M10	0,663	0,379	0,560	0,672	0,829	0,260	0,384
M11	0,570	0,372	0,556	0,661	0,738	0,221	0,260
Q	0,70	0,65	0,71	0,78	0,76	0,65	0,65
Nº Falhas	—	1	1	3	2	3	1

Apêndice 2. Desempenho das opções alternativas para a série no período de 1950 a 1975: TX5075, X-11 ARIMA.

A variação da série neste período fica explicada tanto pela sazonalidade quanto pela tendência; o componente irregular é pequeno, e, é menor quando se usa o filtro 9-H para a tendência (v. tabela 27).

Tabela 27. Estrutura da variância da porção estacionária da série original (%).

Compo.	auto	3x3,9-H	3x5,9-H	s/out 3x5,9-H	3x9,9-H	3x3,23-H
I	8	6	6	6	7	8
C	42	43	45	46	45	41
S	50	51	49	48	48	51

O valor ADR (v. tabela 28) para todas as opções cai no intervalo 1,36 e 1,75, indicando aleatoriedade dos resíduos. O menor valor para o MCD é de 3, que ocorre quando se utiliza o filtro 9-H. Observa-se também que, ao se utilizar o filtro 23-H, este valor supera o limite permitido, indicando que este não consegue captar os movimentos subjacentes da série.

Tabela 28. Duração média de sequências (ADR) e o número de meses para dominação cíclica (MCD).

Estat.	auto	3x3,9-H	3x5,9-H	s/out 3x5,9-H	3x9,9-H	3x3,23-H
CI	1,65	1,71	1,69	1,67	1,73	1,56
I	1,56	1,45	1,47	1,45	1,43	1,53
C	6,48	4,06	4,86	5,98	5,18	10,37
MCD	2,95	2,58	2,71	4,26	2,97	4,77
Meses	3	3	3	4	3	3

A opção automática proporciona um valor de $\bar{I}/\bar{C}$ próximo ao limite permissível (2,0) para esta razão, e a opção que usa o filtro 23-H supera este limite, indicando o perigo de não conseguir a separação satisfatória destes componentes. Estas razões diminuem ao diminuir o comprimento dos filtros para a tendência e para o fator sazonal, principalmente quando se manipula valores extremos (Tabela 29).

Tabela 29. Razão final entre os componentes.

Razão	auto	3x3,9-H	3x5,9-H	s/out 3x5,9-H	3x9,9-H	3x3,23-H
$\bar{I}/\bar{C}$	2,97	2,21	2,35	1,63	2,39	3,26
$\bar{I}/\bar{S}$	3,93	3,17	3,17	2,35	3,84	4,28

Em todos os casos a hipótese de presença de sazonalidade estável é aceita, pois o valor do p-valor para todos os casos é próximo de zero. A hipótese de presença de sazonalidade móvel também é aceita para todas as opções; isto de fato era esperado pela presença da amplitude decrescente da década de 50 (tabela 30).

Tabela 30. Teste para sazonalidade.

Estat.	auto	3x3,9-H	3x5,9-H	s/out	3x9,9-H	3x3,23-H
Ta_{81}	34,9	34,9	35,5	35,5	34,9	35,5
Ta_{08}	40,1	42,2	45,6	46,0	49,3	38,6
$X2_K_W$	207,4	212,0	216,6	215,1	221,0	205,1
Tm	4,4	4,6	4,6	4,7	4,9	4,4
p-valor	0,0	0,0	0,0	0,0	0,0	0,0

Todas as opções mostram indícios de autocorrelação, e, ao se utilizar o filtro 23-H para a tendência, esta autocorrelação é mais forte (v. tabela 31).

Tabela 31. Função de autocorrelação dos resíduos.

lag k	auto	3x3,9-H	3x5,9-H	s/out	3x9,9-H	3x3,23-H
1	0,01	-0,10	-0,09	-0,12*	-0,12*	0,14*
2	-0,17*	-0,16*	-0,19*	-0,21*	-0,22*	-0,07
3	-0,06	0,02	0,02	0,00	0,05	-0,09
4	-0,04	0,03	-0,00	0,00	0,00	-0,14*
5	-0,03	-0,04	-0,05	-0,04	-0,06	-0,13*
6	0,01	0,02	0,01	0,04	0,03	-0,05
7	-0,05	-0,05	-0,06	-0,09	-0,06	-0,01
8	-0,00	-0,11	-0,08	-0,08	-0,06	0,01
9	0,12*	0,08	0,08	0,10	0,13*	0,14*
10	0,02	0,06	0,01	0,01	-0,01	0,04
11	0,07	0,13*	0,12*	0,11	0,12*	0,03
12	-0,14*	-0,19*	-0,14*	-0,12*	-0,10	-0,17*
13	0,05	0,11	0,09	0,06	0,03	0,00
14	0,18*	0,19*	0,16*	0,16*	0,11	-0,14*

A opção automática e as opções dos filtros mais curtos (3x3,9-H e 3x5,9-H) não apresentam indícios de periodicidade escondida nos resíduos; curiosamente, ao se manipular valores extremos, se introduz periodicidade (v. tabela 32).

A única opção com problemas nas estatísticas de controle de qualidade é a do 23-H, mostrando problemas em M3 e M5 que estão relacionados com a variação da tendência em relação à variação do componente irregular, fato que já tinha sido observado anteriormente. A qualidade do ajuste global melhora substancialmente ao se usar filtros mais curtos, principalmente para a tendência (tabela 33).

Tabela 32. Teste de Kolmogorov-Smirnov (%).

		s/out				
Interv.	auto	3x3,9-H	3x5,9-H	3x5,9-H	3x9,9-H	3x3,23-H
-75%	100	82	81	66	59	59
75% 95%	0	12	15	15	25	38
95%	0	0	4	19	16	0

Tabela 33. As estatísticas de controle de qualidade

		s/out				
Estat.	auto	3x3,9-H	3x5,9-H	3x5,9-H	3x9,9-H	3x3,23-H
M1	0,886	0,542	0,602	0,663	0,623	0,960
M2	0,737	0,527	0,612	0,624	0,630	0,808
M3	0,993	0,605	0,673	0,317	0,693	1,131*
M4	0,453	0,383	0,226	0,331	0,488	0,244
M5	0,672	0,498	0,488	0,505	0,461	1,392*
M6	0,027	0,331	0,331	0,660	0,066	0,081
M7	0,501	0,497	0,479	0,478	0,468	0,510
M8	0,737	0,885	0,663	0,647	0,503	0,933
M9	0,796	0,510	0,484	0,487	0,439	0,523
M10	0,663	0,856	0,564	0,529	0,390	0,917
M11	0,570	0,675	0,545	0,505	0,388	0,782
C	0,70	0,56	0,51	0,52	0,52	0,81
Nºfolhas	---	---	---	---	---	2

Apêndice 3. Desempenho das opções alternativas para a série no período de 1950 a 1965:TX5065, X-11.

As razões $\bar{I}/\bar{C}$ e $\bar{I}/\bar{S}$ diminuem ao se manipular valores extremos. A razão $\bar{I}/\bar{C}$ indica que o filtro adequado para o ajuste final da tendência é o 13-H, que de fato é o escolhido, com exceção das opções que escolhem o filtro a priori. Pode-se observar que ao se diminuir o comprimento do filtro, a razão diminui, pois os filtros mais curtos acompanham melhor os movimentos da série (tabela 34).

Tabela 34. Razão final entre os componentes.

					s/out		
Razão	auto	s/out	3x3	3x3,9-H	3x3,9-H	3x5,9-H	3x9,9-H
$\bar{I}/\bar{C}$	2,91	1,90	2,47	2,46	1,69	2,53	2,37
$\bar{I}/\bar{S}$	4,05	3,24	3,47	3,59	2,83	3,68	4,62

A maior parte da variação da série se deve à sazonalidade, que explica mais de 77% da variação total. A tendência quase não participa e a participação do componente irregular vai diminuindo à medida que se encurta o comprimento dos filtros (tabela 35).

Tabela 35. Estrutura da variância da porção estacionária da série original (%).

Compo.	auto	s/out	3x3	3x3,9-H	3x3,9-H	3x5,9-H	3x7,9-H
I	11	11	9	8	8	9	9
C	10	11	11	11	10	10	11
C	79	78	80	81	79	78	77

Os valores do ADR do componente irregular caem no intervalo de aceitação da hipótese de aleatoriedade dos resíduos (tabela 36). A opção 3x7,9-H fornece uma série sazonalmente ajustada mais suave.

Tabela 36. Duração média de sequências (ADR) e o número de meses para dominação cíclica (MCD).

Estat.	auto	s/out	3x3	3x3,9-H	3x3,9-H	3x5,9-H	3x7,9-H
IC	1,65	1,59	1,62	1,65	1,65	1,62	1,74
I	1,57	1,54	1,52	1,49	1,43	1,50	1,52
C	6,59	7,07	6,30	4,90	5,16	4,90	5,16
MCD	5,78	5,14	2,88	4,16	4,35	3,98	2,82
Nmeses	6	6	5	4	4	4	3

As hipóteses de presença de sazonalidade estável e móvel são aceitas para todos os casos ao nível de significância de 0,1%. Isto se deve fundamentalmente ao fato da amplitude sazonal decrescer rapidamente na década de 50 (tabela 37).

Tabela 37. Teste para sazonalidade.

Estat.	auto	s/out	3x3	3x3,9-H	3x3,9-H	3x5,9-H	3x7,9-H
TS_31	42,1	42,1	42,1	42,1	42,1	42,1	42,1
TS_38	49,2	50,5	50,3	56,2	58,9	60,5	68,8
X2_K_W	152,7	153,5	153,7	158,3	158,1	159,9	161,8
Tn	3,5	3,9	3,5	3,2	3,5	3,2	3,4
p-valor	0,0	0,0	0,0	0,01	0,0	0,01	0,01

A função de autocorrelação dos resíduos (tabela 38) indica que a opção automática, a opção com manipulação de valores extremos e a opção 3x7,9-H não mostram presença de autocorrelação, enquanto as outras mostram autocorrelação de lag 12.

O teste K-S mostra que não há presença de periodicidade nos resíduos (tabela 39).

Tabela 38. Função de autocorrelação dos resíduos.

lag k	auto	s/out	3x3	3x3,9-H	3x3,9-H	3x5,9-H	3x9,9-H
1	0,07	0,04	0,02	-0,07	-0,12	-0,07	-0,10
2	-0,18*	-0,21*	-0,22*	-0,17*	-0,25*	-0,23*	-0,25*
3	-0,03	-0,09	-0,07	0,02	-0,01	0,04	0,06
4	-0,06	-0,07	-0,06	0,00	0,01	0,02	-0,02
5	-0,03	-0,08	-0,04	-0,03	-0,03	-0,05	-0,05
6	-0,01	-0,04	0,05	0,03	0,07	0,04	0,04
7	-0,03	-0,08	-0,03	-0,04	-0,07	-0,06	-0,06
8	-0,04	0,03	-0,07	-0,13	-0,12	-0,11	-0,06
9	0,10	0,08	0,10	0,04	0,06	0,03	0,12
10	0,03	0,01	0,03	0,08	0,04	0,03	-0,01
11	0,10	-0,07	0,10	0,17*	0,14	0,15*	0,12
12	-0,11	-0,05	-0,13*	0,21*	0,17*	-0,20*	-0,11
13	0,05	0,04	0,01	0,04	0,03	0,03	-0,01
14	0,21*	0,17*	0,13*	0,24*	0,21*	0,20*	0,13

Tabela 39. Teste de Kolmogorov-Smirnov.

Interv.	auto	s/out	3x3	3x3,9-H	3x3,9-H	3x5,9-H	3x9,9-H
-75%	98	100	100	98	84	92	83
75%-95%	2	0	0	2	16	8	17
>95%	0	0	0	0	0	0	0

Tabela 40. As estatísticas de controle de qualidade.

Estat.	auto	s/out	3x3	3x3,9-H	3x3,9-H	3x5,9-H	3x9,9-H
M1	0,704	0,755	0,599	0,432	0,488	0,486	0,501
M2	1,009*	1,104*	0,876	0,694	0,760	0,705	0,858
M3	0,956	0,452	0,733	0,728	0,346	0,764	0,684
M4	0,370	0,245	0,111	0,022	0,423	0,044	0,111
M5	0,936	1,007*	0,854	0,546	0,560	0,524	0,473
M6	0,020	0,303	0,210	0,164	0,468	0,129	0,247
M7	0,420	0,425	0,412	0,386	0,385	0,372	0,353
M8	0,571	0,514	0,772	0,752	0,694	0,549	0,367
M9	0,389	0,354	0,449	0,397	0,395	0,367	0,306
M10	0,379	0,320	0,444	0,410	0,421	0,431	0,310
M11	0,372	0,317	0,401	0,329	0,340	0,451	0,313
Q	0,65	0,55	0,58	0,48	0,49	0,45	0,45
Wolius	1	2	---	---	---	---	---

Observa-se que a opção automática (ver tabela 40) acusa problemas na estatística M2 que está relacionada com a participação do componente irregular em relação ao fator sazonal. Ao se manipular a

opção de valores extremos se agrava a situação, pois M5 também falha. A qualidade geral do ajuste melhora substancialmente ao se usar o filtro 7-H para a tendência e isto é complementado ao se usar os filtros 3x3 ou 3x9 para o fator sazonal.

Apêndice 4. Desempenho das opções alternativas para a série no período de 1960 a 1975: TX6075, X-11 ARIMA.

Ambas razões diminuem ao se encurtar o comprimento dos filtros, e mais ainda ao se manipular valores extremos (tabela 41).

Tabela 41. Razão Final entre os componentes.

Razão	auto	3x3,9-H	3x5,9-H	3x9,9-H	s/out 13x9,9-H	s/out, 13x9,9-H,24-t
$\frac{I}{C}$	2,78	2,03	2,24	2,35	1,60	1,60
$\frac{I}{S}$	4,65	3,57	3,54	4,10	3,15	3,21

Os filtros mais curtos conseguem reduzir a participação do componente irregular (tabela 42). A variabilidade neste período fica explicada principalmente pela variação da sazonalidade.

Tabela 42. Estrutura da variância da porção estacionária da série original.

Compon.	auto	3x3,9-H	3x5,9-H	3x9,9-H	s/out 13x9,9-H	s/out, 13x9,9-H,24-t
I	12	9	9	10	10	10
C	31	33	34	35	34	34
S	57	58	57	55	56	56

Os valores do ADR caem no intervalo 1,36 a 1,75 para o componente irregular indicando que não há presença de autocorrelação nos resíduos. O uso de filtros mais curtos fornece séries sazonalmente ajustadas mais suaves (v. tabela 43).

Tabela 43. Duração média de sequências (ADR) e o número de meses para a dominação cíclica (MCD).

Estat.	auto	3x3,9-H	3x5,9-H	3x9,9-H	s/out 13x9,9-H	s/out, 13x9,9-H,24-t
IC	1,72	1,69	1,72	1,75	1,72	1,75
I	1,49	1,38	1,45	1,38	1,52	1,49
C	5,97	4,77	4,77	5,03	5,03	5,03
MCD	2,03	2,70	2,55	2,98	2,95	4,45
meses	5	3	3	3	3	4

Em todos os casos a hipótese de sazonalidade estável é aceita para todas as opções, mas a hipótese de presença de sazonalidade móvel é rejeitada ao nível 0,1%. Isto acontece porque neste período a amplitude sazonal não sofre fortes variações como nos anos anteriores (tabela 44).

Tabela 44. Teste para sazonalidade.

Estat.	auto	3x3,9-H	3x5,9-H	3x9,9-H	s/out	s/out, 13x9,9-H,24-t
$F_{1,01}$	24,2	23,0	23,0	23,0	23,0	26,6
$F_{1,05}$	34,8	33,2	32,9	4,01	43,3	43,4
XC_K_W	134,7	138,0	140,0	143,9	143,2	142,9
F_4	2,8	2,3	2,2	1,9	1,8	1,8
p valor	0,9	0,5	0,7	2,0	4,2	4,5

A FAC da opção automática apresenta autocorrelação e a opção baseado nos filtros 3x9,9-H consegue diminuir este fator ainda mais, ao se manipular valores extremos (tabela 45).

Tabela 45. Função de autocorrelação dos resíduos.

lag k	auto	3x3,9-H	3x5,9-H	3x9,9-H	s/out	s/out, 13x9,9-H,24-t
1	-0,03	-0,21*	-0,20*	-0,18*	-0,13	-0,13
2	-0,11	-0,10	-0,10	-0,11	-0,13	-0,13
3	-0,15*	-0,01	-0,01	-0,00	-0,06	-0,07
4	-0,09	0,00	-0,00	-0,00	-0,02	-0,03
5	-0,01	-0,03	-0,03	-0,01	0,02	0,01
6	-0,03	0,02	0,02	0,02	0,04	0,05
7	-0,01	-0,07	-0,07	-0,05	-0,04	0,05
8	0,07	-0,03	-0,04	-0,03	-0,02	-0,03
9	0,19*	0,19*	0,18*	0,17*	0,17*	-0,02
10	-0,05	-0,05	-0,06	-0,03	-0,05	0,16*
11	-0,06	0,00	0,02	0,04	0,04	-0,06
12	-0,16*	-0,20*	-0,14	-0,07	-0,07	0,04
13	0,10	0,20*	0,18*	0,13	0,14	-0,06
14	0,04	0,05	0,05	0,04	0,03	0,14

Tabela 46. Teste de Kolmogorov-Smirnov(%).

Interv.	auto	3x3,9-H	3x5,9-H	3x9,9-H	s/out	s/out, 13x9,9-H,24-t
-75%	100	63	66	75	94	94
75%-95%	0	21	23	22	6	6
+95%	0	16	11	3	0	0

Quando se usa o filtro 9-H para a tendência (tabela 46)

se

introduz periodicidade, mas ao se ampliar o comprimento do filtro sazonal, a periodicidade diminui; isto é, ao se combinar os dois filtros a periodicidade quase desaparece. A manipulação de valores extremos elimina a periodicidade.

A opção automática mostra evidências de participação forte de componente irregular e isto é mostrado através dos altos valores de $M1$ e $M2$ (tabela 47). O fato de usar o filtro 24-lar não consegue melhorar esta situação, o que é conseguido pelo filtro 9-H. Quando este é combinado com o filtro 3x3 para o fator sazonal, a estatística $M10$ cai, evidenciando a incapacidade deste filtro para seguir as mudanças aleatórias da sazonalidade. Esta última deficiência é superada ao se usar os filtros 3x5 e 3x9.

Tabela 47. As estatísticas de controle de qualidade.

Estat.	auto	3x3,9-H	3x5,9-H	3x9,9-H	s/out 13x9,9-H	s/out, 10x9,9-H,24-l
$M1$	1,253*	0,729	0,825	0,850	0,979	1,001*
$M2$	1,169*	0,812	0,899	0,945	1,009*	1,012*
$M3$	0,092	0,514	0,621	0,674	0,298	0,302
$M4$	0,022	0,690	0,289	0,690	0,111	0,022
$M5$	0,071	0,441	0,449	0,439	0,495	0,503
$M6$	0,259	0,172	0,183	0,070	0,339	0,316
$M7$	0,440	0,427	0,415	0,380	0,375	0,376
$M8$	0,691	0,902	0,719	0,509	0,470	0,470
$M9$	0,497	0,578	0,578	0,472	0,431	0,433
$M10$	0,022	1,229*	0,827	0,551	0,522	0,525
$M11$	0,731	0,997	0,809	0,547	0,517	0,520
σ	0,74	0,66	0,56	0,60	0,53	0,52
Falhas	2	1	—	—	1	2

Apendice 5. Desempenho das opções alternativas para a série no período 1970-1975: TX7085, X-11 ARIMA.

O alto valor da razão $\hat{I}/\bar{S}$ indica que a série neste período é altamente irregular e que se deve usar filtros de maior comprimento para o fator sazonal (tabela 48). A manipulação dos valores extremos e o uso de filtros mais curtos diminuem estas razões.

Tabela 48. Razão final entre os componentes.

Razão	autom	3x3,9H	s/o 3x3,9H	s/o 3x5,9H	s/o 3x9,9H	X11 3x5,9H	3x5,9-H
$\hat{I}/\bar{S}$	2,25	1,77	1,21	1,33	1,29	2,02	1,97
$\hat{I}/\bar{S}$	6,84	6,17	4,94	5,20	5,88	6,43	6,31

Na tabela 49 pode-se observar os valores da estatística ADR. Para o irregular observa-se que todos eles caem no intervalo de acei-

testão da hipótese nula, isto é, de não presença de autocorrelação nos resíduos. Com exceção da opção automática, a série sazonalmente ajustada é suave.

Tabela 49. Duração média de sequências (ADR) e número de meses para a dominação cíclica (MOD).

Estat.	autom	3x3,9H	3x3,9H	3x5,9H	3x9,9H	3x5,9H	3x5,9-H
IC	1,77	1,85	1,93	1,82	1,84	1,75	1,75
I	1,65	1,52	1,54	1,54	1,49	1,45	1,45
C	7,64	4,90	5,48	5,16	4,90	4,90	4,66
MOD	5,34	2,59	2,52	2,74	2,02	2,91	2,91
Amostras	4	3	3	3	3	3	3

Pode-se notar que a participação da tendência é muito alta, e quanto à participação do componente irregular, observa-se que na opção automática é superior a 10%. Ao se manipular as opções, esta participação diminui (tabela 50).

Tabela 50. Estrutura da variância da porção estacionária da série original (%).

Compo.	autom	3x3,9H	3x3,9H	3x5,9H	3x9,9H	3x5,9H	3x5,9-H
I	11	7	0	8	8	7	7
C	45	48	48	47	48	47	46
S	44	45	44	45	44	46	47

Em todos os casos a hipótese de presença significativa de sazonalidade estável é aceita, pois o p-valor é quase zero. Para um nível de 0,1% de significância a hipótese nula de presença significativa de sazonalidade móvel é rejeitada em todos os casos. As últimas opções conseguem controlar melhor este tipo de sazonalidade (tabela 51).

Tabela 51. Teste para sazonalidade.

Estat.	autom	3x3,9H	3x3,9H	3x5,9H	3x9,9H	3x5,9H	3x5,9-H
Fs_3f	24,1	24,1	24,1	25,1	25,1	25,1	25,1
Fs_DS	45,1	61,9	50,1	75,2	83,2	80,2	80,0
X2_K_W	144,3	152,1	150,6	157,2	159,0	157,8	157,7
Tn	1,9	1,6	1,6	1,3	0,9	1,1	1,1
p-valor	2,8	8,9	7,3	18,8	55,9	40,6	34,4

A opção 3x3,9-H, s/out consegue eliminar a autocorrelação. Ao se utilizar a opção ARIMA (X-11, 3x5,9-H) se introduz autocorrelação

(tabela 52).

Tabela 52. Função de autocorrelação dos resíduos.

lag			s/o	s/o	s/o	X11	
k	autom	3x3,9H	3x3,9H	3x5,9H	3x9,9H	3x5,9H	3x5,9-H
1	0,18	0,08	0,06	-0,14	-0,16*	-0,17*	-0,19*
2	-0,15*	-0,18*	-0,21*	-0,26*	-0,25*	-0,25*	-0,24*
3	-0,14	-0,04	-0,08	-0,05	-0,04	0,01	0,00
4	-0,10	0,10	0,02	0,06	0,00	0,14	0,11
5	-0,10	-0,04	-0,05	-0,04	-0,01	-0,03	-0,02
6	-0,07	-0,06	-0,04	-0,08	-0,22	-0,06	-0,06
7	-0,01	-0,01	-0,04	-0,04	0,00	-0,01	-0,03
8	0,03	0,02	0,01	-0,02	-0,00	-0,02	-0,02
9	0,17*	0,04	0,10	0,07	0,08	0,03	0,23
10	0,14	0,07	0,12	0,09	0,07	0,06	0,07
11	0,05	0,08	0,10	0,10	0,05	0,11	0,11
12	-0,19*	-0,20*	-0,13	-0,15*	-0,14	-0,20*	-0,20*
13	-0,11	-0,04	-0,05	-0,05	-0,08	-0,03	-0,01
14	-0,02	0,04	0,04	0,08	0,04	0,07	0,05

Tabela 53. Teste Kolmogorov-Smirnov.

			s/o	s/o	s/o	X11	
Kolmog.	autom	3x3,9H	3x3,9H	3x5,9H	3x9,9H	3x5,9H	3x5,9-H
95%	82	95	96	68	68	62	68
75%-95%	18	5	4	23	10	14	22
195%	0	0	0	9	22	24	9

Tabela 54. As estatísticas de controle de qualidade.

			s/o	s/o	s/o	X11	
Estat.	autom	3x3,9H	3x3,9H	3x5,9H	3x9,9H	3x5,9H	3x5,9-H
M1	1,458*	0,821	1,030*	0,943	0,959	0,759	0,766
M2	1,099*	0,709	0,902	0,780	0,796	0,690	0,706
M3	0,624	0,383	0,107	0,164	0,141	0,508	0,485
M4	0,779	0,111	0,245	0,245	0,022	0,289	0,289
M5	0,624	0,364	0,428	0,403	0,387	0,348	0,358
M6	1,136*	0,868	0,375	0,479	0,752	0,972	0,924
M7	0,375	0,307	0,319	0,270	0,242	0,252	0,254
M8	0,398	1,032*	0,923	0,437	0,257	0,465	0,491
M9	0,093	0,260	0,269	0,242	0,215	0,104	0,206
M10	0,260	0,959	0,797	0,429	0,310	0,473	0,475
M11	0,221	0,476	0,434	0,361	0,308	0,377	0,390
Q	0,65	0,50	0,51	0,44	0,33	0,49	0,49
Valias	3	1	1	—	—	—	—

Ao se ampliar o comprimento do "filtro sazonal" se introduz periodicidade nos resíduos, ao se retirar a opção ARIMA produz o mesmo efeito (tabela 53).

O uso da opção 3x3,9-H implica na falha da estatística M0 que controla o movimento aleatório da sazonalidade ao longo da série. Esta falha deve-se fundamentalmente ao uso do filtro 3x3 para o fator sazonal, indicando que é necessário o uso de filtros de maior comprimento. Ao se introduzir a opção de manipulação de valores extremos se elimina este problema, mas não totalmente, pois a estatística M1 falha. O uso dos filtros 3x5 e 3x9 aumenta a qualidade do ajuste.

Apêndice 6. Desempenho das opções alternativas para a série no período de 1960 a 1985: TX6085, X-11 ARIMA.

A maior parte da variação da série se deve à variação na tendência, que é explicada pela tendência decrescente dos últimos 10 anos (tabela 55).

Tabela 55. Estrutura da variância da porção estacionária da série original (%).

Grupo.	auto	3x3,9-H	3x5,9-H	3x9,9-H	3x3,23-H
T	6	3	4	4	6
C	75	75	76	77	72
S	19	22	20	19	22

O valor ADR para todas as opções (com exceção da opção 23-H) está no intervalo 1,34 e 1,75, aceitando-se a hipótese de aleatoriedade dos resíduos em todos os casos. No caso da opção 23-H há presença de autocorrelação nos resíduos. O menor valor para o XCD é de 3, que ocorre quando se utiliza o filtro 9-H. Observa-se também que, ao se utilizar o filtro 23-H, este valor supera o limite permitido, indicando que este filtro não consegue captar os movimentos subjacentes da série (tabela 56).

Tabela 56. Duração média de sequências (ADR) e o número de meses para coincidência cíclica (MCD).

Teste.	auto	3x3,9-H	3x5,9-H	3x9,9-H	3x3,23-H
SI	1,81	1,82	1,86	1,82	1,80
T	1,57	1,53	1,47	1,50	1,76
C	6,76	4,71	4,86	4,71	10,03
XCD	3,07	2,86	2,81	3,09	4,84
Months	5	3	3	3	7

As opções dos filtros mais curtos mostram as menores razões (tabela 57).

Tabela 57. Razão final entre os componentes.

Razão	auto	3x3,9-H	3x5,9-H	3x9,9-H	3x3,23-H
$\frac{F_1}{F_2}$	2,42	1,34	1,90	1,92	2,47
$\frac{F_1}{F_3}$	5,51	4,57	4,02	5,20	5,56

Em todos os casos a hipótese de presença de sazonalidade sazonal é aceita, pois o valor do p-valor é próximo de zero, e a hipótese de presença de sazonalidade móvel é rejeitada (tabela 58).

Tabela 58. Teste para sazonalidade.

Teste	auto	3x3,9-H	3x5,9-H	3x9,9-H	3x3,23-H
Fs_B1	29,5	28,3	29,5	29,5	29,5
Fs_B2	42,4	40,3	56,0	61,1	35,0
X2_K_M	198,7	202,7	212,9	218,6	187,0
Fa	1,7	1,7	1,6	1,3	2,1
p-valor	0,8	2,4	4,46	13,92	0,17

As opções automática e 23-H,3x3 mostram presença de autocorrelação forte nos resíduos. A opção 3x9,9-H é a única que está livre de autocorrelação (v. tabela 59).

Tabela 59. Função de autocorrelação dos resíduos.

lag k	auto	3x3,9-H	3x5,9-H	3x9,9-H	3x3,23-H
1	0,12*	-0,12*	-0,08	-0,09	0,30*
2	-0,12*	-0,20*	-0,15*	-0,15*	-0,03
3	-0,16*	-0,02	0,02	0,04	-0,19*
4	-0,13*	0,02	0,01	0,03	-0,30*
5	-0,12*	-0,05	-0,04	-0,02	-0,27*
6	-0,04	-0,06	-0,01	0,02	-0,13*
7	-0,00	-0,03	0,01	0,02	0,01
8	0,07	-0,04	-0,02	-0,03	0,10
9	0,14*	0,07	0,08	0,09	0,18*
10	0,10	0,06	0,06	0,05	0,09
11	0,01	0,04	0,03	0,02	-0,05
12	-0,09	-0,16*	-0,15*	-0,11	-0,17*
13	-0,10	0,06	0,00	-0,00	-0,09
14	-0,02	0,06	0,04	0,04	-0,01

A opção 3x5,9-H é a única que não possui evidência de periodicidade nos resíduos (tabela 60).

Tabela 60. Teste de Kolmogorov-Smirnov (%).

Interv.	auto	3x3,9-H	3x5,9-H	3x7,7-H	3x9,23-H
- 75%	64	63	56	38	38
75% - 95%	35	23	14	1	5
+ 95%	1	14	0	19	62

A opção que apresenta melhores valores para as estatísticas de controle de qualidade é a 3x9,9-H (tabela 61). Pode-se observar que a melhoria obtida é devida fundamentalmente ao uso do filtro 9-H para a tendência. Neste caso, a opção dos filtros curtos 3x3,9-H mostra da influência na variabilidade da sazonalidade ao longo de todo o período, medidas por M8 e M10.

Ainda na tabela 61, pode-se observar, que a opção 23-H, mostra várias deficiências ao falharem as estatísticas M1, M4, M5, M6 e M10. Isto é, a presença do componente irregular é alta (M1), há alta correlação (M4) e há predominância do componente irregular em relação à sazonalidade (M5). A falha das estatísticas M6 e M10, isto é, a variabilidade do componente sazonal ao longo de todo o período se deve fundamentalmente ao filtro 3x3 para o componente sazonal, indicando que este filtro pela sua curta amplitude é altamente influenciado pela variabilidade aleatória da sazonalidade.

Tabela 61. As estatísticas de controle de qualidade.

Total	auto	3x3,9-H	3x5,9-H	3x9,9-H	3x9,23-H
M1	1,586*	0,736	0,862	0,870	1,659*
M2	0,557	0,301	0,367	0,373	0,641
M3	0,709	0,429	0,446	0,459	0,735
M4	0,505	0,244	0,174	0,005	1,603 *
M5	0,818	0,407	0,393	0,374	1,270 *
M6	0,602	0,288	0,328	0,511	0,623
M7	0,386	0,353	0,323	0,300	0,432
M8	0,565	1,010*	0,626	0,447	1,145 *
M9	0,356	0,464	0,424	0,394	0,390
M10	0,304	1,084*	0,371	0,293	1,073 *
M11	0,268	0,635	0,231	0,280	0,576
S	0,65	0,50	0,42	0,39	0,92
Abatidas	1	2	---	---	5

Anexo 7. Dados originais da taxa mensal e anual de mortalidade infantil

Anos	jan	fev	mar	abr	mai	jun	jul	ago	set	out	nov	dez	anual
1933	192	175	151	142	141	155	145	138	150	219	202	237	181
1934	248	235	185	164	149	135	131	129	137	174	229	240	172
1935	212	200	197	160	145	139	147	140	141	150	212	260	175
1936	219	193	172	162	150	144	144	145	140	232	237	290	190
1937	215	178	164	143	145	129	128	135	134	182	267	249	172
1938	177	175	147	141	130	133	136	137	131	101	250	260	167
1939	190	160	147	135	136	144	140	134	140	100	244	290	171
1940	200	180	146	140	132	135	129	140	137	167	247	252	178
1941	263	231	171	139	150	138	128	132	156	160	214	202	175
1942	177	153	135	123	115	110	115	107	119	171	207	210	145
1943	185	171	145	126	107	111	112	115	112	132	204	204	140
1944	169	160	149	133	114	110	110	111	111	157	222	251	147
1945	185	164	140	123	117	105	101	100	110	120	145	160	130
1946	120	132	109	97	91	98	95	84	90	113	140	156	110
1947	122	127	111	100	95	89	93	90	90	108	121	140	107
1948	125	133	120	107	99	91	93	96	102	110	151	179	117
1949													111
1950	147	129	114	97	95	97	96	96	101	108	155	171	116
1951	126	117	102	92	90	95	99	91	94	125	172	172	114
1952	138	116	101	95	94	94	83	84	94	98	132	129	104
1953	120	101	96	91	83	89	91	81	88	134	142	146	104
1954	115	100	80	95	92	89	80	77	81	93	126	125	95
1955	120	129	99	91	93	84	82	82	82	102	123	147	102
1956	195	129	99	103	96	83	73	68	65	73	104	111	99
1957	110	104	102	85	81	78	78	68	76	76	102	117	91
1958	117	104	93	86	84	84	74	72	75	77	96	102	89
1959	92	80	82	78	83	80	66	71	71	86	92	104	82
1960	93	90	80	75	77	69	68	60	62	70	90	94	77
1961	90	83	80	72	75	75	65	66	70	72	85	90	77
1962	87	78	73	73	76	84	71	64	65	74	84	86	76
1963	82	86	86	71	76	71	62	67	68	80	98	98	79
1964	82	83	73	71	74	72	69	57	60	61	74	75	71
1965	77	73	68	70	67	69	74	70	67	70	77	90	73
1966	88	83	73	71	76	75	71	69	69	69	80	88	76
1967	79	85	82	74	73	80	70	67	70	68	78	81	75
1968	77	69	68	72	75	71	70	66	65	71	84	97	74
1969	96	99	91	85	83	76	78	75	73	79	84	88	84
1970	83	82	83	84	89	91	89	80	77	77	89	87	84
1971	102	101	90	91	90	99	92	82	74	81	88	86	89
1972	96	92	92	92	81	86	85	77	77	79	84	101	86
1973	106	109	107	93	90	91	80	84	76	69	75	80	89
1974	85	95	88	87	82	91	80	77	68	70	83	93	83
1975	74	96	102	84	84	77	86	69	71	75	94	103	86
1976	115	105	92	82	73	74	77	67	64	63	63	76	79
1977	91	97	85	78	76	67	62	57	54	53	58	66	70
1978	70	66	66	71	70	73	75	70	55	52	61	63	66
1979	64	64	53	61	60	69	65	53	58	58	56	57	59
1980	58	63	62	59	54	55	47	46	42	40	45	50	52
1981	59	67	64	50	46	48	46	39	38	44	48	55	50
1982	60	58	50	55	54	48	47	42	40	39	40	44	40
1983	49	52	48	47	41	41	41	35	32	36	37	41	43
1984	48	50	48	43	44	51	57	47	37	35	39	43	45
1985	41	40	41	43	41	42	34	29	29	29	32	34	37

Lista de gráficos.

Gráfico 1.	Taxa anual de mortalidade infantil do Estado de São Paulo no período de 1933 a 1935	58
Gráfico 2.	Taxa mensal de mortalidade infantil do Estado de São Paulo no período de 1933 a 1935	58
Gráfico 3.	Taxa mensal de mortalidade infantil do Estado de São Paulo: 1933 - 1940	59
Gráfico 4.	Desempenho das revisões da tendência: 1936 - 1941	70
Gráfico 4a.	Desempenho das revisões da tendência das opções com manipulação de valores extremos: 1936 - 1941	70
Gráfico 5.	Desempenho das revisões no maior ponto de inflação: 40-41	71
Gráfico 6.	Desempenho das revisões da tendência: 1940 - 1941	72
Gráfico 6a.	Desempenho das revisões da tendência nos anos: 1942-1943	72
Gráfico 7.	Revisão da sazonalidade: 1941 - 1944, opção automática	73
Gráfico 7a.	Revisão da sazonalidade: 1941 - 1944, opção 3x3,7-H ...	73
Gráfico 8.	Desempenho das estimativas da tendência: 1945 - 1946 ..	74
Gráfico 9.	Desempenho das estimativas da sazonalidade: 1945-1946, opção automática	76
Gráfico 9a.	Desempenho das estimativas da sazonalidade: 1945-1946, opção 3x3,7-H	76
Gráfico 10.	Taxa mensal de mortalidade infantil do Estado de São Paulo: 1950 - 1975	80
Gráfico 11.	Taxa mensal de mortalidade infantil do Estado de São Paulo: 1950 - 1965	82
Gráfico 12.	Taxa mensal de mortalidade infantil do Estado de São Paulo: 1960 - 1975	84
Gráfico 13.	Taxa mensal de mortalidade infantil do Estado de São Paulo: 1970 - 1985	84
Gráfico 14.	Taxa mensal de mortalidade infantil do Estado de São Paulo: 1960 - 1985	86
Gráfico 15.	Desempenho das estimativas da tendência: 1968 - 1969 ..	88

Gráfico 16. Desempenho das revisões do fator sazonal: 1960 - 1969 ..	37
Gráfico 17. Desempenho das revisões do fator sazonal para a opção automática	39
Gráfico 18. Desempenho das revisões do fator sazonal para a opção automática	39

Lista de tabelas.

Tabela A. Pesos do filtro sazonal 3x3	22
Tabela B. Tabela do filtro para a tendência	25
Tabela C. Tabela do filtro sazonal	26

2.2.1 Análise dos resultados do ajuste da opção automática: TX3348

Tabela 1. Ajuste dos modelos ARIMA da opção automática	57
Tabela 2. Ajuste dos modelos ARIMA (2,1,2)(0,1,1) às séries TX3344 e TX3348	58
Tabela 3. Ajuste dos modelos ARIMA opção aditiva à série TX3348 ..	58
Tabela 4. Estrutura da variância da porção estacionária da série original (%)	59
Tabela 5. Duração média de seqüências (ADR) e o número de ondas para dominação cíclica (XCD)	60
Tabela 6. Razão final entre os componentes	60
Tabela 7. Teste para sazonalidade	61
Tabela 8. Função de autocorrelação dos resíduos	61
Tabela 9. Teste de Kolmogorov-Smirnov (%)	62
Tabela 10. As estatísticas de controle de qualidade	62

2.2.2 Desempenho das opções alternativas para a série no período de 1933 a 1948: TX3348, X-11.

Tabela 11. Estrutura da variância da porção estacionária da série original (%)	63
Tabela 12. Duração média de seqüências (ADR) e o número de ondas para dominação cíclica (XCD)	64
Tabela 13. Razão final entre os componentes	64
Tabela 14. Teste para sazonalidade	65
Tabela 15. Teste de Kolmogorov-Smirnov (%)	65
Tabela 16. Função de autocorrelação dos resíduos	65
Tabela 17. As estatísticas de controle de qualidade	66
Tabela 18. O erro médio absoluto (MAE) das seis últimas ondas das diversas séries	70

Índice 1. Desempenho da opção automática no período de 1970 a 1982

Tabela 19.	Ajuste dos modelos ARIMA da opção automática	96
Tabela 20.	Estrutura da variância da porção estacionária da série original (%)	97
Tabela 21.	Duração média de sequências (ADR) e o número de meses para dominação cíclica (MCD)	97
Tabela 22.	Razão final entre os componentes	98
Tabela 23.	Teste para sazonalidade	98
Tabela 24.	Teste de Kolmogorov-Smirnov (%)	99
Tabela 25.	Função de autocorrelação dos resíduos	99
Tabela 26.	As estatísticas de controle de qualidade	99

Índice 2. Desempenho da opções alternativas para a série no período de 1950 a 1975: TX5075, X-11 ARIMA

Tabela 27.	Estrutura da variância da porção estacionária da série original (%)	100
Tabela 28.	Duração média de sequências (ADR) e o número de meses para dominação cíclica (MCD)	100
Tabela 29.	Razão final entre os componentes	100
Tabela 30.	Teste para sazonalidade	101
Tabela 31.	Função de autocorrelação dos resíduos	101
Tabela 32.	Teste de Kolmogorov-Smirnov (%)	102
Tabela 33.	As estatísticas de controle de qualidade	102

Índice 3. Desempenho da opções alternativas para a série no período de 1950 a 1965: TX5045, X-11

Tabela 34.	Razão final entre os componentes	102
Tabela 35.	Estrutura da variância da porção estacionária da série original (%)	103
Tabela 36.	Duração média de sequências (ADR) e o número de meses para dominação cíclica (MCD)	103
Tabela 37.	Teste para sazonalidade	103
Tabela 38.	Função de autocorrelação dos resíduos	104
Tabela 39.	Teste de Kolmogorov-Smirnov (%)	104
Tabela 40.	As estatísticas de controle de qualidade	104

Apêndice 4. Desempenho das opções alternativas para a série no período de 1960 a 1975: TX6075, X-11 ARIMA

Tabela 41.	Razão final entre os componentes	107
Tabela 42.	Estrutura da variância da porção estacionária da série original (%)	107
Tabela 43.	Duração média de sequências (ADR) e o número de meses para dominação cíclica (MCD)	107
Tabela 44.	Teste para sazonalidade	107
Tabela 45.	Função de autocorrelação dos resíduos	107
Tabela 46.	Teste de Kolmogorov-Smirnov (%)	107
Tabela 47.	As estatísticas de controle de qualidade	107

Apêndice 5. Desempenho das opções alternativas para a série no período de 1970 a 1985: TX7085, X-11 ARIMA

Tabela 48.	Razão final entre os componentes	107
Tabela 49.	Duração média de sequências (ADR) e o número de meses para dominação cíclica (MCD)	107
Tabela 50.	Estrutura da variância da porção estacionária da série original (%)	107
Tabela 51.	Teste para sazonalidade	107
Tabela 52.	Função de autocorrelação dos resíduos	107
Tabela 53.	Teste de Kolmogorov-Smirnov (%)	107
Tabela 54.	As estatísticas de controle de qualidade	107

Apêndice 6. Desempenho das opções alternativas para a série no período de 1960 a 1985: TX6085, X-11 ARIMA

Tabela 55.	Estrutura da variância da porção estacionária da série original (%)	110
Tabela 56.	Duração média de sequências (ADR) e o número de meses para dominação cíclica (MCD)	110
Tabela 57.	Razão final entre os componentes	111
Tabela 58.	Teste para sazonalidade	111
Tabela 59.	Função de autocorrelação dos resíduos	111
Tabela 60.	Teste de Kolmogorov-Smirnov (%)	112
Tabela 61.	As estatísticas de controle de qualidade	112

Referências bibliográficas.

- AKAIKE, H. (1980) "Seasonal Adjustment by a Bayesian Modeling". *Journal of Time Series Analysis*, 1, Nº.1, 1-13.
- AKAIKE, H. e ISHIGURO, M. (1983) "Comparative Study of the X-11 and BAYSEA Procedures of Seasonal Adjustment". Institute of Statistical Mathematics, Tokyo, Japan, 17-45.
- ALTMANN, A.M.G. (1982) "O Potencial e o Uso das Estatísticas Vitais de São Paulo: A Experiência do GEADE". SEADE, Informe Demográfico Nº 7.
- ALTMANN, A.M.G. e FERREIRA, C.E. (1982) "A Situação dos Registros dos Fatos Vitais no Brasil". SEADE, Informe Demográfico Nº 7.
- BELL, W.R. e HILLMER S.C. (1984) "Issues Involved with the Seasonal Adjustment of Economic Time Series". *Journal of Business and Economic Statistics*, 2, Nº 4, 291-320.
- BOX, G.E.P, HILLMER, S.C. e TIAO, G.C (1978). "Analysis and Modeling of Seasonal Time Series". em *Seasonal Analysis of Economic Time Series*, Ed. Zellner, U.S. Department of Commerce, Bureau of Census, 309-344.
- BOX, G.E.P. e JENKINS, G.M. (1976). "Time Series Analysis: Forecasting and Control". San Francisco. Holden-Day. 2ª Ed.
- BOX, G.E.P. e PIERCE, D.A. (1970). "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models". *JASA*, 65, nº.332, 1505-1526.
- BULLACH, W.M. (1983) "Modelo Bayesiano de Decomposição Sazonal de Séries Temporais". Dissertação de Mestrado, Instituto de Matemática, UFRJ.
- BURMAN, J.P. (1980) "Seasonal Adjustment by Signal Extration". *JRSS*, A, 143, 321-337.
- BURRIDGE P. e WALLIS, K.F. (1984) "Unobserved-Components Models for Seasonal Adjustment Filters". *Journal of Business and Economic Statistics*, 2, Nº 4, 350-359.
- CLEVELAND, W. e TIAO, G.C. (1976). "Decomposition of Seasonal Time Series. A Model for the Census X-11 Program". *JASA*, 71, 581-587.
- DAGUM, E.B. (1974). "A note on the Seasonal Adjustment of Economic Time Series". Seasonal Adjustment and Time Series Staff, Statistics Canada.
- DAGUM, E.B. (1976). "X-11, The Magic Box and Four Golden Rules of Seasonal Adjustment". Seasonal Adjustment and Time Series Staff, Statistics Canada.

- DAGUM, E.B. (1978a). "Foundations of the Two Main Methodologies of Seasonal Adjustment and the X-11-ARIMA Method". Seasonal Adjustment and Time Series Staff, Statistics Canada.
- DAGUM, E.B. (1978b). "Modelling, Forecasting and Seasonally Adjusting Economic Time Series with the X-11-ARIMA Method". *The Statistician*, 27, 203-215.
- DAGUM, E.B. (1979a). "Further Modifications in the Selection of ARIMA Models for the X-11-ARIMA Seasonal Adjustment Method". Seasonal Adjustment and Time Series Staff, Statistics Canada.
- DAGUM, E.B. (1979b). "Data Extrapolation and Smoothing with the X-11 ARIMA Seasonal Adjustment Method". Seasonal Adjustment and Time Series Staff, Statistics Canada.
- DAGUM, E.B. (1980). "The X-11-ARIMA Seasonal Adjustment Method". Ottawa, Statistics Canada, catalogue No. 12-564E.
- DAGUM, E.B. (1981). "Diagnostic Checks for the ARIMA Models of the X-11 ARIMA Seasonal Adjustment Method". Seasonal Adjustment and Time Series Staff, Statistics Canada.
- DAGUM, E.B. (1982). "The effects of Assimetric Filter on Seasonal Factor Revision". *JASA*, 77, Nº 380, 732-738.
- DAGUM, E.B. (1983a). "Spectral Properties of the Concurrent and Forecasting Seasonal Linear Filters of the X-11 ARIMA Method". *The Canadian Journal of Statistics*, 11, Nº.1, 73-90.
- DAGUM, E.B. (1983b). "Revisions of Seasonally Adjusted Data Due to Filters Changes". Seasonal Adjustment and Time Series Staff, Statistics Canada.
- DAGUM, E.B. e MORRY, M. (1983). "Seasonal Adjustment of Labour Force Series During Recession and Non-recession Periods". Seasonal Adjustment and Time Series, Statistics Canada.
- DURBIN, J. e MURPHY M.J. (1975). "Seasonal Adjustment Based on a Mixed Additive-Multiplicative Model". *JRSS, A*, 138, 385-410.
- FERREIRA, C.E.C. e ORTIZ, L.P. (1982) "Avaliação e Correção do Registro de Nascidos Vivos no Estado de São Paulo, 1900-1975". SEADE, Informe Demográfico Nº 8.
- GAIT, N. (1975) "Ajustamento Sazonal de Séries Temporais". Dissertação de Mestrado, Instituto de Matemática e Estatística, USP.
- GIRALDELLI, B.W. (1982) "O Coeficiente de Separação para óbitos de Menores de Um Ano: São Paulo, 1930/1980". SEADE. Informe Demográfico Nº 8.
- GIRALDELLI, B.W. e SAAD, P.H. (1982) "Estudo da Sazonalidade dos Eventos Vitais no Estado de São Paulo no Período 1930-1980". SEADE, Informe Demográfico Nº 7.

- HARVEY, A.C. e TODD, P.H.J. (1983) "Forecasting Economic Time Series with Structural and Box-Jenkins Models: A Case Study". *Journal of Business and Economic Statistics*, 1, Nº 4, 299-306.
- HARRISON P.J. e STEVENS C.F. (1976) "Bayesian Forecasting". *JRSS,B*, 38, 205-227.
- HENDERSON, R. (1916) "Note on Graduation by Adjusted Average", *Transaction of the Actuarial Society of America*, XVIII, 43-48.
- HIGGOISSON, J. (1975) "An F Test for the Presence of Moving Seasonality When Using Census Method II", *Research Paper, Seasonal Adjustment and Time Series Staff Statistics Canada*.
- HIGGOISSON, J. (1977) "User's Manual for the Bell Canada Model Test". *Seasonal Adjustment and Time Series Staff, Statistics Canada*.
- HILLMER, S.C. e TIAO, G.C. (1982) "AN ARIMA Model-Based Approach to Seasonal Adjustment", *JASA*, 77, 63-70.
- HILLMER, S.C., BELL, W.R., e TIAO, G.C. (1983). "Modeling Considerations in the Seasonal Adjustment of Economic Time Series", em *Applied Time Series Analysis of Economic Data*, Ed. Zellner, U.S. Department of Commerce, Bureau of Census, 309-344.
- HOTTA, L.K. (1984) "Ajuste Sazonal de Séries Temporais", apresentado no VI SINAPE.
- HUOT, G., CHIU, K., HIGGINSON, J. e GAIT, N. (1986) "Analysis of Revisions in the Seasonal Adjustment of Data Using X-11 ARIMA Model-Based Filters". *International Journal of Forecasting*, 2, 217-229.
- KENNY, P.B. e DURBIN, J. (1982). "Local Trend Estimation and Seasonal Adjustment of Economic and Social Time Series", *JRSS*, 145, 1-28.
- LEHMANN, E.L. e D'ABRERA H.J.M. (1974) "Nonparametrics Statistical Methods Based on Ranks", San Francisco, Holden-day Inc.
- LOTHIAN, J. (1978). "The Relationship Between the Additive, Multiplicative and Logarithmic Models in the X-11 ARIMA Seasonal Adjustment Program". *Seasonal Adjustment and Time Series Staff, Statistics Canada*.
- LOTHIAN, J. e MORRY, M. (1978a). "A Set of Quality Control Statistics for the X-11-ARIMA Seasonal Adjustment Method". *Research Paper, Seasonal Adjustment and Time Series Staff, Statistics Canada*.
- LOTHIAN, J. e MORRY, M. (1978b). "A Test for the Presence of Identifiable Seasonality When Using the X-11-ARIMA Program". *Research Paper, Seasonal Adjustment and Time Series Staff, Statistics Canada*.
- LOTHIAN, J. e MORRY, M. (1978c). "Selection of Models for the Automated X-11-ARIMA Seasonal Adjustment Program". *Seasonal Adjustment and Time Series Staff, Statistics Canada*.

- MARGONE CAMARGO, A.B. e CUNHA, J.M. P. (1982) "Considerações Preliminares Sobre os Fenômenos de Invasão e Evasão de óbitos em São Paulo". SEADE, Informe Demográfico Nº 7.
- MONTEIRO, C.A. (1982). "Contribuição para o Estudo do Significado da Evolução do Coeficiente de Mortalidade Infantil do Estado de São Paulo, SP (Brasil) nas três últimas décadas (1950-1979)", Rev. Saúde Pública, SP, 16, 7-18.
- MORETTIN, P. e TOLDI, C.M. (1981). "Modelos Para Previsão de Séries Temporais", IMPA.
- ORTIZ, L.P. (1982a) "Evolução da Mortalidade Infantil do Estado de São Paulo Segundo Sexo, Idade e Causa de Morte". SEADE, Informe Demográfico Nº 8.
- ORTIZ, L.P. (1982b) "Avaliação das Estatísticas de óbitos de Menores de 1 ano". SEADE, Informe Demográfico Nº 8.
- PIERCE, D.A. (1980a). "A Survey of Recent Developments in Seasonal Adjustment". The American Statistician, 34, No.3, pp. 125-134.
- PIERCE, D.A. (1980b). "Data Revision with Moving Average Seasonal Adjustment procedures". Econometrics, 14, pp.95-114.
- SHISKIN, J., MUSGRAVE, J.C. e YOUNG, A.H. (1967). "The X-11 Variant of Census Method II, Seasonal Adjustment". Technical Paper No. 15, Bureau of Census, U.S. Dept. of Commerce.
- SHISKIN, J. e PLEWES, T. (1978). "Seasonal Adjustment of the U.S. Unemployment Rate". The Statistician, 27, Nos. 3 e 4, pp. 181-202.
- TIAO, G.C. e HILLMER S.C. (1978). "Some Consideration of Decomposition of a Time Series". Biometrika, 65, 3, 97-502.
- WALLIS, K.F. (1974). "Seasonal Adjustment and Relations Between Variables". JASA, 69, 18-31.
- WALLIS, K.F. (1982). "Seasonal Adjustment and Revision of Current Data: Linear Filters for the X-11 Method". JRSS, A, 145, 74-85.
- WALLIS, K.F. e MOORE, G.H. (1941). "A Significance Test for Time Series". National Bureau of Economic Research, Technical Paper Nº1.
- WHITTLE, P. (1963) "Prediction and Regulation by Linear Least Square Methods". London English Universities Press.
- YOUNG, A.H. (1968) "Linear Approximations to the Census and BLS Seasonal Adjustment Method". JASA, 63, 445-475.