

# ÍNDICE DE ATIVIDADE ECONÔMICA: OS MODELOS DE FILTRO DE KALMAN E BOX-JENKINS COMPARADOS

Vamerson Schwingel Ribeiro  
Instituto Superior de Ciências Aplicadas – ISCA  
Joilson Dias  
Universidade Estadual de Maringá

## Resumo

O presente trabalho tem como objetivo principal a elaboração de um índice composto ponderado, criado como forma de medir o nível de atividade econômica local. Na sua elaboração utiliza-se a técnica de componentes principais para a determinação dos pesos das variáveis que compõem o índice. A sua qualidade é testada com relação a indicadores nacionais. Como resultado, tem-se que o mesmo comporta-se como um índice coincidente das atividades econômicas locais; portando, sua previsão comporta-se como um indicador antecedente destas mesmas atividades. Na previsão são testados dois modelos: Filtro de Kalman e Box-Jenkins. Devido a presença de outliers, o Filtro de Kalman exigiu que usássemos técnicas avançadas para obter estabilização dos coeficientes. A comparação entre as duas projeções é efetuada através do teste estatístico de Diebold e Mariano (1995). Como resultado final, obtivemos que os modelos apresentam-se como idênticos em suas previsões.

Palavras-chave: índice de atividade econômica, índice antecedente, filtro de kalman, ARIMA, box-jenkins.

## Abstract

This paper has as objective to build a composite economic activity index for the local economy, created as form of measuring the economic activity. We use the factor analysis technique to determinate the components and their weights. This local index is then compared to national ones. As a result, the index behaves nicely as a local coincident index of the economic activities. Two techniques are used in their forecast; the first was the Kalman Filter and the second one the Box-Jenkins model. The presence of outliers required that we use a new technique in order the coefficients of the Kalman Filter model to be stable. The two techniques are then compared using a statistic test developed by Diebold and Mariano (1995). As a final result, the two models' forecasting are the same.

Key words: economic activity index, leading index, kalman filter, ARIMA, box-jenkins.

# Índice de Atividade Econômica: os modelos de Filtro de Kalman e Box-Jenkins comparados

## 1. Introdução

Atualmente, os indicadores de atividade econômica estão concentrados, em sua maioria, em nível agregado. Estas estatísticas, apesar de importantes, possuem uma certa defasagem entre a coleta dos dados e a produção do indicador. Esta defasagem é sem dúvida um dos maiores inconvenientes sobre a utilização de indicadores em nível agregado para aferir o comportamento da economia local. Além deste fator, soma-se o caso em que a economia local pode não apresentar as mesmas características da nacional ou de uma outra na qual exista indicador (estado ou região). Desta forma, a falta de um índice composto que quantifique a atividade econômica local com maior rapidez, impede que políticas e/ou decisões econômicas locais sejam implementadas e/ou sejam mais eficientes.

A existência de um índice em nível local permite que as seguintes perguntas sejam respondidas de forma satisfatória: i) qual foi o comportamento da atividade econômica do município durante o último ano?; ii) a atividade econômica do município pode ser considerada expansionista, estável ou contracionista?; iii) que tipo de flutuações o município apresenta?

Além de responder às questões acima, este artigo também trata de dois problemas técnicos relacionados à previsão de indicadores de atividade econômica. Seguindo a técnica proposta por Lin e Guttman (1993), este artigo apresenta uma solução para o problema de previsão com a presença de valores fora do esperado, os chamados *outliers*. O segundo elemento é o teste que compara dois métodos diferentes de projeção baseado em Diebold e Mariano (1995), que nos permite avaliar com maior certeza as previsões dos dois modelos.

A escolha do município de Maringá para se desenvolver o índice de atividade econômica deveu-se, basicamente, por dois motivos: i) disponibilidade de dados históricos; ii) existência de apoio logístico na coleta e atualização dos dados. O município está localizado na região noroeste do Paraná, tendo aproximadamente 289.000 habitantes (97,5% de população urbana), sendo que, sua região de influência é de 450.000 habitantes composto de sete municípios. Sua economia, segundo os ramos de atividades, divide-se em 7,55% de indústrias, 2,61% de comércio atacadista, 43,43% de comércio varejista e 46,41% de prestadores de serviços.<sup>1</sup>

Neste trabalho, o Índice de Atividade Econômica de Maringá (IAEMga), compreende de janeiro de 1996 a junho de 2002. Posteriormente, as séries serão atualizadas, para que seja possível calcular mensalmente o índice.

## 2. Fonte e análise dos dados

A elaboração de índices de atividade econômica no Brasil não é algo comum. O mais conhecido é o Índice de Movimentação Econômica (IMEC), sendo que sua metodologia se encontra em Azzoni e Latif (1995). Este índice tem como principal característica o fato que os agentes econômicos ao se movimentarem (deslocarem-se), o faz com a perspectiva voltada para a execução de uma atividade econômica qualquer. Assim, o aumento ou redução neste agregado de movimentação dita o comportamento do índice. Outros indicadores a nível agregado da economia brasileira foram desenvolvidos e analisados por Contador (1977), Chauvet (2000a, 2000b), Cardim de Carvalho e Hermann (2003), Morais e Portugal (2004) e Duarte, Issler e Spacov (2004).

O presente trabalho propõe um índice com características diferentes, concentrando-se no aspecto da demanda. O índice a ser desenvolvido pressupõe, que as mudanças na renda dos agentes econômicos respondem positivamente com as variações na demanda por bens e serviços. Estas mudanças na demanda são capturadas de forma indireta pelo índice. Desta forma, procuramos selecionar variáveis que

---

<sup>1</sup> Secretaria da Fazenda do Município de Maringá - PMM/SEFA, 1997.

estivessem correlacionadas com a atividade de demanda agregada local e que ao mesmo tempo fossem de fácil mensuração.

Utilizando a forma de convênios com diversas entidades detentoras das séries, foi obtido um grupo de 40 variáveis. Estas séries temporais, candidatas a comporem o Índice de Atividade Econômica de Maringá (IAEMga) foram obtidas junto as seguintes fontes: i) Companhia Paranaense de Energia Elétrica (COPEL): total mensal do consumo em MWh de energia elétrica e do número de consumidores por cada categoria (residencial, industrial, rural, comercial, poder público, iluminação pública, serviço público e próprio) do município de Maringá; ii) Companhia de Saneamento do Paraná (SANEPAR): total mensal do consumo de água em volume micromedido, consumidos pela população do município de Maringá; iii) Associação Comercial e Industrial de Maringá (ACIM): quantidade mensal de consultas efetuadas no SCPC (serviço central de proteção ao crédito) e vídeo-cheque do município de Maringá; iv) Prefeitura do Município de Maringá - Secretaria Municipal do Transporte (PMM-SETRAN): quantidade mensal de passageiros que embarcam e desembarcam na rodoviária de Maringá; v) Prefeitura do Município de Maringá - Secretaria Municipal do Desenvolvimento Urbano, Planejamento e Habitação (PMM-SEDUH): quantidade mensal de alvará (obras licenciadas) e habite-se expedidos no município de Maringá; vi) Departamento de Trânsito do Estado do Paraná (DETRAN): número total mensal da quantidade de veículos automotores (frota) e do primeiro emplacamento do município de Maringá; vii) Ministério da Fazenda - Secretaria da Receita Federal do Município de Maringá: total mensal da arrecadação do IPI, IRPF, IRPJ, IRRF, CPMF, COFINS, PIS e CSLL do município de Maringá; viii) Prefeitura do Município de Maringá - Secretaria Municipal da Fazenda (PMM-SEFA): total mensal da arrecadação do ISSQN do município de Maringá; ix) Secretaria de Estado e da Fazenda do Paraná (SEF-PR): total mensal de ICMS arrecadado pelo município de Maringá; x) Banco Central do Brasil: total mensal de depósitos em Maringá, composto de depósitos à vista, depósitos a prazo e demais depósitos.

Deste conjunto de séries, *a priori*, foram analisados seus comportamentos quanto ao objetivo estabelecido de capturarem algum tipo de demanda. Assim, o grupo de energia elétrica reflete demandas potenciais relacionadas aos consumidores residenciais, industriais e comerciais; portanto, variações nas quantidades em geral de eletrodomésticos e/ou sua maior utilização podem refletir melhoria na renda dos consumidores; da mesma forma, as variações no consumo industrial de energia refletem variações na produção industrial; assim como, o consumo comercial têm as suas variações associadas com uma maior demanda. O consumo de água captura os mesmos efeitos do consumo de energia elétrica, mas relacionados aos elementos que demandam água. A qualidade e quantidade de crédito, capturados pelo SCPC e vídeo-cheque, estão diretamente relacionados as atividades econômicas do comércio, que em geral, são *proxies* das variações de demanda efetiva por bens de consumo. O número de embarques e desembarques na rodoviária são *proxies* para o volume de entrada e saída de consumidores no município e visa capturar elementos sazonais de atividades econômicas temporais: férias, shows, encontros religiosos, vestibulares, etc. O número de obras em construção é considerado extremamente importante como medidor de uma cadeia muito ampla de produtos relacionados à demanda para o comércio e indústria de transformação. A quantidade de primeiro emplacamento e suas oscilações são *proxies* para os bens de consumo duráveis e, portanto, a sua inclusão *a priori* visa capturar este tipo de demanda mais específica. Os impostos federais visam capturar o comportamento da demanda efetiva verificada, ou seja, eles refletem o resultado de atividade ocorrida recentemente. A sua seleção *a priori* teve como objetivo dar consistência aos resultados de demanda potencial verificada. Os impostos municipal sobre serviços (ISSQN) e estadual sobre circulação de mercadorias (ICMS) também foram selecionados dentro do mesmo objetivo dos impostos federais. E, por último, os depósitos à vista e a prazo foram considerados como a disponibilidade de renda potencial a ser gasta, sendo que, o seu aumento significa contração de demanda e a sua redução um aumento de demanda efetiva.

Apesar, desta ser uma seleção *a priori*, em nada garantia a inclusão da variável no índice, tendo em vista que a seleção do grupo seria feita através da análise dos componentes principais. Este próximo passo é extremamente importante, pois considera aspectos qualitativos de grande importância para construção de um índice.

Através das séries de valores brutos, foram construídas séries de números índices com base 100 na média de 2001, sendo que as séries em valores monetários foram deflacionadas através do índice de

preços ao consumidor da Fundação Instituto de Pesquisas Econômicas (IPC-FIPE).<sup>2</sup> As séries são apresentadas apenas em números índices, porque algumas delas requerem completo sigilo sobre seus valores nominais. Das 40 variáveis disponíveis, apenas 18 acabaram compondo o IAEMG. As razões para os cortes das variáveis foram os seguintes: i) variáveis que possuem muito atraso no seu fornecimento: os consumidores de energia elétrica e informações do Banco Central sobre o volume de depósitos; ii) variáveis que foram agregadas: consultas de vídeo-cheque e SCPC e, embarque e desembarque de passageiros; iii) variável que foi eliminada por não ser permanente: a CPMF não foi considerada por tratar-se de um imposto provisório no momento. iv) variável que não possui oscilações: a frota de veículos foi desconsiderada, pois apresenta-se como variável de tendência.

As variáveis resultantes destes cortes, se apresentaram dentro de todos os critérios técnicos requeridos de rapidez e qualidade, conforme Quadro 1.

As séries, número de embarque e desembarque na rodoviária, COFINS, ISSQN, IPI, IRRF e ICMS foram normalizadas para diminuir mudanças de níveis, uma vez que, em alguns casos, as séries tornavam-se explosivas. As normalizações foram realizadas com intuito de reduzir grandes amplitudes que ocorriam no IAEMG, devido ao comportamento que as variáveis citadas apresentavam antes da normalização. Uma vez normalizadas, as séries de tempo aproximam-se do verdadeiro caminho, como é o caso da COFINS, em que, devido a uma alteração da alíquota a partir de fevereiro de 1999 de 2% para 3%, a série passa rapidamente para um nível mais elevado, no qual, o comportamento da série sem a normalização não reflete um aumento verdadeiro na COFINS.

As normalizações foram feitas da seguinte forma: i) número de embarque e desembarque na rodoviária: de janeiro de 1996 a maio de 1998, os números índices foram ajustados conforme à média de janeiro de 1996 a dezembro de 1997; ii) a COFINS foi ajustada de janeiro de 1996 a fevereiro de 1999, sendo que seus valores foram atualizados monetariamente; iii) no caso do ISSQN, os valores monetários anteriores a 2000 foram estimados conforme a sua participação de 2000 e 2001;<sup>3</sup> iv) O número índice do IPI foi ajustado à média de janeiro de 1996 a julho de 1999 e de agosto de 1999 a abril de 2000; v) o número índice do IRRF foi ajustado à média de janeiro de 1996 a novembro de 1999 e de dezembro de 1999 a novembro de 2000; vi) no caso do ICMS, o número índice foi ajustado à média de janeiro de 1996 a fevereiro de 1997.

### Quadro 1 - Séries que Compõem o IAEMG

X01	Consumo de energia elétrica em MWh, residencial;
X02	Consumo de energia elétrica em MWh, industrial;
X03	Consumo de energia elétrica em MWh, comercial;
X04	Consumo de água, volume micromedido;
X05	Número de consultas ao SCPC (serviço central de proteção ao crédito) e vídeo-cheque;
X06	Número de embarque e desembarque na rodoviária (normalizado);
X07	Alvará, área em m <sup>2</sup> ;
X08	Habite-se, área em m <sup>2</sup> ;
X09	Quantidade de primeiro emplacamento de veículos automotores;
X10	Arrecadação de IPI (imposto sobre produtos industrializados, normalizado);
X11	Arrecadação de IRPF (imposto de renda pessoa física);
X12	Arrecadação de IRPJ (imposto de renda pessoa jurídica);
X13	Arrecadação de IRRF (imposto de renda retido na fonte, normalizado);
X14	Arrecadação de COFINS (contribuição para financiamento da seguridade social, normalizado);
X15	Arrecadação de PIS (programa de integração social);
X16	Arrecadação de CSLL (contribuição social sobre o lucro líquido);
X17	Arrecadação de ISSQN (imposto sobre serviço de qualquer natureza, normalizado);
X18	Arrecadação de ICMS (imposto sobre circulação de mercadorias e serviços, normalizado).

<sup>2</sup> IPC geral, índice (jun. 1994 = 100): região metropolitana de São Paulo (RMSP), mensal. Foi escolhido o IPC porque, o período de coleta vai desde o primeiro dia de cada mês até o último dia do mesmo e a publicação acontece perto do dia 10 do mês seguinte.

<sup>3</sup> A partir de 2000 a Secretaria Municipal da Fazenda (SEFA), passou a disponibilizar as informações desagregadas. Antes deste período, era o somatório de IPTU, ISSQN e taxas. A participação do ISSQN para 2000 e 2001 foi de 35,244%.

### 3. Metodologia e resultados dos dados

A seguir, as séries foram ajustadas para a sazonalidade. A eliminação da sazonalidade ocorreu através da técnica X-12-ARIMA, tipo multiplicativo do *US Census Bureau* (2000). Esta técnica consiste em ajustar as séries para os efeitos sazonais através de médias móveis, com a vantagem de estimar os valores que seriam perdidos no início e no fim das séries ajustadas sazonalmente. A extração da sazonalidade é baseada no modelo multiplicativo geral do tipo *Seasonal Autoregressive Integrated Moving Average* (SARIMA), em que, para uma série de tempo  $z_t$  pode ser escrito como:<sup>4</sup>

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D z_t = \theta(B)\Theta(B^s)a_t. \quad (1)$$

onde  $B$  é o operador de defasagem ( $Bz_t = z_{t-1}$ );  $s$  é o período sazonal;  $\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$  é o operador auto-regressivo não-sazonal;  $\Phi(B^s) = (1 - \Phi_1 B^s - \dots - \Phi_p B^{ps})$  é o operador auto-regressivo sazonal;  $\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$  é o operador média móvel não-sazonal;  $\Theta(B^s) = (1 - \Theta_1 B^s - \dots - \Theta_\rho B^{\rho s})$  é o operador média móvel sazonal; e o  $a_t$ s são i.i.d. com média zero e variância constante ( $\sigma^2$ ), denominado erro ruído branco. O  $(1-B)^d(1-B^s)^D$  implica a ordem da diferenciação não-sazonal  $d$  e a ordem da diferenciação sazonal  $D$ . Se  $d = D = 0$  (sem diferenciação), é comum recolocar  $z_t$  em (1) através dos desvios de suas médias, que é, por  $z_t - \mu$ , onde  $\mu = E[z_t]$ .

É comum suavizar os dados antes da construção do índice, pela razão que o ciclo verdadeiro pode ser obscurecido pela volatilidade de curto prazo (Salazar, 1996). Além do que, a eliminação das sazonalidades torna-se necessária para suavizar as séries antes de serem incluídas no índice composto (Brunet, 2000). O resultado da aplicação desta técnica se encontra nos Apêndices, Tabela A1.

Como o índice proposto é um composto ponderado, o próximo procedimento é determinar os pesos para cada variável. Segundo Nilsson (2000), os pesos ótimos podem ser obtidos através da técnica de componentes principais. Esta técnica consiste em criar novas variáveis que são combinações lineares ortogonais das variáveis originais, ou seja, não são correlacionadas entre si (Sharma, 1996).

De acordo com a técnica de componentes principais, os pesos das variáveis são determinados através das suas variâncias. Mais especificamente, a mesma consiste em obter combinações lineares de um conjunto de variáveis, de modo que as primeiras, componentes principais, possam explicar grande parte da variância total do conjunto. Assumindo que temos  $\rho$  variáveis, nós estamos interessados em formar as seguintes  $\rho$  combinações lineares:<sup>5</sup>

$$\begin{aligned} \xi_1 &= w_{11}x_1 + w_{12}x_2 + \dots + w_{1\rho}x_\rho \\ \xi_2 &= w_{21}x_1 + w_{22}x_2 + \dots + w_{2\rho}x_\rho \\ &\vdots \\ \xi_\rho &= w_{\rho 1}x_1 + w_{\rho 2}x_2 + \dots + w_{\rho\rho}x_\rho \end{aligned} \quad (2)$$

onde  $\xi_1, \xi_2, \dots, \xi_\rho$  são os  $\rho$  componentes principais e  $w_{ij}$  são os pesos da  $j$ -ésima variável para a  $i$ -ésima componente principal. Os pesos,  $w_{ij}$ , são estimados de modo que:

i) O primeiro componente principal,  $\xi_1$ , estima a variância máxima nos dados; o segundo componente principal,  $\xi_2$ , estima a variância máxima que não foi computada pelo primeiro componente, e assim em diante;

$$\text{ii) } w_{i1}^2 + w_{i2}^2 + \dots + w_{i\rho}^2 = 1 \quad i = 1, \dots, \rho;$$

$$\text{iii) } w_{i1}w_{j1} + w_{i2}w_{j2} + \dots + w_{i\rho}w_{j\rho} = 0 \quad \text{para todo } i \neq j.$$

<sup>4</sup> U.S. Census Bureau (2000, p.15-16).

<sup>5</sup> Sharma (1996, p.66-7).

Para facilitar a compreensão da obtenção dos componentes principais, apresentamos a seguir um modelo de regressão dos componentes principais, baseado em Judge *et alli* (1985). Para tanto, considere o seguinte modelo:

$$y = X\beta + e. \quad (3)$$

onde  $y$  é uma variável de dimensão  $(T \times 1)$ ,  $X$  é não estocástico de dimensão  $(T \times K)$ ,  $\beta$  é  $(K \times 1)$ , e  $e$  é o erro do ajustamento de dimensão  $(T \times 1)$ , sendo sua distribuição normal  $N(0, \sigma^2 I)$ . Agora considere a seguinte transformação:

$$y = XPP' \beta + e = XP\theta + e = Z\theta + e. \quad (4)$$

onde  $P = (p_1, \dots, p_k)$  é uma  $(K \times K)$  matriz cujas colunas  $(p_i)$  são vetores característicos ortogonais de  $X'X$  ordenado para corresponder às magnitudes relativas das raízes características da matriz definida positiva  $X'X$ ;  $Z = (z_1, \dots, z_k)$  é uma matriz de componentes principais de dimensão  $(T \times K)$ . Assim,  $z_i = Xp_i$  representa o  $i$ -ésimo componente principal, sendo que  $z_i'z_i = \lambda_i$  e  $\lambda_i$  é a  $i$ -ésima maior raiz característica de  $X'X$ .

O estimador dos componentes principais de  $\beta$  é obtido eliminando uma ou mais das variáveis  $z_i$ . A seguir aplicam-se os mínimos quadrados ordinários para o modelo resultante. Mais especificamente, considere  $Z$  em duas partes  $Z_1$ , o  $z_i$  que será armazenado, e  $Z_2$ , o  $z_i$  que será apagado. Esta separação impõe uma separação idêntica em  $P$ . Assim, a equação (4) torna-se a seguinte:

$$y = XP_1\theta_1 + XP_2\theta_2 + e = Z_1\theta_1 + Z_2\theta_2 + e. \quad (5)$$

onde  $X\{P_1 : P_2\} = \{Z_1 : Z_2\}$ .

As propriedades de  $\hat{\theta}_1 = (Z_1'Z_1)^{-1}Z_1'y$ , o estimador de mínimos quadrados de  $\theta_1$  com  $Z_2$  omitido da equação (5), é facilmente obtidos. Especificamente,  $\hat{\theta}_1$  é imparcial, devido a ortogonalidade de  $Z_1$  e  $Z_2$ , e tem matriz de covariância  $\sigma^2(Z_1'Z_1)^{-1}$ . O estimador dos componentes principais é obtido por uma transformação linear inversa. Desde que  $\beta = P\theta = P_1\theta_1 + P_2\theta_2$ , omitindo os componentes em  $Z_2$  significa que  $\theta_2$  foi implicitamente ajustado igual a zero. Consequentemente  $P_2\theta_2 = 0$  e o estimador dos componentes principais de  $\beta$  transforma-se no seguinte:

$$\hat{\beta}^* = P_1\hat{\theta}_1 = P\hat{\theta}^*. \quad (6)$$

onde  $\hat{\theta}^* = (\hat{\theta}_1', 0')$ , sendo que  $0$  é um vetor nulo de dimensão adaptável. As propriedades de  $\hat{\beta}^*$  seguem diretamente de sua equivalência para o estimador de mínimos quadrados restritos obtido através da estimativa da equação (3), sujeito a restrição de que  $P_2'\beta = 0$ . Assim, o estimador dos componentes principais possui as propriedades dos estimadores de mínimos quadrados restritos e é conhecido por ter menor variância da amostra do que o estimador de mínimos quadrados amplo, desde que a restrição  $P_2'\beta = 0$  seja verdade.

O resultado da aplicação da metodologia para os dados (Tabela A1) se encontra na Tabela A2. A interpretação desta tabela é a seguinte: i) as colunas com os valores denominados de Comp 1 referem-se ao primeiro componente principal; Comp 2 refere-se ao segundo componente principal e assim sucessivamente; ii) as raízes características são reportadas da esquerda para a direita na forma descendente; iii) a variância proporcional das raízes características são as próximas; iv) a variância proporcional acumulada refere-se a soma da esquerda para a direita das variâncias dos cinco componentes; e v) os vetores das raízes características ou pesos das variáveis na determinação do componente principal são representados pelas colunas Vetor 1, Vetor 2, ..., Vetor 5.

Usando a regra de Guttman e Kaiser conforme exposto em Koutsoyiannis (1977), selecionam-se os componentes principais ( $P$ 's) de acordo com o critério de que o componente principal é maior que um, ou seja,  $P_m$  é considerado desde que  $\lambda_m > 1$ . De acordo com este critério, foram listados na Tabela A2 somente os cinco componentes principais que satisfazem a condição de serem maior que um.

A partir do critério acima, fica fácil construir o índice de atividade econômica. Assim, usando os coeficientes ( $C_{ij}$ )<sup>6</sup> e a percentagem da variância total explicada pela componente ( $P_j$ )<sup>7</sup>, define-se o peso de cada variável na construção do índice, ou seja:<sup>8</sup>

$$IV_i = \frac{C_{i1}^2 P_1}{(P_1 + P_2 + P_3 + P_4 + P_5)} + \frac{C_{i2}^2 P_2}{(P_1 + P_2 + P_3 + P_4 + P_5)} + \frac{C_{i3}^2 P_3}{(P_1 + P_2 + P_3 + P_4 + P_5)} + \frac{C_{i4}^2 P_4}{(P_1 + P_2 + P_3 + P_4 + P_5)} + \frac{C_{i5}^2 P_5}{(P_1 + P_2 + P_3 + P_4 + P_5)} \quad (7)$$

onde  $IV_i$  é o peso da variável  $i$  no índice;  $C_{ij}$  é o coeficiente da variável  $i$  na componente  $j$ ; e  $P_j$  é a parcela da variância explicada pela componente  $j$ .

A Tabela 1 apresenta o resultado da aplicação da metodologia acima. Verifica-se que a variável com maior peso no índice é o consumo de energia elétrica comercial. Compatível com a distribuição da economia por setor, conforme apresentado na seção 2. O menor peso é da variável imposto de renda pessoa física. Esta implica diretamente na determinação da renda disponível e, portanto, não oscila tanto.

O computo do Índice de Atividade Econômica de Maringá (IAEMga) é efetuado utilizando o  $IV_i$  e o número índice da variável  $V_i$  na seguinte fórmula:

$$IAEMga = \sum IV_i \times V_i. \quad (8)$$

O resultado da série se encontra na Tabela A3. No entanto, para facilitar sua visualização (comportamento) elaborou-se o Gráfico 1, no qual se apresenta as séries do IAEMga com e sem sazonalidade.<sup>9</sup> A diferença entre as duas condições são mínimas, em função de que as variáveis que compõem o índice foram dessazonalizadas *a priori*.

**Tabela 1 - Pesos das Variáveis que Compõem o Índice**

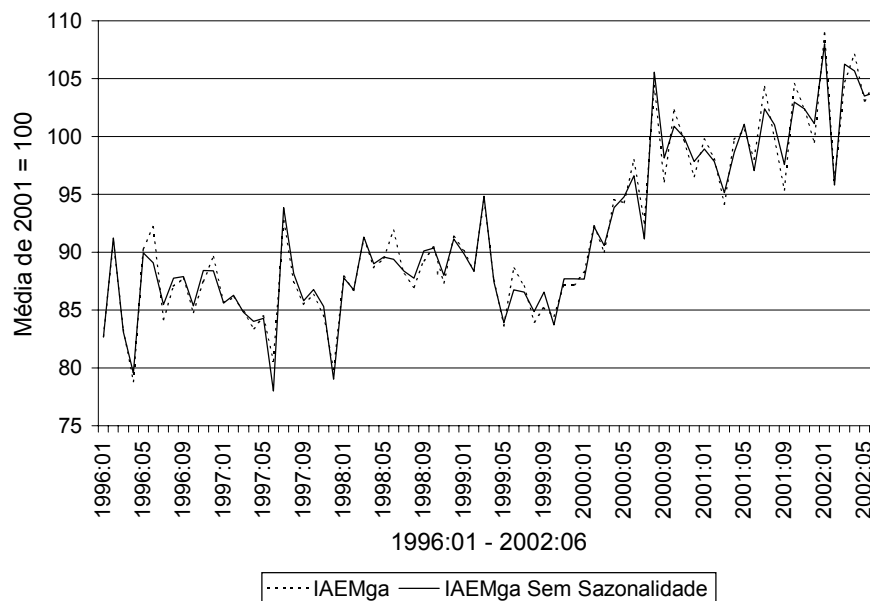
Variável	Peso	Variável	Peso	Variável	Peso
X01	0,0679	X07	0,0488	X13	0,0239
X02	0,0598	X08	0,0476	X14	0,0643
X03	0,0735	X09	0,0557	X15	0,0551
X04	0,0552	X10	0,0535	X16	0,0400
X05	0,0589	X11	0,0515	X17	0,0609
X06	0,0656	X12	0,0583	X18	0,0595

<sup>6</sup> Vetor  $i$ .

<sup>7</sup> Variância Proporcional do componente principal  $i$ .

<sup>8</sup> Azzoni e Latif (1995).

<sup>9</sup> Veja Tabela A3, segunda e terceira coluna.

**Gráfico 1 - Índice de Atividade Econômica de Maringá - IAEMga**

Na próxima seção, efetua-se a extração da tendência e flutuações irregulares, para que a série final resultante seja o ciclo econômico revisado para o município de Maringá.

#### 4. Análise da tendência, ciclo e flutuações irregulares do IAEMga

A metodologia de extração da tendência da série dessazonalizada do índice é efetuada através do filtro de Hodrick e Prescott (1997). A série é então dividida em três componentes: a tendência, os ciclos e a parte residual que compõe as flutuações irregulares. Esta técnica consiste na soma dos componentes de crescimento da série  $g_t$  e de ciclo  $c_t$ , que são minimizadas através da seguinte fórmula:

$$\text{Min}_{\{g_t\}_{t=1}^T} \left\{ \sum_{t=1}^T c_t^2 + \lambda \sum_{t=1}^T [(g_t - g_{t-1}) - (g_{t-1} - g_{t-2})]^2 \right\} \quad (9)$$

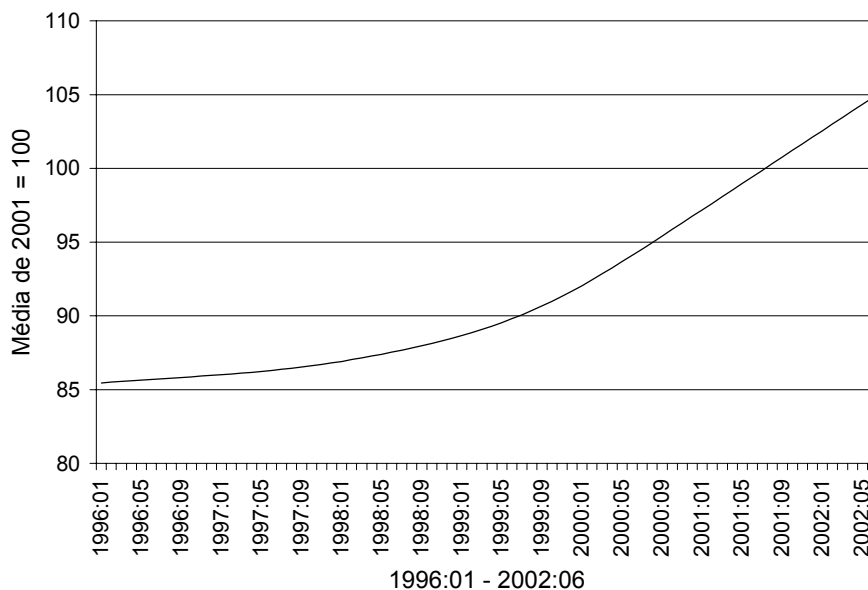
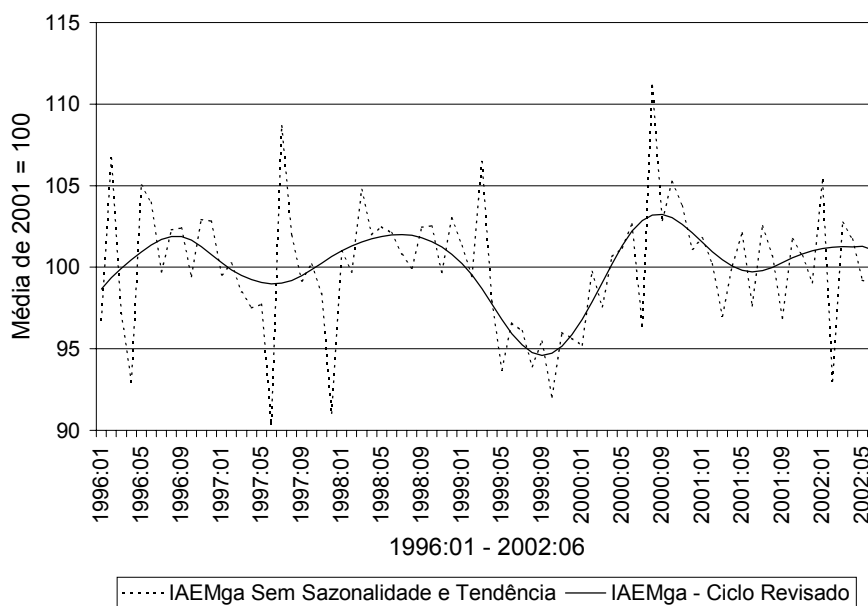
onde temos que  $c_t = y_t - g_t$ ; e  $\lambda$  é um número positivo que penaliza a variabilidade de crescimento da série. Como sabemos que no limite a diferença no tempo do componente de crescimento ( $g_{t+1} - g_t$ ) tende a se aproximar de alguma constante  $\beta$ , temos que o limite desta tendência é dado por  $g_0 + \beta t$ . Assim, para valores de  $\lambda$  próximos ao infinito, o limite da equação (9) é o mesmo que da regressão linear contra a variável tempo.

O Gráfico 2 apresenta o resultado da tendência utilizando a equação (9). A sua mudança na tendência foi significativa após o primeiro trimestre do ano de 1999, quando ocorreu a desvalorização cambial com a mudança para o regime flutuante.<sup>10</sup>

Na eliminação das flutuações irregulares da série do IAEMga sem sazonalidade e sem tendência foi utilizado o procedimento de ajustamento sazonal X-12-ARIMA multiplicativo. Como a tendência foi previamente eliminada, a série estimada após eliminar as flutuações irregulares será o ciclo revisado do IAEMga, que pode ser visualizado no Gráfico 3.

<sup>10</sup> Veja maiores detalhes em Averbug e Giambiagi (2000).



**Gráfico 2 - Tendência Estimada pelo Filtro Hodrick-Prescott****Gráfico 3 - Ciclo Revisado: IAEMga**

A leitura dos ciclos do Gráfico 3 foi transferida para a Tabela 2. A cronologia denotada por P (picos) e T (fundos) foram datadas de acordo com os respectivos períodos. As fases de contração duram em média 11 meses, enquanto que, as de expansão 12 meses.

**Tabela 2 - Cronologia dos Ciclos Revisados: IAEMga**

Datas de referência do IAEMga		Duração (meses)							
T = Fundo	P = Pico	Contração		Expansão		Ciclo			
		P → T	Σ	T → P	Σ	T → T	Σ	P → P	Σ
--	Setembro 1996	--	--	--	--	--	--	--	--
Julho 1997	Julho 1998	10	--	12	--	--	--	22	--
Setembro 1999	Setembro 2000	14	24	12	24	22	--	22	44
Junho 2001	--	09	33	--	--	21	43	--	--
Média		11		12		21,5		22	

Obs.: Σ = acumulado.

Na próxima seção, compara-se o IAEMga sem sazonalidade (IAEMga-Saz) com algumas séries de referência, através da análise de correlação e causalidade de Granger.

## 5. Análise de correlação e de causalidade de Granger

A não existência de dados sobre uma variável de produção e/ou consumo em nível mensal no município, levou a que se utilizassem séries nacionais como comparação da qualidade da série do Índice de Atividade Econômica de Maringá. Assim, utilizam-se as seguintes séries: PIB-Saz (índice do produto interno bruto sem sazonalidade); e IND-Saz (índice da produção industrial sem sazonalidade).<sup>11</sup>

De acordo com a Tabela 3, o IAEMga sem sazonalidade quando comparado com os indicadores agregados nacionais mensais e trimestrais demonstrou possuir excelente qualidade. A correlação de 83% com a série trimestral do PIB brasileiro indicou ser excelente e promissora a sua utilização como forma de prever o comportamento das atividades econômicas locais. Assim, o próximo passo é verificar qual variável antecede a outra por meio do teste de causalidade de Granger. Neste caso, a análise é efetuada entre o IAEMga e as séries de referência do PIB e produção industrial, todas sem sazonalidade.

**Tabela 3 - Matriz de Correlação**

Mensais	IAEMga-Saz	PIB-Saz	IND-Saz
IAEMga-Saz	1,00000	0,77554	0,66874
PIB-Saz	0,77554	1,00000	0,90267
IND-Saz	0,66874	0,90267	1,00000
Trimestrais	IAEMga-Saz	PIB-Saz	IND-Saz
IAEMga-Saz	1,00000	0,83409	0,72222
PIB-Saz	0,83409	1,00000	0,91430
IND-Saz	0,72222	0,91430	1,00000

Fonte dos dados: IPEA e Apêndice.

Este teste consiste em verificar se a variável X causa a variável Y. Isto quer dizer que as mudanças em X devem preceder as mudanças em Y.<sup>12</sup> A Tabela 4, apresenta os testes de causalidade de Granger entre o IAEMga e as variáveis nacionais do PIB e da produção industrial.

**Tabela 4 - Testes de Causalidade de Granger**

Variáveis	Variável Não "Granger Cause" IAEMga					IAEMga Não "Granger Cause" Variável				
	Número de Lags					Número de Lags				
Mensais	1	2	3	4	5	1	2	3	4	5
PIB	<b>11.35</b>	<b>3.24</b>	1.77	1.54	1.65	0.33	<b>3.18</b>	1.77	<b>2.90</b>	2.12
IND	<b>4.28</b>	1.09	1.00	1.44	1.35	1.19	2.81	1.44	2.51	2.12
Trimestrais	1	2	3	4	5	1	2	3	4	5
PIB	<b>5.76</b>	3.51	<b>3.66</b>	<b>5.26</b>	<b>4.86</b>	0.70	2.17	1.29	0.92	0.64
IND	3.02	1.91	1.41	1.14	0.66	0.72	2.14	0.69	1.33	0.86

Obs.: Valores da Estatística F = Estatística Wald; negrito: probabilidade menor que 0,05; mensais, n entre 67 e 71; trimestrais, n entre 19 e 23.

Conforme esperado, o PIB e a produção industrial (IND) nacional antecederem (causam) os movimentos do IAEMga. Ou seja, o comportamento nacional dita o comportamento da economia local. Este efeito de antecipação de dois meses, conforme é confirmado no teste trimestral, no qual o primeiro trimestre

<sup>11</sup> PIB: preços de mercado - índice encadeado (média 2001 = 100). Produção industrial: indústria geral (quantum) índice (média 2001 = 100). Fonte dos dados [www.ipeadata.gov.br](http://www.ipeadata.gov.br).

<sup>12</sup> Granger (1969).

apresenta-se como significativa. A produção industrial antecede seus efeitos em um mês e, portanto, demonstra na série trimestral não ser significativa conforme se espera. Em suma, os benefícios de um ciclo de crescimento positivo em nível nacional levam em média dois meses para afetar a economia local, e da mesma forma aplica-se a um quadro recessivo nacional.

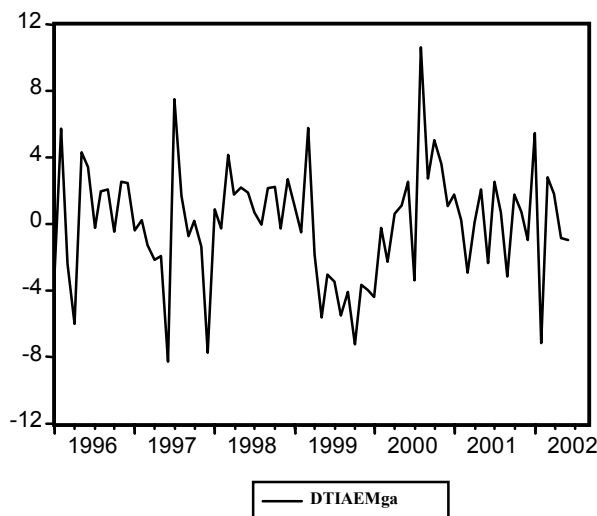
O interesse na próxima seção é fazer previsões do IAEMga sem sazonalidade, uma vez que, a possibilidade de antecipação das flutuações de curto prazo, poderiam fazer com que os agentes econômicos revejam suas expectativas e juntos amenizem as fases de contração, com intuito de fazer com que as fases de expansão sejam mais prolongadas.

## 6. Previsão da série IAEMga: Filtro de Kalman

A qualidade da previsão leva em consideração sempre a minimização dos erros da predição. No entanto, as séries tendem a ter elementos espúrios que causam a existência de valores fora do esperado, os chamados *outliers*. As previsões com a presença de *outliers* tendem a ser instáveis em função da variância ser inflada, especialmente quando usamos a técnica de Kalman (1960) e Kalman e Bucy (1961).

Como forma de ilustração transformamos a série do IAEMga através da retirada da tendência, utilizando o filtro de Hodrick and Prescott (1997), equação (9), na série DTIAEMga.<sup>13</sup> A seguir esta nova série foi testada com respeito a sua estacionariedade usando o teste ADF, proposto por Said e Dickey (1995). O resultado do teste  $ADF = 4,79$  com o valor crítico a nível de 1% de 3,51, indica que a nova série possui ordem de integração  $I(0)$ , ou seja, a série é estacionária. O Gráfico 4 a seguir apresenta a nova variável.

**Gráfico 4 - Série sem Tendência do IAEMga**



Os picos extremos da série podem indicar a existência de *outliers*. As técnicas de detecções sugeridas por Maddala e Kim (2002) apesar de úteis, somente nos informam que os coeficientes são instáveis e que a correção consiste basicamente em retirar as observações que estão causando o problema. Na prática, os *outliers* são facilmente visualizados graficamente e as regressões apresentam coeficientes instáveis. No próximo passo, analisamos esta questão em maiores detalhes.

O método de Filtro de Kalman consiste em um algoritmo aplicado a um sistema de equação, que deve ser representado na forma espaço de estado. Um modelo espaço de estado é representado em geral pelas duas equações seguintes:<sup>14</sup>

$$\text{- equação de medida:} \quad y_t = Z_t \alpha_t + d_t + S_t \varepsilon_t, \quad E(\varepsilon_t) = 0, \quad \text{Var}(\varepsilon_t) = H_t, \quad (10)$$

<sup>13</sup> IAEMga sem tendência.

<sup>14</sup> Baseado em Harvey (2001).

$$\text{- equação de transição: } \alpha_t = T_t \alpha_{t-1} + c_t + R_t \eta_t, \quad E(\eta_t) = 0, \quad \text{Var}(\eta_t) = Q_t, \quad (11)$$

Sendo que as dimensões das variáveis que compõem as equações acima são as seguintes:

$$\begin{array}{lllll} y_t : g \times 1, & Z_t : g \times k, & d_t : g \times 1, & S_t : g \times g, & \varepsilon_t : g \times 1, \\ \alpha_t : k \times 1, & T_t : k \times k, & c_t : k \times 1, & R_t : k \times k, & \eta_t : k \times 1, \end{array}$$

onde  $\varepsilon_t$  é uma perturbação não correlacionada serialmente com média zero e matriz de covariância  $H_t$ , e  $\alpha_t$  é um  $k \times 1$  vetor que não é observável, que é assumido ser gerado por um processo markoviano de primeira ordem, como mostrado na equação (11).  $T_t$  é uma matriz de transição e  $\eta_t$  é um vetor aleatório de perturbações não correlacionadas serialmente com média zero e matriz de covariância  $Q_t$ ,  $Z_t$ ,  $d_t$ ,  $S_t$ ,  $T_t$ ,  $c_t$  e  $R_t$  que pode depender da informação disponível no tempo  $t-1$  se nós assumirmos normalidade para os erros  $\varepsilon_t$  e  $\eta_t$ . A equação (10) é conhecida como equação de medida, observação ou sinal; enquanto que a equação (11) é chamada de equação de transição ou estado. O modelo espaço de estado requer as duas seguintes suposições adicionais: i) o vetor inicial  $\alpha_0$  tem uma média de  $a_0$  e uma matriz de covariância de  $\Sigma_0$ , isto é,  $E(\alpha_0) = a_0$  e  $\text{Var}(\alpha_0) = \Sigma_0$ ; ii) as perturbações  $\varepsilon_t$  e  $\eta_t$  não são correlacionadas por todos os períodos de tempo, e não são correlacionadas com a variável estado inicial, isto é,  $E(\varepsilon_t \eta'_s) = 0$  para todo  $t$  e  $s$ , e  $E(\varepsilon_t \alpha'_0) = E(\eta_t \alpha'_0) = 0$  para  $t = 1, \dots, T$ .

Algumas observações devem ser feitas relativas as suposições anteriores: i) a segunda suposição não garante nenhuma correlação entre  $\varepsilon_t$  e  $\alpha_t$ , e nenhuma correlação entre  $\eta_t$  e  $\alpha_{t-1}$ , isto é,  $E(\varepsilon_t \alpha'_t) = 0$  e  $E(\eta_t \alpha'_{t-1}) = 0$  para todo  $t$ ; ii)  $Z_t$ ,  $d_t$ ,  $S_t$ ,  $T_t$ ,  $c_t$  e  $R_t$  podem depender de um vetor do parâmetro não conhecido  $\theta$ . Neste caso,  $\theta$  tem que ser estimado; iii) os termos de erro  $\varepsilon_t$  e  $\eta_t$  são geralmente assumidos serem normais, mas a suposição de normalidade é necessária se pedido pelo procedimento da derivação do algoritmo recursivo linear. No caso das derivações por *mixed estimation* e *minimum mean square linear estimator* (MMSLE), não há nenhuma suposição relativa a distribuição para os erros  $\varepsilon_t$  e  $\eta_t$ . No entanto, derivar o algoritmo recursivo linear padrão baseado em funções de densidade requererá a suposição de normalidade.

O algoritmo do Filtro de Kalman é derivado das equações (10) e (11). Antes de apresentar as equações que descrevem o algoritmo recursivo para a versão discreta linear do filtro, o processo de filtração será definido.

Se nós considerarmos o problema de estimar  $\alpha_t$  usando a informação até o tempo  $s$ , isto é,  $Y_s = \{y_1, y_2, \dots, y_s\}$  e nós denotamos  $E(\alpha_t | Y_s) \equiv \alpha_{t|s}$  a expectativa condicional de  $\alpha_t$  dado  $Y_s$ , a avaliação de  $\alpha_{t|s}$  tem diferente nomes: filtragem, se  $t = s$ ; suavização, se  $t < s$ ; previsão, se  $t > s$ ; e  $\Sigma_{t|s}$  é a matriz de covariância condicional de  $\alpha_t$  dado  $Y_s$ , isto é,  $\text{Cov}(\alpha_t | Y_s)$ .

O algoritmo do Filtro de Kalman é dado pelas seguintes equações:

i) As equações de previsão:

$$\alpha_{t|t-1} = T_t \alpha_{t-1|t-1} + c_t \quad (12)$$

$$\Sigma_{t|t-1} = T_t \Sigma_{t-1|t-1} T'_t + R_t Q_t R'_t \quad (13)$$

ii) A previsão estimada e sua variância:

$$Y_{t|t-1} = Z_t \alpha_{t|t-1} + d_t \quad (14)$$

$$F_{t|t-1} = Z_t \Sigma_{t|t-1} Z'_t + S_t H_t S'_t \quad (15)$$

iii) O Filtro de Kalman:

$$K_t = \Sigma_{t|t-1} Z'_t F_{t|t-1}^{-1} \quad (16)$$

iv) As equações de atualização:

$$\Sigma_{t|t} = \Sigma_{t|t-1} - K_t F_{t|t-1} K'_t \quad (17)$$

$$\alpha_{t|t} = \alpha_{t|t-1} + K_t (y_t - y_{t|t-1}) \quad (18)$$

onde  $Z_t, d_t, S_t, T_t, c_t, R_t, Q_t, H_t, \alpha_{0|0}$  e  $\sum_{0|0}$  são conhecidos agora e  $t=1,2,\dots,T$ .

As equações (12) e (13) são chamadas de equações de previsão pela razão que essencialmente prevêem o parâmetro  $\alpha_t$  e sua matriz de covariância condicional usando informação até o tempo  $t-1$ . A equação (16) é chamada de filtro de Kalman. Sua interpretação depende do algoritmo de derivação. Quando o algoritmo de filtragem é derivado do MMSLE,  $K_t$  é escolhido tal que a filtragem estimada  $\alpha_{t|t}$  tem variância mínima. As equações (17) e (18) tem a tarefa de combinar a nova observação obtida no tempo  $t$  (isto é,  $y_t$ ) e a informação passada até o tempo  $t-1$  (isto é,  $y_{t-1}$ ).

Conforme se esperava, as estimativas utilizando o Filtro de Kalman geraram várias regressões, na quais, a instabilidade das estatísticas dos parâmetros estavam presentes. Assim, recorreremos a técnica proposta por Lin e Guttman (1993). Esta técnica apresenta-se como simples no sentido de particionar *a priori* em vez de *posteriori*, as distribuições dos erros de previsão, como proposto por Guttman e Peña (1989). Esta técnica pode ser descrita da seguinte forma: i) considere que  $y_t$  advêm de duas distribuições  $y_t \sim N(Z_t\alpha_t, V_1)$  ou  $y_t \sim N(Z_t\alpha_t, V_2)$ , portanto, diferem em variâncias; ii) deixe  $\lambda_1 = 1 - \lambda_2$ , tal que,  $0 < \lambda_1 < 1$ ; iii) especificamos os erros da equação (10) da seguinte forma:  $\varepsilon_t \sim N(0, \lambda_1 V_1 + \lambda_2 V_2)$ ; iv) a variância da equação (10) torna-se  $H_t = \lambda_1 V_1 + \lambda_2 V_2$ , portanto, sem alterar a característica do filtro; v) os valores de  $\lambda$ 's serão aqueles que minimizam um critério da nossa escolha; vi) no computo da probabilidade *a posteriori* dos coeficientes  $\lambda$ 's, os valores das variâncias são reduzidas à metade.

Em palavras simples, esta técnica consiste em encontrar a proporção ideal das observações que advêm das distribuições com e sem *outliers*. Para atingir este objetivo é preciso encontrar as distribuições que serão efetuadas as combinações. Na prática Lin e Guttman (1993) sugerem utilizar o desvio padrão da variável  $y_t$  no lugar de sua variância na combinação acima. Como ambas, a *posteriori*, estão sendo corrigidas pela metade, é como se reduzissemos as variâncias a serem combinadas por um fator de correção. No nosso caso, o desvio padrão da variável é o seguinte  $\sigma_{y_t} = 3,43$ . O desvio padrão dos erros da equação (10) requer que façamos uma regressão de mínimos quadrados utilizando os valores dos AR's e MA's como independentes.<sup>15</sup> O modelo de melhor ajuste foi um ARMA(2,4).<sup>16</sup> O desvio padrão dos erros obtidos foi  $\sigma_e = 2,43$ . Em vez de seguir os autores e reduzir ambos os valores dos desvios à metade, quando da combinação das distribuições, nós reduzimos somente o valor da variável  $y_t$ , mantendo o outro, ou seja, nós utilizamos a combinação  $\varepsilon_t \sim \lambda_1 \times N(0, 1,715) + (1-\lambda_1) \times N(0, 2,43)$ .

O passo a seguir, foi encontrar um critério que pudéssemos efetuar a escolha do valor ótimo de  $\lambda_1$ . Optamos pelo Critério de Informação de Akaike (AIC). De acordo com este critério, quando estamos estimando modelos ARMA, devemos optar pelo menor valor obtido da seguinte formula:

$$AIC(p) = n \log \sigma_e^2 + 2p. \quad (19)$$

onde  $n$  é o tamanho da amostra e representa a variância dos erros da regressão e  $p$  é o número de parâmetros. O resultado deste critério para os valores de  $\lambda$ 's estão na Tabela 5.

Apesar do menor valor, ser o valor de  $\lambda_1 = 0,50$ , nós optamos por apresentar os valores dos coeficientes  $\lambda_1 = 0,60$ , cujo níveis de significância das estatísticas dos parâmetros são superiores. Estes apresentaram-se como significativos a nível de 5%, com exceção do coeficiente c(4), que é significativo a nível de 10%, conforme podemos verificar no resultado da Tabela 6.

<sup>15</sup> AR: Autoregressive, MA: Moving average.

<sup>16</sup> ARMA: Autoregressive moving average.

**Tabela 5 - Critério de Informação de Akaike**

$\lambda_1$	$\lambda_2 = (1 - \lambda_1)$	AIC
0,10	0,90	5,225
0,20	0,80	5,217
0,30	0,70	5,209
0,40	0,60	5,204
0,50	0,50	5,201*
0,60	0,40	5,202*
0,70	0,30	5,208
0,80	0,20	5,220
0,90	0,10	5,238

**Tabela 6 - O Modelo Estado de Espaço**

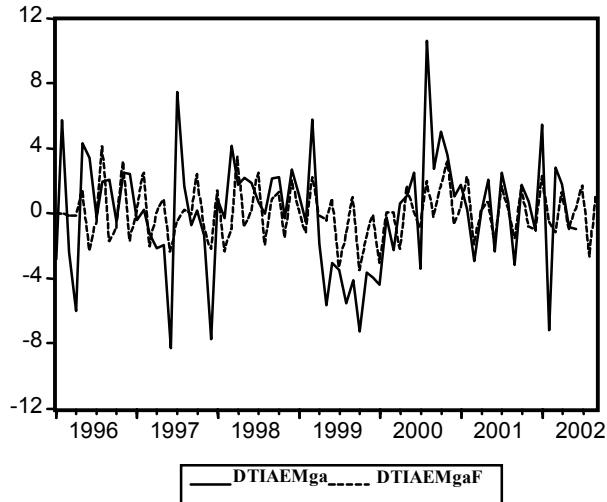
Sspace: DTIAEMga				
Método: Máxima verossimilhança (Marquardt)				
Amostra: 1996:01 2002:09				
Observações incluídas: 81				
Observações válidas: 78				
Ajuste da estimação: tol= 0.00010, derivs=accurate numeric				
Valores Iniciais: C(1)=1.39377, C(2)=1.42355, C(3)=0.45980,				
C(4)=0.26114, C(5)=-1.21549, C(6)=-0.99593				
Convergência alcançada após 20 iterações				
	Coefficiente	Erro Padrão	z-estatístico	Probabilidade
C(1)	1.405430	0.159032	8.837388	0.0000
C(2)	1.439074	0.269818	5.333501	0.0000
C(3)	0.467039	0.202324	2.308367	0.0210
C(4)	0.261264	0.145180	1.799585	0.0719
C(5)	-1.215711	0.017003	-71.49858	0.0000
C(6)	-0.996267	0.012165	-81.89477	0.0000
	Estado Final	Root MSE	z-estatístico	Probabilidade
SV1	-58.39408	11.37175	-5.135014	0.0000
SV2	28.43390	11.41318	2.491321	0.0127
SV3	23.91597	11.12705	2.149354	0.0316
SV4	-57.72430	10.54828	-5.472392	0.0000
SV5	46.43343	10.45343	4.441932	0.0000
Máxima Verossimilhança	-196.8995	Akaike - Critério		5.202550
Paramêtros	6	Schwarz - Critério		5.383835
Diffuse priors	0	Hannan-Quinn - Critério		5.275122

As previsões a partir do resultado da Tabela 6, ou seja, utilizando Filtro de Kalman para a variável DTIAEMga, se encontram no Gráfico 5. Observa-se no gráfico que a variável não minimiza completamente os picos, em função destes representarem provavelmente elementos de choques, cujos valores produzem os chamados *outliers*. A vantagem do Filtro de Kalman é a utilização dos valores passados para a projeção dos valores futuros. A previsão para esta variável utilizando os valores passados conhecidos, produz os resultados do Gráfico 6. Neste caso, os valores passados auxiliam as previsões futuras, entrando, na regressão como fonte de projeção de forma a minimizar os erros da predição. Este mecanismo de autocorreção permite que a projeção seja incorporada ao modelo e, desta forma, implementar novas predições utilizando o valor projetado anteriormente de forma recursiva.

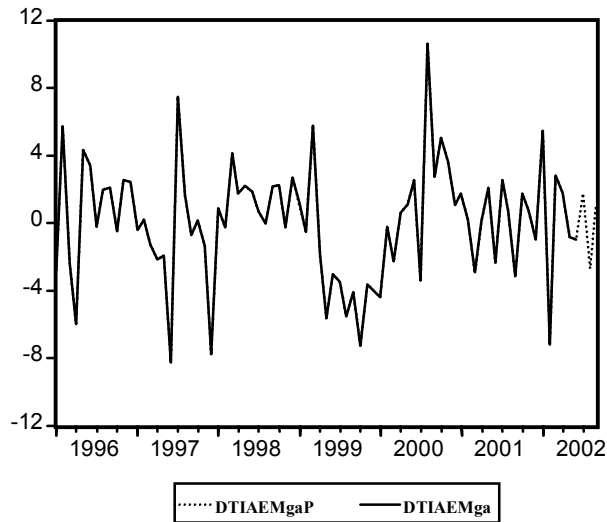
As previsões com base nos valores passados conhecidos, estão representadas pela variável DTIAEMgaP. Estes resultados referem-se somente a parte sem tendência da variável IAEMga.

Para restabelecer os valores a nível agregado da variável IAEMga e, assim, proceder no cálculo do coeficiente de correlação, para termos noção mínima da qualidade das predições, precisamos adicionar as duas partes. Assim, a variável DTIAEMgaF foi adicionada a parte da tendência denominada de IAEMgaTR, para formar a variável IAEMgaF, afim de ser comparada com a original. O resultado foi um coeficiente de correlação de 90,13%. Veja a comparação das duas séries no Gráfico 7.

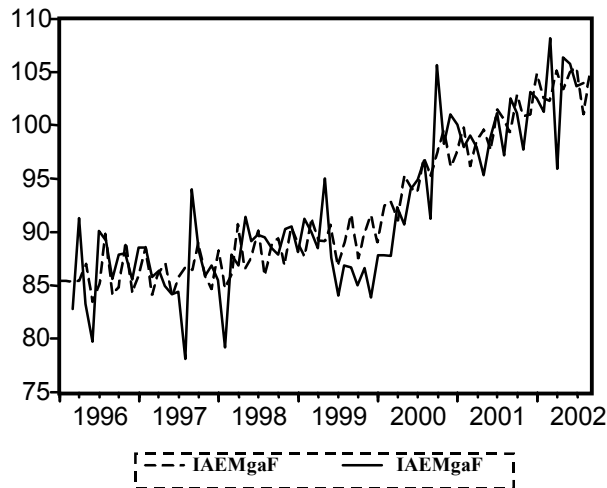
**Gráfico 5 - Previsões do Filtro de Kalman**



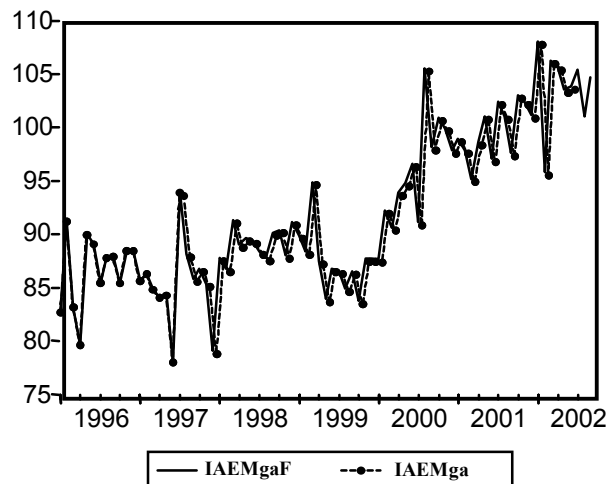
**Gráfico 6 - Previsões do Filtro de Kalman  
Utilizando Valores Passados**



**Gráfico 7 - Previsões do Filtro de Kalman  
Para a Série Original**



**Gráfico 8 - Previsões da Série Original Utilizando as Informações Passadas**



A série IAEMgaF representa as previsões do IAEMga utilizando os valores passados da mesma. Assim, o ajuste é excelente com erro para o passado de zero. Esta formulação facilita a projeção de valores futuros utilizando um método recursivo no qual o próximo valor futuro é incorporado na reestimação e utilizado na projeção de novos valores futuros.

## 7. Previsão com modelos Box-Jenkins

O conhecimento *a priori* da série IAEMga, conforme o Gráfico 1, indica os seguintes comportamentos: i) não-linearidade; e ii) não-estacionariedade. O teste de raiz unitária na diferença da variável IAEMga, ou seja, DIAEMga produziu um  $ADF = 12,01$ , com um valor crítico a nível de 1% de 3,52, o que indica que a diferença da variável IAEMga, ou seja, o DIAEMga é  $I(0)$ , sendo que, a variável em si possui ordem de integração um, ou seja, é  $I(1)$ .

O próximo passo é aprendermos sobre o comportamento da auto-regressividade e mudanças na média da variável. Assim, foram efetuadas regressões de mínimos quadrados considerando os processos auto-regressivos (AR) e de média móvel (MA) como independentes. Dentre uma série de regressões, os resultados da Tabela 7 mostraram-se significantes em seus coeficientes:

**Tabela 7 - Resultado Econométrico dos AR's e MA's**

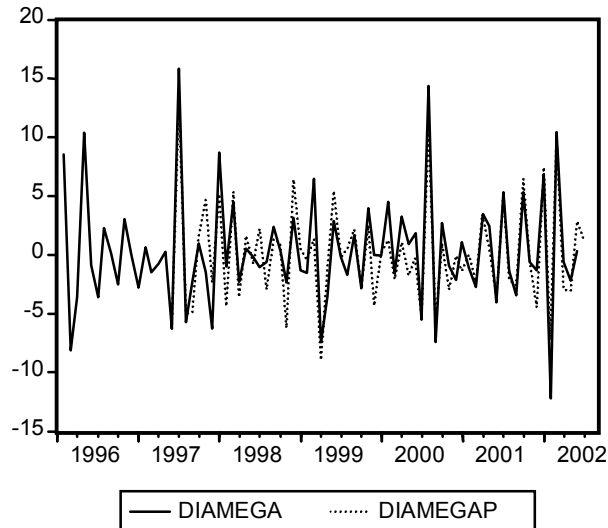
Variável Dependente: D(IAEMga-Saz)				
Método: Mínimos Quadrados				
Amostra: 1997:06 - 2002:06				
Número de observações: 61				
Variável	Coefficiente	Erro Padrão	t-estatístico	Probabilidade
C	0.273640	0.214163	1.277719	0.2067
AR(1)	-0.579661	0.106604	-5.437497	0.0000
AR(13)	-0.317272	0.110047	-2.883053	0.0056
AR(16)	-0.348088	0.103339	-3.368409	0.0014
MA(1)	-0.189665	0.089748	-2.113314	0.0391
MA(17)	0.779758	0.062284	12.51949	0.0000
$R^2$	0.760283	Média da V. D.		0.320080
$R^2$ Ajustado	0.738490	Desvio Padrão da V.D.		4.852083
Soma dos Erros	2.481259	Akaike - Critério		4.748590
SQR	338.6156	Schwarz - Critério		4.956217
Máxima Verossimilhança	-138.8320	F-estatística		34.88740
Durbin-Watson	1.969049	Prob(F-estatística)		0.000000

Obs: i) D(IAEMga-Saz) refere-se a diferença da variável IAEMga-Saz; ii) SQR - Soma do Quadrado dos Resíduos; e iii) V. D.: Variável Dependente.



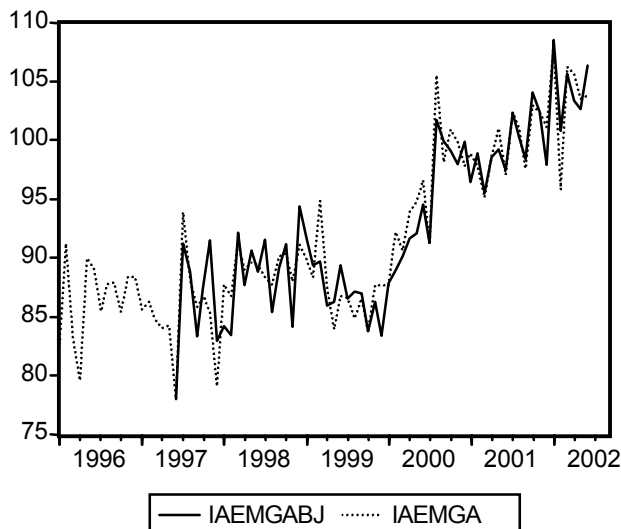
De acordo com este modelo, tem-se uma variável que possui auto-regressão de ordem um, treze e dezesseis, AR (1,13,16); e média móvel um e dezessete, MA(1,17). Com estas informações fica fácil utilizar o modelo Box-Jenkins para a previsão.<sup>17</sup> As previsões utilizando as informações da regressão acima, permitem obter a previsão da diferença (DIAEMga), conforme Gráfico 9.

**Gráfico 9 - Previsão da Diferença do IAEMga**



Como a diferença é estacionária, a variável em nível é de ordem um, I(1). Temos que estes modelos são conhecidos como Auto-regressivo Integrado de Média Móvel (ARIMA). Usando um modelo ARIMA ( $(\|1,13,16\|, 1, \|1,17\|)$ ), de acordo com a Tabela 7, obtivemos o seguinte resultado:

**Gráfico 10 - Projeções do IAEMga: Modelo Box-Jenkins**



O coeficiente de correlação obtido entre as duas variáveis do Gráfico 10 foi de 94,65%. A perda de informações das amostras é uma das características destes modelos. Em especial, neste modelo foi efetuado a previsão para somente um período a frente, enquanto no modelo anterior, trabalhamos com três períodos.

<sup>17</sup> Box e Jenkins (1976).

## 8. Comparando os modelos

A comparação natural entre os dois modelos, Filtro de Kalman e Box-Jenkins, apesar de apresentarem coeficientes de correlação muito próximos, precisam de testes estatísticos. Aqui, usaremos o teste desenvolvido por Diebold e Mariano (1995). De acordo com este teste, devemos construir uma função de perda, onde:

$$f = (e_{bj})^2 - (e_{fk})^2. \quad (20)$$

onde  $\{e_{bj}\}$  é uma sequência de erros de previsão do modelo Box-Jenkins;  $\{e_{fk}\}$  é uma sequência de erros de previsão do modelo de Filtro de Kalman; estes erros referem-se ao período 1997:06–2002:06. O teste estatístico que é assintoticamente distribuído como uma função normal, é o seguinte:

$$S = \frac{f_m}{\sigma_f}. \quad (21)$$

onde  $f_m$  é a média da amostra  $\{f\}$  para o período em questão. A hipótese nula é de que  $f_m = 0$ . Os valores de  $S$  positivos significam que o modelo do Filtro de Kalman é superior, enquanto os negativos dão respaldo ao de Box-Jenkins. Os valores obtidos para as variáveis  $f_m = -7,04$  e  $\sigma_f = 31,01$ , portanto,  $S = -0,22$ . Apesar do resultado informar que o modelo de Box-Jenkins se ajusta melhor aos dados, o mesmo não é estatisticamente diferente de zero, portanto, a hipótese nula é aceita.

## 9. Conclusão

Uma das características dos indicadores agregados, em geral, é a demora na sua produção. Os impactos de suas movimentações, quando tornam-se conhecidos, já ocorreram nas economias locais. Ao produzir um índice que antecipa os efeitos destes movimentos de nível agregado para a economia local cria-se um instrumento de mensuração dos movimentos futuros. A incorporação deste instrumento nas decisões dos agentes locais permitem que estes minimizem perdas e maximizem ganhos. Esta tarefa de incorporação é difícil, mas depende em muito da credibilidade técnica do índice. O índice proposto satisfaz todas as condições técnicas, além de sinalizar com maior rapidez as flutuações de curto prazo. Para tanto, o IAEMga em todas as suas versões, possibilita aos agentes econômicos tomarem decisões de acordo com as suas expectativas, provenientes de suas análises pontuais.

Apesar de não ter sido possível comparar o IAEMga com séries de referência longas para o município de Maringá, pode-se supor, ou auferir, que o IAEMga sem sazonalidade seja um bom índice coincidente e a previsão do IAEMga sem sazonalidade seja um indicador antecedente da atividade econômica de Maringá. Com isso, pode-se utilizá-lo em substituição a outros índices tradicionais; tendo como vantagem, que a sua coleta, manipulação e publicação são muito mais rápidas, o que faz com que os agentes econômicos que necessitem de informações instantâneas possam se utilizar do IAEMga com alto nível de confiança.

O modelo de previsão de Box-Jenkins (ARIMA) foi o que apresentou melhor correlação e menor erro quadrado de previsão na função de perda. No entanto, estatisticamente as previsões dos modelos não diferem. Cabe ressaltar, que o modelo de previsão baseado no Filtro de Kalman permite efetuar previsões incorporando os valores ocorridos no passado. Considerando estritamente do ponto de vista de praticidade na manipulação dos modelos de previsão, o modelo ARIMA oferece maior facilidade na sua utilização.

## 10. Bibliografia

- AZZONI, C. R. e LATIF, Z. A. Indicador de movimentação econômica IMEC-FIPE. *Encontro Brasileiro de Econometria*, 1, p. 53-69, 1995.
- BOX, G. E. P. e JENKINS, G. M. *Time series analysis: forecasting and control*. New York: Holden-Day, 1976.
- BRUNET, O. Calculation of composite leading indicators: a comparison of two different methods. *Center for International Research on Economic Tendency: CIRET Conference in Paris*, 2000.
- CARDIM DE CARVALHO, F. J. e HERMANNY, P. F. Ciclos e previsão cíclica: o debate teórico e um modelo de indicadores antecedentes para a economia brasileira. *Revista Análise Econômica*, ano 21, n.39, 2003.
- CHAUVET, M. A Monthly Indicator of Brazilian GDP. *Mimeo*, 2000a.
- CHAUVET, M. The Brazilian Business Cycle and Growth Cycle. *Mimeo*, 2000b.
- CONTADOR, C. R. *Ciclos Econômicos e Indicadores de Atividade no Brasil*. IPEA, 1977.
- DICKEY, D. A e FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, v. 74, n. 366, p. 427-37, 1979.
- DIEBOLD, F. X. e MARIANO, R. S. Comparing predictive accuracy. *Journal of Business and Statistics*, 13, p. 253-63, 1995.
- DUARTE, A. J. M.; ISSLER, J. V. e SPACOV, A. Indicadores coincidentes de atividade econômica e uma cronologia de recessões para o Brasil. *Ensaio Econômicos*, n.527. Fundação Getúlio Vargas: EPGE, 2004.
- GRANGER, C. W. J. Investigating causal relations by econometrics methods and cross-spectral methods. *Econometrica*, 37, p. 424-38, 1969.
- GREENE, W. H. *Econometric analysis*. Upper Saddle River: Prentice Hall, 1997.
- GUJARATI, D. N. *Econometria básica*. São Paulo: Makron Books, 2000.
- HARVEY, A. C. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press, 2001.
- HODRICK, R. J. e PRESCOTT, E. C. Postwar U.S. Business Cycles: an Empirical Investigation. *Journal of Money, Credit and Banking*, 29, p. 1-16, 1997.
- HOFFMANN, R. Componentes principais e análise fatorial. *Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo: Série Didática*, 90, 1994.
- JACKSON, J. E. *A User's Guide to Principal Componentes*. New York: J. Wiley, 1991.
- JOHNSTON, J. e DINARDO, J. *Econometric Methods*. New York: The McGraw Hill, 1997.
- JUDGE, G. G. *et alli*. *The theory and practice of econometrics*. New York: Wiley, 1985.
- KOUTSOYIANNIS, A. *Theory of econometrics*. London: Macmillan, 1977.
- LIN, D. K. J. e GUTTMAN, I. Handling Spuriousity in the Kalman Filter. *Statistics & Probability Letters*, 16, p. 259-68, 1993.
- MADDALA, G. S. e KIM, I. M. *Unit Roots, Cointegration and Structural Change*. Cambridge: Cambridge University Press, 2002.
- MORAIS, I. A. C. de e PORTUGAL, M. S. Um novo índice coincidente para a atividade industrial do estado do Rio Grande do Sul. *Texto para Discussão*, 2004/01. Programa de Pós-Graduação em Economia: UFRGS, 2004.
- NILSSON, R. Confidence Indicators and Composite Indicators. *Center for International Research on Economic Tendency: CIRET Conference in Paris*, 2000.
- RIBEIRO, V. S. *Elaboração de um Índice de Atividade Econômica: Município de Maringá*. Programa de Pós-graduação em Economia, Universidade Estadual de Maringá, 2003.
- SAID, S. E.; DICKEY, D. A. Testing for unit root in autoregressive - moving average models of unknown order. *Biometrika*, 71, p. 599-607, 1984.
- SALAZAR, E. *et alli*. Leading indicators. *Meeting on OECD Leading Indicators*, 1996.
- SHARMA, S. *Applied multivariate techniques*. John Wiley & Sons, p. 58-89, 1996.
- U. S. CENSUS BUREAU. *X-12-ARIMA: Reference Manual*. Washington, 2000.
- ZARNOWITZ, V. What is a Business Cycle? *National Bureau of Economic Research: NBER Working Paper*, 3863, 1991.



**Tabela A2 - Componentes Principais: Matriz de Correlação**

	Comp 1	Comp 2	Comp 3	Comp 4	Comp 5
Raízes Características	5.753338	2.904304	1.586204	1.215006	1.101643
Variância Proporcional	0.319630	0.161350	0.088122	0.067500	0.061202
Var. Prop. Acumulada	0.319630	0.480980	0.569102	0.636603	0.697805
Vetores das Raízes Características:					
Variável	Vetor 1	Vetor 2	Vetor 3	Vetor 4	Vetor 5
X01	0.331219	-0.273957	-0.002648	-0.004043	0.054413
X02	0.194778	-0.333445	-0.211681	-0.337492	-0.010343
X03	0.386293	-0.146459	0.039154	-0.004753	0.001880
X04	0.280142	-0.225116	0.077565	-0.253428	-0.077449
X05	-0.225034	-0.292258	0.276232	-0.206942	-0.155992
X06	-0.128462	0.281437	-0.333521	-0.087702	-0.533700
X07	-0.104046	0.250519	0.344180	-0.106832	0.388877
X08	0.065931	-0.158115	0.478118	0.290350	-0.179888
X09	0.181445	0.243293	0.387931	0.280602	-0.058177
X10	0.154084	-0.115057	-0.354381	0.233872	0.458673
X11	0.230755	0.125685	0.148016	-0.452099	0.103788
X12	0.180050	0.360583	0.109846	-0.337450	0.097137
X13	0.116062	-0.125658	0.225038	-0.239681	-0.156438
X14	0.311308	0.056509	0.003147	0.291227	-0.353441
X15	0.321194	0.044918	-0.163195	0.171202	0.114600
X16	0.214668	0.245183	0.077419	0.111009	0.186908
X17	0.343120	0.079627	-0.033986	0.050712	-0.241296
X18	0.150489	0.428427	-0.136208	-0.193485	-0.092296

Obs.: Var. Prop. = Variância Proporcional.

**Tabela A3 - Índice de Atividade Econômica de Maringá - IAEMga:  
Séries históricas (números índices, média de 2001 = 100)**

Obs.	♣ IAEMga	♥ IAEMga- Saz	♠ IAEMga- SazTend	♦ IAEMga- Ciclo	Obs.	♣ IAEMga	♥ IAEMga- Saz	♠ IAEMga- SazTend	♦ IAEMga- Ciclo
1996:01	82,711	82,667	96,738	98,660	1999:04	87,595	87,439	97,908	97,752
1996:02	90,913	91,221	106,690	99,363	1999:05	83,632	83,910	93,706	96,805
1996:03	83,300	83,141	97,186	99,929	1999:06	88,574	86,738	96,588	95,934
1996:04	78,809	79,565	92,955	100,463	1999:07	87,076	86,563	96,105	95,232
1996:05	90,305	89,936	105,014	100,941	1999:08	83,970	84,856	93,909	94,771
1996:06	92,195	89,090	103,970	101,359	1999:09	85,219	86,546	95,457	94,598
1996:07	84,163	85,482	99,706	101,698	1999:10	84,440	83,729	92,022	94,733
1996:08	87,076	87,742	102,288	101,888	1999:11	87,185	87,686	96,012	95,168
1996:09	87,736	87,893	102,409	101,888	1999:12	87,195	87,698	95,652	95,871
1996:10	84,778	85,370	99,417	101,667	2000:01	88,220	87,663	95,227	96,791
1996:11	87,461	88,423	102,916	101,272	2000:02	92,267	92,203	99,739	97,871
1996:12	89,621	88,401	102,834	100,778	2000:03	89,996	90,588	97,571	99,036
1997:01	85,631	85,607	99,528	100,274	2000:04	94,541	93,874	100,665	100,211
1997:02	86,088	86,265	100,233	99,831	2000:05	94,266	94,770	101,171	101,307
1997:03	84,856	84,800	98,469	99,487	2000:06	97,963	96,601	102,657	102,216
1997:04	83,411	84,035	97,516	99,236	2000:07	92,512	91,126	96,395	102,861
1997:05	84,442	84,277	97,727	99,066	2000:08	104,396	105,556	111,144	103,204
1997:06	80,622	78,014	90,394	98,984	2000:09	96,092	98,141	102,859	103,246
1997:07	92,720	93,856	108,661	99,018	2000:10	102,290	100,882	105,243	103,037
1997:08	87,389	88,130	101,942	99,185	2000:11	99,618	99,924	103,763	102,638
1997:09	85,490	85,812	99,167	99,474	2000:12	96,521	97,820	101,111	102,107
1997:10	86,331	86,772	100,177	99,847	2001:01	99,741	98,918	101,778	101,529
1997:11	84,533	85,314	98,388	100,254	2001:02	98,043	97,827	100,196	100,963
1997:12	79,855	79,038	91,046	100,651	2001:03	94,140	95,138	96,999	100,458
1998:01	87,881	87,783	100,998	101,009	2001:04	99,763	98,628	100,102	100,064
1998:02	86,683	86,761	99,694	101,310	2001:05	100,557	101,058	102,104	99,809
1998:03	91,193	91,294	104,763	101,554	2001:06	98,019	97,055	97,619	99,720
1998:04	88,677	89,008	101,996	101,742	2001:07	104,307	102,393	102,526	99,793
1998:05	89,461	89,577	102,497	101,878	2001:08	99,826	101,020	100,701	99,997
1998:06	91,892	89,388	102,123	101,969	2001:09	95,368	97,592	96,852	100,282
1998:07	88,089	88,347	100,773	102,000	2001:10	104,542	102,960	101,728	100,574
1998:08	86,982	87,760	99,935	101,949	2001:11	102,367	102,381	100,712	100,821
1998:09	89,286	90,121	102,443	101,808	2001:12	99,500	101,119	99,034	101,018
1998:10	90,442	90,360	102,526	101,573	2002:01	108,920	108,013	105,326	101,150
1998:11	87,341	88,007	99,664	101,243	2002:02	96,213	95,819	93,031	101,236
1998:12	91,344	91,152	103,017	100,805	2002:03	104,730	106,257	102,720	101,274
1999:01	90,152	89,849	101,328	100,233	2002:04	107,085	105,672	101,716	101,250
1999:02	88,360	88,349	99,412	99,526	2002:05	103,075	103,502	99,201	101,284
1999:03	94,535	94,841	106,464	98,687	2002:06	104,593	103,802	99,066	101,076

♣ IAEMga original, ♥ IAEMga sem sazonalidade, ♠ IAEMga sem sazonalidade e tendência, ♦ Ciclo Revisado do IAEMga.