Ministério da
Ciência, Tecnologia
e Inovação

sid.inpe.br/mtc-m18@80/2009/10.22.00.10-RPQ

GERAÇÃO DE AGRUPAMENTOS DE SÉRIES TEMPORAIS

Leandro de Capitani Messias

Relatório da disciplina Princípios e Aplicações de Mineração de Dados (CAP-359) ministrada pelo Dr. Rafael Santos

URL do documento original:
<http://urlib.net/8JMKD3MGP8W/369N938>

INPE
São José dos Campos
2012

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE
Gabinete do Diretor (GB)
Serviço de Informação e Documentação (SID)
Caixa Postal 515 - CEP 12.245-970
São José dos Campos - SP - Brasil
Tel.:(012) 3208-6923/6921
Fax: (012) 3208-6919
E-mail: pubtc@sid.inpe.br

Ministério da
Ciência, Tecnologia
e Inovação

sid.inpe.br/mtc-m18@80/2009/10.22.00.10-RPQ

GERAÇÃO DE AGRUPAMENTOS DE SÉRIES TEMPORAIS

Leandro de Capitani Messias

Relatório da disciplina Princípios e Aplicações de Mineração de Dados (CAP-359) ministrada pelo Dr. Rafael Santos

URL do documento original:
<http://urlib.net/8JMKD3MGP8W/369N938>

INPE
São José dos Campos
2012

Resumo

O foco deste trabalho foi desenvolver um método baseado em agrupamentos de séries temporais que fosse capaz de identificar grupos de indicadores econômicos mostrando quais eram, de fato, relevantes para classificar a situação econômica atual do país. Escolheu-se como métrica de dissimilaridade a DTW (*Dynamic Time Warping*) e como método de geração de agrupamentos foi adotado o método hierárquico do centróide que gerou agrupamentos significativamente interpretáveis. Como resultado percebeu-se que apenas quatro índices são suficientes para determinar a situação econômica do país, e eles são: Índice de Produção intensiva em capital, Índice de consumo, Índice de Insumos para a Produção e Índice de produção para consumo.

Abstract

The focus of this work was to develop a method based on clustering time series that was capable to identify groups of economical indicators in fact showing which they were, relevant to classify the current economical situation of the country. It was chosen as metric of dissimilarity the DTW (Dynamic Time Warping) and as method of generation of groupings it was adopted the hierarchical method of the centróide that generated groupings significantly you interpreted. As result was noticed that only four indexes are enough to determine the economical situation of the country, and they are: Index of intensive Production in capital, consumption Index, Index of Inputs for the Production and production Index for consumption.

Lista de Figuras

Figura 5.1: MDS construído com DTW das 37 séries temporais 20
Figura 5.2: MDS com rótulos nos pontos, construído com DTW para as 37 séries temporais 21
Figura 5.3: Comparação entre os MDS formados por DTW e STS 23
Figura 5.4: dendograma formado pelo método do centróide utilizando a matriz de dissimilaridade por DTW.. 24
Figura 5.5: realce dos sete agrupamentos naturais formados pelo método do centróide utilizando a matriz de dissimilaridade construída por DTW 25
Figura 5.6: realce dos clusters formados pelo método do centróide no gráfico do MDS 26
Figura 5.7: gráfico gerado pelo aplicativo SOM (disposição da folha = paisagem), linha preta é o centróide e as azuis são as séries associadas..... 31
Figura 5.8: lista das series por neurônio da aplicação SOM..... 33

Lista de Tabelas

Tabela 1.1: Clustering of time series data — a survey [Liao, 2005] 10
Tabela 4.1: descrição do banco de dados. 14
Tabela 5.1 Séries que compõem o cluster C1 formado pelo método do centróide 27
Tabela 5.2: Séries que compõem o cluster C2 formado pelo método do centróide 27
Tabela 5.3: Séries que compõem o cluster C3 formado pelo método do centróide 28
Tabela 5.4: Séries que compõem o cluster C4 formado pelo método do centróide 28
Tabela 5.5: Séries que compõem o cluster C5 formado pelo método do centróide 28
Tabela 5.6: Séries que compõem o cluster C6 formado pelo método do centróide 28
Tabela 5.7: Série que compõe o cluster intitulado outlier 29
Tabela 5.8: relaciona o cluster a sua interpretação 36

Sumário

Lista de Figuras 4
Lista de Tabelas 5
Siglas 10
1.0 Introdução 11
2.0 Motivação 13
3.0 Objetivos 13
4.0 Metodologia 13
4.1 Descrição do DTW (Dynamic Time Warping) 18
4.2 Obtenção da matriz de dissimilaridade 19
4.3 Geração do MDS (*multi dimensional scale*) 20
4.3.1 Equações do MDS 20
4.4 Geração de agrupamentos através de métodos hierárquicos. 21
5.0 Análises e resultados 22
5.1 MDS bidimensional 22
5.2 Comparação entre a métrica DTW e a métrica STS 24
5.3 Geração dos dendogramas por métodos hierárquicos. 25
5.4 Resultados da geração de agrupamentos 29
5.5 Algumas conclusões importantes extraídas dos clusters 31
5.6 Geração de agrupamentos com o Aplicativo SOM. 33
5.6.1 Análise do gráfico exportado pelo SOM 36
5.6.2 Vantagens e desvantagens do SOM 36
5.7 Índices obtidos capazes de classificar a situação econômica atual do país .. 38
6.0 Conclusão 40
Bibliografia 42
ANEXOS A – Códigos comentados executados no software R 44
ANEXOS A-I Código para o calculo da matriz de dissimilaridade por Dynamic Time Warping 44

ANEXO A-II MDS COM DTW 46
ANEXO A-III Método de Ward – baseado na DTW e no MDS 47
ANEXOS B - Matriz de Dissimilaridade por DTW 48
ANEXOS C – Visualização das Séries Temporais 51

Siglas

DTW = *Dynamic time warping*, é uma métrica de dissimilaridade.

STS = *Short Time Series distance*, é uma métrica de dissimilaridade.

Cluster = equivale a agrupamento.

SOM = aplicativo que utiliza redes neurais para criar agrupamentos.

SGS = Sistema Gerenciador de Séries Temporais, base de dados do Banco Central do Brasil, que possui diversas séries temporais.

MDS = *multi dimensional scale*, constrói um gráfico bi ou tridimensonal baseado na dissimilaridade entre as observações.

1.0 Introdução

Métodos de geração de agrupamentos vêm ganhando espaço em mineração de dados, sendo aplicados a diversas áreas do conhecimento desde medicina em ressonâncias magnéticas até em geologia em previsão de terremotos.

Existem três grandes grupos de métodos de geração de agrupamentos: os que trabalham os dados crus, sem modificar as series, lembrando que padronização não é considerada um modificação; existem um segundo grande grupo que constroem agrupamentos baseados em características e o terceiro apóia a construção dos agrupamentos em modelos como ARMA, ARIMA entre outros. Nesse trabalho será aplicado o primeiro grupo os que trabalham com os dados crus.

No que diz respeito a métodos capazes de gerar agrupamentos baseado em dados crus, muitos fatores podem mudar de um método para o outro, cada método é construído par atender séries de tamanhos iguais ou desiguais, alguns métodos como aqueles que se utilizam da dissimilaridade por DTW permitem tanto serie de tamanhos iguais quanto desiguais. A escolha da medida de dissimilaridade é um fator que diferencia os modelos, no *paper* do Liao ele apresenta nove medidas de dissimilaridade embora exista um numero ainda maior. Os métodos de geração de algoritmos também variam, podendo ser usado métodos hierárquicos, fuzzy, c-médias, k-médias e redes neurais. Por fim, o critério de avaliação de homogeneidade do cluster é o ultimo fator que diferencia um modelo do outro, alguns se utilizam deles e outros não. Nota-se uma grande gama de possibilidades para a construção de modelos, ainda que só falasse de modelos baseados nos dados crus.

Liao faz uma revisão de métodos de geração de agrupamentos em seu *paper* e mostra as diversas aplicações que estão sendo desenvolvidas utilizando essas

técnicas. A tabela que apresenta as aplicações utilizando dados crus foi retirada do Liao e pode ser vista abaixo.

Summary of raw-data-based time series clustering algorithms

Paper	Variable	Length	Distance measure	Clustering algorithm	Evaluation criterion	Application
Golay	Single	Equal	Euclidean and two cross-correlation-based	Fuzzy *c*-means	Within cluster variance	Functional MRI brain activity mapping
Kakizawa	Multiple	Equal	J divergence and symmetric Chernoff information divergence	Agglomerative hierarchical	N/A	Earthquakes and mining explosions
Košmelj and Batagelj	Multiple	Equal	Euclidean	Modified relocation clustering procedure	Generalized Ward criterion function	Commercial energy consumption
Kumar	Single	Equal	Based on the assumed independent Gaussian models of data errors	Agglomerative hierarchical	N/A	Seasonality pattern in retails
Liao	Multiple	Equal & unequal	Euclidean and symmetric version of Kullback–Liebler distance	*k*-Means and fuzzy *c*-Means	Within cluster variance	Battle simulations
Liao	Single	Equal & unequal	DTW	*k*-Medoids-based genetic clustering	Several different fitness functions	Battle simulations
Möller-Levet	Single	Equal	Short time series (STS) distance	Modified fuzzy *c*-means	Within cluster variance	DNA microarray
Policker and Geva	Single	Equal	Euclidean	Fuzzy clustering by Gath and Geva	Symmetric Kullback–Leibler distance between probability function pairs	Sleep EEG signals
Shumway	Multiple	Equal	Kullback–Leibler discrimination information measures	Agglomerative hierarchical	N/A	Earthquakes and mining explosions
Van Wijk and van Selow	Single	Equal	Root mean square	Agglomerative hierarchical	N/A	Daily power consumption
Wismüller et al.	Single	Equal	N/A	Neural network clustering performed by a batch EM version of minimal free energy vector quantization	Within cluster variance	Functional MRI brain activity mapping

Tabela 1.1: Clustering of time series data — a survey [Liao, 2005]

2.0 Motivação

De acordo com Caiado et al. (2006), em economia, nós podemos estar interessados em classificar a situação econômica de um país apenas olhando em alguns indicadores de séries temporais, tal como o Produto Interno Bruto, Gastos com Investimentos, Gastos com Consumo, ou até mesmo, Taxa de Desemprego.

3.0 Objetivos

O trabalho tem como objetivos principais:

1. Descobrir e Visualizar padrões em Séries Temporais;
2. Identificar Séries semelhantes (seqüências que combinam);
3. Detectar Anomalias;
4. Gerar agrupamentos por métodos hierárquicos;
5. Identificar quais indicadores são relevantes para classificação da situação do país.

4.0 Metodologia

Para iniciar o estudo adquiriu-se a base de dados, foram coletadas trinta e sete séries temporais do “Sistema Gerenciador de Séries Temporais (SGS)” do Banco Central, no site

https://www3.bcb.gov.br/sgspub/localizarseries/localizarSeries.do?method=prepararTelaLocalizarSeries.

No SGS cada série temporal possui um número associada a ela, os números e as respectivas séries coletadas são apresentados abaixo, elas estão também classificadas por tipo.

Séries: Indicadores de antecedentes para nível de atividade

- Produção Industrial:
  - S1 - Bens de Capital (11.067)
  - S2 - Bens Intermediários (11.068)
  - S3 - Bens de Consumo Duráveis (11.070)
  - S4 - Bens de Consumo Não-Duráveis e Semiduráveis (11.071)
- Produto Interno Bruto:
  - S10 - Mensal (4.380)
  - S11 - Acumulado em 12 meses (4.382)
- Volume de Vendas no Varejo:
  - S29 - Combustíveis e Lubrificantes (1.483)
  - S33 - Hipermercados e Supermercados (1.561)
  - S30 - Tecido, Vestuário e Calçado (1.509)
  - S31 - Móveis e Eletrodomésticos (1.522)
  - S32 - Automóveis, Motocicletas, Peças (1.548)
- Consumo de Energia Elétrica:
  - S34 - Comercial (1.402)
  - S35 - Residencial (1.403)
  - S36 - Industrial (1.404)
  - S37 - Outros (1.405)
- Taxa de Desemprego (10.777)
- Tempo Médio de Procura de Emprego:

- S6 - Até 30 dias (10.796)
- S7 - De 31 dias até 6 meses (10.797)
- S8 - De 7 meses a 11 meses (10.798)
- S9 - Mais de 1 ano (10.799)

- Exportações:
  - S22 - Total (kg) (4.193)
  - S23 - Produtos Básicos (kg) (4.194)
  - S24 - Produtos Semifaturados (kg) (4.221)
  - S25 - Produtos Manufaturados (kg) (4.248)
  - S18 - Total (US$) (2.946)
  - S19 - Produtos Básicos (US$) (2.947)
  - S20 - Produtos Semifaturados (US$) (2.974)
  - S21 - Produtos Manufaturados (US$) (3.001)
- Importações:
  - S12 - Total (US$) (3.034)
  - S13 - Matérias-Primas e Produtos Intermediários (US$) (3.050)
  - S14 - Bens de Capital (US$) (3.062)
  - S15 - Total (kg) (4.281)
  - S16 - Matérias-Primas e Produtos Intermediários (kg) (4.297)
  - S17 - Bens de Capital (kg) (4.309)
- Investimentos Diretos:
  - S26 - Total Líquido (2.851)
  - S28 - Estrangeiro Direto (2.860)
  - S27 - Brasileiro Direto (2.852)

As séries podem ser visualizadas nos ANEXOS separadas em grupos.

O processo de geração de agrupamentos reside em 4 etapas:

1. Adquirir a base de dados
2. Escolher a métrica de dissimilaridade
3. Calcular as coordenadas das series em um espaço bidimensional baseado na dissimilaridade
4. Escolher o método hierárquico de geração de agrupamentos mais conveniente

Como o propósito do trabalho é conseguir classificar a situação econômica do país, sempre no dia de hoje, não faz muito sentido trabalhar com séries de tamanhos desiguais apesar do fato da DTW permitir que as séries tenham tamanhos distintos entre si. Assim, as séries estudadas possuem 91 medidas mensais (*time points*) de out/01 até abr/09, não possuindo valores faltantes (*missing values*), como mostra a tabela abaixo.

Quantidade de séries estudadas	37
Comprimento	Equal
Time points de cada série	91 (meses)
Primeiro *time point*	Outubro de 2001
Último *time point*	Abril de 2009
Valores faltantes	Nenhum
Intervalo entre *time points* adjacentes	1 mês

Tabela 4.1: descrição do banco de dados.

Nota: É importante ressaltar que todas as séries foram padronizadas para que tivessem média zero e desvio padrão 1. Isso é importante no caso de geração de

agrupamentos de séries temporais, pois algumas séries são índices, outras tem como unidade R$, outras são em Gigawatts (GW).

A etapa 2 é escolher uma métrica de dissimiliridade capaz de gerar clusters interpretáveis, não existe uma regra bem definida sobre qual métrica é mais indicada para esse ou aquele grupo de séries, geralmente a escolha faz-se após a experimentação.
No caso do presente trabalho foram testadas três métricas:

1. **STS- *short time series*** - ela se baseia parte do pressuposto que entre um *time point* e outro as séries podem ser escritas como segmentos de retas. Assim, essa métrica compara a inclinação desses segmentos para medir a dissimilaridade.
2. **Distância baseado em um índice de correlação cruzada** - através de uma correlação cruzada de duas séries é medido a dissimilaridade entre elas.
3. **DTW**- é um método construído para bioinformática muito eficiente para o calculo da dissimilaridade entre duas séries.

Será feito a descrição apenas do método do DTW, isso, pois foi o que gerou resultados mais relevantes. As métricas 1 e 2 não formaram agrupamentos muito nítidos.
A STS apresentou deficiências quando aplicada a séries temporais econômicas, pois ela considera que séries que possuem movimentos opostos (quando uma cresce a outra diminui) possuem a maior dissimilaridade, contudo isso não é verossímil na pratica, pois séries com comportamento absolutamente opostos são fenômenos regidos pelas mesmas leis, mas que reagem de maneira inversa as variações, devendo ter baixa dissimilaridade e não alta como calcula o STS.

4.1 Descrição do DTW (Dynamic Time Warping)

O DTW (*dynamic time warping*) ou alinhamento temporal dinâmico é a generalização dos algoritmos clássicos para comparar sequências discretas com sequências de valores contínuos, muito usado também para comparar sequências de DNA/RNA . Dado duas séries temporais, *Q=q1, q2, . . . , qi, . . . , qn* e *R =r1, r2, . . . , rj , . . . , rm*, DTW alinha as duas séries minimizando suas diferenças. Para esse fim, é construída uma matriz $n \times m$ onde o elemento (i , j) contem a distância $d\left(q_i, r_j\right)$ entre dois pontos q_i e r_j. A distância euclidiana é normalmente usada.

Depois de construída a matriz de distâncias, é necessário obter-se o caminho *warping (warping path*), afinal a dissimilaridade depende disso.

O caminho *warping*, *W =w1,w2, . . . , wk, . . . ,* wk onde $\max(m,n) \le K \le m+n-1$, é o conjunto dos elementos da matriz de distancia que satisfaz três condições: condição de contorno , continuidade e monotonicidade.

A condição de contorno requer que o caminho path inicie-se em $w_1 = (1,1)$ e termina em $w_k = (m,n)$. Já a condição de continuidade restringe que os passos dados na matriz de distancia para construir o caminho *warping* sejam dados para células adjacentes. Por fim, a condição de monotonicidade força que os pontos no caminho *warping* sejam monotonicamente espaçados no tempo.

Levado isso em conta, o caminho *warping* é aquele que minimiza a distância entre as duas séries temporais de interesse. Matematicamente temos:

$$d_{DTW} = \min \frac{\sum_{k=1}^{K} w_k}{K}$$

Programação dinâmica pode ser usada de maneira muito eficiente para encontrar o caminho *warping*, utilizando o algoritmo abaixo.

$$dcum(i, j) = d(qi, rj) + \min\{dcum(i-1, j-1), dcum(i-1, j), dcum(i, j-1)\}.$$

O algoritmo computacional para calcular o DTW foi feito no software R e é apresentado nos Anexos.

4.2 Obtenção da matriz de dissimilaridade

O método DTW é uma maneira de obter a dissimilaridade entre duas séries temporais, contudo para que a geração de agrupamentos seja possível necessita-se da dissimilaridade de uma com todas e todas com uma. Logo, precisa-se de uma matriz de dissimilaridades, que no caso desse estudo contenha 37×37 elementos, uma vez que temos 37 series.
O algoritmo para o calculo da matriz de dissimilaridade foi feito no software R e é apresentado nos Anexos.
As medidas de dissimilaridade possuem algumas propriedades, elas são:

1. d(x,y) ³ 0
2. Simetria: d(x,y) = d(y,x)
3. Se d(x,y) $\neq$ 0 então x $\neq$ y
4. Se x = y, então d(x,y) = 0

A matriz de dissimilaridade pode ser visualizada no Anexo B.

4.3 Geração do MDS (*multi dimensional scale*)

Tendo como entrada a matriz de dissimilaridade, construiu-se um código no R (apresentado em anexo) capaz de gerar coordenadas e *plotar* um gráfico bidimensional, onde pontos próximos são mais similares que pontos distantes. O MDS é uma boa ferramenta de redução de dimensão e visualização de cluster, pois é uma foto do melhor plano bidimensional para visualizar as diferenças entre as séries.

Existem duas formas de MDS: métrico (assume-se que as dissimilaridades são proporcionais às distâncias Euclidianas) e não-métrico (assume-se que as dissimilaridades são relacionadas às distâncias Euclidianas por alguma transformação monotônica).

4.3.1 Equações do MDS

É construída uma função chamada STRESS que deve ser minimizada, ela é responsável pela decomposição das dissimilaridades.

$$Min\ STRESS = \left[\sum_i \sum_j \left(d_{ij} - D_{ij} \right)^2 \Big/ \sum_i \sum_j {d_{ij}}^2 \right]$$

Em que Dij é a dissimilaridade entre as observações i e j e dij é a distância euclidiana entre as coordenadas das observações i e j no espaço k-dimensional,

calculada por:

$$d_{ij} = \sqrt{\sum_{v=1}^{k<p} \left(x_{iv} - x_{jv} \right)^2}$$

4.4 Geração de agrupamentos através de métodos hierárquicos.

A geração de *clusters* não precisa ser feita por métodos hierárquicos, contudo eles têm uma vantagem que é a formação de um dendograma que permite visualizar agrupamentos naturais. Os dendogramas são construídos baseados na matriz de dissimilaridade, que no caso desse trabalho, foi obtida utilizando DTW.

Nos Anexos consta o código fonte do programa que gerou os dendogramas, o código foi feito em R.

Existem cinco métodos hierárquicos de geração de agrupamentos: Ligação simples, Ligação média, Ligação máxima, Método do centróide e Método de Ward. Como veremos mais adiante o método de ward tende a fazer clusters mais esféricos e com numero semelhantes de observações em cada um deles, e no caso do presente trabalho, essa característica viesou o resultado, dando ao método do centróide o titulo de melhor método hierárquico para a geração de agrupamentos de séries temporais econômicas.

5.0 Análises e resultados

5.1 MDS bidimensional

Como foi dito, a partir da matriz de dissimilaridade construída com DTW foi possível gerar coordenadas espaciais para cada uma das séries temporais, as coordenadas plotadas em um gráfico bidimensional é apresentado abaixo.

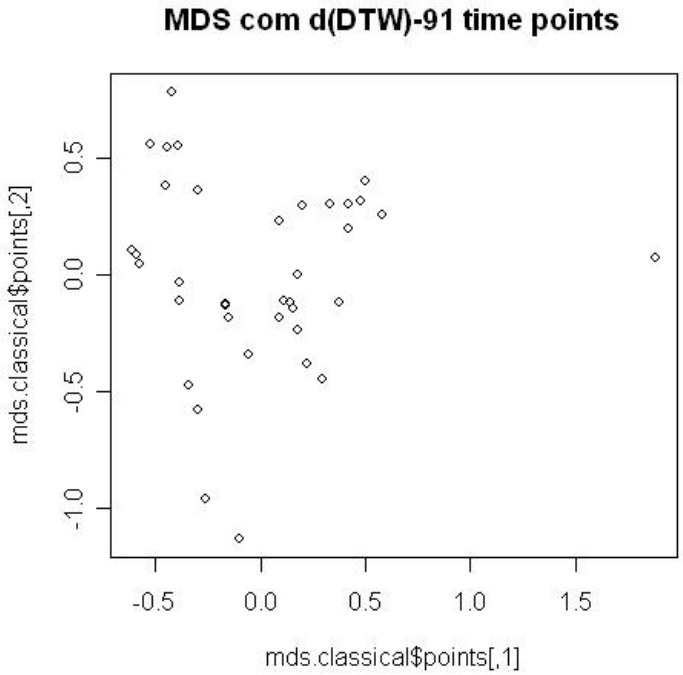


Figura 5.1: MDS construído com DTW das 37 séries temporais

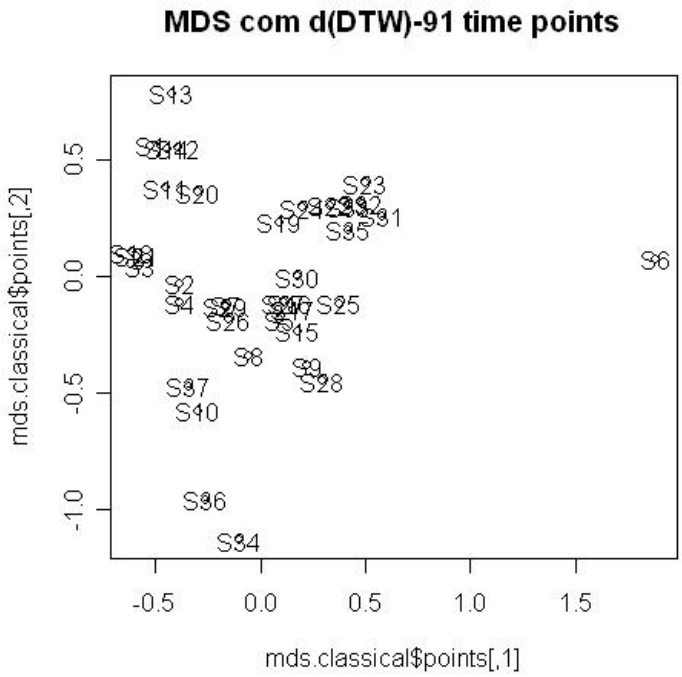


Figura 5.2: MDS com rótulos nos pontos, construído com DTW para as 37 séries temporais.

Os gráficos acima têm a seguinte propriedade: séries temporais que estão próximas são mais semelhantes entre si, que séries temporais que estão distantes. Sem a necessidade de nenhum método adicional notamos que a série S6 comporta-se como um *outlier* sendo distante de todas as outras séries.

S6	10796 - Distribuição de desocupados segundo o tempo de procura de emprego - Até 30 dias - %

Explicação do porquê a S6 é aparentemente um *outlier*: É de pensar ao olhar essa série que se o desemprego aumenta isso deveria refletir uma situação ruim do país, contudo ela aponta pessoas desempregadas até 30 dias, e se uma pessoa não passa mais do que trinta dias desempregada é de se imaginar que o país esteja em boa situação econômica , assim essas duas possíveis

interpretações contraditórias se refletem na matriz de dissimilaridade e mostra que é impossível prever a situação do pais olhando para a porcentagem de desempregados até 30 dias.

5.2 Comparação entre a métrica DTW e a métrica STS

Como foi dito nos itens anteriores a DTW mostrou-se uma métrica mais indicada para gerar agrupamentos, contudo anteriormente não foi dito o porquê ela é mais indicada. Uma boa métrica permite que a dissimilaridade ressalte as diferenças, já uma métrica inadequada não consegue deixar claras as diferenças entre as diversas séries.

Far-se-á agora uma comparação entre a DTW e a STS, onde será nítido que a DTW é uma métrica mais adequada para o estudo em questão.

Não foi feito uma grande explanação sobre a métrica STS, mas isso não é importante, pois o foco desse item não está sobre as propriedades da STS e sim sobre a comparação entre duas métricas. As conclusões das comparações devem advir dos resultados e não de suas propriedades especificas.

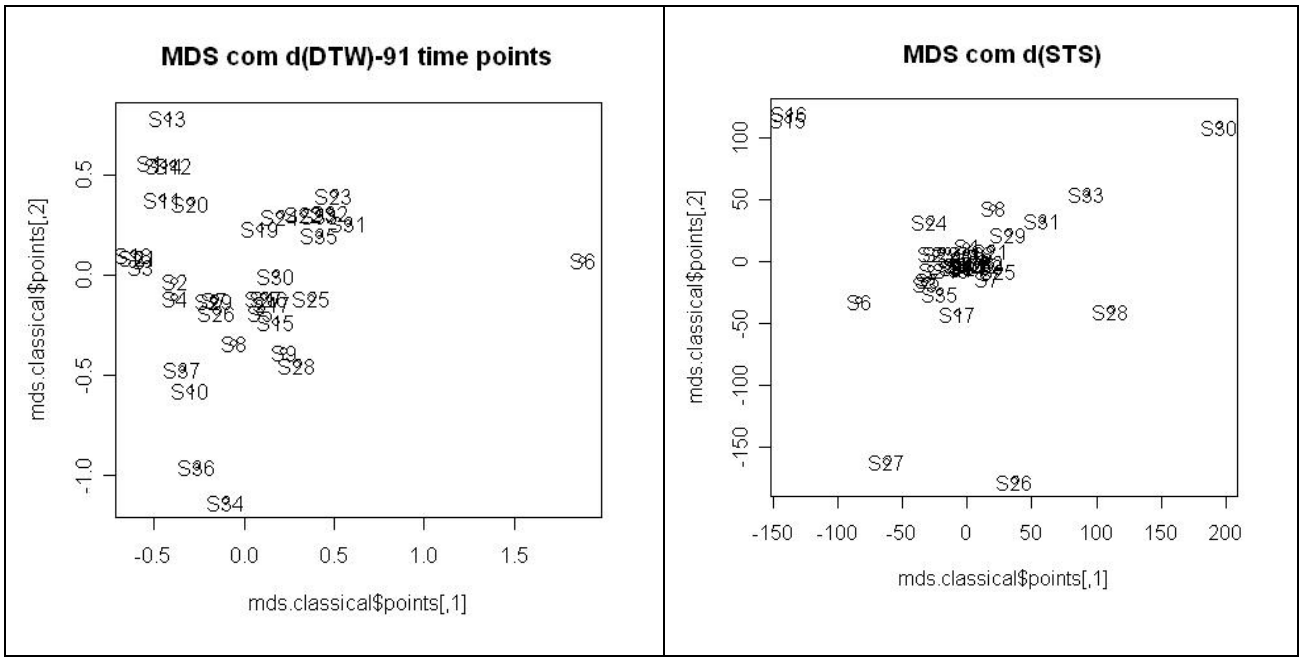


Figura 5.3: Comparação entre os MDS formados por DTW e STS

Note que o MDS formado com a métrica STS não permite visualizar claramente as diferenças entre varias séries, tão pouco é possível enxergar facilmente a existência de clusters naturais. Essa visualização lado a lado permite justificar que a DTW é uma métrica mais aconselhável que a STS para a construção de agrupamentos das séries estudadas.

5.3 Geração dos dendogramas por métodos hierárquicos.

Foi escolhido o método do centróide para a construção dos agrupamentos devido a sua robustez e capacidade de gerar clusters significativamente interpretativos. O código fonte do programa que gerou o dendograma foi executado no software R e pode ser visto nos Anexos .

Abaixo podemos ver o dendograma formado pelo método do centróide.

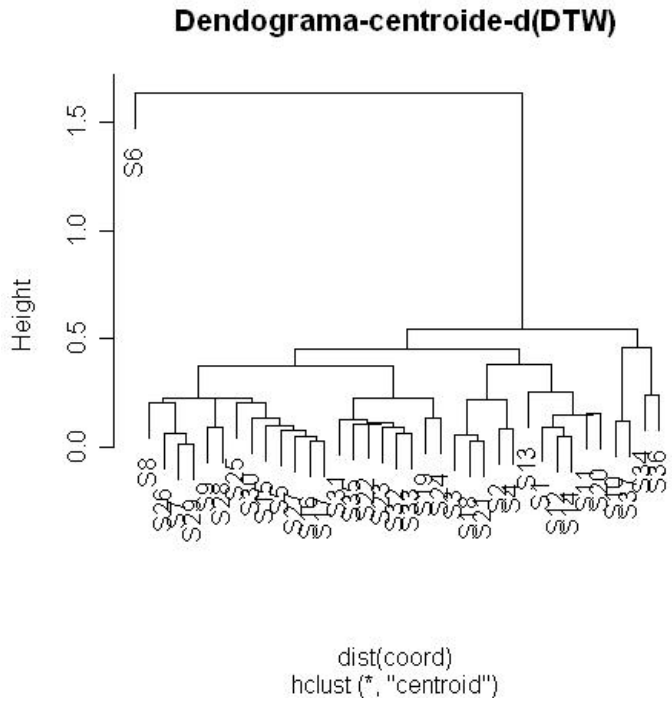


Figura 5.4: dendograma formado pelo método do centróide utilizando a matriz de dissimilaridade por DTW

Esse dendograma reafirma o que havia sido dito sobre a série S6 ser um *outlier*, note que a altura de ligação é proporcional a diferença/dissimilaridade.
Pode-se observar a formação de sete agrupamentos naturais sendo que um deles contém apenas a série S6 que tem características de *outlier*. No gráfico abaixo o realce dos *clusters* naturais é evidenciado.

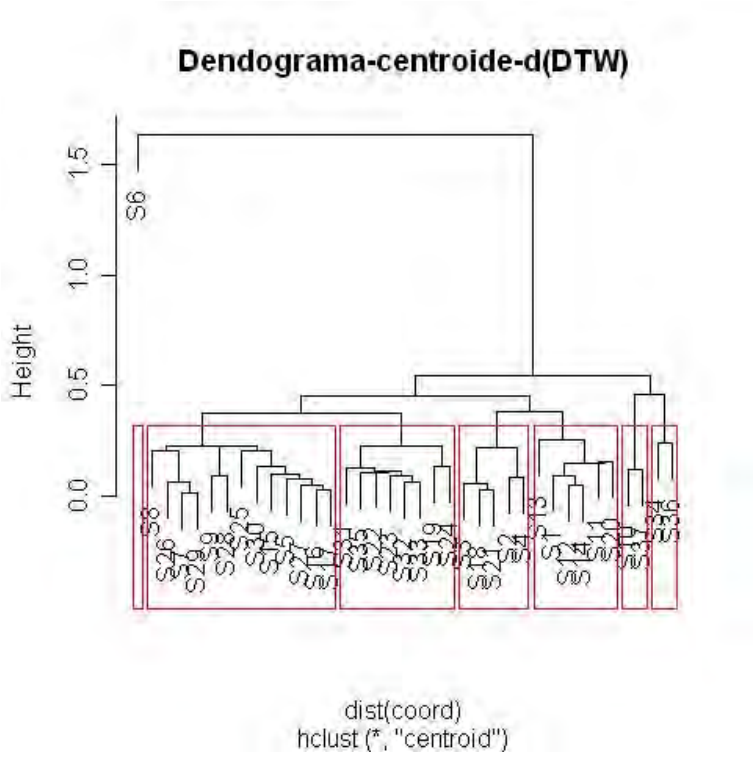


Figura 5.5: realce dos sete agrupamentos naturais formados pelo método do centróide utilizando a matriz de dissimilaridade construída por DTW

É possível unir as informações do dendograma com as informações do MDS gerando o gráfico abaixo:

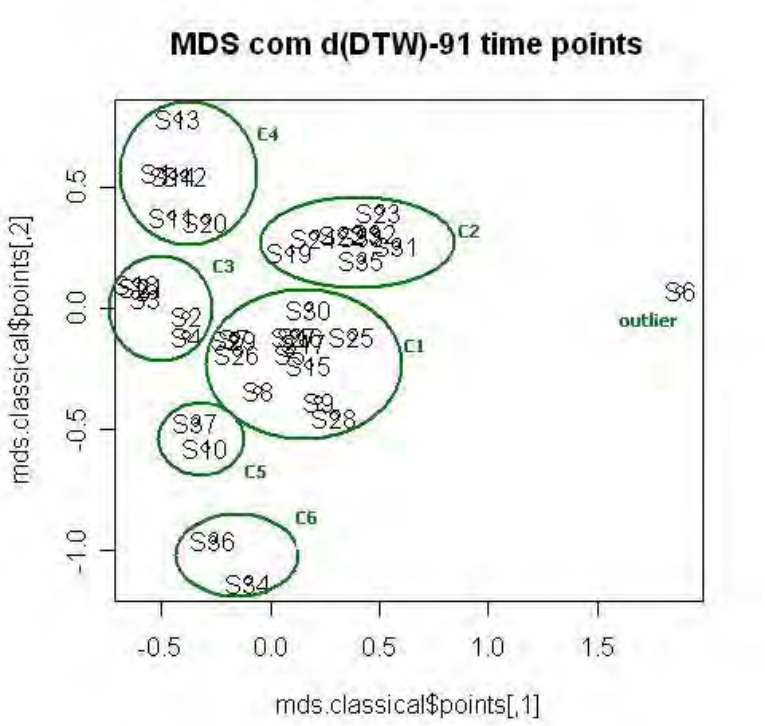


Figura 5.6: realce dos clusters formados pelo método do centróide no gráfico do MDS

5.4 Resultados da geração de agrupamentos

Cluster C1	Nº série	Nome da série
S8	10798	Distribuição de desocupados segundo o tempo de procura de emprego - De 7 a 11 meses - %
S26	2851	Investimento direto total (líquido) - mensal - US$ milhões
S7	10797	Distribuição de desocupados segundo o tempo de procura de emprego - De 31 dias a 6 meses - %
S29	1483	Índice volume de vendas no varejo (2003=100) - Combustíveis e lubrificantes - Brasil - Índice
S9	10799	Distribuição de desocupados segundo o tempo de procura de emprego - Mais de 1 ano - %
S28	2860	Investimento estrangeiro direto - IED (líquido) - mensal - US$ milhões
S25	4248	Exportações (Kg) - Produtos manufaturados – Kg
S30	1509	Índice volume de vendas no varejo (2003=100) - Tecido, vestuário e calçado - Brasil - Índice
S15	4281	Importações (Kg) - Total - Kg
S5	10777	Taxa de desemprego - Região metropolitana - Brasil (na semana) - %
S27	2852	Investimento brasileiro direto - IBD (líquido) - mensal - US$ milhões
S16	4297	Importações (Kg) - Matérias-primas e produtos intermediários - Kg
S17	4309	Importações (Kg) - Bens de capital - Kg

Tabela 5.1 Séries que compõem o cluster C1 formado pelo método do centróide

Cluster C2	Nº série	Nome da série
S31	1522	Índice volume de vendas no varejo (2003=100) - Móveis e eletrodomésticos - Brasil - Índice
S35	1403	Consumo de energia elétrica - Brasil - Residencial - GWh
S22	4193	Exportações (Kg) - Total - Kg
S23	4194	Exportações (Kg) - Produtos básicos - Kg
S32	1548	Índice volume de vendas no varejo (2003=100) - Automóveis, motocicletas, partes e peças - Brasil - Índice
S33	1561	Índice volume de vendas no varejo (2003=100) - Hipermercados e supermercados - Brasil - Índice
S19	2947	Exportações - Produtos básicos - US$
S24	4221	Exportações (Kg) - Produtos semimanufaturados - Kg

Tabela 5.2: Séries que compõem o cluster C2 formado pelo método do centróide

Cluster C3	Nº série	Nome da série
S3	11070	Indicadores da produção (2002=100) - Por categoria de uso - Bens de consumo (duráveis) - Índice
S18	2946	Exportações - Total - US$
S21	3001	Exportações - Produtos manufaturados - US$
S2	11068	Indicadores da produção (2002=100) - Por categoria de uso - Bens intermediários - Índice
S4	11071	Indicadores da produção (2002=100) - Por categoria de uso - Bens de consumo (não-duráveis e semiduráveis) - Índice

Tabela 5.3: Séries que compõem o cluster C3 formado pelo método do centróide

Cluster C4	Nº série	Nome da série
S13	3050	Importações - Matérias-primas e produtos intermediários - US$
S1	11067	Indicadores da produção (2002=100) - Por categoria de uso - Bens de capital - Índice
S12	3034	Importações - Total - US$
S14	3062	Importações - Bens de capital - US$
S11	4382	PIB acumulado dos últimos 12 meses - Valores correntes (R$ milhões) - R$ (milhões)
S20	2974	Exportações - Produtos semimanufaturados - US$

Tabela 5.4: Séries que compõem o cluster C4 formado pelo método do centróide

Cluster C5	Nºsérie	Nome da série
S10	4380	PIB mensal - Valores correntes (R$ milhões) - R$ (milhões)
S37	1405	Consumo de energia elétrica - Brasil - Outros - GWh

Tabela 5.5: Séries que compõem o cluster C5 formado pelo método do centróide

Cluster C6	Nº série	Nome da série
S34	1402	Consumo de energia elétrica - Comercial - GWh
S36	1404	Consumo de energia elétrica - Brasil - Industrial - GWh

Tabela 5.6: Séries que compõem o cluster C6 formado pelo método do centróide

outlier	Nº série	Nome da série
S6	10796	Distribuição de desocupados segundo o tempo de procura de emprego - Até 30 dias - %

Tabela 5.7: Série que compõe o cluster intitulado outlier

5.5 Algumas conclusões importantes extraídas dos clusters

É possível extrair muitas conclusões tendo em mãos a relação das séries que formam os clusters, tantas que seria inviável listar todas. As conclusões dependem do tipo de enfoque, do objetivo da análise e da visão particular do analista. Extrair conclusões não é uma ciência exata, nem tão pouco há como saber, a priori, quantas conclusões são possíveis de se extrair de um dado conjunto de clusters.

Aqui serão apresentadas algumas poucas conclusões para reforçar a robustez do modelo adquirido.

**As conclusões abaixo não estão distribuídas em ordem de importância ou prioridade, portanto a Conclusão 1 não é, necessariamente, mais relevante que a Conclusão 2.

Conclusão 1: Ao se olhar o **cluster C3**, se vê um detalhe da realidade brasileira e uma indicação de como deveria ser norteada a política econômica. Note que esse cluster tem cinco séries, onde duas são exportações e três são indicadores econômicos de diversos bens de consumo, a conclusão direta é que as exportações ajudam a melhorar os índices de consumo. Contudo a conclusão implícita, decorre da série S21 (Exportações de produtos manufaturados- US$) , mostrando que apesar de exportarmos mais produtos básicos, o que gera distribuição de renda são os produtos manufaturados, pois a exportação desses,

gera mais empregos e empregados com poder aquisitivo capaz de influenciar os índices de consumo.

Ao perceber que a exportação de produtos básicos não está presente nesse cluster concluí-se que os trabalhadores desse ramo não são qualificados (em sua maioria), têm salário que serve apenas para a subsistência básica não alterando assim os indicadores de consumo. Nota-se também que a exportação de produtos básicos gera concentração de renda, pois os grandes agricultores apesar de perceberem grandes somas de dinheiro não consomem o suficiente para alterar de maneira significativa os índices de consumo.

Conclusão 2: Ao se olhar o **cluster C1** tem-se a relação de S26,S27,S28 que referem-se a Investimentos diretos, com as séries S7, S8, S9, S5 que remetem a taxas de desemprego de períodos mais longos, isso revela que a falta de investimento direto impacta sobre a qualidade de vida das cidades e sobre os indivíduos que são mão de obra mais qualificada gerando taxa de desemprego mais elevadas no longo prazo.

5.6 Geração de agrupamentos com o Aplicativo SOM.

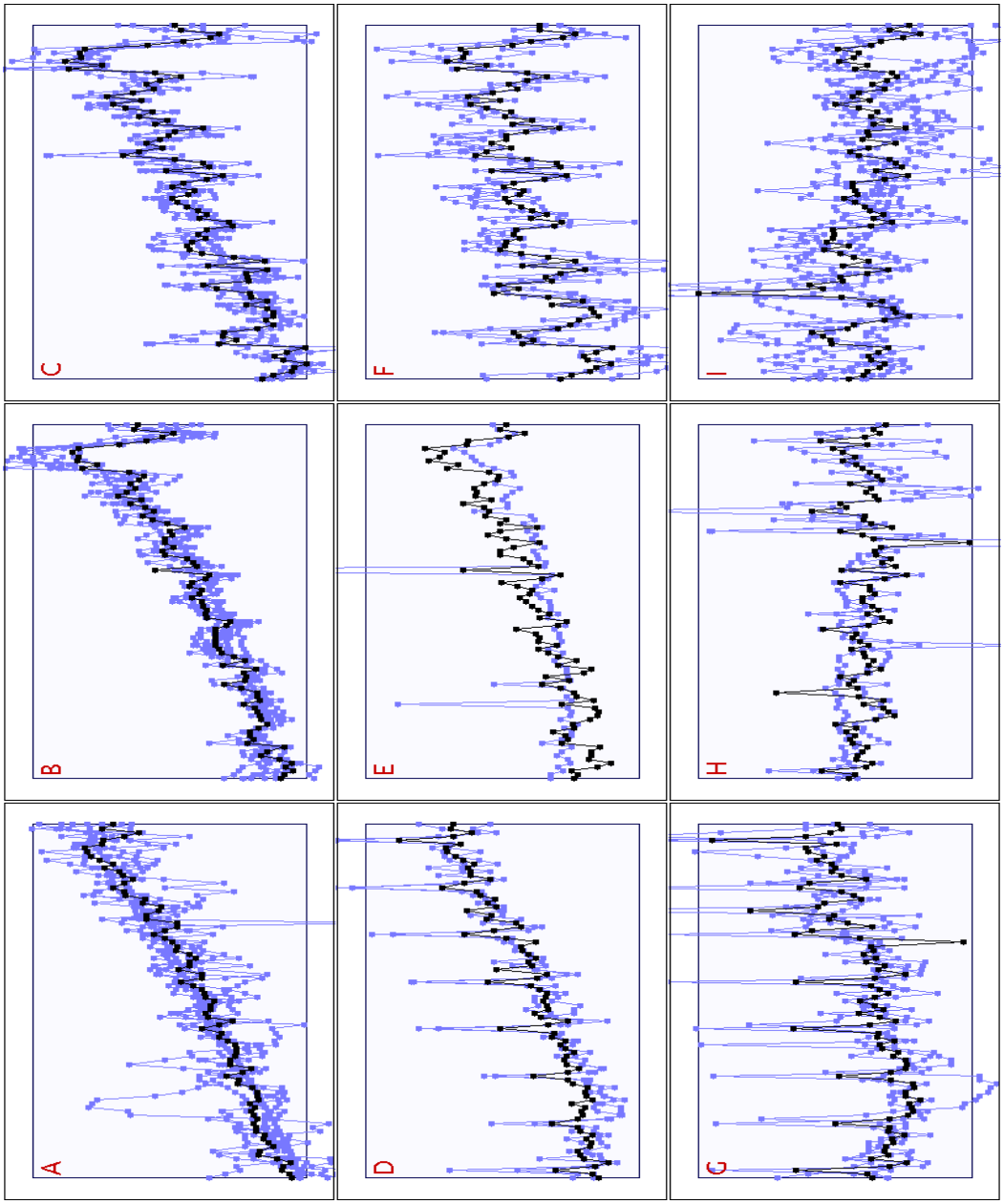


Figura 5.7: gráfico gerado pelo aplicativo SOM (disposição da folha = paisagem), linha preta é o centróide e as azuis são as séries associadas

O SOM é um aplicativo que tenta agrupar as séries em *clusters*, mas sem que seja necessário dizer o número exato de clusters, ele tenta distribuir as séries pelos elementos da rede neural.

O Aplicativo possui neurônios, estes por sua vez possuem centróide e cada neurônio tenta capturar as séries mais semelhantes a eles. É possível que um neurônio não capture série alguma, embora no caso acima termos que cada neurônio capturou pelo menos uma série.

Neuron A	
S7	10797 - Distribuição de desocupados segundo o tempo de procura de emprego - De 31 dias a 6 meses - %
S10	4380 - PIB mensal - Valores correntes (R$ milhões) - R$ (milhões)
S11	4382 - PIB acumulado dos últimos 12 meses - Valores correntes (R$ milhões) - R$ (milhões)
S32	1548 - Índice volume de vendas no varejo (2003=100) - Automóveis, motocicletas, partes e peças - Brasil - Índice
S34	1402 - Consumo de energia elétrica - Comercial - GWh
S35	1403 - Consumo de energia elétrica - Brasil - Residencial - GWh
S37	1405 - Consumo de energia elétrica - Brasil - Outros - GWh
Neuron B	
S1	11067 - Indicadores da produção (2002=100) - Por categoria de uso - Bens de capital - Índice
S12	3034 - Importações - Total - US$
S13	3050 - Importações - Matérias-primas e produtos intermediários - US$
S14	3062 - Importações - Bens de capital - US$
S18	2946 - Exportações - Total - US$
S19	2947 - Exportações - Produtos básicos - US$
S36	1404 - Consumo de energia elétrica - Brasil - Industrial - GWh
Neuron C	
S2	11068 - Indicadores da produção (2002=100) - Por categoria de uso - Bens intermediários - Índice
S3	11070 - Indicadores da produção (2002=100) - Por categoria de uso - Bens de consumo (duráveis) - Índice
S20	2974 - Exportações - Produtos semimanufaturados - US$
S21	3001 - Exportações - Produtos manufaturados - US$
S22	4193 - Exportações (Kg) - Total - Kg
S23	4194 - Exportações (Kg) - Produtos básicos - Kg
Neuron D	
S31	1522 - Índice volume de vendas no varejo (2003=100) - Móveis e eletrodomésticos - Brasil - Índice
S33	1561 - Índice volume de vendas no varejo (2003=100) - Hipermercados e supermercados - Brasil - Índice
Neuron E	
S17	4309 - Importações (Kg) - Bens de capital - Kg
Neuron F	
S4	11071 - Indicadores da produção (2002=100) - Por categoria de uso - Bens de consumo (não-duráveis e semiduráveis) - Índice

S24	4221 - Exportações (Kg) - Produtos semimanufaturados - Kg
S25	4248 - Exportações (Kg) - Produtos manufaturados - Kg
Neuron G	
S6	10796 - Distribuição de desocupados segundo o tempo de procura de emprego - Até 30 dias - %
S28	2860 - Investimento estrangeiro direto - IED (líquido) - mensal - US$ milhões
S30	1509 - Índice volume de vendas no varejo (2003=100) - Tecido, vestuário e calçado - Brasil - Índice
Neuron H	
S26	2851 - Investimento direto total (líquido) - mensal - US$ milhões
S27	2852 - Investimento brasileiro direto - IBD (líquido) - mensal - US$ milhões
Neuron I	
S5	10777 - Taxa de desemprego - Região metropolitana - Brasil (na semana) - %
S8	10798 - Distribuição de desocupados segundo o tempo de procura de emprego - De 7 a 11 meses - %
S9	10799 - Distribuição de desocupados segundo o tempo de procura de emprego - Mais de 1 ano - %
S15	4281 - Importações (Kg) - Total - Kg
S16	4297 - Importações (Kg) - Matérias-primas e produtos intermediários - Kg
S29	1483 - Índice volume de vendas no varejo (2003=100) - Combustíveis e lubrificantes - Brasil - Índice

Figura 5.8: lista das series por neurônio da aplicação SOM

5.6.1 Análise do gráfico exportado pelo SOM

Pode ser visto a partir da figura que o número de séries de um neurônio para outro variam, podemos notar também que o neurônio B mostrou-se mais homogêneo, essa conclusão decorre das séries estarem mais próximas da linha preta que é o centróide. O neurônio E ficou com uma única observação, a série S17 – (4309 - Importações (Kg) - Bens de capital – Kg), para entender melhor o motivo disso um especialista em economia deveria ser consultado.

É possível notar que o neurônio G e o I já são bem poucos homogêneos quando comparados com o neurônio B. Nota-se que o neurônio G é um cluster que possui séries sazonais, devido ao centróide e as séries possuírem picos sazonais.

A rede neural optou por 9 *clusters* quando não deixou nenhum neurônio vazio, essa quantidade de *clusters* foi diferente da encontrada usando a métrica de dissimilaridade DTW, por isso a comparação da interpretação dos *clusters* se tornou difícil, mas isso não diminui o potencial do aplicativo que apresenta vantagens e desvantagens.

Mais abaixo poderá ser vista a lista das séries em cada neurônio.

5.6.2 Vantagens e desvantagens do SOM

Empiricamente nota-se que redes neurais geralmente são inferiores para geração de agrupamentos e classificação, embora tenha alta capacidade preditiva em séries temporais por exemplo.

Na questão especifica do aplicativo SOM ele tem algumas vantagens e desvantagens em relação aos métodos de geração de agrupamentos, elas serão apresentadas abaixo:

- Vantagens:

1. Não carece de escolha de uma métrica de dissimilaridade, como DTW;
2. Permite a visualização das séries inteiras nos clusters e não apenas como um ponto como é o caso do MDS;
3. Não necessita saber o número de clusters a priori, a própria rede neural escolhe quais e quantos neurônios são suficientes para um agrupamento eficaz;
4. Visualmente é possível reconhecer quais clusters são mais homogêneos, sem a necessidade de métricas como no caso de agrupamentos formados pelo Método de Ward.

- Desvantagens:

1. Não possui dendograma como os métodos de geração hierárquicos impossibilitando saber se o número de neurônios escolhidos é o número ideal de agrupamentos;
2. Devido à falta de métricas de dissimilaridade não é possível saber a semelhança entre duas séries que não estejam no mesmo neurônio;
3. Como outras redes neurais os cálculos são como uma caixa preta, não gerando interpretação do por que uma série está em um cluster e não em outro;
4. A falta de métrica de homogeneidade faz com que seja difícil dizer entre dois *clusters* homogêneos qual é o mais homogêneo, ou entre dois difusos qual é o mais espalhado.

Assim podemos simplificar dizendo que o SOM tem seu ponto fraco na "interpretatividade" uma vez que é difícil saber por que uma série ficou em um

neurônio ou não, por outro lado permite de maneira eficiente a visualização concomitante das séries e dos clusters, permitindo a extração de muita informação com apenas um olhar.

5.7 Índices obtidos capazes de classificar a situação econômica atual do país

Desde o inicio do trabalho o objetivo da geração de agrupamentos das séries temporais foi identificar alguns poucos índices capazes de classificar a situação econômica do país no momento atual (dia de hoje), conseguir esses índices permite deixar de usar 37 séries e passar a usar poucos indicadores, uma vez que, séries que estão no mesmo cluster são similares e reagem de maneira similares aos fenômenos econômicos. Levando o que foi dito em conta, extraímos **quatro indicadores relevantes**, eles são citados e descritos abaixo:

Cluster	Nome do Indicador
C1	Produção Intensiva em Capital
C2	Consumo
C3	Produção para Consumo
C4	Insumos para Produção

Tabela 5.8: relaciona o cluster a sua interpretação

Dar-se-á uma breve explicação sobre os indicadores:

Produção Intensiva em Capital - cluster C1: nota-se claramente que esse cluster relaciona os investimentos diretos aos tipos de mão de obra especializada

através das taxas de desemprego, em especial a taxa de desemprego em regiões metropolitanas. O cluster reúne também importações de bens de capital, esse tipo de importação é típico de empresas intensivas em capital. Como o objetivo é reduzir o número de séries à serem analisadas, uma boa série capaz de substituir as treze séries contidas no cluster C1, para o efeito de verificar a atividade da produção intensiva em capital é a S5 - (Taxa de desemprego – Região metropolitana – Brasil (na semana) %).
A S5 é uma boa candidata a ser indicador de atividade produtiva intensivas em capital, pois para as empresas produzirem mais, necessitam de mais mão de obra especializada encontrada nas cidades.

Indicador de Consumo - Cluster C2: Esse *cluster* relaciona índice de volume de vendas no varejo a exportações. Obviamente temos a seguinte relação diretamente proporcional: um aumento nas exportações é devido ao aumento da produção dos produtos destinados a exportação, que traz consigo um aumento dos postos de trabalho, que traz consigo um aumento do poder de compra da população que se reflete em um maior consumo no varejo. Para esse cluster o consumo de energia elétrica residencial (S35) seria uma boa escolha para representar o consumo, nota-se que o consumo de eletricidade pelas casas é fortemente afetado por variações na renda familiar.

Indicador de Produção para Consumo – cluster C3: Esse *cluster* contém essencialmente indicadores de produção de bens de consumo, possui também exportações do mesmo tipo de bens.

Insumos para Produção – cluster C4: Esse *cluster* associa, principalmente, as importações e o indicador de produção de bens de capital, isso pode ser entendido como insumos para a produção, sem tais coisas a indústria não é capaz de produzir.

O cluster C5 essencialmente possui o PIB que é a métrica de avaliar o desempenho de um país, ou seja, um país cresce o que o seu PIB (produto interno bruto) cresce.

O cluster C6 mostra que tanto o consumo de energia elétrica comercial quanto industrial não são bons indicadores para classificar a situação do país, uma explicação possível para esse fato é de que o gasto de energia para esses setores é bastante inelástico com a demanda, por exemplo, um comércio não gasta menos luz se, em um dia, tem o tráfego de clientes reduzidos pela metade, pois, é claro, os clientes que foram até o comércio em questão necessitam de um nível adequado de serviço, não podendo ser atendidos no escuro.

6.0 Conclusão

Concluiu-se que o objetivo do trabalho de gerar agrupamentos hierárquicos com a finalidade de obter um número restrito de indicadores capazes de classificar a situação econômica do país foi alcançado. Obtiveram-se quatro indicadores: Produção Intensiva em Capital, Consumo, Produção para Consumo, Insumos para Produção; esses indicadores substituem a análise de 37 séries temporais para classificar a situação atual do país.

Pode-se observar durante o trabalho que existem métricas de dissimilaridade mais eficientes que outras e que a escolhida foi a DTW (*Dynamic Time Warping*), por gerar clusters mais bem definidos e mais interpretáveis. O MDS por sua vez, mostrou seu poder de reduzir as dimensões do espaço de atributos e gerar coordenadas para as séries temporais onde séries mais próximas eram mais similares que as mais distantes, a relação de proximidade entre as séries pode ser confirmado pelo dendograma.

O trabalho permitiu também constatar que os métodos de geração de agrupamentos de séries temporais são úteis para analistas, pois possibilitam revelar conclusões, antes difíceis de visualizar ou provar.

Utilizou-se também a aplicação SOM baseado em redes neurais, esta apresentou vantagens e desvantagens, que foi discutido durante o trabalho.

O software estatístico R, mostrou-se útil na implementação dos algoritmos, apesar de possuir um compilador fraco em laços de repetição, possui funções estatísticas que facilitam em muito a criação de um algoritmo, reduzindo drasticamente as linhas do código fonte.

Bibliografia

T.W. Liao, Clustering of time series data – a survey, The Journal of the pattern recoginition society (2005)

Caiado, J. , Crato, N. e Peña, D. (2006). A periodogram – based metric for time series classification. Computational Statistics & Data Analysis, 50, 2668 – 2684.

Sistema Gerenciador de Séries Temporais – SGS https://www3.bcb.gov.br/sgspub/localizarseries/localizarSeries.do?method=prepararTelaLocalizarSeries –- origem da base de dados - acessado dia 1 de setembro de 2009

X. Golay, S. Kollias, G. Stoll, D. Meier, A. Valavanis, P.Boesiger, A new correlation-based fuzzy logic clustering algorithm for fMRI, Mag. Resonance Med. 40 (1998)
249–260.

C.S. Möller-Levet, F. Klawonn, K.-H. Cho, O. Wolkenhauer, Fuzzy clustering of short time series and unevenly distributed sampling points, Proceedings of the 5th International
Symposium on Intelligent Data Analysis, Berlin, Germany, August 28–30, 2003.

M. Kumar, N.R. Patel, J. Woo, Clustering seasonality patterns in the presence of errors, Proceedings of KDD ’02, Edmonton, Alberta, Canada.

K. Košmelj, V. Batagelj, Cross-sectional approach for clustering time varying data, J. Classification 7 (1990) 99–109.

T.W. Liao, B. Bolt, J. Forester, E. Hailman, C. Hansen, R.C. Kaste, J. O'May, Understanding and projecting the battle state, 23rd Army Science Conference, Orlando, FL, December 2–5, 2002.

T.W. Liao, Mining of vector time series by clustering,Working paper, 2005.

ANEXOS A – Códigos comentados executados no software R

ANEXOS A-I Código para o calculo da matriz de dissimilaridade por Dynamic Time Warping

#d(DTW)

```
setwd("C:/INPE/Séries Scarpel")
series=numeric()
series = read.csv("series-padronizada-equal-.csv", header = TRUE, dec=".", sep=";", row.names=1)
 # deve ser salvo como .csv separados por virgulas

dissimilaridade = array (0, dim =c(37,37))  # esse array é a matriz de dissimilaridade, é 37x37 pois tem-se 37 séries temporais

for (u in 1:37){      # u são as linhas da matriz de dissimilaridade
for(h  in 1:37) {     # h são as colunas da matriz de dissimilaridade

# o código abaixo calcula a dissimilaridade entre duas séries , já os "for" acima faz isso as vezes necessárias para completar a matriz

S1=series[,u]    # no R, quando não especificamos o numero de linhas ele entende que queremos todas elas, o mesmo acontece com colunas
S2=series[,h]
 d=array ( 0, dim=c(92,92)) # a matriz d é a diferença entre um time point de uma série e os time points da outra serie
for (i in 2:92){
   for ( j in 2:92){
    d[i,j]=S1[i-1]-S2[j-1]
 d[i,j]= d[i,j]^2
 d[i,j]= sqrt(d[i,j])}}
for (j in 1:92){
 d[1,j]=100000}
for(i in 1:92){
```

```
  d[i , 1] = 100000} # o valor 100000 é apenas um controle de borda
a=92  # é a linha de um elemento da matriz d
b=92  # é a coluna de um elemento da matriz d
# (92,92) é o ultimo elemento da matriz d, no caso desse trabalho pois as séries tinham 91 time point e foi acrescentado
uma linha e uma coluna com elementos 100000.
w=numeric()  # w irá receber os valores das células do caminho warping
w=d[a,b]
k=1                    # k é o numero de passos no caminho warping
while ( a!=2 & b!=2)  {     # o “while” faz que o caminho warping que começou no ultimo elemento acabe no primeiro
     {w= w + min( d[a-1,b], d[a-1,b-1], d[a,b-1])}
     { if ( d[a-1,b] == min( d[a-1,b], d[a-1,b-1], d[a,b-1])) {  # os “if” escolhem a próxima célula do caminho warping
                    a=a-1} }

      {if (d[a-1,b-1] == min( d[a-1,b], d[a-1,b-1], d[a,b-1] )) {
                       a= a-1
                       b=b-1}}

      { if (d[a,b-1] == min( d[a-1,b], d[a-1,b-1], d[a,b-1] ) ){
                       b= b-1}
                   k= k+1  }}


DTW=numeric()
DTW= w/k   #isso dá a dissimilaridade entre duas séries apenas


dissimilaridade[u,h]= DTW
}}

write.table(dissimilaridade, file='C:/INPE/Séries Scarpel/d-DTW-.txt') # esse codigo exporta a matriz de dissimilaridade
```

ANEXO A-II MDS COM DTW

```
library(MASS)  # é essa a biblioteca mais confiável para gerar o MDS


setwd("C:/INPE/Séries Scarpel")
dissimilaridade = read.csv( "d-DTW-.csv", header = TRUE, dec=".", sep=";", row.names=1)
# é necessário a matriz de dissimilaridade gerada por DTW, contudo ela tem que ser em .csv e não em txt


# MDS Classical

mds.classical=cmdscale(dissimilaridade, k=2, eig=T)  # k =2, pois o gráfico gerado será em duas dimensões
jpeg(filename = “MDS com d(DTW)-pairs-91 time points.jpg”)    # esse codigo se prepara o R para exporter a figura gerada abaixo
pairs(mds.classical$points)
dev.off()
jpeg(filename = "MDS com d(DTW)-91 time points.jpg")
plot(mds.classical$points, main = "MDS com d(DTW)-91 time points")  # plota as séries temporais em um gráfico bidimensional
dev.off()
jpeg(filename = "MDS com d(DTW)-text-91 time points.jpg")
plot(mds.classical$points, main = "MDS com d(DTW)-91 time points")
text(mds.classical$points, labels = row.names(dissimilaridade), main = "MDS com d(DTW)-91 time points ")
# o comando acima rotula as séries temporais no gráfico bidimensional
dev.off()
pontos = mds.classical$points
write.table(pontos, file='C:/INPE/Séries Scarpel/coord. espaciais das series por MDS com d(DTW)-91 time points.txt’)            # exporta em txt as coordenadas dos pontos
```

ANEXO A-III Método de Ward – baseado na DTW e no MDS

```
setwd("C:/INPE/Séries Scarpel")
coord = read.csv("coord. espaciais das series por MDS com d(DTW)-91 time points.csv", header = TRUE, dec=".", sep=";", row.names=1)
# a variável “coord” deve receber as coordenadas que foram geradas e exortadas pelo MDS, contudo deve estar em .csv separado por virgulas


# CLUSTER

seg=hclust(dist(coord), method=”ward”) # no lugar do “ward” é possível inserir “centroid”, “average” ou outros métodos hierárquicos
jpeg(filename = “Dendograma-ward-d(DTW).jpg”)
plot(seg, main= “Dendograma-ward-d(DTW)-91 Time Points”)  # gera o dendograma
dev.off()
jpeg(filename = “Dendograma-ward-d(DTW)-4clusters.jpg”)
plot(seg)
rect.hclust(seg, k=7, border="red") # k=7 pois apos olhar o dendograma escolheu-se 7 agrupamentos naturais
dev.off()
membros = cutree(seg, k = 7)
write.table(membros, file='C:/INPE/Séries Scarpel/membros-ward-d(DTW).txt’) # exporta as séries que pertencem a cada um dos sete clusters
```

ANEXOS B - Matriz de Dissimilaridade por DTW

Devido à matriz ser 37x37 ela foi divida em três partes para melhor visualização.

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13
S1	0	0,37	0,26	0,92	0,81	2,18	0,50	0,85	0,65	0,86	0,78	0,22	0,20
S2	0,40	0	0,38	0,91	0,32	2,19	0,73	0,72	0,66	0,79	0,79	0,77	0,78
S3	0,26	0,37	0	0,83	0,42	2,15	0,83	1,11	0,92	0,34	0,33	0,77	0,52
S4	0,91	0,92	0,83	0	0,58	2,18	0,78	1,09	0,63	0,54	0,55	0,49	0,51
S5	0,81	0,32	0,39	0,57	0	1,35	0,71	0,56	0,44	1,33	0,95	1,57	1,22
S6	2,17	2,16	2,16	2,15	1,34	0	1,66	1,30	1,05	2,23	2,23	2,23	2,22
S7	0,50	0,72	0,86	0,75	0,72	1,65	0	0,89	0,68	1,10	0,86	0,81	0,71
S8	0,84	0,71	1,12	1,07	0,55	1,34	0,91	0	0,57	0,50	1,28	0,90	0,81
S9	0,65	0,64	0,91	0,65	0,44	1,05	0,70	0,56	0	0,93	2,04	1,35	0,89
S10	0,86	0,80	0,34	0,55	1,34	2,21	1,10	0,50	0,92	0	1,17	1,29	1,83
S11	0,79	0,80	0,34	0,55	0,96	2,23	0,84	1,30	2,04	1,17	0	0,68	0,72
S12	0,20	0,77	0,76	0,49	1,56	2,21	0,82	0,90	1,34	1,30	0,68	0	0,08
S13	0,19	0,79	0,50	0,51	1,20	2,22	0,72	0,81	0,89	1,87	0,72	0,08	0
S14	0,28	0,79	0,47	0,97	1,56	2,23	0,65	0,90	1,46	1,58	1,37	0,26	0,29
S15	0,65	0,68	0,79	0,68	0,75	0,83	0,66	0,70	0,40	0,83	0,64	1,17	0,88
S16	0,61	0,65	0,77	0,62	0,96	0,82	0,59	0,68	0,52	0,66	0,81	0,61	0,85
S17	0,64	0,84	0,70	0,70	1,05	0,70	0,63	0,73	0,89	0,53	0,84	0,79	0,69
S18	0,36	0,82	0,47	0,52	0,56	2,17	0,67	0,71	0,74	0,65	0,64	0,14	0,14
S19	0,29	0,77	0,72	0,91	1,24	0,43	0,72	0,89	1,12	1,30	0,63	0,93	0,93
S20	0,21	0,94	0,51	0,92	0,73	2,17	0,76	0,94	0,85	1,55	0,76	1,55	0,14
S21	0,36	0,77	0,37	0,93	0,52	2,16	0,67	0,50	0,78	0,63	0,58	0,69	0,26
S22	0,97	0,54	1,40	0,69	0,63	0,68	1,15	1,22	0,90	0,55	0,57	0,41	0,86
S23	0,96	1,53	1,51	1,52	0,63	0,60	1,12	1,29	0,94	0,56	0,39	0,35	0,38
S24	0,89	0,75	0,58	0,55	0,45	0,52	0,53	0,54	0,61	0,50	0,86	0,39	0,40
S25	1,40	0,36	0,73	0,52	0,45	0,60	0,55	0,69	0,43	1,02	0,74	0,92	0,83
S26	0,41	1,15	0,41	1,16	0,77	1,18	1,13	0,76	0,61	0,46	0,82	0,35	0,35
S27	0,69	0,71	0,73	0,73	0,75	0,59	0,61	0,76	0,43	0,57	0,53	0,57	0,55
S28	1,37	1,06	1,39	1,06	0,83	0,99	0,77	0,85	0,66	0,58	0,60	1,43	1,40
S29	0,31	0,64	0,87	0,87	0,56	1,51	0,51	0,57	0,64	0,69	0,73	0,69	0,70
S30	0,58	0,75	0,66	0,58	0,59	0,66	0,81	0,68	0,63	0,67	0,65	0,58	0,55
S31	1,24	1,21	1,43	0,42	0,91	0,67	1,17	1,33	0,98	1,49	0,86	1,27	1,29
S32	1,44	0,45	0,52	0,99	1,28	0,40	1,04	1,27	1,30	1,37	1,28	0,72	0,73
S33	0,96	0,94	1,30	0,47	1,28	0,90	1,27	1,20	1,23	1,29	1,28	0,95	0,98
S34	1,97	1,01	0,91	0,95	1,28	2,23	0,55	0,81	1,11	0,90	2,13	1,51	1,89
S35	0,69	1,03	0,67	1,69	1,09	0,61	0,56	0,78	1,06	1,04	1,59	1,01	1,17
S36	1,82	1,12	1,36	0,54	0,66	2,16	0,62	0,40	0,52	0,88	0,56	1,57	1,79
S37	0,83	0,82	0,37	0,91	1,32	2,19	1,01	0,42	1,05	0,61	1,08	1,47	1,34

Tabela contendo matriz de dissimilaridade (parte 1)

	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	S26
S1	0,30	0,66	0,62	0,65	0,37	0,26	0,21	0,38	0,96	0,96	0,87	1,41	0,40
S2	0,77	0,70	0,66	0,84	0,82	0,76	0,94	0,78	0,54	1,56	0,74	0,35	1,18
S3	0,46	0,82	0,78	0,70	0,49	0,74	0,52	0,39	1,40	1,52	0,57	0,74	0,42
S4	0,96	0,68	0,63	0,71	0,53	0,91	0,91	0,89	0,72	1,54	0,56	0,55	1,15
S5	1,56	0,76	0,96	1,06	0,57	1,25	0,74	0,51	0,64	0,65	0,46	0,46	0,78
S6	2,24	0,83	0,82	0,72	2,17	0,45	2,17	2,15	0,67	0,61	0,52	0,58	1,21
S7	0,64	0,66	0,60	0,64	0,66	0,71	0,74	0,68	1,17	1,13	0,50	0,53	1,12
S8	0,89	0,72	0,70	0,75	0,71	0,88	0,95	0,48	1,26	1,33	0,55	0,67	0,78
S9	1,45	0,42	0,54	0,90	0,78	1,12	0,85	0,78	0,92	0,97	0,63	0,45	0,62
S10	1,58	0,83	0,68	0,54	0,66	1,29	1,52	0,64	0,55	0,56	0,48	1,02	0,43
S11	1,37	0,66	0,83	0,84	0,63	0,63	0,76	0,59	0,57	0,39	0,87	0,73	0,82
S12	0,27	1,17	0,61	0,81	0,14	0,91	1,53	0,66	0,39	0,36	0,38	0,91	0,34
S13	0,29	0,87	0,85	0,71	0,15	0,91	0,15	0,28	0,86	0,39	0,39	0,83	0,35
S14	0	1,11	0,73	0,83	0,17	0,77	0,20	0,23	0,50	0,96	0,40	1,48	0,35
S15	1,11	0	0,27	0,34	0,96	0,97	0,98	0,46	1,00	1,05	0,94	0,76	0,55
S16	0,73	0,27	0	0,46	0,58	0,61	0,64	1,01	0,56	0,56	0,76	0,80	0,51
S17	0,81	0,34	0,45	0	0,79	0,74	0,79	0,58	0,64	0,76	0,73	0,95	0,46
S18	0,16	0,95	0,58	0,81	0	0,86	0,14	0,30	1,46	0,92	0,45	1,33	0,39
S19	0,77	0,97	0,61	0,75	0,89	0	0,36	0,47	0,90	0,97	0,53	1,39	0,40
S20	0,21	0,98	0,64	0,81	0,15	0,37	0	0,40	0,90	0,90	0,60	1,36	0,37
S21	0,24	0,46	1,01	0,60	0,32	0,50	0,41	0	0,53	1,51	0,82	0,75	0,40
S22	0,51	1,01	0,56	0,65	1,48	0,93	0,90	0,52	0	0,08	0,46	0,52	1,03
S23	0,97	1,05	0,56	0,78	0,92	1,02	0,89	1,50	0,08	0	0,44	0,52	1,05
S24	0,41	0,94	0,76	0,77	0,46	0,54	0,63	0,82	0,46	0,45	0	0,43	0,86
S25	1,47	0,78	0,81	0,94	1,30	1,40	1,38	0,75	0,52	0,52	0,47	0	1,06
S26	0,35	0,55	0,51	0,44	0,39	0,42	0,38	0,42	1,03	1,05	0,81	1,06	0
S27	0,54	0,36	0,39	0,57	0,56	0,46	0,50	0,64	1,05	1,03	0,67	0,83	0,45
S28	0,59	0,45	0,40	0,49	1,37	1,38	0,47	1,28	1,04	1,06	1,01	1,05	0,57
S29	0,70	0,72	0,59	0,68	0,47	0,91	0,66	0,34	1,21	1,25	0,58	0,79	0,74
S30	0,59	0,60	0,53	0,58	0,57	0,57	0,57	0,67	0,72	0,75	0,76	0,73	0,50
S31	1,31	0,75	0,74	0,86	1,22	1,20	0,81	1,22	0,77	0,72	0,48	0,74	1,27
S32	0,72	1,02	1,03	0,81	1,42	1,43	0,33	1,36	1,09	0,39	0,91	0,58	1,35
S33	0,97	1,01	1,00	0,79	0,93	0,89	0,93	1,28	0,60	0,61	0,79	0,91	1,28
S34	1,63	0,72	0,69	0,51	1,40	1,29	1,40	1,68	1,51	1,50	1,53	1,24	0,47
S35	0,85	0,67	0,99	0,64	0,99	0,83	0,66	1,04	0,70	0,63	0,89	0,96	1,24
S36	1,53	0,78	0,79	0,59	0,25	0,97	1,51	0,44	1,42	1,52	1,35	0,77	0,39
S37	1,21	0,96	0,73	0,51	1,13	1,30	1,17	0,68	0,60	0,60	0,72	1,06	0,39

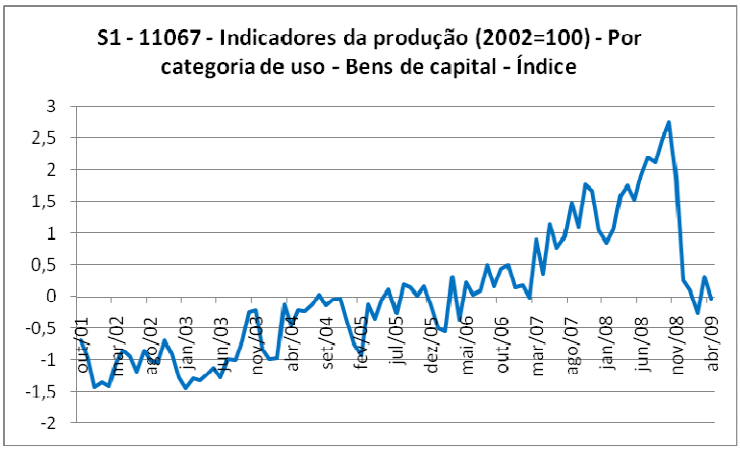
Tabela contendo a matriz de dissimilaridade (parte 2)

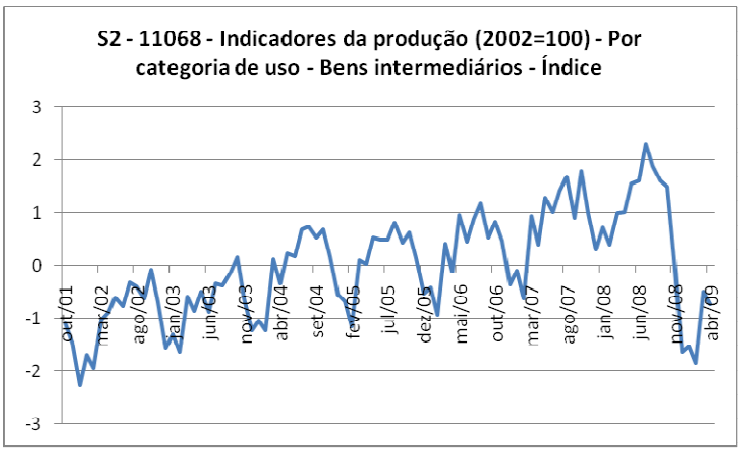
	S27	S28	S29	S30	S31	S32	S33	S34	S35	S36	S37
S1	0,69	1,34	0,32	0,56	1,27	1,46	0,99	1,96	0,67	1,86	0,83
S2	0,74	1,05	0,61	0,77	1,23	0,45	0,97	0,98	1,03	1,12	0,81
S3	0,72	1,40	0,86	0,69	1,45	0,53	1,31	0,90	0,68	1,35	0,36
S4	0,72	1,09	0,88	0,58	0,47	0,97	0,49	0,94	1,70	0,55	0,89
S5	0,73	0,83	0,57	0,60	0,91	1,28	1,29	1,25	1,04	0,67	1,29
S6	0,58	0,97	1,56	0,66	0,69	0,40	0,90	2,27	0,64	2,16	2,21
S7	0,63	0,78	0,50	0,82	1,17	1,01	1,28	0,56	0,55	0,62	1,02
S8	0,78	0,87	0,55	0,71	1,35	1,28	1,23	0,81	0,77	0,40	0,41
S9	0,43	0,67	0,65	0,63	0,99	1,28	1,26	1,14	1,07	0,53	1,04
S10	0,57	0,59	0,68	0,68	1,47	1,38	1,28	0,91	1,04	0,88	0,60
S11	0,52	0,60	0,73	0,65	0,86	1,28	1,29	2,13	1,58	0,55	1,07
S12	0,55	1,43	0,69	0,57	1,28	0,72	0,96	1,50	0,99	1,57	1,47
S13	0,56	1,40	0,70	0,55	1,30	0,74	0,99	1,90	1,17	1,80	1,36
S14	0,55	0,60	0,69	0,58	1,30	0,73	0,96	1,62	0,83	1,52	1,22
S15	0,34	0,44	0,70	0,58	0,74	1,01	1,00	0,70	0,70	0,78	0,94
S16	0,37	0,40	0,59	0,51	0,74	1,02	0,99	0,68	0,98	0,79	0,72
S17	0,57	0,48	0,67	0,56	0,84	0,79	0,77	0,50	0,63	0,57	0,51
S18	0,56	1,40	0,46	0,57	1,27	1,42	0,93	1,40	0,98	0,26	1,14
S19	0,47	1,39	0,91	0,57	1,22	1,43	0,90	1,27	0,81	1,01	1,29
S20	0,51	0,46	0,66	0,57	0,82	0,32	0,95	1,43	0,66	1,52	1,20
S21	0,64	1,27	0,33	0,67	1,23	1,39	1,29	1,69	1,06	0,46	0,69
S22	1,04	1,07	1,16	0,73	0,78	1,09	0,61	1,48	0,70	1,43	0,60
S23	1,03	1,08	1,24	0,75	0,74	0,38	0,63	1,51	0,63	1,53	0,59
S24	0,69	1,00	0,58	0,75	0,49	0,94	0,79	1,56	0,93	1,36	0,73
S25	0,85	1,05	0,82	0,71	0,74	0,58	0,93	1,24	0,95	0,72	1,05
S26	0,47	0,58	0,74	0,52	1,25	1,33	1,26	0,46	1,25	0,40	0,39
S27	0	0,67	0,54	0,69	0,93	0,48	0,91	0,51	0,53	0,60	0,44
S28	0,68	0	0,83	0,42	1,13	0,36	1,11	0,65	1,13	0,52	0,59
S29	0,57	0,85	0	0,71	1,05	0,46	1,28	0,89	0,44	0,39	0,68
S30	0,69	0,43	0,72	0	0,36	0,52	0,37	0,82	0,74	0,64	0,67
S31	0,91	1,12	1,05	0,38	0	0,89	0,19	1,54	0,69	1,43	1,47
S32	0,51	0,39	0,48	0,53	0,89	0	0,64	1,44	0,50	1,42	1,41
S33	0,90	1,10	1,26	0,35	0,18	0,64	0	1,31	0,58	1,29	1,30
S34	0,52	0,66	0,91	0,82	1,54	1,44	1,32	0	1,60	0,46	0,80
S35	0,53	1,15	0,45	0,74	0,67	0,51	0,57	1,62	0	1,07	0,85
S36	0,61	0,51	0,39	0,64	1,43	1,43	1,29	0,45	1,07	0	0,26
S37	0,45	0,59	0,66	0,68	1,45	1,42	1,29	0,80	0,84	0,27	0

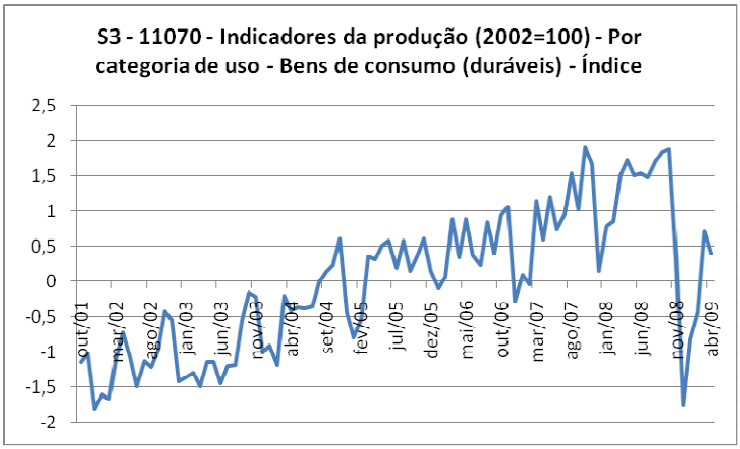
Tabela contendo a matriz de dissimilaridade (parte 3)

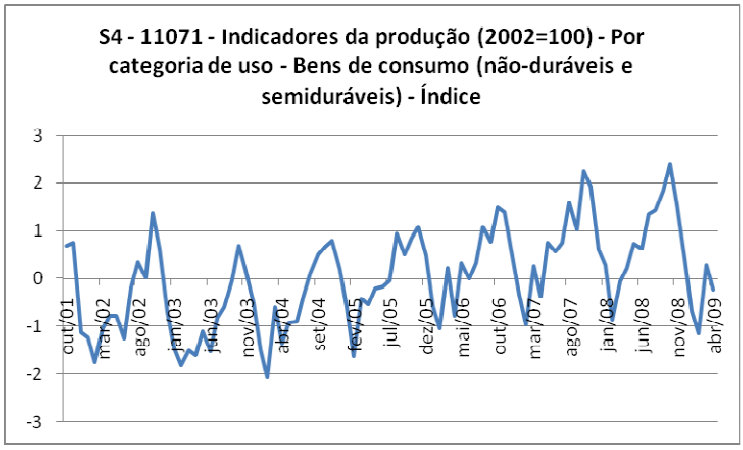
ANEXOS C – Visualização das Séries Temporais

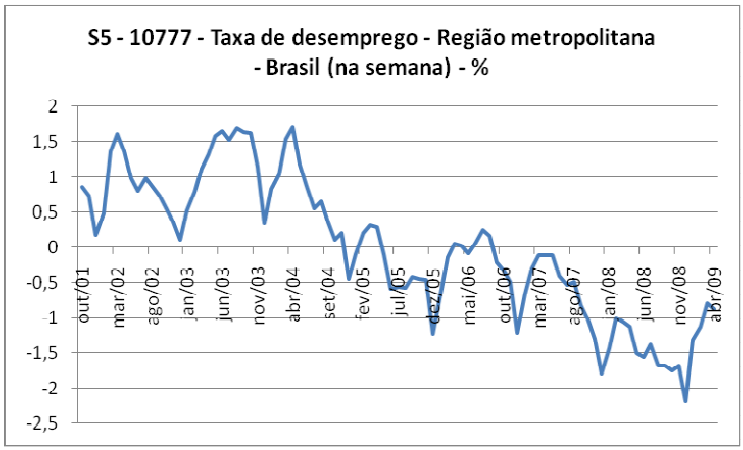
Todas as séries que serão apresentadas estão padronizadas tendo media zero e desvio padrão um.

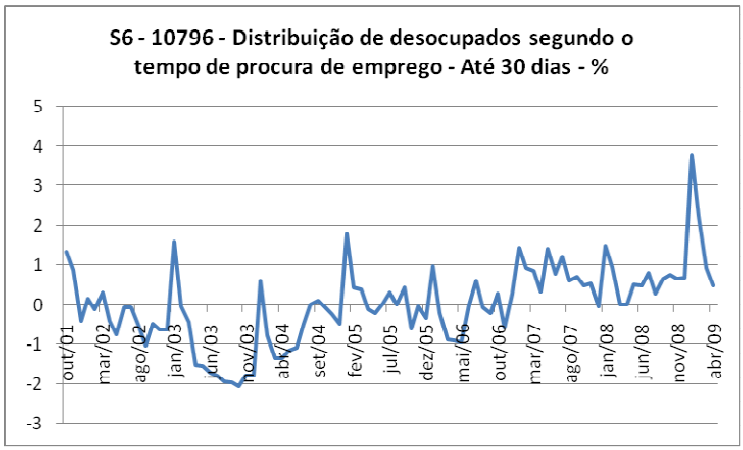


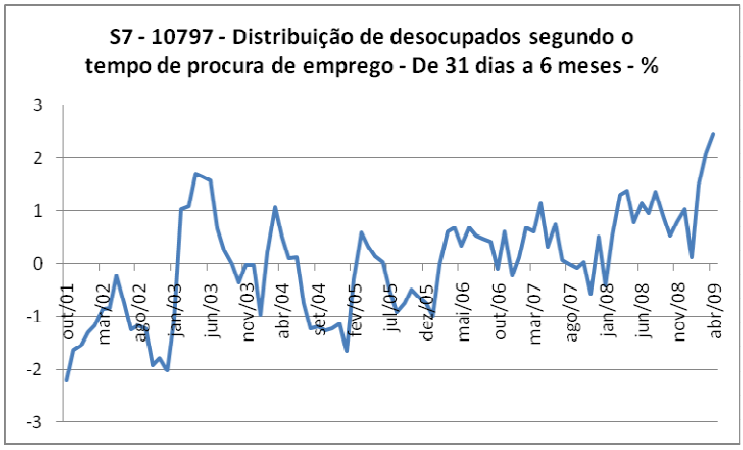


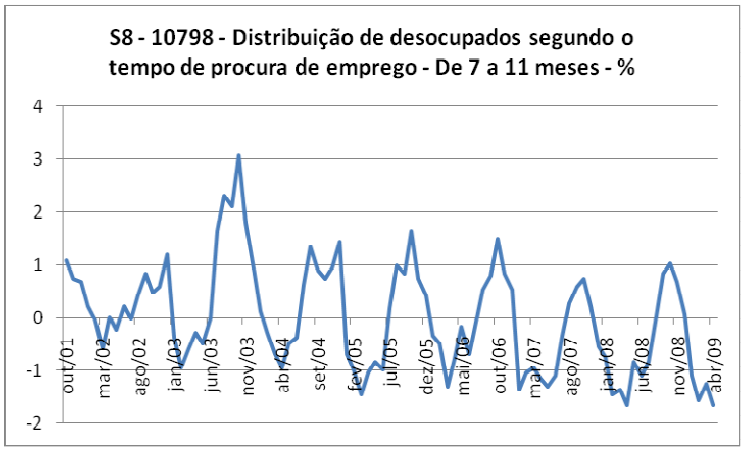

S3 - 11070 - Indicadores da produção (2002=100) - Por categoria de uso - Bens de consumo (duráveis) - Índice
2,5
2
1,5
1
0,5
0
-0,5
-1
-1,5
-2
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09

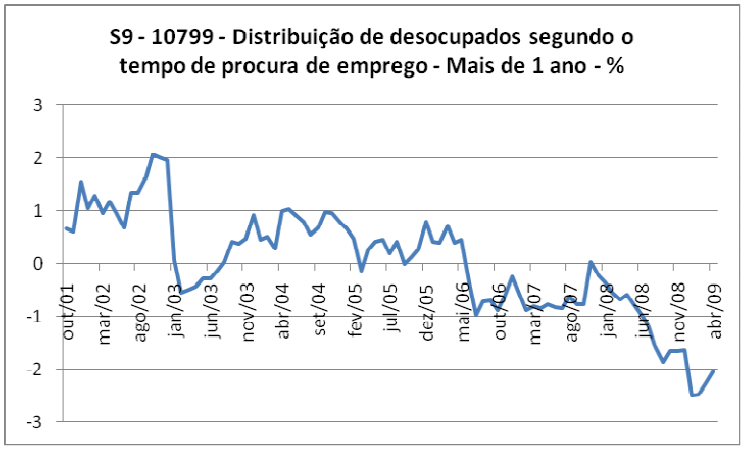

S4 - 11071 - Indicadores da produção (2002=100) - Por categoria de uso - Bens de consumo (não-duráveis e semiduráveis) - Índice
3
2
1
0
-1
-2
-3
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09

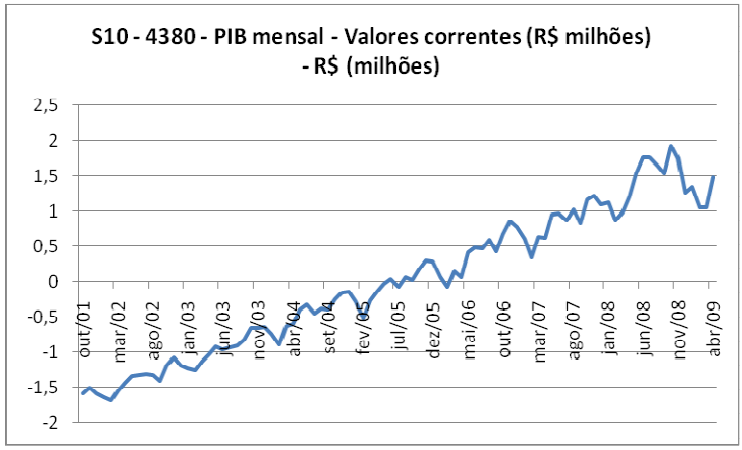

S5 - 10777 - Taxa de desemprego - Região metropolitana - Brasil (na semana) - %
2
1,5
1
0,5
0
-0,5
-1
-1,5
-2
-2,5
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09

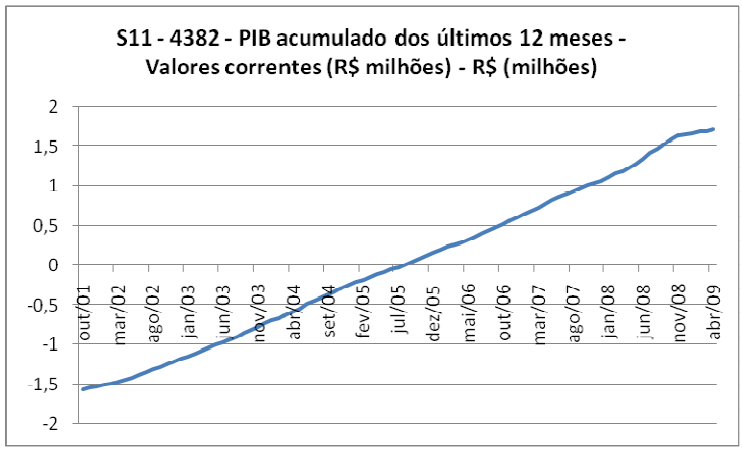

S6 - 10796 - Distribuição de desocupados segundo o tempo de procura de emprego - Até 30 dias - %
5
4
3
2
1
0
-1
-2
-3
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09

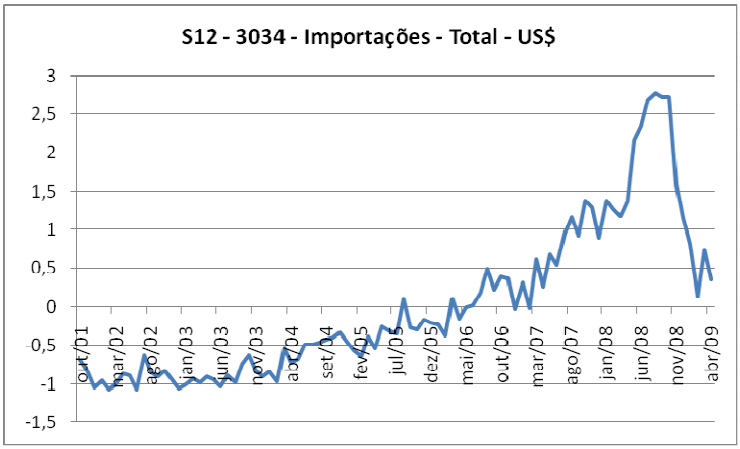

S7 - 10797 - Distribuição de desocupados segundo o tempo de procura de emprego - De 31 dias a 6 meses - %
3
2
1
0
-1
-2
-3
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09

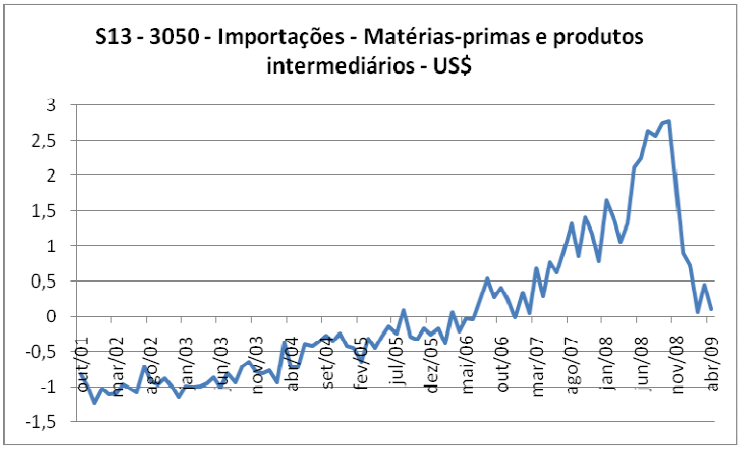

S8 - 10798 - Distribuição de desocupados segundo o tempo de procura de emprego - De 7 a 11 meses - %
4
3
2
1
0
-1
-2
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09

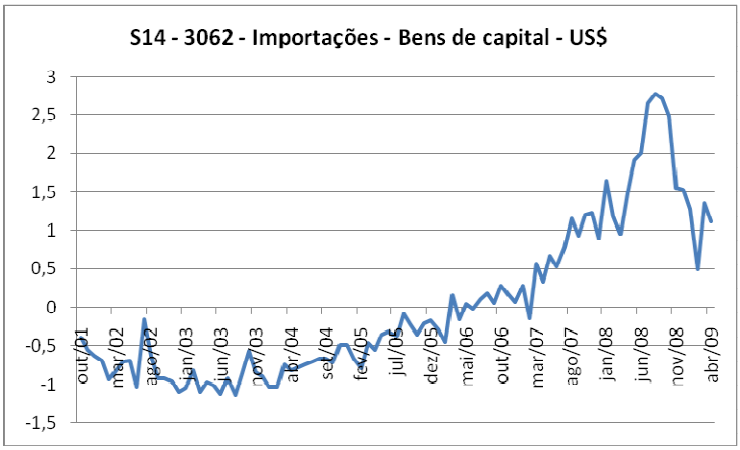

S9 - 10799 - Distribuição de desocupados segundo o tempo de procura de emprego - Mais de 1 ano - %
3
2
1
0
-1
-2
-3
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09

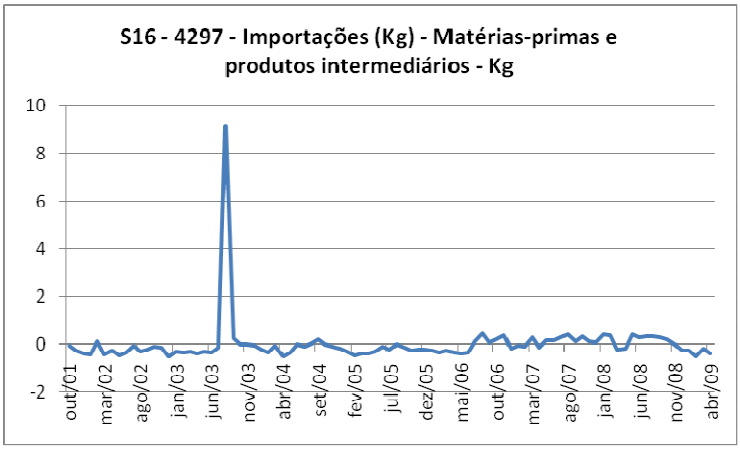

S10 - 4380 - PIB mensal - Valores correntes (R$ milhões) - R$ (milhões)
2,5
2
1,5
1
0,5
0
-0,5
-1
-1,5
-2
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09

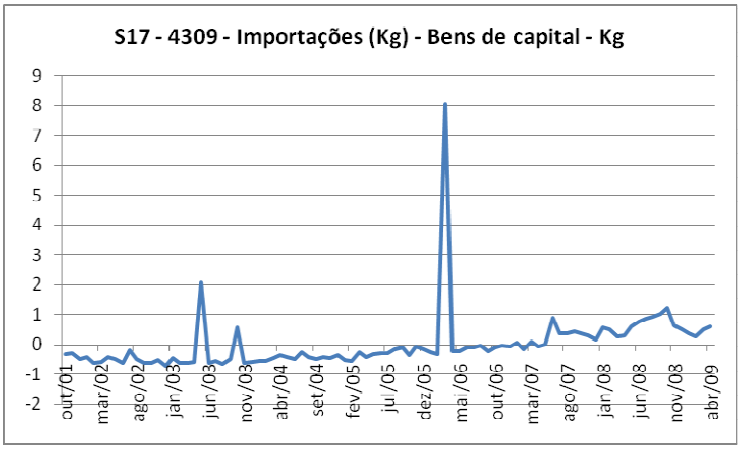

S11 - 4382 - PIB acumulado dos últimos 12 meses - Valores correntes (R$ milhões) - R$ (milhões)
2
1,5
1
0,5
0
-0,5
-1
-1,5
-2
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09

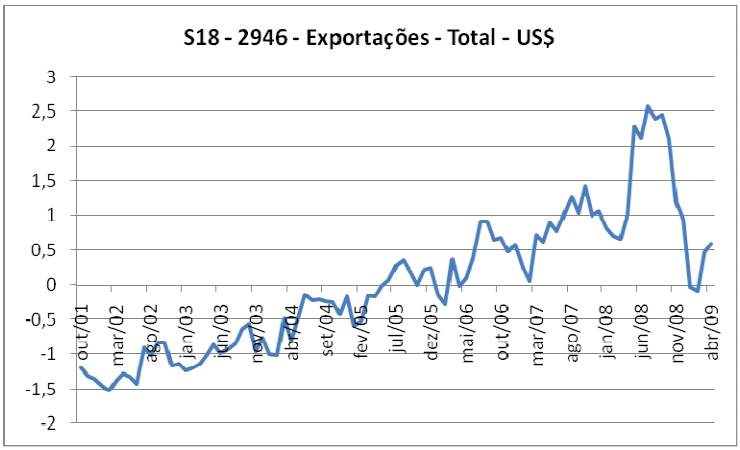

S12 - 3034 - Importações - Total - US$
3
2,5
2
1,5
1
0,5
0
-0,5
-1
-1,5
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09

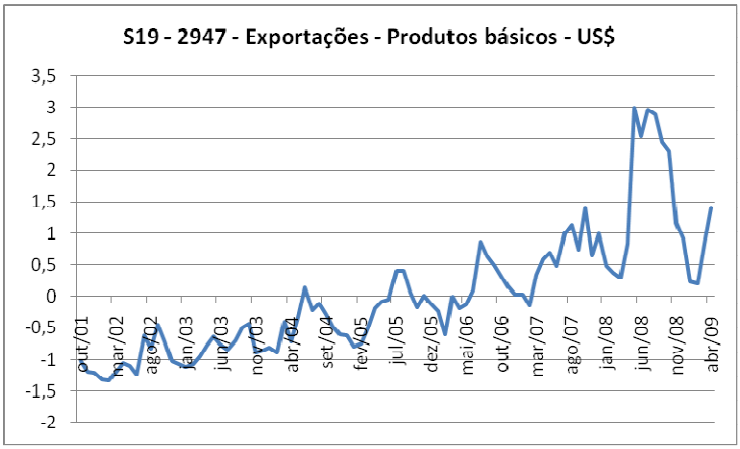

S13 - 3050 - Importações - Matérias-primas e produtos intermediários - US$
3
2,5
2
1,5
1
0,5
0
-0,5
-1
-1,5
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09

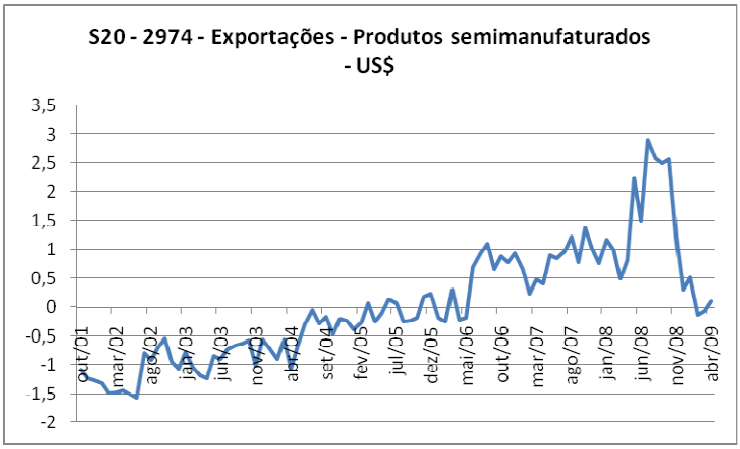

S14 - 3062 - Importações - Bens de capital - US$
3
2,5
2
1,5
1
0,5
0
-0,5
-1
-1,5
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09


S16 - 4297 - Importações (Kg) - Matérias-primas e produtos intermediários - Kg
10
8
6
4
2
0
-2
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09

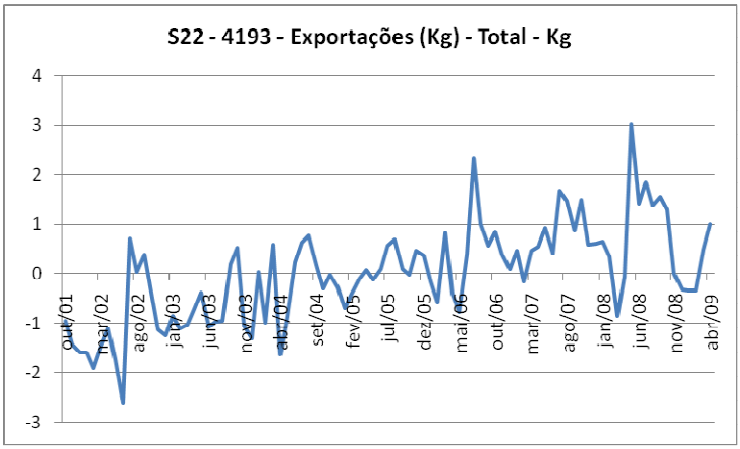

S17 - 4309 - Importações (Kg) - Bens de capital - Kg
9
8
7
6
5
4
3
2
1
0
-1
-2
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09

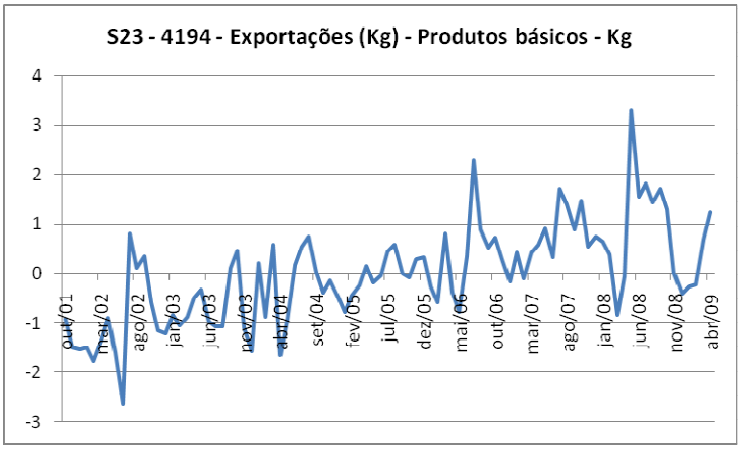

S18 - 2946 - Exportações - Total - US$
3
2,5
2
1,5
1
0,5
0
-0,5
-1
-1,5
-2
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09

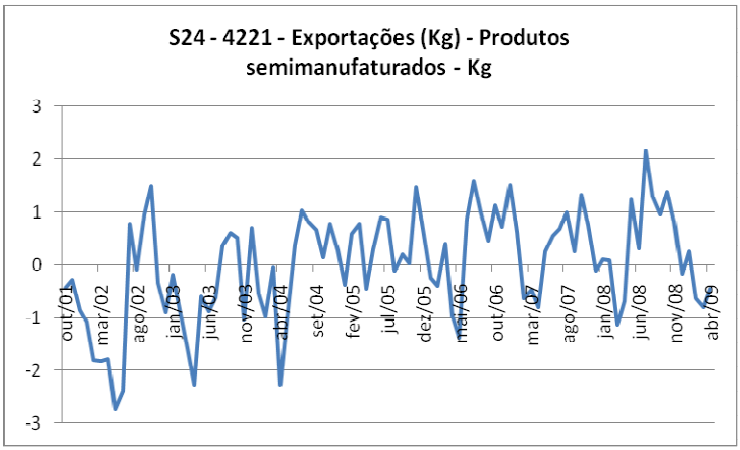

S19 - 2947 - Exportações - Produtos básicos - US$
3,5
3
2,5
2
1,5
1
0,5
0
-0,5
-1
-1,5
-2
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09

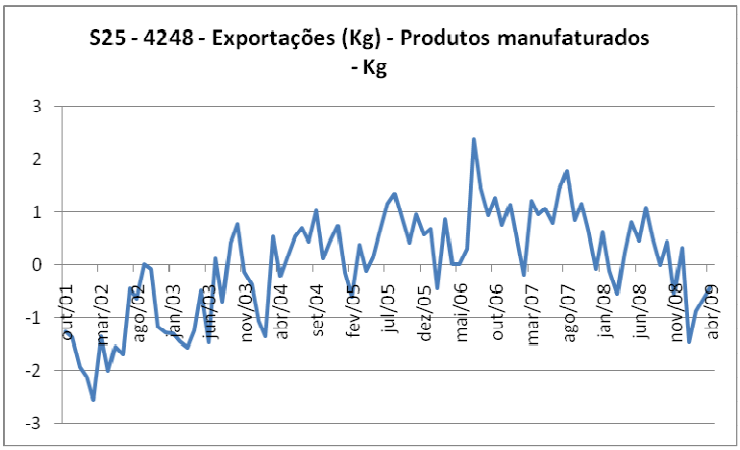

S20 - 2974 - Exportações - Produtos semimanufaturados - US$
3,5
3
2,5
2
1,5
1
0,5
0
-0,5
-1
-1,5
-2
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09

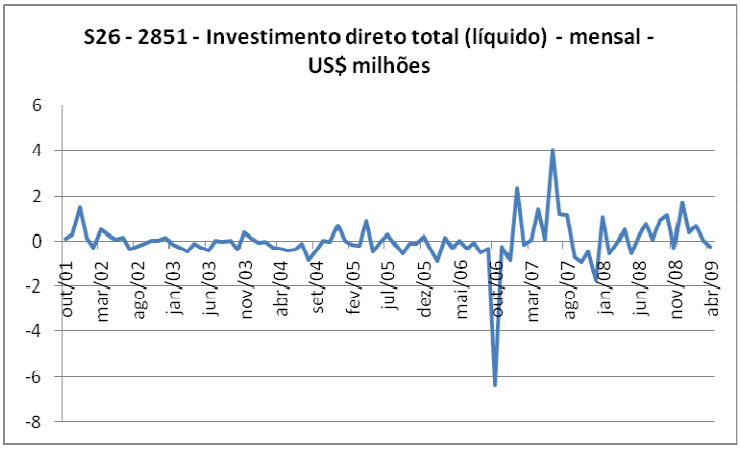

S21 - 3001 - Exportações - Produtos manufaturados - US$
2,5
2
1,5
1
0,5
0
-0,5
-1
-1,5
-2
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09

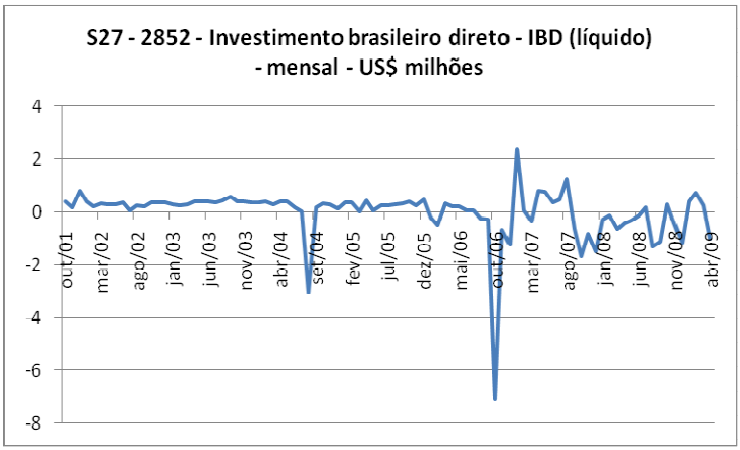

S22 - 4193 - Exportações (Kg) - Total - Kg
4
3
2
1
0
-1
-2
-3
out/01
mai/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09


S23 - 4194 - Exportações (Kg) - Produtos básicos - Kg
4
3
2
1
0
-1
-2
-3
out/01
mai/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09

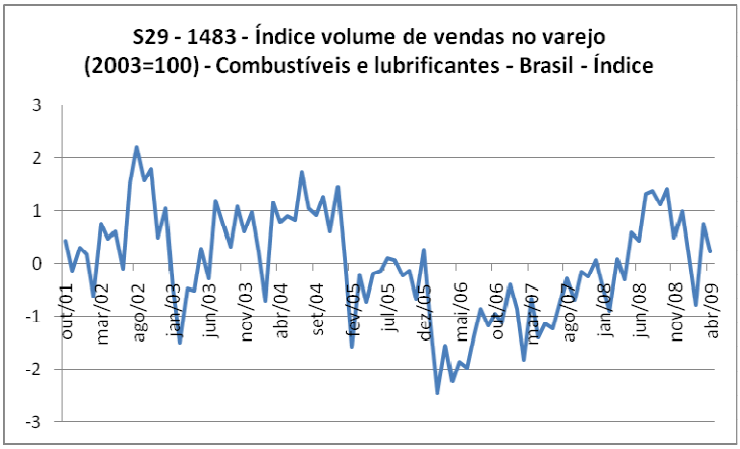

S24 - 4221 - Exportações (Kg) - Produtos semimanufaturados - Kg
3
2
1
0
-1
-2
-3
out/01
mai/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09

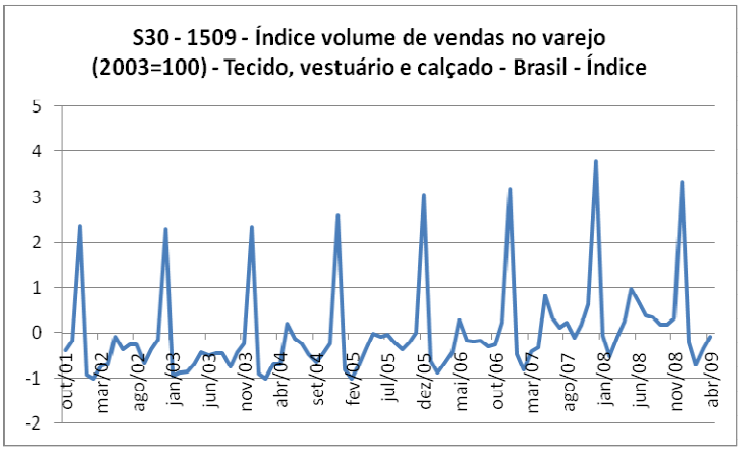

S25 - 4248 - Exportações (Kg) - Produtos manufaturados - Kg
3
2
1
0
-1
-2
-3
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09

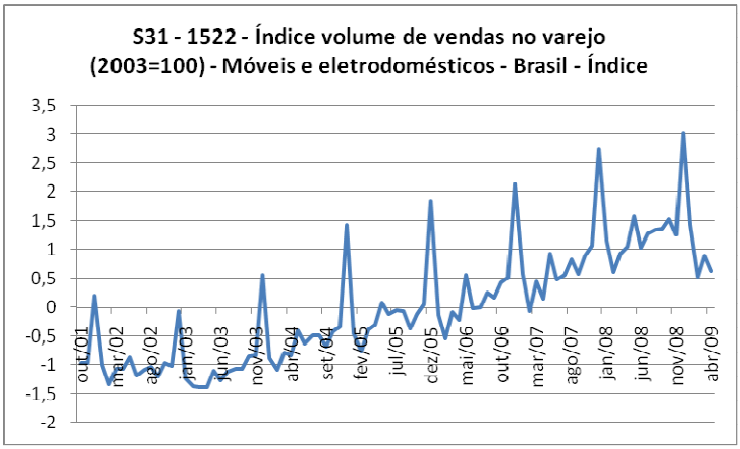

S26 - 2851 - Investimento direto total (líquido) - mensal - US$ milhões
6
4
2
0
-2
-4
-6
-8
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09

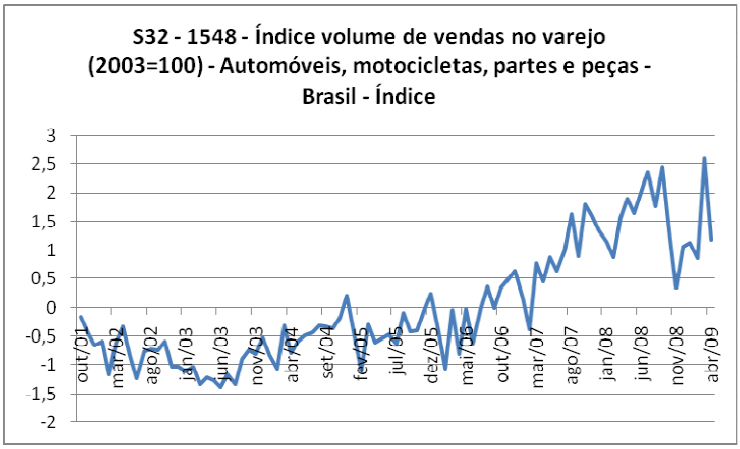

S27 - 2852 - Investimento brasileiro direto - IBD (líquido) - mensal - US$ milhões
4
2
0
-2
-4
-6
-8
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09

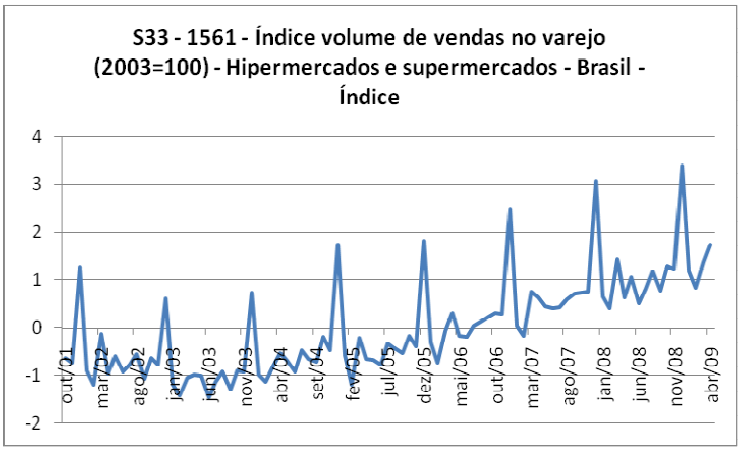

S28 - 2860 - Investimento estrangeiro direto - IED (líquido) - mensal - US$ milhões
6
5
4
3
2
1
0
-1
-2
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09

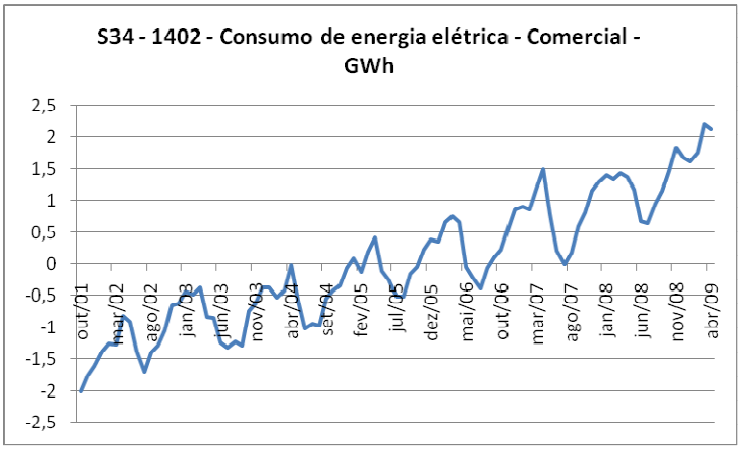

S29 - 1483 - Índice volume de vendas no varejo (2003=100) - Combustíveis e lubrificantes - Brasil - Índice
3
2
1
0
-1
-2
-3
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09

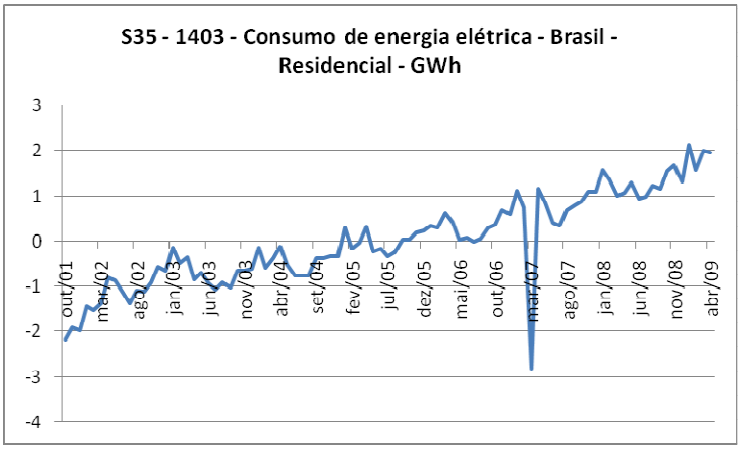

S30 - 1509 - Índice volume de vendas no varejo (2003=100) - Tecido, vestuário e calçado - Brasil - Índice
5
4
3
2
1
0
-1
-2
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09

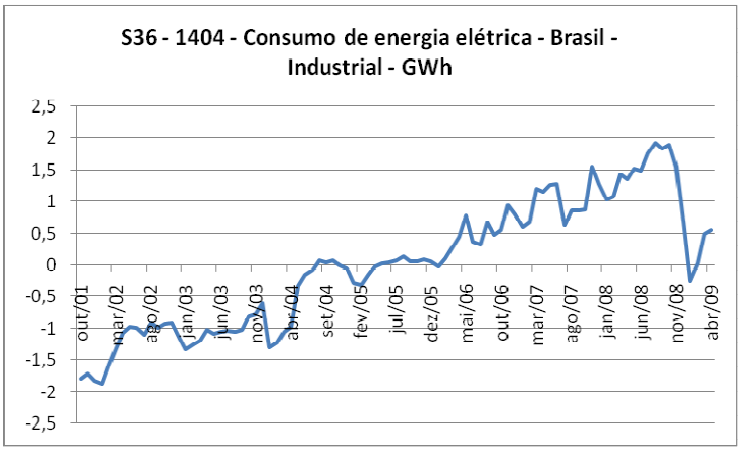

S31 - 1522 - Índice volume de vendas no varejo (2003=100) - Móveis e eletrodomésticos - Brasil - Índice
3,5
3
2,5
2
1,5
1
0,5
0
-0,5
-1
-1,5
-2
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09

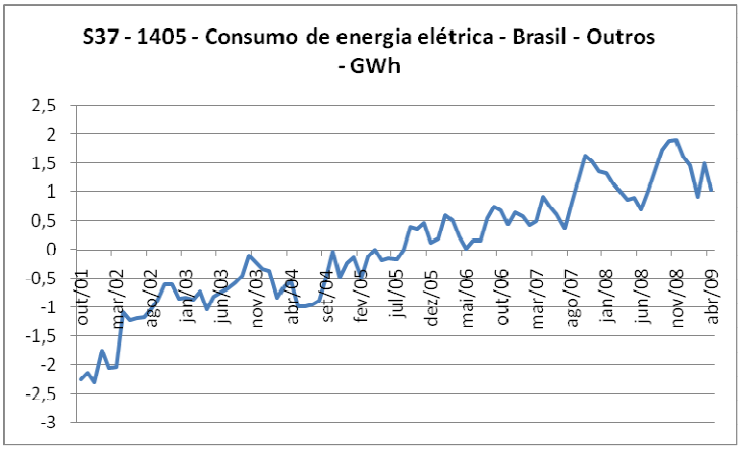

S32 - 1548 - Índice volume de vendas no varejo (2003=100) - Automóveis, motocicletas, partes e peças - Brasil - Índice
3
2,5
2
1,5
1
0,5
0
-0,5
-1
-1,5
-2
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09


S33 - 1561 - Índice volume de vendas no varejo (2003=100) - Hipermercados e supermercados - Brasil - Índice
4
3
2
1
0
-1
-2
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09


S34 - 1402 - Consumo de energia elétrica - Comercial - GWh
2,5
2
1,5
1
0,5
0
-0,5
-1
-1,5
-2
-2,5
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09


S35 - 1403 - Consumo de energia elétrica - Brasil - Residencial - GWh
3
2
1
0
-1
-2
-3
-4
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09


S36 - 1404 - Consumo de energia elétrica - Brasil - Industrial - GWh
2,5
2
1,5
1
0,5
0
-0,5
-1
-1,5
-2
-2,5
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09


S37 - 1405 - Consumo de energia elétrica - Brasil - Outros - GWh
2,5
2
1,5
1
0,5
0
0,5
-1
-1,5
-2
-2,5
-3
out/01
mar/02
ago/02
jan/03
jun/03
nov/03
abr/04
set/04
fev/05
jul/05
dez/05
mai/06
out/06
mar/07
ago/07
jan/08
jun/08
nov/08
abr/09