Simpósio de Inovações Metodológicas e Tecnológicas

As Inovações nas Informações Sociais e Econômicas

Volume 9

Tomo 4
sessões 109 a 110

SIMPÓSIO DE INOVAÇÕES

As Inovações nas Informações Sociais e Econômicas

109 Dessazonalização dos índices de preços ao consumidor

110 Análise longitudinal de dados

ST 109

DESSAZONALIZAÇÃO DOS ÍNDICES DE PREÇOS AO CONSUMIDOR

Coordenador: WANDERLEY RAMALHO (IPEAD / MG)

O correto entendimento da questão inflacionária de uma economia impõe a necessidade de uma correta avaliação da tendência de médio e longo prazo. Do ponto de vista econômico, um dos aspectos relevantes da questão é o extrair-se a direção geral do movimento dos preços que constitui a chamada inflação estrutural ou verdadeira inflação. Com este objetivo tem-se procurado eliminar, dos índices de preços, as chamadas sazonais ou estacionais. Elas surgem da ação de forças socio-econômicas (ou climáticas) de natureza distinta daquelas que determinam a tendência geral da série de preços.

Um indicador relevante não deve incorporar as influências sazonais uma vez que elas dizem respeito apenas a períodos específicos Isto tem conduzido, assim, à busca do chamado 'índice sazonalmente ajustado".

A questão gira em torno da seleção da melhor técnica para obter deste índice que, descontaminado das pressões sazonais, reflita a tendência do processo inflacionário.

Um dos métodos estatísticos bastante utilizados é conhecido por X11, desenvolvido por SHISKIN, J., em 1967 e aprimorado por DAGUN em 1977 e 1978. É baseado essencialmente na utilização de filtros lineares tipo médias móveis. Dagun introduziu modificações que possibilitaram a utilização de filtros simétricos também nos extremos das séries. Para isto foram utilizados modelos ARIMA sazonais.

Se apresentarão aplicações do método X11, realizadas por instituições que trabalham com índices de preços no Brasil, e se examinará a sua contribuição para o entendimento da questão inflacionária no país.

SAZONALIDADE EM ÍNDICES DE PREÇOS : O CASO DO IPC-FIPE

Vera Lucia Fava*
Juarez A. B. Rizzieri*

Resumo

Este artigo aborda a questão do ajustamento sazonal de índices de preços, assunto frequentemente discutido mas sobre o qual não existe consenso. Estuda-se o caso concreto do IPC-Fipe, índice largamente usado como indexador. O período considerado estende-se de janeiro de 1980 a dezembro de 1994. Utilizando duas metodologias alternativas - método X-11 e Modelos Estruturais de Séries de Tempo - foram identificados onze itens do IPC-Fipe com comportamento sazonal. Quando se agrupam os efeitos sazonais desses itens, observa-se que eles se compensam, não transferindo, dessa forma, nenhum padrão de sazonalidade para o índice geral. Por esse resultado, o ajustamento sazonal do IPC-Fipe não é necessário.

1. INTRODUÇÃO

O ajustamento sazonal de séries econômicas não é procedimento aceito de forma irrestrita. A polêmica é ainda maior quando se trata da dessazonalização de séries de índices de preços.

Antes de defender ou criticar a realização do ajustamento sazonal de um índice de preços, convém ter bem claro o objetivo e as possíveis utilizações desse indicador.

Os índices de preços ao consumidor, por exemplo, geralmente visam medir a evolução dos preços de um conjunto de bens e serviços consumidos por uma *população-alvo* previamente escolhida. Encarados dessa forma, não há porque submetê-los a um processo de ajustamento sazonal, *escondendo* suas reais oscilações.

Quando, porém, esses mesmos índices de preços passam a ser utilizados como indexadores pelos mais diversos segmentos da economia, o ajustamento sazonal pode ser defensável como forma de evitar que as oscilações devidas ao comportamento sazonal de um grupo restrito de produtos possam vir a contaminar os preços dos demais bens e serviços.

* Professores da FEA/USP e pesquisadores da FIPE.

O Índice de Preços ao Consumidor calculado pela FIPE (IPC-Fipe) pode ser citado como exemplo dessa dupla função desempenhada por um índice de preços. Originalmente concebido para medir aproximadamente a variação do custo de vida das famílias paulistanas pertencentes à classe de renda modal, o IPC-Fipe foi, até recentemente, utilizado como indexador por diversos agentes econômicos: órgãos governamentais e instituições privadas utilizaram, oficialmente ou contratualmente, o IPC-Fipe como instrumento de atualização monetária. Talvez mais relevante seja seu uso como indicador referencial, tanto pelo Banco Central como pelo mercado financeiro, para a formação da taxa de juros futura que, em última instância, acaba por induzir a própria taxa futura de inflação. Evidentemente, nesse caso também são utilizados outros índices de preços como o IGP-M e o INPC.

Em função, portanto, da utilização do IPC-Fipe como indexador ou como elemento de formação de expectativas inflacionárias, seria ou não necessário proceder ao ajuste sazonal do mesmo?

Analisando a trajetória desse índice durante os últimos quinze anos, constata-se que ele não apresenta comportamento sazonal inequívoco, apesar de vários de seus componentes certamente seguirem padrão de sazonalidade bem definido.

Aparentemente, há duas explicações alternativas para tal característica do IPC-Fipe: a) os efeitos sazonais dos componentes simplesmente acabam se compensando, não provocando sazonalidade no índice geral: b) os efeitos sazonais podem influenciar a formação de expectativas inflacionárias pelos agentes econômicos, à medida que, em períodos subsequentes, se transferem para os preços dos demais itens, impedindo que o índice geral reflita o comportamento verdadeiramente sazonal de alguns de seus componentes.

Se valer a primeira explicação, evidentemente não há necessidade de proceder ao ajustamento sazonal do índice geral. Já no caso de ser correta a segunda explicação, o ajustamento é justificável.

O objetivo deste artigo é testar a presença de sazonalidade em alguns componentes do IPC-Fipe e avaliar seu efeito conjunto sobre o índice geral[1]. Inicialmente buscou-se identificar os componentes do IPC-Fipe que poderiam apresentar padrão de sazonalidade justificável. Foram selecionados treze componentes cujo peso no índice era de 36,61% até 1993 e de 35,30% a partir de janeiro de 1994, em decorrência da implantação dos resultados da última pesquisa de orçamentos familiares.

---

[1]Não é objeto desta discussão o procedimento alternativo de dessazonalização parcial das séries de preços através de mudanças nos pesos orçamentários mensais de produtos que apresentam comportamento sazonal, à semelhança do praticado pelo Desip do IBGE para os hortifrutigrangeiros.

O passo seguinte consistiu na estimação dos efeitos sazonais e na verificação de sua significância estatística. Dois procedimentos foram utilizados: o método X-11 e o Modelo Estrutural para séries de tempo. O período analisado abrange os anos de 1980 a 1994. Além deste, foram também considerados os subperíodos 1980-1985 e 1986-1994. O objetivo de tal divisão foi avaliar a sensibilidade dos resultados aos congelamentos de preços decretados a partir de 1986, além de eventuais alterações nos padrões de sazonalidade.

Por fim, avaliou-se mês a mês o impacto conjunto dos efeitos sazonais associados aos diversos componentes do IPC-Fipe. A seguir, são apresentadas detalhadamente as etapas perseguidas e os resultados obtidos.

2. AVALIAÇÃO DO PADRÃO DE SAZONALIDADE

Os componentes do IPC-Fipe que em princípio teriam comportamento sazonal justificável são os seguintes: arroz, feijão, leite, carnes, aves, frutas, legumes, verduras, tubérculos, ovos, aluguel, vestuário e educação. Os dez primeiros são produtos do setor agropecuário e o padrão sazonal de seus preços pode ser explicado pela escassez nos períodos de entressafra e por relativa abundância nos períodos de safra. É claro que uma política eficaz de estoques reguladores e de importação poderia suavizar os efeitos sazonais e eventualmente eliminá-los. A inclusão do aluguel justifica-se pela *concentração* de contratos ocorrida em consequência das políticas de reajuste de cada plano de estabilização a partir do Cruzado. O vestuário é reconhecidamente um componente com acentuado comportamento sazonal, marcado por substanciais elevações de preços nos meses de lançamento das coleções outono-inverno e primavera-verão e por quedas reais ou nominais de preços nos meses de liquidação. Quanto ao item educação, a sazonalidade tinha seu pico nos primeiros meses do ano quando eram feitas as matrículas para o ano letivo a se iniciar; nos anos recentes o pagamento da matrícula passou a ser feito no final do ano anterior, inclusive a pretexto de reserva de vaga, o que pode ter contribuído para mudanças no padrão de sazonalidade.

2.1 Métodos Utilizados

O estudo da sazonalidade associada aos treze itens do IPC-Fipe mencionados anteriormente foi feito com base na abordagem clássica do Modelos Estruturais de Séries de Tempo e no método X-11.

Os Modelos Estruturais têm por objetivo a estimação dos componentes não-observáveis subjacentes a uma série temporal, quais sejam, tendência, ciclo, sazonalidade e componente irregular, permitindo que todos eles sejam estocásticos. Uma breve exposição dos Modelos Estruturais para séries de tempo pode ser encontrada no apêndice[2]. Esta metodologia é bastante útil pois, além de possibilitar a estimação do componente sazonal, permite testar se o padrão de sazonalidade é constante ou variável no tempo, tomando por base a significância estatística da variância do componente sazonal. Permite também a inclusão de variáveis explicativas e de intervenção.

O modelo estimado para todos os itens considerados foi o denominado *Modelo Estrutural Básico* que contém tendência, sazonalidade e o termo de erro. Para os períodos 1986-1994 e 1980-1994, foram estimados modelos com e sem variável *dummy*, definida de forma a captar o efeito dos planos de estabilização decretados a partir de 1986. A opção por um dos dois modelos foi feita com base na significância do coeficiente estimado da variável *dummy* e nos resultados dos testes de normalidade, autocorrelação e heterocedasticidade.

As estimações foram feitas no domínio do tempo, utilizando o pacote econométrico denominado STAMP-Structural Time Series Analyser, Modeller and Predictor.

O método X-11 foi desenvolvido no *US Bureau of the Census* na década de 1960 e é um dos procedimentos de ajustamento sazonal mais utilizados no mundo. Ao contrário dos Modelos Estruturais, o X-11 não tem por base um modelo econométrico. Ele extrai da série original os componentes de tendência-ciclo, sazonalidade e irregular utilizando médias móveis[3]. O método pode ser aplicado a séries mensais e trimestrais e permite sazonalidade aditiva e multiplicativa. Esta última foi adotada no presente estudo por melhor se adequar às séries trabalhadas. Utilizou-se nesta etapa o pacote SAS.

2.2 Resultados Obtidos

Dos treze itens analisados, aves e feijão não apresentaram sazonalidade em nenhum dos períodos considerados, conforme indicam os resultados do Quadro 1. Para arroz e aluguel, só não foi detectada sazonalidade no período 1980-1985. Já o item carnes não apresentou sazonalidade no período 1986-1994. No caso do leite, as duas metodologias produziram resultados opostos: enquanto o método X-11 rejeitou a existência de sazonalidade apenas no período 1980-1985, os resultados dos Modelos Estruturais só

[2]Para maiores detalhes sobre essa metodologia, consultar Harvey (1989).
[3]Ver a respeito, Butter e Fase (1991) e Hylleberg (1992).

indicaram padrão sazonal neste período. Quanto aos demais itens, a sazonalidade não foi rejeitada em nenhum período por nenhum dos procedimentos empregados.

As conclusões acerca da constância, no tempo, do padrão de sazonalidade, extraídas do Modelo Estrutural, foram as seguintes: no período 1980-1985, cinco itens - leite, frutas, legumes, vestuário e educação - apresentaram sazonalidade variável ou estocástica; para 1986-1994, identificou-se sazonalidade constante em todos os casos e para 1980-1994, apenas os itens frutas e educação caracterizaram-se por sazonalidade variável.,

Os Gráficos 1(a) a 1(ℓ) permitem comparar o padrão de sazonalidade entre períodos e entre métodos de estimação. Naqueles referentes aos Modelos Estruturais, os fatores de sazonalidade significantemente diferentes de 1, ao nível de 10%, estão assinalados com seta[4]. O mesmo não foi feito para os demais gráficos porque o método X-11 não permite a realização de teste análogo. Notam-se várias diferenças, sobretudo entre períodos, o que pode ser devido a efetivas mudanças nos padrões de sazonalidade ou à interferência das sucessivas acelerações do processo inflacionário, seguidas pelos congelamentos de preços decretados pelo governo. Por exemplo, o vestuário, item sazonal de maior peso no IPC-Fipe, apresenta efeitos sazonais positivos e significantes nos mesmos meses, independentemente do intervalo de tempo: abril e maio (coleção outono-inverno) e outubro (coleção primavera-verão). Já os efeitos sazonais negativos e significantes ocorrem em janeiro, fevereiro julho e agosto no período 1980-1985 e em fevereiro, julho, agosto e dezembro no período 1986-1994, sugerindo antecipação da liquidação dos artigos de verão, ou seja, alteração no padrão de sazonalidade.

A análise minuciosa dos gráficos revela que não existe nenhum mês em que todos os itens apresentam efeitos sazonais na mesma direção. Há, portanto, em todos os meses, efeitos sazonais positivos de um subconjunto de itens compensando, ao menos em parte, os efeitos sazonais do outro.

Para se ter uma medida exata dessa compensação e do impacto final da sazonalidade sobre o índice geral, é necessário levar em conta o peso de cada item no IPC-Fipe.

A maneira correta de calcular essa medida está associada à forma como é computado o índice. O IPC-Fipe é calculado através da fórmula geométrica:

$$IPC_t = \prod_i \left(R^i_{t,t-1} \right)^{w^i}$$

onde:

[4]Nos casos de padrão sazonal variável, foi considerada a significância nos últimos doze meses do período.

$R^i_{t,t-1}$ é o relativo de preços do item i entre os instantes t e $t-1$;

w^i é o peso do item i.

O IPC-Fipe ajustado sazonalmente (IPCS), obtido a partir da dessazonalização de seus componentes, seria dado por:

$$IPCS_t = \prod_i \left(R^i_{t,t-1} / FS^i_t \right)^{w^i}$$

onde FS^i_t é o fator de sazonalidade assoaciado ao item i no mês t.

Assim, se $\prod_i \left(FS^i_t \right)^{w^i}$ for próximo da unidade, pode-se concluir que o efeito sazonal agregado é pouco relevante, não implicando em comportamento sazonal do índice de preços global.

Foi então calculado o produtório dos efeitos sazonais ponderados para cada mês do ano. Os resultados encontram-se na Tabela 1.

Quase todos os valores obtidos ficaram muito próximos de um, indicando que os comportamentos sazonais apresentados individualmente pelos itens considerados, quando agregados, não acarretam comportamento sazonal do IPC-Fipe. Todavia, uma observação deve ser feita, especialmente em relação aos resultados do método X-11: até 1985, os desvios da unidade eram insignificantes porém, a partir de 1986, os números de janeiro parecem não ser tão despreziveis. É evidente que se pode associar o comportamento desses fatores sazonais de janeiro aos planos de estabilização visto que, excetuando os plano Bresser e Real, todos os demais congelamentos foram adotados no primeiro trimestre de cada ano, ficando os mais altos índices de preços concentrados nos meses de janeiro, o que acabou se incorporando aos próprios fatores de sazonalidade. Os Modelos Estruturais parecem mais eficientes para isolar esse efeito, uma vez que seus resultados são mais bem comportados em janeiro.

Por consequência dessa diferença metodológica, nota-se que o teste de sazonalidade não rejeita a existência desta na série do IPC-Fipe pelo método X-11 para os períodos 1986-94 e 1980-94 ao passo que pelos Modelos Estruturais a existência de sazonalidade é rejeitada para todos os períodos. Assim, pelo exposto sobre janeiro, deve-se olhar com reserva o resultado positivo de sazonalidade para os períodos que contêm os planos de estabilização.

3. CONCLUSÃO

Os resultados mostraram a presença de padrão sazonal para onze itens do IPC-Fipe, os quais respondem por algo em torno de 32% dos gastos familiares no Município de São Paulo, ficando sem padrão sazonal significativo feijão e aves. O primeiro tem sua produção razoavelmente distrubuída ao longo do ano, o que tem garantido um abastecimento normal, prejudicado apenas por acidentalidades climáticas. O frango já tem uma produção industrializada, razão pela qual o comportamento de seus preços distancia-se cada vez mais dos padrões da carne bovina. Sua oferta também tem sido regularizada em função dos compromissos de exportações a preços previamente contratados. É importante registrar que, apesar da significância estatística dos padrões sazonais, eles em geral não se constituem na maior parcela da oscilação dos preços, ficando esta por conta do componente irregular.

Apesar da existência de sazonalidade para itens do IPC-Fipe, o mesmo não fica garantido para o índice geral, principalmente quando o período pós-Cruzado é incluído na análise. Isso revela que existe um efeito compensatório entre os padrões sazonais quando eles são tomados conjuntamente, isto é, os efeitos se anulam quando considerados mês a mês. Aparentemente, o simples fato da ausência de padrão sazonal para o índice geral leva a crer que a variação sazonal a nível de produto tem sido incapaz de contaminar a formação dos preços dos demais componentes do índice, ou seja, os agentes econômicos têm levado em conta o efeito *conjunto* da sazonalidade e, como este tem sido anulado pelo efeito compensação mencionado, nada tem sido transferido para os demais produtos. Em outras palavras, os agentes econômicos têm sido racionais o suficiente para separar as variações sazonais localizadas em alguns produtos, da variação devida a outros choques inflacionários. Estes representam, em última instância, as ocorrências mais graves porque afetam as expectativas inflacionárias que, a curto prazo, incorporam-se às taxas de juros do mercado e se difundem pela maioria dos contratos da economia, comprometendo a eficácia das políticas econômicas que buscam o equilíbrio macroeconômico num ambiente de preços estáveis.

Apêndice: Modelos Estruturais para Séries de Tempo

Os Modelos Estruturais visam a estimação dos componentes não observáveis subjacentes a uma série temporal. Esses componentes são a tendência, o ciclo, a sazonalidade e o termo irregular, os quais podem, de acordo com essa abordagem, ser estocásticos.

A estimação desses componentes não é prerrogativa dos Modelos Estruturais. Ela já era objeto dos métodos *tradicionais* de análise de séries de tempo, frequentemente utilizados antes da disseminação dos modelos ARIMA.

O Modelo Estrutural Univariado, em sua forma aditiva, é expresso por[5]:

$$y_t = \mu_t + \psi_t + \gamma_t + \varepsilon_t \qquad t = 1, 2, \ldots, T$$

onde y_t é a série observada, μ_t é a tendência, ψ_t é o ciclo, γ_t é a sazonalidade e ε_t é o componente errático.

O Modelo Estrutural acomoda várias possibilidades de evolução dos componentes ao longo do tempo. Apresentar-se-á apenas uma alternativa. Outras formulações podem ser encontradas em Harvey (1989).

TENDÊNCIA

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t$$
$$\beta_t = \beta_{t-1} + \zeta_t$$

CICLO

$$\psi_t = \rho\ [\alpha\ \cos\lambda\, t + \beta\ \operatorname{sen}\lambda\, t] + \kappa_t$$

SAZONALIDADE

$$\gamma_t = -\sum_{j=1}^{s-1} \gamma_{t-j} + \omega_t$$

COMPONENTE ERRÁTICO

$$\varepsilon_t = \text{ruído branco}$$

[5]A notação aqui adotada é a utilizada por Harvey (1989).

onde η_t, ζ_t, κ_t e ω_t são ruídos brancos, β_t é a declividade da tendência, ρ é o fator de amortecimento ($0<\rho \leq 1$), λ é a frequência do ciclo, $\sqrt{\alpha^2+\beta^2}$ é a amplitude do ciclo e s é o número de períodos sazonais.

O modelo expresso pelas equações anteriores tem apenas uma variável observável (y_t) e requer a estimação dos quatro componentes não observáveis em cada instante de tempo t, com base nas informações disponíveis em t-1. Para obter estimativas atualizadas dos componentes não observáveis, deve-se colocar o modelo em forma de Espaço de Estado e utilizar o Filtro de Kalman.

O Modelo Estrutural Básico, que não contém o componente cíclico, tem a seguinte representação em forma de Espaço de Estado, considerando s = 4:

1. EQUAÇÃO DE MEDIDA

$$y_t = [1\ 0\ 1\ 0\ 0]\alpha_t + \varepsilon_t \qquad \text{ou} \qquad y_t = z\alpha_t + \varepsilon_t$$

$$E(\varepsilon_t) = 0; \quad V(\varepsilon_t) = h_t$$

2. EQUAÇÃO DE TRANSIÇÃO

$$\alpha_t = \begin{bmatrix} \mu_t \\ \beta_t \\ \gamma_t \\ \gamma_{t-1} \\ \gamma_{t-2} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ \beta_{t-1} \\ \gamma_{t-1} \\ \gamma_{t-2} \\ \gamma_{t-3} \end{bmatrix} + \begin{bmatrix} \eta_t \\ \zeta_t \\ \omega_t \\ 0 \\ 0 \end{bmatrix}$$

ou

$$\alpha_t = T\alpha_{t-1} + \delta_t$$

$$E(\delta_t) = 0; \quad V(\delta_t) = Q_t \,.$$

3. HIPÓTESES ADICIONAIS

$$E(\alpha_0) = a_0; \qquad V(\alpha_0) = P_0;$$
$$E(\varepsilon_0 \alpha_0') = E(\delta_0 \alpha_0') = E(\varepsilon_0 \delta_0') = 0.$$

As estimativas atualizadas do vetor de estado α_t, ou seja, as estimativas atualizadas dos componentes não observáveis do modelo, podem ser obtidas a partir do Filtro de Kalman.

A efetiva implementação do filtro pressupõe que T, Q_t, z e h_t sejam conhecidos. Na presente formulação, apenas T e z o são; Q_t e h_t precisam ser estimados.

De acordo com a abordagem clássica, estimativas desses parâmetros são obtidas através da maximização da seguinte função de verossimilhança:

$$\ln(y_t\,; Q_t, h_t) = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\sum_t^T \ln(f_t) - \frac{1}{2}\sum_t^T (y_t - z a_{t/t-1})^2 / f_t$$
$$f_t = z P_{t/t-1} z' + h_t \,.$$

Um conjunto de testes e estatísticas permite avaliar a significância dos parâmetros estimados bem como a qualidade do ajustamento do modelo (ver Harvey(1989), cap.5). Em particular, pode-se testar H_0: Var(ω) = 0; caso essa hipótese não seja rejeitada, pode-se inferir que o padrão de sazonalidade é constante ao longo do tempo.

QUADRO 1

Resultados dos testes de sazonalidade no IPC-Fipe (#)

item	modelo estrutural			método X-11		
	1980-85	1986-94	1980-94	1980-85	1986-94	1980-94
arroz	N	S	S	N	S	S
feijão	N	N	N	N	N	N
leite	S*	N	N	N	S	S
carnes	S	N+	S+	S	N	S
aves	N	N	N	N	N	N
frutas	S*	S	S*	S	S	S
verduras	S	S	S	S	S	S
legumes	S*	S	S	S	S	S
tubérculos	S	S	S+	S	S	S
ovos	S	S+	S	S	S	S
aluguel	N	S+	S+	N	S	S
vestuário	S*	S	S	S	S	S
educação	S*	S	S*	S	S	S
índice geral	N	N	N	N	S	S

(#) S (N) indica presença (ausência) de sazonalidade;
* indica padrão variável de sazonalidade;
\+ indica modelo estimado com variável dummy.

TABELA 1
Efeito Agregado dos Itens Sazonais do IPC-Fipe

a) 1980/85

ano	jan	fev	mar	abr	mai	jun	jul	ago	set	out	nov	dez
modelo estrutural												
1980	0.9955	0.9951	1.0017	1.0047	0.9978	0.9974	1.0044	1.0007	1.0036	1.0029	0.9999	0.9957
1981	0.9942	0.9991	0.9997	1.0050	0.9977	0.9974	1.0011	1.0048	1.0041	1.0019	0.9994	0.9964
1982	0.9934	0.9980	1.0006	1.0059	0.9953	0.9997	1.0017	1.0032	1.0048	1.0024	0.9981	0.9975
1983	0.9928	0.9968	1.0035	1.0036	0.9949	1.0007	1.0042	0.9994	1.0056	1.0040	0.9968	0.9970
1984	0.9941	0.9949	1.0054	1.0044	0.9952	0.9977	1.0019	1.0050	1.0033	1.0045	0.9966	0.9966
1985	0.9947	0.9959	1.0047	1.0035	0.9944	0.9956	1.0069	1.0053	1.0019	1.0037	0.9975	0.9951
método X-11												
1980/85	0.9938	0.9997	1.0028	1.0016	0.9945	0.9979	1.0024	1.0044	1.0035	1.0018	0.9972	0.9967

b) 1986/94

ano	jan	fev	mar	abr	mai	jun	jul	ago	set	out	nov	dez
modelo estrutural												
1986/93	1.0098	1.0042	1.0040	1.0090	1.0034	1.0029	0.9868	0.9882	0.9910	1.0044	1.0000	0.9967
1994	1.0144	0.9964	1.0071	1.0081	1.0006	1.0031	0.9877	0.9938	0.9936	1.0057	0.9972	0.9926
método X-11												
1986/93	1.0115	1.0024	0.9956	1.0016	1.0032	0.9975	0.9879	0.9919	0.9934	1.0060	1.0028	1.0010
1994	1.0117	1.0017	0.9978	1.0013	1.0040	0.9965	0.9884	0.9919	0.9930	1.0056	1.0027	1.0006

c) 1980/94

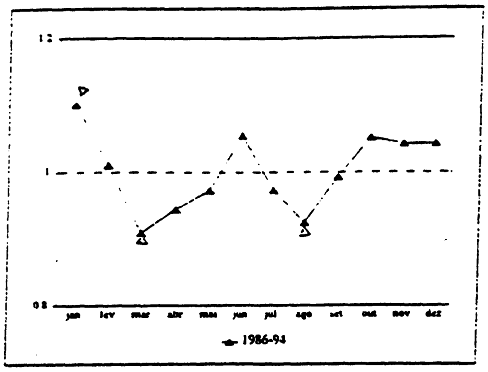
ano	jan	fev	mar	abr	mai	jun	jul	ago	set	out	nov	dez
modelo estrutural												
1980	1.0036	1.0007	1.0043	1.0053	0.9995	1.0023	0.9929	0.9943	0.9968	1.0033	1.0000	0.9971
1981	1.0037	1.0006	1.0043	1.0054	0.9995	1.0023	0.9928	0.9945	0.9967	1.0033	0.9999	0.9971
1982	1.0038	1.0007	1.0041	1.0055	0.9995	1.0022	0.9929	0.9945	0.9966	1.0033	0.9999	0.9970
1983	1.0041	1.0006	1.0039	1.0057	0.9994	1.0022	0.9929	0.9946	0.9965	1.0035	0.9997	0.9969
1984	1.0045	1.0003	1.0038	1.0058	0.9994	1.0021	0.9929	0.9948	0.9963	1.0035	0.9996	0.9970
1985	1.0047	1.0003	1.0035	1.0062	0.9993	1.0020	0.9930	0.9948	0.9962	1.0035	0.9996	0.9969
1986	1.0050	1.0001	1.0034	1.0064	0.9992	1.0019	0.9932	0.9948	0.9961	1.0035	0.9996	0.9968
1987	1.0055	0.9997	1.0035	1.0064	0.9992	1.0018	0.9934	0.9947	0.9961	1.0035	0.9995	0.9966
1988	1.0061	0.9992	1.0036	1.0063	0.9992	1.0018	0.9935	0.9946	0.9962	1.0035	0.9993	0.9966
1989	1.0066	0.9988	1.0038	1.0062	0.9993	1.0019	0.9931	0.9947	0.9964	1.0033	0.9993	0.9965
1990	1.0070	0.9985	1.0040	1.0061	0.9991	1.0022	0.9930	0.9947	0.9964	1.0034	0.9992	0.9964
1991	1.0069	0.9987	1.0037	1.0062	0.9992	1.0022	0.9929	0.9947	0.9965	1.0034	0.9992	0.9963
1992	1.0071	0.9984	1.0040	1.0058	0.9994	1.0025	0.9925	0.9948	0.9966	1.0034	0.9992	0.9964
1993	1.0070	0.9986	1.0038	1.0059	0.9994	1.0025	0.9925	0.9947	0.9968	1.0032	0.9992	0.9965
1994	1.0064	0.9986	1.0049	1.0054	1.0002	1.0021	0.9930	0.9959	0.9962	1.0030	0.9984	0.9961
método X-11												
1980/93	1.0059	0.9988	0.9976	1.0034	0.9982	0.9967	0.9957	0.9981	0.9961	1.0051	1.0010	0.9991
1994	1.0055	0.9992	0.9988	1.0032	0.9990	0.9961	0.9958	0.9994	0.9954	1.0044	1.0003	0.9986

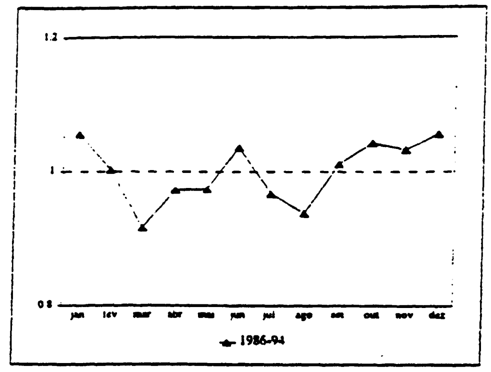
GRÁFICO 1

FATORES DE SAZONALIDADE

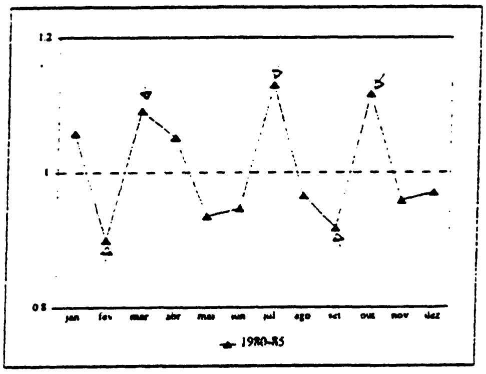
Modelo estrutural (*) — Método X-11

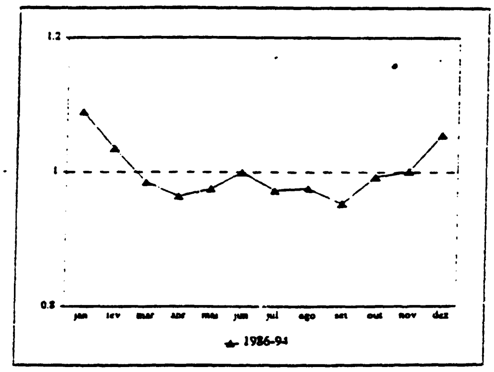
a) arroz



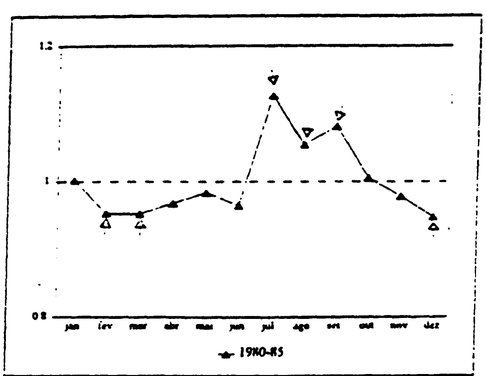


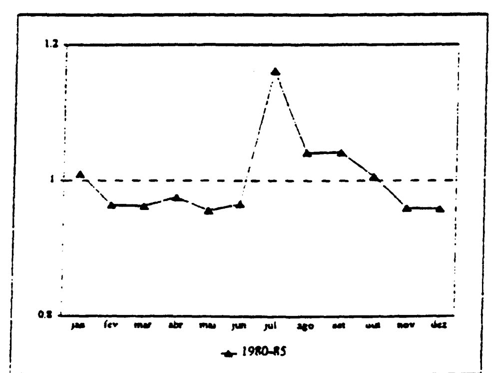
b) leite



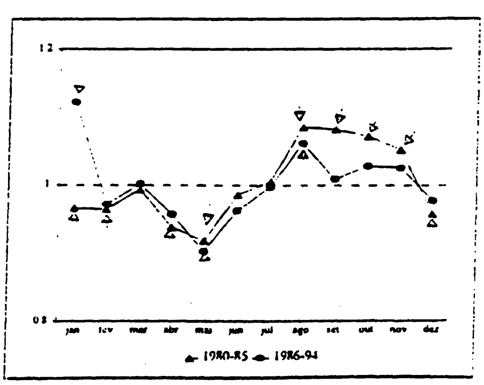


c) carne





d) frutas



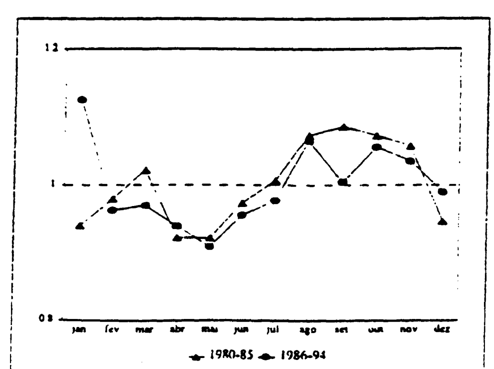
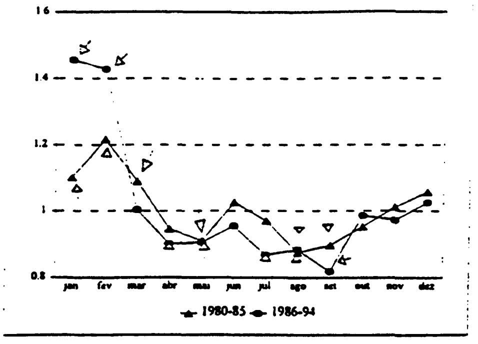


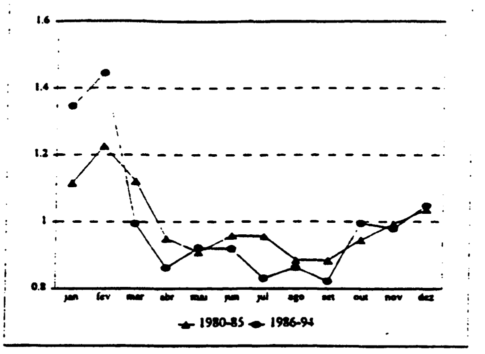
GRÁFICO 1 (cont.)

FATORES DE SAZONALIDADE

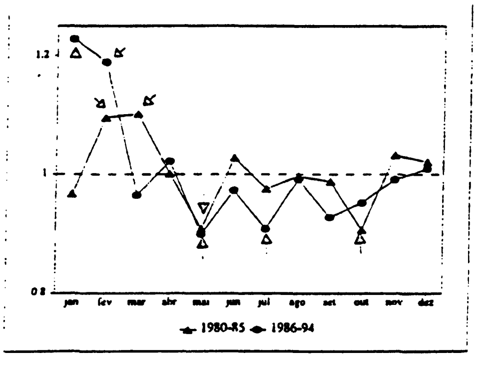
Modelo estrutural (*) — Método X-11

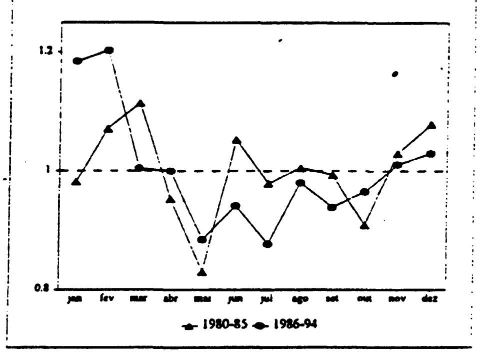
e) verduras



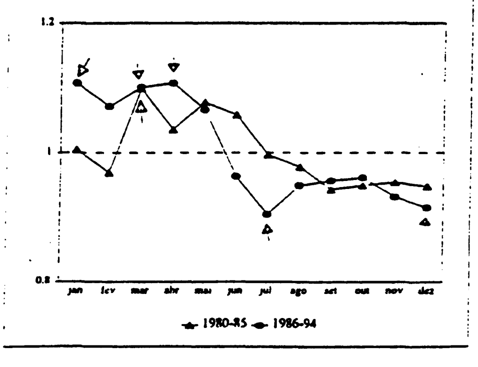


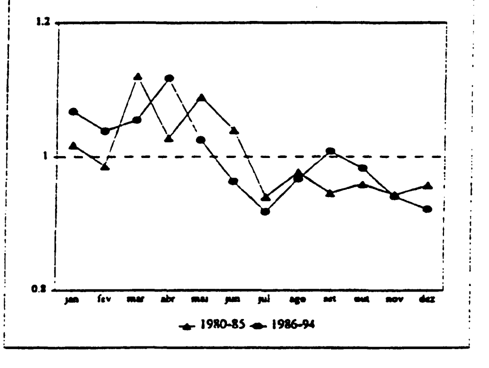
f) legumes



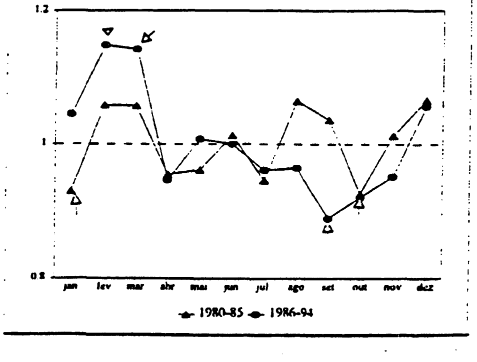


g) tubérculos





h) ovos



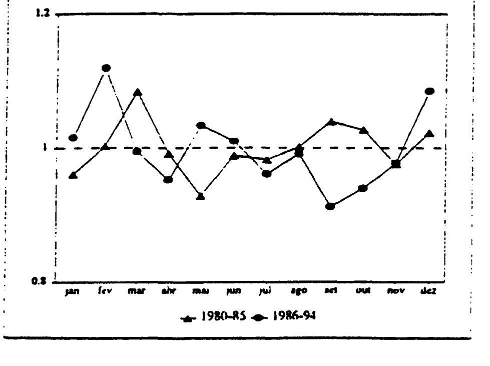
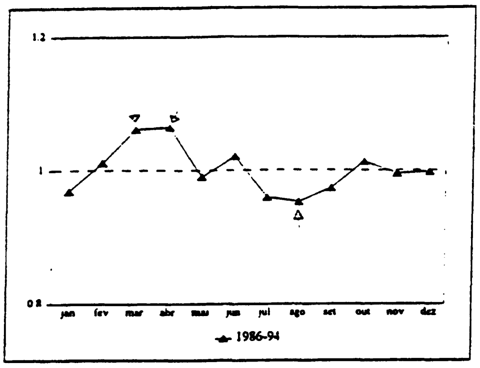


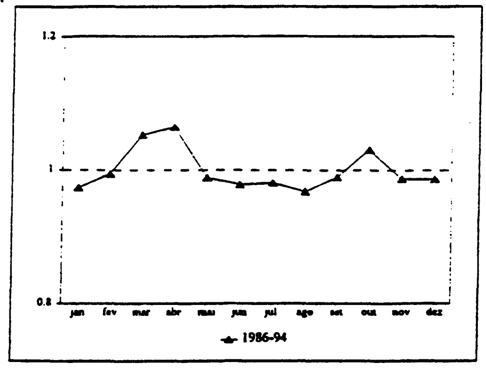
GRÁFICO 1 (cont.)

FATORES DE SAZONALIDADE

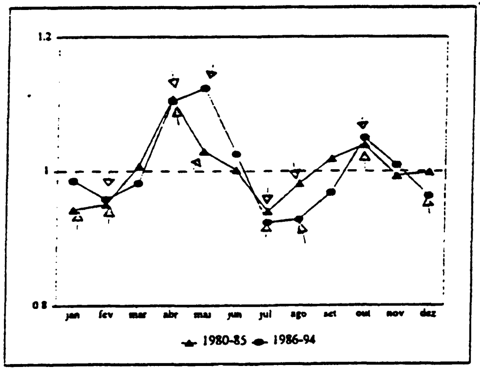
Modelo estrutural (*) | Método X-11

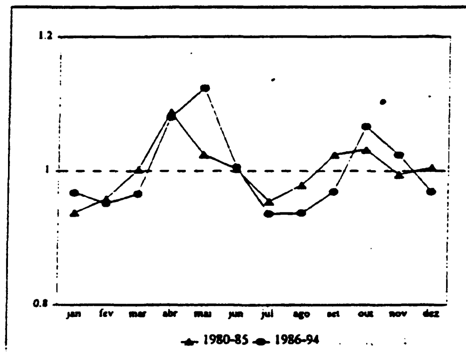
i) aluguel



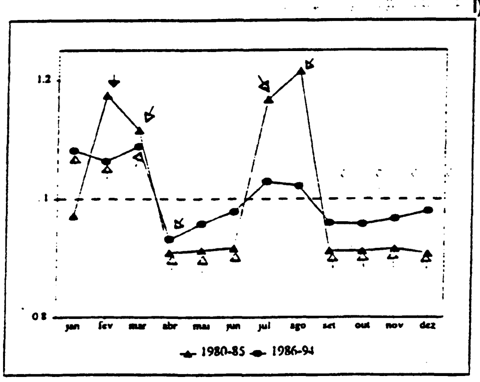


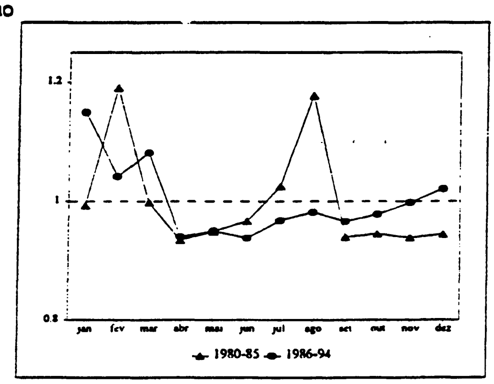
j)vestuário





l) educação





(*) As setas indicam fatores de sazonalidade significantemente diferentes de um, ao nível de10%.

BIBLIOGRAFIA

Bell, W.R. & S. Hillmer, 1984, Issues involved with seasonal adjustment of economic time series, *Journal of Business and Economic Statistics*, vol. 2, n. 4, 291- 320.

Butter, F.A.G. den & M.M.G. Fase, 1991, *Seasonal Adjustment as a Practical Problem*, Elsevier Science Publishers B.V.

Dagum, E.B. & M. Morry, 1984, Basic issues on the seasonal adjustment of the Canadian consumer price index, *Journal of Business and Economic Statistics*, vol. 2, n. 3, 250-259.

Harvey. A.C., 1989, *Forecasting , Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

Hylleberg, S., 1992, The X-11 method, in *Modelling Seasonality*, ed. S. Hylleberg, Oxford University Press.

Jain, R.K., 1989, The seasonal adjustment procedures for the consumer price indexes: some empirical results, *Journal of Business and Economic Statistics*, vol. 7, n. 4, 461-469.

Maravall, A., 1980, Effects of alternative seasonal adjustment procedures on monetary policy, *Journal of Econometrics*, 14, 115-136.

Plosser, C.I., 1979, Short-term forecasting and seasonal adjustment, *Journal of the American Statistical Association*, vol 74, n. 365, 15-24.

AJUSTAMENTO SAZONAL NOS ÍNDICES DE PREÇOS AO CONSUMIDOR AMPLO

IBGE

Angela Maria Puppim Buzanovsky

Lucia Maria Coelho Pinto

Marcelo Martins Cruz

SUMÁRIO

Introdução

Índice de Preços ao Consumidor-IPC é uma medida síntese do movimento de preços de um conjunto de mercadorias, chamado "Cesta de Mercadorias" , representativo do consumo de um determinado grupo populacional, em um certo período de tempo.

O IPC é calculado como sendo uma média ponderada dos relativos de preços, mês t contra mês t-1, tomando como pesos as participações, em valor, dos componentes desta "Cesta de Mercadoria" nos orçamentos das famílias que compõem a população objetivo.

Um dos objetivos principais dos índices de preços ao consumidor é fornecer informações básicas e fundamentais para a compreensão do fenômeno inflacionário do país. Em conjunto com outras séries de indicadores auxiliam a interpretação sobre a natureza e as causas das flutuações da atividade econômica.

A maioria das pesquisas sobre as flutuações macroeconômicas tem trabalhado com dados sazonalmente ajustados. A idéia básica, ressaltada nestas pesquisas, é de que a existência de movimentos sazonais afetam o reconhecimento e a interpretação de importantes movimentos não sazonais numa série, tais como:

a) mudanças de tendência e outros eventos cíclicos;

b) ocorrências inesperadas para as quais possíveis causas são procuradas.

Existem técnicas estatísticas que identificam, isolam e removem das séries históricas as oscilações sazonais e, quando aplicadas aos Índices de Preços, obtém-se o que se poderia chamar de **"Taxa de Inflação Verdadeira"**. Esta seria uma medida relevante para avaliação e formulação de políticas econômicas, daí, a prática de algumas instituições estrangeiras de publicarem, além das séries originais dos índices, índices com ajustes sazonais.

Cabe enfatizar, que todas as incursões técnicas sobre a possibilidade de produção de séries ajustadas sazonalmente têm como pressuposto a produção de uma estatística adicional que melhor reflita a **Tendência Inflacionária** (se a componente irregular for irrelevante).

Neste contexto, a questão sazonal vem sendo estudada no Departamento de Índices de Preços ao Consumidor, do IBGE - Instituto Brasileiro de Geografia e Estatística, visando fornecer aos usuários de Índices de Preços ao Consumidor mais um instrumental analítico de importância significativa.

O objetivo deste documento é apresentar a metodologia do ajustamento sazonal realizado, os resultados das séries de índices de preços ao consumidor sazonalmente ajustadas e, simultaneamente, propiciar o debate a respeito dos procedimentos técnicos utilizados.

Em seu primeiro capítulo será abordada a abrangência do estudo (regional, nível de agregação, períodos estudados).

No capítulo 2 citaremos o método adotado e descreveremos vários indicadores de qualidade do ajustamento sazonal, bem como outros recursos disponíveis no programa.

O capítulo 3 contém uma análise exploratória da possível interferência dos planos econômicos e seus reflexos nos índices de preços ao consumidor, tendo como objetivo investigar se a existência de sucessivos planos inviabilizariam a produção de séries sazonalmente ajustadas para a economia brasileira.

O capítulo 4 destina-se a apresentação dos resultados e o 5 as conclusões finais.

As tabelas e gráficos que subsidiaram as tomadas de decisão, em todas as etapas do estudo, encontram-se em anexo.

Capítulo 1 - Abrangência do Estudo

1.1- O Índice Escolhido

Foram estudadas as séries de números índices do IPCA - Índice de Preços ao Consumidor Amplo. Este índice é o estimador da inflação sob a ótica da demanda familiar agregada, pois abrange em sua população objetivo as famílias com rendimento mensal de 1 a 40 salários mínimos.

1.2- Abrangência Geográfica

O estudo foi desenvolvido para as onze áreas urbanas que integram o SNIPC - Sistema Nacional de Índices de Preços, a saber:

- Regiões Metropolitanas - RM: Belém (BE), Fortaleza (FT), Recife (RE), Salvador (SL), Belo Horizonte (BH), Rio de Janeiro (RJ), São Paulo (SP), Curitiba (CT) e Porto Alegre (PA).
- Brasília (BR) e Goiânia (GO).

1.3- Periodicidade do Estudo

Numa primeira fase, analisou-se as séries de números índices do IPCA de janeiro de 1980 a maio de 1995 (Anexo 3). Posteriormente estendemos o estudo até dezembro de 1995 (Anexo 6).

Com o objetivo de avaliarmos as conseqüências de alterações metodológicas relevantes ocorridas na produção das séries de índices, em especial a partir de 1986, e da influência dos sucessivos planos de estabilização econômica analisou-se as séries em mais dois períodos: 1980 a 1985 e 1986 a 1995.

1.4- Nível de agregação

Foi aplicado o programa ao menor nível de agregação; subitem. Obteve-se os fatores sazonais para este nível, calculou-se em seguida as séries sazonalmente ajustadas, e só a partir destas é que foram obtidos os índices para os sucessivos níveis de agregação, mediante a metodologia de cálculo[1] usualmente adotada.

Em resumo, para cada área estudada foram analisadas em média 215 séries de subitens para cada período, totalizando 645 séries nos três períodos e perfazendo aproximadamente 7095 séries nas onze áreas urbanas.

[1] SISTEMA NACIONAL DE ÍNDICES DE PREÇOS AO CONSUMIDOR: MÉTODOS DE CÁLCULO. Rio de Janeiro: IBGE, 1994, 102p. (Série relatórios metodológicos, v.14).

Capítulo 2 - Análise da Metodologia Proposta para o Ajustamento Sazonal das Séries dos Índices de Preços ao Consumidor Amplo - IPCA

2.1 - Procedimento para o Ajuste Sazonal do Índice de Preços ao Consumidor utilizando o X11-ARIMA [2]

As séries do IPCA analisadas são as séries de Números Índices (Mar86=100). São séries de acumulados mensais que permitem a comparação de todos os meses com a base fixa, e por conseguinte, torna possível, uma vez removidas todas as variações regulares, as comparações $mês_t$ / $mês_{t-1}$, para que estas guardem o correto sentido econômico. Nas análises dos Índices Mensais, é importante estimar o impacto de variações sazonais.

Variações sazonais representam movimentos periódicos (intra-anuais) provocados por efeitos calendários (climáticos ou institucionais) que se repetem, mais ou menos regularmente a cada ano. As variações sazonais podem ser distinguidas das variações cíclicas por suas oscilações características, bem definidas em períodos intra-anuais, bem como das flutuações irregulares por sua característica de previsibilidade.

Quanto às estruturas das séries, todas são multiplicativas, e suas composições: $Z_t=TC_t*S_t*I_t$, onde Z_t são as séries observadas, TC_t a tendência-ciclo, S_t a variação estacional e I_t a componente irregular.

Considerando a ordem de grandeza dos valores destas séries, optamos pelo procedimento padrão de linearização de modelos multiplicativos, adotando portanto o logaritmo neperiano destes valores passando para composição aditiva, isto é, $Z^*_t = TC^*_t + S^*_t + I^*_t$ com $Z^*_t = \ln Z_t$ e assim sucessivamente. Do ponto de vista do ajustamento sazonal, a decomposição adotada fornece melhores fatores estacionais estáveis.

Com relação ao pressuposto básico - a existência de sazonalidade - as técnicas mais comuns utilizadas na tentativa de verificar a presença de sazonalidade são:

i) a inspeção visual do gráfico das séries originais;
ii) correlogramas;
iii) análise espectral para a verificação da existência de sazonalidade e outras periodicidades ;

A inspeção visual do Gráfico 1 não permite verificar a presença da componente sazonal, entretanto, com as técnicas da análise espectral, uma série pode ser decomposta em seus componentes periódicos e a contribuição da variância de cada um dos componentes na variância total da série pode ser determinada.

---

[2] Stela B. Dagum do "Time Series Research And Analysis Division, Statistics Canada" - Versão Microcomputador

A análise espectral tem como aplicação natural identificar movimentos sazonais e avaliar diferentes técnicas de ajustamento . As freqüências correspondentes aos 12, 6, 4, 3, 2.4 e 2 meses são definidas como freqüências sazonais. A componente sazonal é representada pelos picos em torno destes harmônicos estacionais e pode ser levemente observada no Gráfico 2, no harmônico equivalente a 4 meses e facilmente observada no Gráfico 3, em todos os harmônicos, (série em que a tendência foi previamente removida pelo programa - tabela D8).

Gráfico 1

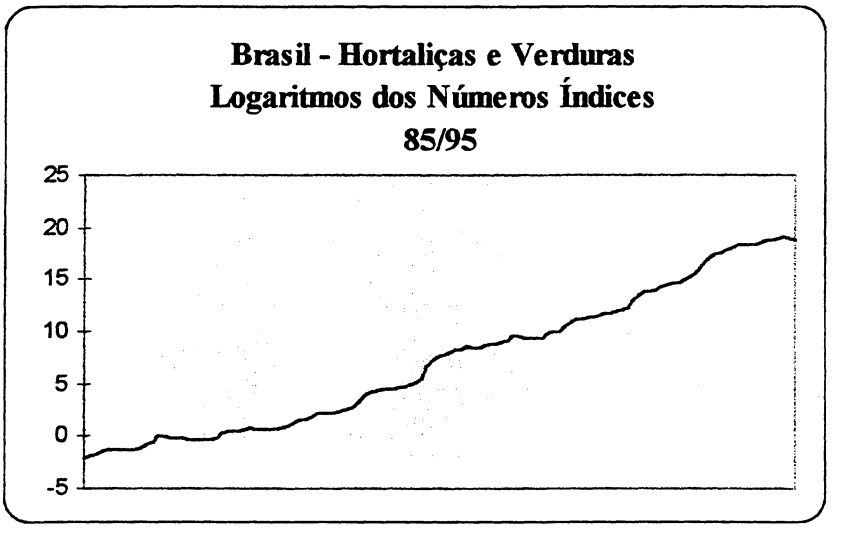


Gráfico 2

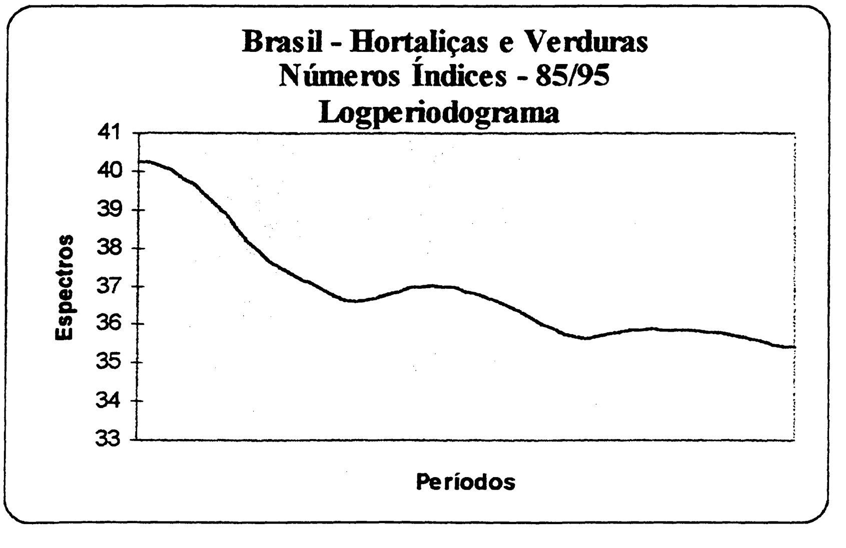
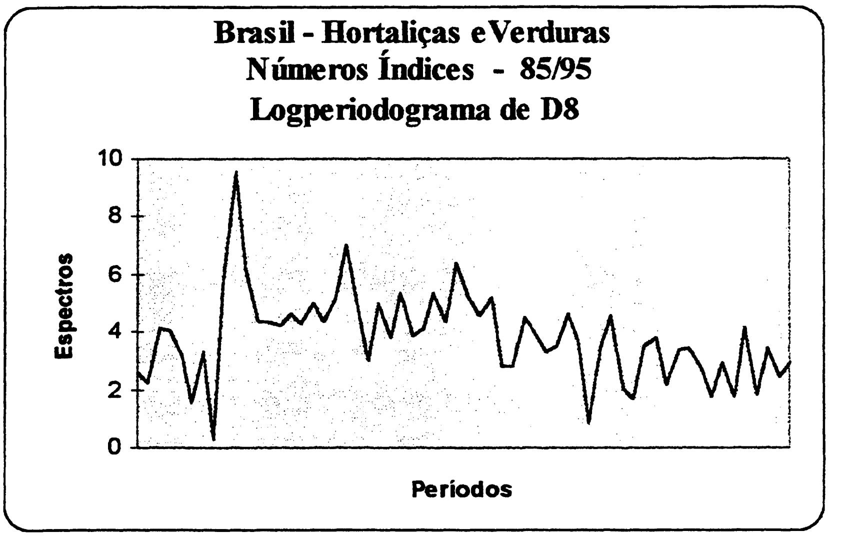


Gráfico 3



Além das técnicas anteriormente descritas, para a identificação da sazonalidade e estimação dos fatores estacionais de todas as séries utilizamos o procedimento desestacionalizador X11-ARIMA, que a seguir descreveremos em seus aspectos mais relevantes.

2.2 - Avaliação da Qualidade do Ajustamento Sazonal

O objetivo da utilização de procedimentos de ajuste sazonal em séries temporais é a decomposição das séries nas componentes tendência-ciclo, sazonal e irregular.

A decomposição de uma série nas citadas componentes é viável se a quantidade da variação sazonal presente for suficientemente grande comparada com a quantidade de irregularidade. Caso contrário, o programa não estima as duas componentes de modo confiável.

O programa X-11 ARIMA oferece vários indicadores da importância relativa dessas duas componentes. Um deles é o teste F para a presença da sazonalidade estável (Fe). É um teste de análise de variância de um fator que mede se as médias mensais são significativamente diferentes umas das outras, dada a flutuação aleatória total dos dados.

Este teste é aplicado às razões irregular e sazonal ($\bar{I}/\bar{S}$). Se o valor de F neste teste for maior que sete, as séries contém uma sazonalidade estável significativa.

Um teste complementar ao citado acima é o teste para a presença de sazonalidade móvel (Fm). Quanto menor forem as mudanças da componente sazonal de ano para ano, com mais segurança ela pode ser estimada pelo programa. Este teste de análise de variância de dois fatores indica se os movimentos de ano para ano são significantes. Do ponto de vista do ajustamento sazonal, valores pequenos de Fm neste teste indicam melhores chances para ajustes confiáveis.

Um outro teste é uma combinação dos dois testes anteriores. O principal propósito deste teste é determinar se a sazonalidade das séries é identificável ou não. Por exemplo: se existe uma pequena sazonalidade estável (entre meses) e a maior parte do processo ao longo dos anos é dominada rapidamente por mudanças sazonais, possivelmente aquela sazonalidade não será corretamente estimada, isto é, não será devidamente identificada pelo método.

Com relação ao monitoramento e controle de qualidade do programa a idéia geral é a seguinte: se a quantidade de sazonalidade estável for suficientemente grande comparativamente à quantidade de sazonalidade móvel, o valor do teste para sazonalidade identificável é menor que um, indicando qualidade aceitável.

Esse número pode ser encontrado na tabela F3 da saída do programa é denotado por medida M7. Se M7 for maior que 1 então não existe sazonalidade identificável na série analisada e consequentemente não é recomendável o ajustamento sazonal. M7 é a medida mais importante, mas existem dez outras taxas estatísticas de qualidade apresentadas na tabela F3. Essas medidas comparam o tamanho relativo das componentes sazonal e irregular, irregular e tendência-ciclo com as que encontramos nas séries sazonais de qualidade média. As medidas variam entre 0 e 3. Valores maiores que 1 sugerem problemas com a qualidade das séries.

No final da tabela F3, é impresso o valor do indicador total de qualidade (Q) que representa uma média ponderada das onze medidas (M1 a M11). Séries com valor de Q maior que um são consideradas de qualidade inaceitável (rejeitadas).

Um outro indicador útil para a avaliação dos dados sazonalmente ajustados é o MCD, isto é, número de meses necessários para o domínio do ciclo sobre a irregularidade. A estatística MCD (impressa na tabela F2.E) nos dá uma idéia da quantidade de flutuações irregulares deixadas na série ajustada para estacionalidade. O MCD é especialmente útil na determinação de quantos meses deverão ser abrangidos na análise para esboçar as mudanças nas séries dessazonalizadas, as quais podem ser atribuídas mais propriamente a tendência-ciclo do que a alguma flutuação aleatória.

Valores do MCD iguais a 1 indicam que certamente até as comparações mês a mês são significantes. Valores baixos do MCD estão associados a séries muito suavizadas ou com fortes movimentos tendenciais. Valores iguais ou maiores que seis indicam que as séries são altamente irregulares e de qualidade inaceitável ou com pouquíssimos movimentos tendenciais.

O programa possibilita ainda a escolha da média móvel sazonal (MMS) mais apropriada com base nas razões irregularidade-sazonalidade ($\bar{I}/\bar{S}$). Estas razões medem as flutuações médias, de ano para ano, da componente irregular em relação a mesma estatística correspondente da componente sazonal.

As médias móveis sazonais (MMS) podem ser de cinco termos (denominada 3X3), sete (3X5) ou onze (3X9), dependendo da quantidade de flutuações irregulares presentes nos dados. Essas médias móveis são escolhidas com base na razão global ($\bar{I}/\bar{S}$) de acordo com os intervalos:

a) Se

$$\bar{I}/\bar{S}_n \leq 2{,}5 \qquad \text{usar MMS } 3x3$$

$$3{,}5 \leq \bar{I}/\bar{S}_n \leq 5{,}5 \qquad \text{usar MMS } 3x5$$

$$\bar{I}/\bar{S}_n \geq 6{,}5 \qquad \text{usar MMS } 3x9$$

b) Se

$$2{,}5 < \bar{I}/\bar{S}_n < 3{,}5 \quad \text{ou}$$

$$5{,}5 < \bar{I}/\bar{S}_n < 6{,}5\text{, repetir a) usando a razão } \bar{I}/\bar{S}_{n-1}$$

Onde n é o último ano completo e n-1 o ano anterior completo.

Observa-se que quanto mais irregular a série, maior será a média móvel sazonal, afim de preservar a estabilidade do padrão sazonal. No presente estudo utilizamos a opção que seleciona diferentes médias móveis sazonais para cada mês baseado na razão I/S mensal, tendo em vista a seqüência dos planos de estabilização da economia ocorridos a partir de 1986.

Outro importante aspecto do programa é a escolha das três médias móveis para isolar a componente tendência-ciclo, chamadas filtros de Henderson. As médias móveis de Henderson (MMH) de 9, 13 e 23 termos são aplicadas nos dados sazonalmente ajustados para obter uma estimativa apropriada da componente tendência-ciclo. Estas médias móveis são escolhidas com base na razão $\bar{I}/\bar{T}\bar{C}$ que relaciona o percentual médio absoluto, mês a mês, da variação a componente irregular (I) comparada com a tendência-ciclo (TC). Esta razão $\bar{I}/\bar{T}\bar{C}$ pode ser encontrada no topo da tabela D12 e na tabela F2.H.

O programa seleciona automaticamente uma MMH apropriada, de acordo com os seguintes intervalos:

Usar:

$$\text{MMH9} \quad \text{se} \quad \bar{I}/\bar{T}\bar{C} \leq 0{,}99$$

$$\text{MMH13} \quad \text{se} \quad 1{,}00 \leq \bar{I}/\bar{T}\bar{C} \leq 3{,}49$$

$$\text{MMH23} \quad \text{se} \quad \bar{I}/\bar{T}\bar{C} \geq 3{,}50$$

Se a razão global $\bar{I}/\bar{T}\bar{C}$ excede 3, a quantidade do movimento irregular é considerado alto.

Com relação aos modelos Box ε Jenkins (modelos autorregressivos - integrados - médias móveis - ARIMA) incorporados ao X11, podemos destacar o seguinte: a principal crítica ao método X11 (basicamente filtros de médias móveis) está no fato de que as estimativas mais recentes não possuem o mesmo grau de confiabilidade, se comparadas com as observações mais centrais da série, o que é uma limitação inerente aos procedimentos de alisamento linear, dado que, as primeiras e as últimas observações não podem ser alisadas com o mesmo conjunto de pesos (filtros) simétricos aplicados às observações centrais. Por este motivo, as estimativas para as observações correntes são revistas a medida que novos dados são adicionados à série original. Com a incorporação dos modelos ARIMA, os dados da série estudada são projetados um ano a frente, através de um modelo ARIMA "adequado", e a seguir é utilizado o procedimento usual dos filtros de médias móveis do X11. A utilização da série com previsões para um ano a frente reduz significativamente os problemas mencionados.

São 4 (quatro) os modelos testados automaticamente pelo programa, são eles:

(1) (0,1,1) (0,1,1)S
(2) (0,1,2) (0,1,1)S
(3) (2,1,0) (0,1,1)S
(4) (0,2,2) (0,1,1)S

Os modelos são testados seqüencialmente na ordem apresentada. Em outras palavras, se o modelo (1) "passa", então o programa não experimenta os demais, mas se o modelo (1) "falha", o programa experimenta o modelo (2) e assim por diante.

No caso em que os quatro modelos falhem, estudos separados são elaborados a fim de se obter um modelo de previsão. Identificados os modelos estes são incorporados ao programa X11-ARIMA.

Visando o aprimoramento deste trabalho, no decorrer de 1996, serão realizados estudos com o objetivo de identificar "o melhor" modelo ARIMA para os subitens sazonais.

Capítulo 3 - Interferência dos Planos Econômicos e seus Reflexos - Processo de Intervenção

A análise das séries de índices de preços (IPCA) através do procedimento desestacionalizador X11-ARIMA forneceu várias informações estatísticas. A seguir apresentaremos alguns resultados da região metropolitana de São Paulo.

Total de séries analisadas: 236
(I) Séries sazonais: 40
(II) Efeitos determinísticos prévios (calendário, páscoa etc) : 0
(III) Todas sofreram os efeitos dos planos econômicos.

A questão que se levanta é esta:

Os planos de estabilização da economia ocorridos desde 1986 afetaram as estimativas da componente estacional? Sim ou não? Se sim, em que grau?

Para responder a estas indagações, estudamos as séries do subgrupo Roupas, pois além de ser sazonal tem peso expressivo no índice (8,27) e do item Hortaliças e Verduras, por sua característica marcadamente sazonal.

Com as informações acima fornecidas, observamos que as séries estudadas não foram afetadas pelos efeitos estacionais determinísticos prévios: dias úteis, carnaval, semana santa e outros feriados intra-semanais (móveis ou não). Estes efeitos, quando presentes, podem perturbar de tal maneira que dificultaria, por exemplo, a obtenção das estimativas das componentes estocásticas, bem como a adequação de um modelo de previsão. Ausentes estas preocupações, resta-nos investigar a perturbação causada pelos efeitos dos planos econômicos (acontecimentos especias pontuais ocorridos a partir de 1986 - resíduos aleatórios atípicos)

A idéia é detectar a presença dos efeitos dos resíduos atípicos pontuais (impulsos), realizando uma regressão da transformação estacionária da série com as variáveis que "pegam" tais efeitos, expressas estas últimas com a mesma ordem de diferenciação da série analisada.

Primeiro passo: Detectar os resíduos atípicos pontuais (impulsos)

Critério estatístico: Estimação robusta (SAS - Statistics Program, The Univariate Procedure)

Primeira série analisada: Subgrupo Roupas - Logaritmos dos Relativos Mensais

Pontos atípicos identificados: nov89, mar90, mai94 e jun94

Variáveis artificiais para captar os valores dos resíduos pontuais: I8911, I9003, I9405 e I9406

Estas quatro variáveis são do tipo impulso, recebem valor 0 em todas as observações salvo nos meses de novembro de 1989, março de 1990, maio de 1994 e junho de 1994, respectivamente, em que assumem o valor 1. Com sua inclusão no modelo se pretende estimar o efeito que determinados acontecimentos especiais, ocorridos nos mencionados meses, exerceram sobre a série analisada.

As variáveis impulsos exercem sua influência sobre a componente irregular e seu efeito é o resultado de se multiplicar o valor do coeficiente estimado - o qual é bastante robusto - pelo valor da variável correspondente.
Fatores prévios que atuam sobre a componente irregular (FPI)

$$FPI_t = Exp\ (\ w_1\ I8911 + w_2\ I9003\ +\ w_3\ I9405 + w_4\ I9406\ )$$

$$FPI_t = Exp\ (0{,}029367\ *\ I8911 + 0{,}197695\ *\ I9003 + 0{,}163757\ *\ I9405 + 0{,}196318\ *\ I9406\ )$$

Assim, por exemplo:

$FPI_{nov\,89}\ =\ e^{0{,}029367} \Rightarrow FPI_{nov\,89} = 1{,}0298$ e assim por diante.

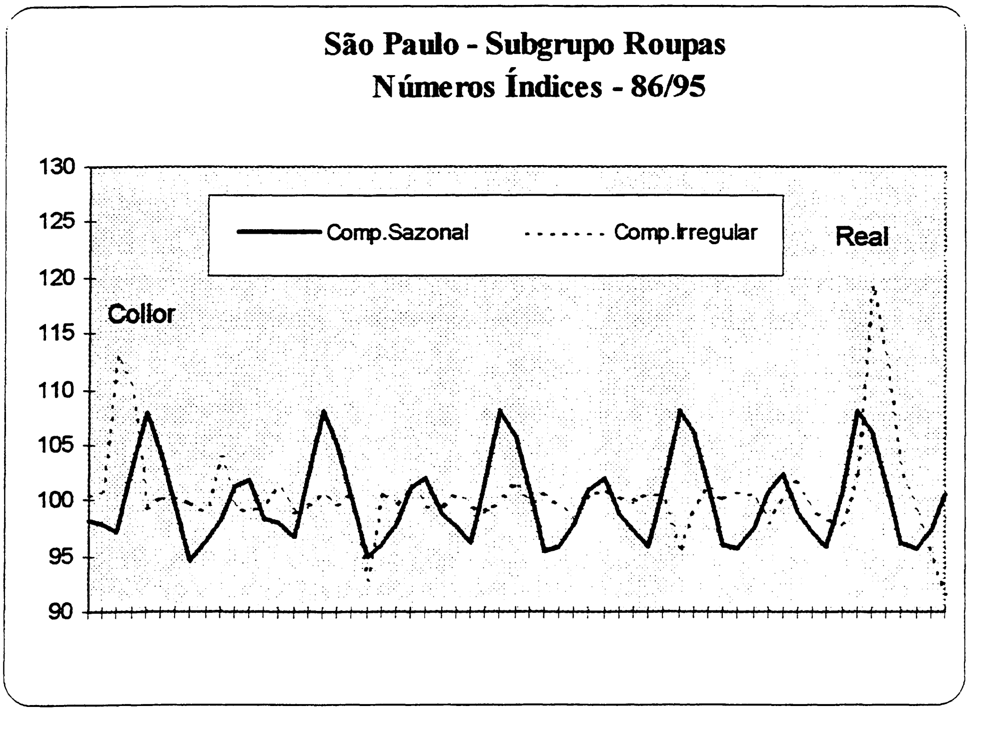
Retiram-se esses efeitos da série Roupas (variações mensais), isto é, divide-se, por exemplo, o relativo mensal de nov89=1,5182 por 1,0298 encontrando o resultado: 1,4743 e assim sucessivamente são obtidas os demais relativos. O próximo passo é montar a nova série de índice base fixa mar86=100.

Em seguida, calculamos o logaritmo da série Roupas corrigida dos tais efeitos que é submetida ao programa X11-ARIMA. Compara-se, então, os fatores estacionais das duas séries de número índice.

Os gráficos no Anexo 1, ano a ano, apresentam a componente estacional da série Roupas com e sem intervenção (C/I, S/I).

A análise dos gráficos permite concluir que os efeitos dos planos econômicos afetaram às estimativas da componente sazonal sim, porém de forma residual, o que não invalida a utilização dos fatores estacionais estimados para qualquer que seja a utilização.

O gráfico abaixo, mostra que a componente irregular foi a que mais captou os efeitos do Planos a partir de 1990, o que reforça a hipótese de que a componente sazonal não foi afetada pelos planos.



Segunda série analisada: Hortaliças e Verduras - logaritmos das variações mensais.

Pontos atípicos identificados:

Período 1986/1995: Fev86, Jan87, Fev88, Jan/Fev89, Jan/Fev90, Jan/Fev93, Jan94 e Jul94.

Período 1980/1995: Idem período 1986/1995.

Período 1980/1985: Mar80, Jan81, Jan/Fev83, Fev84 e Mar85.

Observa-se que no período de não-planos (80/85) os valores atípicos relacionados são devidos exclusivamente à sazonalidade (marcadamente sazonal). No período misto 1980/1995 os valores "atípicos" do período 80/85 são absorvidos pelo período 86/95 e deixam de ser atípicos. O que nos leva a questão: No caso de hortaliças e verduras os planos econômicos introduziram mais sazonalidade ainda?.

Trabalho análogo ao anterior foi elaborado para o item Hortaliças e Verduras e os gráficos, ano a ano, da componente estacional da série com (C/I) e sem intervenção (S/I) encontram-se no Anexo 2.

A análise dos gráficos permite concluir que se aceitarmos a hipótese de que os planos introduziram mais sazonalidade - o que se observa fortemente de 1986 a 1992 - o mesmo não se observa em 1994 e 1995 (a diferença é residual), anos para os quais se pretende divulgar séries sazonalmente ajustadas.

Nas duas intervenções realizadas , verificou-se que:

- para o Sub-Grupo Roupas nenhum efeito dos Planos foi observado;
- para o ítem - Hortaliças e Verduras, cujos subítens são marcadamente sazonais, pode-se admitir efeitos dos Planos nos anos de 1986 a 1992, não sendo significativos seus resultados para os anos de 1993,1994 e 1995.

Em resumo, podemos concluir que para os subitens sazonais identificados, como um todo, houve alguma influência dos Planos de Estabilização nos anos de 1986 a 1992, não se revelando significativos para os anos de 1993,1994 e 1995. Este fato é relevante para afimarmos que os Planos não afetaram as estimativas da componente estacional a partir de 1993.

Capítulo 4 - Apresentação de Resultados

4.1 - Subitens Sazonais

Na primeira fase de identificação dos subítens sazonais foi utilizado o procedimento desestacionalizador - médias móveis do X11-ARIMA, não sendo utilizada, nesta fase, a parte ARIMA (Modelos Box ε Jenkins).

Assim foram identificados, para os três períodos de estudo, os subitens sazonais. A Tabela 1 , no Anexo 3, fornece os resultados das principais estatísticas obtidas a cada nível de agregação do índice, para a RM de São Paulo. Os subitens com a notação tipo (T) 1 são os que foram identificados como sazonais pelo procedimento estatístico utilizado, com 2 são os subitens com sazonalidade identificável provavelmente não presente e os sem notações são os subitens com sazonalidade identificável não presente.

Estes resultados estão disponíveis para todas as onze áreas.

Restava, neste ponto, decidir a respeito do período a ser escolhido para a obtenção dos fatores estacionais, e, ainda, avaliar informações qualitativas adicionais relevantes , a saber:

- alterações metodológicas ocorridas na produção dos IPC's;
- se o padrão sazonal obtido é consistente com o conhecimento que se tem sobre o mercado dos produtos.

A título de ilustração o Anexo 4, contém uma síntese, bastante resumida, das análises qualitativas realizadas.

Desta forma, foram identificados os subitens sazonais segundo o perírdo e análises qualitativas, obtendo-se um conjunto de "subitens sazonais" por área.(Tabela 2-Anexo 5)

A estes "subitens sazonais" aplicamos o Procedimento Desestacionalizador X11-ARIMA utilizando, nesta fase, os modelos Box ε Jenkins para previsão um ano a frente, obtendo-se os fatores estacionais finais. Pretendeu-se com isto, contemplar o máximo possível o período mais recente, o Plano Real.

O quadro 1 revela o número de subitens sazonais identificados, bem como sua importância relativa (pesos) na composição dos índices, por área.

Quadro 1 - Total de Subitens sazonais por Área

	RJ	SP	BH	PA	RE	DF	BE	FT	SL	CT	GO
Subitens Sazonais	33	41	31	35	34	24	24	15	28	30	20
N Subitens R. Metropolitana	230	236	224	226	215	217	205	203	221	214	192
%Subitens Sazonais	15,7	16,9	13,4	15,9	16,7	10,1	12,2	8,4	13,1	14,5	10,4
Peso Subitens Sazonais (%)	5,1	8,1	5,7	12,1	7,0	1,9	4,2	3,7	5,3	8,5	6,5

4.1.2- Principais Estatísticas fornecidas pelo modelo X-11 ARIMA

Para exemplificar, a potencialidade de estatísticas obtidas por subitens sazonais, a Tabela 3, no Anexo 6, para a região metropolitana de São Paulo é ilustrativa, contém o modelo de previsão adotado, o erro padrão médio (%) da estimativa, a probabilidade X^2 (%), a razão global $\bar{I} / \bar{S}$ (quantidade de irregularidade em relação à sazonalidade) e as estatísticas de qualidade Fe, Fm, M7, Q e MCD. Sendo relevante observar os valores de M7 a mais importante, sob a ótica do ajustamento sazonal.

Cabe mencionar que as estatísticas acima descritas, bem como as demais obtidas pelo método, estão disponíveis para todos os subitens sazonais para as onze áreas.

4.1.3- Gráficos dos Fatores Sazonais

Os gráficos, no Anexo 7, mostram os padrões sazonais dos números índices , para a Região Metropolitana de São Paulo, para os anos de 1992 a 1996. Observa-se o comportamento destes padrões para os cinco anos. Isto mostra o grau de confiabilidade dos fatores obtidos, uma vez que, a maioria dos perfis sazonais dos subitens têm praticamente a mesma evolução. Para os subitens Farinha de Mandioca, Laranja Baía, Camarão e Agasalho de Homem, o programa sugere a necessidade de acompanhar seus perfis sazonais.

4.1.4- Fatores Sazonais

No Anexo 8 a tabela 4 relaciona os fatores estacionais, de São Paulo, por subitem para 1995.

4.2 - Séries Ajustadas a nível Brasil

Utilizando os fatores estacionais de 1994 e 1995,para as onze áreas, calculamos as séries ajustadas de 1994 e 1995, conforme procedimento descrito no Capítulo 3, item 1.4. O Anexo 9, apresenta quadros comparativos entre os índices originais e ajustados para o ano de 1995, a nível geral e grupos.

Na última coluna desta tabela, apresentamos o cálculo da variação acumulada no ano, para as duas séries, o que mostra a igualdade em 1995, entre a série original e a ajustada, reforçando desta forma o pressuposto básico para a construção de séries sazonalmente ajustada.

O comportamento destas duas séries pode ser observado através dos gráficos constantes do Anexo 10.

5 - Considerações Finais

- É possível produzir com rigor técnico o IPCA sazonalmente ajustado.

Durante todo o período da realização deste projeto - Dessazonalização dos Índices de Preços ao Consumidor, preocupou-nos sobremaneira as características específicas do processo inflacionário brasileiro, em especial as ocorridas nos últimos quinze anos.

Neste sentido, tomadas de decisões técnicas relevantes foram imperiosas tanto na definição do procedimento estatístico utilizado (X11-ARIMA), quanto às opções oferecidas pelo programa foram intensamente exploradas.

Análises qualitativas de resultados foram realizadas a luz de conhecimentos adquiridos sobre o mercado dos produtos e procedimentos técnicos na produção de IPC's foram avaliados de forma a verificar se provocavam pseudo sazonalidade.

Além disto, foram realizadas análises comparativas de séries com ou sem intervenção, objetivando investigar os possíveis efeitos dos planos nas séries analisadas.

Todos estes cuidados, possibilitam afirmar, com todo o rigor técnico ser possível calcular as séries de Índices de Preços ao Consumidor sazonalmente ajustadas para os anos 1994, 1995 e 1996. Tem-se claro que as séries pretéritas (1994 e 1995) teriam uma utilização restrita ao campo acadêmico.

- Atualização dos fatores sazonais

Para o ano de 1996, serão produzidas as séries dessazonalizadas mês a mês com os conjuntos anuais de fatores estacionais estimados a partir de dezembro de 1995.

Em janeiro de 1997, o trabalho todo será revisado, com incorporação ou perda de subitens sazonais e novos conjuntos anuais de fatores estacionais serão estimados apenas para o ano de 1997 e assim sucessivamente.

- Os resultados obtidos para os diversos níveis de agregação são valiosos para a análise e compreensão do fenômeno sazonal

A abrangência do índice possibilita fornecer aos usuários dos Índices de Preços como um todo, um valioso instrumental para a análise e compreensão do fenômeno sazonal, uma vez que forneceremos não só o perfil sazonal, por subitem (e para todos os níveis de agregação), para as onze áreas e Brasil; como também poderemos fornecer a evolução dos perfis sazonais obtidos.

Estes dados são relevantes para tomadas de decisões a nível setorial e global, fornecendo relevantes subsídios para análise de políticas específicas.

- As Séries Ajustadas e seus Usos

Deve-se enfatizar que o projeto visa a produção do IPCA dessazonalizado, objetivando fornecer, como já mencionado, um estatística adicional que melhor reflita a tendência inflacionária (quando a componente irregular for irrelevante).

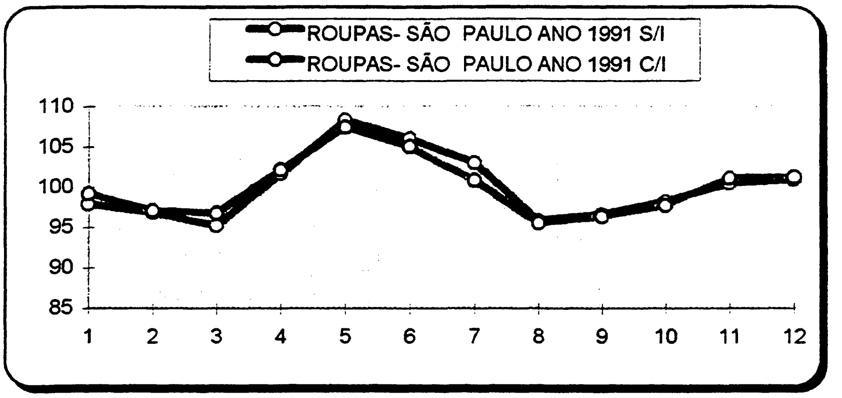
No Brasil, é recorrente surgirem demandas por séries de índices de preços sazonalmente ajustadas às instituições produtoras de índices de preços ao consumidor, por diversos agentes econômicos (governos, entidades de classes, sindicatos, etc), em período pós adoção de planos de estabilização econômica. Este fato, tem sua razão de ser, pois os agentes econômicos objetivam organizar seus possíveis padrões de racionalidade - de acordo com os contextos de incertezas - que vão exprimir-se na formação das expectativas, no direcionamento das tomadas de decisão individual e até mesmo na coordenação das decisões coletivas.

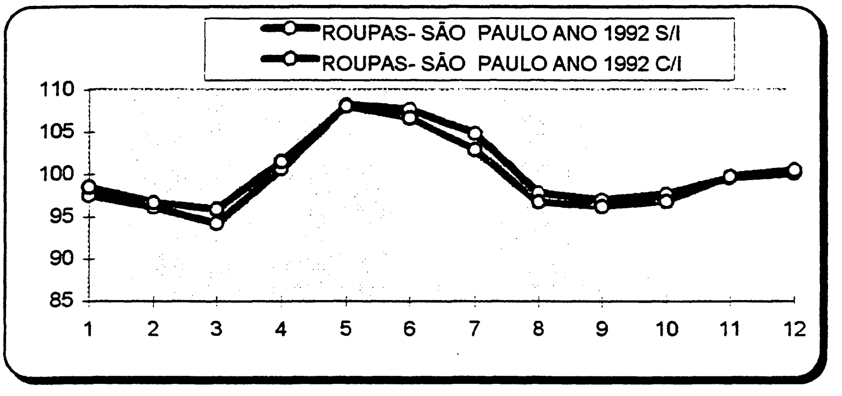
Sabemos que, num contexto de reaver a confiança na eficácia das políticas adotadas é fundamental a avaliação precisa dos indicadores econômicos, em especial, os utilizados para aferir a inflação. A análise correta do fenômeno inflacionário, num cenário de estabilização de preços mais duradouro e persistente, é fundamental para a reversão de expectativas e procedimentos remanescentes da cultura inflacionária.

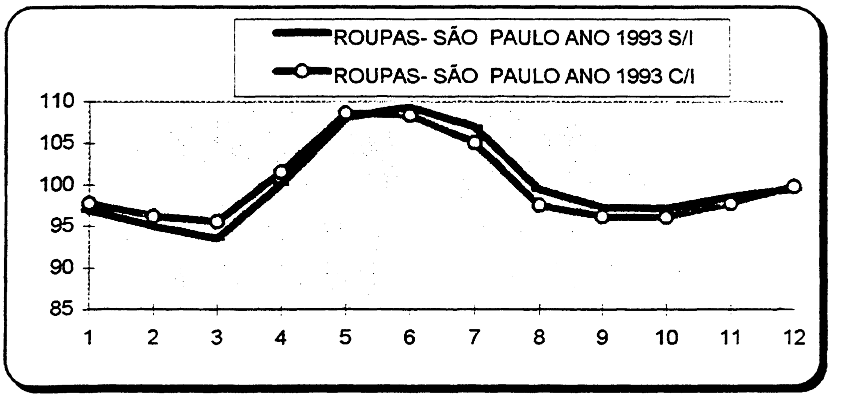
Neste sentido sabemos que a produção de índices de preços sazonalmente ajustados constitui-se em ferramenta importante para a interpretação do processo inflcionário em nosso país.

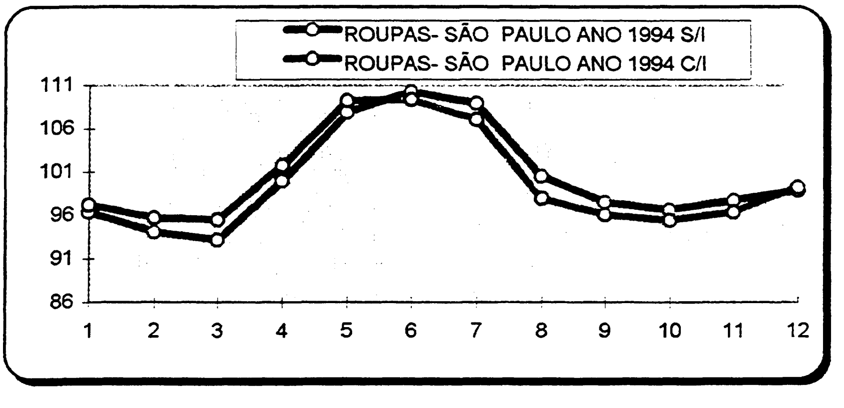
ANEXO 1

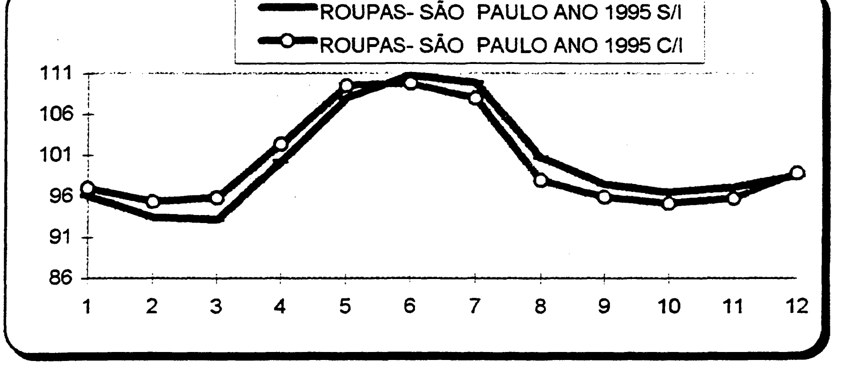
FATORES ESTACIONAIS
NÚMEROS ÍNDICES - IPCA



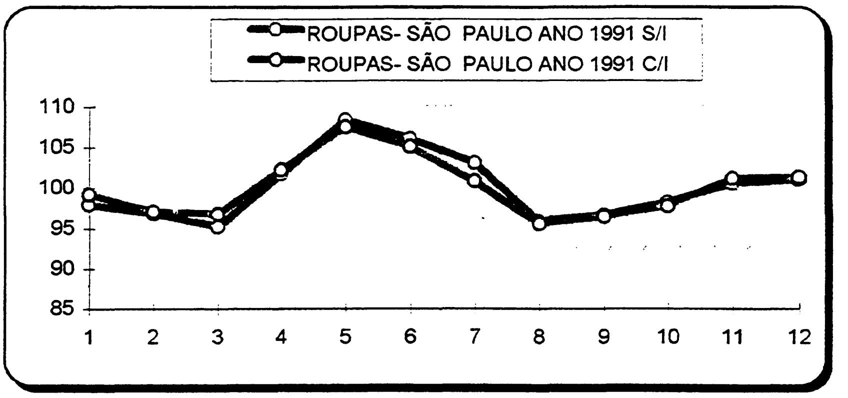


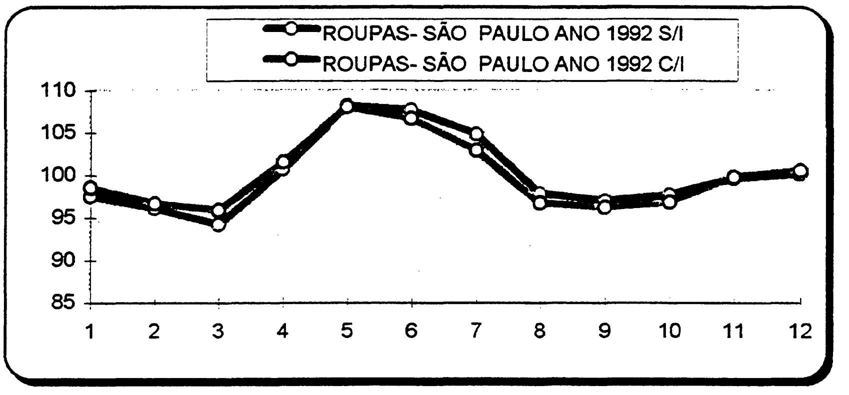


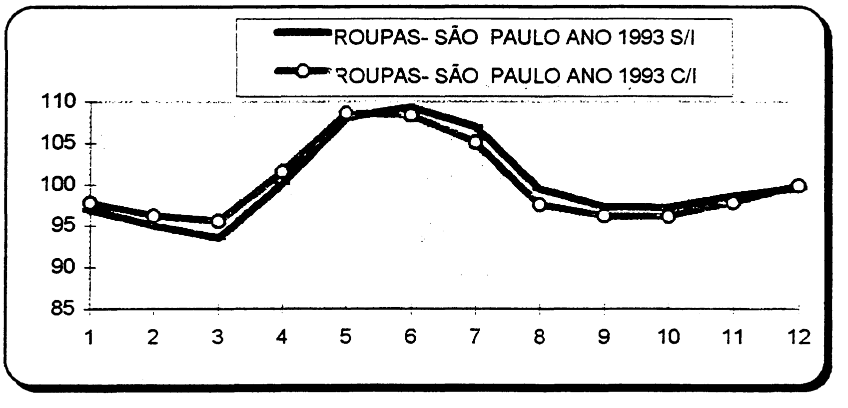


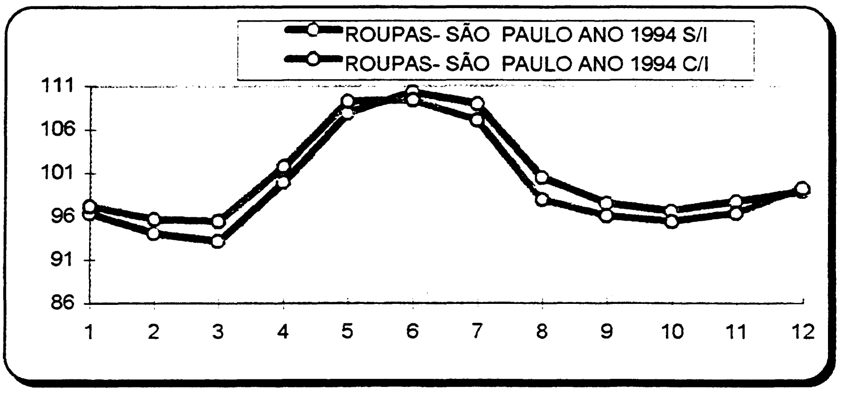


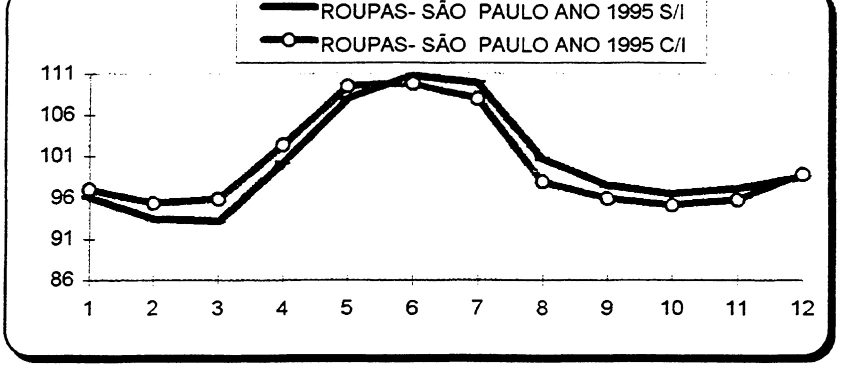
FATORES ESTACIONAIS
NÚMEROS ÍNDICES - IPCA





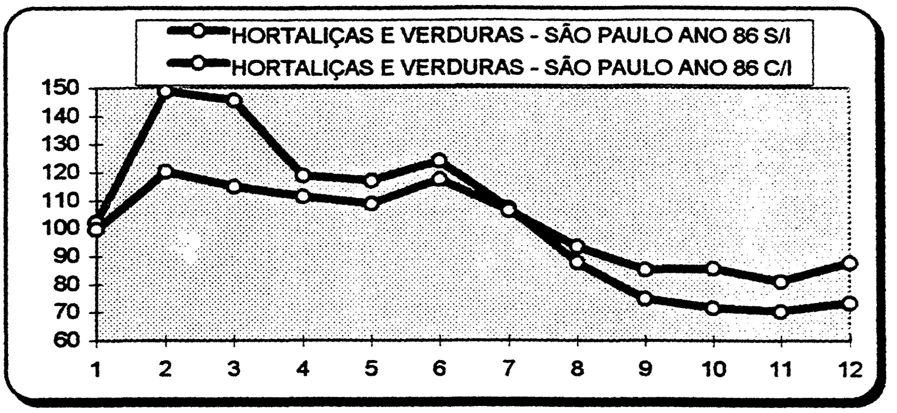


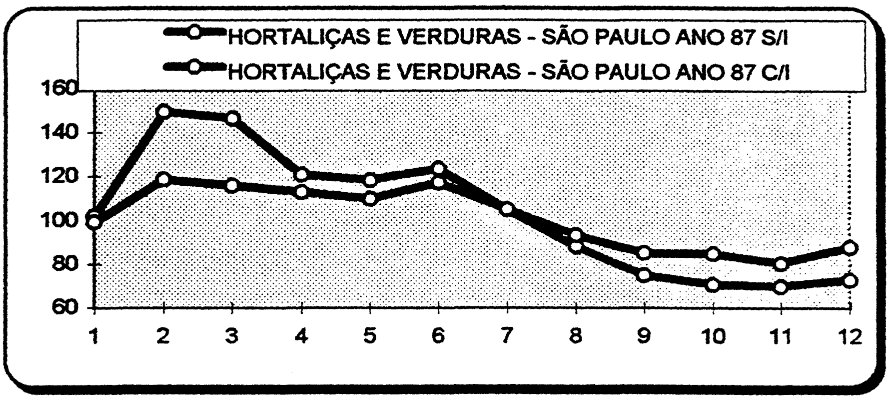


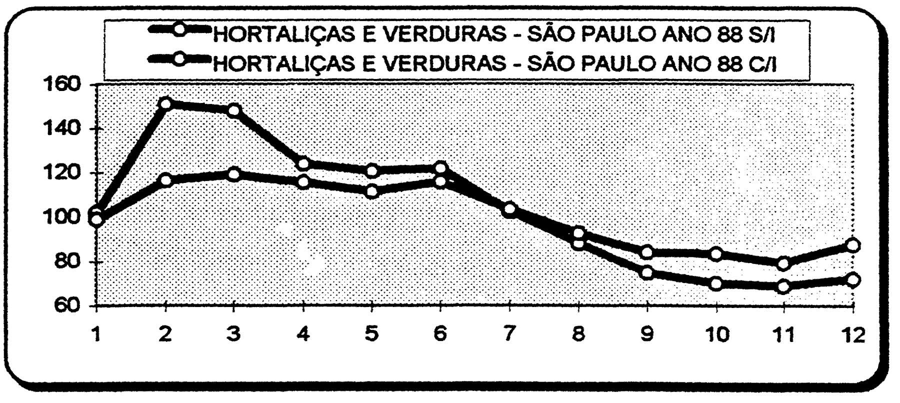


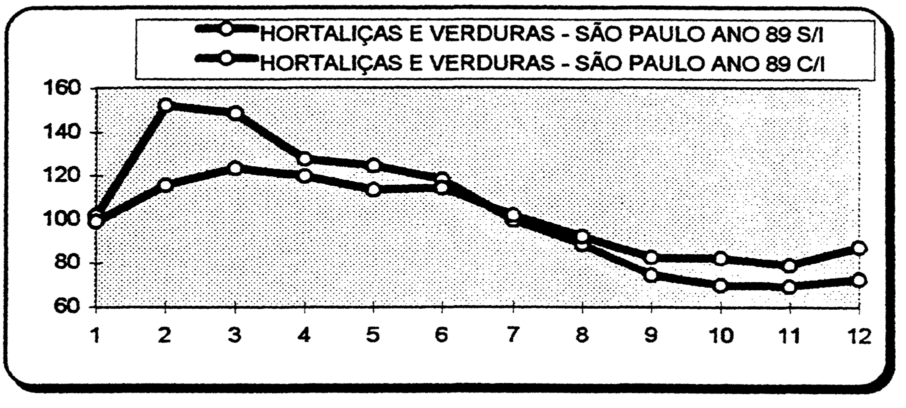
ANEXO 2

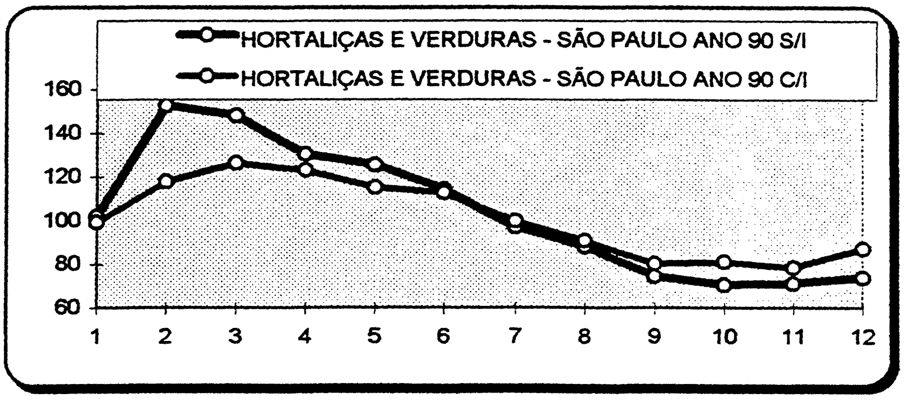
FATORES ESTACIONAIS
NÚMEROS ÍNDICES - IPCA



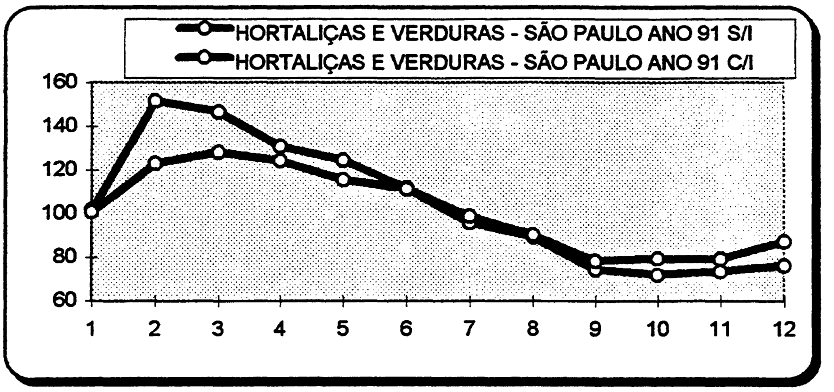


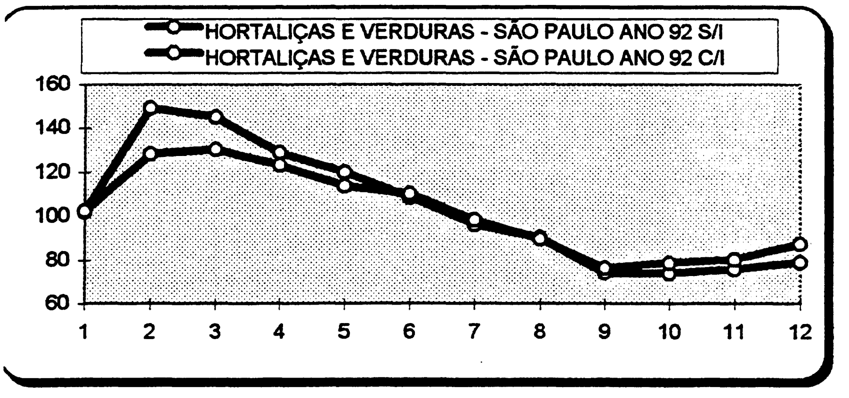


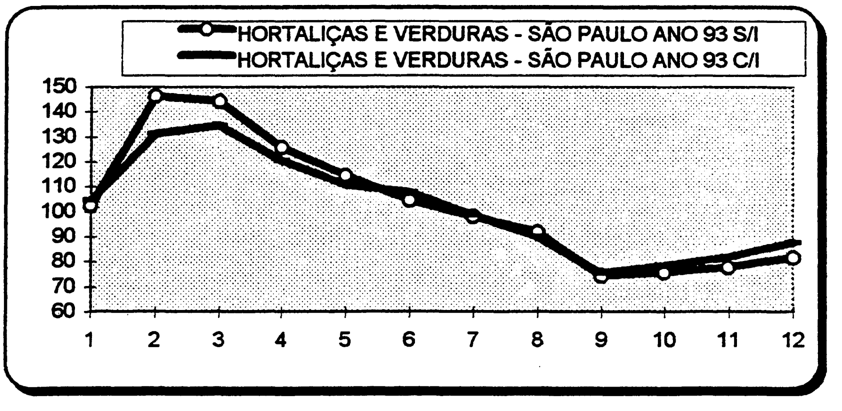


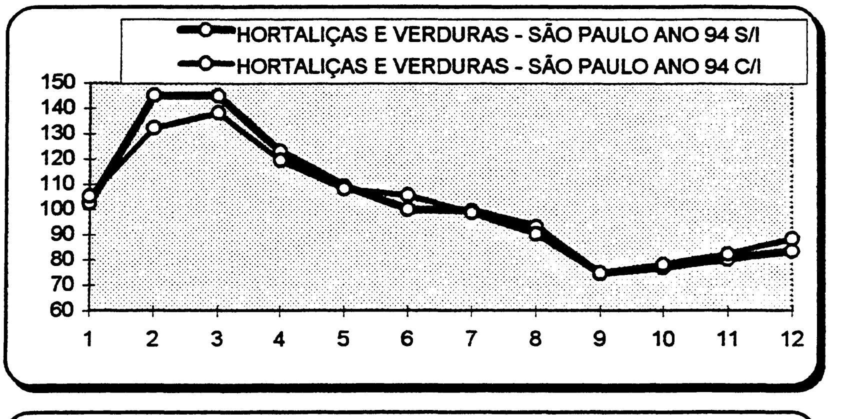


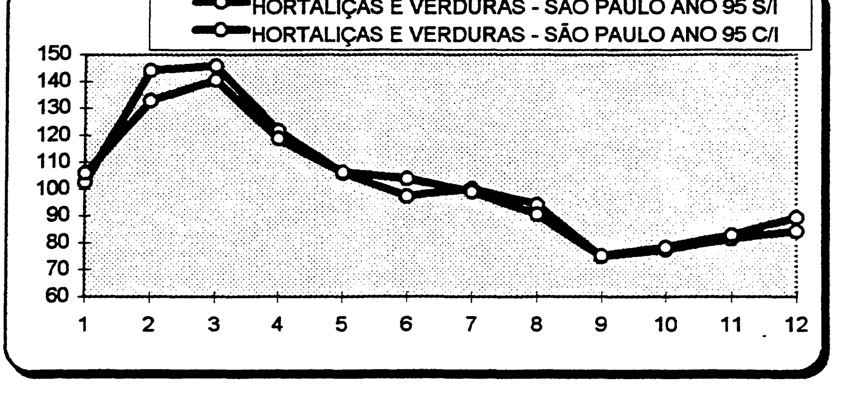
FATORES ESTACIONAIS
NÚMEROS ÍNDICES - IPCA











ANEXO 3

Tabela 1 - Principais Estatísticas fornecidas pelo programa X11-ARIMA (*)
São Paulo - IPCA

Alimentação e Bebidas

Códigos	Descrição	1980 a 1985					1986 a maio 1995					1980 a maio 1995				
		T	Fe	Fm	M7	Q	T	Fe	Fm	M7	Q	T	Fe	Fm	M7	Q
0000000	**Indice Geral**		0.48	2.34	3.00	1.15		7,52	5,81	1,27	0,55		9.14	9.98	1.42	0.61
1000000	**Alimentação e Bebidas**		1.53	1.98	2.06	0.80		5,70	4,39	1,33	0,56		7.37	7.76	1.43	0.68
1100000	Alimentação no Domicílio		1.53	1.66	1.98	0.85		5,58	4,57	1,36	0,61		6.84	8.23	1.52	0.71
1101000	Cereais, Legum.,Oleagin		1.06	2.45	2.60	1.11		2,95	5,25	1,96	0,82		4.64	8.34	1.86	0.84
1101402	Arroz		4.96	2.11	1.16	0.57		4,10	10,71	2,18	0,91		6.11	14.16	2.01	0.88
1101473	Feijão Rajado	1	10.29	2.78	0.86	0.47	2	4,54	1,42	1,11	0,53	2	7.80	2.82	1.00	0.50
1102000	Farinha, Féculas, Massas		1.84	3.53	2.18	0.95		2,05	4,72	2,27	0,86		3.54	8.79	2.17	0.95
1102406	Macarrão		3.14	2.13	1.46	0.76		1,73	3,59	2,27	0,87		2.67	6.19	2.19	0.92
1102408	Fubá de Milho	2	10.45	3.70	0.93	0.47	2	6,60	1,92	0,98	0,51	1	11.44	3.79	0.90	0.49
1102412	Farinha de Trigo		0.77	2.09	2.93	1.25		0,66	5,77	3,00	1,17		1.80	7.46	2.54	1.02
1102423	Farinha de Mandioca		5.61	2.59	1.15	0.59	1	25,93	4,42	0,62	0,25	2	22.58	9.58	0.89	0.44
1103000	Tuberc., Raízes, Legumes		9.25	4.19	1.03	0.55	1	28,97	1,98	0,47	0,20	1	27.43	4.12	0.59	0.38
1103403	Batata Inglesa	1	8.50	0.72	0.73	0.45	1	21,06	1,55	0,53	0,30	1	16.78	3.95	0.75	0.45
1103420	Abobrinha	1	18.36	1.81	0.58	0.50	1	11,01	1,26	0,70	0,42	1	20.20	1.50	0.53	0.37
1103421	Chuchu	1	12.71	0.40	0.57	0.42	1	20,70	2,39	0,58	0,35	1	26.34	1.47	0.47	0.31
1103425	Pepino	1	8.84	2.04	0.86	0.62	1	11,32	1,16	0,68	0,42	1	12.83	1.51	0.67	0.39
1103426	Pimentão		3.24	1.05	1.25	0.62	1	7,49	1,72	0,90	0,41	1	7.91	2.07	0.91	0.45
1103428	Tomate		3.39	2.90	1.52	1.19	1	9,36	1,98	0,83	0,39	1	11.73	2.51	0.79	0.42
1103429	Vagem	1	14.14	3.66	0.80	0.84	1	14,56	1,58	0,63	0,37	1	25.16	2.17	0.52	0.38
1103431	Beringela		6.03	2.78	1.13	0.66		6,28	3,04	1,13	0,48		7.87	3.70	1.07	0.51
1103443	Cebola	1	12.10	1.17	0.66	0.57	1	16,34	1,96	0,63	0,39	1	23.63	1.90	0.52	0.34
1103444	Cenoura	1	20.71	1.25	0.51	0.37	1	22,77	3,78	0,63	0,30	1	44.08	2.47	0.40	0.28
1104000	Açúcares e Derivados	1	7.56	0.82	0.79	0.43		5,61	4,96	1,39	0,57		4.76	6.91	1.71	0.70
1104403	Açúcar Refinado	1	7.09	0.81	0.81	0.45		6,44	3,46	1,16	0,50		4.79	5.13	1.53	0.71
1104418	Balas, Chicletes		2.32	3.62	1.96	0.79		3,96	4,79	1,64	0,64		4.67	7.52	1.78	0.76
1104423	Chocolate em Barra		0.61	1.74	2.96	1.10		2,86	2,45	1,58	0,62		3.57	7.09	1.99	0.85
1105000	Hortaliças e Verduras	1	29.25	3.51	0.55	0.35	1	59,80	1,93	0,33	0,21	1	59.56	7.73	0.50	0.31
1105401	Alface	1	14.16	2.68	0.73	0.51	1	38,00	1,19	0,37	0,24	1	38.85	5.60	0.55	0.36
1105403	Chicória	1	21.63	4.42	0.68	0.44	1	49,20	1,41	0,34	0,25	1	45.12	6.56	0.54	0.31
1105405	Couve	1	22.66	3.78	0.64	0.45	1	62,54	2,08	0,32	0,21	1	57.98	7.12	0.49	0.28
1105406	Couve-Flor	1	41.14	1.31	0.36	0.29	1	47,01	1,38	0,34	0,19	1	66.33	4.20	0.38	0.24
1105410	Repolho	1	41.60	1.41	0.37	0.33	1	71,72	1,61	0,29	0,14	1	82.64	6.40	0.40	0.26
1105413	Agrião	1	37.60	3.24	0.47	0.35	1	55,81	1,93	0,34	0,17	1	61.69	6.49	0.46	0.22
1105419	Brócolis	1	49.29	0.77	0.31	0.26	1	70,27	1,74	0,29	0,14	1	79.21	6.19	0.40	0.20

(*) Sem Modelos de Previsão

continua

Alimentação e Bebidas

Códigos	Descrição	1980 a 1985					1986 a maio 1995					1980 a maio 1995				
		T	Fe	Fm	M7	Q	T	Fe	Fm	M7	Q	T	Fe	Fm	M7	Q
1106000	Frutas	1	37.22	6.25	0.59	0.33	1	16,95	0,63	0,51	0,30	1	25.15	1.44	0.47	0.27
1106403	Abacaxi	1	17.73	4.99	0.79	0.39	1	25,54	1,66	0,48	0,27	1	31.71	5.38	0.60	0.30
1106405	Banana-D'Agua	1	19.09	1.16	0.52	0.31	2	6,51	1,45	0,93	0,42	1	7.21	1.24	0.86	0.39
1106408	Banana-Prata	1	21.53	1.77	0.53	0.31	1	27,67	2,08	0,49	0,27	1	28.43	3.27	0.54	0.28
1106411	Laranja-Baia	2	6.93	1.58	0.92	0.58	1	11,43	2,43	0,79	0,39	1	10.96	1.80	0.75	0.35
1106412	Laranja-Lima	1	69.75	1.19	0.27	0.31	1	125,95	0,67	0,19	0,10	1	155.84	1.15	0.18	0.16
1106415	Limão	1	61.75	3.27	0.37	0.30	1	39,32	3,79	0,48	0,27	1	78.30	3.09	0.32	0.23
1106417	Maçã	1	30.27	1.83	0.45	0.36	1	30,74	4,68	0,58	0,27	1	42.77	5.96	0.54	0.32
1106418	Mamão	1	18.20	0.77	0.50	0.34		2,11	1,37	1,62	0,69		2.28	1.53	1.60	0.69
1106421	Melancia	2	3.59	0.65	1.12	0.66	2	5,92	0,77	0,89	0,46		4.07	1.89	1.25	0.55
1106423	Pera	1	48.45	2.10	0.37	0.30	1	38,45	2,88	0,45	0,24	1	51.29	6.18	0.50	0.27
1106427	Tangerina	1	22.60	3.39	0.62	0.43	1	22.79	0,71	0,45	0,28	1	30.87	1.10	0.41	0.26
1106428	Uva	1	87.66	1.19	0.25	0.25	1	30,20	1,39	0,43	0,26	1	52.96	3.41	0.40	0.27
1106439	Laranja-Pera	1	49.85	0.76	0.30	0.18	1	49,35	5,88	0,50	0,22	1	59.39	5.56	0.45	0.22
1107000	Carnes Frescas, Vísceras		16.03	14.17	1.24	0.60		2,08	1,50	1,66	0,67		5.11	1.91	1.12	0.49
1107418	Carne de Porco		6.36	4.29	1.25	0.57	2	4,40	1,98	1,21	0,56		6.95	2.32	1.00	0.52
1107484	Contrafilé		15.77	13.86	1.24	0.57	2	3,88	0,79	1,10	0,49	1	10.34	1.63	0.76	0.39
1107487	Chã-de-Dentro		17.00	13.99	1.20	0.53	2	4,13	1,11	1,12	0,50	1	10.74	1.76	0.76	0.41
1107488	Alcatra		16.51	14.15	1.22	0.56	2	3,89	0,90	1,12	0,51	1	10.23	1.64	0.76	0.42
1107489	Patinho		16.97	14.27	1.21	0.54	2	3,91	1,10	1,15	0,51	1	10.60	1.77	0.76	0.42
1107495	Acém	2	22.47	10.73	0.93	0.52		1,45	2,32	2,19	0,91		1.79	2.10	1.93	0.80
1108000	Pescado	1	28.28	0.79	0.41	0.22	1	47,18	6,88	0,54	0,28	2	34.28	14.76	0.86	0.41
1108412	Peixe Sardinha	1	25.82	0.46	0.40	0.19	1	47,55	0,50	0,30	0,13	1	37.09	5.55	0.56	0.28
1108413	Camarão	1	15.48	2.33	0.67	0.35	1	14,61	3,74	0,79	0,43	2	18.45	7.86	0.91	0.46
1108438	Peixe Pescada	1	8.66	0.54	0.70	0.35	1	18,05	5,23	0,79	0,29	2	18.26	9.29	0.98	0.37
1109000	Carnes, Peixes Indust.		3.23	1.66	1.36	0.60		7,57	3,48	1,07	0,41		9.79	4.98	1.06	0.45
1109402	Presunto	2	5.28	0.63	0.92	0.50		8,46	5,45	1,17	0,54		12.40	6.09	1.01	0.52
1109407	Salsicha e Salsichão		4.28	4.48	1.54	0.62		4,10	3,31	1,44	0,58		5.34	6.48	1.57	0.68
1109408	Linguiça	2	5.95	1.74	1.01	0.52		6,02	2,60	1,11	0,46	2	9.93	3.54	0.94	0.41
1109410	Mort. Salame, Salaminho		1.31	2.16	2.26	0.86	1	8,39	1,95	0,88	0,36		8.55	5.97	1.21	0.57
1110000	Aves e Ovos		1.79	2.09	1.92	0.85		2,41	2,57	1,75	0,78		3.31	2.41	1.46	0.67
1110409	Frango		3.77	3.80	1.56	0.74		2,15	2,67	1,87	0,81		4.39	2.43	1.28	0.65
1110444	Ovo de Galinha		5.05	2.89	1.24	0.57	1	8,44	1,64	0,84	0,42	2	11.01	4.03	0.93	0.47
1111000	Leite e Derivados		3.03	2.26	1.51	0.60		4,48	2,28	1,24	0,52		3.45	3.87	1.64	0.70
1111404	Leite Pasteurizado		3.61	4.56	1.69	0.68		5,56	2,74	1,17	0,53		5.09	4.23	1.39	0.63

continua

Alimentação e Bebidas

Códigos	Descrição	1980 a 1985					1986 à maio 1995					1980 a maio 1995				
		T	Fe	Fm	M7	Q	T	Fe	Fm	M7	Q	T	Fe	Fm	M7	Q
1111409	Leite em Pó	1	14.33	1.55	0.64	0.36		3.34	3.27	1.59	0.61		1.77	4.78	2.45	1.99
1111419	Yogurt	1	8.79	1.25	0.78	0.38		8,69	5,33	1,15	0,44		8.35	6.67	1.27	0.52
1111423	Queijo Minas	1	7.13	0.69	0.80	0.41		1,67	2,93	2,17	0,91		2.29	4.90	2.18	0.92
1111424	Queijo Prato e Muzzarela	2	4.59	1.64	1.14	0.55		6,58	3,13	1,12	0,53		3.97	4.32	1.59	0.75
1112000	Panificados		2.34	1.65	1.60	0.72		3,59	4,65	1,71	0,75		4.98	6.51	1.63	0.72
1112403	Biscoitos		2.54	3.18	1.80	0.79		5,11	3,53	1,31	0,58		4.63	7.85	1.82	0.75
1112415	Pão Frances		2.21	1.85	1.69	0.83		3,79	4,86	1,69	0,70		5.11	6.12	1.58	0.67
1112417	Pão Doce		2.05	1.67	1.71	0.70		2,13	2,26	1,80	0,77		3.73	3.50	1.53	0.60
1113000	Óleos e Gorduras		3.49	3.43	1.55	0.69		8,88	4,14	1,04	0,52		4.44	7.58	1.83	0.80
1113413	Óleo de Soja		2.74	3.32	1.76	0.82		8,86	4,10	1,04	0,53		4.37	7.13	1.80	0.83
1114000	Bebidas e Infusões		1.92	6.10	2.51	1.08		6,49	9,39	1,65	0,67		11.62	9.05	1.21	0.57
1114422	Café Moído		2.17	6.47	2.47	1.10		4,81	7,93	1,79	0,71		10.52	6.68	1.13	0.50
1115000	Enlatados e Conservas	1	8.17	2.55	0.95	0.49		5,18	2,71	1,21	0,55		9.60	6.34	1.16	0.55
1115406	Ervilha em Lata		4.67	8.55	1.87	0.80		7,30	6,44	1,34	0,53		9.04	11.00	1.49	0.62
1115416	Palmito em Conserva	1	15.40	2.19	0.66	0.42		1,37	1,81	2,13	0,81		5.04	2.61	1.21	0.52
1115439	Sardinha em Lata		14.84	7.63	1.00	0.46		7,71	3,33	1,05	0,55	2	13.44	5.91	0.96	0.47
1115457	Azeitona		2.25	2.36	1.77	0.76		5,60	2,27	1,11	0,44		6.33	7.23	1.51	0.62
1116000	Sal e Condimentos		0.72	1.53	2.77	1.13		2,37	5,35	2,20	0,87		2.89	9.27	2.39	0.97
1116405	Massa de Tomate		2.45	5.22	2.15	0.80		6,85	6,48	1,39	0,59		5.97	13.69	2.01	0.81
1116410	Alho		0.43	1.79	3.00	1.26	2	6,09	0,77	0,87	0,49		4.50	5.05	1.57	0.75
1116433	Maionese		2.09	10.67	2.49	1.00		4,36	6,21	1,71	0,72		3.81	9.88	2.19	0.90
1200000	Alimentação Fora Domicílio		5.42	5.84	1.50	0.60		6,71	4,46	1,23	0,53		9.82	7.31	1.21	0.61
1201000	Alimentação Fora Domicílio		5.42	5.84	1.50	0.60		6,71	4,46	1,23	0,53		9.82	7.31	1.21	0.61
1201401	Refeição		6.97	5.95	1.33	0.52		5,47	2,62	1,16	0,44		8.49	3.99	1.06	0.39
1201403	Lanche		1.72	3.58	2.27	0.94		4,83	3,96	1,40	0,59		7.82	5.90	1.26	0.56

continua

Habitação

Códigos	Descrição	1980 a 1985					1986 a maio 1995					1980 a maio 1995				
		T	Fe	Fm	M7	Q	T	Fe	Fm	M7	Q	T	Fe	Fm	M7	Q
2000000	**Habitacao**		1.51	2.24	2.13	0.81		7.54	5.66	1.26	0.51		6.84	10.39	1.67	0.63
2100000	Encargos e Manutenção	2	3.49	0.97	1.19	0.51	2	10.36	3.83	0.95	0.46		7.65	9.76	1.54	0.68
2101000	Habitação	2	5.67	0.70	0.90	0.42		8,11	5,12	1,17	0,47		5.49	10.94	1.90	0.76
2101401	Aluguel Residencial		1.39	2.13	2.20	0.94		7,81	4,21	1,12	0,54		5.64	9.36	1.76	0.72
2101402	Condomínio		4.00	13.60	2.32	0.92		4,89	6,56	1,65	0,66		4.19	11.91	2.26	0.90
2101404	Taxa de Água e Esgoto		2.49	2.28	1.67	0.81		4,13	2,19	1,28	0,56		5.36	4.20	1.35	0.56
2103000	Reparos		1.34	2.04	2.21	0.86		8,20	4,10	1,08	0,50		7.22	9.94	1.60	0.73
2103405	Ferragens		1.79	2.17	1.94	0.75		5,27	2,81	1,21	0,58		5.85	4.91	1.36	0.63
2103409	Mat. de Pintura		2.49	2.58	1.72	0.73		7,11	4,03	1,16	0,52		8.16	8.93	1.44	0,67
2103414	Tinta para Casa		0.80	6.67	2.98	1.26		9,00	5,99	1,18	0,55		7.46	13.16	1.76	0.77
2104000	Artigos de Limpeza		3.64	2.88	1.46	0.60		6,75	4,87	1,26	0,62		5.63	13.12	2.03	0.82
2104408	Detergente		2.67	3.43	1.80	0.75		8,34	7,29	1,32	0,57		6.10	16.89	2.17	0.92
2104409	Sabão (Pó e Barra)		5.32	2.73	1.19	0.58		5,08	6,54	1,62	0,68		4.00	12.92	2.32	0.98
2104412	Desinfetante		2.84	2.72	1.63	0.62	1	9,00	2,55	0,90	0,35		7.94	9.78	1.51	0,66
2200000	Combustíveis, Energia		1.09	0.67	2.03	0.79		4,65	8,01	1,82	0,76		5.16	10.64	1.94	0.84
2201000	Combustíveis Uso Domést.		1.83	0.33	1.48	0.62		3,21	7,78	2,17	0,87		3.81	9.67	2.17	0.90
2202000	Energia Elétrica		0.66	2.97	3.00	1.11		4,42	7,90	1,86	0,76		5.19	11.31	1.99	0.80
2202403	Energia Elétrica		1.43	2.84	2.33	0.88		4,25	7,19	1,83	0,76		5.22	10.08	1.89	0.77

continua ...

Artigos de Residência

Códigos	Descrição	1980 a 1985					1986 a maio 1995					1980 a maio 1995				
		T	Fe	Fm	M7	Q	T	Fe	Fm	M7	Q	T	Fe	Fm	M7	Q
3000000	**Artigos de Residencia**		1.18	4.87	2.73	1.10	2	4.63	2.07	1.19	0.46		4.35	5.28	1.62	0.67
3100000	Móveis e Utensílios		2.32	1.18	1.51	0.59		3,92	2,54	1,37	0,53		4.00	6.14	1.78	0.71
3101000	Mobiliário		1.72	9.66	2.55	0.94		5,98	2,91	1,15	0,43		7.74	7.81	1.40	0.58
3101402	Móveis para Sala		1.89	3.48	2.14	0.77		6,17	2,30	1,06	0,41		7.60	4.65	1.17	0.44
3101403	Móveis para Quarto		2.37	4.55	2.09	0.82	1	10,31	3,28	0,90	0,35		7.77	7.20	1.36	0.49
3101415	Móveis p/ Copa e Cozinha		1.52	3.25	2.34	0.90	2	3,81	1,88	1,29	0,56		4.33	5.45	1.64	0.63
3102000	Utensílios e Enfeites	1	12.23	2.77	0.79	0.33		1,27	1,81	2,21	0,85		0.75	2.63	3.00	1.17
3102406	Cortinas	1	13.94	0.55	0.56	0.27	2	3,78	2,03	1,31	0,51		6.94	2.77	1.05	0.44
3103000	Cama, Mesa e Banho		2.79	2.10	1.54	0.64	2	6,68	2,04	0,99	0,39		7.18	5.75	1.30	0.47
3103401	Roupa de Cama		2.42	1.51	1.54	0.64	1	7,38	2,17	0,96	0,43		7.44	6.30	1.32	0.57
3103403	Roupa de Banho	2	3.69	2.31	1.37	0.50		3,85	3,05	1,45	0,55		5.22	5.52	1.50	0.61
3200000	Aparelhos Elétricos		3.34	4.26	1.72	0.72	2	4,93	1,41	1,07	0,43		4.70	5.04	1.53	0.71
3201000	Eletrodom. e Equipam.		2.48	5.98	2.24	0.88	1	9,01	1,78	0,83	0,31		8.68	6.37	1.23	0.50
3201401	Refrigerador		4.42	4.83	1.56	0.67	1	9,35	1,92	0,83	0,32		11.90	6.35	1.05	0.43
3201403	Máquina de Costura	1	9.22	1.51	0.79	0.33	1	9,15	0,71	0,71	0,29		9.17	4.77	1.08	0.45
3201406	Máquina Lavar/Secar		2.23	4.76	2.19	0.92	2	6,79	1,97	0,97	0,41		5.57	5.68	1.47	0.60
3201421	Fogão	1	9.50	1.26	0.75	0.36	1	8,20	2,57	0,95	0,42		6.21	5.92	1.41	0.57
3202000	TV e Som		2.20	6.80	2.47	1.06		1,82	1,18	1,70	0,71		1.51	3.03	2.31	0.94
3202401	Televisor	1	10.94	1.28	0.70	0.41		1,88	2,11	1,88	0,77		1.30	3.47	2.58	1.05
3202402	Rádio		0.81	4.01	2.96	1.13		0,70	1,81	2,90	1,09		0.88	3.87	2.91	1.09
3202403	Aparelho de Som	2	3.45	1.81	1.34	0.56		2,04	1,21	1,61	0,64		2.63	3.17	1.77	0.72

continua

Vestuário

Códigos	Descrição	1980 a 1985					1986 a maio 1995					1980 a maio 1995				
		T	Fe	Fm	M7	Q	T	Fe	Fm	M7	Q	T	Fe	Fm	M7	Q
4000000	**Vestuario**	1	27.68	5.99	0.67	0.37	2	5.75	1.48	1.00	0.42	2	9.44	3.86	0.99	0.47
4100000	Roupas	1	35.69	7.89	0.66	0.32	1	8,83	0,78	0,73	0,34	1	15.64	3.37	0.74	0.39
4101000	Roupa de Homem	1	17.66	4.86	0.78	0.34	2	5,71	1,52	1,01	0,39	1	11.14	3.50	0.89	0.38
4101402	Calça Comprida	1	12.54	0.97	0.63	0.32	2	3,83	0,46	1,04	0,45	1	8.96	2.38	0.89	0.43
4101404	Terno		6.43	6.32	1.42	0.64	1	9,43	1,87	0,82	0,33	2	15.75	5.41	0.86	0.37
4101405	Agasalho	1	159.88	0.58	0.16	0.16	1	30,77	8,05	0,71	0,44	1	42.34	8.11	0.61	0.40
4101409	Camisa	1	9.50	0.60	0.68	0.37		5.38	3,32	1,25	0,50	2	9.47	3.70	0.98	0.41
4101410	Camiseta	1	8.29	0.37	0.70	0.38		6,10	2,22	1,06	0,40		4.25	3.49	1.43	0.55
4102000	Roupa de Mulher	1	42.37	2.57	0.42	0.24	1	11,85	0,30	0,58	0,24	1	20.97	1.85	0.55	0.28
4102402	Calça Comprida	1	24.25	2.54	0.55	0.31	1	9,93	1,23	0,73	0,33	1	19.68	1.93	0.57	0.31
4102403	Agasalho	1	46.82	6.41	0.53	0.39	2	32,75	11,35	0,79	0,34	1	36.80	8.05	0.65	0.31
4102404	Saia	1	26.39	1.62	0.47	0.29	1	25,89	0,88	0,43	0,26	1	28.19	1.60	0.46	0.26
4102405	Vestido	2	10.02	3.47	0.93	0.47		5,13	2.51	1,19	0,45	1	10.01	2.71	0.87	0.39
4102408	Camiseta, Blusa	1	7.15	1.60	0.91	0.41	1	7,99	1,82	0,88	0,37	1	12.62	2.17	0.73	0.35
4102410	Lingerie		2.31	1.37	1.55	0.65		2,48	2,65	1,73	0,67		2.70	4.46	1.94	0.74
4103000	Roupa de Criança	1	26.32	0.95	0.43	0.24	1	15,08	1,52	0,62	0,28	1	24.90	2.97	0.56	0.28
4103402	Calça Comprida	1	32.98	3.03	0.49	0.20	2	4,64	0,99	1,04	0,39		5.01	2.05	1.15	0.45
4103405	Agasalho	1	200.09	1.69	0.17	0.15	1	39,86	7,97	0,62	0,35	1	66.54	8.72	0.50	0.29
41034012	Camiseta		1.58	1.01	1.78	0.72	2	3,50	2,02	1,37	0,52		3.07	3.33	1.66	0.67
4103418	Roupa de Bebê	2	6.28	1.64	0.97	0.38	1	14,49	0,83	0,57	0,26		12.80	6.36	1.01	0.41
4200000	Calçados e Acessórios	1	8.21	1.12	0.79	0.40		4,66	2,43	1,24	0,53		3.20	4.82	1.83	0.75
4201000	Calçados e Acessórios	1	8.21	1.12	0.79	0.40		4,66	2,43	1,24	0,53		3.20	4.82	1.83	0.75
4201402	Sapato Homem	1	13.78	0.79	0.58	0.23	2	3,46	1,98	1,37	0,55		1.79	3.29	2.17	0.83
4201403	Sapato Mulher	1	54.10	8.57	0.55	0.32		4,38	2,51	1,29	0,53	1	7.96	2.03	0.91	0.35
4201404	Sapato Criança		2.91	7.06	2.20	0.86		3,14	2,32	1,49	0,55		2.89	5.46	2.01	0.78
4201407	Sandália Mulher	1	79.13	1.48	0.27	0.12	1	55,47	8,98	0,55	0,26	2	37.63	15.33	0.84	0.33
4201415	Bolsa de Mulher		1.54	1.89	2.03	0.81		1,67	2,65	2,12	0,81		1.93	5.48	2.46	1.02
4300000	Jóias, Relógios de Pulso	2	4.70	0.31	0.92	0.44		5,34	7,51	1,66	0,65		7.45	11.02	1.64	0.67
4301000	Jóias, Relógios de Pulso	2	4.70	0.31	0.92	0.44		5,34	7,51	1,66	0,65		7.45	11.02	1.64	0.67
4301402	Jóias		3.22	1.63	1.36	0.57		4,23	4,77	1,59	0,66		5.27	6.16	1.55	0.68
4301414	Relógios de Pulso	2	4.38	0.32	0.95	0.45		5,20	6,77	1,62	0,63		7.48	12.53	1.73	0.68
4400000	Tecidos e Armarinho		5.26	6.86	1.62	0.71		5,30	2,17	1,13	0,45		8.50	4.67	1.11	0.48
4401000	Tecidos e Armarinho		5.26	6.86	1.62	0.71		5,30	2,17	1,13	0,45		8.50	4.67	1.11	0.48
4401401	Tecidos		5.12	8.17	1.75	0.73	2	4,15	2,01	1,25	0,53		7.18	5.40	1.27	0.52
4401402	Artigos de Armarinho		2.42	0.27	1.27	0.54		6,30	2,79	1,10	0,41		6.28	6.53	1.45	0.59

continua

Transporte e Comunicação

Códigos	Descrição	1980 a 1985					1986 a maio 1995					1980 a maio 1995				
		T	Fe	Fm	M7	Q	T	Fe	Fm	M7	Q	T	Fe	Fm	M7	Q
5000000	**Transp. e Comunicacao**	1	10.29	0.74	0.67	0.31		10.33	5.77	1.08	0.52		12.22	8.69	1.16	0.55
5100000	Transporte	1	10.11	0.76	0.68	0.33		10,50	5,69	1,07	0,51		12.34	8.58	1.15	0.57
5101000	Transporte Público	1	8.89	1.71	0.83	0.40		4,77	2,92	1,28	0,54		6.73	2.89	1.08	0.51
5101401	Ônibus Urbano	2	6.08	2.37	1.08	0.61	2	4,54	1,15	1,07	0,47		6.38	2.17	1.03	0.49
5101402	Táxi	2	5.43	0.50	0.88	0.51		3,16	2,32	1,49	0,60		4.97	2.21	1.17	0.47
5101406	Ônibus à Distância	1	18.00	0.86	0.52	0.30		4,32	3,53	1,43	0,56		6.79	2.94	1.08	0.49
5101410	Avião	1	15.03	1.32	0.60	0.28		5,65	6,38	1,52	0,58		5.35	8.06	1.71	0.65
5102000	Veículo Próprio	2	9.93	3.48	0.94	0.44	2	13,89	5,79	0,94	0,40		14.18	9.58	1.12	0.50
5102401	Automóveis Novos		1.63	6.38	2.58	1.05		10,98	5,94	1,06	0,53		9.22	13.00	1.58	0.68
5102404	Emplacamento e Licença		5.89	8.85	1.69	0.73		6,13	7.76	1,57	0,86		6.89	15.70	1.98	0.82
5102405	Seg. Volunt. Veículos	1	7.19	1.02	0.84	0.48		4,21	2,69	1,34	0,53		4.51	3.17	1.35	0.54
5102407	Óleo		1.42	3.40	2.46	1.03		5,13	5,88	1,55	0,67		5.71	8.72	1.70	0.76
5102409	Acessórios e Peças		2.50	0.42	1.29	0.60		8,05	3,37	1,03	0,40		8.76	7.46	1.29	0.54
5102410	Pneus e Câmara de Ar	2	4.04	0.51	1.03	0.50		5,07	3,58	1,32	0,53		5.80	7.33	1.58	0.67
5102411	Conserto de Automóveis	1	11.44	1.86	0.74	0.31		7,09	3,30	1,09	0,43		9.91	5.51	1.09	0.48
5102413	Estacionamento		2.42	6.53	2.34	0.96		6,77	2,21	1,00	0,40		6.50	3.42	1.15	0.46
5102419	Lubrificação e Lavagem		3.65	3.49	1.55	0.68	2	5,56	1,91	1,07	0,47		6.53	3.45	1.15	0.48
5102420	Automóveis Usados	1	36.47	3.55	0.49	0.28		8,46	4,24	1,08	0,45	2	16.77	6.59	0.89	0.41
5104401	Gasolina		1.83	0.33	1.48	0.62		4.21	6.56	1.78	0.70		4.73	8.79	1.88	0.73
5200000	Comunicações		2.83	3.24	1.72	0.69		4,55	5,91	1,65	0,63		4.31	7.92	1.89	0.71
5201000	Comunicações		2.83	3.24	1.72	0.69		4,55	5,91	1,65	0,63		4.31	7.92	1.89	0.71
5201402	Telefone		2.63	3.20	1.78	0.72		4,31	5,62	1,66	0,63		4.27	7.66	1.87	0.71

continua

Saúde e Cuidados Pessoais

Códigos	Descrição	1980 a 1985					1986 a maio 1995					1980 a maio 1995				
		T	Fe	Fm	M7	Q	T	Fe	Fm	M7	Q	T	Fe	Fm	M7	Q
6000000	**Saude e Cuid. Pessoais**		3.00	1.93	1.46	0.65		7.27	6.06	1.32	0.58		6.67	11.41	1.76	0.75
6100000	Prod. Farmac., Óculos,Lentes		3.00	3.07	1.64	0.71		8,11	13,31	1,70	0,68		6.32	16.19	2.10	0.86
6101000	Produtos Farmacêuticos		2.79	3.45	1.76	0.79		8,21	13,66	1,71	0,65		6.26	17.13	2.16	0.88
6101401	Antiinfecciosos		2.36	4.08	2.02	0.85		9,97	14,88	1,61	0,64		8.07	17.85	1.94	0.75
6101402	Analgésicos		3.01	7.81	2.25	1.18		7,73	18,21	2,00	0,75		5.78	20.90	2.26	0.92
6101404	Antigrip. Antitussígenos		2.62	2.99	1.74	0.81		9,75	7,56	1,23	0,53		7.92	13.03	1.71	0.74
6101409	Antiácidos, etc.		2.68	3.17	1.75	0.98		5,89	6,83	1,53	0,60		4.30	11.65	2.21	0.92
6101410	Fort. e Vitam. (exc. B12)		2.44	2.08	1.65	1.04		5,50	9,34	1,78	0,73		3.82	10.95	2.28	0.91
6101412	Antiespasmódicos		2.53	3.27	1.82	0.82		8,98	7,94	1,31	0,60		5.84	12.83	1.97	0.81
6101413	Psicotrop., Anorex.		2.38	4.17	2.02	0.97		9,08	11,52	1,51	0,60		5.94	15.69	2.13	0.84
6101414	Anticoag., Cardiovasculares		2.46	3.84	1.94	0.85		4,39	8,15	1,89	0,72		4.46	10.89	2.11	0.86
6102000	Óculos e Lentes	1	7.44	2.09	0.94	0.43		5,71	5,80	1,46	0,57		6.81	8.44	1.54	0.66
6102401	Óculos e Lentes	1	7.44	2.09	0.94	0.43		5,71	5,80	1,46	0,57		6.81	8.44	1.54	0.66
6200000	Atendimento e Serviços		5.95	3.70	1.23	0.57		6,41	2,30	1,04	0,46		6.97	6.16	1.35	0.59
6201000	Atendimento		2.67	4.13	1.91	0.75	1	7,47	1,36	0,86	0,32		8.46	4.16	1.07	0.40
6201402	Médico		4.63	3.55	1.38	0.59		4,42	2,25	1,25	0,52		6.54	3.87	1.19	0.51
6201403	Dentista		3.34	5.04	1.82	0.79	1	7,84	1,28	0,83	0,36		8.69	4.04	1.05	0.41
6201405	Aparelho Dentário		3.43	2.47	1.45	0.64	1	8,39	2,46	0,93	0,39		7.45	5.20	1.23	0.46
6202000	Serviços Médicos	1	7.68	1.82	0.90	0.41		5,14	4,43	1,40	0,52		4.27	7.64	1.87	0.75
6202403	Exame de Laboratório		6.12	3.14	1.16	0.54		3,48	3,57	1,59	0,63		6.98	5.79	1.32	0.59
6202404	Hospitalização e Cirurgia	1	8.41	0.89	0.76	0.30		3,10	3,97	1,75	0,67		1.92	4.73	2.35	0.85
6202405	Mensalidade de Clínica		6.33	3.55	1.18	0.49		3,45	3,68	1,62	0,61		4.08	7.99	1.95	0.77
6300000	Cuidados Pessoais		3.48	4.41	1.70	0.72		6.74	5,62	1,33	0,51		5.98	11.33	1.85	0.74
6301000	Higiene Pessoal		3.48	4.41	1.70	0.72		6.74	5,62	1,33	0,51		5.98	11.33	1.85	0.74
6301409	Produtos para Pele		2.40	3.18	1.85	0.71		6,43	4,15	1,23	0,53		5.70	11.47	1.90	0.81
6301414	Desodorante, Perfume		3.98	2.89	1.40	0.60		8,77	5,37	1,15	0,56		7.62	10.62	1.60	0.74
6301416	Sabonete		1.45	1.77	2.06	0.87		7,48	7,69	1,42	0,57		7.17	14.56	1.88	0.80
6301417	Papel Higiênico		3.30	3.52	1.63	0.71		4,82	6,47	1,65	0,68		4.04	10.68	2.20	0.90

continua

Despesas Pessoais

Códigos	Descrição	1980 a 1985					1986 a maio 1995					1980 a maio 1995				
		T	Fe	Fm	M7	Q	T	Fe	Fm	M7	Q	T	Fe	Fm	M7	Q
7000000	**Despesas Pessoais**	1	16.20	0.17	0.48	0.27		9.00	3.84	1.01	0.42	2	16.26	6.02	0.88	0.43
7100000	Serviços	2	5.73	1.16	0.96	0.45		6,32	4,01	1,23	0,52		7.70	7.87	1.41	0.57
7101000	Serviços Pessoais	2	5.73	1.16	0.96	0.45		6,32	4,01	1,23	0,52		7.70	7.87	1.41	0.57
7101404	Tinturaria	1	10.05	2.14	0.82	0.39		5,30	4,31	1,37	0,50		7.22	7.69	1.44	0.59
7101408	Barbeiro	2	4.56	1.62	1.14	0.43	2	8,59	3,19	0,98	0,44		9.90	5.78	1.11	0.49
7101409	Cabeleireiro e Manicure	1	17.58	1.00	0.53	0.28	1	9,96	2,54	0,86	0,31	2	12.73	4.49	0.90	0.39
7101410	Empregado Doméstico	2	4.32	1.18	1.11	0.47		5,38	7,30	1,64	0,68		5.69	12.09	1.95	0.75
7200000	Recreação e Fumo	1	14.05	0.59	0.56	0.40		8,37	4,06	1,07	0,43	1	16.91	4.84	0.80	0.42
7201000	Recreação	1	54.51	1.33	0.32	0.15		5,61	3,00	1,19	0,49	1	17.82	2.36	0.63	0.35
7201401	Cinema	2	4.78	1.39	1.08	0.58		1,11	4,40	2,77	1,02		1.99	5.09	2.36	0.92
7201406	Clubes	1	69.60	1.31	0.28	0.18		6,70	5,39	1,31	0,54	1	19.02	4.31	0.72	0.31
7201408	Discos e Fitas		1.37	1.08	1.93	0.79		4,12	3,08	1,40	0,55		4.15	7.19	1.86	0.73
7201413	Acessórios para Fotografia		2.17	1.71	1.67	0.72		6,22	7,30	1,52	0,61		6.96	12.40	1.78	0.78
7201423	Brinquedos	1	39.71	5.03	0.53	0.27	2	4,89	1,23	1,05	0,45	1	8.54	1.51	0.82	0.44
7202000	Fumo	2	4.53	0.66	1.00	0.67	2	8,90	3,37	0,98	0,41		10.98	5.45	1.03	0.45
7202441	Cigarros	2	4.42	0.64	1.00	0.72	2	7,91	3,04	1,01	0,42		10.54	4.78	1.01	0.45
7300000	Educação e Leitura	1	27.41	1.40	0.45	0.26		6,83	3,55	1,14	0,47		10.98	5.39	1.03	0.42
7301000	Educação	1	25.32	1.59	0.48	0.25		7.58	3,63	1,09	0,45		10.55	4.87	1.01	0.39
7301403	Artigos de Papelaria		3.13	4.38	1.79	0.65		5,55	4,17	1,33	0,56		7.96	8.27	1.41	0.59
7301406	Cursos Formais	1	17.21	3.60	0.72	0.44		3,35	2,87	1,53	0,61		4.26	4.07	1.50	0.59
7301407	Cursos Diversos	1	42.52	1.86	0.38	0.24		9,19	4,52	1,06	0,41	2	11.40	7.78	0.97	0.49
7302000	Leitura	2	6.83	1.83	0.96	0.42		5,37	2,53	1,16	0,52	2	8.47	3.22	0.99	0.40
7302401	Jornal Diário	1	7.49	1.30	0.85	0.36	1	7,27	2,26	0,97	0,43	1	12.22	3.38	0.84	0.40
7302403	Revistas não Técnicas		1.85	2.96	2.07	0.81		5,13	4,96	1,46	0,57		6.53	7.42	1.50	0.63
7302405	Livro (Bolso, Broch.)		2.20	3.64	2.02	0.82		1,16	1,38	2,19	0,86		1.02	2.80	2.74	1.03

conclusão

ANEXO 4

ANÁLISE QUALITATIVA DOS SUBITENS IDENTIFICADOS COMO SAZONAIS PELO PROGRAMA X11-ARIMA

SUBITENS ACEITOS APÓS A ANÁLISE
SÃO PAULO

SUBITEM	JUSTIFICATIVA	PERÍODO	JUSTIFICATIVA
Farinha de Mandioca	Principal insumo - Produto agrícola c/entressafra	86/95	Período em que o programa permite medir a componente sazonal com mais segurança
Batata Inglesa	Produtos Agrícolas c/ entressafra	86/95	Painel peso POF (1)
Chuchu			
Pepino			
Vagem			
Cebola			
Cenoura			
Alface			
Chicória			
Couve			
Couve-Flor			
Repolho			
Agrião			
Brócolis			
Abacaxi			
Banan Prata			
Laranja Baía			
Laranja Lima			
Laranja Pera			
Limão			
Maçã			
Pera			
Tangerina			
Uva			
Melão		89/95	Produto a partir da POF
Contra Filé	Produtos com entressafra (2)	80/95	Período em que o programa permite medir a componente sazonal com mais segurança
Chã de Dentro			
Alcatra			
Patinho			
Peixe Sardinha	Apresenta periodicidade de pesca (desova, fecundação e fatores climáticos)	86/95	Alteração do Painel de produtos com a POF
Camarão			
Peixe Pescada			
Ovo de Galinha	Baixa produção de ovos no verão	86/95	Período em que o programa permite medir a componente sazonal com mais segurança
Refrigerador	Sazonal de verão	86/95	Período em que o programa permite medir a componente sazonal com mais segurança
Agasalho (homem)	Produtos do grupo Vestuário com sazonalidade marcante (verão/inverno)	86/95	Alteração dos métodos de imputação a partir da POF
Saia			
Agasalho (criança)			
Roupa de Bebê			
Sandália de Mulher			
Brinquedos	Promoções Natalinas e dia da Criança	80/95	Período em que o programa permite medir a componente sazonal com mais segurança

(1) - Pesquisas de Orçamentos Familiares - 1988
(2) - As carnes vem diminuindo a amplitude das variações sazonais

SUBITENS NÃO ACEITOS APÓS A ANÁLISE

Subitens	Justificativas
Banana D'água	As estatísticas de 86/94 não permitem isolar a componente sazonal O período 80/94 mistura painel ENDEF (3) com POF
Mortadela Desinfetante Jornal Diário	Sazonalidade não esperada e não explicada
Móveis para Quarto Máquina de Costura Fogão	Amostras de locais e produtos revisadas
Cabelereiro e Manicure	Sazonalidade não esperada e afetada pela revisão do painel de produtos em 89

(3) - Pesquisa Estudo Nacional de Despesa Familiar - 1974

ANEXO 5

TABELA 2
SUBITENS SAZONAIS POR REGIÃO METROPOLITANA
VESTUÁRIO

SUBITENS	RJ	SP	BH	PA	RE	BR	BE	FT	SL	CT	GO	C/PESO	SAZONAL
CALÇA COMPRIDA (H)				X								11	1
TERNO				X								8	1
AGASALHO (H)		X	X	X								5	3
SHORT, CALÇÃO, BERMUDA				X								8	1
CAMISA (H)				X						X		11	2
CALÇA COMPRIDA (M)				X						x		11	2
AGASALHO (M)			X	X						X		4	3
SAIA (M)	X	X	X	X						X		11	5
VESTIDO (M)	X			X	X				X			11	4
CAMISETA, BLUSA (M)				X						X		11	2
ROUPA DE BANHO					X		X					3	2
UNIFORME							X					10	1
CALÇA COMPRIDA (C)				X								11	1
AGASALHO (C)		X		X						X		5	3
CAMISA (C)				X								11	1
CAMISETA (C)										X		10	1
ROUPA DE BEBÊ		X								x		7	2
SAPATO DE MULHER				X						X		11	2
SANDÁLIA DE MULHER		X	X									11	2

SUBITENS SAZONAIS POR REGIÃO METROPOLITANA

TUBER. RAÍZES E LEGUMES

SUBITENS	RJ	SP	BH	PA	RE	BR	BE	FT	SL	CT	GO	C/PESO	SAZONAL
BATATA DOCE				X								1	1
BATATA INGLESA	X	X	X	X	X	X	X	X	X	X	X	11	11
INHAME					X							1	1
MANDIOCA(AIPIM)				X		X					X	4	3
ABÓBORA			X		X				X			6	3
ABOBRINHA		X				X						3	2
CHUCHU	X	X	X		X	X			X			7	6
PEPINO		X								X		3	2
PIMENTÃO	X		X		X	X	X	X	X	X		9	8
QUIABO	X								X		X	4	3
TOMATE	X		X		X	X	X	X	X		X	11	8
VAGEM	X	X	X							X	X	6	5
BETERRABA			X			X		X		X	X	5	5
CEBOLA	X	X	X	X	X	X	X	X	X	X		11	10
CENOURA	X	X	X	X	X	X	X	X	X	X	X	11	11

SUBITENS SAZONAIS POR REGIÃO METROPOLITANA
HORTALIÇAS E VERDURAS

SUBITENS	RJ	SP	BH	PA	RE	BR	BE	FT	SL	CT	GO	C/PESO	SAZONAL
ALFACE	X	X	X	X	X	X	X		X	X	X	10	10
CHICÓRIA		X										1	1
COENTRO					X			X	X			3	3
COUVE	X	X	X	X			X				X	6	6
COUVE FLOR	X	X	X	X		X				X		6	6
ESPINAFRE	X											1	1
REPOLHO	X	X	X	X	X	X	X	X	X	X	X	11	11
CHEIRO VERDE	X			X	X	X	X					8	5
AGRIÃO	X	X										2	2
ALMEIRÃO			X									1	1
BRÓCOLIS		X										1	1
CARURU							X					1	1

SUBITENS SAZONAIS POR REGIÃO METROPOLITANA
FRUTAS

SUBITENS	RJ	SP	BH	PA	RE	BR	BE	FT	SL	CT	GO	C/PESO	SAZONAL
BANANA DA TERRA					X							2	1
ABACAXI	X	X	X	X	X			X	X	X	X	10	9
ABACATE							X					1	1
BANANA D'ÁGUA										X		7	1
BANANA MAÇÃ										X		3	1
BANANA PRATA	X	X	X		X	X	X	X	X			9	8
BANANA BAÍA		X		X								2	2
LARANJA LIMA	X	X										2	2
LARANJA SELETA	X											3	1
LIMÃO	X	X	X	X			X		X			6	6
MAÇÃ	X	X	X	X	X	X	X	X	X	X		11	10
MAMÃO					X		X				X	11	3
MARACUJÁ					X				X			3	2
MELANCIA					X	X				X		8	3
MELÃO	X	X				X				X		6	4
PÊRA	X	X	X			X					X	5	5
TANGERINA		X	X	X						X		4	4
UVA	X	X	X		X	X	X		X		X	8	8
LARANJA PÊRA	X	X	X	X	X	X	X	X	X	X		11	10
BANANA MARMELO											X	1	1

SUBITENS SAZONAIS POR REGIÃO METROPOLITANA
GÊNEROS ALIMENTÍCIOS, ARTIGOS DE PAPELARIA, ELETRODOMÉSTICOS

SUBITENS	RJ	SP	BH	PA	RE	BR	BE	FT	SL	CT	GO	C/PESO	SAZONAL
OVO DE GALINHA	X	X	X	X		X	X	X	X	X	X	11	10
ALHO						X					X	9	2
COLORAU					X							3	1
PIMENTA DO REINO									X			1	1
REFRIGERADOR		X			X							11	2
CONDICIONADOR DE AR					X							3	1
BRINQUEDOS		X	X									11	2
CADERNOS					X		X					11	2
LIVROS DIDÁTICOS					X				X			7	2
FEIJÃO MULATINHO									X			3	1
FEIJÃO MACASSAR								X				1	1
FARINHA DE MANDIOCA	X	X			X				X			8	4
LEITE PASTEURIZADO											X	11	1
QUEIJO MINAS											X	5	1

SUBITENS SAZONAIS POR REGIÃO METROPOLITANA
CARNES E PEIXES

SUBITENS	RJ	SP	BH	PA	RE	BR	BE	FT	SL	CT	GO	C/PESO	SAZONAL
ARNE DE PORCO					X							11	1
ONTRA FILÉ		X										11	1
HÃ DE DENTRO		X	X	X								11	3
LCATRA		X		X								11	2
ATINHO		X	X	X								9	3
AGARTO PLANO			X									2	1
ILÉ DE PEIXE									X	X		7	2
EIXE ANCHOVA	X				X							2	2
EIXE CORVINA	X				X	X			X			4	4
EIXE CAVALINHA					X							2	1
EIXE SARDINHA	X	X							X			5	3
AMARÃO		X		X								6	2
EIXE VERMELHO									X			1	1
EIXE PARGO								X				1	1
EIXE PESCADA	X	X					X			X		5	4
EIXE CIOBA					X							2	1
EIXE PIRAMUTABA							X					2	1
EIXE SURUBIM						X					X	3	2
AMARÃO SECO SALGADO							X					1	1

ANEXO 6

Tabela 3 - Principais Estatísticas fornecidas pelo programa X11- ARIMA

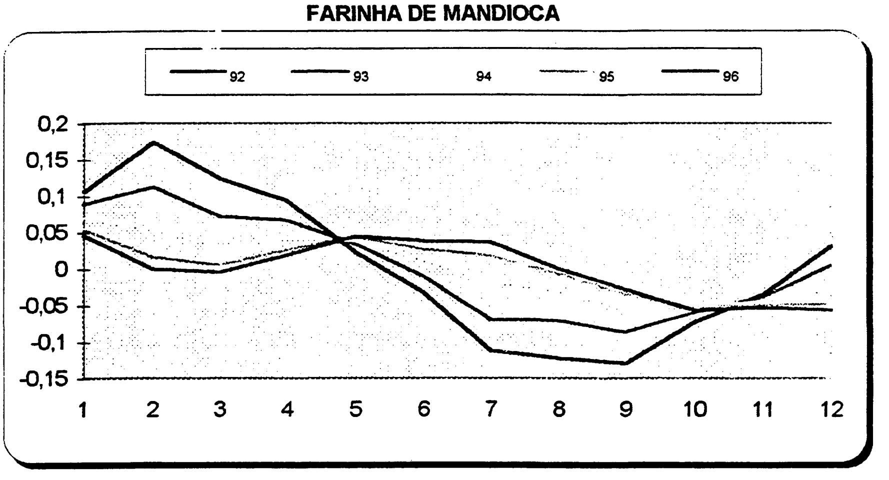
Região Metropolitana de São Paulo

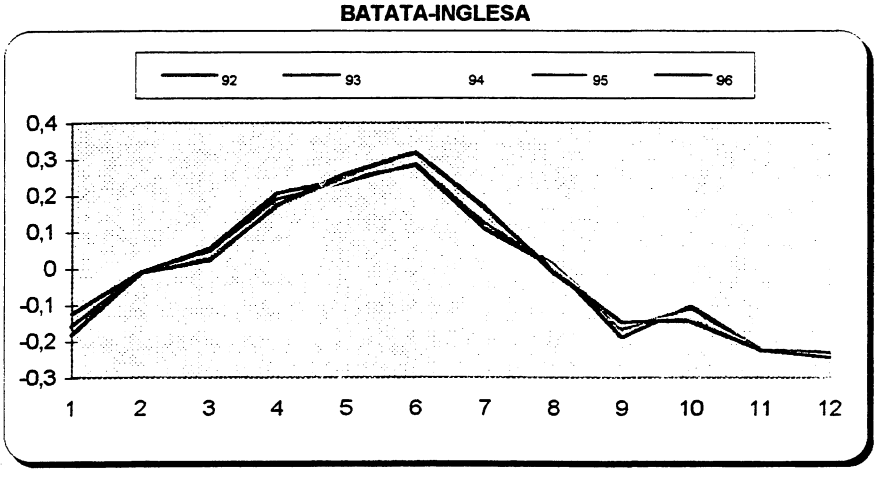
Código	Descrição	Modelo Arima	Erro Médio Padrão (%)	X^2 - Prob (%)	I/S	Fe	Fm	M7	Q	MCD
1102423	FARINHA DE MANDIOCA	(0,1,2)(0,1,1)	3,91	5,62	3,69	15,78	5,48	0,86	0,37	1
1103403	BATATA INGLESA	(0,1,2,)(0,1,1)	4,07	11,26	4,13	17,5	2,31	0,61	0,38	1
1103420	ABOBORINHA	(0,1,1)(0,1,1)	2,79	83,27	5,42	11,46	1,5	0,71	0,43	1
1103421	CHUCHU	(0,1,1)(0,1,1)	3,18	79,6	4,26	17,04	1,59	0,59	0,37	1
1103425	PEPINO	(0,1,1)(0,1,1)	3,08	36,62	5,74	9,89	1,34	0,75	0,45	1
1103429	VAGEM	(0,1,1)(0,1,1)	3,77	97,42	4,72	12,97	2,6	0,76	0,48	1
1103443	CEBOLA	(0,1,1)(0,1,1)	3,44	36,69	6,18	18,46	1,73	0,57	0,39	1
1103444	CENOURA	(0,1,1)(0,1,1)	3,52	13,56	5,11	20,82	4,33	0,69	0,34	1
1105401	ALFACE	(0,1,1)(0,1,1)	3,3	79,06	3,69	30,64	1,54	0,44	0,32	1
1105403	CHICÓRIA	(0,1,1)(0,1,1)	3,61	97,27	4,2	40,2	1,63	0,38	0,31	1
1105405	COUVE	(0,1,1)(0,1,1)	3,51	77,34	6,15	60,71	1,68	0,31	0,25	1
1105406	COUVE-FLOR	(0,1,1)(0,1,1)	3,25	88,91	4	38,91	1,75	0,4	0,23	1
1105410	REPOLHO	(0,1,1)(0,1,1)	3,14	59,23	4,53	59,39	2,27	0,34	0,25	1
1105413	AGRIÃO	(0,1,1)(0,1,1)	3,13	27,58	4,45	51,12	2,65	0,38	0,21	1
1105419	BRÓCOLIS	(0,1,1)(0,1,1)	2,7	29,41	4,43	64,84	2,16	0,32	0,19	1
1106403	ABACAXI	(0,1,1)(0,1,1)	2,73	48,13	3,52	18,97	1,53	0,55	0,27	1
1106408	BANANA-PRATA	(0,1,2)(0,1,1)	2,97	39,21	3,37	20,98	2,65	0,6	0,28	1
1106411	LARANJA-BAIA	(0,1,1)(0,1,1)	3,01	64,91	3,42	10,96	2,19	0,79	0,38	1
1106412	LARANJA-LIMA	(0,1,1)(0,1,1)	3,14	52,71	3,75	92,2	0,68	0,22	0,14	1
1106415	LIMÃO	(0,1,2)(0,1,1)	2,63	47,56	3,42	34,77	6,44	0,62	0,3	1
1106417	MAÇÃ	(0,1,1)(0,1,1)	2,87	23,32	3,56	22,35	4,46	0,67	0,32	1
1106423	PERA	(0,1,1)(01,1)	2,51	37,58	4,44	31,98	2,45	0,47	0,26	1
1106427	TANGERINA	(0,1,1)(0,1,1)	3,08	31,16	3,8	20,89	0,65	0,46	0,28	1
1106428	UVA	(0,1,1)(0,1,1)	3,25	10,4	5,95	34,36	1,41	0,4	0,3	1
1106439	LARANJA-PERA	(0,1,2)(0,1,1)	2,63	47,56	3,42	34,77	6,44	0,61	0,3	1
1107484	CONTRA-FILÉ	(0,1,2)(0,1,1)	3,06	10,37	5,32	11,04	1,7	0,74	0,4	1
1107487	CHÃ-DE-DENTRO	(0,1,2)(0,1,1)	3,13	7,25	5,17	11,1	1,91	0,76	0,4	1
1107488	ALCATRA	(0,1,2)(0,1,1)	3,13	5,88	5,36	10,7	1,76	0,78	0,42	1
1107489	PATINHO	(0,1,2)(0,1,1)	3,18	7,45	5,23	11,04	1,85	0,75	0,42	1
1108412	PEIXE SARDINHA	(0,1,1)(0,1,1)	2,68	48,99	3,02	43,13	0,68	0,32	0,21	1
1108413	CAMARÃO	(0,1,1)(0,1,1)	2,39	42,53	3,76	11,91	3,31	0,84	0,4	1
1108438	PEIXE PESCADA	(0,1,2)(0,1,1)	1,79	70,47	3,1	17,78	4,67	0,77	0,29	1
1110444	OVO DE GALINHA	(0,1,1)(0,1,1)	4,74	52,8	6,18	7,67	1,7	0,89	0,45	1
3201401	REFRIGERADOR	(0,2,2)(0,1,1)	2,1	9,1	3,49	8,79	1,51	0,81	0,36	1
4101405	AGASALHO (H)	(0,1,2)(0,1,1)	1,97	6,71	2,77	20,52	7,68	0,86	0,48	1
4102404	SAIA	(0,1,2)(0,1,1)	2,58	5,46	6,05	23,65	3,38	0,6	0,27	1
4103405	AGASALHO (C)	(0,1,1)(0,1,1)	1,75	14,45	2,95	32,27	8,7	0,72	0,37	1
4103418	ROUPA DE BEBÊ	(0,1,2)(0,1,1)	2,21	16,97	3,5	11,86	1,41	0,69	0,31	1
4201407	SANDÁLIA DE MULHER	(0,1,2)(0,1,1)	2,36	8,94	3,97	47,15	6,36	0,53	0,3	1
7201423	BRINQUEDOS	(0,2,2)(0,1,1)	2,22	14,58	4	8,27	1,42	0,82	0,41	1

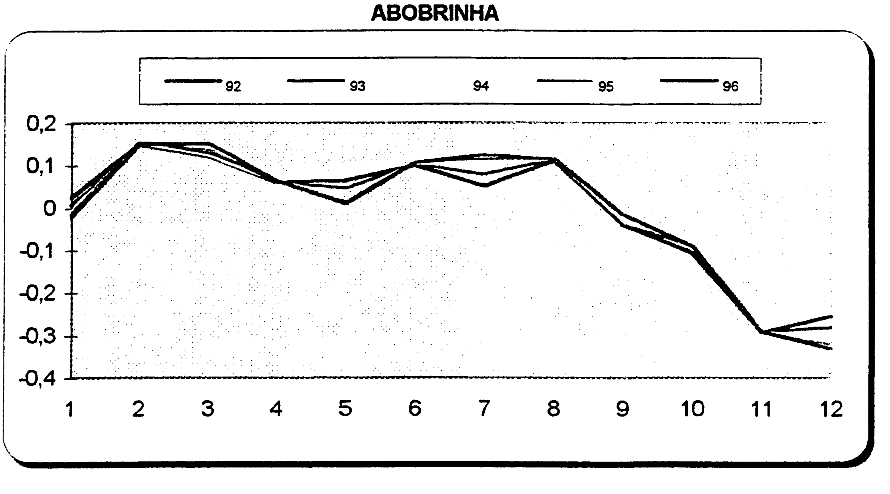
ANEXO 7

FATORES SAZONAIS DE SÃO PAULO

NÚMERO ÍNDICE - BASE MAR86=100



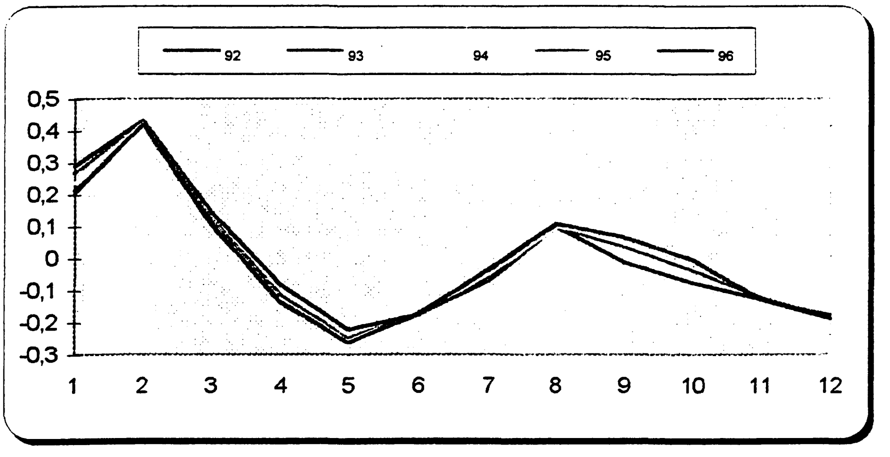




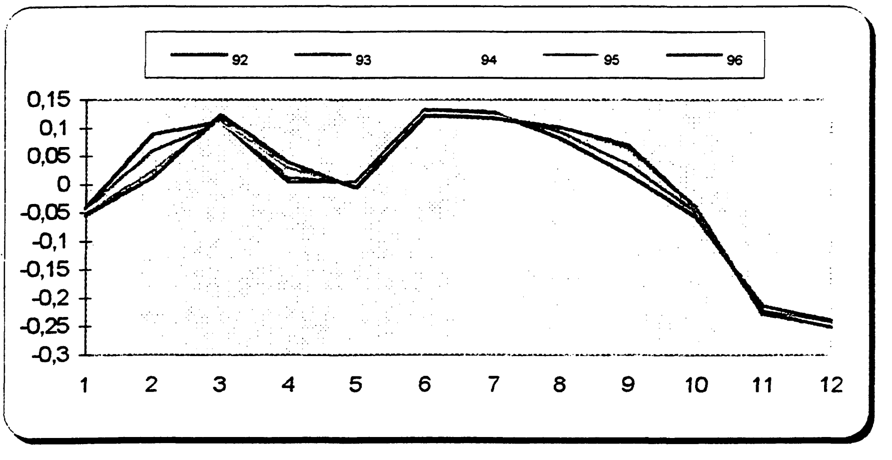
FATORES SAZONAIS DE SÃO PAULO

NÚMERO ÍNDICE - BASE MAR86=100

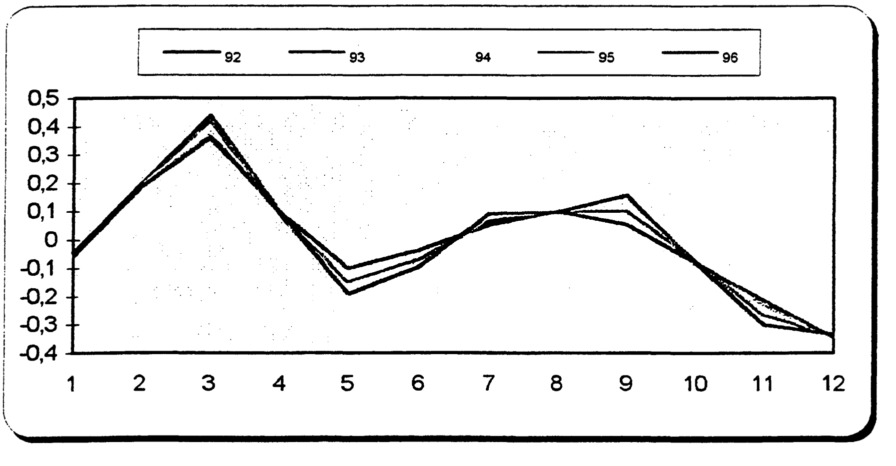
CHUCHU



PEPINO

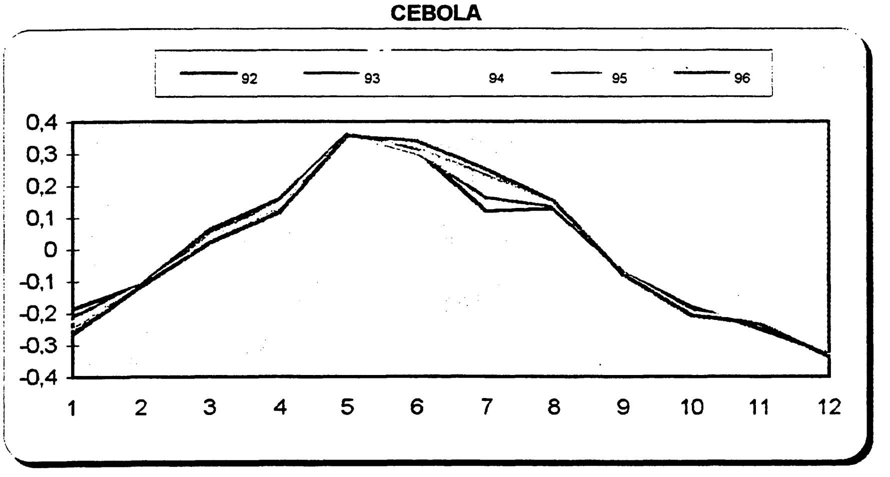


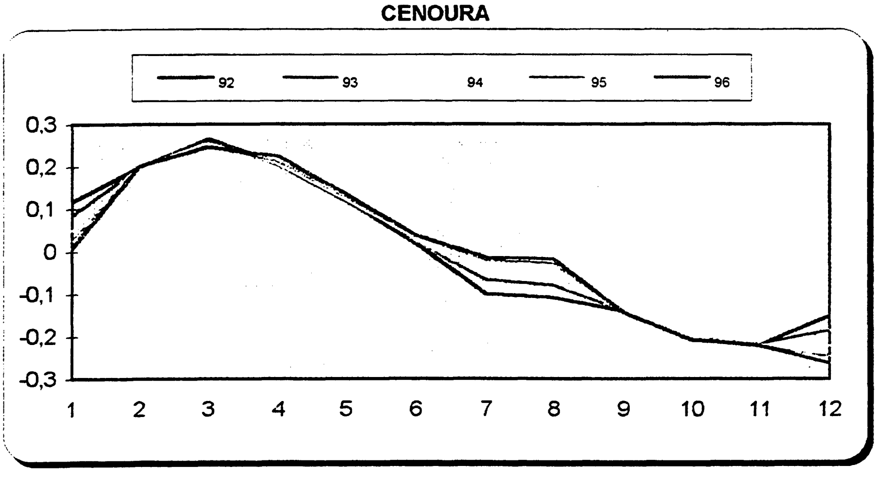
VAGEM

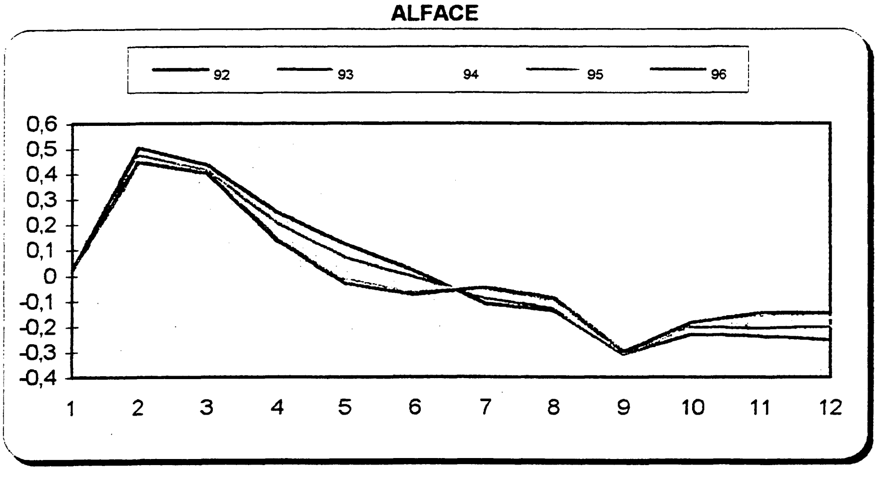


FATORES SAZONAIS DE SÃO PAULO

NÚMERO ÍNDICE - BASE MAR86=100



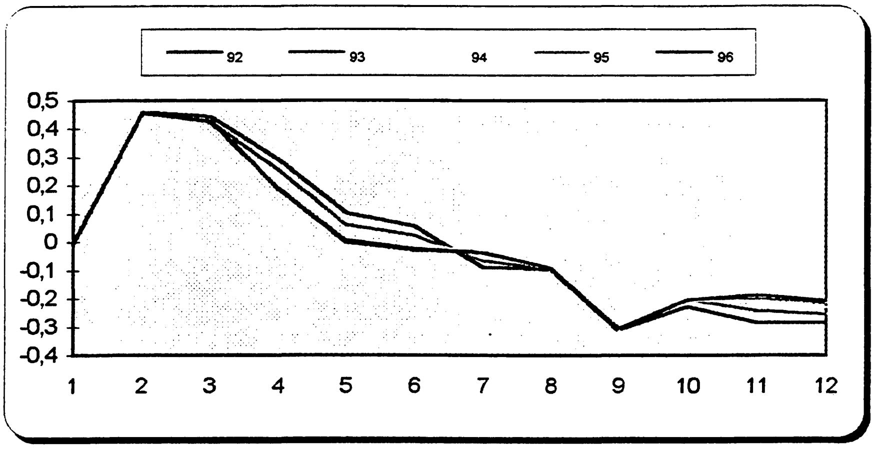




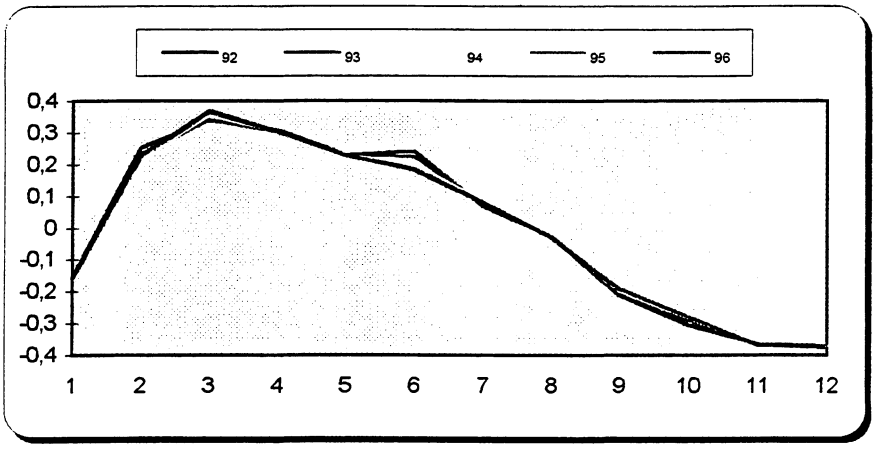
FATORES SAZONAIS DE SÃO PAULO

NÚMERO ÍNDICE - BASE MAR86=100

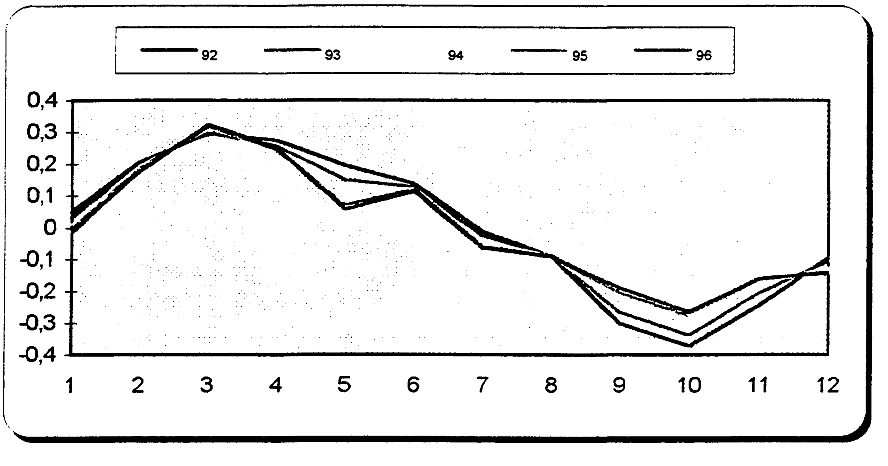
CHICÓRIA



COUVE



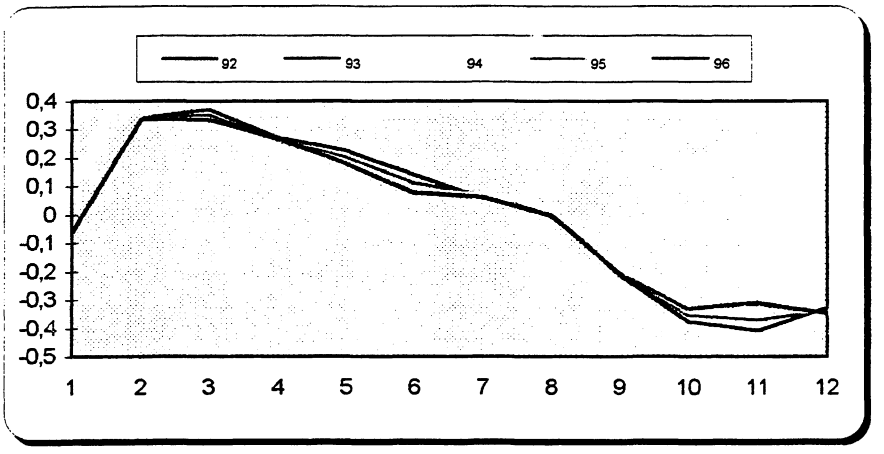
COUVE-FLOR



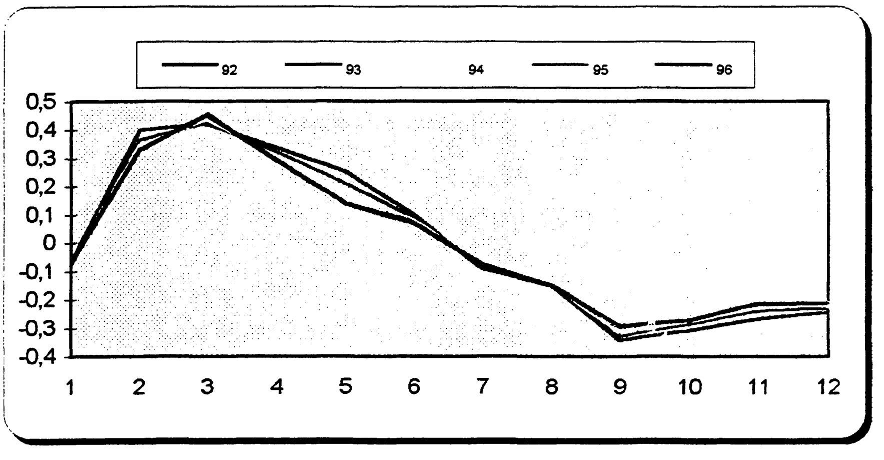
FATORES SAZONAIS DE SÃO PAULO

NÚMERO ÍNDICE - BASE MAR86=100

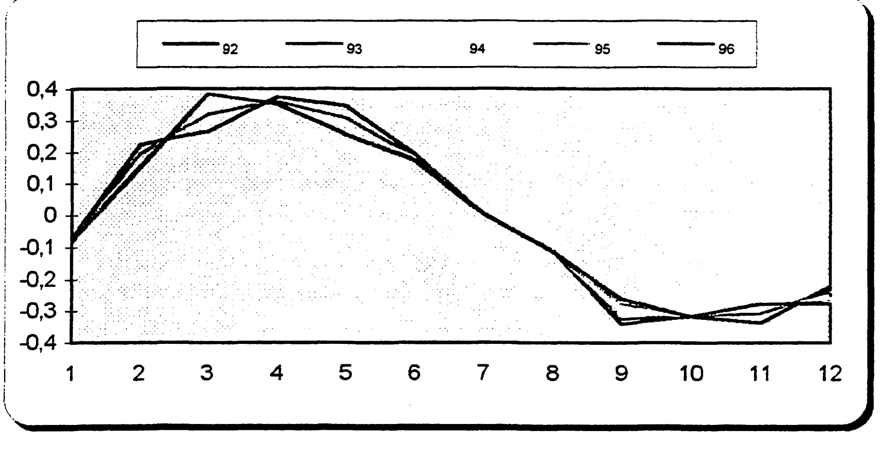
REPOLHO



AGRIÃO

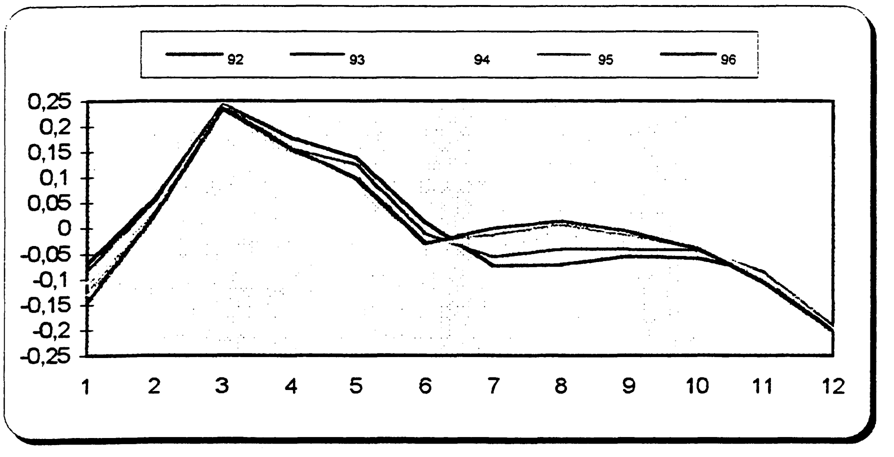


BROCOLIS

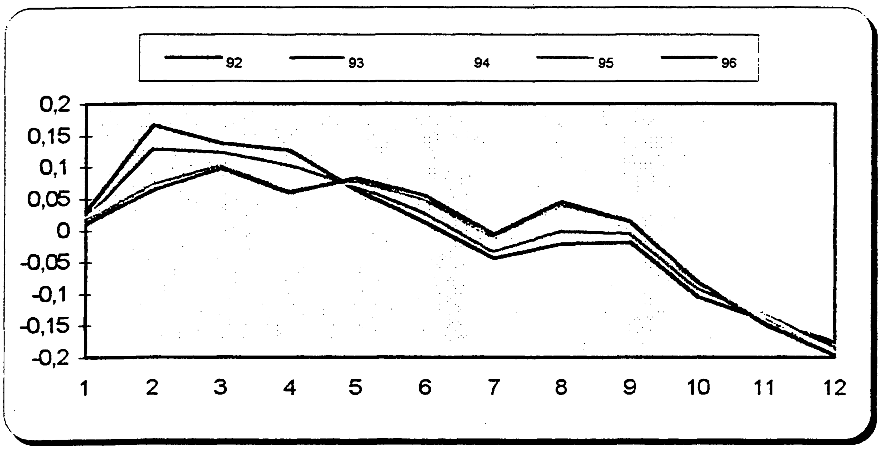


FATORES SAZONAIS DE SÃO PAULO
NÚMERO ÍNDICE - BASE MAR86=100

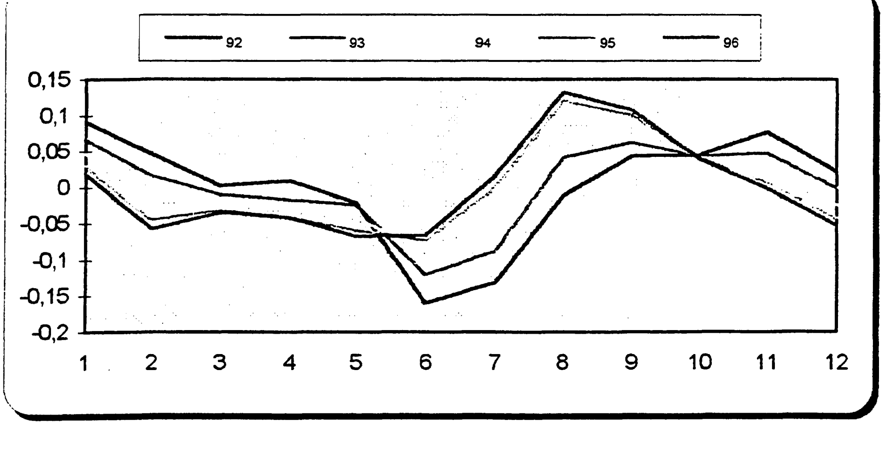
ABACAXI



BANANA-PRATA



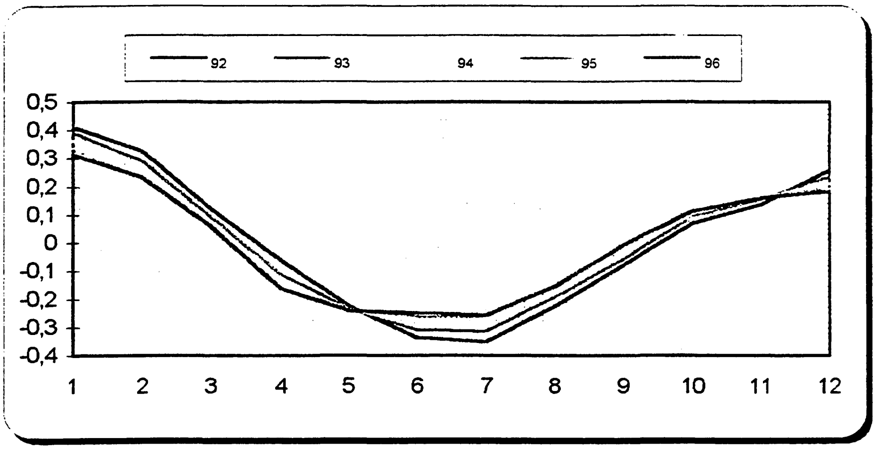
LARANJA-BAÍA



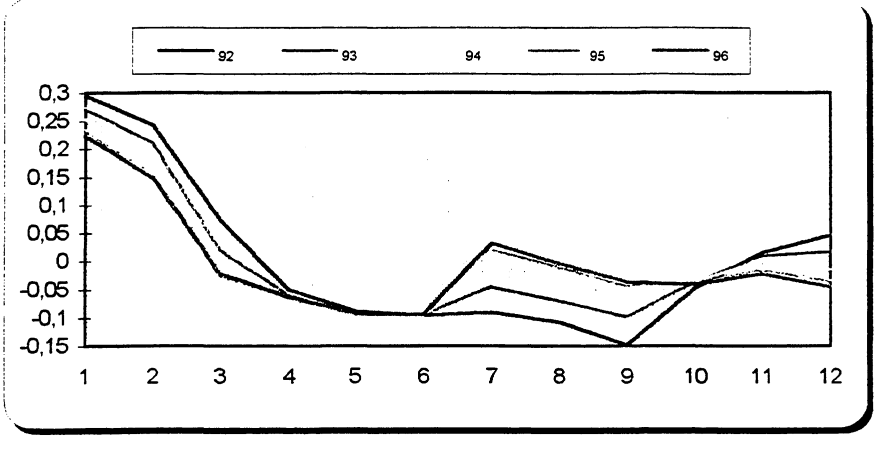
FATORES SAZONAIS DE SÃO PAULO

NÚMERO ÍNDICE - BASE MAR86=100

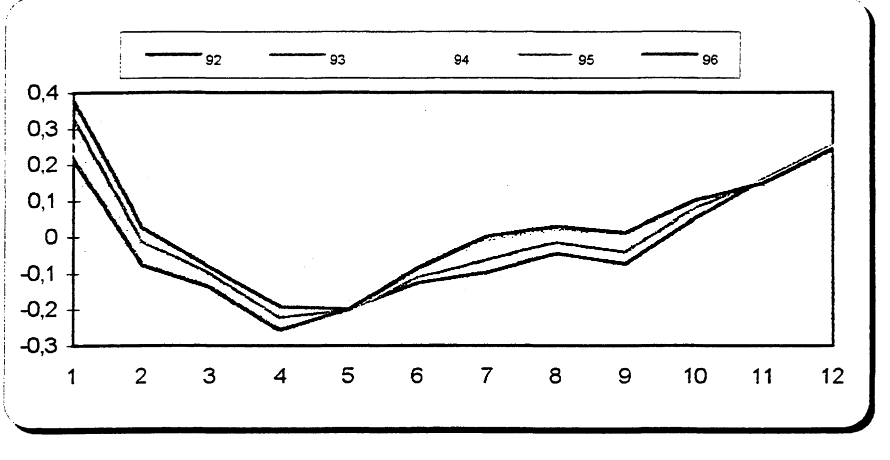
LARANJA-LIMA



MAÇÃ

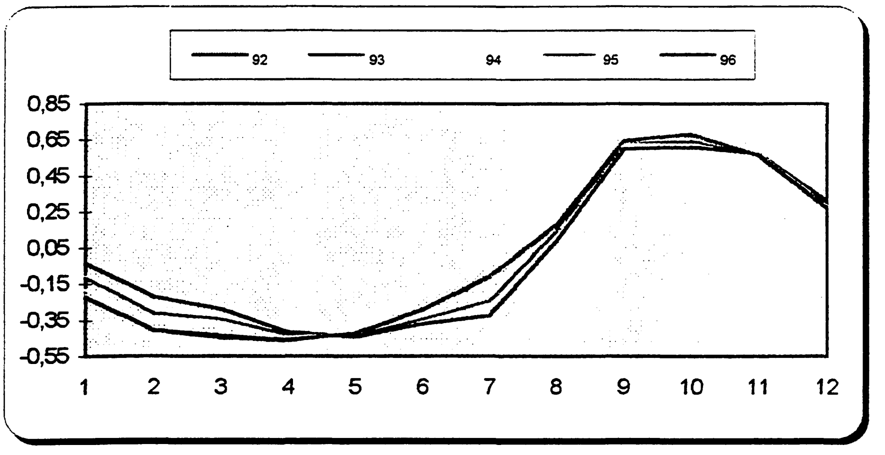


PERA

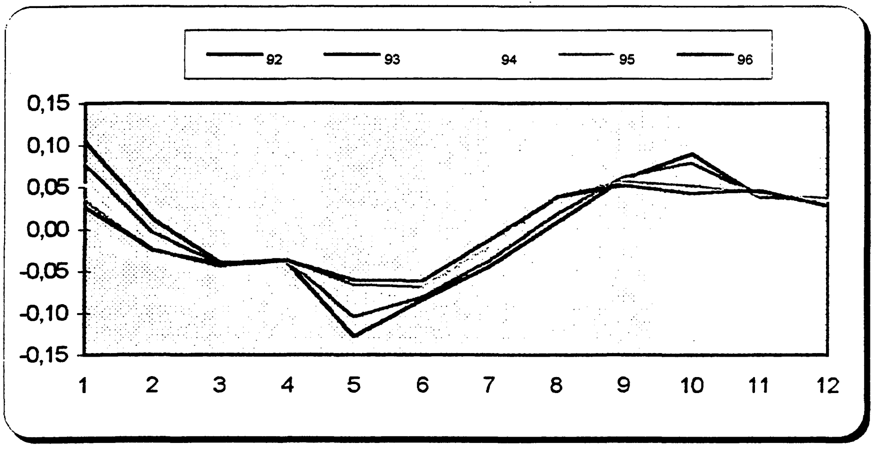


FATORES SAZONAIS DE SÃO PAULO
NÚMERO ÍNDICE - BASE MAR86=100

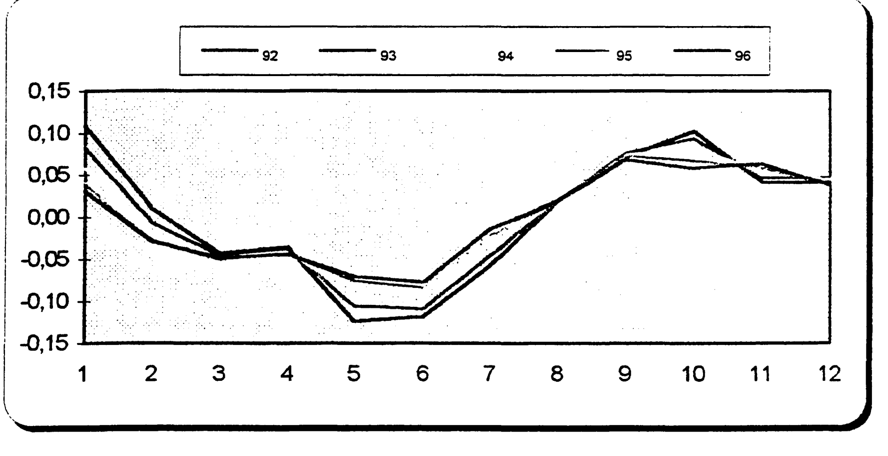
LIMÃO



CONTRAFILÉ



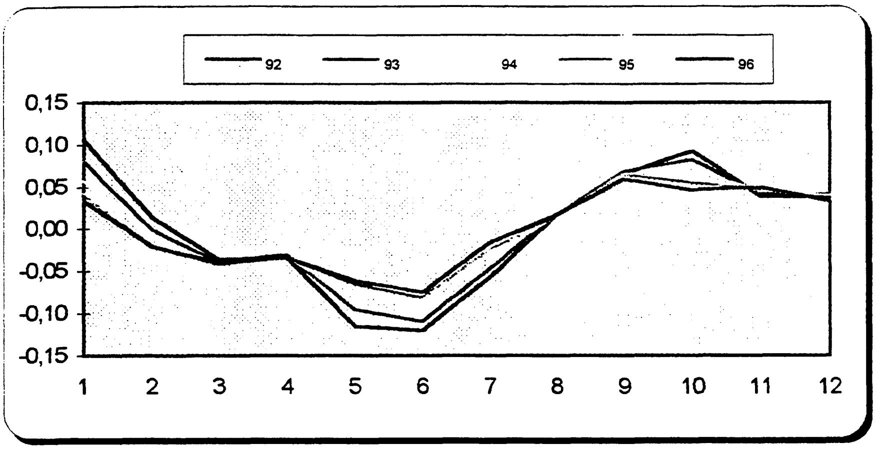
CHÃ-DE-DENTR0



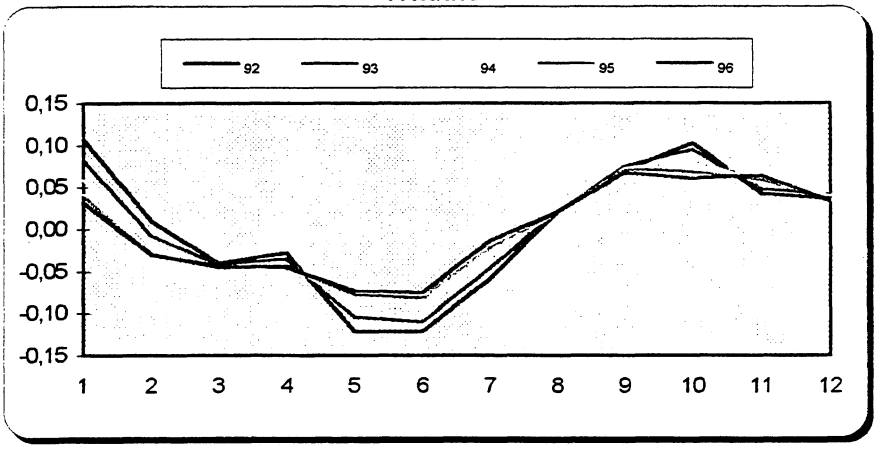
FATORES SAZONAIS DE SÃO PAULO

NÚMERO ÍNDICE - BASE MAR86=100

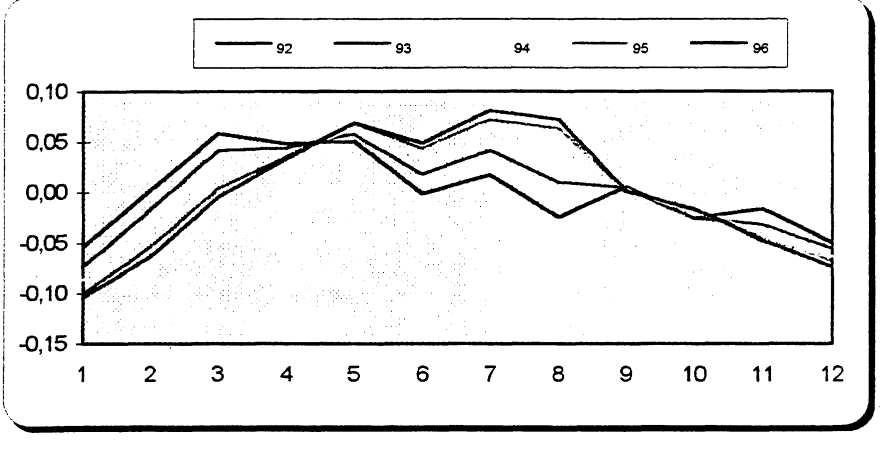
ALCATRA



PATINHO

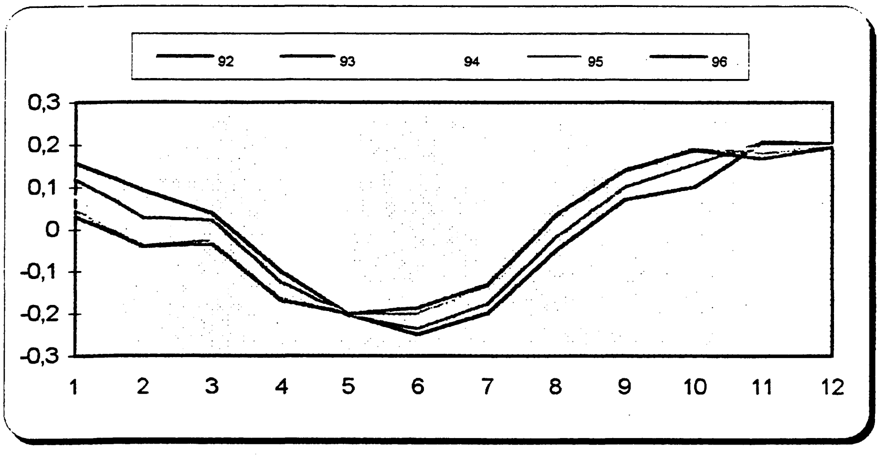


BRINQUEDOS

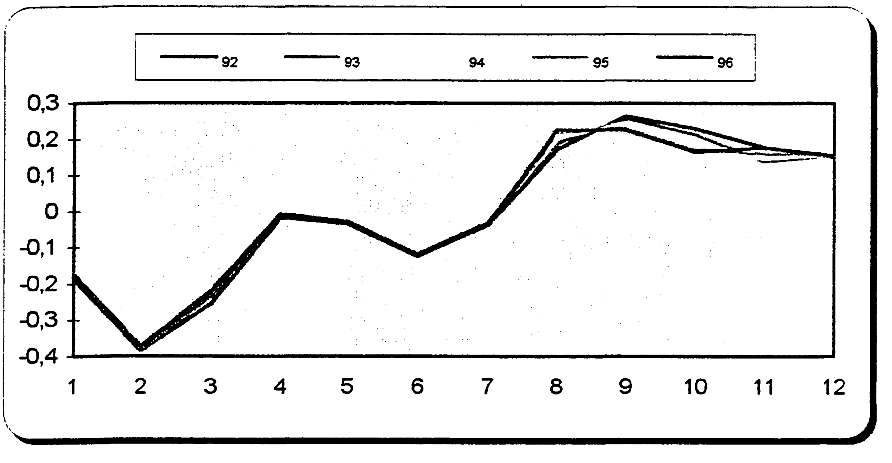


FATORES SAZONAIS DE SÃO PAULO
NÚMERO ÍNDICE - BASE MAR86=100

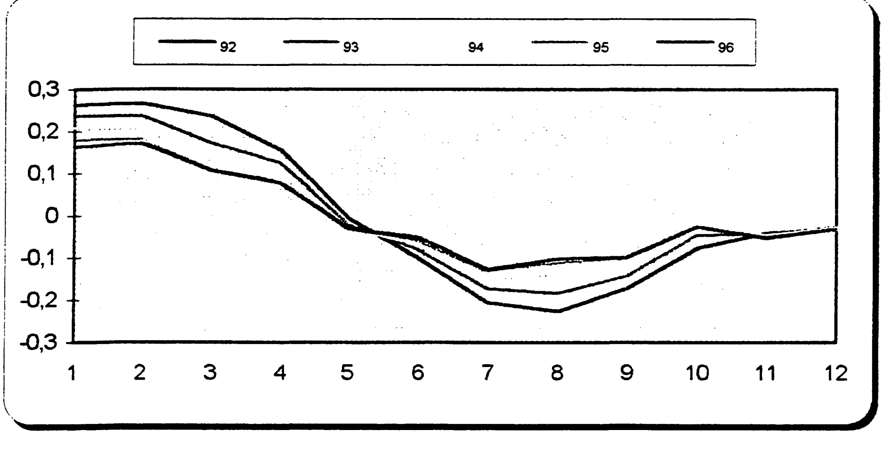
TANGERINA



UVA



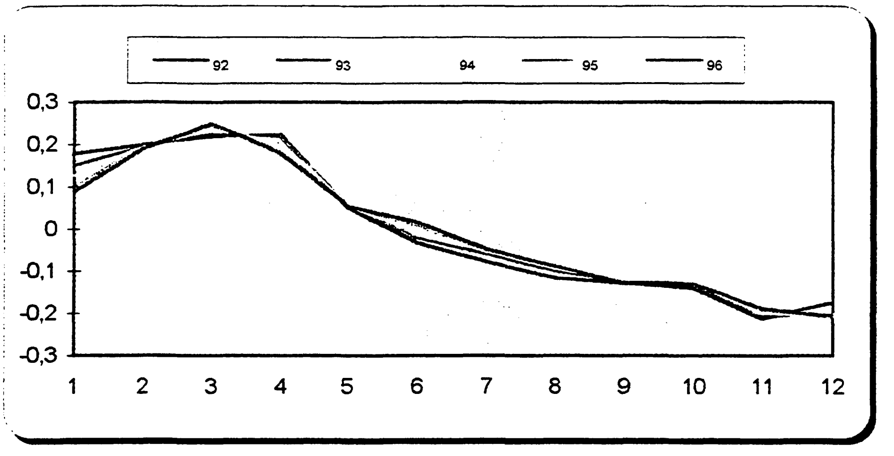
LARANJA-PERA



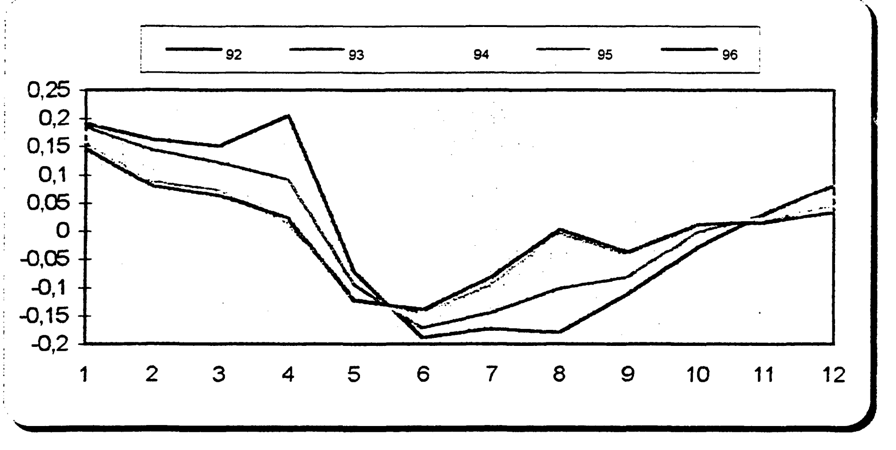
FATORES SAZONAIS DE SÃO PAULO

NÚMERO ÍNDICE - BASE MAR86=100

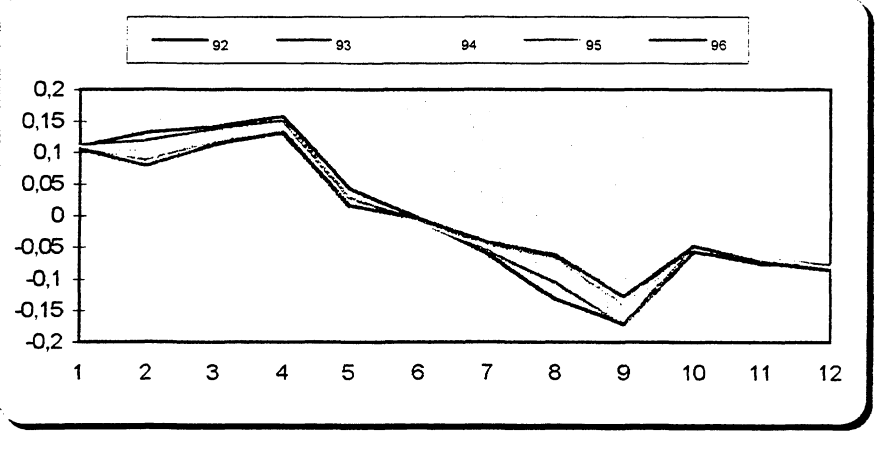
PEIXE SARDINHA



CAMARÃO



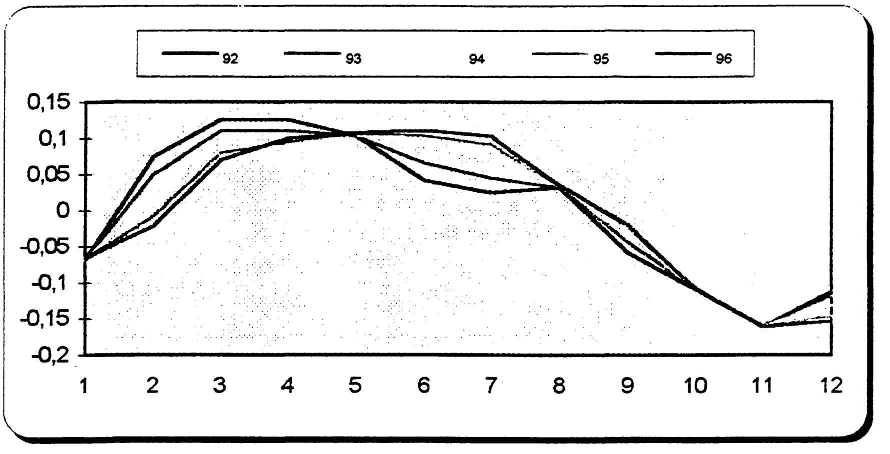
PEIXE PESCADA



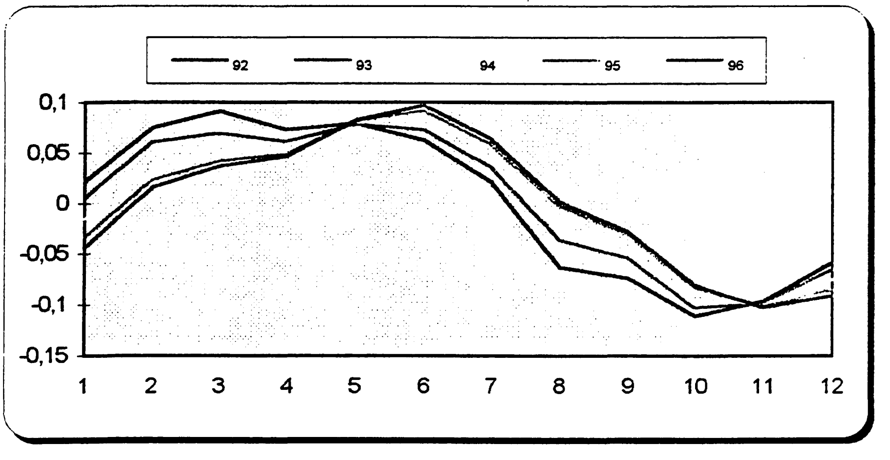
FATORES SAZONAIS DE SÃO PAULO

NÚMERO ÍNDICE - BASE MAR86=100

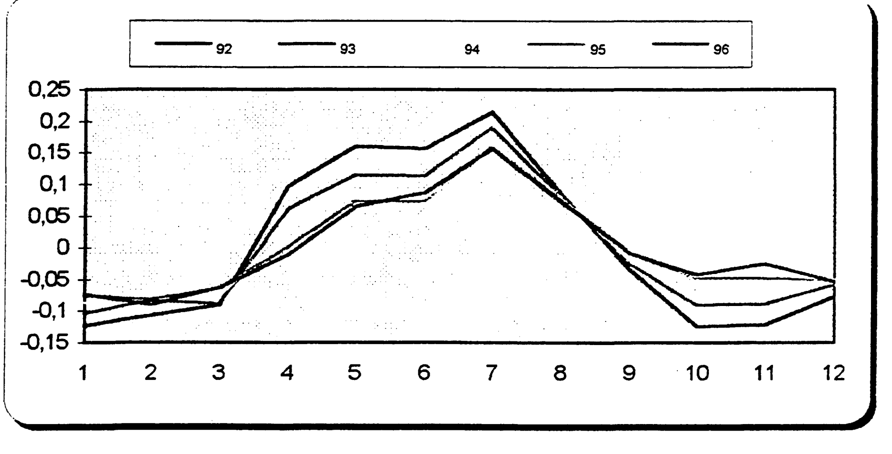
OVO DE GALINHA



REFRIGERADOR



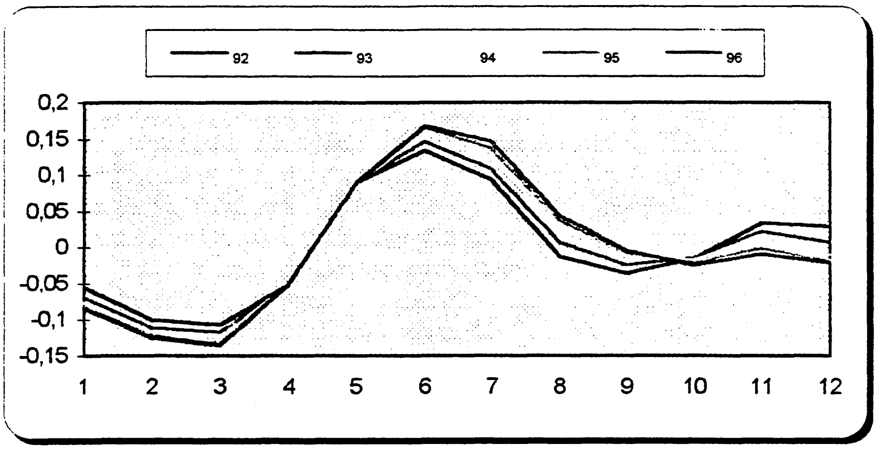
AGASALHO DE HOMEM



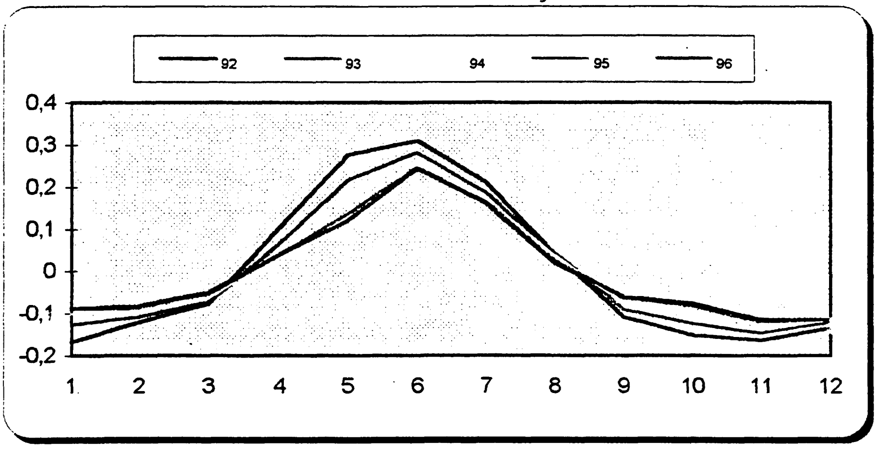
FATORES SAZONAIS DE SÃO PAULO

NÚMERO ÍNDICE - BASE MAR86=100

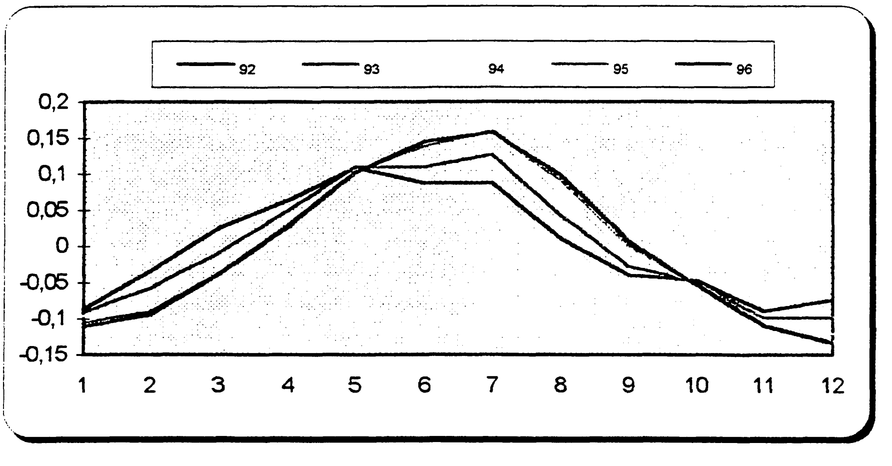
SAIA



AGASALHO DE CRIANÇA



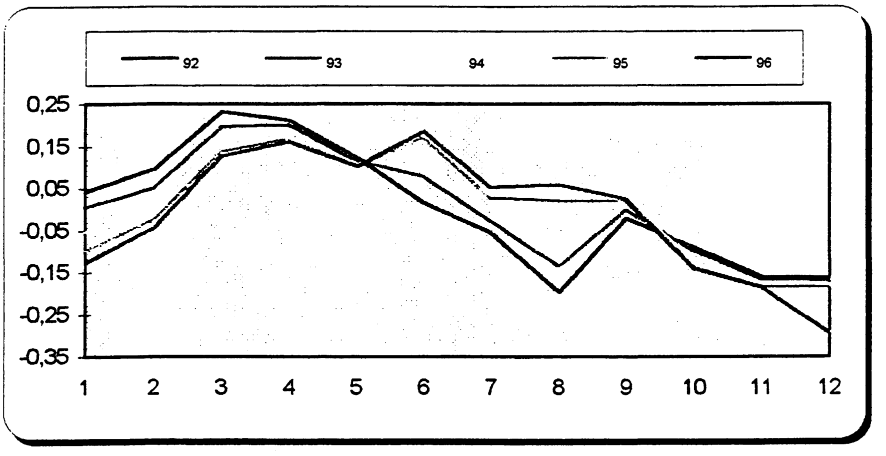
ROUPA DE BEBE



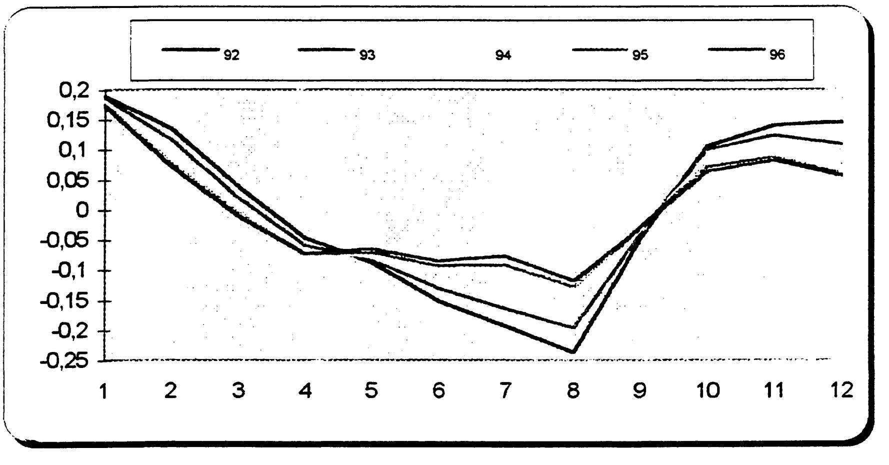
FATORES SAZONAIS DE SÃO PAULO

NÚMERO ÍNDICE - BASE MAI89=100

MELÃO



SANDÁLIA DE MULHER



ANEXO 8

TABELA 4

REGIÃO METROPOLITANA DE SÃO PAULO - 1995

FATORES SAZONAIS PARA AS VARIAÇÕES MENSAIS

DESCR.	JAN	FEV	MAR	ABR	MAIO	JUN	JUL	AGO	SET	OUT	NOV	DEZ
FAR. MANDIOCA	108,87	96,39	98,92	102,17	101,93	98,54	99,19	97,65	97,39	98,18	100,53	100,25
BATATA INGL.	105,26	117,71	103,70	114,44	108,19	106,17	85,26	83,91	86,13	100,64	91,11	97,47
ABOBRINHA	132,48	117,87	97,57	92,29	94,55	108,74	100,25	99,32	87,12	92,13	81,19	96,49
CHUCHU	146,36	120,73	71,61	76,84	86,66	107,50	112,14	113,21	94,63	90,94	86,85	92,53
PEPINO	121,03	107,23	109,18	91,36	96,01	113,13	99,00	97,98	95,68	89,71	82,30	97,39
VAGEM	130,54	125,01	118,14	74,31	74,01	107,72	118,01	98,91	103,55	78,21	84,18	87,41
CEBOLA	107,54	113,48	113,85	110,13	124,92	95,57	90,99	91,34	78,68	87,51	96,86	89,11
CENOURA	127,32	119,11	104,59	96,01	91,57	91,02	94,02	98,86	88,95	93,40	98,13	97,02
ALFACE	118,99	150,55	93,80	76,01	83,84	93,26	99,45	93,64	80,21	110,53	100,64	99,10
CHICÓRIA	122,85	156,21	94,98	77,90	81,46	95,13	96,36	92,58	79,00	108,79	98,63	96,10
COUVE	122,25	145,45	113,28	92,79	91,58	95,16	88,57	88,27	83,86	90,56	90,24	98,00
COUVE-FLOR	112,36	120,15	113,79	92,18	83,32	104,38	83,41	96,10	88,85	92,30	111,03	102,11
REPOLHO	130,81	147,48	98,40	91,75	90,66	89,21	97,12	91,76	80,68	86,56	100,15	95,42
AGRIÃO	114,39	148,30	110,75	84,55	84,79	91,88	84,79	91,44	84,63	101,27	104,13	99,07
BRÓCOLIS	118,34	125,60	124,77	95,83	90,41	91,20	83,23	88,24	84,24	95,13	102,97	100,05
ABACAXI	106,42	117,10	122,11	91,72	94,51	87,33	101,12	101,54	97,39	97,29	93,55	89,91
BANANA PRATA	122,90	105,60	102,71	95,85	101,20	96,83	93,91	104,99	97,08	90,79	93,84	94,29
LARANJA BAÍA	105,53	92,91	101,31	99,00	98,21	98,68	107,42	112,97	97,99	94,15	96,72	95,10
LARANJA LIMA	113,11	90,27	83,39	79,60	92,02	97,23	98,97	110,78	114,75	112,73	104,42	102,71
LIMÃO	59,41	80,29	94,37	94,79	99,82	110,31	115,50	131,51	154,61	100,09	87,03	72,28
MAÇÃ	128,08	92,36	83,03	95,80	96,55	99,69	111,92	96,40	96,45	100,13	101,92	97,67
MELÃO	107,81	107,67	116,98	102,84	93,28	106,82	86,25	99,10	99,46	85,53	94,66	99,61
PÊRA	97,83	74,08	92,92	87,84	105,08	110,91	107,81	102,47	98,09	109,14	104,49	109,35
TANGERINA	85,90	91,39	100,95	87,02	95,68	99,80	105,79	117,56	110,72	105,68	98,31	101,20
UVA	70,95	80,91	114,82	122,61	96,81	90,04	107,79	126,48	100,73	92,92	97,56	98,40
LARANJA PÊRA	123,21	99,92	92,63	97,13	88,96	97,02	92,63	101,73	100,76	107,34	97,18	101,51
CONTRAFILÉ	100,18	94,29	97,99	100,70	97,15	99,69	104,83	106,15	101,90	99,41	99,16	98,56
CHÃ DE DENTRO	99,63	93,42	97,77	100,69	96,79	99,12	106,29	104,34	105,44	99,41	99,06	98,03
ALCATRA	100,23	94,00	97,75	100,89	96,88	98,38	105,76	104,36	104,76	99,09	99,10	98,80
PATINHO	100,25	93,22	98,30	100,16	96,69	99,50	106,04	104,48	105,12	99,67	99,06	97,51
PEIXE SARDINHA	135,33	109,12	104,53	94,15	86,83	94,90	93,70	95,65	95,64	99,28	93,21	97,65
CAMARÃO	110,23	93,28	98,43	94,32	87,38	97,21	105,04	109,35	96,29	105,27	100,51	102,68
PEIXE PESCADA	120,58	97,82	102,53	101,49	88,63	97,54	95,98	97,57	92,63	109,56	97,30	98,37
OVO	106,78	106,00	109,16	101,45	101,19	99,76	98,74	94,54	94,49	91,84	94,78	101,28
REFRIGERADOR	104,35	105,90	101,86	100,72	103,26	101,11	96,75	94,07	97,25	94,89	98,18	101,66
AGASALHO H.	97,56	99,19	101,61	106,54	107,43	99,79	108,97	91,83	91,74	95,99	99,91	99,44
SAIA	92,90	95,98	98,72	108,40	114,96	107,71	97,03	90,35	95,44	98,46	101,90	98,16
AGASALHO C.	102,28	99,86	103,05	109,42	109,95	111,35	91,82	86,83	91,19	97,76	96,19	100,31
ROUPA DE BEBÊ	101,46	101,51	105,27	106,93	107,60	103,49	102,02	93,37	91,39	94,71	94,55	97,70
SANDÁLIA M.	110,27	90,80	91,75	93,34	100,26	97,64	100,16	96,26	110,27	110,47	101,60	97,17

ANEXO 9

SÉRIES DE ÍNDICES DE PREÇOS - IPCA

ORIGINAIS X AJUSTADAS - BRASIL E REGIÕES METROPOLITANAS

GERAL - PERÍODO: JAN95 A DEZ95

ÁREAS GEOGRÁFICAS	JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ	ACUM
BRASIL	**1,41**	**0,95**	**1,46**	**2,23**	**2,57**	**2,28**	**2,51**	**1,27**	**1,18**	**1,50**	**1,64**	**1,74**	**22,81**
	1,70	1,02	1,55	2,43	2,67	2,26	2,36	0,99	0,99	1,41	1,47	1,56	22,41
RIO JANEIRO	**1,29**	**0,94**	**1,47**	**1,97**	**2,57**	**3,11**	**1,74**	**1,31**	**1,41**	**1,34**	**1,95**	**1,74**	**22,93**
	1,89	1,02	1,29	2,15	2,59	3,08	1,47	0,99	1,19	1,20	1,85	1,60	22,30
PORTO ALEGRE	**2,10**	**1,05**	**1,83**	**1,84**	**2,57**	**2,39**	**2,22**	**1,60**	**0,50**	**1,56**	**1,50**	**1,63**	**22,87**
	1,72	0,30	2,20	2,94	3,29	2,68	2,02	0,78	0,60	1,47	1,23	1,20	22,40
BELO HORIZONTE	**1,60**	**0,56**	**1,19**	**2,40**	**3,64**	**1,61**	**1,84**	**1,64**	**1,57**	**1,86**	**1,64**	**1,67**	**23,38**
	1,91	0,82	1,32	2,49	3,71	1,51	1,65	1,44	1,36	1,63	1,59	1,55	23,09
RECIFE	1,44	0,92	1,61	1,47	2,56	1,64	3,56	1,63	0,88	1,27	1,54	1,97	22,49
	2,02	1,16	2,08	1,97	2,69	1,59	2,95	1,17	0,40	1,04	1,22	1,86	22,08
SÃO PAULO	**1,25**	**1,22**	**1,65**	**2,78**	**2,30**	**2,29**	**3,14**	**0,92**	**1,08**	**1,51**	**1,72**	**2,06**	**24,23**
	1,51	1,37	1,56	2,69	2,29	2,29	3,18	0,83	0,90	1,53	1,54	1,90	23,83
BRASÍLIA	**1,57**	**0,61**	**0,71**	**1,89**	**3,09**	**2,36**	**2,33**	**1,22**	**1,03**	**1,99**	**1,66**	**1,17**	**21,46**
	1,91	0,81	0,78	1,99	3,08	2,24	2,04	1,06	0,99	1,90	1,62	1,17	21,42
BELÉM	**0,97**	**1,20**	**1,59**	**1,24**	**2,43**	**1,61**	**1,85**	**1,28**	**2,16**	**1,23**	**0,98**	**1,29**	**19,35**
	1,55	1,43	1,74	1,49	2,41	1,37	1,58	0,97	1,89	1,02	0,87	1,26	19,06
FORTALEZA	**1,47**	**-0,05**	**1,03**	**1,65**	**2,69**	**2,14**	**2,18**	**1,36**	**1,18**	**1,34**	**1,48**	**0,62**	**18,46**
	1,96	0,23	1,42	1,77	2,66	2,14	1,82	0,86	0,90	1,08	1,39	0,69	18,27
SALVADOR	**1,03**	**0,45**	**0,87**	**1,66**	**2,59**	**3,05**	**3,01**	**1,52**	**1,04**	**1,05**	**1,76**	**1,34**	**21,14**
	1,59	0,71	1,07	1,84	2,63	3,04	2,69	1,14	0,71	1,02	1,56	1,13	20,86
CURITIBA	**1,83**	**0,97**	**1,53**	**2,27**	**2,02**	**1,53**	**2,09**	**1,53**	**1,36**	**1,61**	**1,28**	**1,41**	**21,25**
	1,95	0,74	2,19	2,71	2,36	1,48	1,87	1,04	1,26	1,50	0,98	1,05	20,87
GOIÂNIA	**1,54**	**0,57**	**0,99**	**1,81**	**2,96**	**2,20**	**1,22**	**1,66**	**1,12**	**1,56**	**1,49**	**1,84**	**20,67**
	1,62	0,67	1,33	2,27	3,13	2,09	1,16	1,45	0,92	1,33	1,21	1,57	20,42

[ ] AJUSTADO   [ ] ORIGINAL

SÉRIES DE ÍNDICES DE PREÇOS - IPCA

ORIGINAIS X AJUSTADAS - BRASIL E REGIÕES METROPOLITANAS

ALIMENTAÇÃO - PERÍODO: JAN95 A DEZ95

ÁREAS GEOGRÁFICAS	JAN	FEV	MAR	ABR	MAIO	JUN	JUL	AGO	SET	OUT	NOV	DEZ	ACUM
BRASIL	**-0,44**	**-0,56**	**1,40**	**1,98**	**0,82**	**0,01**	**1,46**	**0,92**	**0,41**	**0,49**	**1,56**	**1,34**	**9,77**
	0,73	-0,06	1,44	1,99	0,51	-0,25	0,98	0,57	-0,03	0,32	1,07	0,86	8,42
RIO DE JANEIRO	**-0,49**	**0,00**	**1,39**	**1,36**	**0,66**	**-0,71**	**0,80**	**1,19**	**1,05**	**0,89**	**1,85**	**1,63**	**10,02**
	1,32	0,40	1,13	1,72	0,23	-0,87	-0,08	0,36	0,37	0,42	1,53	1,12	7,89
PORTO ALEGRE	**0,89**	**-0,59**	**1,29**	**2,92**	**2,08**	**0,82**	**1,42**	**0,83**	**-0,92**	**0,96**	**0,67**	**0,87**	**11,77**
	1,12	-0,07	1,52	2,90	2,03	0,58	1,40	1,05	-0,89	0,69	0,10	0,30	11,21
BELO HORIZONTE	**-1,31**	**-1,91**	**0,99**	**1,73**	**1,93**	**-0,47**	**0,00**	**0,82**	**0,79**	**1,34**	**1,15**	**1,48**	**6,65**
	-0,09	-0,92	1,06	1,60	1,70	-0,97	-0,80	0,60	0,25	0,80	0,95	1,24	5,50
RECIFE	**-0,48**	**0,18**	**1,86**	**1,72**	**1,93**	**-0,42**	**1,40**	**2,96**	**0,06**	**0,17**	**1,38**	**1,07**	**12,42**
	0,89	0,72	3,04	3,13	2,01	-0,62	0,01	1,96	-1,21	-0,35	0,56	0,69	11,27
SÃO PAULO	**-0,49**	**-0,21**	**1,74**	**2,26**	**-0,12**	**0,36**	**2,26**	**0,20**	**0,12**	**-0,13**	**2,01**	**1,39**	**9,74**
	0,47	0,14	1,13	1,55	-0,84	0,20	2,31	0,40	-0,04	0,12	1,53	0,91	8,13
BRASÍLIA	**-0,87**	**-0,42**	**-0,17**	**1,41**	**2,32**	**0,28**	**0,03**	**1,06**	**-0,58**	**0,41**	**0,81**	**1,16**	**5,53**
	0,43	0,38	0,13	1,79	2,28	-0,16	-1,15	0,37	-0,76	-0,01	0,57	1,18	5,11
BELÉM	**-0,90**	**-0,47**	**2,71**	**0,82**	**1,47**	**-1,54**	**1,21**	**1,10**	**1,37**	**0,65**	**2,11**	**0,79**	**9,64**
	0,51	0,06	3,09	1,50	1,39	-2,26	0,43	0,22	0,73	0,27	1,79	0,79	8,77
FORTALEZA	**0,35**	**-1,93**	**0,54**	**1,26**	**0,52**	**1,01**	**2,11**	**1,63**	**1,47**	**1,48**	**1,21**	**0,18**	**10,22**
	1,70	-1,14	1,62	1,62	0,49	1,05	1,05	0,18	0,63	0,70	0,93	0,37	9,56
SALVADOR	**-0,91**	**-0,81**	**1,05**	**2,07**	**2,09**	**1,00**	**2,25**	**0,98**	**0,39**	**0,27**	**0,90**	**1,69**	**11,47**
	0,80	0,12	1,71	2,56	1,98	0,90	1,33	-0,14	-0,61	0,24	0,30	0,97	10,60
CURITIBA	**0,82**	**-1,29**	**1,24**	**3,08**	**0,38**	**-0,14**	**1,76**	**1,55**	**0,88**	**1,08**	**0,56**	**1,45**	**11,91**
	2,05	-1,07	2,31	2,93	0,41	-0,76	1,18	1,17	0,77	0,94	-0,10	0,63	10,89
GOIÂNIA	**-1,22**	**-1,34**	**-0,27**	**1,30**	**0,33**	**-1,35**	**-1,37**	**2,61**	**1,05**	**0,85**	**2,45**	**1,65**	**4,67**
	-0,90	-0,94	1,23	3,34	1,14	-1,70	-1,58	1,59	0,10	-0,21	1,08	0,33	3,41

AJUSTADA (shaded)   ORIGINAL

SÉRIES DE ÍNDICES DE PREÇOS - IPCA
ORIGINAIS X AJUSTADAS - BRASIL E REGIÕES METROPOLITANAS
ARTIGOS DE RESIDÊNCIA - PERÍODO: JAN95 A DEZ95

ÁREAS GEOGRÁFICAS	JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ	ACUM
RASIL	**3,05**	**2,83**	**1,57**	**2,35**	**1,95**	**1,54**	**0,64**	**-0,43**	**-0,16**	**-0,25**	**-0,08**	**0,92**	**14,76**
	3,15	2,99	1,65	2,36	2,05	1,56	0,55	-0,57	-0,22	-0,36	-0,14	0,93	14,77
ECIFE	**2,83**	**2,00**	**1,80**	**0,16**	**0,96**	**1,49**	**2,80**	**-0,72**	**1,22**	**0,07**	**0,89**	**1,62**	**16,14**
	3,20	3,25	2,67	0,29	1,84	1,53	1,95	-1,63	0,80	-0,45	0,30	1,23	15,92
ÃO PAULO	**3,77**	**3,54**	**1,20**	**2,86**	**2,92**	**1,76**	**0,32**	**-0,87**	**0,11**	**0,07**	**0,09**	**0,80**	**17,74**
	3,98	3,81	1,31	2,88	3,06	1,79	0,17	-1,14	-0,02	-0,18	0,00	0,88	17,67

AJUSTADA (shaded) ORIGINAL (unshaded)

SÉRIES DE ÍNDICES DE PREÇOS - IPCA

ORIGINAIS X AJUSTADAS - BRASIL E REGIÕES METROPOLITANAS

VESTUÁRIO - PERÍODO:JAN95 A DEZ95

ÁREAS GEOGRÁFICAS	JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ	ACUM
RASIL	**1,52**	**-0,09**	**-0,94**	**1,12**	**1,20**	**0,63**	**1,10**	**-1,15**	**0,32**	**0,94**	**0,42**	**0,02**	**5,17**
	0,99	-1,19	-0,69	2,62	2,64	1,21	0,91	-2,62	0,11	0,82	0,20	-0,34	4,63
IO DE JANEIRO	**2,18**	**-1,01**	**-0,42**	**1,08**	**1,10**	**2,05**	**0,53**	**-1,00**	**0,83**	**-0,17**	**0,98**	**-0,07**	**6,19**
	2,03	-1,53	-1,39	1,74	2,71	2,29	0,47	-1,72	0,59	-0,25	0,93	0,00	5,90
ORTO ALEGRE	**4,21**	**0,82**	**-0,46**	**-1,63**	**-1,75**	**-0,97**	**0,89**	**2,34**	**-0,08**	**0,70**	**1,72**	**1,75**	**7,63**
	1,14	-5,40	1,52	6,19	3,35	1,63	-0,44	-4,23	0,61	0,56	0,75	-0,53	4,74
ELO HORIZONTE	**2,14**	**2,01**	**-3,38**	**2,22**	**2,37**	**0,52**	**0,06**	**-1,42**	**-1,38**	**1,78**	**0,43**	**1,04**	**6,39**
	1,96	1,15	-3,51	3,29	3,37	0,61	0,06	-2,49	-1,55	1,82	0,41	1,08	6,11
ECIFE	**3,95**	**-0,61**	**1,21**	**1,34**	**0,34**	**0,72**	**2,52**	**-0,73**	**0,38**	**-0,21**	**2,01**	**-0,55**	**10,76**
	4,04	-1,22	1,12	1,29	0,56	0,84	2,43	-0,96	0,32	-0,13	2,28	-0,30	10,63
ÃO PAULO	**0,12**	**-0,61**	**-0,29**	**2,57**	**1,78**	**0,53**	**1,89**	**-2,78**	**0,88**	**1,95**	**0,37**	**-0,68**	**5,75**
	0,02	-0,98	-0,30	3,28	2,90	1,07	1,89	-3,89	0,33	1,81	0,26	-0,89	5,42
ELÉM	**-1,13**	**2,59**	**-1,49**	**-1,76**	**1,41**	**0,26**	**2,00**	**-1,71**	**0,15**	**-1,70**	**-2,51**	**0,22**	**-3,76**
	-1,08	2,83	-1,35	-1,73	1,50	0,32	1,90	-1,84	0,07	-1,91	-2,57	0,15	-3,80
ALVADOR	**3,27**	**-0,11**	**-2,59**	**0,08**	**2,43**	**0,09**	**0,05**	**-1,20**	**0,60**	**1,19**	**-0,12**	**-0,22**	**3,39**
	2,84	-0,42	-2,81	0,27	2,98	0,35	-0,13	-1,32	0,52	1,19	-0,09	0,03	3,32
URITIBA	**0,77**	**0,01**	**-0,91**	**0,02**	**0,82**	**0,60**	**-0,69**	**2,28**	**1,37**	**1,30**	**-0,16**	**-0,11**	**5,38**
	-0,74	-2,02	1,67	3,40	3,19	1,49	-1,07	-0,64	0,88	0,78	-1,19	-1,29	4,37

AJUSTADA (shaded) ORIGINAL

SÉRIES DE ÍNDICES DE PREÇOS - IPCA
ORIGINAIS X AJUSTADAS - BRASIL E REGIÕES METROPOLITANAS
DESPESAS PESSOAIS - PERÍODO: JAN95 A DEZ95

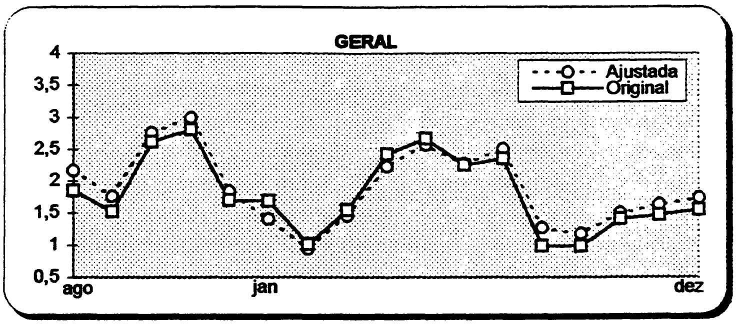
REAS GEOGRÁFICAS	JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ	ACUM
ASIL	**4,82**	**0,58**	**1,85**	**3,97**	**8,40**	**3,54**	**2,56**	**1,91**	**1,64**	**1,66**	**1,71**	**1,02**	**39,04**
	4,71	0,80	2,14	4,08	8,51	3,47	2,64	1,88	1,38	1,50	1,60	0,90	38,96
LO HORIZONTE	**5,25**	**-0,17**	**3,30**	**4,21**	**11,03**	**2,43**	**1,55**	**1,76**	**1,90**	**1,38**	**1,34**	**1,16**	**40,78**
	4,94	0,30	4,09	4,22	11,05	2,62	1,65	1,63	1,60	0,80	1,31	0,81	40,60
CIFE	**4,99**	**0,55**	**0,60**	**2,26**	**9,44**	**3,64**	**2,25**	**1,77**	**1,04**	**2,34**	**1,21**	**1,29**	**35,88**
	5,34	0,68	0,53	2,25	9,48	3,80	2,04	1,62	1,01	2,13	1,05	1,39	35,78
O PAULO	**4,92**	**0,73**	**1,45**	**4,72**	**7,64**	**3,96**	**3,09**	**1,89**	**1,74**	**1,90**	**2,45**	**0,98**	**41,55**
	4,56	1,18	2,02	5,00	7,93	3,69	3,31	1,86	1,17	1,73	2,19	0,77	41,45
LÉM	**4,62**	**0,76**	**2,57**	**4,42**	**7,06**	**3,02**	**1,91**	**1,57**	**0,58**	**3,11**	**1,06**	**0,02**	**35,12**
	5,21	0,91	2,56	4,45	7,06	3,06	1,89	1,53	0,24	2,73	1,06	-0,17	34,87
LVADOR	**3,92**	**2,49**	**1,16**	**3,28**	**6,16**	**4,60**	**5,39**	**3,07**	**2,26**	**1,10**	**3,10**	**0,62**	**43,96**
	4,20	2,46	1,21	3,22	6,19	4,64	5,41	3,07	2,23	0,96	2,99	0,57	43,94

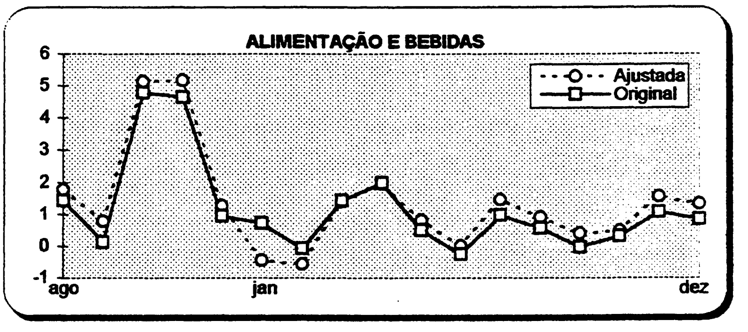
AJUSTADA ORIGINAL

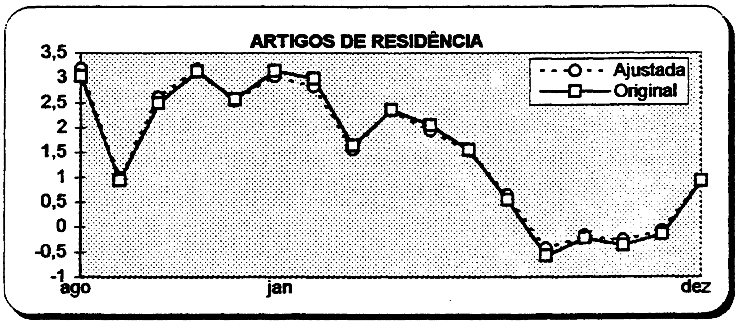
ANEXO 10

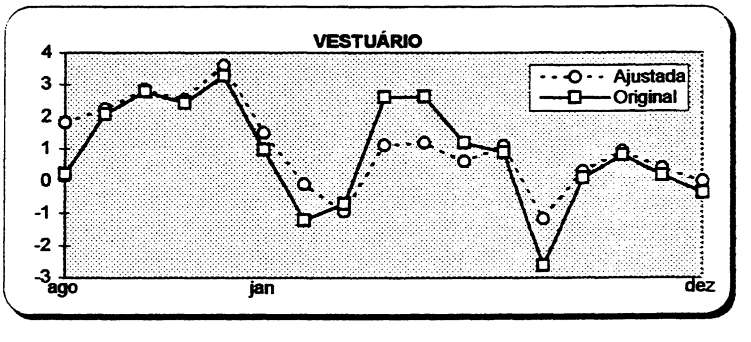
SÉRIES DE ÍNDICES DE PREÇOS - IPCA
ORIGINAL X AJUSTADA PERÍODO: AGO/94 A DEZ/95

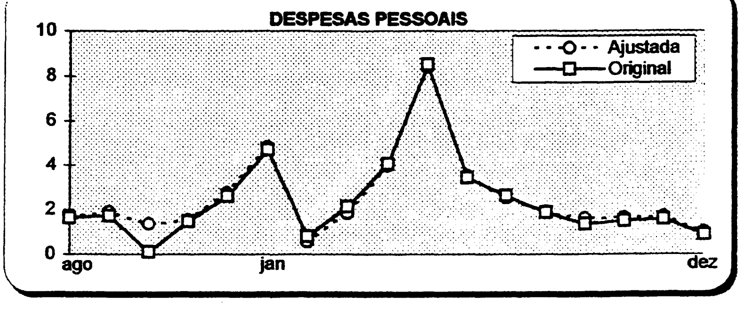
BRASIL











ANÁLISE LONGITUDINAL DE DADOS

Coordenadora: CARMEM FEIJÓ (IBGE)

Nesta Sessão temática deverão ser tratados os temas seguintes:

- A relevância para instituições produtoras de estatística em gerar arquivos de microdados para análise dos próprios dados.

- Recomendações sobre a montagem de bancos de microdados de empresas e de estabelecimentos, baseadas na experiência internacional.

- Como tratar a questão do sigilo.
  Técnicas de análise.

Análise estatística de dados longitudinais

Julio da Motta Singer

Departamento de Estatística
Instituto de Matemática e Estatística
Universidade de São Paulo

Resumo

Apresentamos uma breve caracterização de estudos com dados longitudinais salientando as diferenças relativamente a outros tipos de planejamentos. Indicamos os principais modelos estatísticos utilizados na análise desse tipo de dados, mostrando que o esforço analítico adicional está concentrado na modelagem da estrutura de correlação entre as observações realizadas na mesma unidade amostral. Discorremos sobre os métodos de estimação e implementação computacional e finalmente ilustramos esses tópicos com um exemplo numérico.

Palavras-chave: Análise de perfis, Correlação intra-unidades amostrais, Curvas de crescimento, Dados longitudinais, Medidas repetidas, Planejamento do tipo painel.

1. Introdução

Em muitas situações, há interesse em se estudar o comportamento de uma ou mais características (aqui chamadas de variáveis respostas) dos elementos de uma ou mais populações ao longo de uma certa escala ordenada. Um exemplo envolveria a comparação da evolução do faturamento de diferentes tipos de empresas ao longo de um certo período de tempo. Num outro campo, um estudo em que o objetivo é avaliar os níveis de poluição em regiões situadas a distâncias crescentes de diferentes tipos de fontes poluidoras constitui outro exemplo. Para efeito de simplificação, referir-nos-emos a essa dimensão ao longo da qual são realizadas as observações, genericamente como **tempo**.

Nesse contexto, em geral, há duas estratégias de coleta de dados: a primeira envolve a observação das variáveis respostas para uma amostra (de empresas, por exemplo) de cada uma das populações em cada instante (e somente nesse instante) ao longo do tempo. A segunda envolve a observação dessas variáveis respostas para os elementos de uma mesma amostra de cada uma das populações durante dois ou mais instantes do período em questão. No primeiro caso dizemos que o estudo é **tranversal** ("cross-sectional") e no segundo caso, dizemos que o estudo é **longitudinal**. Nas áreas de Administração, Economia ou Sociologia, essa forma de coleta de dados também é conhecida como **painel**. Neste ponto, vale ressaltar que o tipo de problema considerado sob a denominação de análise de dados longitudinais difere daquele usualmente tratado na literatura estatística por **análise de séries temporais** pelo fato de que, neste caso, em geral, dispomos de uma única unidade amostral com muitas observações ao longo do tempo (e.g. 100 ou mais), ao passo que naquele, lidamos com várias unidades amostrais (e.g. 5 ou mais) observadas em poucos instantes (e.g. 2 a 10). Vale também lembrar que os planejamentos longitudinais podem ser inseridos na classe mais ampla dos chamados planejamentos com **medidas repetidas** ("repeated measures"), que incluem, entre outros, os planejamentos do tipo "**split-plot**" e planejamentos com **intercâmbio** ("crossover"). Mais detalhes sobre a conceituação de estudos longitudinais e sua relação com outras estruturas de coleta de dados podem ser encontrados em Goldstein (1979), Singer e Andrade (1986), Duncan and Kalton (1987) ou Diggle, Liang and Zeger (1994), por exemplo.

Estudos longitudinais são de particular interesse quando o objetivo é avaliar **variações** globais ou individuais ao longo do tempo. Em primeiro lugar, esse tipo de planejamento permite observação da(s) variável(eis) resposta(s) sob condições de exposição uniforme das unidades amostrais relativamente a diferentes covariáveis. No exemplo acima, as mudanças no faturamento ao longo do tempo consideradas num estudo longitudinal estariam (pelo menos parcialmente) dissociadas de possíveis diferenças nas políticas administrativas das diversas empresas selecionadas em cada instante de observação de um estudo transversal. Essa característica tem especial interesse quando a variabilidade **entre unidades amostrais** (entre empresas, por exemplo) é maior que a variabilidade **intra-unidades amostrais** (dentro da mesma empresa ao longo do tempo, por exemplo). Em segundo lugar, estudos longitudinais dão subsídios para a avaliação de padrões individuais de variação dos níveis da(s) variável(eis) resposta(s). Finalmente, alguns parâmetros de interesse podem ser estimados de forma mais eficiente sob planejamentos longitudinais do que sob planejamentos transversais com o mesmo número de observações. Essencialmente, essa diferença na eficiência dos estimadores pode ser explicada através do seguinte exemplo.

Consideremos uma situação em que o interesse é comparar as médias de uma variável resposta observada antes e após uma certa intervenção. Denotemos por X a variável observada antes da intervenção e por Y a mesma variável observada após a intervenção. Num planejamento transversal, essa comparação seria realizada a partir dos dados provenientes de duas amostras independentes, cada uma com n unidades amostrais diferentes, digamos $(X_1,\ldots,X_n)$ e $(Y_1,\ldots,Y_n)$ através da estatística $t=(\overline{X}-\overline{Y})/s\sqrt{2/n}$ onde $\overline{X}$ e $\overline{Y}$ são as médias amostrais de X e Y respectivamente e s^2 é uma estimativa da variância de $X-Y$, ou seja, de $\sigma_X^2+\sigma_Y^2$ onde σ_X^2 e σ_Y^2 são respectivamente as variâncias de X e Y. Num planejamento longitudinal, a comparação seria realizada com a observação das variáveis X e Y nas mesmas n unidades amostrais através da utilização da chamada estatística t pareada, dada por $t_d=(\overline{X}-\overline{Y})/s_d\sqrt{1/n}$ onde s_d^2 é uma estimativa da variância de $X-Y$ que neste caso é dada por $\sigma_X^2+\sigma_Y^2-2\sigma_{XY}$ com σ_{XY} denotando a covariância entre X e Y.Convém lembrar que no caso anterior, $\sigma_{XY}=0$, em função da independência estatística entre X e Y.Quando σ_{XY} é positiva, espera-se que o denominador de t_d seja menor que o de t e consequentemente que o teste tenha mais poder para detectar diferenças entre as médias da variável resposta antes e após a intervenção.

As duas principais desvantagens associadas a estudos longitudinais estão relacionadas com seu custo (pois, em geral, é difícil garantir que as unidades amostrais selecionadas sejam observadas nas épocas designadas) e com as dificuldades técnicas de análise dos dados (como veremos adiante, os modelos estatísticos adequados para esse tipo de planejamento são, em geral, mais complexos do que aqueles adequados para estudos transversais).

Essencialmente, os problemas em que estamos interessados quando consideramos estudos longitudinais são os mesmos com que nos deparamos em estudos tranversais e podem ser classificados do ponto de vista estatístico como problemas de análise de variância (ANOVA) ou, mais geralmente, de regressão (linear ou não-linear). A diferença entre eles, está no fato de que no caso transversal lidamos com dados (estatisticamente) independentes, ao passo que no caso longitudinal é necessário considerar uma possível dependência (estatística) entre eles. O esforço adicional empregado na análise de dados longitudinais relativamente à análise de dados transversais está praticamente concentrado na modelagem da estrutura dessa dependência.

O objetivo da próxima seção é apresentar alguns modelos que permitem incorporar possíveis dependências entre as observações. Concentramo-nos em situações com uma única variável resposta contínua com distribuição gaussiana; modelos para variáveis respostas discretas ou categorizadas podem ser encontrados em Diggle, Liang and Zeger (1994) ou Koch, Singer and Stokes (1992), por exemplo.

2. Principais modelos

Num estudo longitudinal, os dados correpondentes a cada unidade amostral podem ser expressos através de um vetor contendo as respostas em cada um dos instantes de observação e de uma matriz contendo os valores das variáveis explicativas. De uma forma simbólica, o vetor com as p_i respostas associadas à i-ésima unidade amostral (chamado de **perfil de respostas**) pode ser escrito como

$$\mathbf{y}_i = (y_{i1}, \ldots, y_{ip_i})^t$$

onde y_{ik} denota a resposta obtida no k-ésimo instante, $(k = 1, \ldots, p_i)$ e $\mathbf{a}^t$ simboliza o vetor $\mathbf{a}$ transposto. Os modelos usualmente empregados na análise têm a forma

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i \qquad (2.1)$$

onde $\mathbf{X}_i$ é uma matriz de especificação, cujas colunas contêm os valores das variáveis explicativas (em geral essas variáveis estão associadas à potências do tempo ou são variáveis indicadoras, embora outros tipos de covariáveis também possam ser consideradas), $\boldsymbol{\beta}$ é o vetor de parâmetros que desejamos estimar e $\boldsymbol{\varepsilon}_i$ é um vetor de erros aleatórios com vetor de médias $\mathbf{0}$ e matriz de covariância Σ_i e para o qual geralmente se admite uma distribuição gaussiana.

Num caso em que o objetivo é modelar os dados através de uma reta para explicar a variação das p_i respostas da i-ésima unidade amostral ao longo do tempo, teríamos

$$\mathbf{X}_i = \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_{p_i} \end{pmatrix} \qquad (2.2)$$

onde t_k representa o tempo (em dias, meses etc.) decorrido entre o início do estudo e o instante em que foi realizada a k-ésima observação, e

$$\boldsymbol{\beta} = (a, b)^t$$

onde a e b representam respectivamente o coeficiente linear e o coeficiente angular da reta que pretendemos ajustar. Segundo esse modelo, o valor esperado para uma observação realizada no k-ésimo instante é dado por $a + bt_k$.

Vale apena notar que os elementos do modelo apresentado acima são os mesmos que utilizaríamos para representar a reta caso o estudo fosse tranversal. O componente que diferencia os modelos para estudos longitudinais é a estrutura imposta à matriz de covariância intra-unidades amostrais, Σ_i. No caso transversal temos $\Sigma_i = \sigma^2 \mathbf{I}_{p_i}$ onde σ^2 representa a variância comum a todas as observações e $\mathbf{I}_{p_i}$ denota uma matriz identidade de dimensão p_i. Nesse caso a covariância (ou correlação) entre quaisquer duas observações (mesmo que realizadas na mesma unidade amostral) é nula. No caso longitudinal há várias alternativas para a estrutura de Σ_i, tanto mais sofisticadas quanto mais complexa a relação de dependência entre as observações intra-unidades amostrais. A mais simples (conhecida na literatura estatística como **estrutura uniforme)** é aquela em que todas as variâncias são iguais e todas as covariâncias são iguais, ou seja

$$\Sigma_i = \sigma^2 \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{pmatrix} \tag{2.3}$$

onde ρ é o coeficiente de correlação entre quaisquer duas observações realizadas na mesma unidade amostral. A estrutura de covariância mais complexa é aquela em que todas as variâncias podem ser diferentes entre si e todas as covariâncias podem ser diferentes entre si, ou seja

$$\Sigma_i = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1p_i} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2p_i} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p_i} & \sigma_{2p_i} & \cdots & \sigma_{p_i}^2 \end{pmatrix} \tag{2.4}$$

onde σ_k^2 representa a variância das observações realizadas no k-ésimo instante e σ_{kl} representa a covariância entre as observações realizadas na mesma unidade amostral nos instantes k e l. Nesse caso dizemos que a matriz de covariância é **não estruturada**. Se por um lado, a estrutura uniforme tem o atrativo da simplicidade, por outro, ela peca por não permitir a incorporararação de um padrão bastante comum em dados longitudinais, onde as variâncias crescem com o tempo e as correlações decrescem com o espaçamento entre as observações intra-unidades amostrais (ver Kenward (1987), por exemplo). Embora o modelo com matriz de covariância não estruturada admita esse padrão, ele tem a desvantagem de envolver um número muito grande de parâmetros, mesmo nos casos em que o número de medidas intra-unidades amostrais é da ordem de três ou quatro. Esse fato causa problemas tanto de estimação quanto de interpretação. É para resolver esse dilema, que muitos autores têm concentrado esforços para apresentar modelos intermediários que incorporem o padrão de dependência mencionado acima com um número reduzido de parâmetros. Nessa categoria, os modelos mais comuns são os modelos baseados em **processos auto-regressivos** (ver Rao (1967) ou Rochon and Helms (1989), por exemplo), onde

$$\Sigma_i = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{p_i} \\ \rho & 1 & \rho & \cdots & \rho^{p_i-1} \\ \rho^2 & \rho & 1 & \cdots & \rho^{p_i-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{p_i} & \rho^{p_i-1} & \rho^{p_i-2} & \cdots & 1 \end{pmatrix} \tag{2.5}$$

ou os chamados **modelos de efeitos aleatórios** (ver Laird and Ware (1982), por exemplo), para os quais a matriz Σ_i não admite uma expressão de interpretação tão simples como aquelas apresentadas acima, mas também pode ser escrita em função de um número reduzido de parâmetros. Mais especificamente, no caso de ajuste de retas como no exemplo mencionado, um possível modelo de efeitos aleatórios é dado por

$$\Sigma_i = \mathbf{X}_i \Delta \mathbf{X}_i^t + \sigma^2 \mathbf{I}_{p_i} \tag{2.6}$$

onde $\mathbf{X}_i$ é dada por (2.2) e

$$\Delta = \begin{pmatrix} \sigma_a^2 & \sigma_{ab} \\ \sigma_{ab} & \sigma_b^2 \end{pmatrix} \tag{2.7}$$

com σ_a^2 e σ_b^2 denotando respectivamente as variâncias dos coeficientes lineares e angulares das retas associadas às diferentes unidades amostrais (consideradas como efeitos aleatórios) e σ_{ab} denotando a correspondente covariância entre esses coeficientes.

De uma forma geral as classes de modelos para a estrutura de covariância envolvem a expressão das matrizes Σ_i como função de um vetor de parâmetros desconhecidos, Θ. No caso do modelo de efeitos aleatórios descrito acima, os parâmetros da estrutura de covariância intra-unidades amostrais, Σ_i correspondem aos quatro elementos do vetor $\Theta = (\sigma_a^2, \sigma_b^2, \sigma_{ab}, \sigma^2)^t$, independentemente do número de observações realizadas em cada unidade amostral.

3. Estimação e implementação computacional

Os métodos de estimação e obtenção de estatísticas de teste sob os modelos utilizados para a análise de dados longitudinais não diferem, substancialmente, daqueles considerados no caso transversal. A maneira mais simples de se analizarem dados longitudinais é através das chamadas **medidas resumo** ("summary measures"); nesse caso, o perfil de respostas de cada unidade amostral é substituido por um único valor (a medida resumo) e o conjunto dessas medidas é analizado transversalmente, em geral através de ANOVA. Exemplos de medidas resumo são a resposta média num determinado período, a resposta máxima, a taxa de variação da resposta nesse período etc. O leitor poderá consultar Domenech (1989) para maiores detalhes. Esse tipo de análise só tem interesse, entretanto, quando os objetivos estão centrados na comparação as populações investigadas com relação a certas características das distribuições da variável resposta e não incluem a avaliação do comportamento da variável resposta ao longo do tempo. Nesse caso, outros métodos de análise, cuja essência descrevemos abaixo, são necessários.

Basicamente, se a estrutura de covariância $\Sigma_i = \Sigma_i(\Theta)$ fosse conhecida, o estimador de mínimos quadrados generalizados (ou de máxima verossimilhança) do vetor β corresponderia a

$$\bar{\beta} = (\sum_{i=1}^{n} \mathbf{X}_i^t \Sigma_i^{-1} \mathbf{X}_i)^{-1} (\sum_{i=1}^{n} \mathbf{X}_i^t \Sigma_i^{-1} \mathbf{y}_i) \tag{3.1}$$

Mesmo em algumas situações onde $\Sigma_i = \Sigma_i(\Theta)$ não é conhecida, é possível obter expressões explícitas para o estimador do vetor β a partir de (3.1). Esse é o caso de situações em que além de o planejamento prever que todas as unidades amostrais sejam observadas nos mesmos instantes, não existam observações incompletas e que as matrizes de covariância sejam do tipo uniforme ou não estruturada. Diz-se, então, que os dados são **balanceados em relação ao tempo**. Estão nessa categoria os problemas tratados na literatura como **Análise de Perfis** ("Profile analysis") e **Análise de Curvas de**

Crescimento ("Growth Curve Analysis"). Maiores detalhes podem ser obtidos em Singer e Andrade (1986) ou Crowder and Hand (1990), por exemplo. Esses estimadores com forma explícita também podem ser obtidos para alguns modelos de efeitos aleatórios quando os dados são balanceados em relação ao tempo. O leitor poderá consultar Graybill (1976) para maiores detalhes. Nesse contexto, a distribuição (exata) do estimador $\hat{\boldsymbol{\beta}}$. é Normal multivariada com vetor de médias $\boldsymbol{\beta}$ e matriz de covariância $(\sum_{i=1}^{n} \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1} \mathbf{X}_i)^{-1}$.

No caso geral, em que as observações podem ser feitas em instantes não especificados, podem ser incompletas ou podem ter estrutura de covariância regida por outros modelos, temos que recorrer a métodos iterativos para obtenção dos estimadores de máxima verossimilhança ou máxima verossimilhança restrita (que produz estimadores menos viesados). Entre eles, destacamos o método de Newton-Raphson, o método "Scoring" de Fisher ou o método EM. Em particular, o método de Newton-Raphson envolve a especificação de valores iniciais $\boldsymbol{\beta}^{(0)}$ e $\boldsymbol{\Theta}^{(0)}$ para os parâmetros do modelo e o cálculo iterado (para $k = 1,2,\ldots$) da expressão

$$\begin{pmatrix} \boldsymbol{\beta}^{(k)} \\ \boldsymbol{\Theta}^{(k)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}^{(k-1)} \\ \boldsymbol{\Theta}^{(k-1)} \end{pmatrix} + \begin{pmatrix} \mathbf{J}_{\beta\beta} & \mathbf{J}_{\beta\Theta} \\ \mathbf{J}_{\Theta\beta} & \mathbf{J}_{\Theta\Theta} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{u}_{\beta} \\ \mathbf{u}_{\Theta} \end{pmatrix} \qquad (3.2)$$

onde $\mathbf{u}_{\beta}$ e $\mathbf{u}_{\Theta}$ denotam as derivadas parciais de primeira ordem da função de verossimilhança relativamente a $\boldsymbol{\beta}$ e $\boldsymbol{\Theta}$ calculadas no ponto $(\boldsymbol{\beta}^{(k-1)}, \boldsymbol{\Theta}^{(k-1)})$ e $\mathbf{J}_{\beta\beta}, \mathbf{J}_{\beta\Theta}, \mathbf{J}_{\Theta\beta}$ e $\mathbf{J}_{\Theta\Theta}$ denotam as correspondentes derivadas parciais de segunda ordem calculadas no mesmo ponto. As iterações são interrompidas quando a distância (segundo algum critério) entre as estimativas obtidas em dois passos consecutivos do processo iterativo atinge um valor pré-especificado. No caso do método "Scoring" de Fisher, o procedimento iterativo pode ser simplificado e a expressão (3.2) pode ser utilizada apenas para estimar os parâmetros referentes à estrutura de covariância, $\boldsymbol{\Theta}$, sendo os parâmetros de posição, $\boldsymbol{\beta}$, estimados através da expressão (3.1).

Inferências sobre o vetor de parâmteros $\boldsymbol{\beta}$ podem ser feitas a partir da distribuição (aproximada) do estimador $\hat{\boldsymbol{\beta}}$, que é Normal multivariada com vetor de médias $\boldsymbol{\beta}$ e uma matriz de covariância que pode ser estimada por $(\sum_{i=1}^{n} \mathbf{X}_i' \boldsymbol{\Sigma}_i^{-1}(\hat{\boldsymbol{\Theta}}) \mathbf{X}_i)^{-1}$ onde $\hat{\boldsymbol{\Theta}}$ é um estimador consistente de $\boldsymbol{\Theta}$.

Para dados balanceados em relação ao tempo, detalhes sobre os métodos de estimação podem ser encontrados em diversas fontes como Winer (1971), Morrison (1976) ou mais recentemente em Crowder and Hand (1990). Nos casos mais gerais, sugerimos os trabalhos de Jennrich and Schluchter (1986), Laird, Lange and Stram (1987), Andreoni (1989), ou Jones (1993), por exemplo. Procedimentos para o ajuste de modelos estatísticos a dados longitudinais estão incorporados na maioria dos pacotes de "software" estatístico. No BMDP, as subrotinas 2V e 4V podem ser utilizadas para analizar dados balanceados em relação ao tempo; isso também pode ser feito através da subrotina GLM do SAS. No caso geral, a análise pode ser conduzida através da subrotina 5V no BMDP ou MIXED no SAS, por exemplo. Em Andreoni (1989) podem-se encontrar programas fontes para implementação dos algoritmos mencionados para modelos de efeitos aleatórios; esses programas estão escritos na linguagem matricial (CM) do pacote NTIA da EMBRAPA.

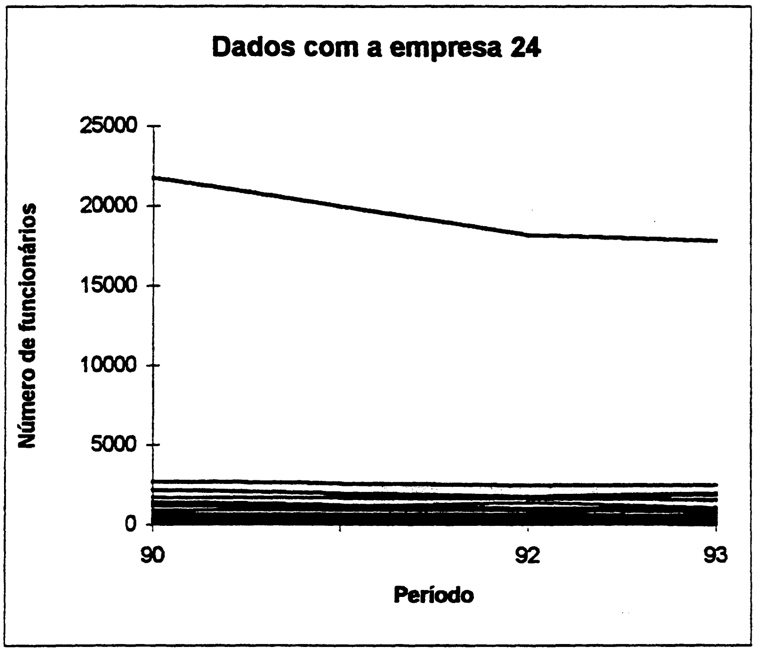
4. Exemplo

Os dados apresentados em anexo correspondem ao número de funcionários empregados em 38 empresas durante os anos de 1990, 1992 e 1993. Eles foram obtidos dos arquivos do IBGE e são utilizados aqui apenas com a finalidade de ilustrar uma análise de dados longitudinais. O objetivo é estudar a evolução do número de funcionários ao longo do período mencionado, assumindo que essas 38 empresas possam ser consideradas como uma amostra aleatória simples de uma população de interesse.

	Número de funcionários no ano de		
Empresa	1990	1992	1993
1	817	1402	1024
2	640	520	482
3	1172	1041	941
4	1444	1016	1059
5	85	23	84
6	596	681	483
7	337	289	229
8	1671	1719	1499
9	462	544	414
10	1263	963	830
11	264	483	549
12	619	542	491
13	633	598	592
14	106	175	161
15	385	442	404
16	189	207	201
17	295	250	243
18	288	247	237
19	533	449	438
20	122	135	112
21	1732	1635	1564
22	658	627	577
23	2716	2469	2450
24	21761	18186	17812
25	2178	1705	1898
26	238	280	105
27	266	294	325
28		417	332
29	2177	1763	1964
30	51	71	14
31	311	371	466
32	136	136	12
33	354	275	238
34	254	272	243
35	349	355	355
36		103	0
37			0
38	922	904	777

Nas Figuras 4.1 e 4.2 apresentamos os chamados **diagramas paralelos de dispersão** ("parallel plots") com e sem a inclusão de uma das empresas. Esses gráficos podem ser utilizados com a finalidade de identificar possíveis modelos para a estrutura de covariância e/ou de detectar perfis de resposta discrepantes. O leitor poderá consultar Rao and Rao (1966), Weiss and Lazaro (1992) ou Suyama (1995) para detalhes sobre a utilização de métodos gráficos na identificação de modelos para análise de dados longitudinais. O perfil correspondente à empresa desconsiderada na construção na Figura 4.2 (identificada pelo número 24 no conjunto de dados) pertence a essa categoria; os dados correspondentes foram eliminados do restante da análise.

Figura 4.1: Diagrama paralelo de dispersão para as 38 empresas.

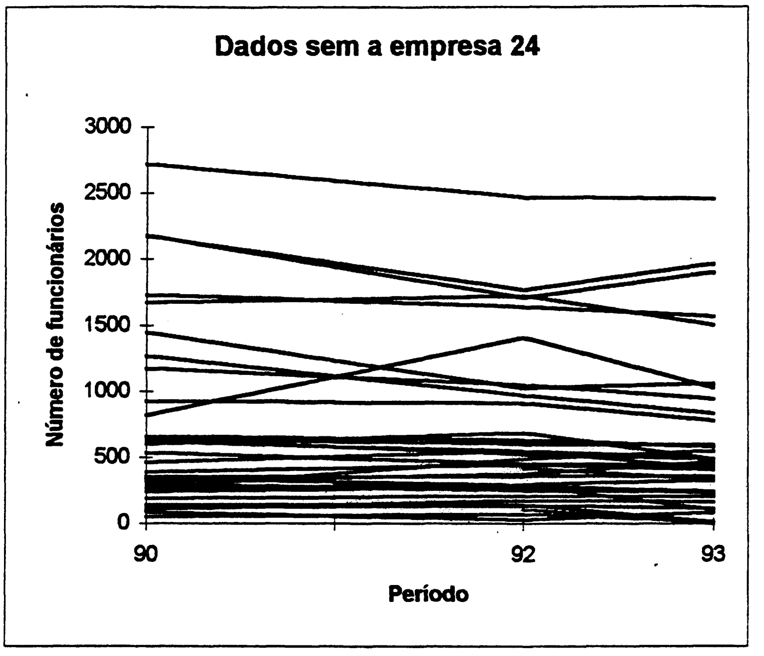


Os números médios de funcionários e correspondentes erros padrões para as 37 empresas consideradas nos anos de 1990, 1992 e 1993 são respectivamente 713 ± 118, 650 ± 98 e 589 ± 98. Pode-se observar uma tendência decrescente das médias nesses três anos. Se desconsiderarmos a possível correlação entre as observações intra-unidades amostrais, a comparação dessas médias via uma ANOVA não evidencia que as diferenças observadas sejam significativas ($p=0.7031$). A matriz de correlações intra-unidades amostrais é

	1990	**1992**	**1993**
1990	1.000	0.966	0.982
1992	0.966	1.000	0.984
1993	0.982	0.984	1.000

indicando que não só os dados são altamente correlacionados como também que um modelo com estrutura de covariância uniforme poderia ser considerado na análise. Esse modelo também é sugerido pelo padrão de paralelismo dos perfis de resposta da Figura 4.2. Em função dessas indicações comparamos as médias de interesse através de uma ANOVA para medidas repetidas

Figura 4.2: Diagrama paralelo de dispersão excluindo a Empresa de número 24.



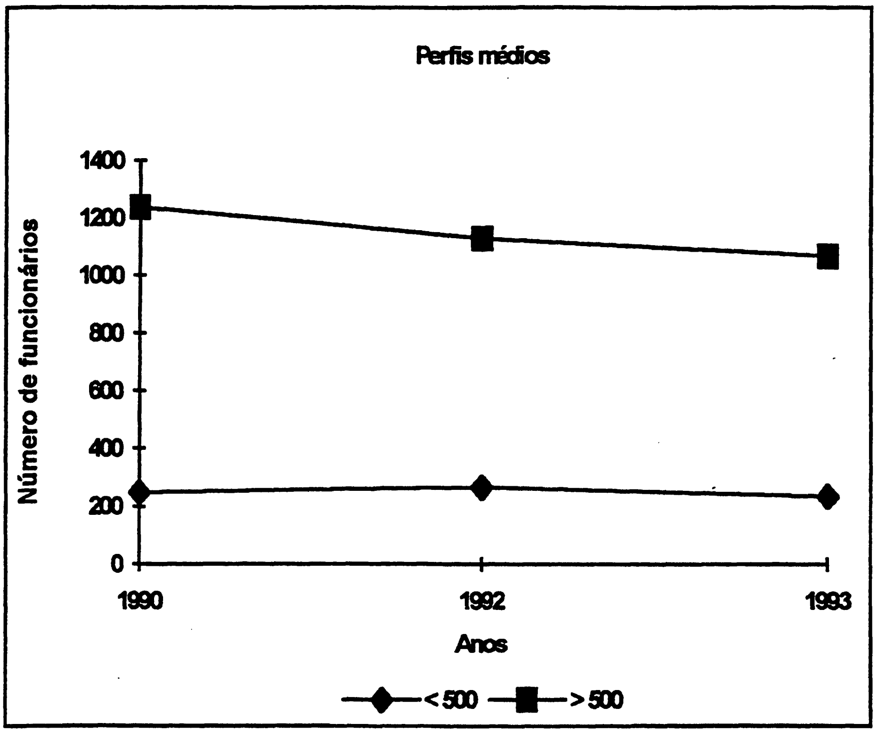
utilizando o procedimento GLM do pacote estatístico SAS. Para efeito de ilustração, as matrizes de especificação seriam dadas por $\mathbf{X}_i = \mathbf{I}_3$ e o vetor de parâmetros por $\beta = (m_{90}, m_{92}, m_{93})^t$ onde m_j representa o número médio de funcionários no ano j, $j = 90, 92, 93$. Esse procedimento elimina as empresas com observações incompletas da análise; a inclusão dessas empresas pode ser considerada através de outros procedimentos mais sofisticados; optamos pela alternativa mais simples em função do pequeno número de observações incompletas e do caráter puramente ilustrativo do exemplo. Neste caso, o resultado detecta uma diferença altamente significativa (p=0.0096) entre as médias em questão, sugerindo uma tendência decrescente para o número médio de funcionários empregados.

Uma análise complementar, também com propósitos puramente ilustrativos, envolve a classificação das empresas como **pequenas** ou **grandes** consoante o número de funcionários em 1990 (no exemplo, adotamos arbitrariamente, 500 como ponto de corte). Os números médios de funcionários nesses dois grupos estão indicados abaixo.

Tamanho da empresa em 1990	**Número médio de funcionários em**		
	1990	**1992**	**1993**
Pequena	250	269	244
Grande	1236	1127	1067

Na Figura 4.3 estão representados os chamados **perfis médios**, que essencialmente correspondem aos conjuntos de valores médios associados a cada grupo.

Figura 4.3: Número médio de funcionários



Um exame desses perfis médios indica uma tendência decrescente do número médio de funcionários para as empresas grandes, em contraposição a uma certa estabilidade para as empresas pequenas ao longo do período investigado. Uma Análise de Perfis (ver Winer (1971) ou Singer e Andrade (1986), por exemplo) para esses dados confirma a significância estatística dessa diferença de comportamento (com $p<0.0082$ para o teste da hipótese de paralelismo dos perfis médios, que corresponde à hipótese de inexistência de interação entre Tamanho e Tempo no jargão estatístico) e sugere que o número médio de funcionários das empresas pequenas pode ser considerado estável ($p>0.6560$) com valor 254 entre 1990 e 1993 ao passo que decresce ($p<0.0003$) com uma taxa de 56 por ano para as empresas grandes no mesmo período. A Análise de Perfis é baseada num modelo do tipo "split-plot" (com dois fatores - Tamanho e Tempo) que induz uma estrutura uniforme para a matriz de covariância intra unidades amostrais. Os testes obtidos sob esse modelo continuam válidos mesmo sob condições menos restritivas (i.e. de **esfericidade**) para a estrutura de covariância das observações realizadas nas mesmas unidades amostrais. Além disso, testes aproximados para as hipóteses de interesse podem ser obtidos mediante correções nos graus de liberdade dos testes originais, mesmo sob condições ainda mais gerais. Uma ANOVA usual com os mesmos dois fatores não detecta as diferenças acima.

Agradecimentos

Este trabalho foi parcialmente finaciado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq. O autor agradece a colaboração de Adriana Sañudo na análise dos dados.

Referências Bibliográficas

Andreoni, S. (1989). Modelos de efeitos aleatórios para análise de dados longitudinais não balanceados em relação ao tempo. Dissertação de mestrado. São Paulo: Departamento de Estatística, Instituto de Matemática e Estatística, USP.

Crowder, M.J. and Hand, D.J. (1990). **Analysis of Repeated Measures**. London: Chapman and Hall.

Diggle, P.J., Liang, K.Y. and Zeger, S.L. (1994). **Analysis of longitudinal data**. Oxford: Clarendon Press.

Domenech, C.H. (1989). Métodos exploratórios e utilização de medidas resumo para análise de dados longitudinais. Dissertação de mestrado. São Paulo: Departamento de Estatística, Instituto de Matemática e Estatística, USP.

Duncan, G.J. and Kalton, G. (1987). Issues of design and analysis of surveys acroos time. **International Statistical Review, 55**, 97-117.

Goldstein, H. (1979). **The design and analysis of longitudinal studies**. London: Academic Press.

Graybill, F.A. (1976). **Theory and applications of the linear model**. North Scituate, Mass.: Duxbury Press.

Jennrich, R.L. and Schluchter, M.D. (1986). Unbalanced repeated-measures models with structured covariance matrices. **Biometrics, 42**, 805-820.

Jones, R.H. (1993). **Longitudinal Data with Serial Correlation: a State Space Approach**. London: Chapman and Hall.

Kenward, M.G. (1987). A method for comparing profiles of repeated measurements. **Applied Statistics, 36**, 296-308.

Koch, G.G., Singer, J.M. and Stokes, M.E. (1992). Some aspects of weighted least squares analysis for longitudinal categorical data. In **Statistical Models for Longitudinal Studies of Health**, eds. J.H. Dwyer, M. Feinleib, P. Lippert and H. Hoffmeister. (Monographs in Epidemiology and Biostatistics. Vol. 16). New York: Oxford University Press, 215-258.

Laird, N.M., Lange, N. and Stram, D. (1987). Maximum likelihood computations with repeated measures: application of the EM algorithm. **Journal of the American Statistical Association, 82**, 97-105.

Laird, N.M. and Ware, J.H. (1982). Random effects models for longitudinal data. **Biometrics, 38**, 963-974.

Morrison, D.F. (1976). **Multivariate Statistical Methods, 2nd. edition**. New York: McGraw-Hill.

Rao, C.R. (1967). Least squares theory using estimated dispersion matrix and its application to measurement of signals. In **Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability,1,** eds. L.M. LeCam and J. Neyman. 355-372. Berkeley: University of California Press.

Rao, M.N. and Rao, C.R. (1966). Linked cross-sectional study for determining norms and growth-rates - a pilot survey of Indian school-going boys. **Sankhya, B28**, 237-258.

Rochon, J. and Helms, R.W. (1989). Maximum likelihood estimation for incomplete repeated measures experiments under an ARMA covariance structure. **Biometrics, 45**, 207-218.

Singer, J.M. e Andrade, D.F. (1986). **Análise de Dados Longitudinais**. São Paulo: Associação Brasileira de Estatística.

Suyama, E. (1995). Identificação de um modelo de efeitos aleatórios. Tese de doutoramento. São Paulo: Departamento de Estatística, Instituto de Matemática e Estatística, USP.

Weiss, R.E. and Lazaro, C.G. (1992). Residual plots for repeated measures. **Statistics in Medicine, 11**: 115-124.

Winer, B.J. (1971). **Statistical Principles in Experimental Design**, 2nd edition. New York: McGraw Hill.

Draft: April 11,1996

The Development and Use of Longitudinal Microdata:
The U.S. Census Bureau's Center for Economic Studies' Experience With Confidential Survey Microdata

by
Robert H. McGuckin and Arnold P. Reznek[1]

Abstract

This paper discusses lessons learned from the experience of the Center for Economic Studies (CES) of the U.S. Census Bureau with the development and use of micro-level longitudinal data sets. While the research has improved our understanding of the workings of the economy, we focus here on the practical aspects of the experience. Perhaps the primary lesson that has been learned is that programs providing analytic users access to the microdata enable the U.S. Census Bureau and the researchers to become partners working together to improve the basic information that society uses to make decisions. An access program is a simple, cost effective way to take advantage of the coincidence of interests of statistical agencies and analytic researchers from bcth academic and policy areas.

1 Center for Economic Studies, U.S. Census Bureau. The views expressed herein do not necessarily reflect those of the U.S. Census Bureau.

1. Introduction

In a typical year, the U.S. Census Bureau, like statistical agencies around the world, carries out numerous surveys of firms and establishments. These data are routinely processed, tabulated, and published as cross-section aggregates -- the minimum level of aggregation being determined primarily by considerations of confidentiality. The effort is directed toward tabular output with surveys designed to produce aggregate point-in-time cross-section estimates for use in the National Income and Product Accounts (NIPA), input-output tables, business cycle indicators, and productivity measures for various sectors.

Historically, the microdata that form the basis for these aggregate statistics were considered expendable once the tabulations had been made.[2] Moreover, even though construction of the first longitudinal establishment-level microdata database at the U.S. Census Bureau -- known today as the Longitudinal Research Database (LRD) -- dates from the early 1980s, only within the last 2 or 3 years has it been designated as one of the core assets at the U.S. Census Bureau. This recognition coincides with sharply increased demands for microdata from analytic subject matter specialists in both the academic and policy communities.

This paper discusses lessons learned from the experience of the Center for Economic Studies (CES) of the U.S. Census Bureau with the development and use of micro-level longitudinal data sets. While CES research has generated a large number of basic insights about the workings of the economy, we focus here on the practical aspects of the experience.[3] We place particular emphasis on access by analytic users to the confidential microdata, the partnerships that researchers form with the U.S. Census Bureau when they work at CES, and the advantages of the program to both groups.

Widespread support for microdata panels from the research and policy communities is a relatively new phenomenon in the area of "economic" statistics -- the data obtained from business establishments and firms. The predominant pattern, until recently, has been that academic economists worked primarily with theory, and business and empirical economists worked primarily with macroeconomic models and aggregate data. Aside from practical considerations -- such as the high costs of using microdata for research and policy analysis before the arrival of low cost computer hardware and software -- analysts and policy makers hoped that the losses introduced by use of aggregate data were small. While researchers recognized that aggregate

---

2 Microdata from the U.S. Census of Manufactures exists for five years, four of them during the great depression, prior to 1963.

3 For summaries of substantive work at CES see McGuckin (1995), McGuckin and Jensen (1996), and the last several CES Annual Reports.

statistics meant the sacrifice of information[4], they hoped these losses were an acceptable price to pay for the simplicity and tractability that aggregate data affords.

A key reason for the increased appreciation for microdata is the growing awareness that the information loss from use of aggregate data is significant. In study after study at CES, researchers are finding that most of the variation in behavior of plants is observed within traditional sectorial aggregations -- industries, geography, size classes, etc.[5] This means that, for example, the establishment -- not the firm or industry -- is often the most appropriate unit of analysis for understanding economic performance. And, in many cases, the loss of information from use of firm- and industry-level aggregates leads to seriously distorted and incorrect conclusions.[6]

The growing awareness that answers to certain types of questions require longitudinal data is also an important aspect of the increased demands for microdata. It is virtually impossible to sort out cause and effect relationships without panel data.[7] Moreover, evaluation of the effects of policy interventions is seriously compromised without observations of the performance of the affected entity before and after the imposition of the policy.[8] It is also critical that a sample of "control" observations be included in the empirical model that evaluates the effects of the policy. Microdata often makes this aspect of the empirical specification feasible.

---

4 See Theil (1954) for example.

5 Similar results are being obtained for other countries. See, for example, Davis, Haltiwanger, and Schuh (1996), especially Chapter --, where establishment based statistics on job creation and destruction are compared for various countries.

6 See, for example, McGuckin and Nguyen (1995b) where it is demonstrated that aggregation to the firm level obscures the relationship of ownership change and productivity change. As another example, Baily, Bartelsman and Haltiwanger (1994) show that aggregate statistics for the manufacturing sector provide an incorrect picture of the relationship of downsizing in productivity growth. There are many other examples from recent CES studies.

7 See Bernard and Jensen (1996) and McGuckin, Streitwieser, and Doms (1995) for discussion of this point in the context of empirical examinations of exporter performance and technology adoption, respectively.

8 See Olley and Pakes (1992) for an example of the application of longitudinal microdata to an industry undergoing policy change. Also see Jarmin (1995, 1996) for studies involving evaluation of a specific Government program.

The U.S. Census Bureau, particularly in the area of business and economic statistics, has become more supportive of direct access to microdata and the development of longitudinal panels.[9] Some of this support reflects the increased demands of analytic users. Some of it is due to increased sensitivity to customer needs and requirements as budgets for traditional statistical data products are reduced.

An important, and often overlooked, factor has been recognition of the importance of analytical subject matter research to a world-class statistical program. Until recently, most managers at the U.S. Census Bureau saw subject matter research as a drain on resources, rather than an activity with potential benefits. The products generated by the CES research program have increased support among Census officials for preservation of and research with the microdata. Analytic use of the microdata has generated invaluable insights and improved the quality and usefulness of U.S. Census Bureau survey programs. In addition, microdata panels are leading to new data products that can be marketed in traditional ways, and they have become an asset in attracting survey business. In short, it is becoming more widely acknowledged that the microdata are, in themselves, a valuable statistical agency asset.

The attitudes of the U.S. Census Bureau have also been swayed by recent experiences with the data access program at CES. Having witnessed literally hundreds of researchers from government, academia, and research organizations use the data without incident over the past ten years has eased concerns about protecting the confidentiality of the data. Nevertheless, such concerns are important and remain a central feature of the rules and procedures that guide CES operations.[10]

The CES program offers researchers access to sampling frames and general data collection programs that provide detailed and generally representative data for very detailed geographic areas, products, and industries. Such access is tightly controlled: Projects are chosen to ensure fairness and scientific merit, and the confidentiality of data is maintained. However,

---

[9] The "demographic" side of the U.S. Census Bureau is also feeling increased demands for local area microdata for which public use databases are difficult to construct because of disclosure concerns.

[10] There are well understood legal (Title 13 U.S.C) and moral obligations to protect the confidentiality of the data that survey respondents provide. We believe that without confidentiality protection, the survey response rates would decrease to unacceptable levels (at least in the United States).

research results are widely disseminated and peer review of research findings is fostered in many ways.

The CES access program ensures that the data are used to their fullest extent. This is not just a matter of the researchers exploiting data already collected. While access means a bigger proportion of the data become available for analysis, access also means that researchers effectively become efficient agents for change and quality improvement at the U.S. Census Bureau. By working with the underlying data, the researchers suggest improvements to the basic data collection programs and design and measurement changes far more effectively than is possible when their contact with the data is in the form of aggregative statistics used in endeavors far removed from the producers of the data -- the statistical agency.

While the information collected by the Census provides a very detailed and complete profile of the structure of the economy, it does not by itself provide sufficient detail to understand many research and policy problems.[11] In fact, in a substantial portion of research projects at CES researchers bring new data sets and merge them with the Census microdata. This not only improves individual research projects, it also provides important feedback to the U.S. Census Bureau on the quality of its data and possibilities for more efficient survey designs. A related recent development is survey sponsorship by research organizations on the basis of researchers identifying important gaps in the information provided by the U.S. Census Bureau, sponsor design in collaboration with U.S. Census Bureau professionals and using U.S. Census Bureau sampling frames, with analysis of the survey results (perhaps combining with other data) undertaken in collaboration with the sponsor at U.S. Census Bureau facilities.[12]

The paper is organized as follows. We begin by describing why analytic subject matter research is crucial to statistical agencies and why statistical agencies are unlikely to be able to obtain the benefits from subject matter research in the absence of an access program. We also argue that the cost of providing access is low at a statistical agency, which already has collected the basic microdata for other purposes, and can cover many of the costs of the program with user fees. To provide a concrete example of a program that works, we next describe the model for access developed by CES. The CES model offers a practical program that can be adapted for use in a variety of circumstances. Next, we briefly summarize some main themes from the

---

[11] See Davis, Haltiwanger, and Schuh (1996) and Jensen and McGuckin (1996) for discussions of the relatively small percentage of variation in plant behavior explained by the variables typically collected by statistical agencies.

[12] While such merged data sets often create problems of bias and representativeness, these problems are no worse, and probably less severe at the U.S. Census Bureau.

substantive research at CES. This serves as a background for a discussion of how access and the resulting analytic research has benefitted the U.S. Census Bureau. These benefits derive from creation of new data products, improvements and extensions to existing data programs, and, most recently, new survey business for the U.S. Census Bureau.

In spite of these successes, all is not perfect, and we conclude with some unsolved problems and possible changes for the future.

2. Analytic Subject Matter Research and Access Are Important to Statistical Agencies.

A capability in analytical subject matter research is essential for statistical agencies for many reasons. First, it provides the agency with an understanding of the problems with its data: Without using the microdata it is difficult to understand what is wrong with them or how to improve them. Second, with an analytic capability it is possible for a statistical agency to provide the public, in an non partisan way, with an understanding of the economic forces that affect them. Third, analytic work provides an understanding of the needs of customers, including both the executive and legislative branches of the government. Fourth, new data products are most often generated as a part of analytic research and problem solving. Thus, analytic research provides "essential" research and development services to the agency. Finally, analytical research can lead to new surveys since analysis helps to identify areas where information is incomplete and new data are needed for resolution of an issue. In fact, the U.S. Census Bureau has attracted several recent surveys because of the possibility of analytic work with the survey microdata, coupled with the LRD and the program for data access.

A comparison of the experience of demographic programs that deal with household and individual data with that of economic programs relying on business data illustrates the value of access to microdata. No known theorem argues that the supply side of the labor market is less important than the demand side. Yet, for many years, labor market researchers have had microdata on individual workers (the supply side of the market) available through public use files. Moreover, funds have been available for longitudinal household surveys such as the Survey of Income and Program Participation (SIPP). There are other examples, such as the National Longitudinal Survey, conducted outside the U.S. Census Bureau.

A key difference between the economic and demographic areas has been the significant presence of analytical researchers with access to microdata in the demographic areas, and the absence of them, until recently, in economic areas. Access to demographic data is related directly to the existence of public use microdata files, which are released and widely used. Researchers can be accommodated on the demographic side because the large sample sizes involved make it possible to release masked microdata without fear of disclosure. In contrast, business data is characterized by small numbers of observations and extremely skewed size distributions at appropriate levels of aggregation and thus cannot support useful public use databases. See McGuckin and Nguyen (1990) and Fuller (19__). This difference is a key reason why analytic use of business data must rely on direct access programs, at least in the United States.[13]

The use of microdata panels is not restricted to large area demographic data. On the business side, longitudinal microdata have been used extensively by finance economists: They have ready access to panels of daily security prices and income and balance sheet information on publicly traded companies. Finance specialists have used these panels to understand the operation of security markets and to develop new financial products (e.g., options, index mutual funds and other "derivatives") allowing greater flexibility for businesses seeking capital, as well as providing investors with new tools to manage their risks. It is probably not a coincidence that finance, like labor economics, has been one of the more dynamic areas of social science research in the past 20 years.

Regardless of how it is accomplished, access to the data is essential if the benefits of microdata panels are to be fully realized. It might be argued that agencies could simply develop internal programs of analytic subject matter research and thereby limit the possibilities for access and exposure to disclosure risk. It is always true that an increase in the number of people who have access to data increases the possibilities of disclosure. There are two strong arguments against this reasoning. First, the empirical evidence is against it. The CES program has offered analytic researchers access to economic microdata for some time without incident. Moreover, there are good reasons to expect these good results to continue.

Second, statistical agencies are not likely to develop analytic research programs on their own. Aside from difficulties in attracting the best academic talent on a permanent basis, internal research programs tend to be either co-opted by the political process or starved because of inadequate funding. By developing independent research associates, the possibility for scientifically sound research is maximized.

---

13 There are public use data files of establishment data outside the United States. For example, Millward (1993) discusses the British Workplace Industrial Relations Surveys (WIRS) and the conditions under which they are made available. However, the United States has a different legal environment, and the WIRS data sets are based on small samples. For large samples, it is still unlikely that public use files will be released in the foreseeable future in the United States .

Beyond this, public use samples do not allow researchers to link the public use data to data from other sources except at often unacceptable levels of aggregation.

An important ingredient of such a program is independence and wide dissemination of research results. Peer review of these research results preserves independence and helps CES to recruit quality staff members and research associates. In the absence of freedom to publish findings independently of the statistical agency, researchers are not likely to become a part of the program. While there is always some risk that CES research output could be confused with the agency's output and that the agency's reputation may suffer as a result, in practice this has not been a problem. We believe this is an important distinction that has made the CES program relatively popular with analytic researchers.

3. The Center for Economic Studies – A Vehicle for Access

The mission of CES is centered on three closely related activities: the creation and maintenance of longitudinal microdata, the development and implementation of procedures and mechanisms to provide access to the confidential microdata, and analytic subject matter research with the microdata. Access is instrumental in maximizing the value of longitudinal microdata panels and the procedures used to accomplish it are a major issue for many statistical agencies both in the U.S. and abroad. In this section we focus on the ways CES provides data access and satisfies the requirements of analytic customers, while meeting the legal, moral, and political requirements that the U.S. Census Bureau operates under.

Models for Access – CES Research Model. Under United States law, microdata from the U.S. Census Bureau data programs are confidential and may only be used for statistical purposes at secure sites by U.S. Census Bureau employees or by individuals who have obtained Special Sworn Status (SSS) from the U.S. Census Bureau. The law provides specific penalties for violations.[14] The primary reason is concern for the privacy rights of companies that fill out survey forms, but the publication (intended or not) of microdata would also reduce co-operation with data collection programs and cause response rates to fall. so the agency has practical reasons to ensure confidentiality.

Researchers who use confidential U.S. Census Bureau microdata must obtain SSS status. They are subject to the same prohibitions, rules, and legal penalties as regular U.S. Census Bureau employees and contractors. In fact, researchers probably face greater penalties than employees because of the potential for loss of their professional credentials, as well as denial of future access to the microdata if they disclose confidential information. Once they are sworn in, researchers work at a Census secure site -- a site where appropriate measures to ensure physical

---

[14] The relevant law is Title 13, U.S.C., Section 214. Violations are punishable with a fine of not more than $5,000 and imprisonment of not more than five years, or both.

security are in place and where a Census staff member is in charge. The researchers work directly with the individual or firm specific microdata required for their project. For most practical purposes they function just like regular staff researchers: They take part in internal research seminars and offer advice and discuss their projects with operating Division staffs and statistical professionals. Aside from developing and enforcing basic operating procedures, CES employees protect data confidentiality by performing "disclosure analysis" on any results the researchers wish to publish or remove from U.S. Census Bureau grounds. This analysis ensures that U.S. Census Bureau policies are met with respect to release of data.

It is worth noting in this regard that the range of information that can be released without violations of confidentiality is much greater for analytic results (e.g. regression coefficients) than for tabular data. Indeed, CES procedures suggest researchers minimize tabular output because secondary disclosure is very difficult: The operating divisions release as much tabular data as possible, which limits the amount of additional tabulations that can be released. Use of regression coefficients which offer protection (the r-squares are never close to 1.00) and qualitative descriptions is possible because CES stands ready to verify the specific statistical results reported.

In addition to obtaining SSS and paying of a laboratory fee to cover their project costs, research associates' projects must benefit the data collection activities of the U.S. Census Bureau. Meeting this criteria involves two basic considerations. First, proposed research projects must use U.S. Census Bureau microdata and have scientific merit. Second, as part of their projects, the researchers must be willing to try to improve the micro databases at CES, produce new data products, or provide the U.S. Census Bureau with recommendations for improving its data programs -- improved survey concepts or questionnaires, better survey processing procedures, new data products, etc.

CES employees support the research projects by becoming closely involved with them. CES staff researchers, augmented by a small computer staff, also undertake an internal program of database development and economic research independently and jointly with the "research associates" (as SSS researchers are called at CES). CES staff researchers are an important part of the project choice process and become agents for change within the U.S. Census Bureau. With the knowledge gained through their research and interaction with the research associates, CES employees are an important part of the transmission of research results within the U.S. Census Bureau. This is how the U.S. Census Bureau learns how to improve its data programs. We give several examples in Section 5. But, we also emphasize that in our experience research associates themselves also often have substantial contacts with operating division personnel.[15]

Research Data Center Program. Since space and resources are limited at CES headquarters and it is often very expensive for researchers to relocate for the time required for major projects , we have begun to establish secure facilities, called Research Data Centers (RDCs), away from CES headquarters. The pilot RDC is the Boston Research Data Center (BRDC), established in January 1994 under a grant from the National Science Foundation (NSF). This RDC has proved successful, and final negotiations are underway to establish a new RDC this spring or summer at the Carnegie Mellon University Heinz School of Public Policy and Management in Pittsburgh. In addition, we have had serious talks with researchers and potential sponsors in Los Angeles, San Francisco, New York, Chicago, and Atlanta about setting up other RDCs. With the exception of San Francisco, each of these areas has a U.S. Census Bureau Regional Office capable of accommodating an RDC. The Regional Directors have been very supportive of the RDC concept, following the outstanding lead of Arthur Dukakis in Boston, who has been instrumental in the success of the first RDC. An important feature of the RDC program is its flexibility and ability to accommodate local variations. For example, an RDC can focus on specific studies of the regional economy (sometimes comparing it to the nation or other regions) and local economic issues. Or an RDC could develop specialties in areas like health, crime, or environmental issues. In fact, the Boston RDC has developed a concentration in environmental economics research. An RDC prospectus is now available to parties interested in creating new RDCs.

4. Subject Matter Benefits From Data Access

A major theme that has emerged from the CES research program is that heterogeneity in the distribution of business units is pervasive along a wide variety of dimensions. Firms differ dramatically, even within the same geographic areas and similar regions, and within four-digit industries and five-digit product classes as defined by the Standard Industrial Classification (SIC). Heterogeneity is observed across time as well as in the cross section. Not only does the growth process differ across firms, it is characterized by large discrete movements rather than smooth or continuous changes even for those establishments and firms in continuous operation. During any time interval, observed changes are "lumpy" and uneven. Some business units open and some grow, while others shrink and die. These facts raise the issue of what is the appropriate level of aggregation for analytic research.

---

15 The number of CES permanent staff ranges from 20-30. With over 50 projects underway during any year, CES staff can not undertake all such contacts with the other operating divisions.

Research Without Representative Firm Assumptions. The representative firm assumption helps to reduce the myriad of economic activity to manageable proportions --and it provides confidentiality protection -- but it assumes that the behavior of all agents is alike. This assumption can be quite restrictive when there is substantial heterogeneity in the distribution of firms within a sector. The research shows that the behavior of firms and establishments varies greatly within typical sectoral classifications (it is "idiosyncratic"). This is true no matter what variables are analyzed (for example, output, employment, investment, or productivity), no matter what sectors are used to classify the data (for example, industry, size or location) and no matter what topics are examined (for example, merger policy, job turnover, business cycle analysis, research and development, energy consumption, pollution emissions, or pollution abatement expenditures). In the face of idiosyncratically behaving agents, aggregation error is introduced. Thus, a primary strand of research at CES is the evaluation of aggregation error.

Examining individual changes is necessary if particular components within an aggregate move differently from each other. An important example of the pitfalls in relying on aggregate data alone involves the relationship between productivity growth and employment change in U.S. manufacturing. In the aggregate, the manufacturing sector has increased productivity and shed workers over the past 10-15 years. The conventional wisdom is that rising productivity in manufacturing is due to firm downsizing. But the evidence suggests that this conclusion is misleading. Cross-sectoral reallocations of jobs (from manufacturing to services, for example) are small relative to reallocations within manufacturing. And within manufacturing, almost half of the productivity growth among continuously operating plants is associated with growing plants.

A similar situation can occur at the level of the firm: Even when the firm is the ultimate decision maker, data on the behavior of components of the firm, such as the plants it owns, may be required to understand the firm's performance. In recent work McGuckin and Nguyen (1995) show that the productivity enhancing effects of ownership change on manufacturing plants are obscured when firm level data is used in place of a model based on plant level observations.[16] The reason for this result appears to be that large multi-unit firms -- the kind most often examined by researchers because their data are public -- have diverse activities. Ownership changes in such companies typically involve large changes in the composition of the activities undertaken by the firm, and the use of firm level data can obscure the effects of the ownership change on the acquired properties.

---

[16] The plant-level model was estimated with a significant firm fixed effect so that characteristics of the firm are important determinants of plant performance. Thus, the failure of the model specified on firm-level responses is due to aggregation.

Both of these examples illustrate the need for longitudinal microdata for certain types of problems. It is virtually impossible to sort out the effects of ownership changes on performance unless one observes the economic unit before and after the ownership change (mergers, divestitures, leveraged buyouts, etc.). It is also impossible to compare and contrast the role of upsizing and downsizing plants without using longitudinal data to classify plants into each category.

In the absence of longitudinal data, identifying cause and effect relationships is problematic. For example, if well managed firms are the ones that adopt productivity enhancing technology, it is not possible from cross-section estimates that show intensive use of technology associated with high productivity to deduce whether the primary relationship is one of good businesses adopting technology or technology making businesses good. The problem is that both technology and productivity are correlated with an unobservable factor(s) called good management. See McGuckin, Streitwieser, and Doms (1996).

Heterogeneity Does Not Guarantee Aggregation Bias. It is important to note that the mere fact that establishments behave idiosyncratically is not sufficient to invalidate the use of aggregate data. Under certain conditions, the use of aggregate variables will introduce only negligible bias in an estimated relationship. Unfortunately, a long line of research has demonstrated that these conditions are quite restrictive. Even if interest centers only on aggregate responses to alternative policies (such as the effect of changes in pollution regulations, defense expenditure reductions, or tariff increases), responses will not be captured by simple linear functions of an average or representative firm if the responses of individual firms to changes are very different. In such cases, industry responses will be a weighted average of individual responses, and the weights can change over time.

Moreover, the effect of heterogeneity is not simply a matter of differences in firms and plants that continuously operate in an industry. Entry and exit decisions also generate aggregate industry responses that are not simple linear functions of the representative firm. Thus, the process determining survival is important for determining the proper level of data aggregation in a study. Work at CES by Olley and Pakes (1992), specifically investigating these issues, demonstrated significant errors in using aggregate data to estimate productivity relationships in telecommunications, an industry with substantial entry and exit.[17]

---

[17] Pakes and colleagues are constructing similar models to investigate the effects of policy changes (like energy tax increases) in the automobile industry. See McGuckin (1995) for a discussion and references.

This point is an important one: Recent empirical work at CES provides overwhelming evidence that not only is heterogeneity observable in cross-section data, but that plant and firm level responses to economic change are heterogeneous. In fact, most of the observed variation in the data is *within* industries. Moreover, the vast majority of this variation is *not* associated with traditional observables such as location, industry, size, age, or capital. Rather, this variation is associated with unobserved factors specific to the firm or business unit, many of which appear to be permanent attributes of the unit. Thus, linking basic operating data of the type typically collected -- shipments, value added, materials, labor, and capital -- is not sufficient for understanding the dynamics that govern economic growth. See Jensen and McGuckin (1996) for a summary of CES work. This body of work, as well as work from other sources (Bertin, Bresnahan and Raff 1992; Bresnahan and Raff 1991; Bresnahan and Ramey 1992) shows striking heterogeneity in the levels and movements of productivity, employment, growth, output, product structure, investment, and ownership change among establishments in similar markets, industries, and cohorts.

5. How Analytic Research Has Helped the U.S. Census Bureau -- Specific Examples

Aside from the importance of the body of research results illustrated in the previous section, data access has provided direct benefits to the U.S. Census Bureau. At the U.S. Census Bureau, research access has been an important factor in the development of the LRD and new data products such as job destruction and creation series that are derived from it. The research program has also helped a launch a major effort to improve economic classification systems in North America, has provided suggestions for improvements and extensions to data programs, and has increased the U.S. Census Bureau's survey business. These benefits can be anticipated for any statistical agency -- perhaps any data gathering organization -- with a microdata access program. The benefits flow from the partnerships formed with the analytic research community.

Research Databases. From the very beginning, researchers with access to the microdata have played a central role in developing the unique databases at CES. Research needs have been important factors guiding the development of these databases. In carrying out their projects, researchers provide essential contributions to the development and improvement of the databases. We illustrate with three of our databases -- The LRD, the Research and Development (R&D) Database, and the Worker Establishment Characteristics Database (WECD).

Longitudinal Research Database (LRD). The oldest and still the most used database at CES, the LRD consists of annual cost and output data on manufacturing establishments (plants) from the Census of Manufactures (1963, 1967, 1972, 1977, 1982, 1987, and 1992) and from the

Annual Survey of Manufactures (since 1972), linked to form an unbalanced longitudinal panel[18]. Over one million plants have appeared in the LRD in at least one year. Thus, the LRD is one of the most ambitious and comprehensive datasets available for the study of manufacturing. It has given rise to a large and varied body of research and policy analysis, and has formed the basis for several new statistical products (described below), the most notable of which is a series of annual (and quarterly) measures of job creation and job destruction.

The construction of the LRD was a major achievement. It grew from work in the late 1970's by the U.S. Census Bureau that was carried out under the direction of Richard and Nancy Ruggles of Yale University and funded by the national Science Foundation (NSF) and the Small Business Administration (SBA). In fact, CES was created in 1982 to facilitate research projects with the longitudinal data -- then called the Longitudinal Establishment Database (LED) -- and to improve, maintain, and update the basic data.

In the early years, work on the longitudinal panel focused on plants in continuous operation. This database of annual observations was called the time-series file. Its structure made it easy to carry out many types of analysis. For example, productivity growth was studied since plants were observed over the entire time interval. However, it soon became apparent that although balanced panels are very useful for many research issues, a balanced panel strategy for development of a longitudinal panel was inappropriate. Exits due to plant closings continually reduced the number of plants available for study. In addition to such losses from natural attrition, rotating sample designs also made big inroads in the size of the panel. Aside from sampling issues, the analysis of births and deaths has direct policy and research interest. Moreover, it is not enough to simply look at surviving plants, even if, for example, interest centers on production function elasticities.[19]

For these reasons, efforts shifted to creation of an unbalanced panel, called the LRD.[20] At the present time, CES research associates and CES regular staff are working together in a project

---

18 A balanced panel would include data for all establishments in all years. The LRD is unbalanced because establishments are born and die, and because the ASM, as a sample survey, does not cover all establishments.

19 See, for example, Olley and Pakes (1992).

20 The name LRD was chosen to distinguish the new unbalanced panel from the LED. It also served to forestall controversy and complaints from U.S. Census Bureau managers who were concerned when industry totals from the LRD did not exactly match previously published figures. Such differences arise from new edits and different imputation procedures adopted for particular projects. See McGuckin and Pascoe (1988) for a more complete description of the LRD and its history.

-- actually a series of projects -- that will ultimately lead to a Longitudinal Business Database (LBD) that will cover virtually the entire economy, not just the manufacturing sector. The additional basic data will come from the Standard Statistical Establishment List (SSEL) -- the U.S. Census Bureau's master list of domestic establishments[21] -- and the quinquennial Censuses of Wholesale Trade, Retail Trade, Service Industries, Mineral Industries, and selected Transportation Industries.

Research and Development (R&D) Database. This database includes annual data from 1972 through 1993 from the RD-1 survey, a U.S. National Science Foundation-sponsored survey on firms performing R&D in the United States. The database is well suited for studies of firms' investments in technology. Several research projects have developed and refined this database. For more detail, see Adams and Peck (1994).

Worker Establishment Characteristics Database. The WECD is the result of a match between data from the 1990 Decennial Census long form (1 in 6 sample) and the _________ (Troske 1995). The motivation for constructing this database is that theoretical models in labor economics stress the importance of employer-employee matching in determining labor market outcomes, but most empirical work relies on either worker surveys with little information about employers or establishment surveys with little information about workers. With almost 200,000 workers matched to over 16,000 manufacturing establishments where they work, the WECD is the largest worker-firm matched data set in the United States. It has been used in several research projects on measuring the tendency of larger plants to pay higher wages (Troske 1994), wage determination based on worker characteristics and productivity (Hellerstein, Neumark, and Troske 1994); and the effect of technology use on the wages and skill mix of workers in plants (Doms, Dunne, and Troske 1994).

While the WECD contains a substantial number of workers and plants, it relates to a single year, covers only the manufacturing sector, and is not a random sample of workers or plants -- it contains more data on senior male workers in large plants in urban areas. Current research focuses on creating larger, more representative, longitudinal versions.

New Data Products. ***Job Creation and Destruction.*** A new book by this title (Davis, Haltiwanger, Schuh 1996) will appear in April 1996. As part of their research, the authors created annual and quarterly time serie s for job creation and destruction from 1973-1988 for a large number of sectors of the economy. The basic data series as well as an extended series for 1988-93 are now part of the products produced by CES. WE plan to continue this series for the next few years, but we anticipate that these data series, as well as extended series covering

---

21 See U.S. Census Bureau (1979).

nonmanufacturing sectors, will soon become a regular U.S. Census Bureau data product. Aside from benefitting the U.S. Census Bureau, this work has spawned the development of similar products at other statistical agencies in both the U.S. (BLS) and around the world.

The work by Davis, Haltiwanger, and Schuh has broad implications for economic research and policy making in many areas. The issues dealt with are fundamental and cut across a wide variety of fields in both macro and micro economics. For example, the high rates of job destruction documented in virtually every sector of the economy argue strongly that workers need the flexibility to adapt to changes in the location and skill requirements of jobs. The authors also show that the job reallocation rate is countercyclical: the job destruction rate shows greater cyclical variability than the job creation rate in the United States manufacturing sector. This fact runs counter to many business cycle theories and work is now underway to extend these insights. One possibility is that this work will provide a foundation for new leading indicators to improve forecasters ability to predict economic turning points. Part of our optimism in this regard is based on past successes, but a good portion can be traced to the explicit use of a non-representative firm paradigm as part of the research strategy.

Diversification Indexes. An index of manufacturing product diversification (Gollop and Monahan (1988,1991)) has several desirable properties and uses detailed product -level data from the LRD to measure diversification at the plant and firm level for the five Census years from 1963 to 1982. The index, together with some more recent evidence (Streitwieser (1991)), shows that over the last 30 years, firms became more diversified, but plants became more specialized in the products they produce.

Advanced Technology Products Series. CES has helped the Foreign Trade Division at the U.S. Census Bureau to develop improved statistics showing the volume of trade in advanced technology products (ATP). See McGuckin, Abbott, Herrick, and Norfolk (1992). Since January 1989, the ATP series has been a part of the U.S. Census Bureau's monthly trade statistics.

Support for Improved Economic Classification. One important set of longstanding measurement issues concerns the problems the Standard Industrial Classification (SIC) system has in trying to classify economic activity. Researchers working with regular staff provided several studies in this area. For example, a number of microdata examinations examined the feasibility and design of alternative systems (Abbott and Andrews 1990, McGuckin 1992, Mattey 1993). A vital part of this work involved the product heterogeneity component of the Gollop and Monahan (1991) manufacturing product diversification index. This quantitative evidence supports qualitative judgments of experts in classification. These computer- and staff-intensive computations could not have been carried out without the LRD.

Collaboration With U.S. Census Bureau Survey Programs. Researchers develop working relationships with U.S. Census Bureau production division staff members, as well as workers in other statistical agencies. These working relationships lead to informal (and sometimes formal) consultation and advice that benefits all parties involved. Although it is difficult to document all of these effects (particularly the informal ones), the following are some major projects and surveys to which CES has contributed.

Characteristics of Business Owners (CBO) Survey. The 1982 and 1987 CBO Surveys, sponsored by the U.S. Small Business Administration (SBA) and the U.S. Minority Business Development Agency (MBDA), provide data on the demographic and economic characteristics of business owners and the economic performance of their firms. Since the CBO over samples firms owned by minorities and women, it is particularly useful for studying small businesses owned by these groups. For more detail, see Nucci (1992). CES researchers worked with the CBO survey staff on the 1987 and 1992 surveys. In particular, research with the 1982 CBO microdata -- supported in part by the SBA and the MBDA -- resulted in a variety of improvements to the 1987 survey questionnaire and sampling scheme. CES also helped to develop longitudinal linkages between the 1982 survey data and the 1987 survey universe. Extensions of these linkages to the 1992 universe have begun.

Research and Development (R&D) Survey. Research projects by James Adams and William Long, working with SuZanne Peck, resulted in the construction and documentation of a longitudinal research data set for the R&D survey (Adams and Peck 1994). This research has generated suggested improvements to the sampling scheme and questionnaire for this survey (Adams and Champion 1992). Research associates Long and Bronwyn Hall are continuing to evaluate the data by comparing R&D data (from Form RD-1) with data from a matched set of 800-1,000 companies that also file R&D data with the Securities and Exchange Commission (filed on form 10-K).

Annual Survey of Manufactures (ASM). CES has been especially interested in the ASM, which is one of the two main sources of the LRD data (along with the Census of Manufactures). CES researchers have supplemented statistical research on the ASM by pointing out certain problems of bias resulting from ASM procedures. Davis, Haltiwanger and Schuh (1991) gave a complete description of the differences between published and sample (LRD) statistics in the ASM. McGuckin and Peck (1993) analyzed the effects on published statistics of the ASM rules under which establishments' industry classifications can change. Research associates from the Federal Reserve Board (FED) have examined adjustment factors for the ASM undercount. This work is expected to lead to a joint U.S. Census Bureau/FED/Bureau of Economic Analysis project to develop new adjustment factors.

Pollution Abatement Costs and Expenditures (PACE) Survey. The PACE database covers the years 1979-93 and is fully integrated with the Longitudinal Research Database, with annual additions as data becomes available. A new report by Mary Streitwieser (1996) describes the survey design and how it has evolved over time, evaluates the suitability of the data for research, and suggests changes to the survey. Future data users, in turn, will use the report and add their own suggested changes.

Increasing the U.S. Census Bureau's Survey Business

The CES access program provides survey sponsors with a way to bring their analytic capabilities to the survey microdata, even after linking their survey data to the existing Census databases. The capability to provide survey sponsors with direct access to survey microdata linked to the LRD helped generate new survey business for the U.S. Census Bureau. In 1994, the U.S. Census Bureau conducted the *National Employers Survey (NES)* survey of over 3,000 business establishments in the United States (about half in manufacturing) and their training practices. The NES provides information on who pays for training, how training and its returns are evaluated by firms, how much is spent on training versus recruitment of already trained workers, and who provides training - outside vendors and educational institutions. Survey analysis has been carried out at both the BRDC and CES headquarters, and the survey data for the manufacturing sector have been linked with the LRD to allow investigation of the effects of training on establishments' success over time.

6. Outstanding Issues

The CES research program has demonstrated the mutual benefits that accrue when researchers and the U.S. Census Bureau work together to exploit the microdata collected in its regular programs. In spite of this impressive record, there are ways to increase these benefits.

Support for Longitudinal Panels Should be Increased. The success of many studies at CES using longitudinal data sets like the LRD has led the U.S. Census Bureau to declare longitudinal economic microdata to be one of its core strengths. This is evidence that the U.S. Census Bureau and its customers have come to recognize the value of these panels in producing information valuable for public and private policy decisions. However, survey designs at the U.S. Census Bureau have yet to reflect this. For example, many manufacturing surveys, such as the Annual Survey of Manufactures, select their samples with probability proportional to size. The five-year ASM panels before 1979 included all plants of a firm if one establishment was selected for the firm. To reduce respondent burden, the 1979 panel and following panels no longer automatically select these other plants. Thus, the ASM includes a smaller percentage of small establishments than large establishments and often does not include all of the plants of multi-establishment firms. In addition, editing and imputation routines typically use a limited number of adjacent years of data for a plant. Although this design and processing system has worked well for producing aggregate data, it is not ideal for longitudinal microdata research.

Need to Supplement Basic Census Data. Out of necessity, U.S. Census Bureau surveys are designed within limited resources to fill a variety of needs while minimizing respondent burden, and as a result much information is left out. Filling in these gaps can be accomplished in two ways: First, the basic microdata sets that form the LRD can be merged with data from both within and outside the U.S. Census Bureau. Particularly for merging internal data sets, it would be helpful if certain surveys were designed as supplements to other, more fundamental surveys, in much the same way as CPS supplemental questions are handled. Thus, for examples, surveys of energy use, capacity, pollution abatement expenditures, plant occupation distributions, and technology use would be designed as subsamples of the ASM. Also, if surveys were viewed in this way ownership changes might be more consistently carried forward across data sets.[22]

The EQW/NES training survey, sponsored by the DOE, provides another promising avenue. This establishment survey was designed to maximize the possibilities of matches with the LRD, thus providing time-series data on the matched establishments (and their firms) as well as additional cross-section information. The U.S. Census Bureau and other government agencies are now investigating other possibilities.

Need to Develop Methods of Handling Longitudinal Data -- not part of Survey Designs. Existing panel datasets such as the LRD are constructed by concatenating survey and U.S. Census Bureau data collected to serve another purpose. Linking complete business censuses offers few statistical problems. When, as is often the case, the data have been collected as part of a probability sample designed to develop a point-in-time estimate of some aggregate like GNP or industry output there are many problems in interpreting the results.

The lack of true probability designs raises tensions between survey statisticians and analytical researchers. Most panel data at CES comes from non-random designs. This is also true for linked data coming from several cross-section surveys, whether there is a longitudinal aspect or not. Such data cause many problems of representativeness that Government survey statisticians are often uncomfortable with. The problem is that such non-probabilistic datasets make statistical inference complicated. Moreover, the methodologies used by economists and other social scientists to handle such data are explicitly model based and rely heavily on

---

22 Ownership changes have been a difficult problem in developing a longitudinal panel of R&D firms and in merging LRD plant-level data with R&D survey firm-level data.

economic theory. As such they are often outside the realm of standard statistical practices used by survey statisticians, most of whom spend their professional time dealing with the problem of point-in-time probability surveys.

The answer is not simply to try and develop true panel designs. They are very costly (attrition is a real problem) and difficult to process. In the face of budget constraints, this solution is unlikely to be adopted. Nor is it necessary.

The crux of the matter involves the proper use of non-random samples. Work on methodologies for use of non-random samples is crucial. The virtual absence of truly random probability samples bearing on the most important research issues in economic policy, as well as other social science issues, is one reason the issue of the proper analysis of non-random data will not go away. In such cases, one must be content with careful documentation, good theory to guard against spuriousness, and replication of the analysis with different datasets. Replication across a wide range of non-representative samples enables more general conclusions to be drawn. In fact, this is standard practice in the social sciences (Smith 1983). But it is not just social scientists that use non-random data. So do statisticians. Non-response problems are technically equivalent to (though perhaps less severe than) those facing a researcher who has only a non-random dataset to work with. The required adjustments complicate matters, and in some cases, are controversial. But these adjustments are workable (Laaksonen 1992).

In developing panels, statistical agencies need to be aware of these problems. More importantly, they must make it a priority to bring together -- amicably and productively --analytic users of the data with survey statisticians. This is particularly important because the professional standards and goals of the two groups can be quite different. Of course, in reporting the results of analysis using data, economists and other social scientists must carefully discuss the limitations of the data and not overgeneralize the results. This is, of course, not always done by researchers and such practices raise concerns with survey statisticians whether or not the data are from a panel.

7. A Concluding Comment

The CES program has resulted in a great deal of important and innovative analytic research; new published data products, suggestions for improving U.S. Census Bureau data programs, and a large and growing list of micro databases. It is also a key input of a new data collection strategy -- collection of data for a sponsor, while providing the sponsor with access to the survey microdata, linked to other microdata bases. These benefits flow from a strategy that exploits the natural coincidence of interests between analytic subject matter researchers, the U.S. Census Bureau, and survey sponsors, including governmental units. Since the understandings of the economy and insights into its operation rest heavily on longitudinal microdta, survey value added is dependent on access to the microdata for analytic research purposes. In turn, the access program creates partnerships between the data collectors and users that generate dividends well beyond the sum of their individual contributions.

While there are still many problems and issues to tackle, based on our past experience they will be solvable if analytic researchers and statistical officials continue to work together to encourage partnerships such as those formed at CES.

References

Abbott, T.A. and S. Andrews (1990), "The Classification of Manufacturing Industries: An Input-Based Clustering of Activity," *Proceedings, Census Bureau Annual Research Conference*, pp. 26-45, *CES Discussion Paper #90-7.*

Adams, J.D. and E.J. Champion (1992), "Restructuring Research and Development Statistics at Census: A Blueprint for Change," paper presented to the Census Advisory Committee of the American Economic Association, October, 1992.

Adams, J.D. and S. Peck (1994), "A Guide to R&D Data at the Center for Economic Studies, U.S. Bureau of the Census," *CES Discussion Paper #94-9.*

Baily, M.N., E. Bartelsman, and J. Haltiwanger (1994), "Downsizing and Productivity Growth: Myth or Reality?," *CES Discussion Paper #94-4.*

* Baldwin, J. (19__)

Bernard, A. and J.B. Jensen (1994), "Exporters, Skill Upgrading, and the Wage Gap," *CES Discussion Paper #94-13.*

Berry, S., J.A. Levinsohn, and A. Pakes (1992), "Automobile Prices in Market Equilibrium," *Working Paper, Yale University Department of Economics.*

Bertin, A.L, T.F. Bresnahan, and D.M.G. Raff (1992), "Localized Competition and the Aggregation of Plant-Level Increasing Returns: Blast Furnaces 1929-1935," mimeo, Stanford University Department of Economics.

* Boyd, G., J.B. Jensen, and R.H. McGuckin (1996)

Bresnahan, T.F. and D.M.G. Raff (1991), "Intra-Industry Heterogeneity and the Great Depression: The American Motor Vehicles Industry 1929-1935," *Journal of Economic History*, Volume 51, pp. 317-331.

Bresnahan, T.F. and V. Ramey (1992), "Output Fluctuations at the Plan Level," *Working Paper #4105, National Bureau of Economic Research.*

Caballero, R. (1992), "A Fallacy of Composition," *American Economic Review*, Volume 82, pp. 1279-1292.

__________ and E.M.R.A. Engel (1992a), "Beyond the Partial Adjustment Model," *American Economic Review, Papers and Proceedings*, Volume 82, pp. 360-364.

__________ (1992b), "Microeconomic Adjustment Hazards and Aggregate Dynamics," *Working Paper #4090, National Bureau of Economic Research.*

__________ and J. Haltiwanger (1995a) "Plant Level Adjustment and Aggregate Investment Dynamics," *Brookings Papers on Economic Activity, Number 2*, pp. 1-54.

__________ (1995b), "Plant Level Adjustments and Aggregate Investment," unpublished working paper.

Davis, S. and J. Haltiwanger (1992), "Gross Job Creation, Gross Job Destruction, and Employment Reallocation," *Quarterly Journal of Economics*, Volume 107, pp. 819-864, *CES Discussion Paper #90-4.*

__________ and S. Schuh (91-1), "Published Versus Sample Statistics from the ASM: Implications fro the LRD," *Proceedings of the American Statistical Association Business and Economics Section, CES Discussion Paper #91-1.*

__________ (1996), Job Creation and Destruction, Cambridge, MA and London, England: The MIT Press, forthcoming.

Doms, M.E., Dunne, T., and Troske, K. (1994), "Workers, Wages, and Technology," mimeo, Center for Economic Studies, U.S. Bureau of the Census.

Dunne, T. and M.J. Roberts (1993), "The Long-Run Demand for Labor: Estimates from Census Establishment Data," *CES Discussion Paper #93-13.*

Dunne, T. and J. Schmitz (1992), "Wages, Employer Size-Wage Premia and Employment Structure: Their Relationship to Advanced-Technology Usage at U.S. Manufacturing Establishments," *CES Discussion Paper #92-15.*

* Gilchrist, S. (19__)

* Gollop, F. and J.L. Monahan (1988)

__________ (1991), "A Generalized Index of Diversification: Trends in U.S. Manufacturing," *Review of Economics and Statistics*, Volume 73, pp. 318-330.

Haltiwanger, J. (1993), "Proposal for Census Fellows Program," mimeo, U.S. Bureau of the Census.

Hellerstein, J., D. Neumark, and K. Troske (1994), "Wages, Productivity, and Worker Characteristics," mimeo.

Jarmin, R. (1995), "Using Matched Client and Census Data to Evaluate the Performance of the Manufacturing Extension Partnership," *CES Discussion Paper #95-7.*

* Jensen, J.B. (1996)

__________ and R.H. McGuckin (1996), "Firm Performance and Evolution: Empirical Regularities in the Microdata," mimeo.

Laaksonen, S. (1992), "Adjustment Methods for Non-Response and Their Application to Finnish Income Data, *Statistical Journal of the United Nations*, pp. 124-137.

__________ (1993), "Creation of the European Enterprise Panel: Strategies," *Proceedings, International Conference on Establishment Surveys.*

Lichtenberg, F. (1992), Corporate Takeovers and Productivity, Cambridge, MA: The MIT Press.

__________ and Siegel, D. (1987), "Productivity and Changes in Ownership of Manufacturing Plants," *Brookings Papers on Economic Activity*, Volume 3, pp. 643-673.

* Long, W.F. and B. Hall (19__)

Long, W.F. and D.J. Ravenscraft (1992), "The Financial Performance of Whole Company LBOs," *CES Discussion Paper #93-16.*

__________ (1993), "LBOs, Debt, and R&D Intensity," in The Deal Decade, M. Blair, ed., Washington, D.C.: The Brookings Institution, *CES Discussion Paper # 93-3.*

Mattey, J. (1993), Evidence on IO Technology Assumptions from the Longitudinal Research Database," *CES Discussion Paper # 93-8.*

McGuckin, R.H. (1992), "Analytic Use of Economic Microdata: A Model for Researcher Access With Confidentiality Protection," *Proceedings, International Seminar on Statistical Confidentiality*, pp. 83-97, *CES Discussion Paper #92-8.*

__________ (1995), "Establishment Microdata for Economic Research and Policy Analysis: Looking Beyond the Aggregates," *Journal of Business and Economic Statistics*, Volume 13, Number 1, pp. 121-126.

__________, T.A. Abbott, P. Herrick, and L. Norfolk (1989), "Measuring the Trade Balance in Advanced Technology Products," *Proceedings, 1989 American Statistical Association*, CES Discussion Paper #89-1.

McGuckin, R.H. and Nguyen, S.V. (1993), "Post-Reform Industrial Productivity Performance of China: New Evidence from the 1985 Industrial Census Data," *Economic Inquiry*, Volume 31, pp. 323-341.

__________ (1995a), "On Productivity and Ownership Change: New Evidence from the LRD," *RAND Journal of Economics*, Volume 26, Number 2, pp. 257-276, *CES Discussion Paper #93-15.*

__________ (1996), "Exploring the Role of Acquisition in the Performance of Firms: Is the 'Firm' the Right Unit of Analysis," mimeo.

__________ and S.H. Andrews (1991), "The Relationships Among Acquiring and Acquired Firms' Product Lines," *Journal of Law and Economics*, Volume 34, pp. 477-502, *CES Discussion Paper #90-12.*

McGuckin, R.H., S.V. Nguyen, and A.P. Reznek (1994), "The Impact of Ownership Change on Employment, Wages, and Labor Productivity in U.S. Manufacturing, 1977-1987," *Proceedings, NBER Conference on Income and Wealth, 1995, CES Discussion Paper #95-8.*

McGuckin, R.H. and G. Pascoe (1988), "The Longitudinal Research Database: Status and Research Possibilities," *Survey of Current Business*, Volume 68, pp. 30-37, *CES Discussion Paper #88-2.*

McGuckin, R.H. and S. Peck (1993), "Manufacturing Establishments Reclassified Into New Industries: The Effect of Survey Design Rules," *Journal of Economic and Social Measurement*, Volume 1, Number 2, *CES Discussion Paper # 92-14.*

McGuckin, R.H., M.L. Streitwieser, and M.E. Doms (1995), "The Effect of Technology Use on Productivity Growth," mimeo, CES.

Millward, N. (1993), "How British Labor Economists Benefit from Establishment Surveys," unpublished working paper.

Nucci, A.R. (1992), "The Characteristics of Business Owners Database," *CES Discussion Paper #92-7.*

Olley, G.S. and A. Pakes (1992), "The Dynamics of Productivity in the Telecommunication Equipment Industry," *Working Paper 3977, National Bureau of Economic Research, CES Discussion Paper #92-2.*

* Phillips, G. (19__)

Smith, B.T. (1983), "On the Validity of Inference from Non-Random Samples," *Journal of the Royal Statistical Society*, 146, 4, pp. 394-403.

Streitwieser, M.L. (1991), "The Extent and Nature of Establishment Level Diversification in Sixteen U.S. Manufacturing Industries," *Journal of Law and Economics*, Part 2, pp. 503-534, *CES Discussion Paper #90-8.*

__________ (1996), "Evaluation and Use of the Pollution Abatement Costs and Expenditures Survey Micro Data ," *CES Discussion Paper #96-1.*

Theil, H. (1954), Linear Aggregation of Economic Relations, Amsterdam: North Holland Publishing Company.

Troske, K. (1995), "The Worker-Establishment Characteristics Database," *CES Discussion Paper #95-10.*

__________ (1994), "Evidence on the Employer Size-Wage Premium From Worker-Establishment Matched Data," *CES Discussion Paper #94-10.*

Zakrajsek, Egon (1995), "Retail Inventories, Internal Finance, and Aggregate Fluctuations: Evidence from Firm-Level Panel Data," *CES Discussion Paper #95-9.*

EXPLORING THE LONGITUDINAL ASPECT OF PME EARNINGS DATA: SIMPLE NON-PARAMETRIC EVIDENCE

Marcelo Neri.

(VERY PRELIMINARY AND INCOMPLETE VERSION)

ABSTRACT:

The main purpose of this paper is to exemplify through specific aplications some basic problems related to the use of longitudinal information on earnings extracted from Pesquisa Mensal do Emprego (PME). The paper assesses problems and advantages found using PME to study earnings short-run movements in a disaggregated manner. In particular, the combination of large sample sizes with the possibility of following the same dwelling through short periods of time provided by PME allows us to estimate the cross-sectional distribution of longitudinal statistics based on observed earnings (i.e., monthly and annual rates of variation at individual and family levels). The visual analysis of the distributions of earnings variations proves useful in highlighting a series stylized facts that may guide future empirical work.

EXPLORING THE LONGITUDINAL ASPECT OF PME EARNINGS DATA: SIMPLE NON-PARAMETRIC EVIDENCE

Marcelo Neri[1][2]

1. Introduction

High inflation makes the recent Brazilian experience a good laboratory to study earnings short-run dynamics. Moreover, the availability of monthly surveys on labor markets across the main Brazilian metropolitan regions during the last 15 years allows to study earnings short-run dynamics at a disaggegated level. In particular, the combination of large sample sizes with the possibility of following the same worker through short-periods of time allows us to estimate the cross-sectional distribution of longitudinal statistics based on observed earnings (ie., monthly and annual rates of variation). The analysis of the cross-sectional distribution of longitudinal statistics generated from actual earnings data provide visual evidence on the heterogeneity of earnings adjustments patterns.

This paper has two objectives: First, to assess some basic problems related to the use of longitudinal information extracted from Pesquisa Mensal do Emprego (PME). Second, to provide evidence

---

1 From UFF and PNPE/IPEA.

2 I would like to thank Alexandre Pinto for superb research assistance. All remaining errors are my own.

on the patterns of earnings adjustments adopted during the recent Brazilian experience. The main characteristic of the methodology applied here is a high level of disaggregation in the analysis of individual and family earnings variations. The idea is to take into account existing heterogeneity on earnings adjustments patterns by plotting the cross-sectional of earnings variations.

Most of the paper is devoted to providing an assessment of the difficulties found when one attempts to capture earnings short-run movements. It discusses advantages and disadvantages found using some of the data sets available in Brazil. This discussion ranges from the earnings concepts used in the surveys to the level of aggregation used in the analysis. The problems covered in the paper include non-reponse biases, attrition, selectivity biases and measurement error on reported earnings.

The paper is organized as follows: The first five sections discusses data problems related to the use longitudinal information on earnings dynamics extracted from PME. The problems include non-reponse biases, attrition, selectivity biases and measurement error on reported earnings. The seventh section reviews the previous Brazilian time-series evidence on the impact of wage regulation on market wages and points out a few heterogeneity problems intrinsic to this aggregate approach. This section provides an example of advantages and disadvantages of using longitudinal information at individual level. The example explored is related to the quantification of the impacts of the Brazilian Wage Law on earnings short-run dynamics. The eight section assesses the compliance with the Wage Law. The next section assesses the effectivity of the Wage Law. The tenth section evaluates the operation of official indexation clauses that prescribed regressive wage adjustments. The last section presents the main conclusions of the paper and discusses extensions.

2. An Overview of Data Issues

Pesquisa Mensal do Emprego (PME) will be the main source of data. During the 1982-95 period, PME sampled monthly an average of 44 thousand dwellings in the six largest Brazilian Metropolitan Areas. PME uses a rotating panel methodology similar to the one adopted in the US Current Population Survey (CPS). In particular, PME sampling scheme attempts to gather information on the same dwellings at months t, t+1, t+2, t+3, t+12, t+13, t+14 and t+15. One advantage of PME over CPS is that it collects earnings information during all eight observations of the sampling scheme, while CPS gathers earnings information only at the first and the fifth interview. Longitudinal information was generated, specially for this study, by concatenating information on the same individuals and the same families at different points in time. The analysis here uses two types of longitudinal samples: a) individuals/dwellings that were followed during all eight observations of PME rotating panel scheme. b) individuals /dwellings that were successfully concatenated during at least one of the two groups of four consecutive observations. In the case the individual /dwelling is observed all eight times, each group of four consecutive observations is treated separately.

The longitudinal aspect of individual and family earnings data extracted from PME will provide the basic empirical evidence on the patterns of earnings variations. The empirical analysis will be based on two longitudinal statistics: a) the ratio between nominal earnings one month apart (i.e., one plus the one month earnings nominal rate of variation): RR_1, for short. b) the ratio between real earnings one year apart (i.e., one plus the 12-month earnings real rate of variation): RR_{12}, for short.

In order to compensate for the number of observations lost during the concatenation process, the procedure adopted here was to use all earnings information available from all concatenation groups of individuals present in any two months under analysis. For example, in order to calculate the RR_1 statistic between months t and t+1, all concatenation groups of four consecutive observations started in months t-3, t-2 and t-1 will be used. Similarly, to calculate the RR_{12} statistic between months t and t+12, all concatenation groups of eight observations started in months t-3, t-2, t-1 and t will be used.

The empirical work done here is based on the analysis of the cumulative distribution function of RR_1 and RR_{12} statistics in different time periods and for different groups of individuals/dwellings (e.g., divided by initial earnings level, by metropolitan regions and so on).

The smoothness of the CDF used here makes the estimation process much easier than in the case of density estimation. Nevertheless, given the requirements of the CDF estimation in terms of number of observations, the number of interactions between different sub-groups will be kept at a minimum. In sum, the visual approach adopted here may be viewed as a first approximation to the evaluation of the impacts of the Wage Law on earnings dynamics. It is inconclusive by nature but it may be helpful to present some guide lines to future empirical work.

3. Attrition and Selectivity Bias

One should account for attrition and selection biases introduced in the final longitudinal samples. These problems may be serious even if the analysis is to be restricted to the cross-sectional aspect of PME, since not all dwellings in the sample originally selected ends up being interviewed. Graph 2.1 illustrates, the number of dwellings originally selected and those actually interviewed month by month from January 1982 to October 1995. Graphs 2.2 to 2.4 identify the main filters on dwellings. Graphs 2.2 and 2.3 present the evolution of the proportion of successful interviews, the proportion of refusals, the proportion of nonexistent dwellings, and the proportion of dwellings occupied by others. Graph 2.8 present the distribution across metropolitan regions of planned and observed dwellings.

Note that, despite the non-response rate being on average 79% (see, graph 2.2), at each point in time the proportion of dwellings which are not interviewed at least once during the eight (four consecutive) observations of PME rotating panel scheme can be much higher. Graphs 2.2 show that on average 43%(73%) of the dwellings of the set of successful interviews (ie., the 'original' sample) remain in the corresponding eight (four consecutive) observations of the constructed longitudinal samples.

Table 2.1 provides a comparison of basic demographic and economic statistics corresponding to the 1982-95 period for the group of working age individuals belonging to the following samples: a)PME original sample; b)first observation of the longitudinal samples with four observations; c)first observation of the longitudinal samples with eight observations. The analysis does not show any major differences in demographic and economic aspects of each sample. As table 3.1 shows, the profiles of the cross-sectional and longitudinal samples of working age employees also do not present significant differences.

4. PME in 1980-82

During 1982, there was a major revision in PME questionnaire, the census sectors used to select the rotating samples were updated using the 1980 census. The comparison between attrition rates from the PME samples covering the periods from 1980-82 and 1982-95 shows a major improvements in statistics related to non-response biases and attrition: 1)proportion of dwellings successfully interviewed rises (graphs 2.2 and 3.1). 2)proportion of dwellings and individuals with four consecutive observations rises (graphs 2.6 and 3.4). 3)proportion of dwellings and individuals with eight observations also rises (graphs 2.7 and 3.5).

There were also introduced in PME questionnaire during 1982, items related to schooling, payments' frequencies, occupation and the possession of a working identification from the labor ministry (carteira de trabalho). Despite these major improvements, PME questionnaire lost two items that are specially useful to the work at hand:

a) the question if the employee kept the same job during the previous month.

b) the normal earnings concept which excludes the 13th wage and the payments for extra hours.

Items a) and b) will, respectively, help us to restrict the analysis to individuals that were continuously employed at the same job and to evaluate the extend of measurement error on contractual earnings. These issues are important to study the impacts of the Brazilian Wage Law on earnings short-run dynamics (i.e. our first application). We believe that these are good reasons to explore PME 1980-82 sample[3].

5. Continuously Employed Workers

The purpose of this section is to study how to restrict the analysis to the group of continuously employed individuals. This restriction is important to study the impacts of the Brazilian Wage Law on earnings short-run dynamics for at least two reasons: first, except for the minimum wage clause, the wage policy does not regulate the earnings dynamics of workers moving between jobs. For workers above the minimum wage threshold, firms can lower the

---

3 Tables 2.2 and 3.2 present summary statistics of different samples extracted from PME 1980-82.

wage bill through turn over. Another reason to focus on continuously employed workers is to avoid comparing earnings data of individuals that started the job after the month started. That is, we would like to use earnings data refering to a full working month.

Until the major reformulation occurred in 1982, PME questionnaire had a straight question whether employees kept the same job during the previous month. This question allows to test with precision different filters that attempt to restrict the analysis to continuously employed workers. From 1982 onwards, we considered as continuously employed those employees who had positive effective earnings4 and kept the same sector of activity during all observations of the longitudinal samples. Graphs 4.1 to 4.5 allows to compare the size of the samples for alternative filters chosen.

Graph 4.5 shows that during the 1980-82 period and within the group of active age individuals observed four times, 38.04% were workers that did not go through an observable job change. Alternatively, graph 4.5 shows that imposing the filter of positive effective earnings, of constancy of sector of activity and of working class during all observations this statistic corresponds to 38.23%. This value is reasonably close to the ones found with the straight question on job change. However, if we impose all requirements simultaneously (ie., employees with positive effective earnings, same sector of activity and non-reported job changes) during four consecutive observations, the

---

[4] Rates of variation, that will be extensively throughout the paper, are only well defined for strictly positive variables. However, as section 4.D.1 argued the filter on positive effective earnings is also useful to restrict the analysis to the universe of continuosly employed individuals.

proportion of active age individuals drop four percentile points to 32.09%. Assuming that the straight question on job change is correct, this number may be viewed as a first approximation to classification errors related to the filter of continuously employed individuals adopted here for the sample from 1982 onwards.

Graph 4.6, provide information on the proportion of the concatenated sample on working age individuals that were classified as continuously employed from 1982 onwards (ie., employees with positive effective earnings and same sector of activity during four consecutive observations). The key figure is that according to our definition, 19.3% of the active age population was considered on average continuously employed workers during four consecutive observations.

In terms of the sample of eight observations, given the eight-month interruption between the fourth and the fifth observation of PME rotating panel scheme, it is not possible to assure that individuals were continuously employed during the whole sixteen months period. The way to minimize this problem was to impose the most stringent non-mobility conditions during the eight observations that each individual in the sample is actually observed. According to the definition used before (ie., employees with positive effective earnings and same sector of activity during four consecutive observations), on average 13% of active age individuals observed eight times were treated as continuously employed workers during the 1982-95 period.

Table 4.1 provide a comparison of basic demographic and economic characteristics of the following samples of active age individuals taken out of PME rotating panel scheme:

a) longitudinal sample of individuals that were employees during at least one month out of eight observations.

b) longitudinal sample of individuals that were employees during at least one month out of four consecutive observations.

c) longitudinal sample of individuals that were continuously employed during all months of each group of four consecutive observations.

d) longitudinal sample of workers that were continuously employed during all months of each group of eight observations.

The analysis of table 4.1 reveals that as we move in the direction of the sample with more stable jobs: the sample becomes older and more educated, the proportion of males and of heads both rise, the share employed in manufacturing rises and the share of employees with legal contracts also rises. The number of hours worked per week is the only item that does not present major differences across different samples.

Table 4.2 refering to PME 1980-82 sample does not present major differences with respect to the data refering to the period from 1982 onwards, presented in table 4.1. The only exception is that during 1980-82, the share employed in the manufacturing sector declines as jobs become more stable.

6. Measurement Error on Earnings[5]

The effective earnings concept used in PME from 1980 onwards presents some problems to evaluate the influences of the Wage Law on market wages because it includes other items besides contractual earnings such as the payments for extra-hours, payments related to profit sharing and the so-called 13th wage. However, as it was mentioned before, during its first two years of operation, PME questionnaire included besides the effective earnings concepts, the normal earnings concept which is closer to contractual earnings. The comparison between the two concepts

---

[5] DiNardo, Fortin and Lemieux (1995) provide an example of the importance of using the right earnings concepts.

allows us to understand some limitations of the effective earnings concept that was used alone from 1982 onwards.

Graph 5.1 illustrate the differences between normal and effective of active age employees earnings by plotting their CDF for São Paulo during January 1981 (NOTE: the axis are inverted in relation to the usual plot of distribution functions).

6.A. Number of Zeros

A first difference between effective and normal earnings presented in graph 5.1 is the number of zero answers. In the universe of employees during the 1980-82 period around 7% of those employees that reported positive normal earnings reported zero effective earnings. However, if we restrict the analysis to employees that are heads of the household this statistic rises to 10%. Moreover, the probability of reporting zero normal earnings conditioned on reporting positive effective earnings is zero in both samples.

The difference in the number of reported zero earnings between the effective and the normal concepts is at least partially explained the way the effective earnings question was made: How much did you earn last month in the job that you have today? This question induces employees that moved between jobs during the previous month to report zero earnings. Graph 5.2 confirm this fact for January 1981, around 83% of the sample of employees that declared moving between jobs in the month previous to the interview reported zero effective earnings and positive normal earnings. As graph 5.3 shows, this statistic drops to zero in the sample of continuously employed workers where both distributions are very alike. In other words, the restriction that effective earnings are always positive in a given longitudinal sample may be helpful to restrict the analysis to the universe of continuously employed workers[6,7].

The comparison between normal ans effective earnings CDF's during December 1980 shown in Graph 5.4 reveals sharp differences. These differences occur only in December and on a smaller scale during November (not shown here). These differences can be better analyzed in terms of the cross-sectional distribution of longitudinal statistics that will be used throughout the paper.

6.B. the RR_1 Statistic

There is also a timing difference between the two earnings concepts used in PME during 1980-81: while the normal earnings concept reflects nominal values during the current month of the interview, the effective earnings concept reference period is the month previous to the interview. Graphs 5.1 to 5.24 in appendix B, compare month by month the CDF of the RR_1 statistic according to the effective and to the normal earnings concepts during the period ranging from March 1980 to February 1982. These statistics refer only to continuously employed workers extracted from the longitudinal sample with four consecutive interviews. These statistics are also adjusted for the mentioned timing difference between both earnings concepts. The main results are:

a) The cumulative distribution of the RR_1 statistic of each concept does not change much when the same months of different years are compared (eg., March 1980 and March 1981).

b) The cumulative distribution of the RR_1 statistic using the

---

7

Graphs 4.1 and 4.2 show that within the normal concept, there are no differences between the proportion of employees with always positive earnings in relation with the proportion of those with always positive earnings and hours. The inter-concept comparison reveals that the number of zero answers is greater using the effective concept.

effective and the normal earnings concepts are very similar during the following months: February, March, June, July, August and September. The only difference noteworthy during these months is that the normal RR_1 statistic presents smaller absolute variations at the extremes of the distributions. During April 1980, October 1980, April 1981 and May 1981 the cumulative distributions of the RR_1 statistic present some differences in the upper part of the distribution.

c) The sharp differences of the distributions observed during November, December and January are influenced by the 13th salary. This last issue is analyzed in detail next section.

6.C. November, December, January and the 13th Salary

The 13th salary was initially thought and is usually perceived as a Christmas bonus but it can be paid in one or more payments until December every year. According to the law, the 13th salary equals to the nominal contractual wage refering to November. The CDF of the effective earnings RR_1 distribution first-order stochastically dominates the equivalent distribution based on the normal concept during November and December while the reverse is true during January. The differences between the effective and the normal concepts distributions provide a proxy for payments that are not included in contractual wages like those associated with the 13th salary. The observed differences between these two concepts suggest that some caution should be taken with the effective earnings rate of variation during November, December and January.

6.D. the RR_{12} Statistic

PME rotating panel scheme does not yield a matching rate of 50% across any 12 months period as in CPS. All concatenation groups of four consecutive observations that started in odd years are designed to match 100% of the sample 12 months ahead. On the other hand, concatenation groups of four consecutive observations that started in odd years are expected to present a null matching rate with respect to the sample observed 12 months ahead. As a result, the RR_{12} statistic can only be calculated between two months starting in an odd year.

Graphs 5.25 to 5.36 in appendix B compare the CDF distributions of the RR_{12} statistic according to which the denominator was observed in the period from January 1980 to January 1981. The timing convention of the graph corresponds to the initial month which earnings data (ie., the numerator of RR_{12}) of the earnings data used was observed.

The different CDF's of the RR_{12} using the effective and normal earnings concepts are much more alike than the RR_1 CDF's. This is true for inter-concepts comparisons and for temporal comparissons. This similarity between the RR_{12} CDF's give us confidence to use their CDF's based on effective earnings from 1982 onwards.

7. The Effects of the Brazilian Wage Law on Earnings Dynamics: An Application Using the Longitudinal Aspect of PME Earnings Data

The purpose of the remaining part of the paper is to exemplify through specific some problems and potentialities related to the use of the longitudinal aspect of PME data. The example used here apply the methodology described so far to quantify the influences of the Brazilian Wage Law on earnings short-run dynamics. The main characteristic of the approach methodology is a high level of disaggregation in the comparison between observed earnings and artificial earnings data generated from minimum requirements imposed by wage regulation schemes. The idea is to take into account existing heterogeneity between agents vis-a-vis specific wage adjustments clauses of the law.

The empirical analysis benefits from the large variety of official wage indexation rules experienced in Brazil from December 1979 to December 1992. Appendix A gives a brief review of institutional aspect of Brazilian the labor market. The analysis gives a special attention to the first half of the sample when the Brazilian Wage Law prescribed infrequent and staggered wage adjustments that varied across wage levels. These characteristics induce heterogeneity problems that the methodology developed here proposes to address.

The quantification of the Brazilian Wage Law departs from three basic questions: a)Do market wages comply with the Wage Law? In the sense of satisfying the minimum requirements imposed by the law. b) Is the Wage Law binding? In other words do market wages behave exactly as the minimum requirements imposed by the Wage Law? c)How is the Wage Law evaded? What are the main patterns of wage adjustments by those not following exactly the law prescriptions.

8. The Brazilian Wage Law Impact on Average Wages and Aggregation Biases

The Brazilian Wage Law prescriptions can be summarized by an institutional wage variable. The institutional wage represents the path of wages were the indexation clauses of the law exactly fulfilled (i.e., a binding Wage Law). It is an index that combines indexation clauses and ex-post price indexes. The previous institutional wage literature [Gonzaga (1988) and Camargo (1990)] has followed an aggregated time series approach to assess the influences of the law on labor market outcomes. Since the Wage Law prescriptions are specific to settlement dates and wage levels, these previous studies assumed a given joint distribution for these variables in order to generate an average institutional wage time series. The institutional wage is placed along with other explanatory variables in regressions against observed market wages. However, the institutional wage literature was not able to provide a clear assessment about the effectivity of the Brazilian Wage Law.

The time series approach that dominates the institutional wage literature tends to average out measurement error found in individual answers. However, it also averages out some of the true variance to be explained. The combination of infrequent nominal wage adjustments and staggering generates a discrepancy between individual and aggregate earnings processes. Infrequent adjustments produce individual real earnings processes with a saw-toothed pattern, while a staggered structure of wage adjustments tends to smooth out these patterns at an aggregate level. Graph 1 illustrates this point replicating the legal wage adjustment prescriptions from 1966 to 1985. The smooth line corresponds to an average wage real index of all wage settlement categories assuming a uniform distribution of settlements dates throughout the year. The sawtooth patterned line corresponds to the wage index of a representative worker from the January settlement date group.

The solution to the aggregation bias on earnings variability, illustrated on graph 1, within the time series approach is to use data of specific wage settlement groups. A proxy for that may be found with sectoral data where one should expect a more homogeneous wage settlements distribution.

Taking for granted the information on wage settlement dates, the data should account for other sources of heterogeneity. The Wage Law during various sub-periods prescribed wage variations that differed across wage levels. This means that to test the Wage Law one should know the evolution of the cross-sectional distribution of wages.

As the two previous paragraphs pointed out comparisons between observed wages and institutional wages require controls for wage levels and for settlement dates. This suggests difficulties to address our basic questions on a pre-fabricated time series environment. One should be able to adapt observed wages' time series to the specificities of the law. An additional complication is that firms are allowed by the law to fire a worker and hire another one with lower wages (ie., if they are paid above the minimum wage, of course). In other words, turn over may be used legally as a labor cost saving device. Since, the law only regulates the wage dynamics of continuously employed workers at the same job, the analysis should be restricted to this group of individuals. This means that if the analysis is to be done at a time series level, a non-mobility condition based on longitudinal information should be imposed on the sample of observed wages.

A more fundamental source of heterogeneity that makes the time series approach incapable to address the questions at hand is that it aggregates information of those that are evading the law from those that are complying with it. For example, in the case when there is a large proportion of workers well apart but symmetrically distributed across the mean requirements of the law, empirical tests based on the means will lose power to reject the null hypothesis that the law is not obeyed or the null that the law requirements are not binding. In the presence of substantial heterogeneity, the comparison of the mean of the law requirements and the mean of the observed data provides scarce information on the actual compliance with the law and the actual effectivity of the law. One should compare, as well, broader measures of the distributions of artificial and actual wage adjustments, or preferably compare the distributions themselves.

The problems with the average institutional wage approach discussed in this section are listed below:

a)heterogeneity by wage level,

b)heterogeneity by settlement date,

c)aggregation bias on earnings variability,

d)work only with continuously employed workers,

e)heterogeneity related to the compliance with the law,

f)heterogeneity related to how binding is the law for those that complied with it,

All the problems posed above favor the use of flexible data sets where one can adapt the data to the specificities of the Wage Law. While the two first problems can be dealt within the time series environment, the remaining problems require the use of longitudinal information at an individual level. However, the advantage of longitudinal information has its costs. First, the longitudinal earnings information possible to be extracted in Brazil comes from concatenated samples from rotating panels subject to selectivity biases. Second, measurement error is not averaged out. This is specially problematic since the earnings concept used in this paper is influenced by the payment of extra hours and the 13th wage which produce extra noise on the earnings measure available. These problems intrinsic to the use of longitudinal information closes the check list of problems to evaluate the effects of the Wage Law on earnings that are addressed in this paper:

g)Selectivity Biases,

h)Measurement Error on Earnings.

9. The Compliance with the Wage Law

The Wage Law has two main characteristics: first, it fixes a minimum wage level. Second, it fixes minimum earnings variations for all employees according to earnings levels and settlement dates. The present section evaluates the extend and the patterns of compliance with the second aspect of the Wage Law. The basic statistic to be analyzed here is the ratio between real earnings observed one year apart: RR_{12}.

A first advantage of the RR_{12} statistic is to allow comparisons across wage indexation regimes with different frequencies. There is one official settlement date per year and by no coincidence, the frequency of official wage adjustments were always fractions of this yearly cycle (e.g., 12/12, 6/12, 4/12, 3/12, 2/12 and 1/12). As a result, within a fixed frequency of wage adjustments, independently of the settlement date, any 12 months period covers the same number of automatic legal wage adjustments. Furthermore, the numerator and the denominator of the RR_{12} statistic will be in the same phase of the wage adjustment cycle, despite the existing settlement dates heterogeneity among individuals. In other words, this statistic allows us to abstract from individual workers information on settlement dates which is hard to come by[8].

More generally, the RR_{12} statistic is not affected by any event that keeps a regular yearly cycle like the payment of the 13th wage, official wage adjustments at the settlement date or even a stable pattern of anticipations of wage adjustments outside the settlement date. The similarities between the RR_{12} distributions calculated using the effective and the normal earnings concepts presented in Graphs 5.25 to 5.36 in appendix B, illustrate this point. As the analysis of the RR1 statistic in Graphs 5.1 to 5.24 showed in appendix B, these earnings concepts present different seasonal patterns.

On the negative side, the RR_{12} statistic does not provide the whole picture of how earnings are affected by inflation because it

---

8 This is true even in the case of the state dependent indexation rule adopted during 1986-87 (ie., the 20% trigger), since the residual inflation (the part that did not reach the 20% figure) would be compensated for at each workers group settlement date.

abstracts from earnings short-run adjustments. For given initial wages and RR_{12}, the more frequent are wage adjustments the higher is the present value of earnings during the 12-month interval. Another problem of the 12-month interval is that it encompasses during the period under analysis more than one official wage adjustments cycle. So for example, a firm may evade the law by not giving required automatic wage adjustments at some stage of the 12 months period. If the firm reverts this movement by over-indexing wages at some stage of the 12 months period, the RR_{12} statistic will not capture the illegal wage adjustments practices employed in the latter period. In sum, the RR_{12} statistic only assures a necessary condition of compliance with the Wage Law.

Graph 6.1 incorporates two parameters representing the prescriptions of the Wage Law to the graph RR_{12} CDF based on effective earnings and normal earnings, discussed in the previous section. The example of graph 6.1 refers to May 1980 that is according to the timing convention adopted, it represents changes occurred in earnings during the period from May 1980 to May 1981.

The upper horizontal line of graph 6.1 corresponds to the maximum value of the RR_{12} requirement imposed by the Wage Law over a universe of 60 simulations. These simulations of institutional wages combine a set of five initial wage levels (1, 2, 5, 12 and 25 minimum wages) with 12 possible settlement dates. Similarly, the lower horizontal line in graph 6.1 corresponds to the minimum value of the RR_{12} Wage Law requirement over the same set of institutional wages and time intervals. The part of the distribution under the lower horizontal line would be evading the Wage Law adjustment prescriptions, if the assumption that workers are continuously employed during the 12-month period started in May 1981 is valid. According to graph 6.1, the hypothesis of compliance with the Wage Law cannot be rejected for 18% and 23% of employees according to normal and effective earnings concepts, respectively.

Graph 6.2 present the evolution of the maximum and of the minimum RR_{12}

prescriptions of the Wage Law during the 1979-82 period (ie., the evolution of the lower and the upper horizontal lines of graph 6.1 across different months)[9].

Graph 6.3 synthesizes the information presented in Graphs 5.25 to 5.36 in appendix B and Graph 6.2, through an index of the noncompliance with the Wage Law 12 months adjustments prescriptions calculated for both earnings concepts. That is, the proportion of actual earnings variations observed which are below the lowest of all 60 adjustments requirements simulated at each month. According to the information contained in graph 6.3 for normal earnings, an average of 19% of the continuously employed workers appears to be breaking a necessary condition to comply with the Wage Law 12 month adjustments requirements. This statistic raises to 23% when the effective earnings concept is considered. Moreover, graph 6.3 shows a downward trend in the index of the noncompliance with the Wage Law during 1980.

Graph 6.4 replicates the information contained in graph 6.3 for the sample of employees in general and for the sample of employees working in the manufacturing sector, using the effective concept for all possible months between 1980 and 1984 where the RR_{12} statistic can be calculated. The analysis reveals that the average noncompliance index is slightly smaller for employees in manufacturing (17% against 18% for overall employees). The proportion of workers not complying with the Wage Law presents a downward trend during the period under analysis. However, considering the fall of the institutional wage real vale overtime, it is not clear what is causing the fall in the noncompliance rates observed, whether wages are becoming more adjusted to the Wage Law or vice-versa.

10.A. The Effectivity of the Wage Law

This section evaluates the extend observed wages follow exactly the

---

9 For a detailed description of the procedures adopted in these simulations see Jovita, Barros and Neri (1996).

prescriptions of the Wage Law. In other words, it investigates how binding is the Wage Law. This exercise consist in a more decisive step to evaluate the impact of the law on earnings short-run dynamics.

The analysis will focus on the ratio between nominal earnings one month apart: RR_1. This statistic allows us to capture anticipations of wage adjustments between official adjustments dates. However, the analysis of RR_1 is somewhat more complex than the analysis of the RR_{12}, developed last section because it is subject to problems related to the heterogeneity of individuals across wage adjustments dates.

The six-month adjustments' frequency during the 1980-86 period implies a bimodal probability distribution function (pdf) of the ratio of institutional wages nominal monthly adjustment factors (RR_1). The first part of the RR_1 pdf would be an atom at one, corresponding to the zero nominal adjustments for individuals that are not in their respective adjustments dates. In terms of the CDF of the RR1 statistic, there will be a plateau at RR1 equals to one. The remaining part of the institutional wage RR1 pdf would correspond to the wage adjustments prescriptions of the law. Since during the 1980-86 period, official wage adjustments prescriptions varied across earnings brackets, there will be some dispersion of the institutional wage RR_1 pdf in the upper tail of the distribution. This dispersion will depend on the degree of differentiation of wage adjustments prescribed by the law for workers located in their respective wage adjustments dates. As a consequence, the CDF of the institutional wage RR1 statistic, will not necessarily present a plateau at the upper tail of the distribution.

In order to capture the existing dichotomy between adjusting and not adjusting institutional wages and the dispersion of wage adjustments prescriptions for the latter group, the analysis of the RR_1 statistic will use three horizontal lines representing the Wage Law adjustment prescriptions in a given month:

a) a lower horizontal line indicating the minimum wage adjustments assumed by the institutional wage nominal monthly factors across 60 simulations (ie., five wage brackets times 12 settlement dates (see previous section, for details)). During the 1980-86 period, this line will be typically situated at RR1 equals to one, corresponding to the majority group with fixed nominal institutional wages across two consecutive months under analysis.

b1) an intermediary horizontal line corresponding to the smallest legal wage adjustment prescription for those situated in an official wage adjustment date. This line corresponds to the smaller RR_1 statistic of the group located in an official wage adjustment date. Given the regressive wage adjustments prescriptions of the law during the 1980-86 period, the values assumed by this line will coincide with the RR_1 statistic of the group located in a wage adjustment date with higher initial wage levels (in our simulations 25 minimum wages).

b2) an upper horizontal line corresponding to the highest wage adjustment prescription of the Law. During the 1980-86 period, this line corresponds to the group earning one minimum wages that are located in an automatic wage adjustment date. Graph 7.1 presents the evolution of the values representing these three lines across the 1980-92 period.

Graphs 7.2 presents the distribution of normal and effective earnings RR1 statistic for May 1980 coupled with the three lines mentioned above, representing the Wage Law prescriptions. The comparison between artificial and actual earnings behavior reveals the following patterns of wage adjustments during May 1980:

a) 20% (18%) of observations present a reduction in nominal earnings, according to the effective (normal) concept.

b) 28% (28%) of observations present constant nominal earnings. These focal points correspond to the prescription of the Wage Law for the majority of the labor force.

c) 24% (25%) of observations present earnings variations in excess of the minimum adjustments prescriptions of the law for those located in a wage adjustment date.

d) 20% (19%) of observations present earnings variations in excess of the maximum adjustments prescriptions of the law for those located in a wage adjustment date.

e) around 28% (29%) of the observations present variations in between the Wage Law no adjustment prescription and its minimum earnings adjustment prescription. These numbers could be interpreted as a proxy of anticipations of wage adjustments outside official wage dates.

f) finally, around 4% (6%) of the observations present earnings variations in between the minimum and the maximum values prescribed by the law for those located in a wage adjustment date (ie., a direct consequence of items c) and d)). These figures plus the universe of individuals with zero wage adjustment (ie., 28% (28%) of item b)) correspond to 32% (34%) part of the sample which the hypothesis that the Wage Law is exactly followed cannot be rejected.

Graphs 7.3 to 7.8 present the evolution of items a) to f) above for effective and normal earnings during the period ranging from March 1980 to February 1982. The relative importance of the different ranges chosen from the distribution of monthly wage adjustments present different seasonal patterns across the two earnings concepts used.

The normal earnings RR_1 presents a smoother path throughout the year, with the exception of a more pronounced concentration of positive wage adjustments during May and November, specially for those ranges above the minimum adjustment prescription of the law (see, graphs 7.4, 7.5 and 7.6, respectively).

Effective earnings present a richer variety of wage adjustment patterns which includes those observed for normal earnings plus other patterns observed during November, December and January. For example, the share of workers with constant nominal earnings fall sharply during these months (graph 7.3). The increase of the share with negative nominal adjustments (graph 7.5) during January and the increase of the share with higher than the maximum wage adjustments prescriptions of the law during December and November (graph 7.6) are consistent with the operation of transitory shocks on earnings due to the payment of the 13th wage during

November and December.

10.B. Effectivity Conditioned on Compliance

The question "How binding is the Wage Law? is more precisely answered if the analysis is conditioned on the compliance with the Law. Or at least conditioned in the fact that some of the law minimum adjustments requirements were fulfilled. This is a way to reduce the problem of heterogeneity of individuals across official wage adjustments dates.

The first step of the analysis was to divide the RR_1 distribution statistic in two groups depending on the non-compliance with the Wage Law 12 months prescriptions, as analyzed in the section 5. That is, the RR_1 distributions of individuals followed all eight times of PME sampling scheme will be split in two parts according to the RR_{12} statistic: a)the group whose RR_{12} are below the minimum requirements (ie., those which the hypothesis of non-compliance with the law can not be rejected). b)the group which RR_{12} is equal or above the Wage Law minimum requirements[10]. The analysis of the RR_1 statistic for normal earnings during 1980, presented in graphs 8.1 in the end of the paper and graphs 8.12 to 8.24 in appendix B reveals that:

a) as expected, the RR_1 distribution of the non-compliers is first-order stochastically dominated (non-strictly) by the RR_1 distribution of those which the hypothesis of non-compliance was not rejected during all months under analysis.

b)perhaps more interesting, the latter distribution presents a more pronounced concentration of mass at the range within the Wage Law parameters. There are "plateaus" during May and November at the maximum monthly adjustments of the law.

In order to get an even finer picture of the effectivity of the Wage

---

10 Graphs 6.2 and 6.4 present the evolution of the proportion located in each part of the sample.

Law, the analysis of monthly adjustment factors is further restricted to the group of employees with 12 months adjustments above the Wage Law minimum requirements (those which the hypothesis of compliance with the Wage Law can not be rejected), the analysis of this group is divided in two parts: a) those above the law 12 months adjustment prescription. b) those within the Wage Law 12 months adjustments range (ie., those which the hypothesis that the Wage Law 12 months prescription is exactly followed cannot be rejected). The analysis of graph 8.2 and of graphs 8.1 to 8.12 in appendix B reveals that[11]:

a) the distribution of those above the law 12 months adjustments prescriptions first-order stochastically dominates the distribution of the other groups.

b) the distribution of monthly adjustment factors of those within the law 12 months adjustments prescriptions range presents even clearer "plateaus" at the maximum monthly adjustment prescribed during May and November.

11. Wage Adjustments by Wage Levels

The indexation clauses of the Brazilian Wage Law during 1980-86, 1989 and 1991-92 prescribed higher wage adjustments for lower wage brackets. More specifically, during the first sub-period, the Wage Law adopted the so-called "cascade effect", according to which the lower part of individual wages received higher adjustments rates. Taking the "cascade-effect" prescriptions at face value, the earnings inequality would be reduced while the ordering of the earnings distribution would be preserved. The magnitudes of the marginal adjustments factors by earnings brackets of the "cascade effect" are presented in table 5. These factors should multiply the six-month consumer price index factor (ie., one plus the inflation rate accumulated over a six month period). A curious

---

11

Graphs 8.1 to 8.12 present the same distributions for effective earnings concepts.

characteristic of the "cascade-effect" is that its equalizing potential increases with the inflation rate. In the case of stable prices prescriptions of the Wage Law would be homogeneous across different wage brackets.

Graph 9 exemplifies the ex-post equalizing potential of the "cascade-effect" over institutional wages during the 1980-86 period. The graph shows the behavior of the institutional wage of representative workers all with a May settlement date that earned during January 1980 respectively 1, 2, 4, 8, 16 and 32 minimum wages. The earnings ratio between the higher and the lower earnings agents fall to one third of its initial value during this interval. The speed of convergence increases as inflation rises in the middle of the sample. Therefore, in the absence of labor turn over or a noncompliance with the Wage Law, the distributive potential of the "cascate-effect" should not be underestimated.

The "cascade effect", as any of the idiosyncratic components of the Wage Law, creates some problems to the use of aggregated time series of observed wages. The average institutional wage would mix earnings information of individuals subject to different wage adjustments prescriptions.

The analysis of actual wage adjustments by wage level compares the CDF of two groups using the initial earnings value of two minimum wages as the border line dividing the basic samples of employees under analysis. The short-run dynamics of these two initial earnings groups of employees during the 1980-86 period are explored in appendix B: Graphs 9.1 to 9.12 and Graphs 9.36 to 9.48 present all the possible month by month RR_{12}-CDF based on normal earnings and effective earnings, respectively. Similarly, Graphs 10.1 to 10.24 and Graphs 10.25 to 10.96 present the RR_1-CDF based on normal earnings and effective earnings, respectively. Graphs 10.1 to 10.6 in the end of the paper present a resume of the main ranges of the RR_1 CDF's for the two groups of employees considered.

The most striking feature of these graphs is that the wage adjustments distributions for the lower earnings brackets always first-order stochastically dominates (non-strictly) the wage adjustments distribution for the higher wage adjustment brackets. That is, according to the two earnings concepts considered and to all periods under analysis earnings (ie., from March 1980 to February 1986) continuously employed workers with lower initial earnings get higher wage increases than higher initial earnings groups[12]. This conclusion is robust for monthly and for annual adjustment periods considered.

Graph 9.2 provide a measure of the distances between the distributions as a cardinal complementary to the concept of stochastic dominance. This measure corresponds to the average ratio between all corresponding percentiles of the RR_1 and RR_{12} distributions[13].

12. Conclusions

The paper attempted to give a first step toward a methodology to quantify the influences of regulation on earnings short-run dynamics. The analysis developed provided as a by product evidence on the patterns of wage adjustments adopted during the recent high inflationary experience in Brazil.

The main contribution of the paper was providing an assessment of the main difficulties found to evaluate the effect of regulation on earnings. The discussion ranged from the earnings concepts used in the surveys to the level of aggregation used in the analysis. The paper pointed out problems with the earnings concepts used by PME from 1982 onwards. The paper also pointed out advantages of working with flexible data sets where one can adapt actual earnings data to specific aspects of the Wage Law.

The large variety of official wage indexation rules adopted combined

---

12 Barros, Neri and Jovita (1995) report that the probability of a worker with earnings below 2 minimum wage being fired was 3 times higher than for workers with higher wage levels during the 1980.

13 This measure uses the lower distribution in the denominator.

with the availability of monthly surveys on labor markets during the last 15 years generate exceptional conditions to test how regulation affects earnings dynamics. In particular, the combination of large sample sizes with the possibility of following the same worker through short periods of time allows to estimate the cross-sectional distribution of longitudinal statistics based on observed earnings (ie., monthly and annual rates of variation). The main research strategy adopted here was to compare the distribution of actual and artificial earnings short run movements.

The empirical analysis gave an special attention to the 1980-86 period. The analysis revealed a large heterogeneity of wage adjustments patterns which confirmed the adequacy of a more disaggregated approach. Some stylized facts extracted from the empirical analysis are presented below:

a) an average rate of non-compliance with the Wage Law 12-month adjustment prescriptions of approximately 19%. The non-compliance rate declined over the 1980-85 period.

b) an average proportion of fixed nominal contractual wages between consecutive two months of approximately 30%.

c) a concentration of adjustments in contractual wages during May and November, specially for those complying with the Wage Law 12-month prescriptions and for low wage earners. These groups presented proportions ranging from 5% to 20% at the exact rate of variation given to the minimum wage.

d) the distribution of monthly and annual rates of variation of low wage earners first-order stochastically dominates the corresponding distributions of high wage earners during all months under analysis.

e) The occurrence of transitory shocks associated with the payment of the 13th wage during November and December.

In sum, the visual approach adopted here may be viewed as a first approximation to evaluate the impacts of the Wage Law on earnings dynamics. It is inconclusive by nature but it may be helpful to present some guide lines for future empirical work.

APPENDIX A: INSTITUTIONAL BACKGROUND

During the last fifty years, Brazilian workers have been compulsorily organized in associations according to activity and region. Each association has a date of collective wage negotiation. These dates set the time of the year when nominal wage values and other non wage benefits are objects of discussion between firms and employees. The Wage Law circumscribes these negotiations by fixing a floor to nominal wage adjustments.

Each wage settlement date also determines the schedule for within-year wage adjustments. For example, in the case of half-yearly adjustments, a worker group with a May settlement date will also receive automatic wage adjustments every November. Or in the case of quarterly adjustments, a worker group with a May settlement date will also have automatic wage adjustments every August, November and February and so on.

The duration of labor contracts is not specified. The initial wage at a monthly rate is registered and it has to meet the minimum wage legal requirement. The minimum wage adjustment policy follows closely the indexation clauses prescribed to all other wages. The minimum wage settlement date is on May every year. Once a year, workers also receive an additional payment known as the 13th salary, normally viewed as a Christmas bonus.

2.B Wage Indexation Clauses: An Overview

In contrast with most of the literature, the indexation clauses found in the Brazilian Wage Law are not the spontaneous outcome of firms and workers strive to reduce labor income risk. These indexation clauses are the result of government' discretionary power. In particular, the Brazilian Wage Law attempts to interfere not only with the inflationary risk that falls upon earnings but also with the wage level itself.

A regime of wage indexation is defined as a series of deterministic rules that specify the relationship between price indexes and nominal wages. There are two major attributes of wage indexation regimes:

a) The Frequency of Adjustments relies on whether the rule is time dependent or state dependent and on the specific trigger points set for each type of rule. Most of the indexation rules found in the recent Brazilian experience were time dependent. However, as inflation rose over the years, the successive time dependent regimes of indexation adopted shorter wage adjustment intervals. The only genuine state dependent rule used in Brazil had its trigger point given by the cumulative variation of a price index (ie. as inflation reaches a certain point wages were automatically adjusted). The main feature of state dependent rules is that the frequency of adjustments becomes endogenous to the system within the same regime of wage indexation.

b) The Wage Index aspect is multidimensional. It involves not only the choice among the price index technologies available but also the decision of how to apply it to wages. The characteristics of price indexes include the implicit lag between price indexes movements and effective price movements[14] and, of course, the universe used to find the price index weights (ie: regions, earnings brackets etc.).

Once the price index is chosen there are numerous ways to use it to adjust wages. One possibility is setting the delay between the moment the price index is available and the moment it adjusts wages. Another possibility is setting how much of a given price level variation is translated into wages. A distinct feature of various regimes of wage indexation found in Brazil is prescribing regressive adjustments rates by wage level.

Appendix A describes in detail the evolution of the official wage indexation rules during the 1980-92 period. Table 1 below pictures the stylized facts associated with the eight regimes of wage indexation discussed in the appendix A.

---

14 In practice, there was no such thing as the perfect indexation paradigm found in most of the literature (Gray (1976), Fisher (1977)).

TABLE 1
Stylized Facts of Wage Indexation Regimes in Brazil
1980-92

Starting date	Dec 79	Mar 86	Jun 87	Jan 89	May 89	Mar 90	Set 91	Dec 92
Duration in months	63	15	20	4	10	18	15	9
Stab.Plan Name (If any)	-	Cruzado	Bresser	Summer	-	Collor	-	-
Transition(1) Phase	Inst.	Inst,Avg	Gradual	Inst,Av	Inst.	Inst.	Gradual	Gradual
Rule Type(2):	Time,RA	State	Time	Time	Time,R	-	Time,RA	Time,RA
Trigger Point:	6 m	20%	1 m	-	1 m	-	4;2 m	4;2 m
Avg Lag(3):	8 m	E(4)	4 m	-	1 m	-	4 m	4 m

Notes :

(1) The first atribute indicates the transition to the new wage indexation regime was done in a Instantaneous or a Gradual manner. The second attribute indicates whether there was a conversion to previous real average values.

(2) For the Time dependent rules the trigger point is specified in terms of months between adjustments. For the State dependent rule the trigger point was specified in terms of accumulated price index variation between adjustments. RA means Regressive Adjustments.

(3) Refers to the average lag between price rises and their incorporation to wages.

(4) In the state dependent case the lag is endogenous.

References:

BARROS, R., NERI, M. & JOVITA M. (1995). "De Volta para o Futuro da Política Salarial", Anais do XVI Encontro Brasileiro de Econometria da Sociedade Brasileira de Econometria, Salvador.

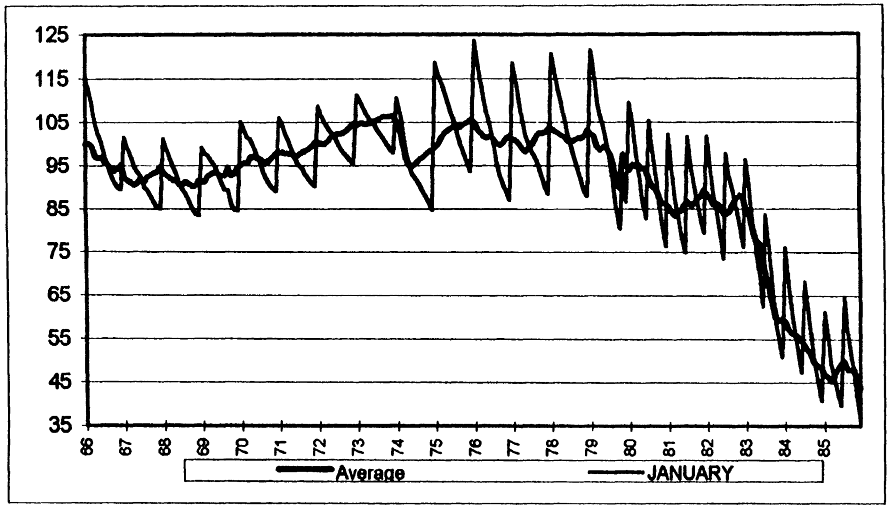
CAMARGO, J. M. (1990). Salários nominais, política salarial e ativismo sindical. PNPE 36/90.

CAMARGO, J. M. (1990). Salários e negociações coletivas. TD 244, PUC/RJ.

DINARDO, J., Fortin, N. & Lemieux, T. (1995). "Labor market institutions and the distribution of wages, 1973-1992: a semiparametric approach" in Working Paper Series 5093.

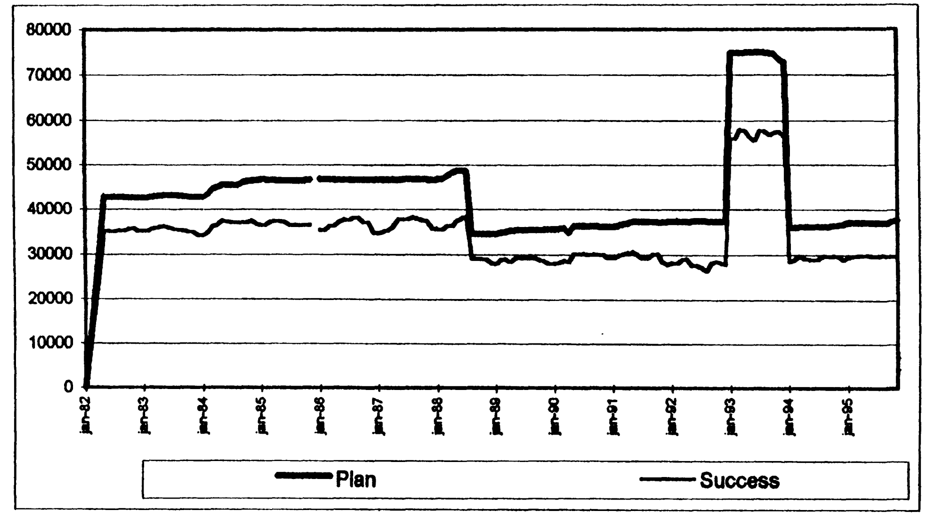
GONZAGA, G. M. (1988). Efetividade da Política Salarial no Brasil: 1964-1985. Dissertação de Mestrado, PUC/RJ.

Graph 1 - Institutional Wages - Aggregate X Individual -1966-85

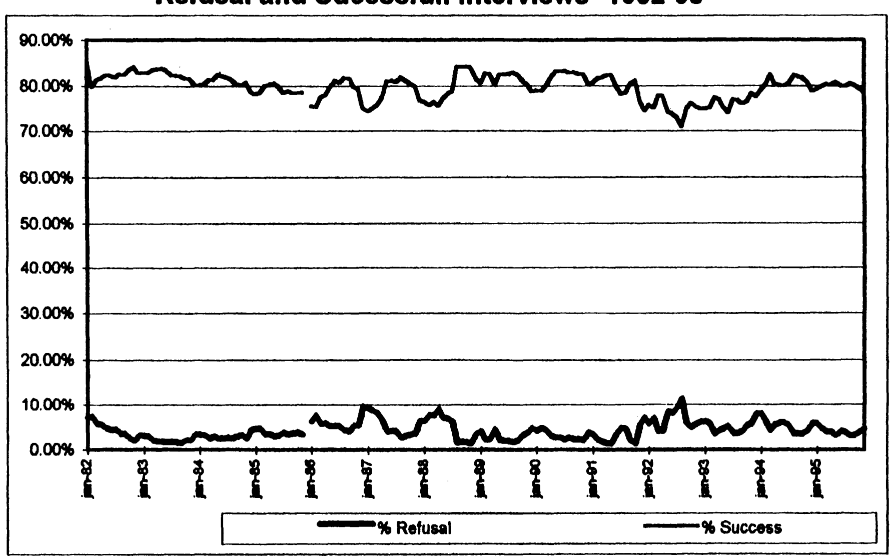


Source:Gonzaga (1988)

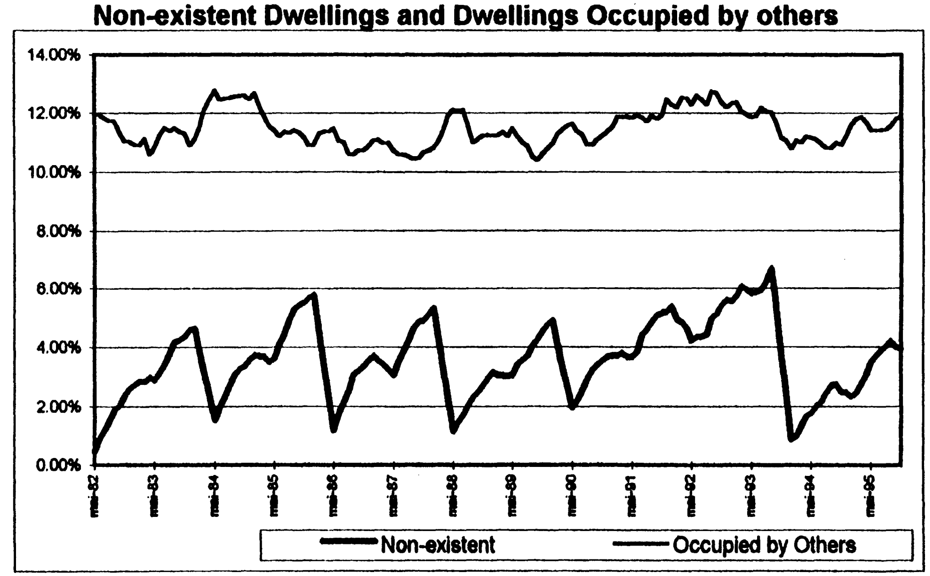
Graph 2.1 - Non-Response Bias - Number of Interviews: Plan and Sucessfull Interviews



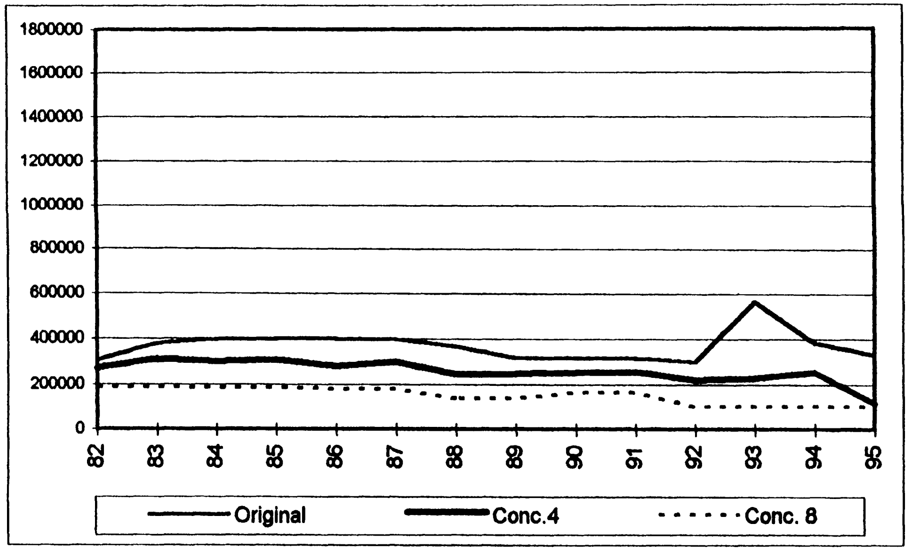
Graph 2.2 - Non-Response Bias - Proportion of Interviews: Refusal and Sucessfull Interviews -1982-95



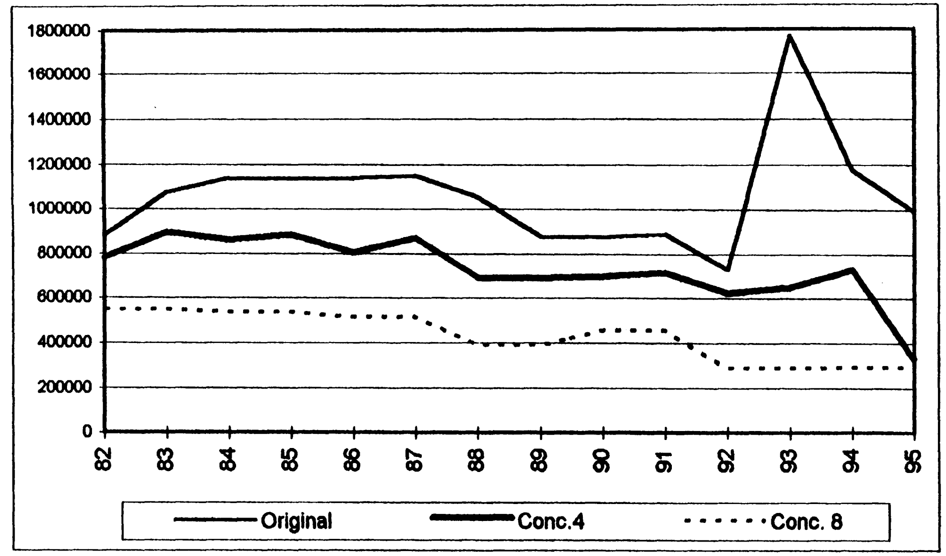
Graph 2.3 - Non-Response Bias - Proporti Non-existent Dwellings and Dwellings Occupied by others



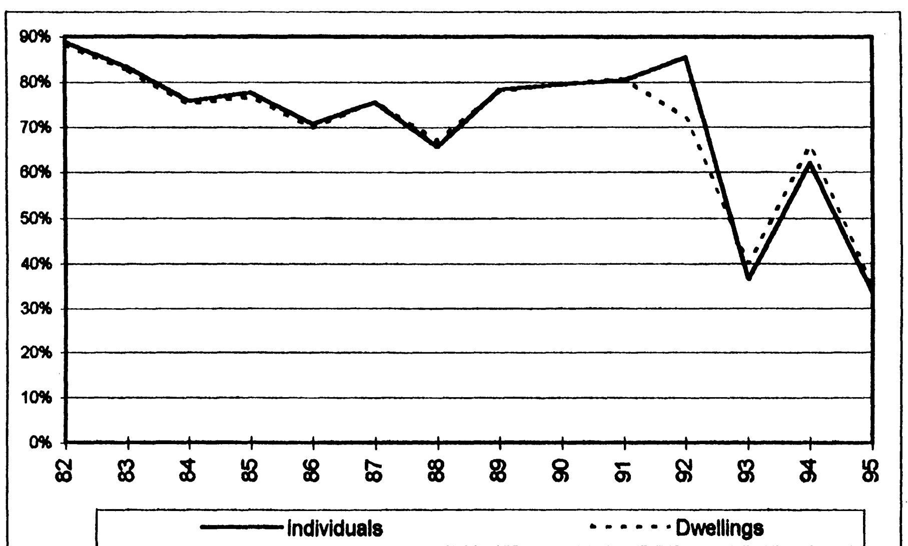
Graph 2.4 Number of Observations (Dwellings per Year) - 1982-95



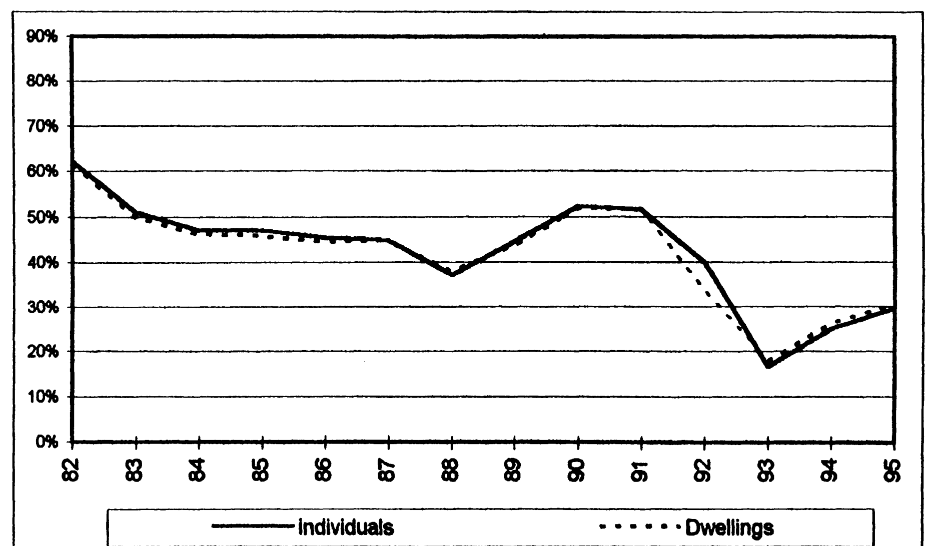
Graph 2.5 Number of Observations (Working Age Individuals per Year)



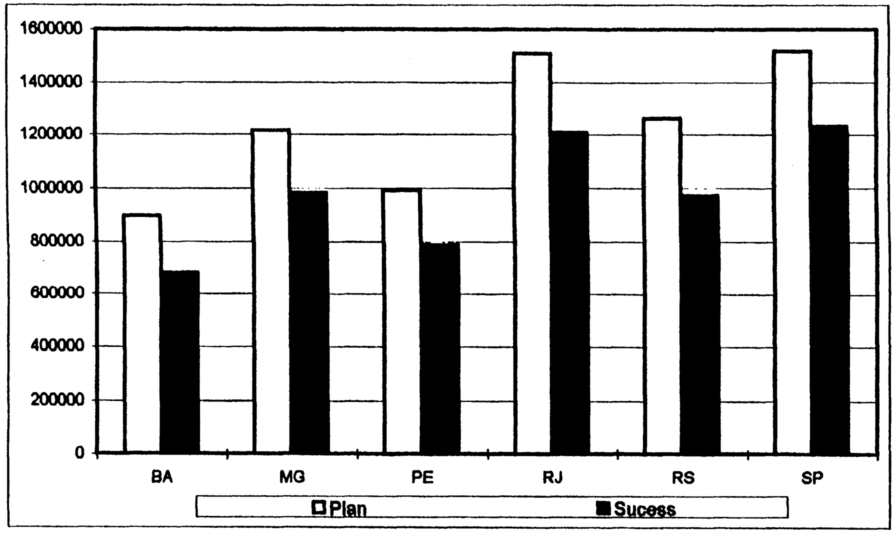
Graph 2.6 - Attrition - Proportion of Sucess in the Concatenation 4 Observations - 1982-95



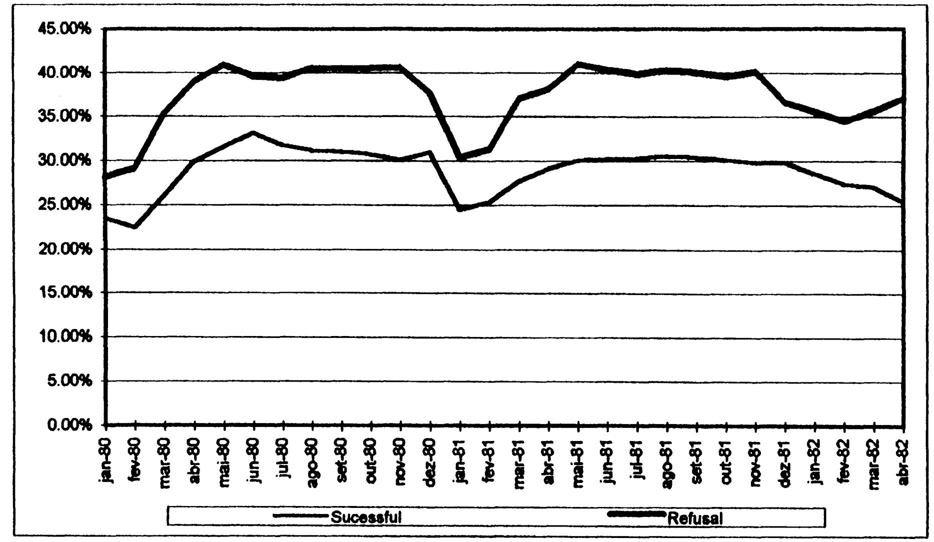
Graph 2.7 - Attrition - Proportion of Sucess in the Concatenation 8 Observations - 1982-95



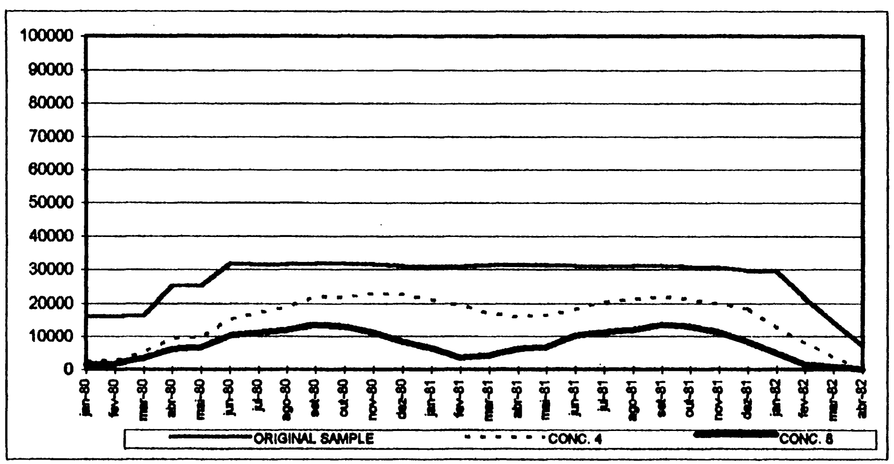
Graph 2.8 - Number of Dwellings: By Metropolitan Region - 1982-95



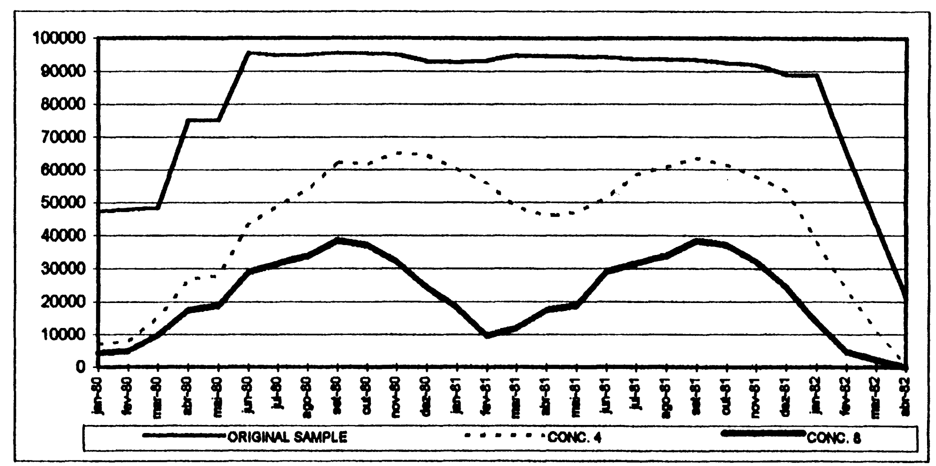
Graph 3.1-Number of Dwellings:Plan and Sucessfull Interviews-1980-82



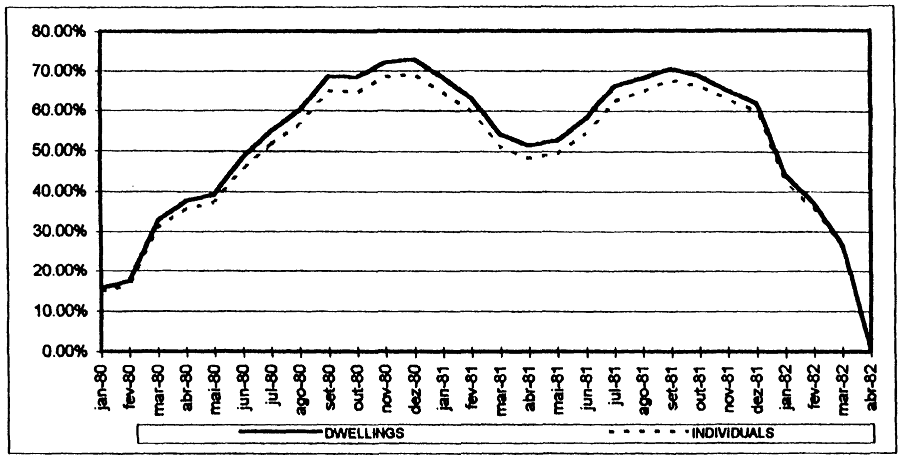
Graph 3.2 Number of Observations (Dwellings per Month) - 1982-95



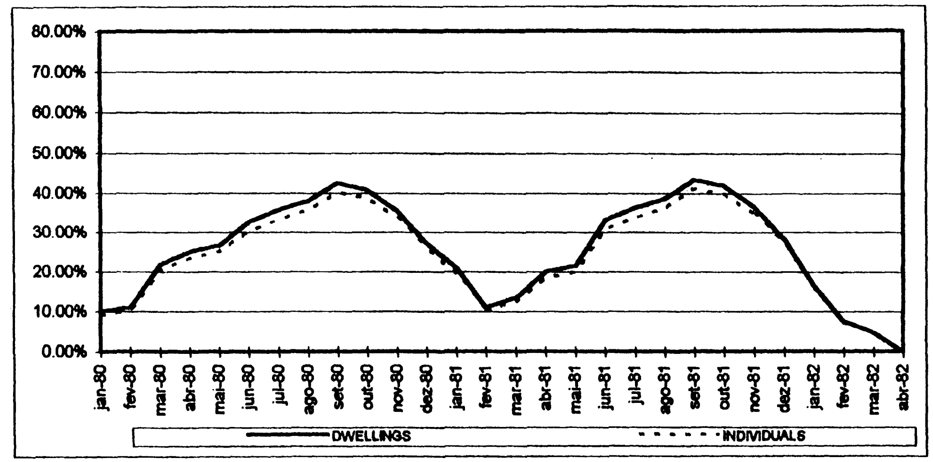
Graph 3.3 Number of Observations (Individuals per Month) - 1982-95



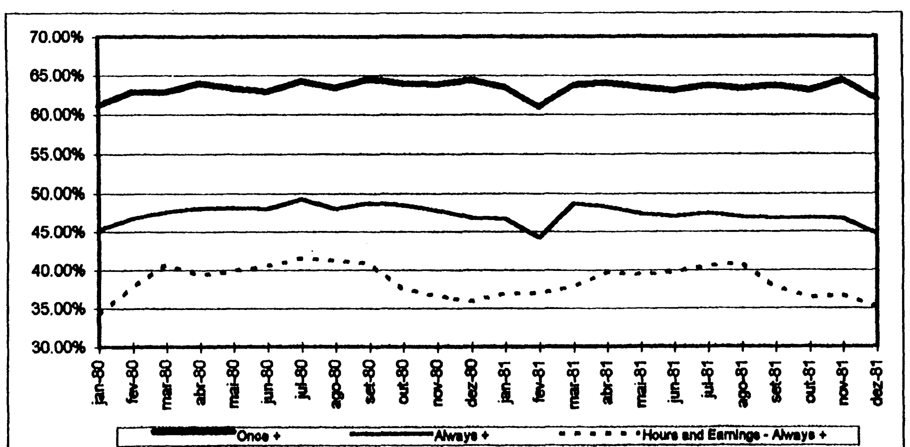
**Graph 3.5 - Attrition - Proportion of Sucess in the Concatenation
4 Observations - 1980-82**



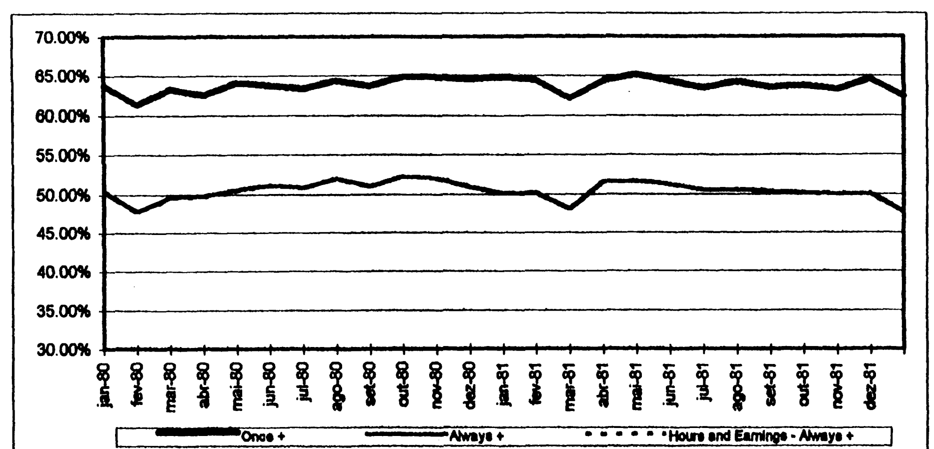
**Graph 3.5 - Attrition - Proportion of Sucess in the Concatenation
8 Observations - 1980-82**



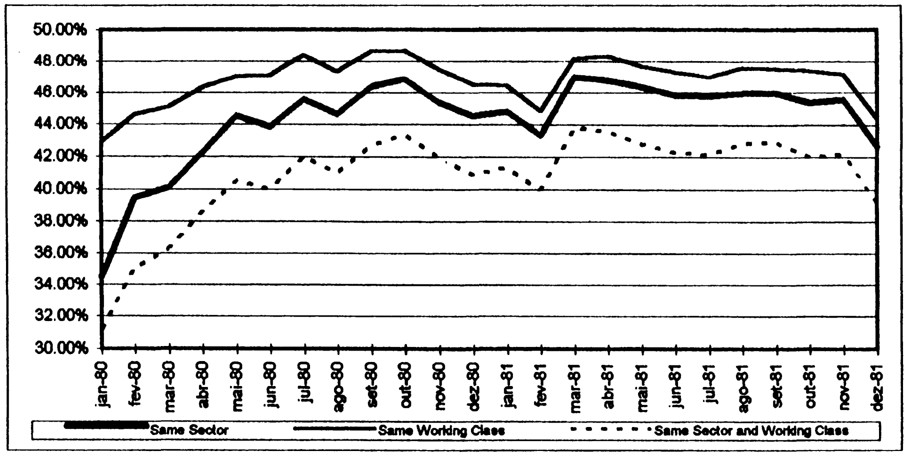
**Graph 4.1 - Proportion of Individuals with Positive Earnings
4 Consecutive Observations - Using the Effective Earnings Concept**



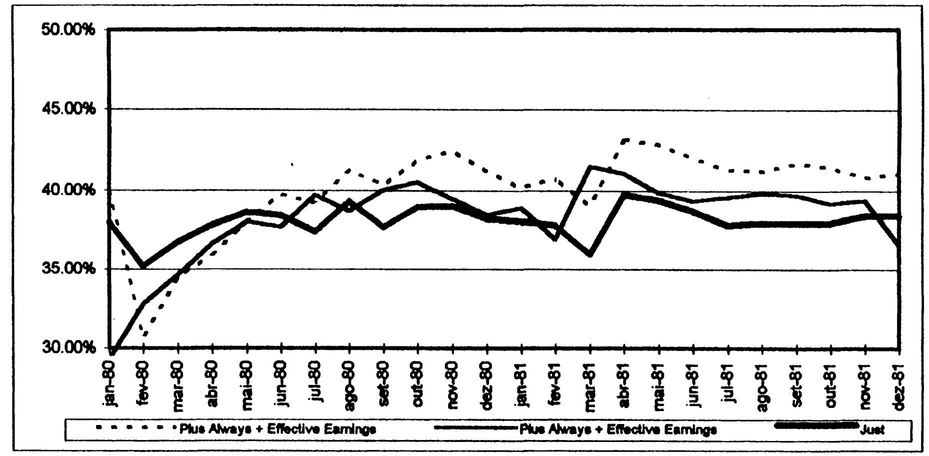
**Graph 4.2 - Proportion of Individuals with Positive Earnings
4 Consecutive Observations - Using the Normal Earnings Concept**



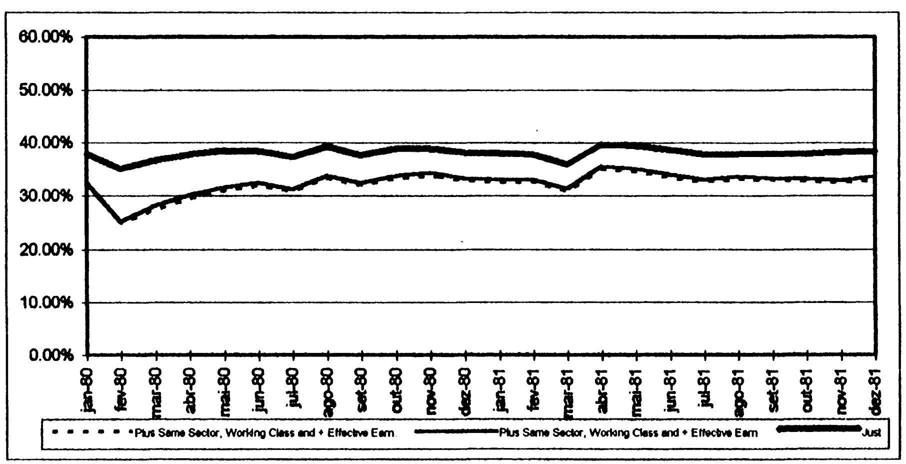
Graph 4.3 - Proportion of Individuals with Same Sector of Activity and/or Working Class - 4 Consecutive Observations - 1980-82



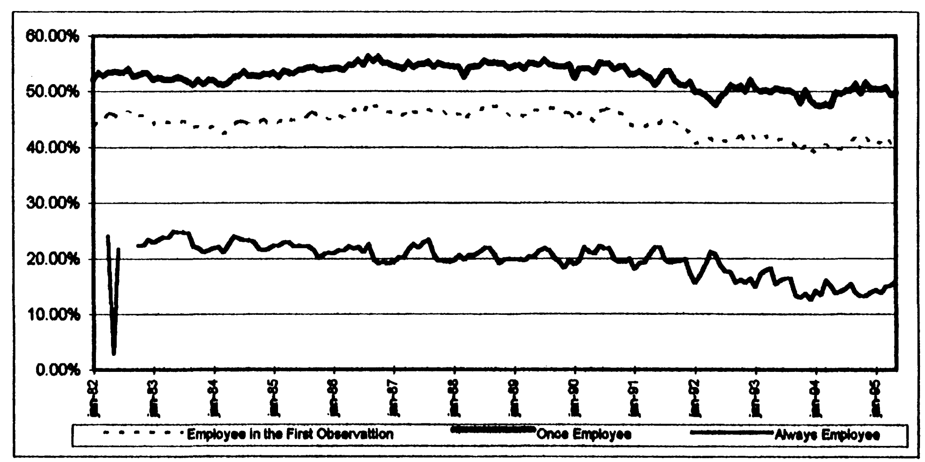
Graph 4.4 - Proportion of Individuals with Same Sector of Activity , Working Class and Proportion of Individuals Just Continuously Employed - 4 Consecutive Observations - 1980-82



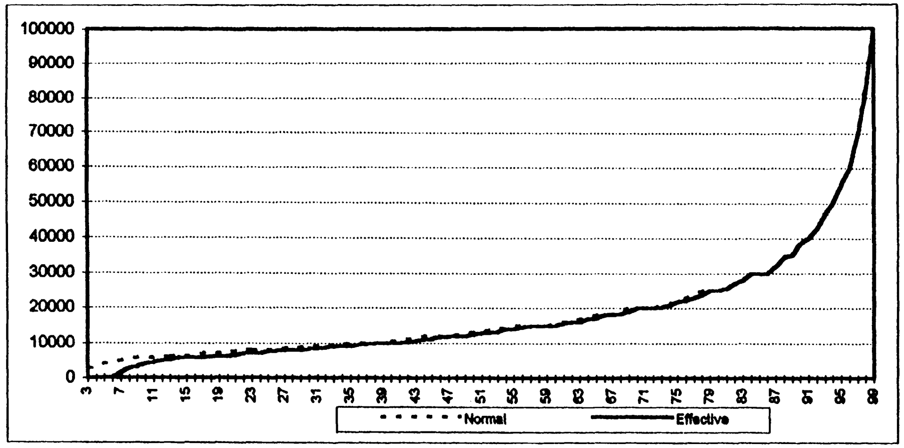
Graph 4.5 - 4 Consecutive Observations - Proportion Proxied as Continuously Employed Individuals (CE)



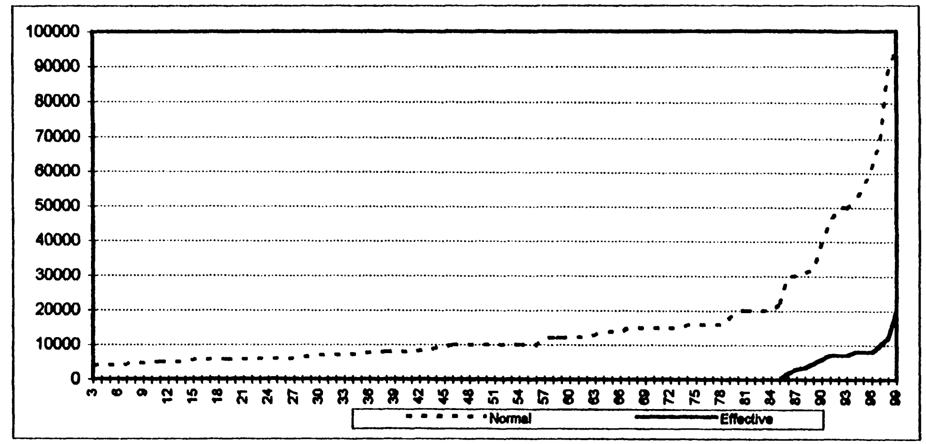
Graph 4.6 - Proportion of Working Age Individuals: Once Employees in 4 Observations, Continuously Employed in 4 Observations and Employees in a given Observation in the of Sample 4 Observations



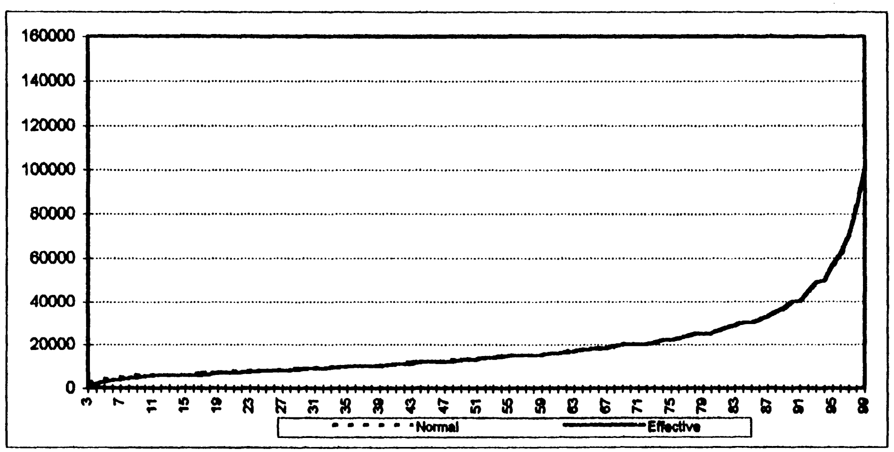
Graph 5.1 - CDF Earnings Level - Universe of Active Age Employees Normal Earnings and Effective Earnings - January 1981 -São Paulo



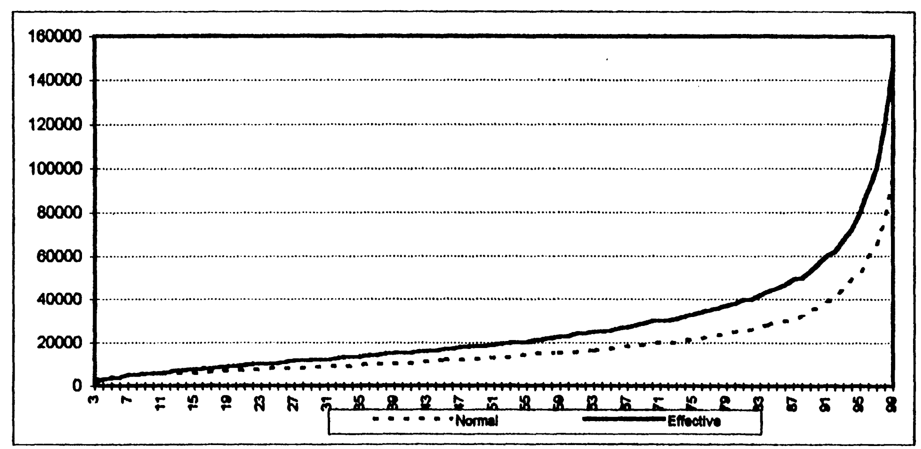
Graph 5.2 - CDF Earnings Level - Employees Moving Between Jobs Normal Earnings X Effective Earnings - January 1981 - São Paulo



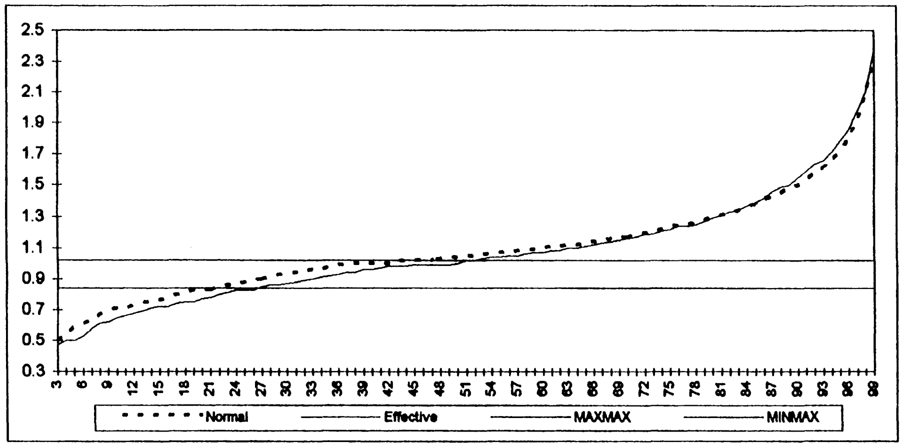
Graph 5.3 - CDF Earnings Level - Continuously Employed Workers Normal Earnings X Effective Earnings - January 1981 - São Paulo



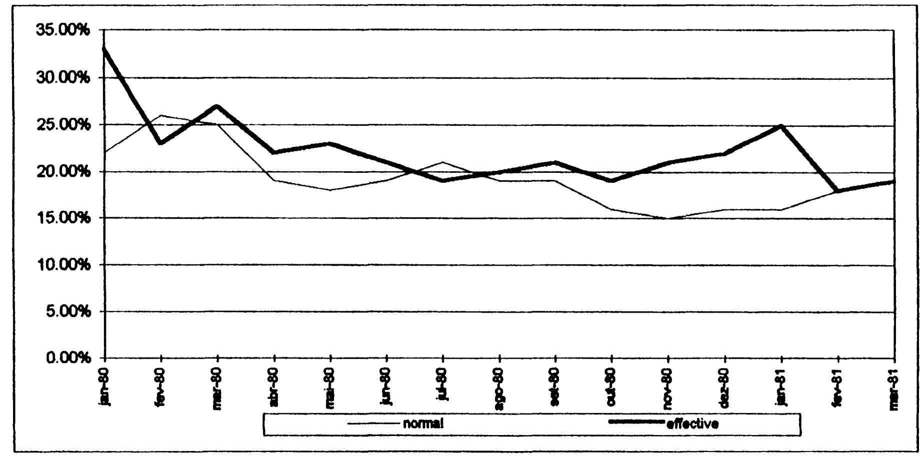
Graph 5.4 - CDF Earnings Level - Continuously Employed Workers Normal Earnings X Effective Earnings - December 1980 - São Paulo



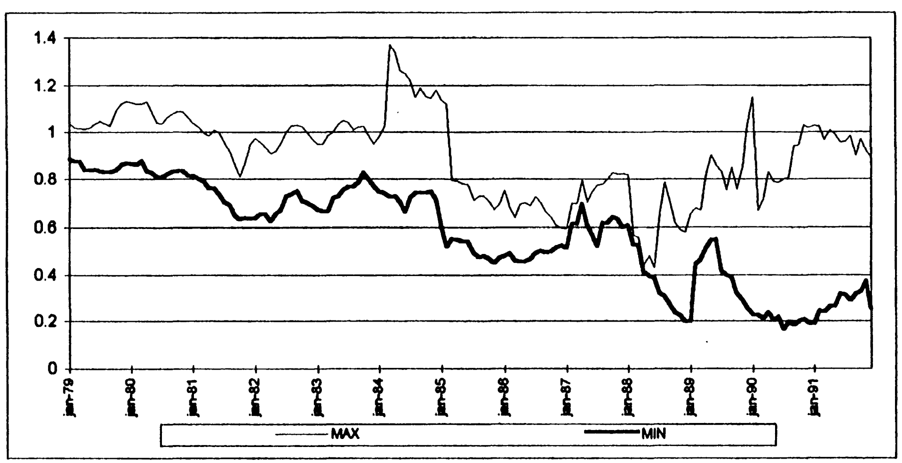
Graph 6.1 - CDF Real Earnings Adjustment Factors - May 1980
12 Months Adjustments - By Earnings Concepts



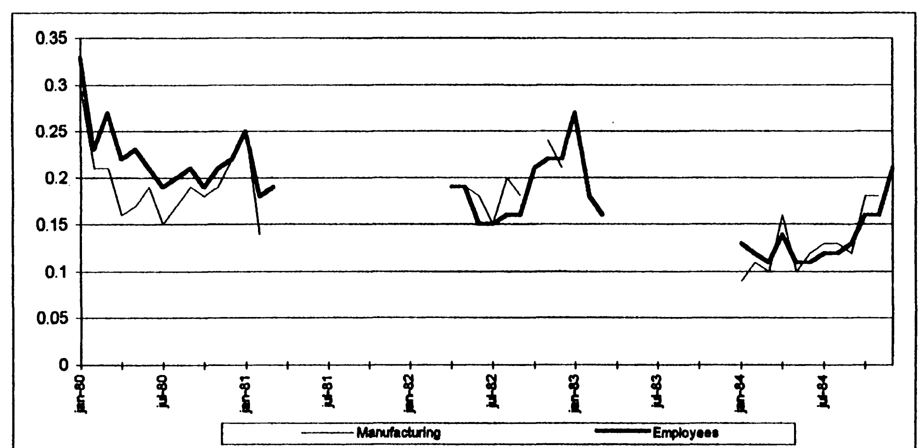
Graph 6.2 - Non-Compliance with the Wage Law
12 Months Prescriptions - 1980-82 - By Earnings Concepts



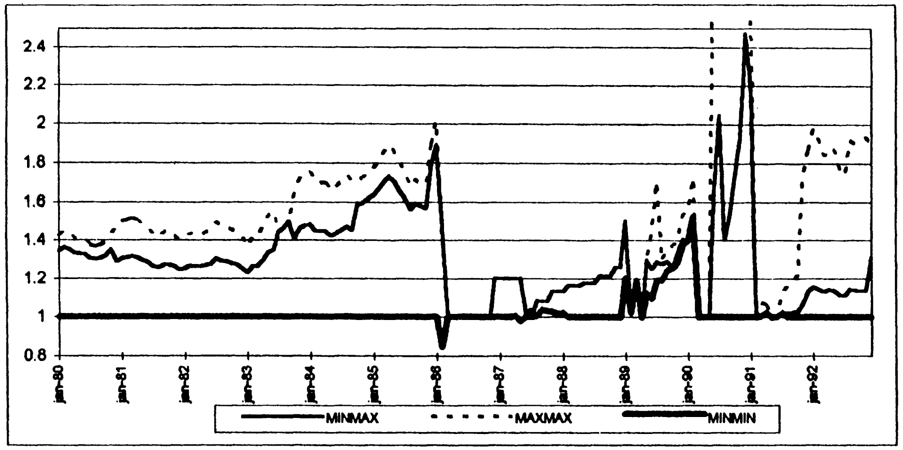
Graph 6.3 - 12-Month adjustment Factors of the Wage Law -1979-91



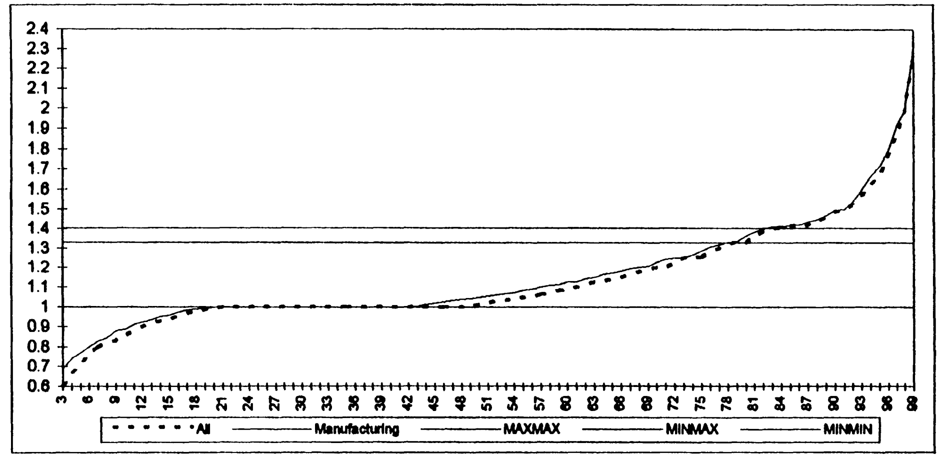
Graph 6.4 - Non-Compliance with the Wage Law
12 Months Prescriptions - 1980-84 -



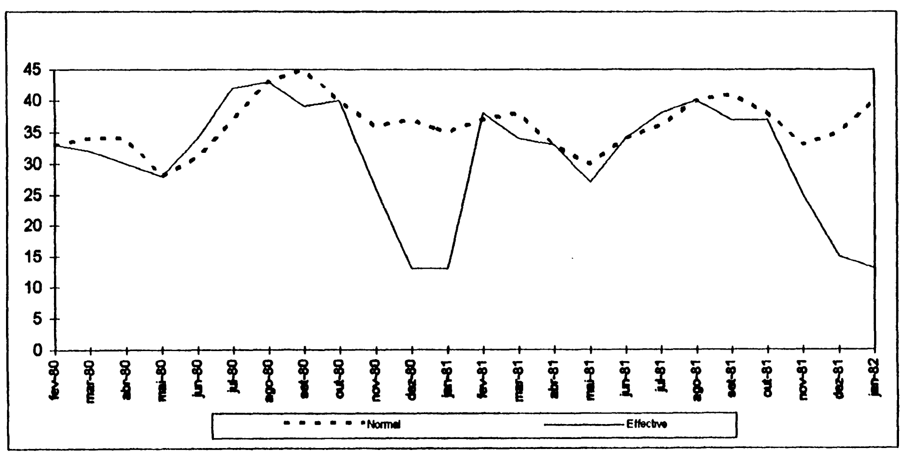
Graph 7.1 - Monthly Adjustments Factors of the Wage Law - 1980-92



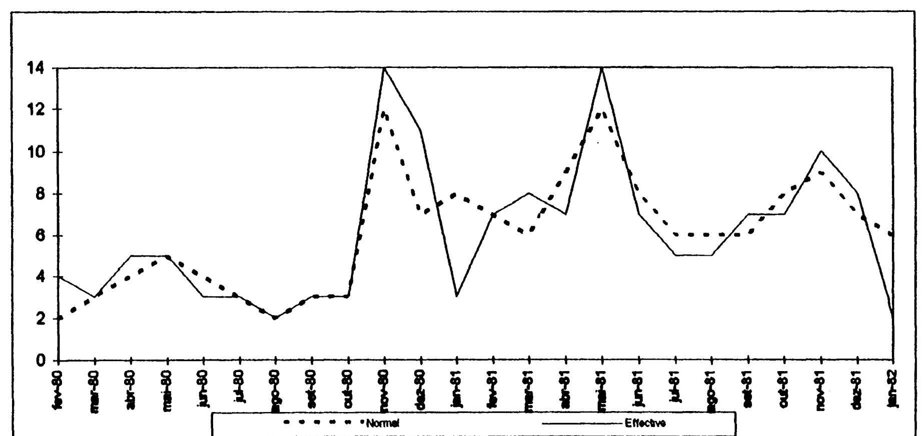
Graph 7.2 - CDF Nominal Earnings Monthly Adjustment Factors May 1980 - By Earnings Concepts



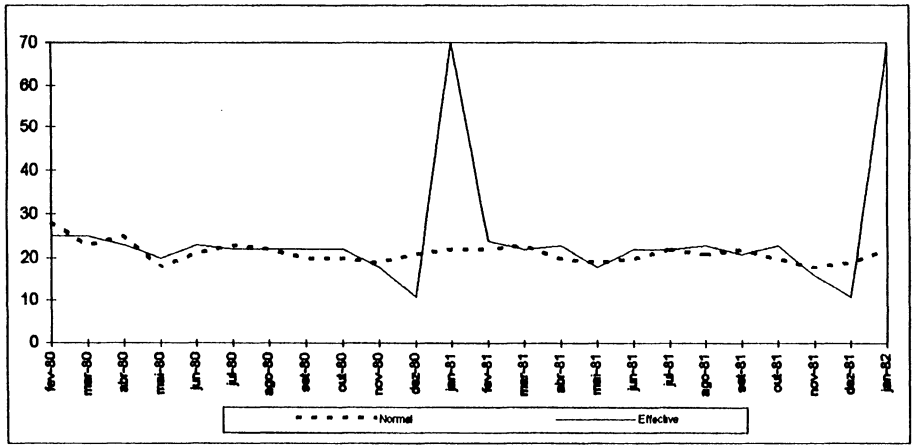
Graph 7.3 - Employees - Feb1980-Jan 1992 - By Earnings Concepts Share with Nominal Earnings Constant



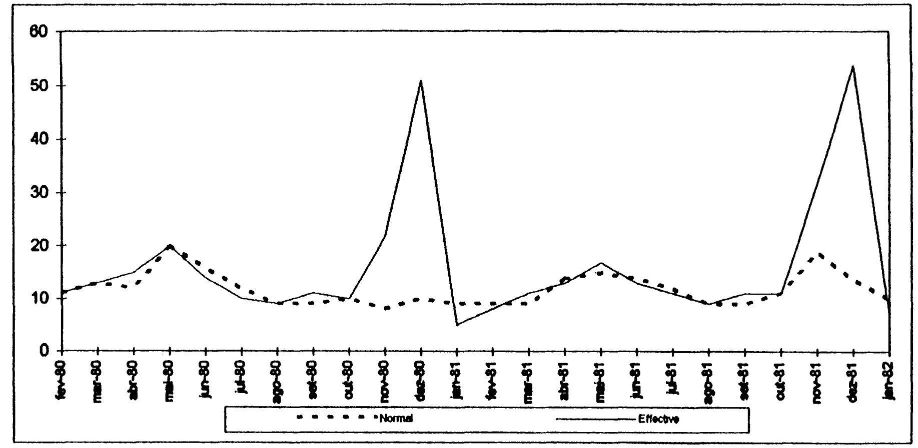
Graph 7.4 - Employees - Feb1980-Jan 1992 - By Earnings Concepts Share Within Wage Law SD Adjustments Range



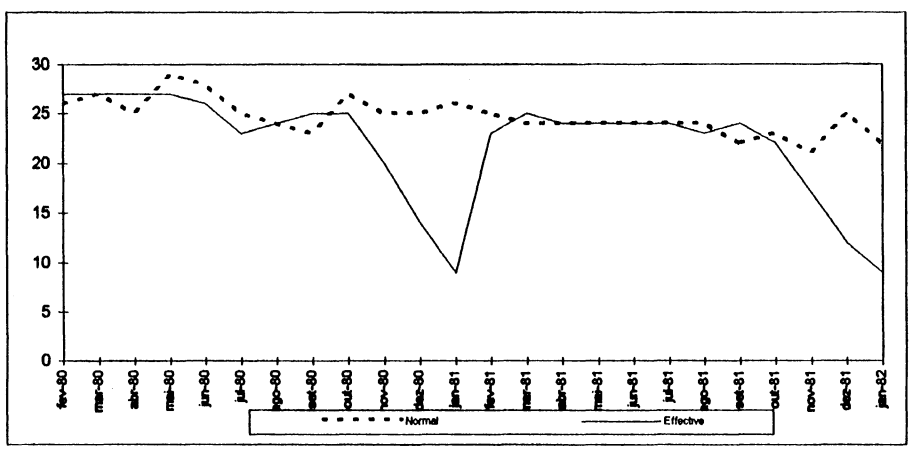
Graph 7.5 - Employees - Feb1980-Jan 1992 - By Earnings Concepts Share with Nominal Earnings Falling



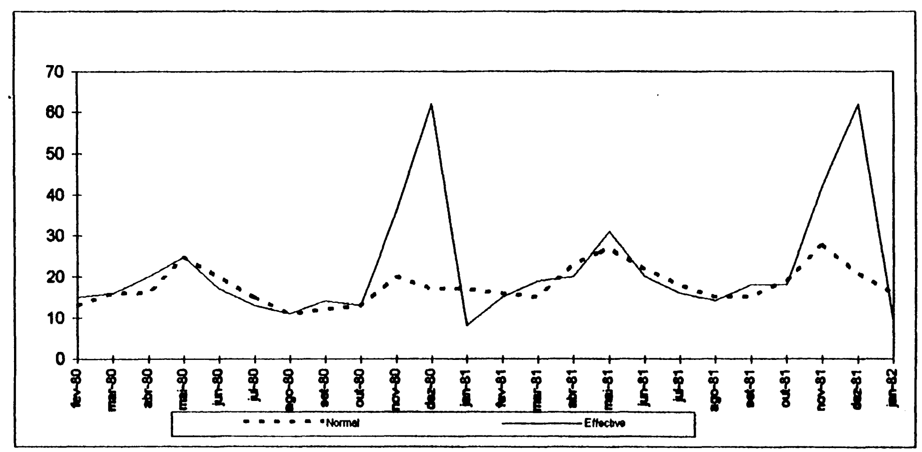
Graph 7.6 - Employees - Feb1980-Jan 1992 - By Earnings Concepts Share Above Wage Law SD Max Adjustments



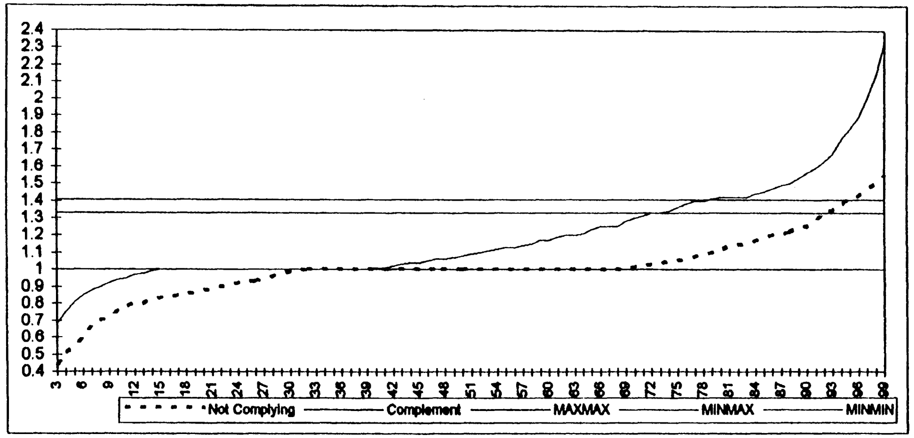
Graph 7.7 - Employees - Feb1980-Jan 1992 - By Earnings Concepts Share Within Wage Law No Adjustment and SD Min. Adjustment.



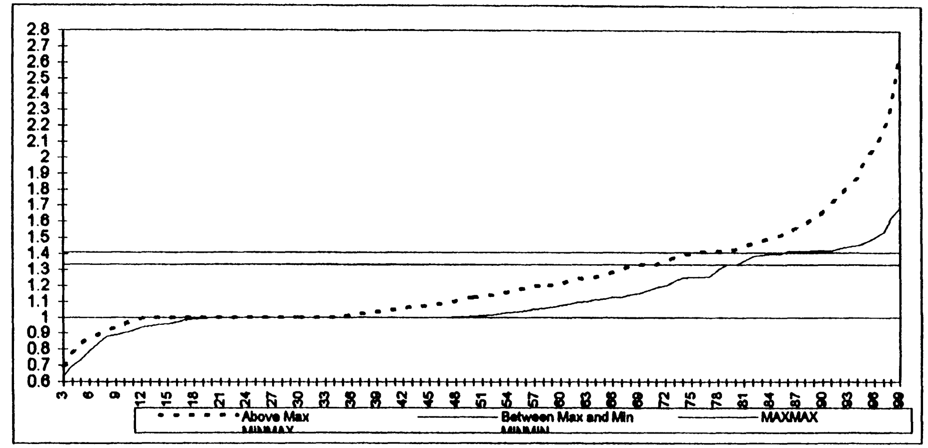
Graph 7.8 - Employees - Feb1980-Jan 1992 - By Earnings Concepts Share Above Wage Law SD Min Adjustments



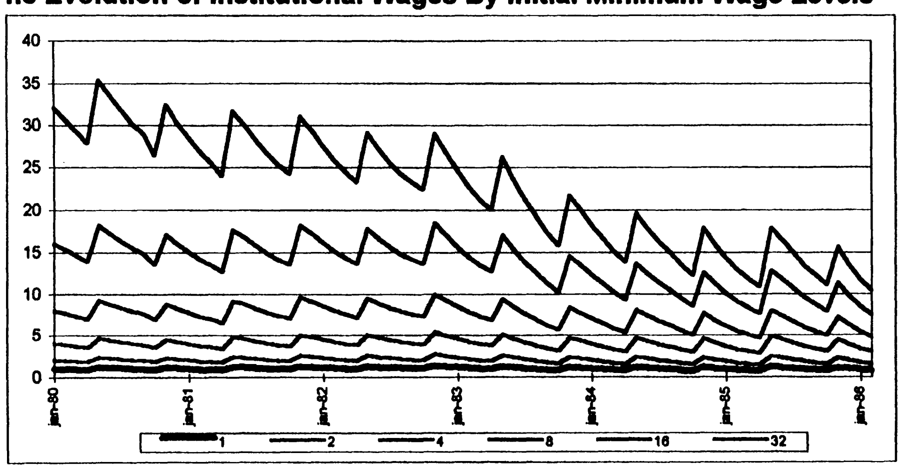
**Graph 8.1 - CDF Nominal Monthly Adjustment Factors-
Normal Earnings - May 1980
Non-Compliers with the Law 12 Months Prescription X Complement**



**Graph 8.2 - CDF Nominal Monthly Adjustment Factors
Normal Earnings - May 1980 - Complement of Non-Compliers
Above 12 Months Max Prescription X In Between Max-Min Prescrip.**



**Graph 9.1 - The "Cascade-Effect" - 1980-86
The Evolution of Institutional Wages By Initial Minimum Wage Levels**



**Graph 9.2 - Average Relative Distance Between Montlhy Adj. Factors
Between Employees Below 2 MW and Employees Above 2 MW**

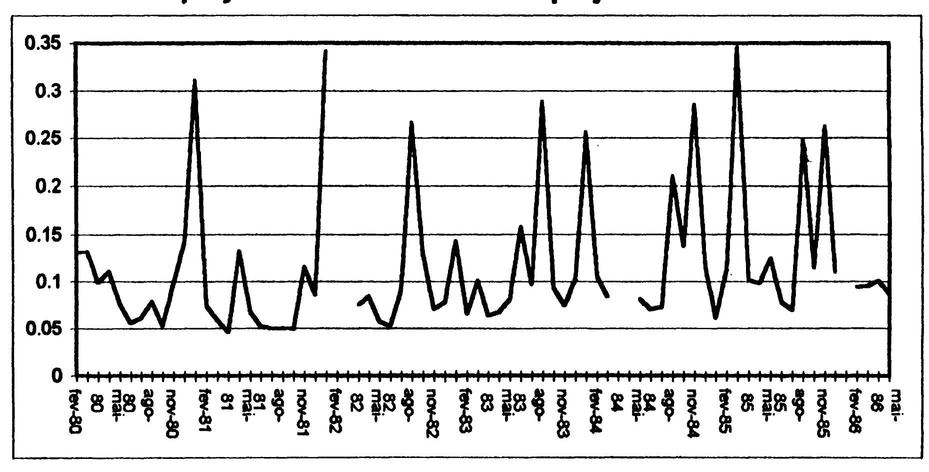


TABLE 2.1
Profile of Samples - Active Age Individuals - 1982-95

	Original PME Sample	Longitudinal Sample 4 Obs	Longitudinal Sample 8 Obs
Share of Males	47.47%	47.37%	47.44%
Average Age	33.78	34.16	34.54
Share of Heads	34.71%	34.99%	34.73%
Share with less than Finished High School Education	61.01%	60.02%	60.94%
Share in Manufacturing	12.91%	12.20%	11.90%
Participation Rates	58.04%	60.01%	59.31%
Share Searching a Job	3.18%	3.44%	3.44%
Average Unemployment Spell	40.85	41.85	44.18
Average Hours Worked Per week	43.61	41.96	41.92
Share of Employees	36.33%	37.41%	36.78%
Share of Legal Employees	75.13%	78.59%	76.58%
Number of Observations	14911309	10236752	6077712

TABLE 2.2
Profile of Samples - Active Age Individuals - 1980-82

	Original PME Sample	Longitudinal Sample 4 Obs	Longitudinal Sample 8 Obs
Share of Males	47.66%	46.20%	46.28%
Average Age	33.14	34.27	35.10
Share of Heads	33.25%	35.00%	35.23%
Share in Manufacturing	23.58%	24.16%	24.01%
Share of Individuals Working	59.71%	59.36%	57.80%
Share of Employees	78.95%	78.19%	77.07%
Number of Observations	2328274	1221800	605896

TABLE 3.1

Profile of Active Age Employees Samples 1982-93

	Original PME Sample	Longitudinal Sample 4 Obs	Longitudinal Sample 8 Obs
Share of Males	62.24%	63.16%	64.43%
Average Age	32.11	32.55	32.99
Share of Heads	45.40%	45.87%	46.56%
Share with less than Finished High School Education	51.20%	52.27%	52.90%
Share in Manufacturing	53.14%	25.09%	24.97%
Average Hours Worked Per week	42.74	42.38	42.33
Share of Legal Employees	72.13%	76.04%	76.58%
Number of Observations	3853502	4547980	2645256

TABLE 3.2

Profile of Active Age Employees Samples 1980-82

	Original PME Sample	Longitudinal Sample 4 Obs	Longitudinal Sample 8 Obs
Share of Males	61.39%	62.61%	64.30%
Average Age	31.54	32.56	33.45
Share of Heads	44.11%	46.33%	47.80%
Share in Manufacturing [(2)]	28.24%	29.04%	29.16%
Share that Would Work More Hours	61.68%	60.88%	58.90%
Number of Observations	1079123	567108	269904

TABLE 4.1

Profile of LongitudinalEmployees Samples 1982-95

	Conc. 8 Obs. Once Employee	Conc. 4 Obs. Once Employee	Conc. 4 Obs. Cont. Employed	Conc. 8 Obs. Cont. Employed
Share of Males	61.46%	62.30%	62.31%	65.19%
Average Age	32.14	32.2	33.08	34.00
Share of Heads	41.52%	43.63%	49.15%	53.40%
Share with less than Finished High School Education	55.75%	54.39%	50.02%	48.49%
Share in Manufacturing (2)	18.99%	21.52%	26.88%	29.30%
Average Hours Worked Per week	41.99	42.16	42.72	42.68
Share of Legal Employees	76.58%	76.01%	88.20%	94.60%
Share of Employees	82.45%	84.19%	1	1
Number of Observations	3589264	5402292	1960276	731496

TABLE 4.2

Profile of LongitudinalEmployees Samples 1982-95

	Conc. 8 Obs. Once Employee	Conc. 4 Obs. Once Employee	Conc. 4 Obs. Cont. Employed	Conc. 8 Obs. Cont. Employed
Share of Males	61.39%	61.76%	62.00%	63.63%
Average Age	32.67	32.21	33.09	34.56
Share of Heads	42.96%	44.20%	48.20%	51.64%
Share in Manufacturing (2)	33.82%	31.02%	30.74%	30.72%
Share of Individuals Working	80.37%	91.08%	100.00%	100.00%
Share that Would Work More Hours	59.35%	55.81%	59.66%	56.88%
Number of Observations	354144	653564	408144	137144

Table 5 - The Cascade-Effect

Marginal Adjustment Factors By Earnings Brackets

Law(L), Decreed L	Initial Period	Earnings Brackets in Minimum wages						
		0-1	1-3	3-7	7-10	10-15	15-20	20-
L 6708/79	Nov79	1,10	1,10	1,00	1,00	0,80	0,80	0,80
L 6886/80	Jan/81	1,10	1,10	1,00	1,00	0,80	0,50	0,00
DL 2012	Feb/83	1,00	1,00	0,95	0,80	0,80	0,50	0,00
DL 2024	Jun/83	1,00	1,00	1,00	0,80	0,80	0,50	0,00
DL 2045	Aug/83	0,80	0,80	0,80	0,80	0,80	0,80	0,80
DL 2065	Nov/83	1,00	1,00	0,80	0,60	0,60	0,50	0,50
L 7238/84	Nov/84	1,00	1,00	0,80	0,80	0,80	0,80	0,80
L 7450/85	Jan/86	1,00	1,00	1,00	1,00	0,80	0,80	0,80