



## Monografia de Final de curso

Criação de Diferentes Modelos de Previsão para a Produção Industrial Doméstica.

Aluno: Eduardo Papazian de Azevedo Santos

Nº. de Matrícula: 0313358

Orientador: Marco Antonio F.de H. Cavalcanti

Tutor: Márcio Gomes Pinto Garcia

Junho de 2007

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO  
DEPARTAMENTO DE ECONOMIA

MONOGRAFIA DE FINAL DE CURSO

Criação de Diferentes Modelos de Previsão para a Produção Industrial  
Doméstica.

---

Eduardo Papazian de Azevedo Santos

Nº de matrícula: 0313358

Orientador: Marco Cavalcanti

Junho de 2007

“Declaro que o presente trabalho é de minha autoria e que não recorri para realizá-lo, a nenhuma forma de ajuda externa, exceto quando autorizado pelo professor tutor”.

“As opiniões expressas neste trabalho são de  
responsabilidade única e exclusiva do autor.”

**À minha família, por todo apoio que me deram durante minha graduação;**  
**A todos os professores, que contribuíram para minha formação acadêmica;**  
**Aos amigos, que sempre me ajudaram e tornaram o aprendizado mais agradável.**

## **Sumário:**

- 1. Introdução**
- 2. O Índice de Produção Industrial**
  - 2.1 Como é elaborada a pesquisa e de que maneira o índice é construído
  - 2.2 A importância de antecipar a Produção Industrial
- 3. Os métodos a serem utilizados**
  - 3.1 ARIMA
  - 3.2 VAR
  - 3.3 ADL
- 4. Rodando os modelos**
  - 4.1 ARIMA
  - 4.2 VAR
  - 4.3 ADL
- 5. Comparando os resultados**
  - 5.1 Comparando as projeções com os dados reais
  - 5.2 Qual método possui maior capacidade preditiva
  - 5.3 Importância de cada resultado
- 6. Conclusão**
- 7. Bibliografia**

## Capítulo 1. Introdução.

Desde o início da década de 1970 o Instituto Brasileiro de Geografia e Estatística divulga, mensalmente, um índice de produção industrial. Este índice é obtido através da Pesquisa Industrial Mensal-Produção Física. A variável a ser levantada na pesquisa é a produção física de um determinado grupo de bens, para que assim possa, aproximadamente, chegar a uma evolução de curto prazo do valor adicionado da indústria. Portanto, a geração desse indicador tem como medida primordial fornecer variações de períodos curtos de tempo do produto real industrial.

O acompanhamento deste índice é de grande importância para o cenário macroeconômico doméstico, uma vez que movimentos no nível de atividade têm impacto sobre as decisões de política econômica, que por sua vez afeta todos os agentes da economia. Se acompanhá-lo é necessário, melhor ainda antecipá-lo. Essa antecipação pode ser feita através de modelos econométricos, com base em observações anteriores da própria série e de outras.

Este estudo propõe através de três modelos diferentes a projeção para a série em questão. Os modelos são simples e podem vir apresentar características semelhantes. Irei me concentrar em um integrado auto-regressivo e de médias móveis (ARIMA), um vetorial auto-regressivo (VAR) e auto-regressivo de defasagens distribuídas (ADL).

O modelo ARIMA(p,d,q) pode ser aplicado a qualquer série temporal não-estacionária homogênea. A dificuldade de tal processo se encontra na escolha dos valores de p, d e q, ou seja, na especificação do modelo ARIMA.

O VAR nada mais é do que a forma reduzida de um modelo estrutural onde todas as variáveis são endógenas, e só dependem de suas defasagens e das defasagens das demais.

O ADL é formado pelas próprias defasagens da variável, bem como os valores contemporâneos e defasados de variáveis exógenas. Diversos modelos utilizados com frequência são casos particulares de um ADL, como por exemplo, uma auto-regressão.

Através deste arcabouço, esclarecendo as peculiaridades de cada um, criei as devidas previsões para o ano de 2006 e assim irei compará-las com a série original, divulgada pelo IBGE. Entretanto, o objetivo central deste trabalho não é apenas atingir qual dos modelos obterá melhor resultado, e observar as dificuldades e os caminhos para evitá-las a fim de obter uma boa estimativa.

## Capítulo 2. O índice de Produção Industrial.

### 2.1 Como é elaborada a pesquisa e de que maneira o índice é construído

A Pesquisa Industrial Mensal de Produção Física (PIM-PF) é realizada pelo IBGE com a finalidade de construir índices capazes de representar o comportamento das industriais, tanto para o Brasil como para alguns estados específicos. Por se tratar de uma informação mensal, a divulgação dos indicadores possui certa defasagem em relação ao mês de referência, não mais do que seis semanas.

Entender como é elaborada a PIM-PF significa compreender alguns processos utilizados na pesquisa, dentre eles as unidades de investigação, os níveis geográficos, os detalhamentos de atividade para os quais são calculados os indicadores e, posteriormente, os processos de seleção de produtos e informantes da PIM-PF, bem como o questionário eletrônico utilizado na captação das informações.

As pesquisas industriais são direcionadas as empresas que possam ser caracterizada como industriais, não obstante de uma pequena parte das atividades industriais serem processadas em empresas não-industriais. As unidades de investigação são, atualmente, as próprias empresas e o que é chamado de unidade local, um espaço físico onde é praticada uma ou mais atividades econômicas, que corresponde à localização onde a empresa atua ou ao sufixo do Cadastro Nacional de Pessoa Jurídica (CNPJ).

As amostras regionais e nacionais da PIM-PF são selecionadas de forma independente, ou seja, existem diferentes considerações para cada uma. Logo, existem produtos que são objetos de uma delas, mas que podem não estar inclusos na outra. Como a série que fará parte do estudo é a de caráter nacional, me concentrarei apenas nesta.

Os produtos que fazem parte da amostra da PIM-PF são escolhidos por meio de seleção intencional, onde é utilizada como referência a variável Valor Bruto da Produção Industrial-VBPI, de modo a obter amostras representativas de, no mínimo, 50% do conjunto das atividades selecionadas. A seleção dos produtos é dada de forma que são escolhidos



apenas os produtos que são responsáveis por no mínimo 80% do VBPI. Para apresentar de forma clara o VBPI indica que o valor da produção é obtido pela multiplicação de um valor unitário médio das vendas do produto por sua produção, conforme especificação a seguir.

$$VBPI_{ik} = \left( \frac{VV}{QV} \right)_{ik} * QP_{ik} ,$$

onde:

$VV_{ik}$  é o valor de vendas do produto  $i$  na unidade local  $k$ ;

$QV_{ik}$  é a quantidade vendida do produto  $i$  na unidade local  $k$ ; e

$QP_{ik}$  denota a quantidade produzida do produto  $i$  na unidade local  $k$ .

A maneira como são construídos os indicadores é baseada em um sistema de ponderações, evidenciando a estrutura da produção industrial, elaborada a partir da importância relativa de cada produto no VTI, valor da transformação industrial, do triênio de 1998-2000. Assim, o peso do produto na Indústria Geral ( $w$ ) é dado pela relação a seguir:

$$w_{il} = \frac{\tilde{VTI}_{il}}{VTI_l}$$

onde:

$w_{il}$  denota o peso associado ao produto  $i$  no nível geográfico  $l$ ;

$VTI_l$  é o valor da transformação industrial da Indústria Geral no nível geográfico  $l$ ; e

$\tilde{VTI}_{il}$  é o valor da transformação industrial associado ao produto  $i$  no nível geográfico  $l$ .

Como, por definição, não existe VTI por produto o,  $\frac{\tilde{VTI}_{il}}$  neste caso é um cálculo derivado.

Os cálculo dos índices, nos diversos níveis de agregação, são baseados no método de Laspeyres, definido, para o caso das quantidades, como a razão entre o valor da produção

no período t e o mesmo valor no período 0 (zero), que serve de base ao índice, ambos valorados aos preços do período base.

Analicamente, o índice (L) é dado por:

$$L_{o,t} = \frac{\sum_{i=1}^n p_{io} q_{it}}{\sum_{i=1}^n p_{io} q_{io}}$$

onde:

$p_{io}$  é o preço do produto i no período 0 (período base: 1998/2000);

$q_{io}$  é a quantidade do produto i no período 0; e

$q_{it}$  é a quantidade do produto i no período t.

Como para que seja possível alcançar os objetivos dos índices de produção física é necessário expressar a ponderação em função do VTI de cada produto e o VTI da industrial geral. Através de alguns artifícios matemáticos podemos chegar a tal, como descrito abaixo.

$$L_{o,t} = \sum_{i=1}^n \frac{p_{io} q_{io}}{\sum_{i=1}^n p_{io} q_{io}} * \frac{q_{it}}{q_{io}}$$

Na expressão acima, que chamaremos de (1) para facilitar, nada mais é do que a anterior multiplicada a anterior pela razão  $\frac{q_{io}}{q_{io}}$  rearranjando (1), e considerando que:

$$w_{io} = \frac{p_{io} q_{io}}{\sum_{i=1}^n p_{io} q_{io}}$$

. Esta chamaremos de (2), que por sua vez é muito similar a expressão que trata do peso do produto i no período 0. Considere, também, que:

$r_{it} = \frac{q_{it}}{q_{i0}}$ . A expressão ao lado (3), trata da razão entre as quantidades produzidas no

período t e 0. É válido observar que:  $\sum_{i=1}^n w_{i0} = 1$ .

Pela expressão (1), o índice é uma média ponderada de relativos de quantidades, calculado nos pontos extremos da série, ou seja, nos períodos 0 e t.

O Índice Mensal de base fixa, que também é conhecido como o número-índice de *quantum* é representado por:

$$I_{ot} = \sum_{i=1}^n w_{i0} r_{it}$$

É essencial ter a percepção de que o cálculo do índice, baseado nos relativos de quantidades, pressupõe painéis idênticos entre um período e outro. Porém,, os painéis, pela própria evolução da pesquisa, tendem a diferir ao longo do tempo. Logo, para contornar esse problema, o cálculo é feito em relativos em cadeia (r), considerando-se, para cada par de meses, os mesmos painéis. Dessa forma,

$$r_{it} = \frac{q_{i1}}{q_{i0}^*} * \frac{q_{i2}}{q_{i1}^*} * \dots * \frac{q_{it}}{q_{i,t-1}^*}, \text{ onde}$$

$q_{it}$  é a quantidade produzida, do produto i, no mês t, na interseção dos painéis dos meses t-1 e t; e

$q_{i,t-1}^*$  é a quantidade produzida, do produto i, no mês t - 1, na interseção dos painéis dos meses t - 1 e t.

Assim torna-se possível uma comparação entre cada relativo mensal, uma vez que passa a existir um controle mensal dos painéis de produtos e informantes.

Além do número-índice, o sistema gera resultados denominados mensais, quando são comparados quaisquer dois meses, e, ainda, os acumulados, quando as médias de dois ou mais meses são o objeto da comparação. O cálculo do índice mensal (mês t relativamente

ao mês anterior,  $t - 1$ ) é feito pela relação entre os dois números-índices correspondentes, isto é:

$$I_{t-1,t} = \frac{I_{ot}}{I_{ot-1}}$$

Logo, os índices mensais podem ser interpretados como médias Ponderadas de relativos de quantidades cujos pesos são definidos pelo valor de cada produto, estimado a partir das quantidades vigentes no mês de comparação ( $t - 1$ ) e dos preços do período-base. Em outras palavras, quando um produto apresenta uma mudança superior à média dos demais do mesmo gênero sua importância dentro do seu segmento industrial aumenta.

## 2.2: A importância de antecipar a produção industrial

O índice produção industrial tem como fundamental objetivo o acompanhamento de curto prazo do valor adicionado da indústria. Logo, possui forte correlação com o PIB industrial. Ao mudar a periodicidade da série de produção de mensal para trimestral, através da média, podemos observar tal correlação, como demonstrado pelo correlograma abaixo.

Sample: 1991Q1 2007Q1  
 Included observations: 65  
 Correlations are asymptotically consistent approximations

PIB,PRODIND(-i)	PIB,PRODIND(+i)	i	lag	lead
.  *****	.  *****	0	0.9845	0.9845
.  *****	.  *****	1	0.7135	0.6947
.  *****	.  *****	2	0.4536	0.4519
.  *****	.  *****	3	0.5094	0.5233
.  *****	.  *****	4	0.5771	0.5824
.  ****	.  ****	5	0.3790	0.3708
.  **	.  **	6	0.1801	0.1765
.  **	.  **	7	0.2352	0.2585
.  ***	.  ***	8	0.3250	0.3215
.  **	.  **	9	0.1794	0.1571

.   .	.   .	10	0.0355	0.0262
.   *	.   *	11	0.1160	0.1286
.   **	.   **	12	0.2038	0.2071
.   *	.   *	13	0.0947	0.0926
.   *	.   .	14	-0.0435	-0.0158
.   .	.   *	15	0.0143	0.0594
.   *	.   *	16	0.1057	0.1323

A correlação contemporânea evidenciada acima, de 0,98, é extremamente relevante. Evidentemente, ao prever a série em questão possuiremos clara noção sobre o PIB industrial de nossa economia. E, tendo em vista que tal possui peso de suma significância sobre o PIB total torna-se muito útil a antecipação em questão.

Existe ainda uma outra correlação relevante, a série objeto se relaciona fortemente com o índice de ações INDX. Este é composto por ações de empresas que atuam no setor industrial, e é fruto de uma parceria entre a Bolsa de Valores de São Paulo e a Fiesp. Sua carteira teórica é construída pelas ações das empresas que possuem maior representatividade na indústria, e são selecionadas dentre as que possuem maior liquidez na bolsa, e são ponderadas pelo valor ponderados das ações disponíveis à negociação<sup>1</sup>.

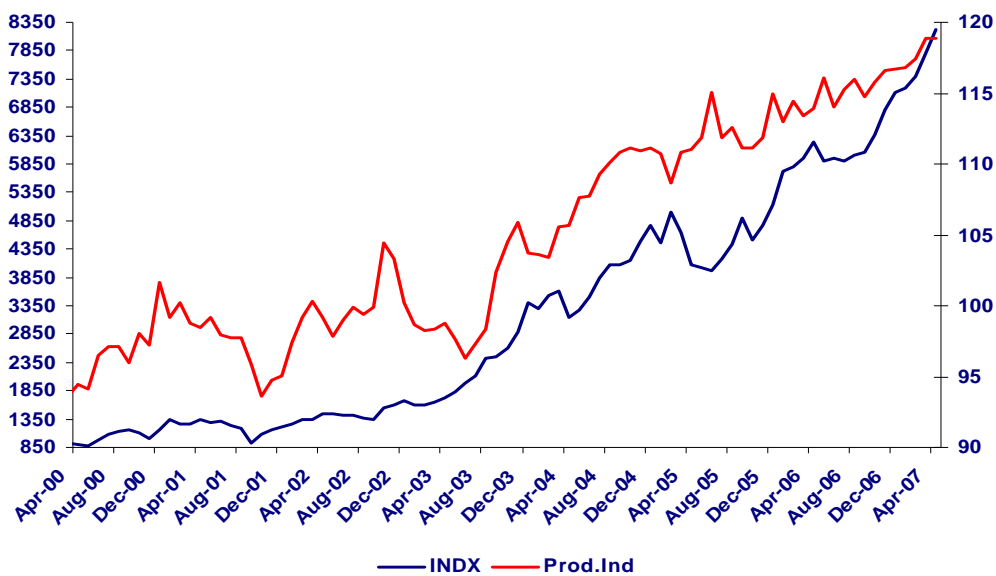
A correlação entre as séries é vista também através de um correlograma gerado pela produção industrial mensal dessazonalizada e o fechamento mensal do índice INDX, embora através de um gráfico é fácil enxergar tal propriedade.

---

1- Informações recolhidas junto ao site da Bolsa de valores de São Paulo, BOVESPA. Em [www.bovespa.com.br](http://www.bovespa.com.br)

Sample: 1991M01 2007M04  
 Included observations: 88  
 Correlations are asymptotically consistent approximations

INDX,PI_DESS(-i)	INDX,PI_DESS(+i)	i	lag	lead
.  *****	.  *****	0	0.9540	0.9540
.  *****	.  *****	1	0.8777	0.8600
.  *****	.  *****	2	0.8126	0.7771
.  *****	.  *****	3	0.7511	0.7104
.  *****	.  *****	4	0.6943	0.6553
.  *****	.  *****	5	0.6339	0.6007
.  *****	.  *****	6	0.5805	0.5499
.  *****	.  *****	7	0.5386	0.4975
.  *****	.  *****	8	0.5016	0.4521
.  *****	.  ****	9	0.4646	0.4152
.  ****	.  ****	10	0.4358	0.3850
.  ****	.  ****	11	0.3981	0.3572
.  ****	.  ***	12	0.3583	0.3392



A partir dos argumentos descritos, vê-se uma gama de motivos para que o índice de produção industrial seja antecipado. Assim partiremos, agora, para apresentação dos modelos que serão utilizados pra realizar diferentes previsões da série em questão.

## Capítulo 3: Os métodos a serem utilizados

### 3.1: ARIMA

Na prática muita das séries temporais com as quais nos deparamos são não-estacionárias, de modo que a característica do processo estocástico se altera ao decorrer do tempo. Porém, podemos fazer uso de modelos para séries que apresentem tal característica, transformando-as em estacionárias diferenciando-as uma ou mais vezes.

O modelo ARIMA (p, d, q) pode ser aplicado a qualquer série temporal não-estacionária homogênea. Quando a série observada não possuir a condição de estacionariedade, deve-se diferenciá-la quantas vezes for necessário para que se atinja tal propriedade. O número de vezes que a diferenciamos a fim de alcançar tal objetivo nos dará o parâmetro “d”. E caso já seja estacionária esse parâmetro será zero e, nos deparamos simplesmente, com um modelo ARMA (p, q), que pode ser apresentado por:

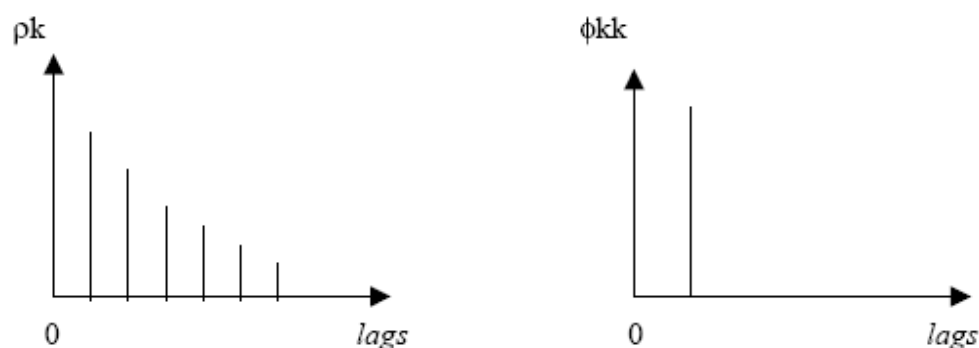
$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

A dificuldade de tal processo se encontra na escolha dos valores de p, d e q, ou seja, na especificação do modelo ARIMA. Primeiramente usaremos testes adequados a fim de descobrir se tal série é estacionária. Caso não seja, vê-se necessário defasá-la quantas vezes forem necessárias para que atinja tal propriedade, assim definimos o parâmetro “d”. Para especificar “d”, examinaremos primeiro a função de autocorrelação da série original, considere-a como  $y_t$ , para definir se ela é estacionária. Caso não seja, temos que diferenciá-la e examinar a função de autocorrelação para  $\Delta y_t$ . Repetindo o processo até que se alcance a estacionariedade, ou seja, até encontrar um “d” para o qual  $\Delta^d y_t$  seja estacionária.

Agora o problema se concentra na definição de “p” e “q”. No entanto, esse problema pode ser resolvido, em parte, através da análise de sua função de correlação e a função de autocorrelação parcial da série objeto.

A ordem autorregressiva  $p$  é determinada pela verificação da função de autocorrelação parcial (FACP)  $\phi_{kk}$  da série estudada. Se a série for unicamente autorregressiva ARIMA  $(p,d,0)$ , sua função de autocorrelação parcial sofrerá uma queda repentina após o lag  $k$ . Se não, efetua-se uma análise dos estimadores  $\phi_{kk}$  para verificar até que ordem de defasagem do correlograma desta função ele é estatisticamente significativa. Essa será sua ordem autorregressiva, “ $p$ ”.

Exemplo de uma FAC e FACP para um modelo AR(1):



Dessa forma, através da utilização do EViews, será rodado o modelo e os passos para tal serão apontados no capítulo 4.

### 3.2: VAR

O VAR nada mais é do que a forma reduzida de um modelo estrutural onde todas as variáveis são endógenas, e só dependem de suas defasagens e das defasagens das demais.

A análise dinâmica de sistemas de variáveis econômicas através de modelos VAR foi introduzida por Sims<sup>2</sup>.

---

2- As auto-regressões vetoriais foram introduzidas como uma alternativa aos modelos de equações múltiplas através do trabalho de C.A. Sims.



A abordagem VAR ao tratar todas as variáveis simetricamente sem impor qualquer restrição quanto à independência e dependência entre elas, permite descrever cada uma das variáveis endógenas no sistema como uma função dos valores defasados de todas as variáveis endógenas.

Sejam  $Y_1, \dots, Y_n$  as variáveis endógenas e  $X_1, \dots, X_m$  as variáveis exógenas. Uma autoregressão vetorial será dada por um conjunto de  $n$  equações lineares, entretanto, pode-se representar tal modelo de forma mais compacta usando notação matricial.

$$Y_t = A_0 + A_1 Y_{t-1} + \dots + A_p Y_{t-p} + B_1 X_{t-1} + \dots + B_r X_{t-r} + e_t, \quad (1)$$

onde,  $A_0$  é um vetor  $n \times 1$  de termos de intercepto,  $A_1, \dots, A_p$  são matrizes  $n \times n$  de coeficientes que relacionam valores defasados das variáveis endógenas a valores correntes dessas variáveis,  $B_0, \dots, B_r$  são matrizes  $n \times m$  de coeficientes que relacionam valores correntes e defasados das variáveis exógenas a valores correntes das endógenas, e  $e_t$  é um vetor  $n \times 1$  de termos de erro. Assim,  $p$  e  $r$  denotam o número de defasagens para as variáveis endógenas e exógenas, respectivamente. Esse modelo pode ser estimado por mínimos quadrados ordinários. Tendo em mente que apenas aparecem valores defasados de variáveis endógenas nos segundos membros das equações do VAR, a aplicação do método dos mínimos quadrados ordinários (MQO) na estimação de cada equação do sistema (1) vai produzir estimativas consistentes, mesmo que as perturbações  $e_t$  possam estar contemporaneamente correlacionadas, já que todos os regressores dos segundos membros são idênticos.

Ao escolher  $p$  e  $r$ , buscam-se defasagens suficientemente longas para capturar a dinâmica do sistema que está modelado. Por outro lado, quanto mais longas forem as defasagens, maior o número de parâmetros a serem estimados e menor o número de graus de liberdade. A partir de tal *trade-off*, vemos o ponto fraco das autoregressões vetoriais. Na prática, é necessário limitar o número de defasagens a uma quantidade menor do que seria ideal, dada a natureza dinâmica.

A determinação do número de defasagens pode ser facilitada pelo uso do  $R^2$  corrigido ou do critério de informação de Akaike (AIC). AIC é dado por:

$$AIC = \log \left( \frac{\sum \hat{e}_i^2}{N} \right) + (2k/N),$$

onde  $\Sigma$  é  $i^2$  é a soma dos quadrados dos resíduos.

Tanto o  $R^2$  corrigido quanto o AIC são medidas da qualidade do ajustamento que fazem uma correlação que leva em conta a perda de graus de liberdade decorrente da adição de defasagens adicionais ao modelo. A partir de tais critérios o modelo será rodado em um programa econométrico, para que assim as previsões possam ser realizadas, o que será apresentado nos próximos capítulos.

### 3.3: ADL

O modelo ADL (auto-regressivo com defasagens distribuídas) é formado pelas próprias defasagens da variável, bem como os valores contemporâneos e defasados de variáveis exógenas. Diversos modelos utilizados com frequência são casos particulares de um ADL, como por exemplo, uma auto-regressão.

Em modelos de série temporal pode passar um tempo substancial entre o período em que são tomadas decisões e o impacto dessas mudanças na variável de política econômica. Se o período entre decisão e resposta for suficientemente longo, variáveis explanatórias defasadas deveriam ser incluídas explicitamente no modelo.

Assim, mudanças econômicas podem ser distribuídas ao longo de uma série de períodos no tempo, chegamos à base do modelo com defasagens distribuídas. Nesse uma série de variáveis explanatórias defasadas responde pelo processo de ajustamento no tempo. De forma mais geral, tal modelo pode ser expresso como:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \dots = \alpha + \sum \beta_s X_{t-s} + \varepsilon_t$$

A não ser que explicitemos de forma diferente, segundo supomos, o erro possui distribuição normal, é independente de  $X$  e não apresenta correlação serial nem heterocedasticidade<sup>3</sup>.

---

3- Quando a variância condicional do erro dada uma variável, por exemplo,  $x$ , depende de  $x$ , o erro apresenta heterocedasticidade. De outra forma, quando  $\text{Var}(\varepsilon/x)$  depende de  $x$ ,  $\varepsilon$  apresenta tal propriedade.

Dada a especificação do modelo ADL a primeira pergunta que se apresenta é como estimar e testar hipóteses nestes modelos. Dado que o modelo ADL é de uma única equação esta simplificação pode implicar em perda de generalidade.

Mas esta simplificação, sob certas condições (exogeneidade fraca), não implica em perda de informação relevante. A questão prática que se coloca é relacionada a quais e quantas variáveis incluir no modelo e também quantas defasagens usar.

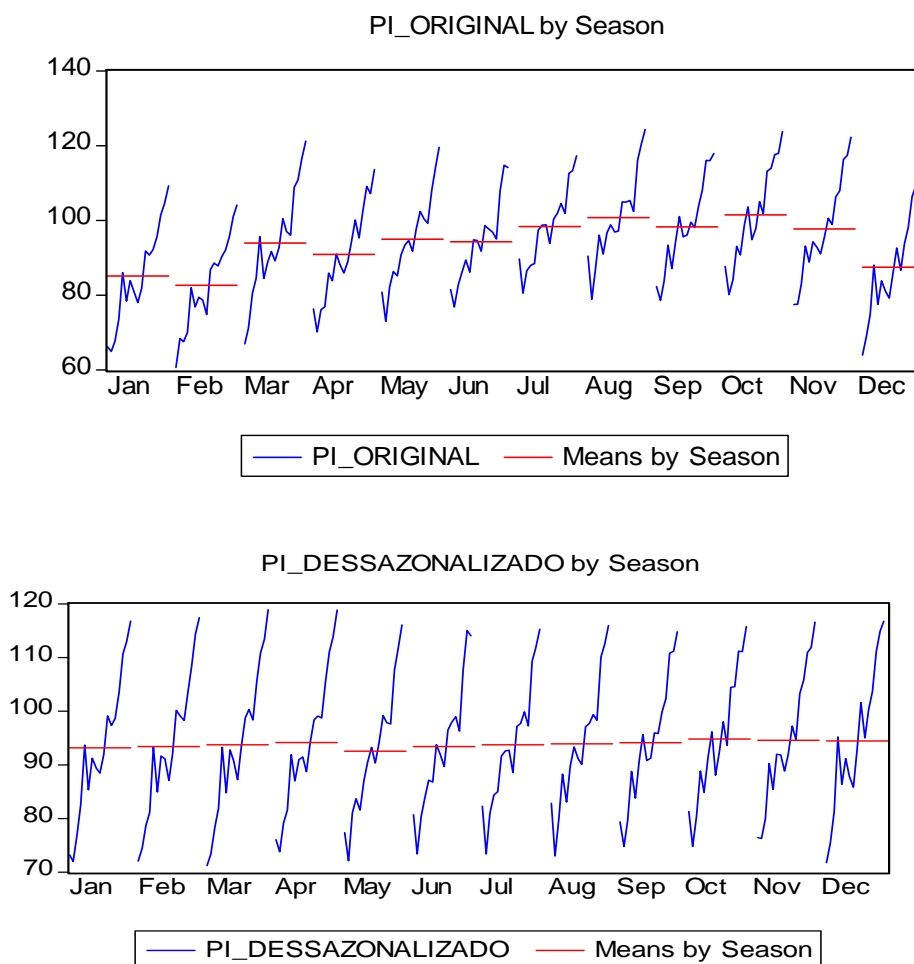
Um procedimento que era usado no passado era começar de um modelo simples, na maioria das vezes estático, e caso a parte não sistemática, isto é, os erros não fossem um processo ruído branco modificava-se o modelo ou através de correção por autocorrelação ou por inclusão de novas variáveis ou por inclusão de dinâmica, até que as hipóteses sobre a parte não sistemática fossem alcançadas. Este tipo de procedimento foi e é criticado uma vez que excluir informação relevante, ou por omissão de variáveis ou de dinâmica, implica em perda de eficiência e em algumas situações em perda de consistência dos estimadores dos parâmetros.

Agora passaremos a utilização do EViews, a fim de rodar os modelos e criar previsões consistentes para a série objeto.

## Capítulo 4: Rodando os modelos

Primeiramente, devemos esclarecer que, devido ao fato da série de produção industrial original possuir grande sazonalidade, usaremos a série dessazonalizada pelo próprio órgão responsável pelo índice, o IBGE.

Através dos gráficos abaixo<sup>4</sup> se vê de forma clara a diferença entre a série original e a que sofreu ajuste sazonal.



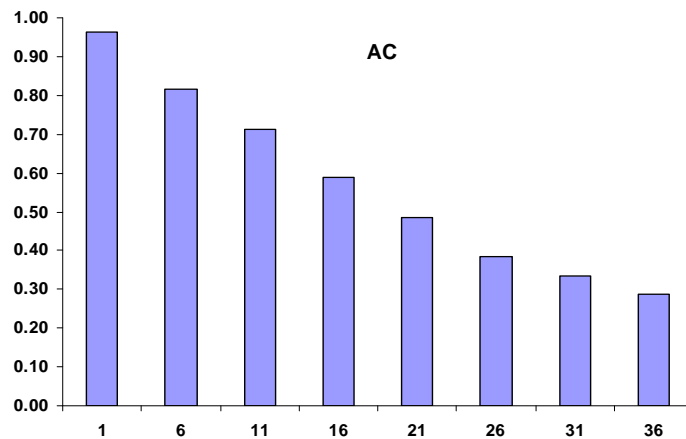
---

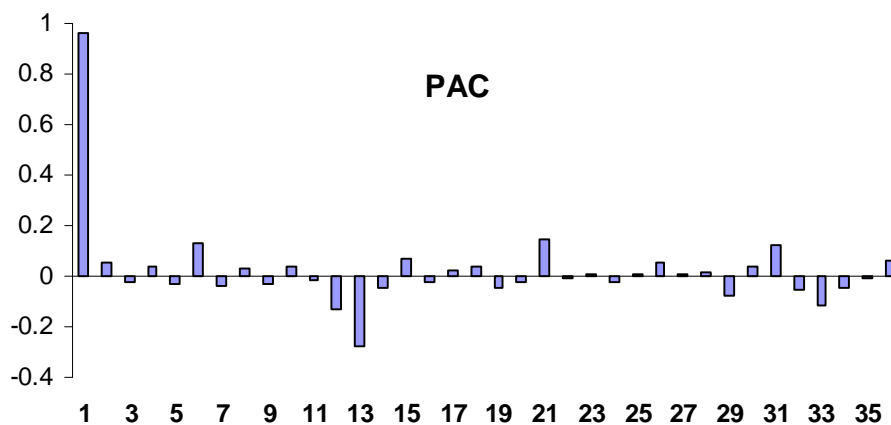
4 – Os dados a partir de dezembro de 2005 não serão incluídos nos modelos, entretanto, foram utilizados nos gráficos.

## 4.1: ARIMA

Para iniciar o processo de estimação pelo método autorregressivo integrado e de média móvel, analisaremos as funções de autocorrelação (AC) e autocorrelação parcial (PAC) da série de produção industrial com ajuste sazonal.

Sample: 1991M01 2005M12						
Included observations: 180						
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
*****	*****	1	0.962	0.962	169.36	0
*****	.	2	0.929	0.055	328.34	0
*****	.	3	0.896	-0.023	476.96	0
*****	.	4	0.867	0.035	616.78	0
*****	.	5	0.836	-0.033	747.47	0
*****	*	6	0.816	0.133	872.75	0
*****	.	7	0.793	-0.035	991.69	0
*****	.	8	0.773	0.034	1105.6	0
*****	.	9	0.751	-0.032	1213.6	0
*****	.	10	0.734	0.041	1317.3	0
*****	.	11	0.713	-0.014	1415.9	0
*****	*	12	0.686	-0.129	1507.6	0
*****	.	13	0.66	0.026	1593.2	0
*****	.	14	0.635	-0.03	1672.8	0
*****	.	15	0.612	0.029	1747.1	0
****	.	16	0.587	-0.05	1815.9	0
****	.	17	0.564	-0.012	1879.7	0
****	.	18	0.545	0.062	1939.7	0



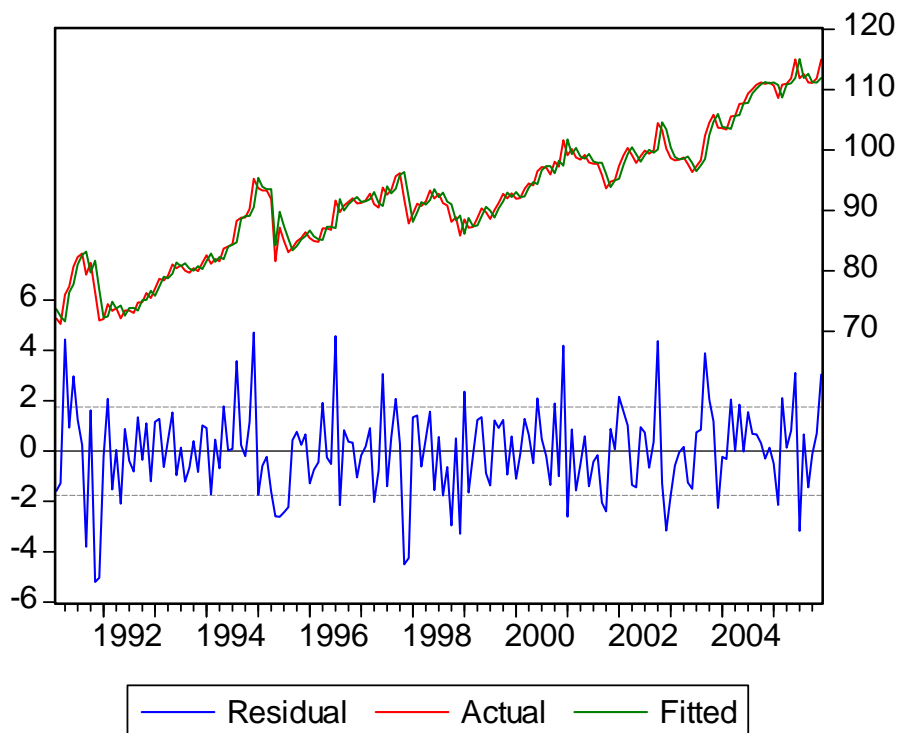


Após a análise das funções acima, foi decidido rodar um ARIMA (1,0,0), obtendo:

Dependent Variable: PI\_DESS  
 Method: Least Squares  
 Date: 06/26/07 Time: 03:39  
 Sample (adjusted): 1991M02 2005M12  
 Included observations: 179 after adjustments  
 Convergence achieved after 5 iterations

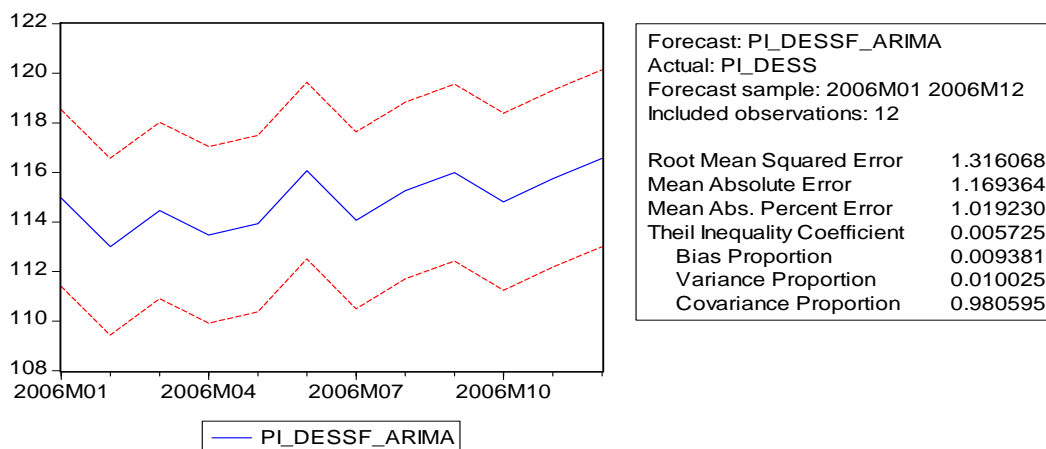
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	117.3090	37.62338	3.117982	0.0021
D9505	-7.941169	1.246208	-6.372265	0.0000
AR(1)	0.990867	0.012457	79.54525	0.0000
R-squared	0.973079	Mean dependent var		92.00101
Adjusted R-squared	0.972773	S.D. dependent var		10.63149
S.E. of regression	1.754253	Akaike info criterion		3.978581
Sum squared resid	541.6231	Schwarz criterion		4.032001
Log likelihood	-353.0830	F-statistic		3180.840
Durbin-Watson stat	1.984349	Prob(F-statistic)		0.000000
Inverted AR Roots	.99			

O gráfico da estimação, mostrado abaixo, evidencia a adequação do modelo.



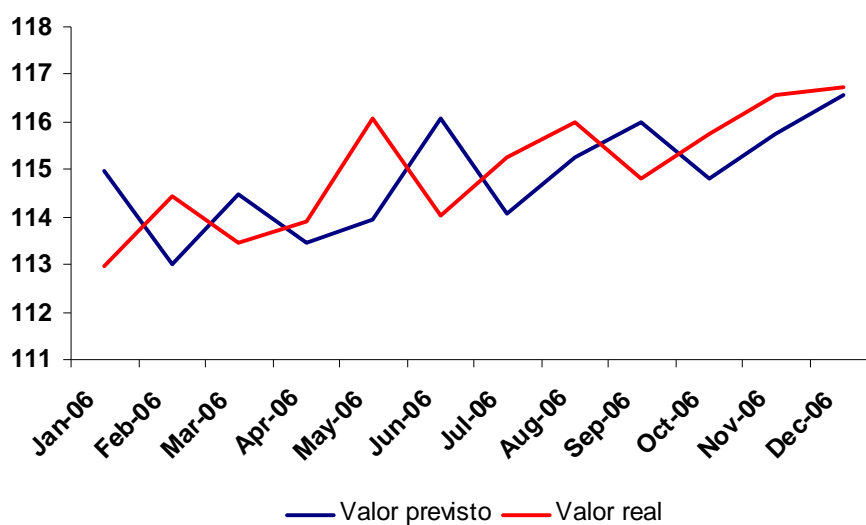
É importante citar que a introdução de uma dummy para o mês de maio de 1995 foi necessária.

Passaremos agora a etapa de previsão para os 12 meses de 2006. Ao pedir um “forecast”, obtemos o seguinte resultado:



A tabela e o gráfico abaixo evidenciam os valores previstos e os realmente observados entre janeiro e dezembro de 2006.

	Valor previsto	Valor real
jan-06	114.9616	112.97
fev-06	113.0096	114.44
mar-06	114.4662	113.44
abr-06	113.4753	113.9
mai-06	113.9311	116.06
jun-06	116.0714	114.04
jul-06	114.0699	115.25
ago-06	115.2688	115.98
set-06	115.9921	114.79
out-06	114.813	115.73
nov-06	115.7444	116.56
dez-06	116.5668	116.74
REQM	<b>0.86601</b>	




---

5- RMSE = Raiz do erro quadrático médio



## 4.2: VAR

No modelo do vetor auto-regressivo será incluída mais uma variável, a fim de prever a série de produção industrial. A variável a ser acrescentada será o nível de utilização da capacidade instalada, medido pela CNI. Os dados utilizados, assim como o índice de produção industrial, são dessazonalizados pela instituição responsável por sua criação e divulgação.

Primeiramente, devemos atentar para a escolha do número de defasagens a ser utilizado no mesmo. O critério escolhido para tomar tal decisão foi o Akaike (AIC). Através do software, faremos uso de uma ferramenta que disponibiliza quatro critérios de escolha. Os dados abaixo, e após escolher o método para definir o número de defasagens, nos leva à 6 diferenciações.

VAR Lag Order Selection Criteria  
 Endogenous variables: PI\_DESS NUCI  
 Exogenous variables: C  
 Date: 06/26/07 Time: 00:33  
 Sample: 1991M01 2005M12  
 Included observations: 132

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-654.2674	NA	71.35678	9.943446	9.987125	9.961195
1	-378.0887	539.8038	1.154639	5.819526	5.950562*	5.872773
2	-369.9467	15.66722	1.084465	5.756768	5.975162	5.845513*
3	-366.0838	7.316001	1.086858	5.758846	6.064598	5.883090
4	-364.3196	3.287920	1.124559	5.792721	6.185831	5.952463
5	-361.5797	5.023122	1.146638	5.811814	6.292281	6.007054
6	-353.1761	15.15197	1.073173*	<b>5.745092*</b>	6.312917	5.975830
7	-350.6329	4.508437	1.097882	5.767165	6.422347	6.033401
8	-349.9184	1.244997	1.154962	5.816945	6.559485	6.118679
9	-348.8196	1.881274	1.208252	5.860902	6.690800	6.198135
10	-345.0457	6.347006	1.214128	5.864328	6.781583	6.237059
11	-343.7709	2.105347	1.267504	5.905619	6.910232	6.313848
12	-339.8024	6.433773	1.270742	5.906097	6.998067	6.349823
13	-337.4210	3.788608	1.305482	5.930621	7.109949	6.409845
14	-334.7089	4.232505	1.335016	5.950134	7.216820	6.464857
15	-331.1802	5.399939	1.349032	5.957276	7.311319	6.507497
16	-329.8319	2.022403	1.409663	5.997454	7.438854	6.583173
17	-328.6299	1.766615	1.477058	6.039847	7.568606	6.661064
18	-322.9560	8.166917	1.447102	6.014486	7.630602	6.671201
19	-321.2871	2.351683	1.507361	6.049805	7.753279	6.742018
20	-320.6353	0.898655	1.595554	6.100535	7.891367	6.828247
21	-317.4950	4.234646	1.627550	6.113561	7.991750	6.876771
22	-316.2182	1.683046	1.709011	6.154822	8.120369	6.953530

23	-314.3317	2.429677	1.779471	6.186844	8.239748	7.021050
24	-311.5357	3.516152	1.829050	6.205086	8.345348	7.074791
25	-310.1721	1.673454	1.923022	6.245032	8.472652	7.150235
26	-306.0925	4.883167	1.942212	6.243826	8.558804	7.184527
27	-304.3508	2.032009	2.034455	6.278043	8.680378	7.254241
28	-303.3137	1.178485	2.156370	6.322936	8.812628	7.334632
29	-302.3134	1.106404	2.289608	6.368385	8.945436	7.415580
30	-296.3922	6.369848	2.259304	6.339275	9.003683	7.421968
31	-293.1725	3.365995	2.325744	6.351099	9.102864	7.469290
32	-285.2723	8.019949	2.233512	6.292004	9.131127	7.445693
33	-278.8172	6.357290	2.195882	6.254806	9.181286	7.443993
34	-274.2581	4.351808	2.225569	6.246335	9.260174	7.471021
35	-273.5471	0.657132	2.395433	6.296169	9.397365	7.556353
36	-261.3230	10.92765*	2.169790	6.171561	9.360114	7.467242

\* indicates lag order selected by the criterion

LR: sequential modified LR test statistic (each test at 5% level)

FPE: Final prediction error

AIC: Akaike information criterion

SC: Schwarz information criterion

HQ: Hannan-Quinn information criterion

Passaremos agora, a estimação do VAR.

Vector Autoregression Estimates

Date: 06/26/07 Time: 00:35

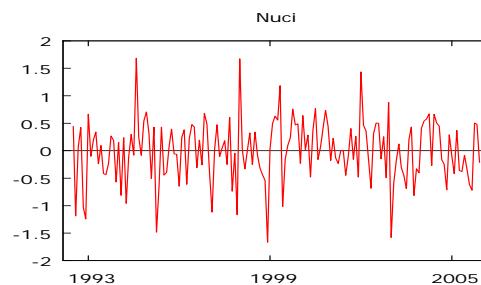
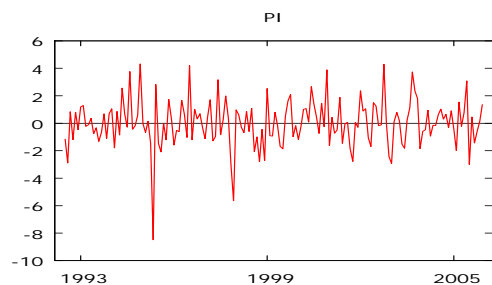
Sample (adjusted): 1992M07 2005M12

Included observations: 162 after adjustments

Standard errors in ( ) & t-statistics in [ ]

	PI_DESS	NUCI
PI_DESS(-1)	0.720122 (0.08696) [ 8.28073]	0.078812 (0.02850) [ 2.76526]
PI_DESS(-2)	0.132150 (0.10406) [ 1.26994]	-0.018347 (0.03410) [-0.53796]
PI_DESS(-3)	0.030422 (0.10515) [ 0.28933]	-0.034823 (0.03446) [-1.01053]
PI_DESS(-4)	0.053862 (0.10536) [ 0.51123]	0.042137 (0.03453) [ 1.22033]
PI_DESS(-5)	-0.185129 (0.10503) [-1.76269]	-0.043824 (0.03442) [-1.27319]
PI_DESS(-6)	0.294921 (0.08999) [ 3.27731]	-0.010217 (0.02949) [-0.34643]

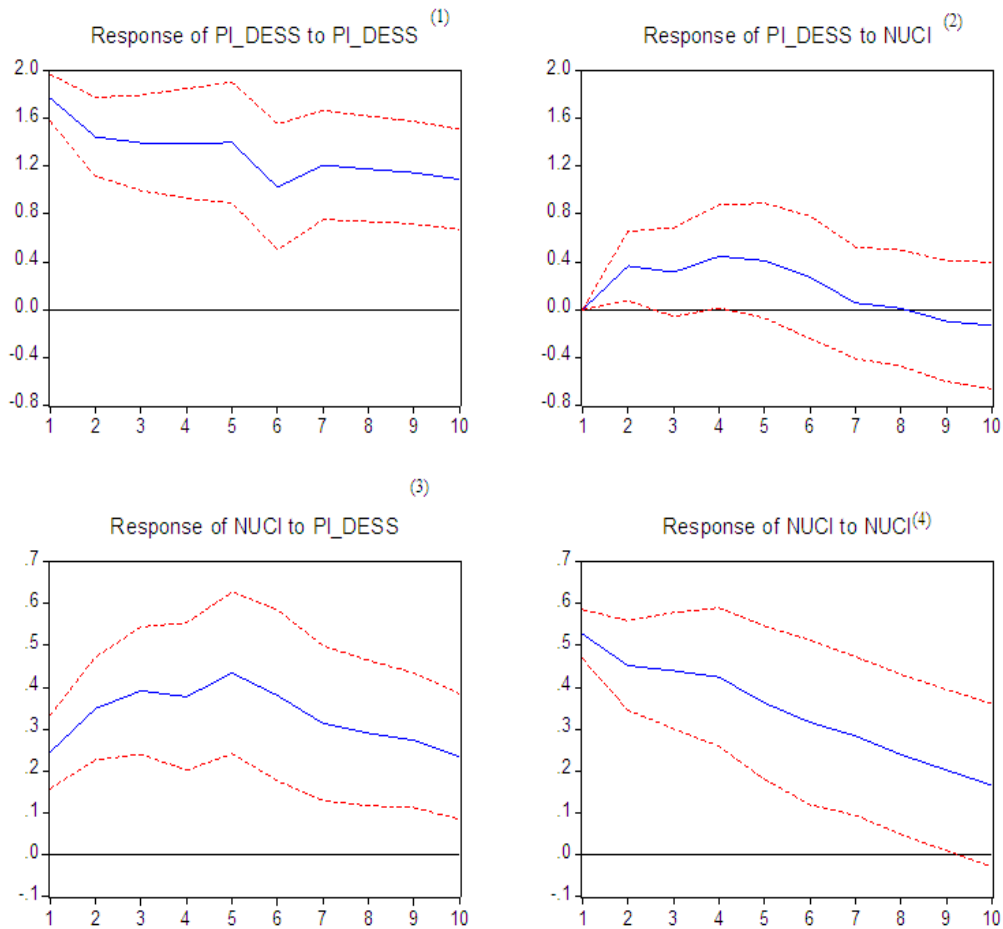
NUCI(-1)	0.691855 (0.27333) [ 2.53118]	0.857464 (0.08958) [ 9.57204]
NUCI(-2)	-0.496768 (0.34933) [-1.42206]	0.043919 (0.11449) [ 0.38361]
NUCI(-3)	0.174246 (0.34938) [ 0.49873]	0.018274 (0.11450) [ 0.15960]
NUCI(-4)	-0.222417 (0.33197) [-0.67000]	-0.086198 (0.10880) [-0.79228]
NUCI(-5)	-0.246000 (0.32334) [-0.76082]	-0.021345 (0.10597) [-0.20142]
NUCI(-6)	-0.110910 (0.25004) [-0.44358]	0.077325 (0.08195) [ 0.94361]
C	12.58300 (8.07097) [ 1.55905]	7.427270 (2.64512) [ 2.80791]
R-squared	0.968765	0.936908
Adj. R-squared	0.966250	0.931827
Sum sq. resids	468.9358	50.36792
S.E. equation	1.774041	0.581412
F-statistic	385.1113	184.3868
Log likelihood	-315.9605	-135.2404
Akaike AIC	4.061240	1.830129
Schwarz SC	4.309011	2.077899
Mean dependent	93.68691	78.73970
S.D. dependent	9.656619	2.226783
Determinant resid covariance (dof adj.)		0.875511
Determinant resid covariance		0.740635
Log likelihood		-435.4160
Akaike information criterion		5.696494
Schwarz criterion		6.192034



Após estimar um VAR e analisar os resíduos, é importante ser capaz de caracterizar nitidamente sua estrutura dinâmica. As funções de resposta a impulso fazem isso ao mostrar como um choque em uma variável se filtra através do modelo, afetando todas as demais variáveis, e eventualmente retroage sobre a própria variável original.

Analisaremos os gráficos abaixo com as respostas a impulso em cada situação. Destacaremos aqui os gráficos (2) que representa a resposta do nível de utilização da capacidade instalada a um choque de um desvio-padrão na produção industrial, e (3) que representa a resposta da produção industrial quando sujeita a um choque de um desvio-padrão no NUCI.

### Response to Cholesky One S.D. Innovations $\pm 2$ S.E.

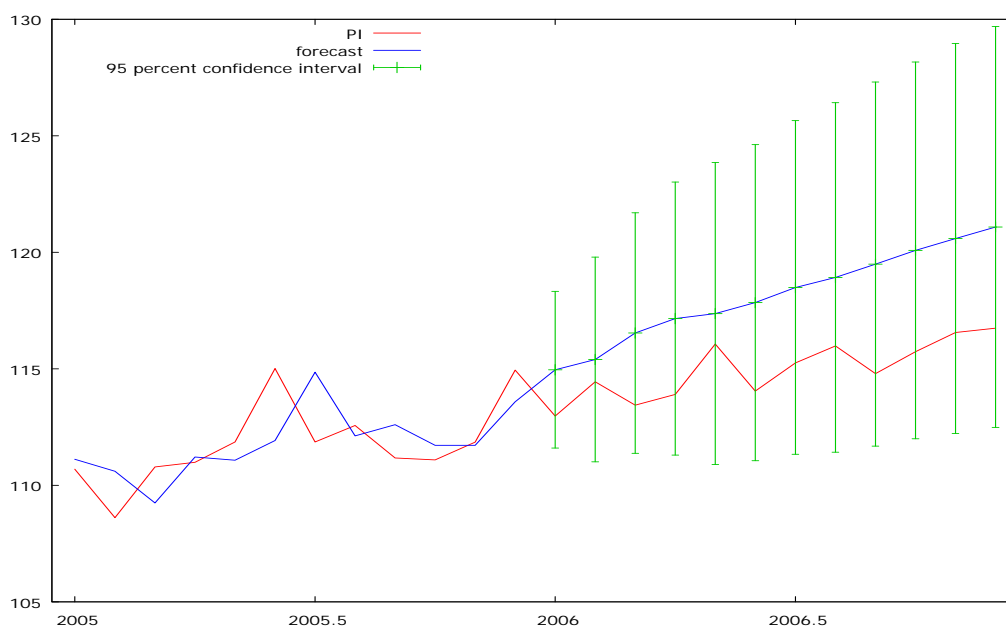


Existe ainda, um outro método de caracterizar o comportamento dinâmico do modelo, através da decomposição da variância. Esse decompõe a variância do erro da previsão para cada variável em componentes que podem ser atribuídos a cada uma das variáveis endógenas. Entretanto, este não será tratado neste trabalho.

Assim, avançaremos para a fase onde serão gerados valores para a série de produção industrial no período de janeiro a dezembro de 2006. Ao rodarmos as previsões para o período desejado obtemos a tabela abaixo.

For 95% confidence intervals,  $t(149, .025) = 1.976$

Obs	PI	prediction	std. error	95% confidence interval
2006:01	112.970	114.958	1.70137	(111.596, 118.320)
2006:02	114.440	115.399	2.22314	(111.006, 119.792)
2006:03	113.440	116.534	2.61284	(111.371, 121.697)
2006:04	113.900	117.155	2.96468	(111.296, 123.013)
2006:05	116.060	117.370	3.27782	(110.893, 123.847)
2006:06	114.040	117.839	3.43231	(111.057, 124.621)
2006:07	115.250	118.488	3.62355	(111.328, 125.648)
2006:08	115.980	118.921	3.79544	(111.421, 126.420)
2006:09	114.790	119.493	3.95273	(111.682, 127.303)
2006:10	115.730	120.082	4.09083	(111.998, 128.166)
2006:11	116.560	120.593	4.23372	(112.227, 128.959)
2006:12	116.740	121.085	4.35308	(112.484, 129.687)
RSME	<b>5.6885</b>			



### 4.3: ADL

Neste modelo será rodada uma regressão onde a produção industrial é explicada por suas defasagens, pelos valores contemporâneos e defasados da NUCI, além de um distúrbio aleatório.

O número de defasagens foi escolhido através do método das tentativas, ou seja, iniciando com uma quantidade “grande” de valores defasados de ambas, de modo que a regressão estivesse adequada, com os resíduos próximos da estacionariedade. A partir daí foram sendo excluídas as defasagens de forma igual, ou seja, cada uma das variáveis continha o mesmo número, até que a regressão atingisse a iminência da perda de suas características relevantes.

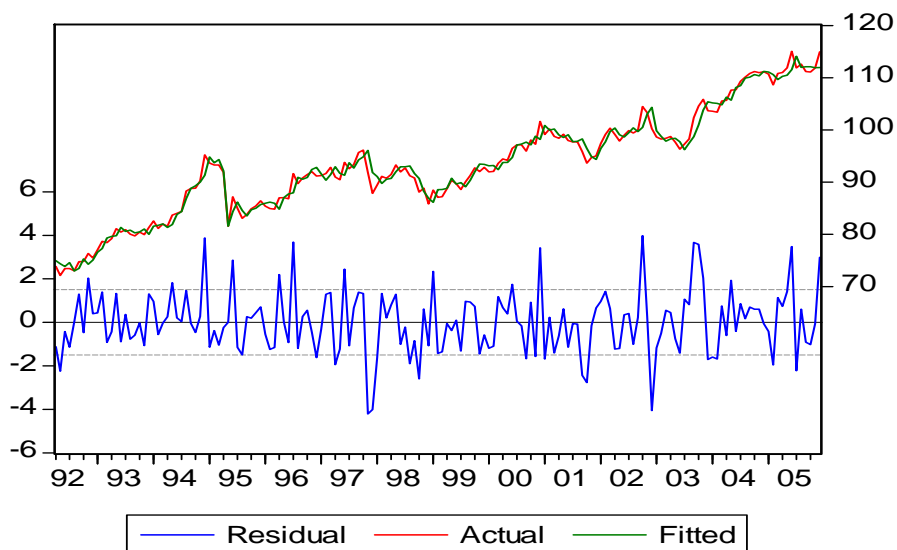
É válido ressaltar que a introdução de uma dummy para o mês de maio do ano de 1995 foi necessária, a fim de solucionar um problema na série objeto deste estudo.

Assim, foi atingida tal regressão:

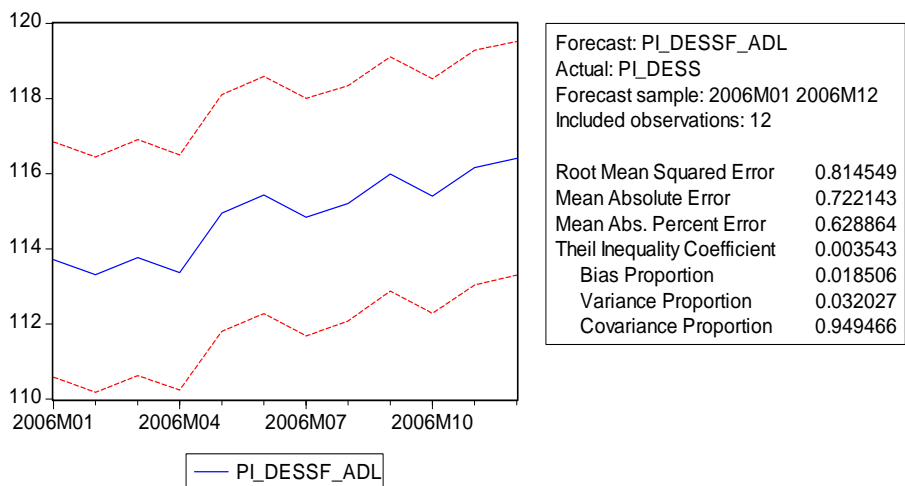
Dependent Variable: PI\_DESS  
Method: Least Squares  
Date: 06/26/07 Time: 03:05  
Sample (adjusted): 1992M04 2005M12  
Included observations: 165 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
NUCI	1.104253	0.208419	5.298225	0.0000
C	-5.424413	6.204110	-0.874326	0.3833
NUCI(-1)	-0.659829	0.273456	-2.412924	0.0170
PI_DESS(-1)	0.671660	0.073423	9.147827	0.0000
PI_DESS(-2)	0.177799	0.086330	2.059517	0.0411
PI_DESS(-3)	0.132759	0.073649	1.802603	0.0734
NUCI(-2)	-0.099929	0.272610	-0.366563	0.7144
NUCI(-3)	-0.250456	0.207214	-1.208683	0.2286
D9505	-9.712268	1.552329	-6.256580	0.0000
R-squared	0.978323	Mean dependent var		93.31321
Adjusted R-squared	0.977212	S.D. dependent var		9.956970
S.E. of regression	1.503082	Akaike info criterion		3.705914
Sum squared resid	352.4438	Schwarz criterion		3.875329
Log likelihood	-296.7379	F-statistic		880.0865
Durbin-Watson stat	1.833129	Prob(F-statistic)		0.000000

Os resíduos da regressão apresentam-se, bem comportados, conforme o gráfico abaixo.



Passemos agora para a previsão do índice de produção industrial pelo último método, ADL. Ao prever a série durante os 12 meses de 2006 encontramos o seguinte resultado, evidenciado no gráfico e na tabela a seguir.





	<b>Valor previsto</b>	<b>Valor real</b>
<b>jan-06</b>	113.7141	112.97
<b>fev-06</b>	113.3121	114.44
<b>mar-06</b>	113.7674	113.44
<b>abr-06</b>	113.3689	113.9
<b>mai-06</b>	114.9525	116.06
<b>jun-06</b>	115.4326	114.04
<b>jul-06</b>	114.8444	115.25
<b>ago-06</b>	115.211	115.98
<b>set-06</b>	115.9938	114.79
<b>out-06</b>	115.4048	115.73
<b>nov-06</b>	116.1594	116.56
<b>dez-06</b>	116.4092	116.74
<b>RSME</b>	<b>0.33174</b>	

## Capítulo 5: Comparando os resultados

A partir da estimação dos três modelos, podemos, em fim, compará-los. Por mais que seja estabelecido um critério para a escolha do que foi capaz de prever com maior sucesso a série de produção industrial, esse não é o único objetivo do trabalho em questão.

Em quase todos os problemas de previsão, existem vários métodos competitivos para projetar uma série. E existem diversos problemas a serem sanados, como quais variáveis usar, quantas diferenças, dentre outros.

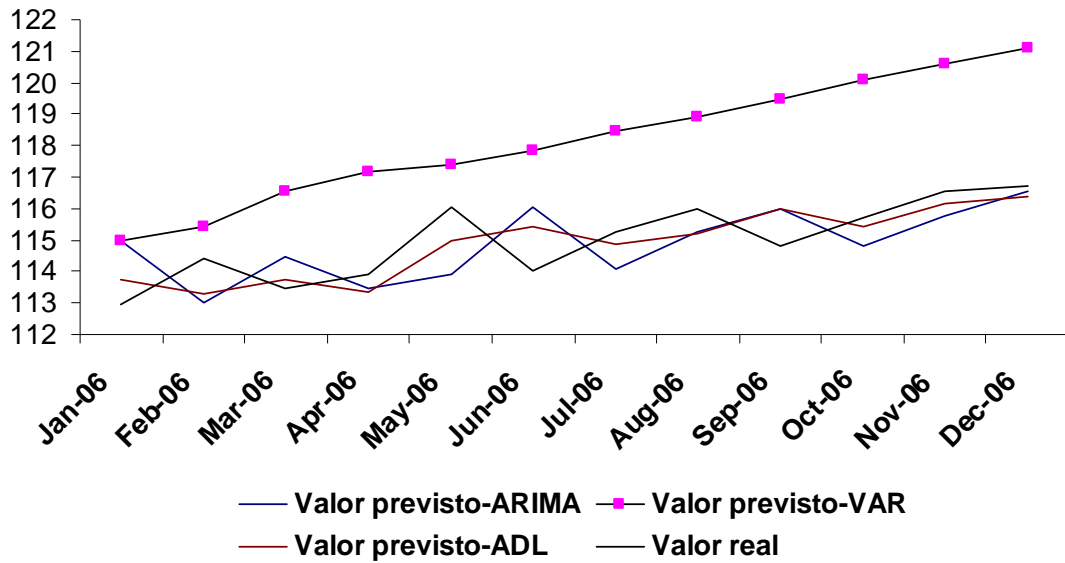
O critério a ser usado para avaliar a capacidade preditiva de cada modelo será a primeira raiz do erro quadrático médio, REQM (ou RSME). Este método, aplicado ao requerido número de observações a serem projetadas, se dá da seguinte maneira:

$$RSME = [1/12 \cdot \sum (\hat{Y}_{2006:01} - Y_{2006:01})^2 + \dots + (\hat{Y}_{2006:12} - Y_{2006:12})^2]^{1/2}$$

Isso é essencialmente o desvio-padrão amostral dos erros de previsão, sem ajuste dos graus de liberdade. Se for computada a raiz erro quadrático médio de métodos distintos de previsão, é preferível aquele que apresenta o menor RSME.

Em nossos três modelos, ao apresentar os valores das previsões também foi exibido o RSME de cada um. A tabela abaixo contém todas as estimativas, que estarão em um gráfico logo a seguir, bem como os RSME de todos os modelos rodados para que possamos compará-los utilizando o critério estabelecido.

Valor previsto- ARIMA	Valor previsto- VAR	Valor previsto- ADL	Valor real	Data
114.9616	114.958	113.7141	112.97	Jan-06
113.0096	115.399	113.3121	114.44	Feb-06
114.4662	116.534	113.7674	113.44	Mar-06
113.4753	117.155	113.3689	113.9	Apr-06
113.9311	117.37	114.9525	116.06	May-06
116.0714	117.839	115.4326	114.04	Jun-06
114.0699	118.488	114.8444	115.25	Jul-06
115.2688	118.921	115.211	115.98	Aug-06
115.9921	119.493	115.9938	114.79	Sep-06
114.813	120.082	115.4048	115.73	Oct-06
115.7444	120.593	116.1594	116.56	Nov-06
116.5668	121.085	116.4092	116.74	Dec-06
<b>0.86601</b>	<b>5.68852</b>	<b>0.33174</b>		<b>RSME</b>



O modelo que foi capaz de prever melhor a série do IBGE foi o modelo ADL, seguido do ARIMA. O VAR ficou bem distante dos demais, o uso da variável NUCI pode não ter sido ideal. Além do mais poderiam ter sido incluídas mais variáveis no modelo. E ainda, os outros modelos são semelhantes. Como já foi dito o processo auto-regressivo, nada mais é do que um caso particular do modelo ADL, basta impor uma restrição que zere os coeficientes da variável dependente. Entretanto, vale ressaltar que o objetivo central do estudo não foi a escolha do melhor modelo e sim a análise de cada um.

## Capítulo 6: Conclusão

Ao longo desse trabalho, construímos três métodos capazes de prever o comportamento da série de produção industrial. O primeiro através de um processo integrado auto-regressivo e de médias móveis, o segundo um vetor auto-regressivo, e o terceiro e último, um auto-regressivo com defasagens distribuídas. Mostramos a dificuldade na estimação deles bem como as saídas alternativas para tais problemas. A partir da análise, e descrição de cada um dos modelos foi possível uma melhor percepção sobre a geração de valores futuros, bem como tornou clara a importância da série de produção industrial.

Quando comparados os modelos se distanciam um pouco, o modelo do vetor auto-regressivo sofreu diversos problemas, o que incorreu em uma raiz do erro quadrático médio muito elevada, o que o levou ao último lugar segundo o critério de avaliação escolhido, o de Akaike. Nem por isso deixou de ser proveitosa a estimação através do VAR, foi importantes mostrar claramente os passos para sua estimação.

Em fim, é importante deixar claro que este trabalho não teve como objetivo a simples e única comparação entre os métodos descritos. Mas sim apresentá-los passo a passo, para que esses sejam esclarecidos.

## Referências Bibliográficas:

PINDYCK, Robert S. e Rubinfeld, Daniel L. Econometria (Elsevier Editora Ltda, 2004, tradução da quarta edição).

Wooldridge, Jeffrey M., Introdução à Econometria: uma abordagem moderna (Thomson Learning Ltda, 2006).

Instituto Brasileiro de Geografia e Estatística. Pesquisa Industrial Mensal Produção Física – Brasil Disponível em: <<http://www.ibge.gov.br/>>.

IPEADData. Disponível em: <<http://www.ipeadata.gov.br>>

BOVESPA. Disponível em: <<http://www.bovespa.com.br>>