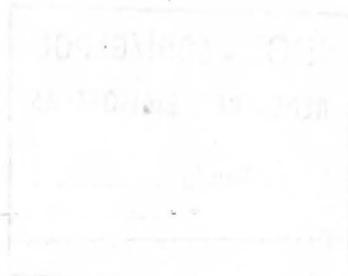


INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA - IBGE

ESCOLA NACIONAL DE CIÊNCIAS ESTATÍSTICAS - ENCE



RELATÓRIOS TÉCNICOS

Nº 05/89

UMA REVISÃO DO BATS - BAYESIAN ANALYSIS OF TIME SERIES
Pacote para Microcomputadores tipo PC

H.S. MIGON - ENCE/IBGE & IM/UFRJ
E. SUYAMA - ICEx/UFMG

RIO DE JANEIRO
Dezembro/89

GE 0000 9879.1

IBGE - CDDI/GEDOC
REDE DE BIBLIOTECAS
N.º de Reg: <u>493</u>
Data: <u>16/3/90</u>

31 (047.3)

R 382 N

S189

UMA REVISÃO DO BATS - BAYESIAN ANALYSIS OF TIME SERIES

Pacote para Microcomputadores tipo PC

H.S.Migon - ENCE/IBGE & IM/UFRJ

E. Suyama - ICEX / UFMG

1 - INTRODUÇÃO

O BATS é um pacote para análise de séries temporais, desenvolvido por West, Harrison & Pole(1988), usando métodos Bayesianos de previsão. Para um bom entendimento de seu funcionamento e para discutir criticamente suas facilidades e dificuldades de uso será necessário resumir as principais características da modelagem Bayesiana de séries temporais. Assim na sessão 2 apresentaremos os aspectos centrais da modelagem dinâmica, enfatizando a característica sequencial do método, as possibilidades de intervenções, o tratamento de observações perdidas e o uso de fatores de desconto. A sessão-3-será dedicada à descrição do BATS e incluirá aspectos de documentação, entrada e saída de dados, manipulação de dados, facilidades gráficas, etc. Um exemplo será apresentado na sessão 4, e, finalmente as vantagens e desvantagens do pacote serão discutidas na sessão 5. Além disto seremos abrangentes nas referências bibliográficas visando facilitar os novos usuários desta poderosa e promissora metodologia para a análise de séries temporais.

2 - ASPECTOS BÁSICOS DA MODELAGEM BAYESIANA DINÂMICA

A modelagem Bayesiana dinâmica foi introduzida por Harrison & Stevens (1976) e desde então várias facilidades foram desenvolvidas. As principais referências bibliográficas nesta área são : West (1986), West & Harrison (1986), West, Harrison & Migon (1985), Migon & Harrison (1985), e o texto de West & Harrison (1989) recém publicado. O uso limitado dessa técnica deve-se à pouca divulgação da metodologia, à falta de um pacote de caráter geral, às dificuldades de natureza técnica no entendimento e uso do método, o qual demanda um conhecimento profundo para operá-lo satisfatoriamente. Tudo isto implica

num alto investimento por parte dos usuários, traduzido pelo custo de aprendizado da técnica, desenvolvimento de programas, etc.

Na década de 80 as idéias Bayesianas ressurgiram vigorosamente na comunidade estatística refletindo-se na área de modelagem dinâmica. Problemas metodológicos foram solucionados introduzindo-se o uso de fatores de desconto, estimação sequencial da variância observacional, e introdução de técnicas não lineares.

O desenvolvimento de um pacote simples de uso e que contivesse as facilidades inerentes ao método Bayesiano, como tratamento de observações perdidas, análise do tipo 'what/happened', monitoramento dos erros de previsão, facilidades de intervenção, uso de fatores de desconto e modelagem da variância observacional, seria o último elemento necessário para a materialização das aplicações do método Bayesiano.

2.1. Uma Classe de Modelos Lineares Dinâmicos

No BATS implementou-se um subconjunto de modelos lineares dinâmicos que se propõe a tratar problemas de séries temporais univariadas e de regressão com parâmetros variando. As séries temporais univariadas são observadas em pontos equiespaçados do tempo.

Denotando-se por y_t a série temporal de interesse, em cada instante de tempo podemos representá-la por duas componentes: nível subjacente μ_t e erro observacional ν_t ; assim:

$$y_t = \mu_t + \nu_t \quad (1)$$

O nível, μ_t , compreende tres componentes representando modelos locais para a tendência (p_t), sazonalidade (s_t) e efeitos de regressores (r_t). A tendência é modelada por polinômios de baixa ordem e a sazonalidade pode ser descrita por fatores sazonais (forma livre) ou por uma representação através de harmônicos. Destacam-se duas alternativas para relacionar essas componentes:

$$\mu_t = \begin{cases} (p_t + r_t) + s_t & \text{aditivo} \\ (p_t + r_t) * s_t & \text{multiplicativo} \end{cases} \quad (2)$$

Desta forma, a série temporal é decomposta em: erro observacional, tendência, sazonalidade e componentes de regressão.

A análise Bayesiana de séries temporais será condicional ao modelo especificado e aos dados observados. Denote-se por D_t as informações históricas disponíveis até o tempo t . A evolução futura de uma série será descrita pelo valor esperado de cada uma de suas componentes.

A seguir descreveremos com mais detalhes as componentes de tendência, sazonalidade, regressão e erro observacional do modelo especificado.

Tendência polinomial

Esta componente representa uma aproximação local para a tendência subjacente à série temporal y_t . Destacam-se os polinômios de 1º e 2º graus conhecidos como modelo estável e modelo de crescimento linear, uma vez que suas trajetórias futuras são, respectivamente, constante e linear. A característica local da modelagem impõe que os parâmetros do modelo sejam considerados como variáveis aleatórias. A título de exemplo considere-se o modelo de crescimento linear, modelo polinomial de 2ª ordem. Esta trajetória é descrita por $p_{t+k} = a + bk$. Definindo-se as equações de recursão:

$$\begin{aligned}\theta_{1,t} &= \theta_{1,t-1} + \theta_{2,t-1} \\ \theta_{2,t} &= \theta_{2,t-1}\end{aligned}\tag{3}$$

é fácil obter, por substituições sucessivas, que:

$$\begin{aligned}\theta_{1,t+k} &= [\theta_{1,t+k-1}] + \theta_{2,t+k-1} \\ &= [\theta_{1,t+k-2} + \theta_{2,t+k-2}] + \theta_{2,t+k-1} \\ &\vdots \\ &= \theta_{1,t} + [\theta_{2,t+1} + \dots + \theta_{2,t+k-1}] \\ &= \theta_{1,t} + k\theta_{2,t}\end{aligned}\tag{4}$$

implicando que $a = \theta_{1,t}$ e $b = \theta_{2,t}$.

Em geral temos:

$$\begin{aligned}y_t &= p_t + \nu_t && \text{Equação da observação} \\ p_t &= \theta_{1,t} && \text{Tendência local} \\ \theta_{1,t} &= \theta_{1,t-1} + \theta_{2,t-1} + \omega_{1,t} && \text{Equações dos parâmetros} \\ \theta_{2,t} &= \theta_{2,t-1} + \omega_{2,t}\end{aligned}\tag{5}$$

onde: $(\omega_1, \omega_2)_t \sim N[0, W_t]$ é a componente estocástica da evolução dos parâmetros.

Sazonalidade

De forma análoga, a componente sazonal será de natureza estocástica para adaptar-se suavemente no tempo. Se o período sazonal for 'p' podemos, então modelar a sazonalidade através de p fatores sazonais, os quais representarão desvios em torno de um nível (considerando ou não os regressores) e portanto satisfazendo a restrição de que sua soma seja nula. Alternativamente poderíamos modelar a sazonalidade em forma multiplicativa. Neste caso os fatores sazonais multiplicativos deveriam somar exatamente 'p'. Formas restritas e mais parcimoniosas (menor número de parâmetros) podem ser obtidas usando-se uma representação através de harmônicos. Por exemplo, em uma particular aplicação pode ser que somente o 1^o e 2^o harmônicos sejam necessários para modelar os fatores sazonais.

Regressores

Para cada variável independente (exógena) - x_t - teremos um parâmetro - β_t - o qual pode variar suavemente no tempo. Parâmetros variando são úteis para modelar omissões de variáveis, mudanças estruturais, modelagem de relações não lineares através de aproximações locais, etc. O caso de regressão múltipla é obtido simplesmente pela superposição dos diversos regressores simples.

Componente Residual

A componente residual ou erro observacional será modelada por uma distribuição normal de média zero e variância descrita pela lei do tipo potência:

$$Var(\nu_t) = c\mu_t^a \quad (6)$$

onde: c é o fator de escala da lei de variância e $a > 0$ assegura que a variância residual seja função do nível da série. É claro que $a = 0$ corresponderá à variância constante, e também estamos supondo que o nível é estritamente positivo quando se aplica a lei de variância do tipo potência. Uma relação entre a família de transformações de Box-Cox (1964), isto é

$$y_t^{(\lambda)} = \begin{cases} y^\lambda / \lambda & \text{se: } \lambda \neq 0 \\ \log(y) & \text{se: } \lambda = 0 \end{cases} \quad (7)$$

e a lei de variância acima é dada por: $\lambda = 1 - a/2$. Do ponto de vista prático é fundamental evitar-se transformar os dados originais de forma a facilitar a interpretação dos resultados e permitir que o previsor se comunique com o modelo incorporando eventuais informações subjetivas.

2.2. Principais Passos da Modelagem Dinâmica Bayesiana

Nesta sessão apresentaremos um sumário dos passos envolvidos na análise Bayesiana. A natureza sequencial do método se adapta muito bem aos problemas de séries temporais. As principais operações são: evolução e atualização. Na primeira operação faz-se evoluir temporalmente as distribuições que descrevem os parâmetros do modelo, ou seja são obtidas as distribuições a priori dos parâmetros. Após observar-se uma nova realização da série temporal a distribuição a priori é misturada com a verossimilhança, gerando a distribuição a posteriori dos parâmetros. Duas outras operações serão também descritas: predição e suavização. A operação de predição é a própria razão do sistema e fornece a distribuição de uma observação futura com base nos dados passados. Esta distribuição denomina-se distribuição preditiva e é obtida integrando-se sobre os parâmetros a distribuição conjunta da observação futura e dos parâmetros, condicional aos dados passados. Por outro lado, a suavização permite descrever a evolução dos parâmetros considerando-se todos os dados observados. É razoável conceber-se que os dados futuros trazem informações relevantes sobre os parâmetros no presente. Um dos usos da suavização diz respeito a análises do tipo 'what happened', como em Harrison(1989).

Modelo Dinâmico Geral

Neste artigo suporemos que as observações se distribuem segundo uma normal. Uma classe mais geral de distribuições pode ser utilizada na modelagem dos dados, como proposto nos modelos lineares generalizados (Migon (1984), West, Harrison & Migon (1985), e Migon & Gamerman(1989)).

Nossas observações, condicionalmente aos parâmetros, serão descritas pela seguinte distribuição de probabilidade:

$$(y_t/\theta_t, \phi_t) \sim N[\mu_t, \mu_t^a \phi_t^{-1}] \quad (8)$$

onde: $\phi_t^{-1} = c_t$ e $c_t \mu_t^a$ representa a lei de variância do tipo potência (6). O subscrito em ϕ_t indica o ganho de informação sobre o parâmetro c_t ao longo do tempo, ao contrário do que ocorre com outros parâmetros que podem evoluir dinamicamente. O nível do processo está relacionado com um vetor de parâmetros θ através da expressão (2), a qual pode, genericamente, ser representada por:

$$\mu_t = F_t(\theta_t) \quad (9)$$

onde: $F_t(\cdot)$ é, em geral, uma função linear. Quando este não for o caso, usa-se uma aproximação de primeira ordem.

Os parâmetros de estado do sistema evoluem segundo a função de transferência, sem perda de generalidade, linear:

$$\theta_t = G_t \theta_{t-1} + \omega_t \quad (10)$$

onde: $\omega_t \sim N[0, W_t/\phi_t]$ e G_t é uma matriz adequada para representar as componentes anteriormente descritas. Por exemplo, no modelo polinomial de 2ª ordem, para obter as relações (3), faz-se:

$$G_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

As informações necessárias no tempo $t - 1$, para implementar a análise Bayesiana estão descritas pelas seguintes distribuições à posteriori:

$$\begin{aligned} (\theta_{t-1}/D_{t-1}, \phi_{t-1}) &\sim N[m_{t-1}, C_{t-1}/\phi_{t-1}] \\ (\phi_{t-1}/D_{t-1}) &\sim \Gamma[n_{t-1}/2, s_{t-1}/2] \end{aligned} \quad (11)$$

No instante inicial o usuário deverá fornecer os valores m_0, C_0, n_0 e s_0 , os quais caracterizam as distribuições a posteriori dos parâmetros no instante zero. O próximo passo na análise será descrever a evolução temporal usando (10).

Evolução Temporal

Os parâmetros de estado, que descrevem os aspectos essenciais de uma série temporal, evoluirão do período $t - 1$ para o período t . A expressão (10) descreve, através da matriz G_t , a evolução sistemática do modelo, enquanto que o termo estocástico $\omega_t \sim N[0, W_t/\phi_t]$ produz um incremento da incerteza associada aos parâmetros de estado. Assim:

$$\begin{aligned} (\theta_t/D_{t-1}, \phi_t) &\sim N[a_t, R_t/\phi_t] \\ (\phi_t/D_{t-1}) &\sim \Gamma[\beta n_{t-1}/2, \beta s_{t-1}/2] \end{aligned} \quad (12)$$

onde: $a_t = G_t m_{t-1}$, $R_t = G_t' C_{t-1} G_t + W_t$ e β é um número entre 0 e 1, denominado fator de desconto (Ameen & Harrison (1985)), representando a perda de informação ao longo do tempo. Vale notar que a evolução da precisão ϕ_t é tal que mantém a média, ampliando somente a incerteza associada à sua distribuição. Estas distribuições são denominadas distribuições a priori. Por integração é possível obter-se a distribuição a priori marginal de θ_t , dadas as informações passadas D_{t-1} , que será uma t-Student.

Atualização

Após observar-se y_t combinamos a distribuição a priori conjunta de θ_t e ϕ_t , dado D_{t-1} , com a verossimilhança para se obter as distribuições a posteriori de θ_t e ϕ_t , isto é:

$$\begin{aligned} (\theta_t / \phi_t, D_t) &\sim N[m_t, C_t / \phi_t] \\ (\phi_t / D_t) &\sim \Gamma[n_t / 2, s_t / 2] \end{aligned} \tag{13}$$

Estas expressões são obtidas como função do erro de previsão um passo à frente e dos parâmetros da distribuição a priori e podem ser encontradas no texto de West & Harrison(1989). Vale comentar que este processo de estimação não envolve inversões de matrizes o que o torna computacionalmente eficiente.

Predição

O objetivo central da análise Bayesiana de séries temporais é gerar modelos robustos de previsão. Uma observação futura é descrita por uma distribuição condicional aos dados passados, denominada distribuição preditiva: (y_{t+k} / D_t) a qual é t-Student, no caso de observações normais.

O previsor, usando sua função de utilidade, escolherá como previsão o valor que maximize sua utilidade esperada. Esta esperança será com respeito à distribuição preditiva.

Outro ponto importante é que, dispondo-se da distribuição conjunta das previsões futuras, y_{t+1}, \dots, y_{t+k} , pode-se facilmente obter a distribuição de funções desses valores futuros. Por exemplo, poderíamos estar interessados na soma das previsões da produção industrial ao longo do ano de 1990, baseada em dados mensais até dezembro de 1989.

Intervenção e Observações perdidas

A natureza subjetiva do método Bayesiano permite que se combine informações objetiva com outras de natureza subjetiva. Um exemplo simples seria considerar alterações

estruturais que ocorreriam no futuro e que tivessem sido assessadas por especialistas. Assim o nível do processo poderia, por exemplo, passar de μ_t para $\mu_t + \delta_t$, onde δ_t é uma variável aleatória de média m_δ e variância V_δ . Note que através desta componente podemos alterar o nível do processo bem como aumentar a variabilidade desta componente.

O tratamento das observações perdidas é muito mais simples, pois basta que nas evoluções (passagem da priori para posteriori) não se incorpore nenhuma informação nova. Isto equivale a manter a descrição dos parâmetros inalterada a menos de ampliar as incertezas.

Suavização

Como comentado anteriormente, os parâmetros do modelo estrutural dinâmico têm interpretação física muito clara e variam ao longo do tempo. A título de ilustração, no modelo de crescimento linear com dados mensais, um dos parâmetros representa o crescimento mensal do processo em estudo; outra componente do modelo poderia ser uma particular variável exógena, e seu coeficiente interpretado como uma elasticidade. Permitir que estes parâmetros variem no tempo é compreender que os aspectos qualitativos dos modelos são mais estáveis do que os aspectos quantitativos. Assim, os parâmetros estarão aptos a acomodar mudanças estruturais que venham a ocorrer. Prosseguindo com este raciocínio não é difícil conceber a importância de se rever os parâmetros no instante t após observar os dados futuros, $t+1, \dots, T$. A distribuição dos parâmetros condicionada em todos os dados futuros será chamada de *suavização* e denotada por:

$$(\theta_t / D_T, \phi_t) \sim N[m_t^s, C_t^s / \phi_t] \quad (14)$$

onde: m_t^s é o vetor de médias suavizadas e C_t^s é a matriz de variâncias-covariâncias suavizadas. Estes valores permitem avaliar a importância dos modelos dinâmicos, descrever efeitos de intervenções, etc.

3 - DESCRIÇÃO DO BATS

O BATS, desenvolvido pelo 'Bayesian Forecasting Group' da Warwick University, é um pacote, em linguagem APL, para análise de séries temporais em micro-computador do

tipo PC. Trata-se de um sistema iterativo, dirigido através de menus conectados hierarquicamente. Em qualquer estágio de uma sessão se tornam disponíveis ao usuário várias facilidades, que podem ser executadas através da seleção de um menu apropriado. Em cada estágio ficam disponíveis automaticamente apenas as operações possíveis.

3.1. *Entrada / Saída de Dados*

Os dados que podem ser lidos/gravados pelo BATS são arquivos sequenciais APL ou nativos DOS (ASCII). Permitida também a entrada de dados através do teclado.

3.2. *Manipulação dos Dados*

Uma vez lido um arquivo de dados, várias operações de manipulação se tornam possíveis:

a) criar novas variáveis através dos comandos:

Plus, Minus, Times, Div, Log, Exp, Abs, Difference, Lag, Std (para padronização), e **Power**.

b) agregar dados, por exemplo: dados mensais em trimestrais, semestrais, etc.

c) deletar variáveis do arquivo.

d) modificar ou acrescentar observações.

e) selecionar os dados correspondentes a um período de tempo para a análise.

3.3. *Análise Exploratória*

São disponíveis: **histograma** com sumário estatístico, **qqplot**, **periodograma**, **função de autocorrelação**, **gráficos** da série temporal e **diagrama de dispersão**.

Os gráficos são iterativos: após criação na tela, podem ser adicionadas notas, linhas, setas, etc., antes de serem impressos. Permite-se também sua gravação para posterior uso ou revisão.

3.4. *Escolha do Modelo*

Podem ser selecionados os modelos estruturais com:

a) tendência: estável ou polinomial;

b) regressores;

c) sazonalidade aditiva ou multiplicativa;

d) lei de variância do tipo $c\mu_t^a$, onde se especifica o parâmetro a , enquanto o parâmetro c é estimado sequencialmente.

Para cada um desses itens podem ser especificados pelo usuário dois fatores de desconto, um para estimação em condições normais, e outro para os períodos de exceção.

3.5. Estimação

A estimação é sequencial, podendo ser iniciada com a especificação da distribuição a priori (em geral, média e desvio padrão de uma normal), ou com o uso da priori de referência (default), para cada parâmetro do modelo.

Todo o processo de estimação é acompanhado na tela, onde num gráfico temporal aparecem os pontos observados e previstos. Após a estimação, são mostradas três medidas de ajuste do modelo: erro médio quadrático e erro médio absoluto, com base nas previsões um passo à frente e a verossimilhança do modelo.

Os resultados da estimação se tornam, então, disponíveis em forma gráfica, ou em tabelas. Podem ser obtidos os resultados: da previsão um passo à frente, ou de uma análise retrospectiva (suavização). Caso os dois sejam desejados, a análise retrospectiva deve ser requisitada por último.

Os gráficos dos valores estimados aparecem com limites de 90% de probabilidade, exceto do parâmetro de escala 'c' e do desvio ($c\mu^a$)⁵. Nos gráficos de previsão da série, de tendência, e do efeito de um regressor (coeficiente - β_t - versus valor do regressor) sobrepõem-se os valores observados da série. Outros gráficos que podem ser requisitados: parâmetros da tendência (por ex. fator de crescimento), coeficientes de regressão, fatores sazonais, resíduos (padronizados ou não).

As principais tabelas são: previsão ou ajuste da série com seus desvios padrões, e componentes do modelo; sumário da distribuição preditiva final das componentes do modelo.

3.6. Previsão

O BATS produz previsões passos à frente a partir de qualquer ponto, usando a priori corrente, ou especificada pelo usuário. Isto pode ser usado também para verificar as implicações de uma priori escolhida pelo usuário, antes de se ajustar um modelo. Limites de 90% de probabilidade são dispostos em gráfico (e tabela) até o horizonte desejado, junto com um número pré-especificado de observações passadas.

3.7. Intervenção

Observações individuais podem ser rejeitadas como 'outliers', ou adaptadas para mudanças estruturais na série através de modificações nos parâmetros do modelo, através de especificação do usuário, ou automaticamente pelo acréscimo à(s) variância(s) da(s) priori(s) da(s) componente(s) afetada(s).

Modificações através do usuário incluem: reespecificar a priori, alterar os fatores de desconto, reespecificar pontos como 'outliers' ou pontos de intervenção, suspender ou abortar a sessão.

Existe também a possibilidade de monitorar a previsão, através de fatores de Bayes (West-1986), os quais são semelhantes às somas cumulativas no caso de normalidade. O usuário deve optar por intervir ou não a cada momento sinalizado pelo monitor.

4 - EXEMPLOS DE USO DO BATS

No modelo de índice de Sharpe (1963) a rentabilidade (R), de uma ação se relaciona com o índice (I), do mercado através de uma função linear:

$$R_t = \alpha_t + \beta_t I_t + \epsilon_t \quad (15)$$

Por outro lado, o índice de mercado é suposto variar aleatoriamente em torno de uma média:

$$I_t = \mu_M + \epsilon_{M,t} \quad (16)$$

A formulação dinâmica deste modelo pode produzir previsões mais realistas ao admitir variação dos parâmetros envolvidos ao longo do tempo como exemplificado em Johnston & Harrison (1983).

O exemplo apresentado pode ser examinado com mais detalhes em Migon & Suyama (1990). Para obter normalidade, a rentabilidade foi definida como $(p_t - p_{t-1})/p_t$, onde p_t é o preço do título no tempo t . Note que no denominador o índice do tempo é t , e não $t - 1$ como o usual.

Exemplo 1

Neste exemplo, tratamos da relação entre a rentabilidade da ação do Banco do Brasil PP (RBB) e a rentabilidade do índice IBOVESPA (RIBVSP). Os gráficos das duas séries, entre 02/80 e 08/88 encontram-se na figura 1, e o gráfico do diagrama de dispersão na figura 2, onde se pode denotar haver correlação positiva entre as variáveis.

Os fatores de desconto e os valores iniciais dos parâmetros utilizados no modelo de regressão dinâmica foram:

Parâmetros	Média	D.Padrão	F.Desconto
tendência	0	.20	.98
regressor	0	.25	.95
variância obs.	default	default	default

O modelo é estimado, de forma sequencial, produzindo-se, passo a passo, a figura 3. A qualidade de previsão do modelo é dada pelas estatísticas, baseadas nas últimas 100 observações :

erro médio quadrático (MSE): .01855

desvio absoluto medio (MAD): .1014

log da verossimilhança: -5031

As figuras 4 a 9 se referem às estimativas sequenciais dos parâmetros e das previsões baseadas nas informações parciais até o instante t, incluindo limites de confiança de 90% de probabilidade.

A distribuição a posteriori dos parâmetros do modelo, no final dos dados pode ser requisitada, fornecendo, neste exemplo:

Parâmetros	Média	D.Padrão
tendência	.015	.020
regressor	.907	.156

As figuras que se seguem, de 10 a 15, são análogas às já apresentadas, tratando-se das

estimativas suavizadas, usando todos os dados disponíveis.

Exemplo 2

Neste exemplo, tratamos da série de rentabilidade do índice IBOVESPA (RIBVSP), com intervenção em 03/86, por ocasião do plano cruzado.

Os fatores de desconto e os valores iniciais dos parâmetros são:

Parâmetros	Média	D.Padrão	F.Desconto
tendência	0.10	.05	.85/.10
variância obs.	default	default	.98/.90

Intervenção: automática em 03/86, considerando a observação uma exceção.

A mudança de tendência foi significativa, após a intervenção, como se pode verificar na fig. 16. A variância, no entanto, parece crescer ao longo do tempo (fig. 17).

5 - VANTAGENS E DESVANTAGENS

Na versão 1.3 do BATS, pacote para análise de séries temporais em micro-computador do tipo PC, destacamos algumas vantagens e desvantagens que serão expostas nesta seção.

Vantagens

portabilidade – ocupa 210 K, isto é, cerca de 2/3 de um disquete de 5' 1/4, e depende do interpretador do APL.

custo – é desprezível tanto para pesquisadores como para pequenas empresas ou firmas de consultoria visto que depende somente de um micro tipo PC com pelo 256K de memória RAM.

simplicidade de uso – por ser baseado em menus organizados de forma hierárquica, independentemente de conhecimentos específicos de APL.

documentação – bastante clara e completa, contendo um breve resumo metodológico e vários exemplos ilustrativos. O manual é escrito de forma clara, dispensando o

conhecimento de qualquer linguagem de computador para sua leitura. Está dividido em 14 capítulos, agrupados em tres seções e uma introdução. Na seção A: **Modelo estatístico inicialização e resultados**, descrevem-se sucintamente : as componentes do modelo, a idéia de previsão/suavização, definições iniciais e resultados finais de uma análise. É na seção B: **Facilidades do BATS**, que se encontram os menus e as opções disponíveis no Bats. A seção C: **Ilustração**, traz exemplos de análise com o Bats. O primeiro exemplo, usando priori não informativa, é bastante detalhado, servindo como um 'tutorial' para o usuário. Outros exemplos mostram o uso de priori informativa, e usos de intervenções, tanto em pontos pré-especificados pelo usuário, quanto em pontos detectados pelo monitor.

facilidades gráficas - os gráficos são de boa resolução e podem ser impressos numa ploter ou numa impressora matricial comum. É possível também colocar-se títulos e símbolos especiais para melhor documentar as figuras.

acesso á dados - geral o suficiente para permitir uma fácil comunicação com outros pacotes.

completitude - por permitir que o usuario faça desde análise exploratória dos dados até a elaboração de modelos razoavelmente complexos.

Desvantagens

A primeira e mais importante crítica é a limitação do pacote no que diz respeito a **salvar uma particular aplicação**. Para melhor fixar as idéias, o usuário após experimentar diferentes modelos para uma particular série de tempo não tem como salvar esta especificação para uso futuro. É verdade que são raros os produtos para microcomputador do tipo PC desenvolvidos segundo esta concepção; no entanto, é também verdade que as aplicações reais demandam fortemente tal possibilidade. É natural, portanto desejar-se que uma sessão de trabalho seja salva para posterior implementação e complementação, isto é salvar a 'pilha' de modelos especificados (parâmetros e componentes), os dados lidos e criados, bem como os gráficos, etc.

Outra crítica, em essência similar a anterior, é que o usuário não pode dar **continuidade a uma análise** a medida que chegam novos dados. Suponha, po exemplo, que numa aplicação real deve-se prever, no curto prazo, as vendas de uma centena de produtos para fins de controle de estoque. A cada mes obtêm-se uma nova informação e deseja-se

rever as previsões futuras a partir deste ponto. A natureza seqüencial do método permite resumir toda informação passada na posteriori - m e C - e recalculas as previsões a partir da consideração da nova observação e eventualmente novos fatores de desconto.

Algumas outras críticas de natureza mais acadêmica seriam a dificuldade de:

a julgar modelos com base nos erros de predição para horizontes maiores do que 1;

b se construir a **distribuição preditiva** para indicadores derivados da série originalmente predita. É comum nas aplicações se necessitar fazer cálculos a partir das previsões futuras, por exemplo acumulando-as ou calculando taxas de crescimento. Seria interessante dispor-se de facilidades de construir a distribuição de probabilidade dessas funções do vetor de previsões.

Finalmente vale ressaltar que são vários os artigos recentes que buscam modelar a variância observacional. O Bats poderia fornecer outras **formas de lei de variância** além do tipo potência. Por exemplo em aplicações financeiras seria interessante poder usar uma variável auxiliar na definição dessas variâncias.

Agradecimentos

O primeiro autor agradece o apoio da FAPERJ, projeto E29/170954/89.

6 - BIBLIOGRAFIA

- Ameen, J. R. M and Harrison, P. J. (1985) Normal discount Bayesian models, in: J. M. Bernardo et al. (Eds.) *Bayesian Statistics 2* (University Press, Valencia).
- Box, G. E. P. and Cox, D. R. (1964) An analysis of transformations, *Journal of the Royal Statistical Society, Series B*, 26, 211-252.
- Harrison, P. J.(1989) Bayesian Forecasting in O.R. *Tech. Report- Warwick University*.
- Harrison, P. J. and Stevens, C. F. (1976) Bayesian forecasting. *Journal of the Royal Statistical Society, Series B*, 38, 205-247.
- Harrison, P. J. and Johnston, F. R. (1983) Discount Weighted Regression. *Journal of Operational Research Society*, vl.35, no. 10, 923-932.
- Migon, H. S. (1984) An approach to non-linear Bayesian forecasting problems with applications, *unpublished Ph. D. thesis*, University of Warwick.
- Migon, H. S. and Harrison, P. J. (1985) An application of non- linear Bayesian forecasting to television advertising, in: *J. M. Bernardo et al. (Eds) Bayesian Statistics 2*, 681-696.
- Migon, H. S. and Gamerman, D. (1989) Generalized exponential growth models: a Bayesian approach, technical report no. 41, *Laboratório de Estatística*, Federal University of Rio de Janeiro.
- Migon, H.S. and Suyama, E. (1990) Análise de portfolio usando regressão dinâmica , *relatório tecnico no. 001/90 - ENCE/IBGE*.
- Sharpe, W. F. (1963) A simplified model for portfolio analysis, *Managment Science* 9, 277-293.
- West. M. (1986) - Bayesian Model Monitoring. *J.Royal Statisitcal Soc.(ser.B)*, 48,70-78.
- West, M., Harrison P. J. and Migon, H. S. (1985) Dynamic generalized linear models and Bayesian forecasting, *Journal of the American Statistical Association*, 80, 73-97.
- West, M. and Harrison, P. J. (1986) Monitoring and adaptation in Bayesian forecasting models, *Journal of the American Statistical Association*, 81, 741-750.
- West, M., Harrison P. J. and Pole, A. (1988) Bats - A user guide - Bayesian Analysis of Time Series - *release 1.9 - Jun 88, University of Warwick*.
- West, M., and Harrison P. J. (1989) Bayesian Forecasting and Dynamic Models, *Springer Verlag*.

FIG.1: SÉRIES DE RENTABILIDADE DA AÇÃO DANCO DO BRASIL ...
(RBB E RENTABILIDADE DO IBOVESPA (RIBVSP))

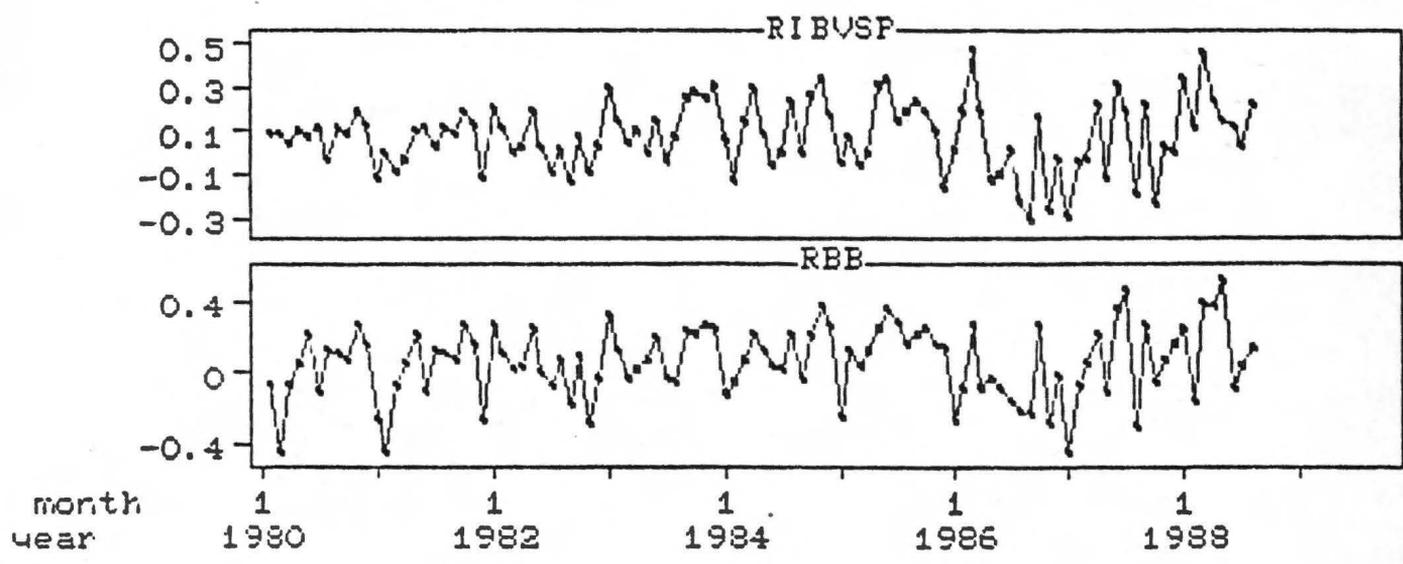


FIG.2: DIAGRAMA DE DISPERSÃO RBB X RIBVSP

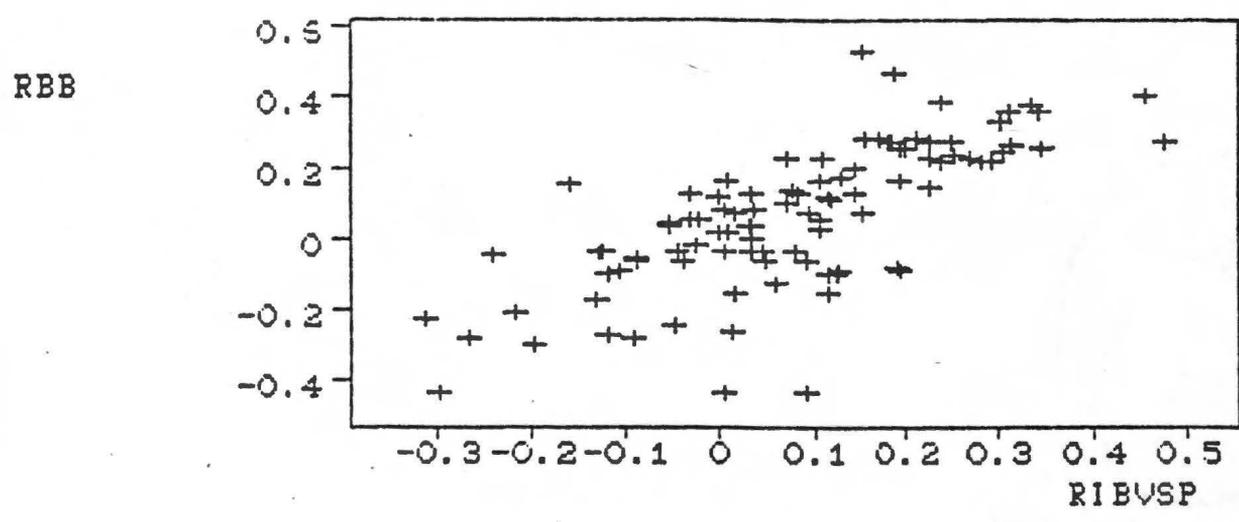


FIG.3: DADOS ORIGINAIS E PREVISÃO UM PASSO A FRENTE DE RBB

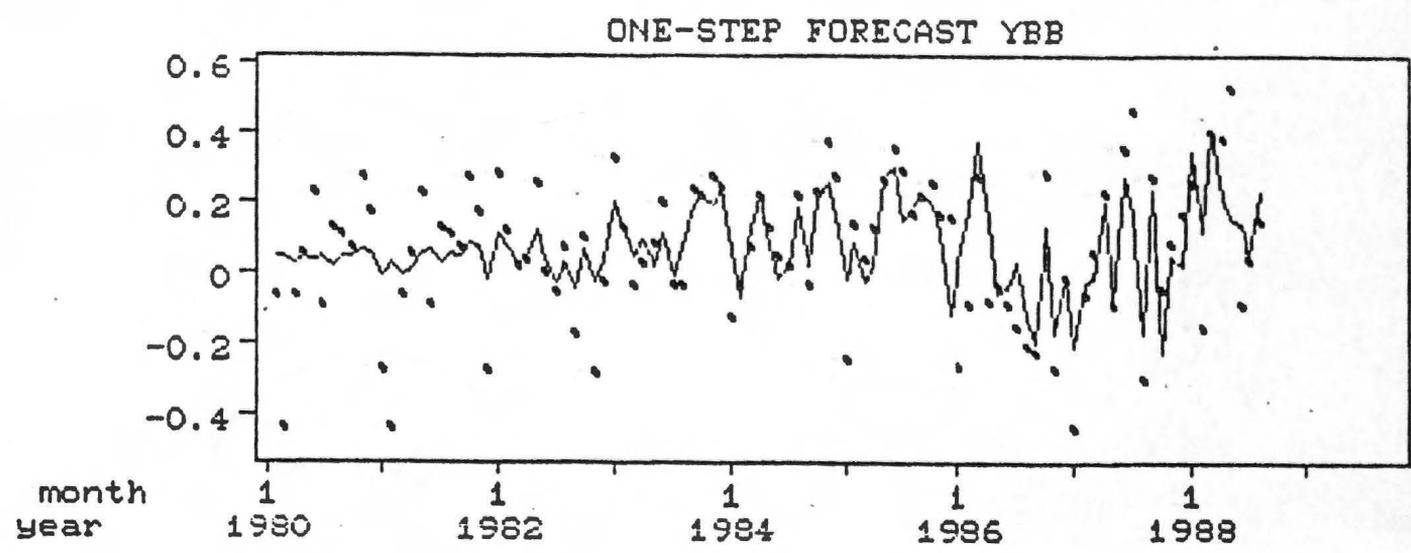


FIG.4: PREVISÃO E LIMITES DE 90% DE PROBABILIDADE DE PREVISÃO 1 PASSO A FRENTE

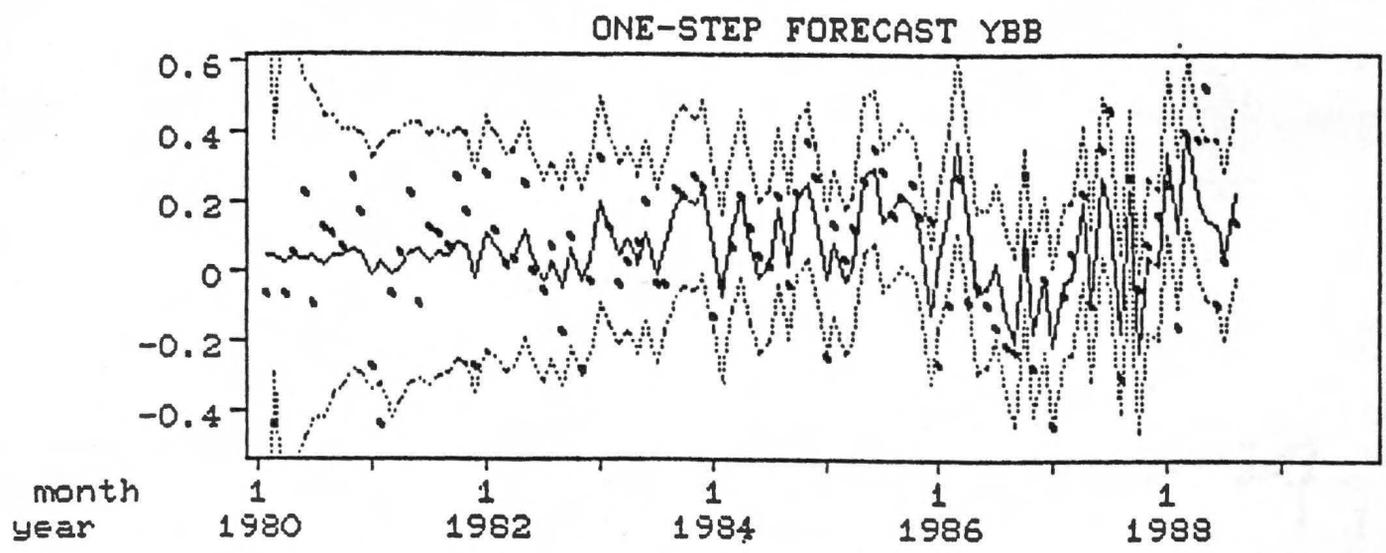


FIG.5: INTERCEPTO ESTIMADO ON-LINE COM LIMITES DE 90% DE PROBABILIDADE

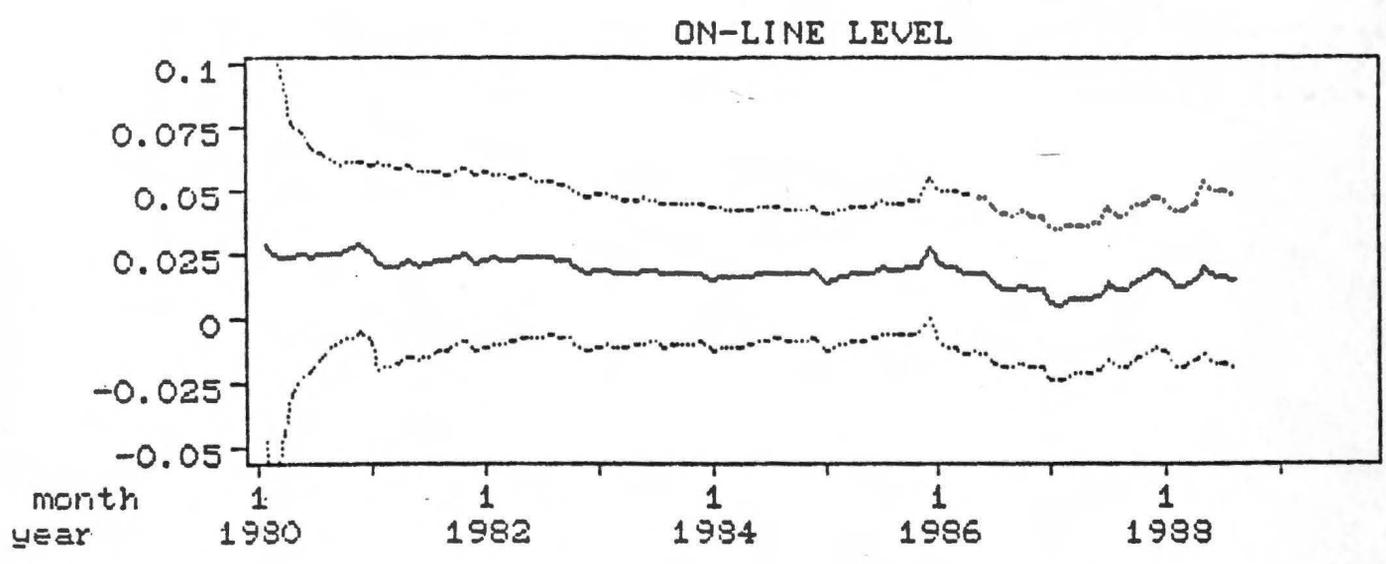


FIG.6: COEFICIENTE DE REGRESSÃO ESTIMADO ON-LINE COM LIMITES DE 90% DE PROBABILIDADE

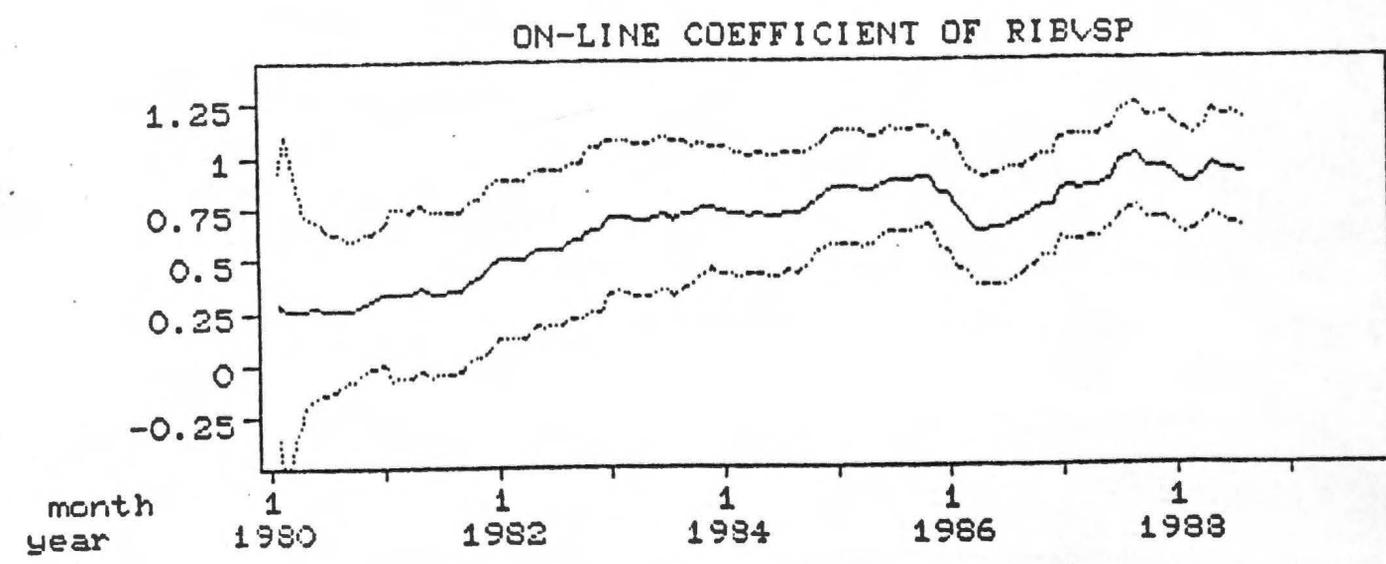


FIG.7: ERRO DE PREVISÃO UM PASSO A FRENTE

RAW ON-LINE RESIDUALS

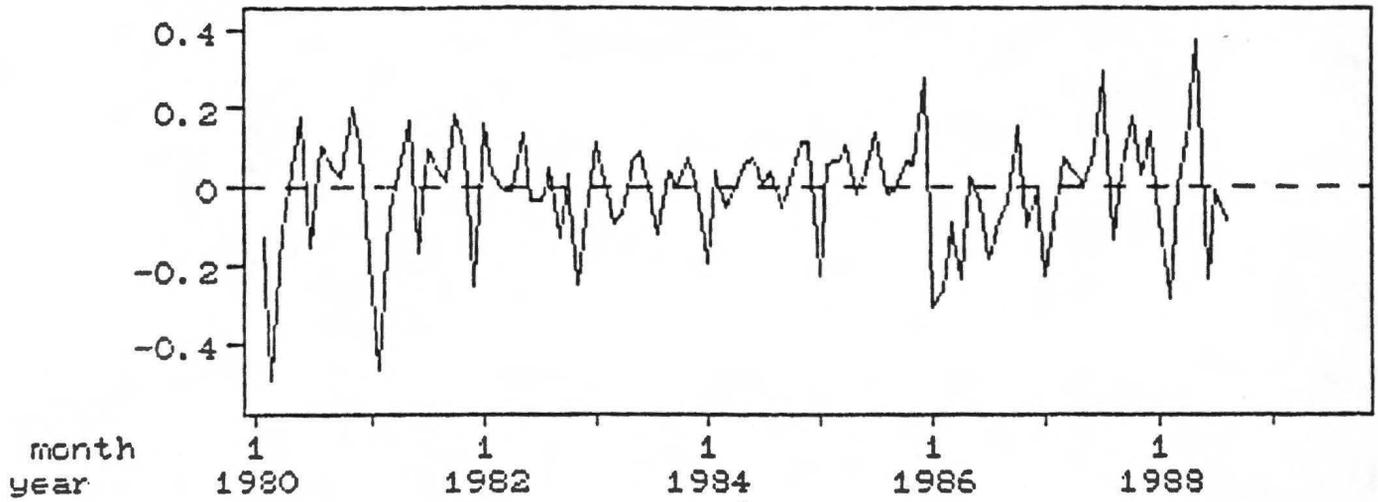


FIG.8: RESIDUO PADRONIZADO, COM INTERVALO DE 90% DE PROBABILIDADE

SCALED ON-LINE RESIDUALS

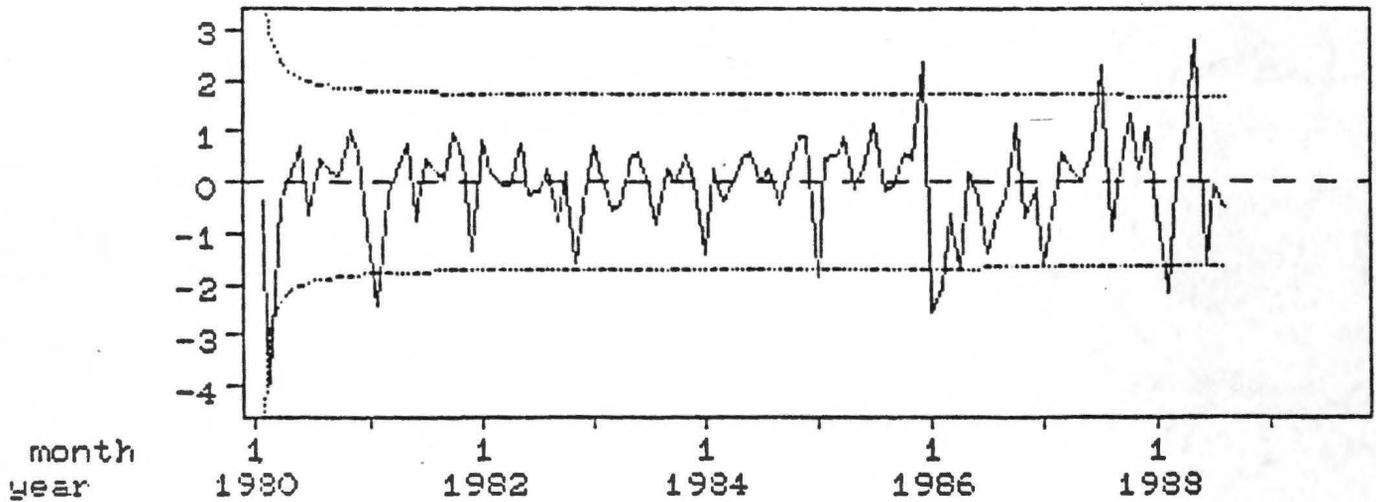


FIG.9: VARIÂNCIA ESTIMADA ON-LINE

ON-LINE SCALE PARAMETER

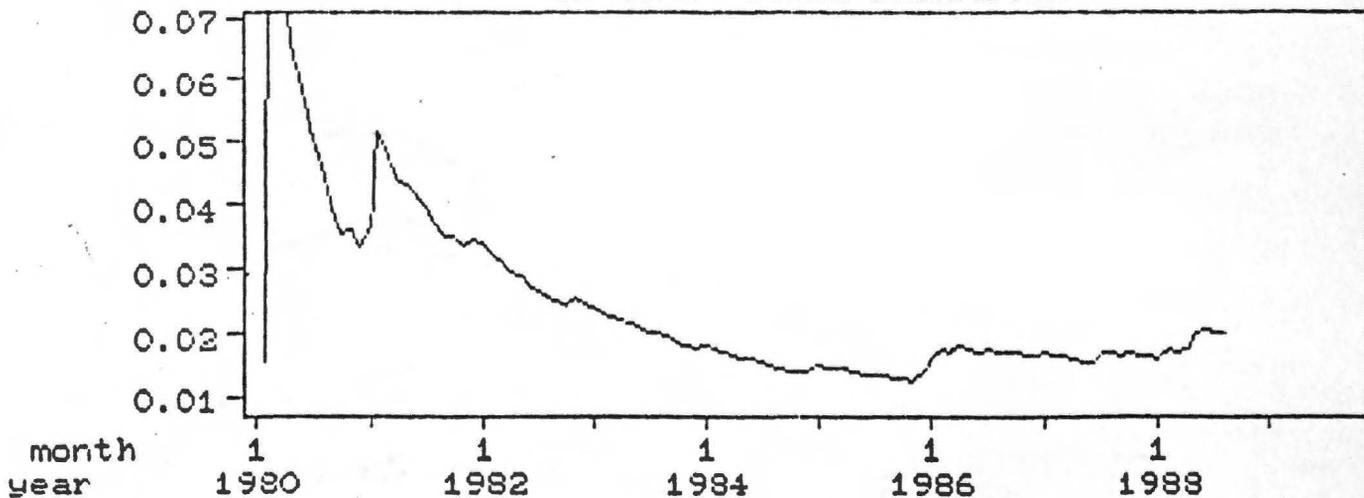


FIG.10: SÉRIE SUAVIZADA E DADOS ORIGINAIS (RBB)

POSTERIOR SUMMARY		
Component	Mean	S.F.
Level	0.015	0.020
RIBVSP	0.907	0.156

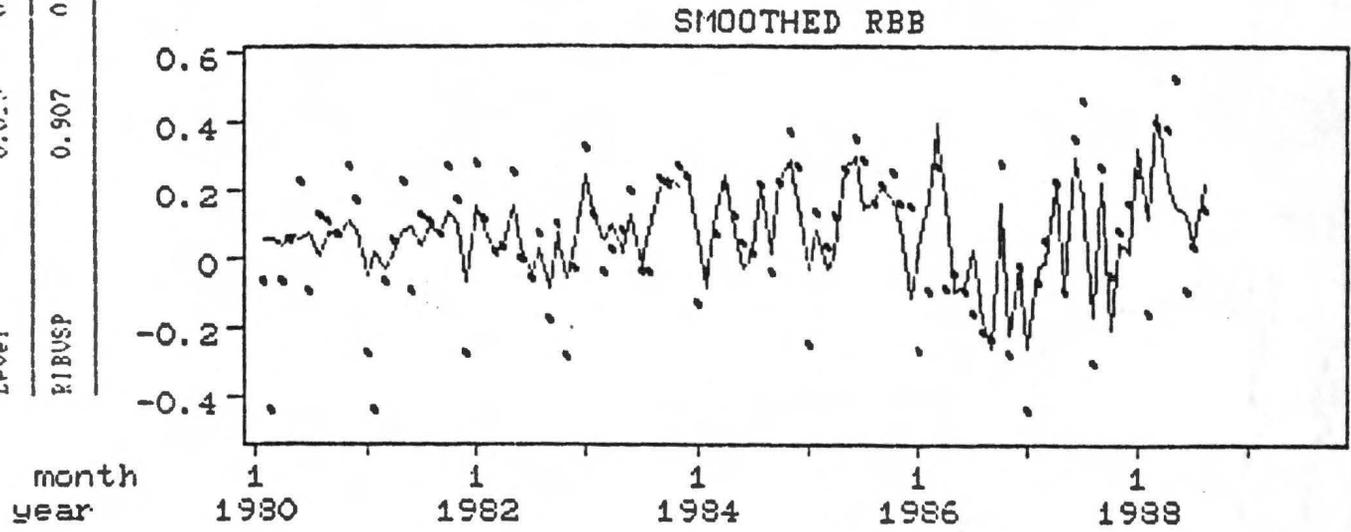


FIG.11: INTERCEPTO ESTIMADO RETROSPECTIVAMENTE .

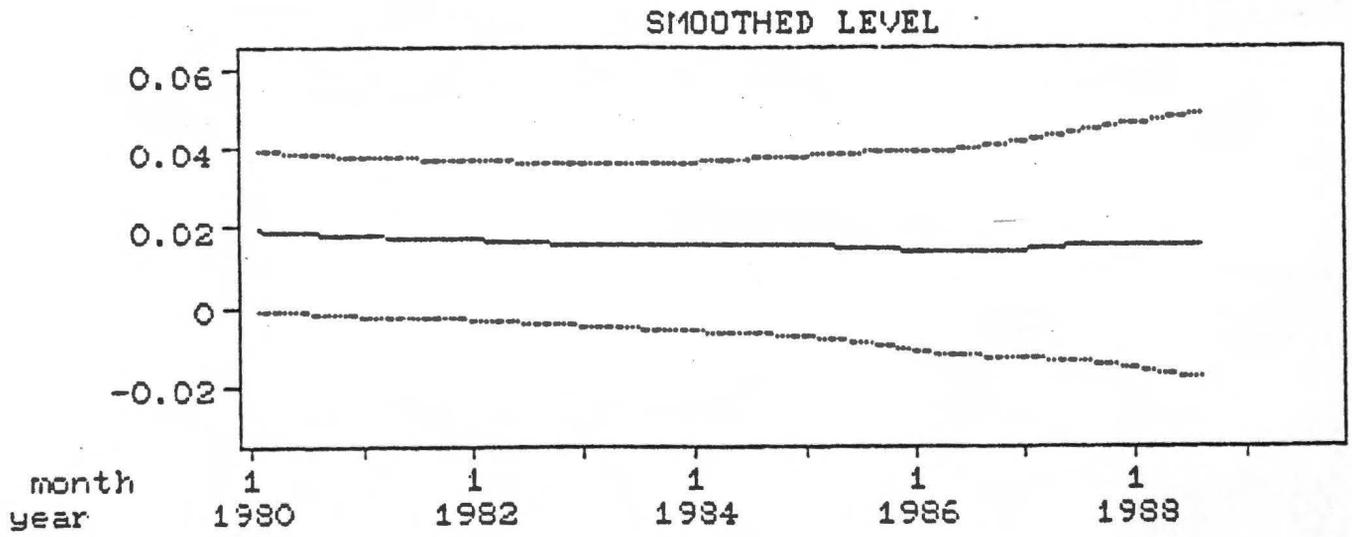


FIG.12: COEFICIENTE DO RIBVSP RETROSPECTIVO

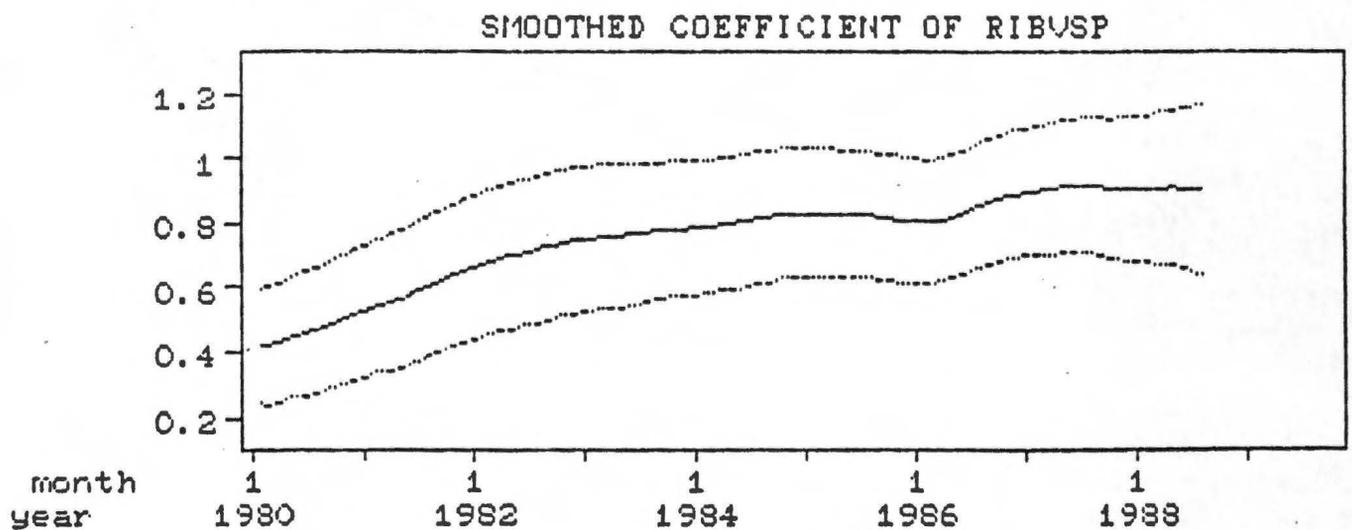


FIG.13: RESIDUO DO AJUSTE RETROSPECTIVO

RAW SMOOTHED RESIDUALS

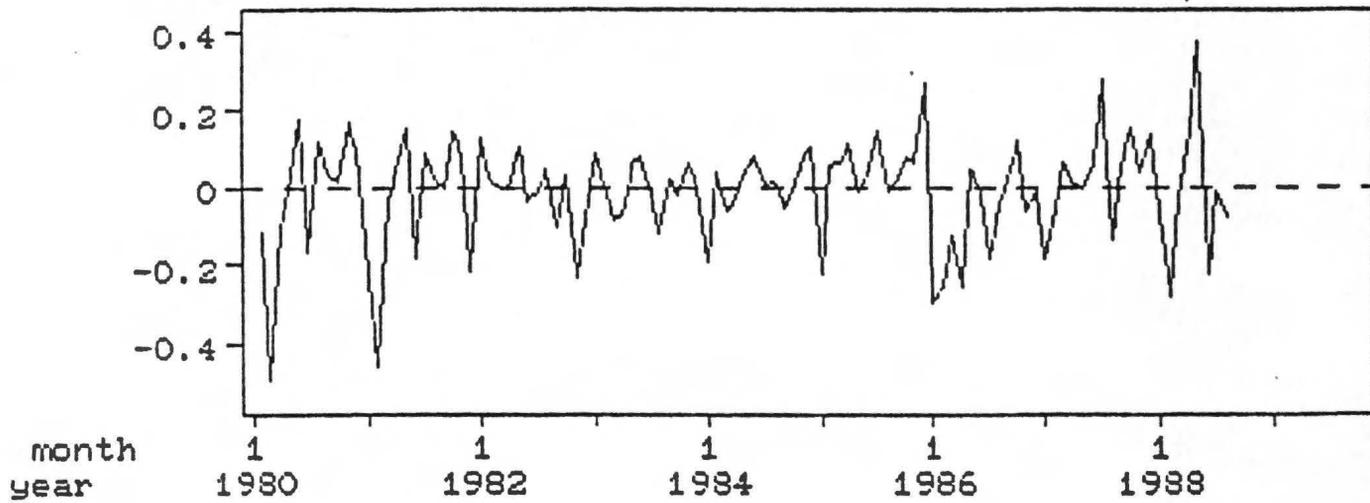


FIG.14: RESIDUO PADRONIZADO DO AJUSTE RETROSPECTIVO

SCALED SMOOTHED RESIDUALS

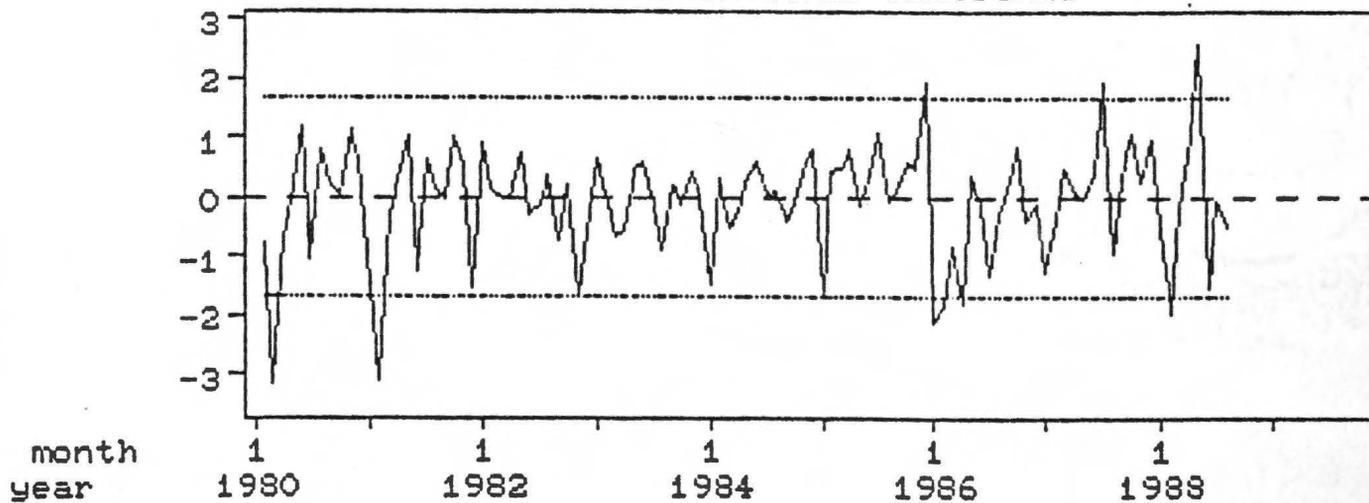


FIG.15: VARIANCIA OBSERVACIONAL SUAVIZADA RBB

SMOOTHED SCALE PARAMETER

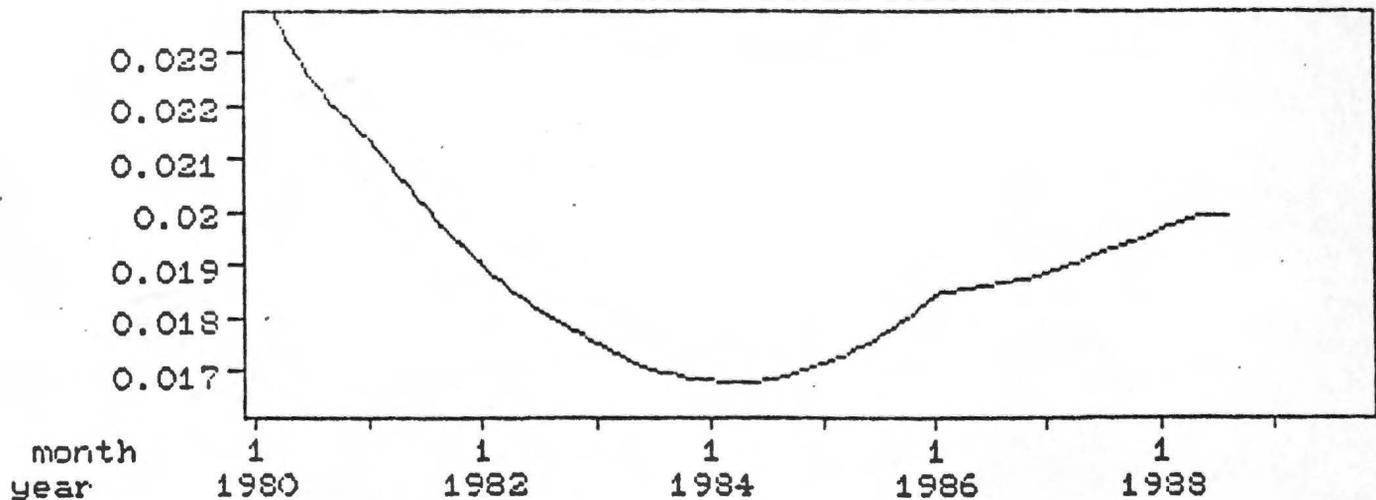


FIG.: TENDÊNCIA DO RIBVSP SUAVIZADA

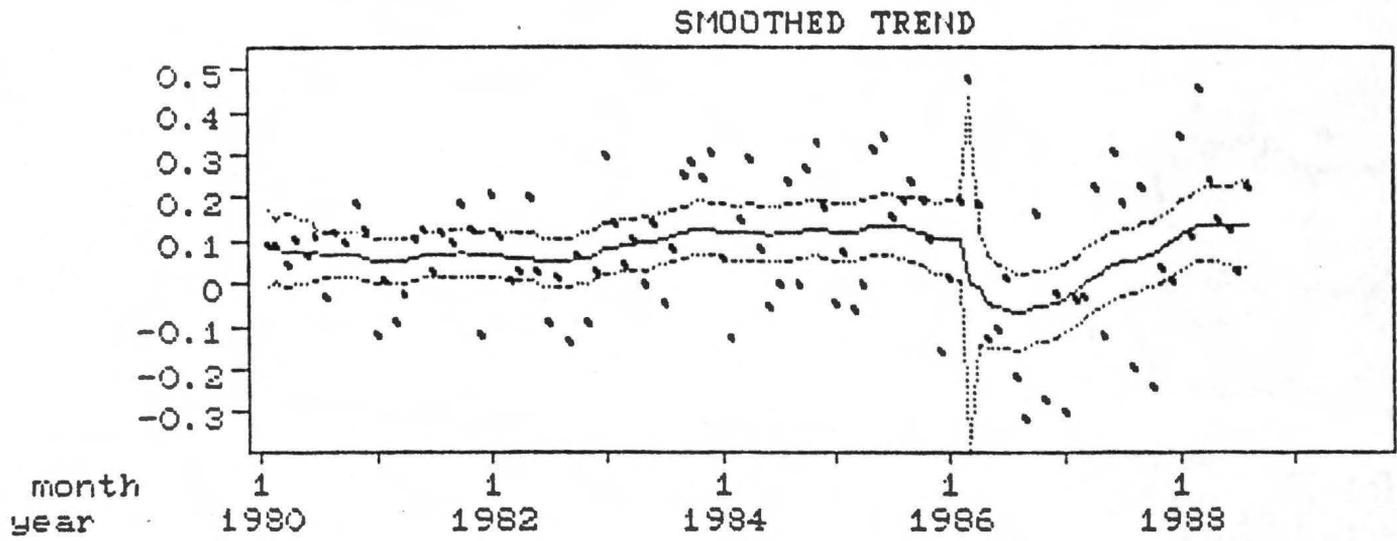
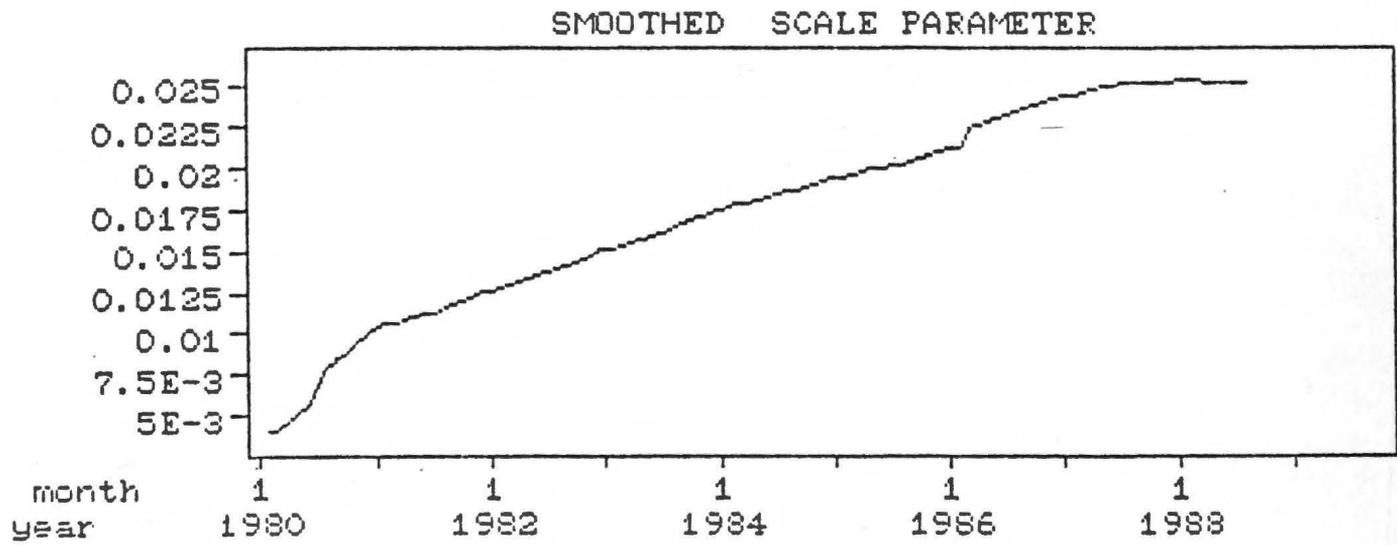


FIG.18: VARIANCIA OBSERVACIONAL SUAVIZADA RIBVSP



NÚMEROS JÁ PUBLICADOS:

- 01/88 - CRÍTICA DE RAZÕES NO CENSO ECONÔMICO
Renato Martins Assunção (ENCE/IBGE)
Rosana de Fretas Castro (DEIND/IBGE)
José Carlos R.C. Pinheiro (ENCE/IBGE)
- 02/88 - USO DE AMOSTRAGEM EM SIMULAÇÃO DE LEGISLAÇÃO TRIBUTÁRIA
José Carlos R.C. Pinheiro (ENCE/IBGE)
Manuel Martins Filho (DISUL/SERPRO)
- 03/88 - FORECASTING THE NUMBER OF AIDS CASES IN BRAZIL
Dani Ganeman (IME/UFRJ)
Helio S. Migon (ENCE/IBGE & IME/UFRJ)
- 04/88 - AVALIAÇÃO DOS EFEITOS DE REDUÇÃO DA FRAÇÃO DE AMOSTRAGEM
NO CENSO DEMOGRÁFICO
José Carlos R.C. Pinheiro (ENCE/IBGE)
José Matias de Lima (DPE/IBGE)
- 01/89 - MIGRAÇÕES ANUAIS RURAL-URBANO-RURAL - Período 70/80
Kaizô Iwakami Beltrão (ENCE/IBGE)
Helio S. Migon (ENCE/IBGE & IME/UFRJ)
- 02/89 - CURVA DE LORENZ & ÍNDICE DE GINI PARA DISTRIBUIÇÕES DE
RENDÁ
José Paulo Q. Carneiro (ENCE/IBGE)
Jorge Luiz Rangel Costa (CIDE)
- 03/89 - METODOLOGIA DE CRÍTICA DE EQUAÇÕES DE FECHAMENTO NOS CENSOS
ECONÔMICOS DE 1985
José Carlos R.C. Pinheiro (ENCE/IBGE)
Renato Martins Assunção (ENCE/IBGE)
- 04/89 - UMA PROPOSTA DE DIMENSIONAMENTO DA AMOSTRA DO CENSO DEMOGRÁ-
FICO DE 1990
José Carlos R.C. Pinheiro (ENCE/IBGE)
José Matias de Lima (ENCE/IBGE)
- 05/89 - UMA REVISÃO DO BATS - BAYESIAN ANALYSIS OF TIME SERIES
Helio S. Migon (ENCE/IBGE e IM/UFRJ)
E. Suyama (ICEX/UFMG)