

## 4

### A Base de Dados

#### 4.1

##### A variável dependente

O candidato natural a ser previsto seria o PIB. No entanto, devido à sua periodicidade trimestral, optou-se por utilizar o ciclo da produção industrial. Apesar da crescente participação do setor terciário no PIB do Brasil, seguiu-se a orientação de Nilsson (2000)<sup>5</sup>, na qual afirma que o ciclo da produção industrial apresenta um comportamento semelhante ao dos ciclos econômicos.

Segundo Mankiw (2003, pág. 246), o Produto Potencial ou a tendência histórica pode ser definida como “o nível do produto no qual os recursos existentes numa economia estão empregues na sua totalidade ou, mais realisticamente, no qual o desemprego é igual a sua taxa natural”. Assim, será utilizado neste trabalho o Hiato do Produto Industrial. A metodologia para seu cálculo se encontra abaixo:

$$Hiato\_Ind\_Geral_t = \frac{(P\_Ind_t - P\_Pot_t)}{P\_Pot_t} \quad (33)$$

onde  $P\_Ind$  é a Produção Industrial (*quantum*) da Indústria Geral<sup>6</sup> e  $P\_Pot$  é o Produto Potencial da Produção Industrial, obtido via extração da tendência pelo Filtro Hodrick-Prescott<sup>7</sup>. A figura a seguir apresenta o gráfico da série mensal do Hiato do Produto Industrial estimado para o Brasil no período de Agosto de 1994 a Março de 2009.

---

<sup>5</sup> Autores como Zarnowitz, Ozyildirim e McGuckin (2003) argumentam que a produção do segundo setor da economia já foi um bom indicador do PIB. Para eles, atualmente, o setor terciário representa uma importante parcela, crescendo a cada ano.

<sup>6</sup> Produção Industrial - Indústria Geral (*quantum*) - índice (média 2002 = 100), ajustado sazonalmente. Fonte: IBGE.

<sup>7</sup> Foi utilizado o parâmetro  $\lambda = 14.400$ , como sugerem Holdrick e Prescott (1997).

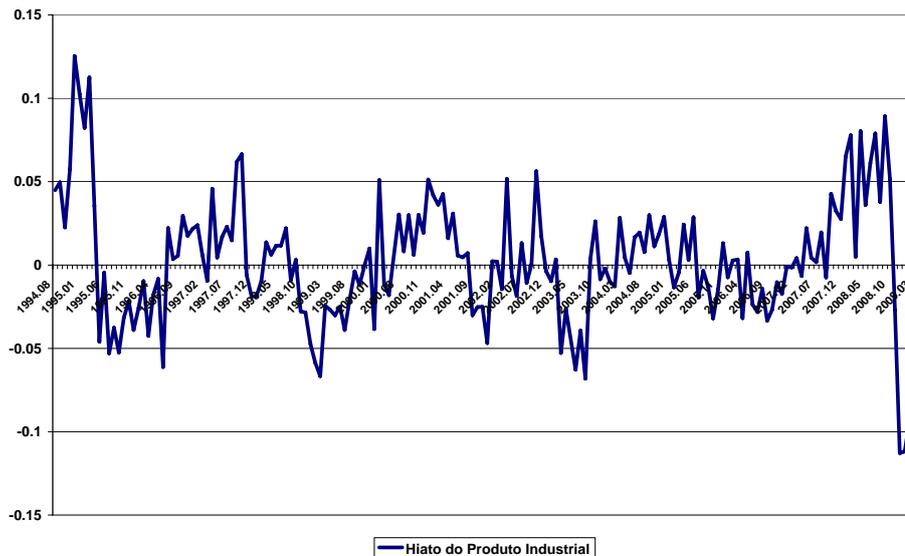


Figura 1: Hiato do Produto Industrial.

A Figura 1 mostra o comportamento do Hiato do Produto na amostra considerada. Esta será a série de referência a ser utilizada no Modelo de Índice de Difusão. Para efeito comparativo com o PIB<sup>8</sup>, transformou-se a série de referência em trimestral<sup>9</sup>. A Figura 2 mostra o gráfico destas séries.

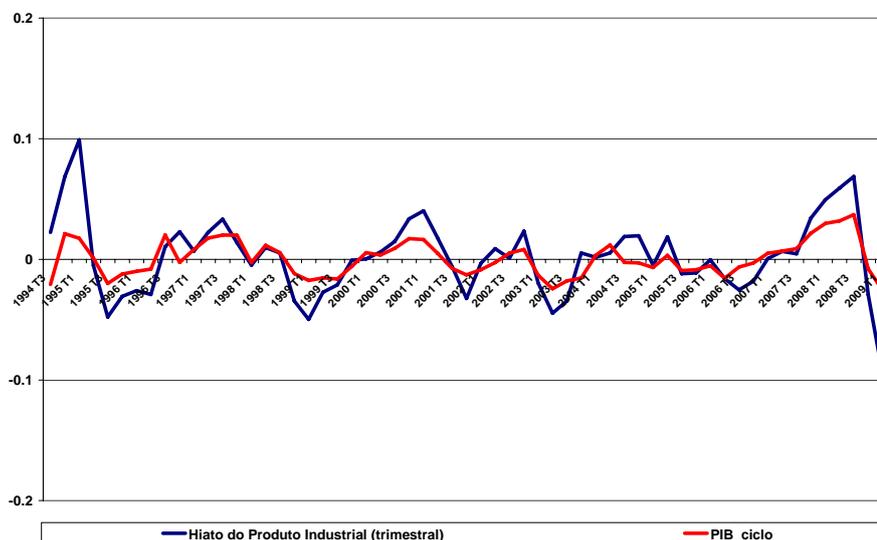


Figura 2: Hiato do Produto Industrial X Hiato do PIB.

<sup>8</sup> Produto Industrial Bruto a Preços de Mercado - Índice Encadeado (média 1995 = 100) - dessazonalizado. Fonte: IPEADATA.

<sup>9</sup> A mudança de periodicidade do Hiato do Produto Industrial consistiu na média trimestral histórica referente de cada período.

Apesar da maior volatilidade do Hiato do Produto Industrial, o comportamento de ambas as séries é bastante semelhante, tendo um comportamento cíclico similar. Com efeito, a correlação contemporânea é de 0,82, reforçando a significância estatística de seu uso para a modelagem.

## 4.2

### As variáveis selecionadas para o estudo

Em um estudo sobre ciclo econômico deve-se atentar à seleção das variáveis para abranger toda a conjuntura econômica. Neste trabalho, seguiu-se a divisão das séries históricas, *a priori*, por tema (ou seja, pelos setores da economia).

Assim, tendo como ponto de partida as séries propostas por Camba-Mendez *et. al.* (1999) e Stock e Watson (2002 a e b), partiu-se para a análise de 490 variáveis macroeconômicas no período de Agosto de 1994 a Março de 2009, configurando uma amostra de 176 observações.

Apesar da amostra não estar presente no período inflacionário, devido aos problemas inerentes às séries históricas brasileiras como dados faltantes e instabilidade financeira, decidiu-se, a princípio, após os critérios de seleção descritos abaixo, utilizar 74 séries históricas, listadas no Anexo A.

Critérios de seleção:

- Séries existentes em outros países e não verificadas no Brasil, como *Help Wanted Index* (Índice de Novos Empregos) ou *New Car Registrations* (Registro de Novos Carros).
- Séries que apresentavam grande instabilidade no início do Plano Real (início da amostra) e que, pela falta de séries de referência para aprendizado, não apresentaram resultados satisfatórios com o uso do Algoritmo EM. As séries financeiras de *Dividend Yield* e Razão Preço/Lucro das empresas que compõem o IBOVESPA, são exemplos.
- Utilizar séries de domínio público encontradas nos sites do IPEADATA, Banco Central e Fundação Getúlio Vargas.

A tentativa de selecionar variáveis que buscassem representar a conjuntura apresenta aquelas que representam o setor financeiro, industrial e macroeconômico. A seleção dos temas foi baseada em Camba-Mendez (1999), buscando segui-lo o mais fiel possível.

O caminho natural a ser seguido seria o processo de análise estatística das séries e modelagem econométrica. Entretanto, foram verificados problemas de informações faltantes na base de dados, necessitando de preenchimento. Logo, o uso do Algoritmo EM, especializado em imputação, apresentado na seção 3.3, fez-se necessário.

### 4.3

#### **O preenchimento dos dados faltantes nas variáveis selecionadas**

Como afirmado anteriormente, as séries históricas brasileiras não apresentam grande longevidade, resultando em muitas delas com periodicidade recente em relação à amostra estudada. Além disso, no período inflacionário, verificam-se, para a maioria delas, um comportamento volátil incomum quando comparada a uma economia estável<sup>10</sup>. Com isso, apesar de testada<sup>11</sup>, decidiu-se pelo uso da amostra de agosto de 1994 a março de 2009.

É importante frisar que o aprendizado foi realizado por temas do Anexo A, apesar de existirem algumas variáveis a serem preenchidas por outras presentes em outro tema. Esses são os casos de Spread, Índice de Otimismo e Confiança do Consumidor em que o primeiro utiliza as séries referentes à Taxa de Juros de Curto Prazo e, as duas últimas, a representante da Tendência da Produção além da Produção Industrial (indústria geral e bens de consumo duráveis).

Para alcançar o objetivo de se ter as séries completas com os dados próximos da realidade, foram realizadas exaustivas simulações. As séries que necessitaram o uso do algoritmo EM, as que serviram de referência para

---

<sup>10</sup> Essas foram as principais razões para a não utilização de uma amostra com maior amplitude.

<sup>11</sup> Inicialmente, a amostra utilizada era de Janeiro de 1985 a Março de 2009. No entanto, devido às razões apresentadas, o uso do algoritmo EM não apresentou bons resultados.

aprendizado na imputação, a amostra original de cada uma e as transformações realizadas quando necessário, encontram-se no Anexo B.

#### 4.4

#### **Tratamento estatístico das variáveis selecionadas**

Com a matriz de dados completa, buscou-se a transformação estatística das variáveis, quando necessária. A literatura sugere a estacionariedade e, em alguns casos específicos, algum tipo de transformação, seja ela logarítmica ou a simples diferenciação, dentre outras.

Para cada série referente à Taxa de Câmbio Efetiva Real, Termos de Troca, Tendência da Produção, Índice de Otimismo, Confiança do Consumidor, Residências em Construção/Autorizadas, Nível de Novos Pedidos e Horas Pagas e Trabalhadas foi realizada a transformação logarítmica, sendo que para as últimas três foi utilizado ainda o método de dessazonalização X-12 ARIMA Aditivo. A seguir, foi verificada a presença de raiz unitária em cada uma das séries aplicando os testes de Dickey-Fuller Aumentado e Phillips-Perron. Todas as séries que se mostraram integradas de ordem 1 foram diferenciadas.

O Anexo C mostra como ficaram as variáveis transformadas que serviram de base para o prosseguimento do estudo, com as transformações utilizadas quando necessário. No Anexo D pode-se verificar os testes de raiz unitária para cada uma das variáveis.