

PRODUÇÃO INDUSTRIAL NO BRASIL: UMA ANÁLISE DE DADOS EM TEMPO REAL

RAFAEL TIECHER CUSINATO *
ANDRÉ MINELLA †
SABINO DA SILVA PÔRTO JÚNIOR ‡

Resumo

A recente literatura sobre análise de dados em tempo real tem mostrado que diversas medidas de atividade econômica passam por importantes revisões de dados ao longo do tempo, implicando limitações relevantes para o uso dessas medidas. Organizamos um conjunto de dados de produção industrial (PI) em tempo real para o Brasil e avaliamos a extensão na qual as séries de crescimento e hiato da PI são revisadas. Encontramos revisões substanciais de crescimento mensal, da média móvel trimestral e do hiato da PI.

Palavras-chave: Dados em tempo real; Produção industrial; Hiato do produto; Ciclo econômico; Brasil.

Abstract

Recent literature on real-time data analysis has shown that several economic activity measures go through important revisions over time, limiting the use of those measures. We organize a real-time data set for industrial production (IP) in Brazil and assess the revisions of the series of IP growth and gap. We find that the revisions of month-over-month growth, three-month moving average growth and IP gap are substantial.

Keywords: Real-time data; Industrial production; Output gap; Business cycle; Brazil.

JEL classification: C82, E32

* Banco Central do Brasil. E-mail: rafael.cusinato@bcb.gov.br

† Banco Central do Brasil. E-mail: andre.minella@bcb.gov.br

‡ Universidade Federal do Rio Grande do Sul. E-mail: sabino@ppge.ufrgs.br

1 Introdução

A produção industrial (PI) é considerada uma das medidas mais importantes do nível de atividade econômica no Brasil. Ainda que o Produto Interno Bruto (PIB) seja a sua principal medida, a produção industrial apresenta um importante diferencial. Enquanto o PIB é uma medida trimestral, divulgada com uma defasagem superior a dois meses, a PI é mensal e é divulgada com uma defasagem um pouco superior a um mês. Além disso, o componente cíclico da PI é bem correlacionado com o ciclo econômico brasileiro. Assim, a PI é uma alternativa natural tanto para trabalhos de pesquisa que utilizam dados mensais quanto para análises efetuadas pelos agentes econômicos, que tomam decisões em tempo real e precisam obter informações recentes sobre o estado da economia.

Entretanto, assim como acontece com outras variáveis de atividade econômica, os dados de produção industrial estão sujeitos a revisões. Isso significa que, potencialmente, a primeira divulgação de um dado de PI pode levar os agentes econômicos a tomarem decisões diferentes do que tomariam se tivessem acesso a divulgações posteriores do mesmo dado. Essa possibilidade é potencialmente relevante se as revisões de PI forem de fato relativamente grandes. A existência de revisões de dados de magnitude elevada é uma preocupação central da recente literatura sobre dados em tempo real. Croushore & Stark (2000, 2001) elaboraram um conjunto de dados em tempo real para os Estados Unidos e encontraram revisões de crescimento de PIB relevantes para a economia norte-americana. Utilizando basicamente esses dados de PIB, Orphanides & Van Norden (2002) construíram diversas séries históricas em tempo real do hiato do produto e calcularam vários indicadores de revisão. Os autores encontraram revisões substanciais e sugeriram que as estimativas de hiato do produto em tempo real tendem a ser pouco confiáveis. McKenzie & Park (2006) analisaram a revisão de produção industrial para vários países e encontraram revisões relevantes. Para o Brasil, eles utilizaram as séries entre 2000:10 e 2004:11 do “OECD Main Economic Indicators Original release data and revisions database” e calcularam alguns indicadores de revisão, mas sem discutir separadamente o caso brasileiro.

Neste artigo, organizamos um conjunto de dados de produção industrial em tempo real para o Brasil. O conjunto de dados é formado por 127 séries de dados mensais, que contêm todas as divulgações de PI compreendidas no período entre 1998:4 e 2008:10. Os dados de cada série iniciam no primeiro mês de 1991. A partir desses dados, estudamos as revisões de dados de PI no Brasil. Em comparação a McKenzie & Park (2006), além de utilizarmos uma amostra maior, analisamos mais detalhadamente as revisões de PI no Brasil, incluindo vários indicadores adicionais. Além disso, analisamos as revisões de estimativas do hiato da produção industrial implicadas pelas séries de PI em tempo real, obtidas pela aplicação de três métodos de extração de tendência: filtro de Hodrick-Prescott, tendência linear e tendência quadrática. Assumindo que a nossa última série de dados (referente a 2008:10) é a melhor estimativa que possuímos, calculamos vários indicadores sobre as revisões de crescimento, aceleração e hiato da PI. Assim, dada a suposição de que as revisões melhoram as estimativas, parcela do erro de medida da PI e do hiato da PI em tempo real é corrigida por meio das revisões. No caso do hiato da produção industrial, decompomos as revisões em duas partes: (i) revisões do hiato provenientes da própria revisão dos dados da PI e (ii) revisões do hiato

provenientes da inclusão de novas observações de PI na amostra.

Analisando os indicadores de revisão, nosso objetivo é investigar a relevância das revisões da produção industrial e do hiato da PI. Mostramos que as revisões de crescimento da PI (mês/mês anterior) e da média móvel trimestral da PI são substanciais (revisão absoluta média de 0,91 p.p. e 0,39 p.p., respectivamente), embora essas revisões tornem-se menos importantes à medida que o período de agregação aumenta (por exemplo, crescimento em doze meses). Além disso, mostramos que as revisões da aceleração da PI (variação da taxa de crescimento da PI) são também importantes (revisão absoluta média da aceleração mensal de 1,48 p.p. e da aceleração da média móvel trimestral de 0,43 p.p.). Na análise do hiato da PI, encontramos que todos os métodos apresentam revisões de magnitude elevada. A revisão absoluta média dos diferentes hiatos foi de 1,26 p.p. a 3,18 p.p. Em dois dos três métodos analisados, a revisão implicou mudança do sinal do hiato em 30% ou mais das vezes. Em geral, tanto a revisão de dados da PI como o aumento da amostra mostraram-se fontes relevantes das revisões do hiato da PI, embora os resultados sugiram certa predominância das revisões provenientes do aumento da amostra.

O artigo é organizado da seguinte maneira. Na seção 2, fazemos uma breve introdução à literatura e aos conceitos da área de análise de dados em tempo real. Na seção 3, esclarecemos alguns aspectos sobre a elaboração do conjunto de dados de produção industrial em tempo real para o Brasil e analisamos as revisões do crescimento e da aceleração da PI. Na seção 4, apresentamos os métodos utilizados para calcular os hiatos da PI, a metodologia para decompor as revisões dos hiatos e a análise das revisões. Na seção 5, concluímos.

2 Dados em tempo real

A literatura de análise de dados em tempo real está relacionada a análises nas quais a revisão de dados é relevante ou o próprio momento no qual os dados são divulgados tem importância. Os trabalhos na área têm estudado as propriedades das revisões de dados, o impacto da revisão de dados sobre a pesquisa macroeconômica e a política monetária, e o efeito da revisão de dados sobre previsões econômicas. Diebold & Rudebusch (1991), por exemplo, estudaram a capacidade de indicadores antecedentes de prever os movimentos da produção industrial. Eles perceberam que o índice de indicadores de antecedentes calculados com dados em tempo real apresentava um desempenho muito inferior ao mesmo índice calculado com os dados já revisados.

Embora o surgimento da literatura de análise em tempo real possa ser datado da década de 1950, a literatura ganhou consistência, efetivamente, com o desenvolvimento de um grande conjunto de dados em tempo real para a economia norte-americana. Esse conjunto de dados foi desenvolvido por Dean Croushore e Tom Stark em meados da década de 1990 e disponibilizado na *internet* a partir de 1999.¹ Este conjunto de dados é ainda atualizado e conta com a cooperação do *Federal Reserve Bank* da Filadélfia e da Universidade de Richmond. O tema subjacente a toda esta literatura (Croushore 2008) é: as revisões de dados são suficientemente grandes em termos econômicos para nos preocupar? A revisão de dados gera algumas dificuldades para a análise de previsões e de política monetária. Por exemplo, pode ser inadequado utilizar

¹<http://www.philadelphiafed.org/research-and-data/real-time-center/real-time-data/>.

o último conjunto de dados disponível para confrontar previsões de um novo modelo a outras previsões efetuadas em tempo real. Nesse caso, enquanto as previsões em tempo real teriam sido efetuadas com dados não-revisados, as previsões do novo modelo teriam o benefício de utilizar dados já revisados, provavelmente mais precisos. Outro exemplo: revisões de dados de magnitude elevada implicam dificuldades para os formuladores de política, pois, nesse caso, suas decisões são baseadas em dados que revelam uma “fotografia” não muito fidedigna do estado da economia.

Basicamente, há três razões por que os dados são revisados ao longo do tempo: (i) o conjunto de informação disponível para estimar determinado dado aumenta, permitindo melhorar a estimativa da variável; (ii) os fatores sazonais (quando há ajuste sazonal) são reestimados, modificando a série histórica; (iii) periodicamente, visando aprimorar a qualidade dos dados, há uma mudança metodológica de cálculo da variável, e a série histórica é reconstruída. De fato, as instituições que produzem dados se defrontam com um *trade-off* entre rapidez de divulgação dos dados e precisão. Por um lado, a instituição pode produzir dados melhores ao esperar que seu conjunto de informação aumente; por outro, os formuladores de política e agentes econômicos, que precisam tomar decisões, demandam certa rapidez pela disponibilidade dos dados.

Um dado em tempo real pode ser definido como um dado que é tal como existia antes de ser subsequentemente revisado (Stark 2002). De acordo com Croushore & Stark (2000, 2001), utilizamos o termo “*vintage*” para designar o conjunto de informação disponível de uma variável em uma data particular (a última série disponível na data, com as revisões mais atualizadas daquele momento). À coleção dessas *vintages*, chamamos de “conjunto de dados em tempo real”.

Para ilustrar esses conceitos, na Tabela 1, mostramos um conjunto de dados em tempo real, em número índice, da produção industrial do Brasil com ajuste sazonal. Na segunda coluna (*vintage* 2008:5), temos os dados disponíveis para os agentes econômicos na ocasião da (primeira) divulgação do dado referente a maio de 2008; na terceira coluna (*vintage* 2008:6), temos os dados disponíveis para os agentes econômicos na ocasião da (primeira) divulgação do dado referente a junho de 2008; e assim por diante. Note que todas as *vintages* iniciam com dados a partir do mesmo período — nesse exemplo, a partir de setembro de 2007. Tomando o conjunto de dados de uma coluna qualquer, obtemos a mais recente série histórica disponível em uma determinada data. Por outro lado, tomando o conjunto de dados de uma linha qualquer, podemos observar como um dado particular foi revisado ao longo do tempo. A diagonal inferior da Tabela 1 é chamada de série de dados em tempo real e contém os dados tais como foram inicialmente calculados ao longo do tempo. A última coluna da Tabela 1 apresenta o que é usualmente chamado de série de dados final. A série de dados final é uma série que possui os dados referentes ao mesmo período da série de dados em tempo real, mas os dados são os mais recentes disponíveis no momento da pesquisa. A série de revisão de dados é obtida pela diferença entre a série de dados final e a série de dados em tempo real.

Tabela 1: Conjunto de Dados em Tempo Real - Produção Industrial com Ajuste Sazonal (em número índice)

Período	Vintage					
	2008:5	2008:6	2008:7	2008:8	2008:9	2008:10
2007:9	122,99	122,94	122,88	122,98	122,70	122,36
2007:10	127,16	127,19	127,16	127,23	127,12	127,60
2007:11	124,66	124,71	124,68	124,77	124,71	124,96
2007:12	123,68	123,92	123,91	123,98	123,93	124,15
2008:1	125,96	126,16	126,15	126,16	126,14	126,29
2008:2	125,34	125,59	125,64	125,67	125,67	125,74
2008:3	125,90	126,20	126,32	126,44	126,45	126,43
2008:4	126,19	126,60	126,59	126,66	126,71	126,66
2008:5	125,54	125,80	125,70	125,59	125,68	125,47
2008:6		129,17	129,29	129,19	129,39	129,23
2008:7			130,64	131,03	131,04	131,02
2008:8				129,28	129,49	129,09
2008:9					131,66	131,01
2008:10						128,79

Fonte: IBGE (Ver seção 3.1).

3 Produção Industrial

3.1 Elaboração de um conjunto de dados de produção industrial em tempo real

Utilizando as publicações do Instituto Brasileiro de Geografia e Estatística (IBGE),² digitamos e organizamos um conjunto de dados de produção industrial em tempo real para o Brasil. Os dados são mensais e com ajuste sazonal. Os dados iniciam em 1991:1 e a primeira *vintage* do conjunto de dados é referente a 1998:4.³ A última *vintage* é de 2008:10.

Sempre que o IBGE faz a primeira divulgação do dado de PI referente ao mês de dezembro do ano a , ele divulga também os dados “definitivos” referentes ao ano $a - 1$. Esses dados do ano $a - 1$ não são mais alterados, exceto quando há uma mudança metodológica. Assim, os dados de cada mês são potencialmente revisados de 12 a 23 vezes, dependendo do mês a que se refere. No período compreendido pela nossa amostra de *vintages*, houve apenas uma mudança metodológica, implementada a partir do dado de fevereiro de 2004.

3.2 Análises das revisões da produção industrial

Embora nossas *vintages* de produção industrial sejam em número índice, analisaremos a revisão de dados a respeito do crescimento da PI (variação percentual em relação ao i -ésimo mês anterior), calculado da seguinte maneira:⁴

²“Indicadores IBGE — Pesquisa Industrial Mensal — Produção Física — Brasil”. As publicações utilizadas neste trabalho estão disponíveis no sítio do IBGE: <http://www.ibge.gov.br/home/download/estatistica.shtm>.

³Refere-se à primeira *vintage* disponível no sítio do IBGE.

⁴Utilizamos a aproximação logarítmica pela conveniência analítica que ela oferece. Por exemplo, a taxa de crescimento acumulada da PI em relação ao i -ésimo trimestre anterior pode ser facilmente decomposta em i parcelas aditivas. Para nos certificar de que a aproximação é razoá-

$$\Delta PI_{n,t} = 100 \log \left(\frac{PI_{n,t}}{PI_{n,t-i}} \right) \quad (1)$$

onde $PI_{n,t}$ é o índice da produção industrial para o período t segundo a *vintage* n , $\Delta PI_{n,t}$ é o crescimento da produção industrial, e \log é o logaritmo natural.

A Figura 1 apresenta o crescimento da PI nos conceitos final (última *vintage* disponível) e em tempo real. Embora as taxas sejam bem correlacionadas, as diferenças podem ser substanciais. A importância das revisões de dados da PI é mais clara na Figura 2, onde apresentamos o crescimento da PI de abril de 1998 tal como evoluiu ao longo das revisões (escolhemos esse dado de PI porque foi divulgado inicialmente na *vintage* mais antiga que dispomos). Quando o dado foi inicialmente divulgado, o crescimento da PI de 1998:4 era $-0,99\%$. Um mês depois, na *vintage* de 1998:5, o dado foi revisado para $-0,76\%$. O valor mínimo ocorreu na *vintage* de 1999:7 ($-1,40\%$). Na *vintage* de 2004:2, com a introdução de uma mudança metodológica, o valor do dado foi alterado para $0,82\%$ e manteve-se constante a partir de então.

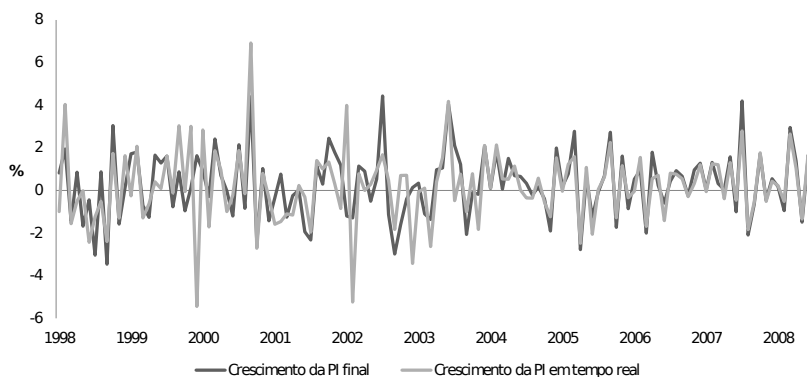


Figura 1: Crescimento da PI final e em tempo real (%) (Mês/mês anterior)

Um panorama mais geral é mostrado na Figura 3, que apresenta a frequência relativa dos valores das revisões da PI. O gráfico mostra que somente em 23% das vezes as revisões foram próximas de zero (entre $-0,25$ p.p. e $0,25$ p.p.). Em 33% das vezes, a magnitude de revisão foi superior a 1 p.p. Na Figura 4, podemos observar todos os valores de PI tomados em tempo real (eixo x) e os respectivos valores finais (eixo y). Sempre que a observação da produção industrial estiver exatamente sobre a linha de 45, a revisão da PI foi zero. Observações acima da linha de 45 indicam revisões positivas e observações abaixo indicam revisões negativas. Observações que estão nos quadrantes superior esquerdo e inferior direito revelam troca de sinal depois das revisões, fato que ocorreu em 19% das vezes.

A média da taxa de crescimento da PI em tempo real foi $0,18\%$, enquanto foi $0,28\%$ para a série final. A revisão média foi $0,10$ p.p. Embora a revisão média seja útil como um indicador do viés das revisões, ela é limitada como um indicador da magnitude das revisões, uma vez que revisões negativas contrabalançam revisões positivas e vice-versa. A Tabela 2 apresenta alguns indicadores adicionais de revisão. Dois indicadores são particularmente

vel, fizemos também os cálculos sem a aproximação, e os resultados foram muito semelhantes.

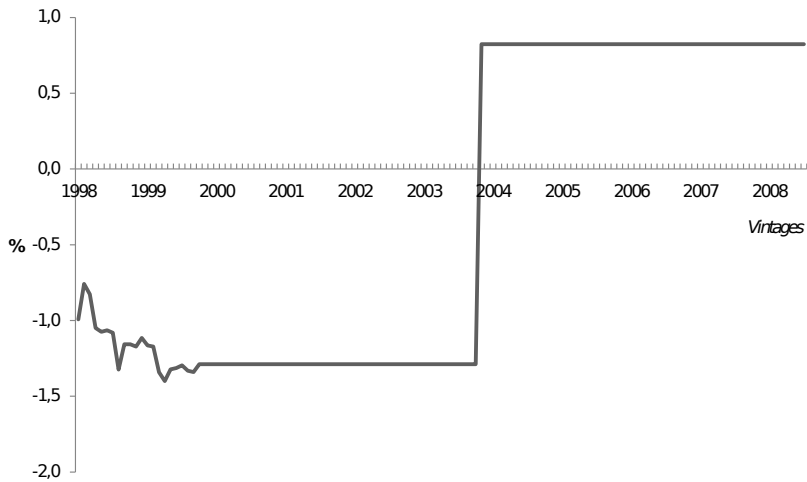


Figura 2: Crescimento da PI de 1998:4 ao longo das revisões (%) (Mês/mês anterior)

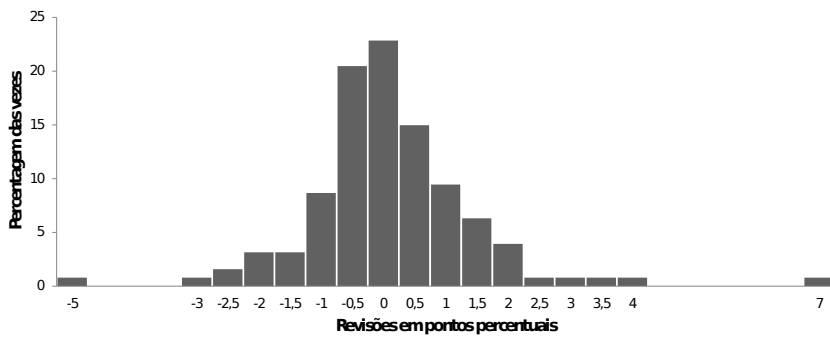


Figura 3: Frequência relativa dos valores das revisões da produção industrial

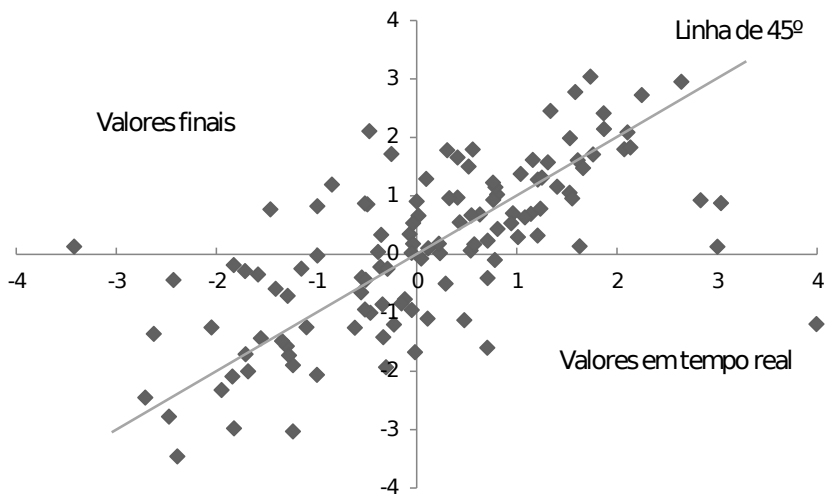


Figura 4: Valores da produção industrial (tempo real versus finais)

adequados para capturar a magnitude das revisões: a revisão absoluta média (RAM) e a raiz da revisão quadrática média (RRQM). Podemos observar que a revisão absoluta média do crescimento mensal foi 0,91 p.p. Isto significa que, em média, o crescimento da produção industrial foi revisado em 0,91 p.p. acima ou abaixo do valor inicialmente divulgado. A RRQM, que penaliza mais fortemente as revisões de maior magnitude, foi 1,37 p.p.

A Tabela 2 mostra também os indicadores R/S, CORR, SIOP, FRMA e AR. R/S é uma *proxy* para a razão ruído-sinal, obtida pela razão entre a RRQM e o desvio-padrão da estimativa final do crescimento da PI. Essa medida capta a magnitude das revisões em relação ao desvio-padrão da série final. O valor da R/S é 0,88, mostrando que a magnitude das revisões não é muito diferente da variabilidade da série. CORR representa a correlação entre as séries em tempo real e final, a qual é de 0,65. SIOP é a frequência na qual o crescimento da produção industrial em tempo real tem sinal oposto ao do crescimento da produção industrial final. Em 19% dos casos, o dado de crescimento da PI é revisado de tal forma que o seu sinal é alterado. E, finalmente, FRMA mede a frequência na qual a revisão do crescimento da PI é maior do que o crescimento da PI final, ambos em valores absolutos, fato que ocorreu em 33% dos casos.

Os indicadores de revisão sugerem que as revisões de crescimento da produção industrial são elevadas. Porém, há um atenuante — a correlação serial de primeira ordem da série de revisão (AR) é negativa: $-0,37$. Isso revela que revisões positivas de um mês são usualmente seguidas por revisões negativas no mês seguinte e vice-versa. Portanto, ao se agregar a série de revisões, as revisões devem perder parte de sua importância. Para verificar se isso de fato ocorre, calculamos as revisões da taxa de crescimento da produção industrial em relação ao i -ésimo mês anterior. Para obter a taxa, comparamos a PI do mês corrente com os valores da PI de um, três, seis ou doze meses atrás. Como podemos observar na Tabela 2, há uma forte melhora de todos indicadores à medida que a agregação aumenta. Há uma redução acentuada da revisão absoluta média (RAM/mês) e da raiz da revisão quadrática média (RRQM/mês), ambas expressas em pontos percentuais por mês.⁵ Além disso, ao longo do aumento da agregação, há uma forte redução da razão ruído-sinal (R/S), da SIOP e da FRMA. Finalmente, há também um aumento da correlação da série de PI em tempo real com a série de PI final. Assim, podemos concluir que as revisões da PI tornam-se menos importantes quando o período de agregação aumenta.

Alternativamente, podemos avaliar o efeito do aumento da agregação por meio da taxa de crescimento da PI acumulada em i meses em relação aos i meses anteriores. Matematicamente,⁶

$$\Delta PI_{n,t} = 100 \log \left(\frac{\sum_{j=0}^{i-1} PI_{n,t-j}}{\sum_{j=0}^{i-1} PI_{n,t-i-j}} \right) \quad (2)$$

Os indicadores da Tabela 3 corroboram os achados da Tabela 2. Há uma forte redução dos indicadores RAM/mês e RRQM/mês à medida que o período de agregação aumenta. Além disso, há uma forte melhora dos indicadores R/S, SIOP e FRMA.

⁵A RAM/mês e a RRQM/mês são obtidas pela divisão da RAM e da RRQM, respectivamente, pelo número de meses em análise.

⁶Note que, quando $i = 1$, as equações (1) e (2) são equivalentes.

Tabela 2: Indicadores de Revisão — Crescimento da Produção Industrial (Variação % em relação ao *i*-ésimo mês anterior) — 1998:4-2008:10

Taxa de crescimento acumulada em	Revisão Média/mês	RAM/mês	RRQM/mês	AR	R/S	CORR	SIOP	FRMA
1 mês	0,10	0,91	1,37	-0,37	0,88	0,65	0,19	0,33
3 meses	0,05	0,39	0,57	0,28	0,71	0,76	0,12	0,27
6 meses	0,03	0,24	0,33	0,38	0,64	0,81	0,13	0,24
12 meses	0,01	0,07	0,11	0,00	0,33	0,95	0,06	0,11

Notas: Série de produção industrial mensal com ajuste sazonal. Revisão Média/mês é a revisão média dividida pelo número de meses em análise. RAM/mês é a revisão absoluta média dividida pelo número de meses em análise. RRQM/mês é a raiz da revisão quadrática média dividida pelo número de meses em análise. AR é a correlação serial de primeira ordem da série de revisão. R/S é uma proxy da razão ruído-sinal (obtida pela razão entre RRQM e o desvio-padrão da estimativa final do crescimento da produção industrial). CORR é a correlação entre o crescimento da produção industrial em tempo real e o crescimento da produção industrial final. SIOP é a frequência na qual o crescimento da produção industrial em tempo real tem sinal oposto do crescimento da produção industrial final. FRMA é a frequência na qual a revisão do crescimento da produção industrial é maior do que o crescimento da produção industrial final, ambos em valores absolutos.

Entre os agentes econômicos, é comum a percepção de que a série mensal de produção industrial apresenta muito ruído. Assim, uma medida muito utilizada para atenuar esse problema é o crescimento da média móvel trimestral da produção industrial, a qual pode ser calculada da seguinte maneira:

$$\Delta PI_{n,t} = 100 \log \left(\frac{PI_{n,t} + PI_{n,t-1} + PI_{n,t-2}}{PI_{n,t-1} + PI_{n,t-2} + PI_{n,t-3}} \right) \quad (3)$$

Os resultados apresentados na segunda linha da Tabela 2 (taxa de crescimento acumulada em três meses) equivalem aproximadamente aos indicadores de revisão para o crescimento da média móvel trimestral.⁷ Por exemplo, a RAM da taxa de crescimento da média móvel trimestral é 0,39 p.p. e a razão ruído-sinal é 0,71. Assim, as revisões da média móvel trimestral, embora sejam menores do que as revisões da série mensal original, ainda mostram-se importantes.

Finalmente, outra ótica relevante de análise das revisões é a aceleração da PI. Chamamos de aceleração da PI à variação da taxa de crescimento da PI entre um mês e o i -ésimo mês anterior. Matematicamente,

$$\Delta^2 PI_{n,t} = \Delta PI_{n,t} - \Delta PI_{n,t-1} \quad (4)$$

onde $\Delta^2 PI_{n,t}$ é a aceleração da produção industrial para o período t , segundo a *vintage* n , e $\Delta PI_{n,t}$ é o crescimento da produção industrial. Uma aceleração positiva (negativa) significa que o crescimento da produção industrial está aumentando (diminuindo), sugerindo um fortalecimento (enfraquecimento) da atividade econômica. Os indicadores de revisão apresentados na Tabela 4 indicam que a aceleração da PI apresenta revisões também elevadas. A RAM, por exemplo, foi 1,48 p.p. para a aceleração da taxa de crescimento da PI em relação ao mês anterior. A variação da taxa de crescimento acumulada em três meses equivale aproximadamente à variação da taxa de crescimento da média móvel trimestral. Assim, no caso da média móvel trimestral houve uma redução da RAM para 0,43 p.p., mas não houve melhora substancial dos indicadores R/S, CORR, SIOP e FRMA (os dois últimos indicam inclusive uma piora).

4 Hiato da produção industrial

A definição usual do hiato da produção industrial é a diferença entre a produção industrial e a produção industrial potencial (PI potencial): $x_t = y_t - y_t^*$. Em termos práticos, o hiato pode ser obtido por meio de métodos de extração de tendência e é calculado como o desvio entre a produção industrial e sua tendência.

4.1 Componentes das revisões do hiato da produção industrial

Baseando-se na metodologia de Orphanides & Van Norden (2002), analisamos o comportamento das estimativas do hiato da PI de final de amostra e a revisão

⁷Para $i = 3$, o lado direito da equação (1), dividido por 3, é uma aproximação para o lado direito da equação (3), isto é, $\frac{100}{3} \log \left(\frac{PI_{n,t}}{PI_{n,t-3}} \right) \approx 100 \log \left(\frac{PI_{n,t} + PI_{n,t-1} + PI_{n,t-2}}{PI_{n,t-1} + PI_{n,t-2} + PI_{n,t-3}} \right)$ para $(PI_{n,t} - PI_{n,t-1})$, $(PI_{n,t-1} - PI_{n,t-2})$ e $(PI_{n,t-2} - PI_{n,t-3})$ pequenos. Efetuamos o cálculo dos indicadores também sem utilizar a aproximação, e os resultados foram virtualmente idênticos.

Tabela 3: Indicadores de Revisão — Crescimento da Produção Industrial (Variação % da Produção Industrial acumulada em i meses em relação aos i meses anteriores) 1998:4-2008:10

Taxa de crescimento acumulada em	Revisão Média/mês	RAM/mês	RRQM/mês	AR	R/S	CORR	SIOP	FRMA
1 mês	0,10	0,91	1,37	-0,37	0,88	0,65	0,19	0,33
3 meses	0,03	0,28	0,40	0,69	0,63	0,81	0,09	0,24
6 meses	0,02	0,12	0,16	0,78	0,41	0,92	0,03	0,09
12 meses	0,00	0,01	0,02	0,58	0,07	0,99	0,02	0,04

Notas: Série de produção industrial mensal com ajuste sazonal. Revisão Média/mês é a revisão média dividida pelo número de meses em análise. RAM/mês é a revisão absoluta média dividida pelo número de meses em análise. RRQM/mês é a raiz da revisão quadrática média dividida pelo número de meses em análise. AR é a correlação serial de primeira ordem da série de revisão. R/S é uma proxy da razão ruído-sinal (obtida pela razão entre RRQM e o desvio-padrão da estimativa final do crescimento da produção industrial). CORR é a correlação entre o crescimento da produção industrial em tempo real e o crescimento da produção industrial final. SIOP é a frequência na qual o crescimento da produção industrial em tempo real tem sinal oposto do crescimento da produção industrial final. FRMA é a frequência na qual a revisão do crescimento da produção industrial é maior do que o crescimento da produção industrial final, ambos em valores absolutos.

Tabela 4: Indicadores de Revisão — Aceleração da Produção Industrial (Variação do crescimento em p.p. em relação ao *i*-ésimo mês anterior) — 1998:4-2008:10

Variação da taxa de crescimento acumulada em	Revisão Média/mês	RAM/mês	RRQM/mês	AR	R/S	CORR	SIOP	FRMA
1 mês	0,07	1,48	2,24	-0,63	0,93	0,63	0,15	0,34
3 meses	0,02	0,43	0,66	-0,38	0,85	0,66	0,24	0,36

Notas: Série de produção industrial mensal com ajuste sazonal. Revisão Média/mês é a revisão média dividida pelo número de meses em análise. RAM/mês é a revisão absoluta média dividida pelo número de meses em análise. RRQM/mês é a raiz da revisão quadrática média dividida pelo número de meses em análise. AR é a correlação serial de primeira ordem da série de revisão. R/S é uma proxy da razão ruído-sinal (obtida pela razão entre RRQM e o desvio-padrão da estimativa final do crescimento da produção industrial). CORR é a correlação entre o crescimento da produção industrial em tempo real e o crescimento da produção industrial final. SIOP é a frequência na qual o crescimento da produção industrial em tempo real tem sinal oposto do crescimento da produção industrial final. FRMA é a frequência na qual a revisão do crescimento da produção industrial é maior do que o crescimento da produção industrial final, ambos em valores absolutos.

dessas estimativas ao longo do tempo. Para avaliar as revisões, efetuamos três grupos de estimativas do hiato: (i) estimativas finais; (ii) estimativas em tempo real; (iii) estimativas “quase-reais”. As estimativas finais dos hiatos da PI são efetuadas com a última *vintage* de dados de produção industrial utilizada no presente trabalho (2008:10). A série resultante é composta pelos hiatos finais. Essa é a maneira usual de calcular hiatos, utilizada em trabalhos que não levam em conta a existência da revisão de dados. As estimativas em tempo real dos hiatos da PI são efetuadas em duas etapas. Em uma primeira etapa, são calculados os hiatos da PI para todas as *vintages* disponíveis. Para cada *vintage*, uma série de hiatos é estimada. Em uma segunda etapa, toma-se a última observação de cada série de hiatos. A série resultante é composta pelos hiatos em tempo real. Essa série contém, para cada ponto no tempo, a primeira estimativa de hiato da PI que os agentes econômicos poderiam ter efetuado.⁸

A revisão total dos hiatos da produção industrial estimados em cada ponto do tempo é a diferença entre os hiatos finais e os hiatos em tempo real. Essa revisão total pode ser decomposta em duas fontes: (i) as revisões provenientes das revisões dos dados da PI e (ii) as revisões provenientes do aumento das observações da PI (à medida que o tempo passa, o número de observações de PI aumenta, modificando as estimativas dos hiatos).⁹ Para isolar a importância desses fatores, calculamos um terceiro grupo de estimativas de hiato: as estimativas quase-reais. As estimativas quase-reais do hiato da PI são calculadas utilizando o mesmo período amostral das estimativas em tempo real, mas, em vez de utilizar os dados em tempo real, utilizam-se os dados finais. Assim, inicialmente procede-se a uma estimação utilizando os dados finais até 1998:4; em um segundo momento, procede-se a uma estimação utilizando os dados finais até 1998:5; e assim, sucessivamente, até efetuar uma estimação que utilize os dados finais até 2008:10. Em seguida, toma-se a última observação de cada estimação. A série resultante é composta pelos hiatos quase-reais. A diferença entre os hiatos quase-reais e os hiatos em tempo real é completamente devido a revisões de dados de PI, uma vez que as estimativas das duas séries em qualquer ponto específico do tempo são realizadas com amostra de dados que cobrem exatamente o mesmo período. Por outro lado, a diferença entre as estimativas finais e quase-reais do hiato capta o efeito do aumento da amostra.

Implicitamente, supomos que as revisões melhoram as estimativas de hiato da produção industrial. Contudo, é razoável também assumir que algum grau de incerteza permanece na última *vintage* estimada de hiato da produção industrial. Essa *vintage* provavelmente será revisada ou, mesmo não o sendo, ela terá as limitações metodológicas de qualquer estimativa de PI. Além disso, mesmo com uma medida perfeita de PI, o hiato da PI é uma variável não-observável, implicando que qualquer estimativa sua carrega um grau de incerteza não-desprezível. Assim, a revisão total capta apenas parte do erro de medida associado aos hiatos calculados em tempo real.

⁸Cabe ressaltar que a *vintage* e , portanto, a estimativa do hiato da PI referentes ao mês t só estarão disponíveis para os agentes econômicos no início do mês $t + 2$.

⁹Parte do efeito de cada uma das duas fontes acima é proveniente da reestimação dos fatores sazonais.

4.2 Análises das revisões totais do hiato da produção industrial

A Figura 5 apresenta as três séries em tempo real de hiato da produção industrial, calculadas a partir do conjunto de dados em tempo real. A região sombreada da Figura 6 mostra, para cada período do tempo, o intervalo no qual se encontram os hiatos calculados a partir dos diversos métodos. O intervalo pode ser interpretado como uma medida de incerteza no sentido de “thick modeling” (Granger & Jeon 2004). Os três métodos apresentam correlações positivas. Nas Figura 7 e 8, temos dois gráficos análogos aos anteriores, mas apresentando as séries finais de hiato da PI. Novamente, todas as correlações são positivas. Comparando as Figuras 6 e 8, podemos perceber que a incerteza dos hiatos no sentido de thick modeling é maior nas estimativas em tempo real. De fato, as diferentes estimativas de hiatos em tempo real apresentam uma amplitude média igual a 4,21 p.p., enquanto essa amplitude é 2,57 p.p. para as estimativas finais.

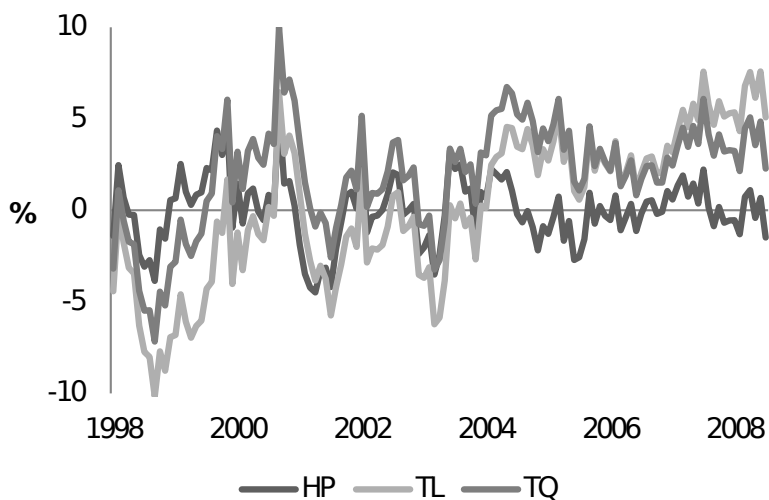


Figura 5: Hiatos em tempo real

As séries de hiato em tempo real podem apresentar diferenças importantes em relação às séries finais. A Figura 9 contrapõe a série de hiatos em tempo real à série final, obtidas pelo método de Hodrick-Prescott. A correlação entre ambas as séries é 0,55, e, em 30% das vezes, as observações da série em tempo real apresentam sinais opostos aos da série final.

Para ilustrar a importância da revisão das estimativas de hiato da PI, apresentamos, na Figura 10, a estimativa do hiato da PI de abril de 1998, calculada pelo método do filtro HP, tal como evoluiu ao longo das revisões. Quando o hiato foi inicialmente calculado, ele era $-1,41\%$ da PI. Na *vintage* de 1998:7, ele foi revisado para $-1,77\%$, atingindo seu mínimo. Na *vintage* de 2001:3, ele foi revisado para $0,02\%$, atingindo seu valor máximo antes da mudança metodológica de 2004:2. Na *vintage* de 2004:2, ele foi revisado para $0,62\%$, mantendo-se praticamente constante a partir de então. Utilizando todo o conjunto de dados, a Figura 11 mostra que, em apenas 13% das vezes, as revisões foram próximas de zero (entre $-0,25$ p.p. e $0,25$ p.p.). Em 49% das vezes, a magnitude da revisão foi superior a 1 p.p. A Figura 12 revela que houve troca

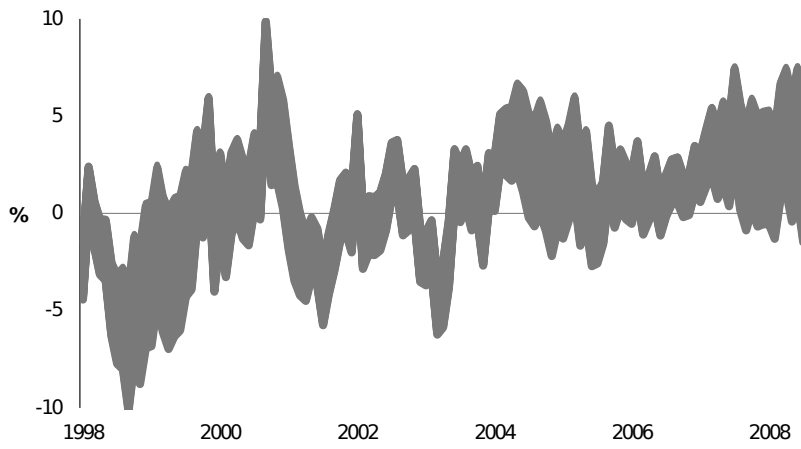


Figura 6: Hiatos em tempo real — “Thick Modeling”

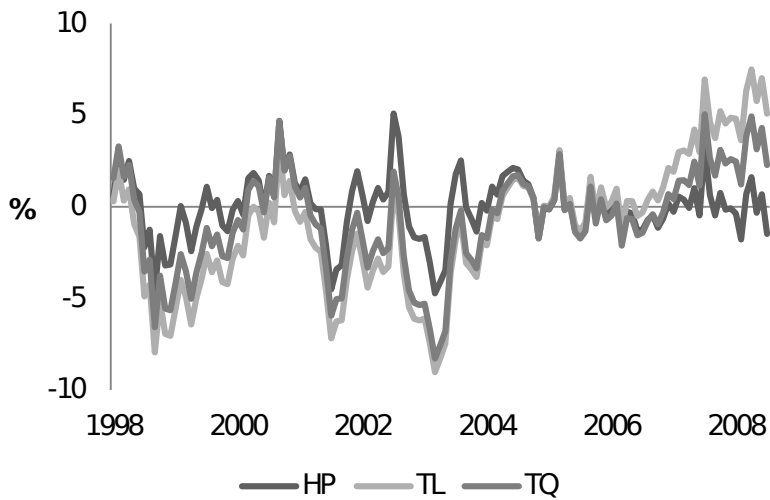


Figura 7: Hiatos finais

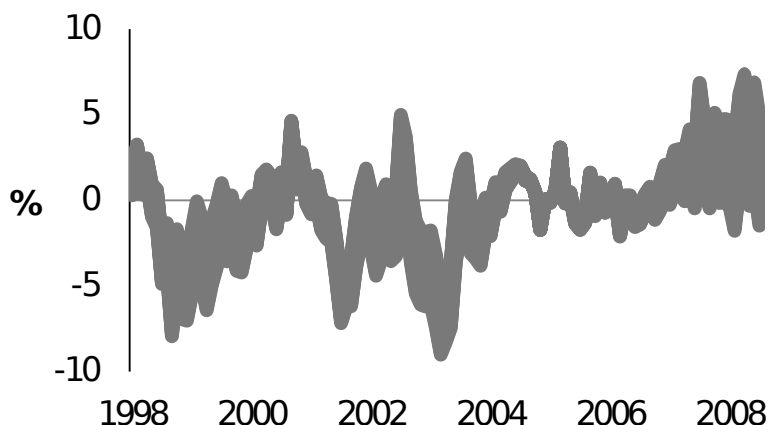


Figura 8: Hiatos finais — “Thick Modeling”

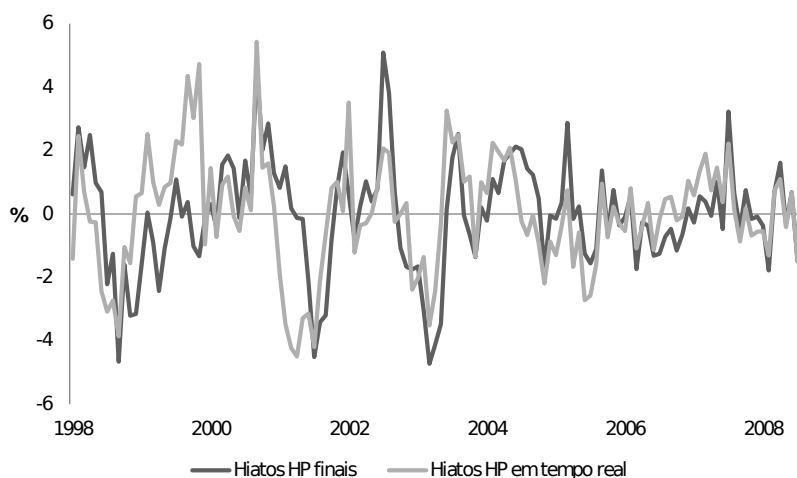


Figura 9: Hiatos HP finais e em tempo real

do sinal do hiato do produto em 30% das vezes (observações que estão nos quadrantes superior esquerdo e inferior direito).¹⁰

A Tabela 5 mostra os indicadores de revisão para o hiato da PI. A revisão média indica valores elevados, em magnitude, para os métodos TQ e TL, mas baixo para o método HP. Por outro lado, observando os indicadores RAM e RRQM pode-se perceber que todos os métodos analisados apresentam revisões substanciais. O método HP, que teve uma revisão média próxima de zero, apresentou uma RAM de 1,26 p.p. O método TQ apresentou a maior RAM (3,18 p.p.), enquanto o método TL registrou uma RAM de 2,02 p.p.

A correlação serial de primeira ordem (AR), que mostra o grau de persistência das revisões, ficou no intervalo entre 0,68 e 0,78. Uma persistência elevada revela que os “erros” das estimativas de hiato em tempo real se es-

¹⁰Gráficos análogos aos das Figuras 9 a 12 para os métodos de tendência linear e tendência quadrática podem ser solicitados aos autores.

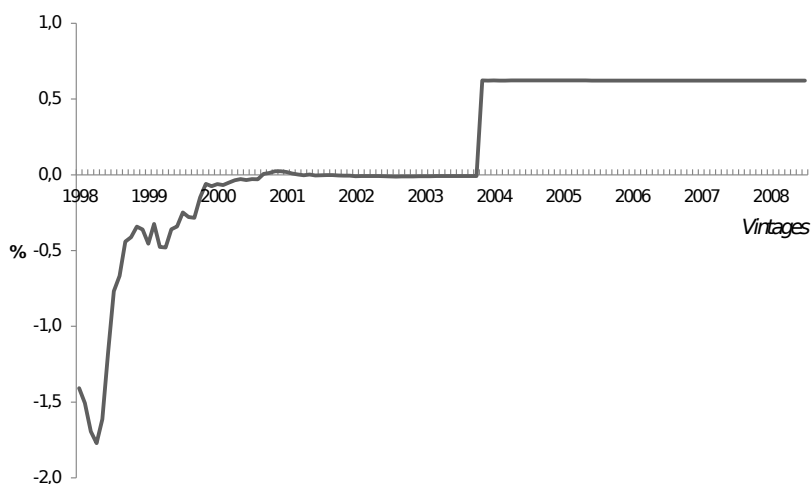


Figura 10: Hiato da PI (Filtro HP) de 1998:4 ao longo das revisões (%)

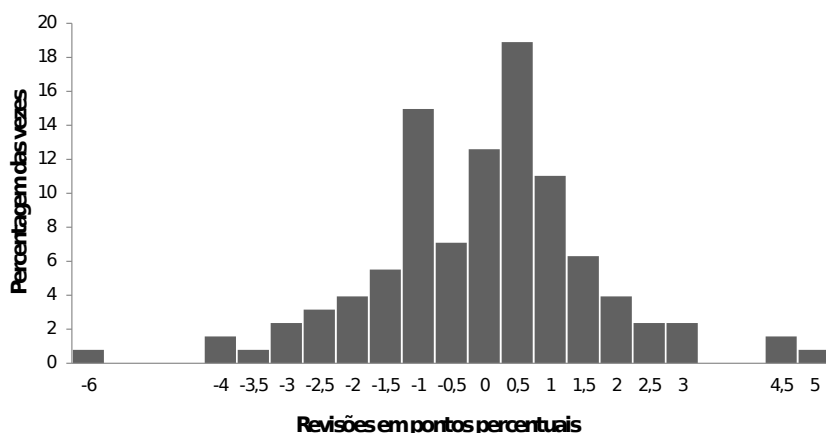


Figura 11: Frequência relativa das revisões dos hiatos HP

Tabela 5: Indicadores de Revisão — Hiatos da Produção Industrial (%) 1998:4-2008:10

	Revisão Média	RAM	RRQM	AR	R/S	CORR	SIOP	FRMA
Hodrick-Prescott (HP)	-0,04	1,26	1,69	0,68	0,97	0,55	0,30	0,49
Tendência Linear (TL)	-1,21	2,02	2,36	0,72	0,66	0,87	0,19	0,42
Tendência Quadrática (TQ)	-2,78	3,18	3,61	0,78	1,31	0,68	0,44	0,69

Notas: A amostra de dados utilizada para as estimações dos hiatos inicia em 1991:1. RAM é a revisão absoluta média. RRQM é a raiz da revisão quadrática média. AR é a correlação serial de primeira ordem da série de revisão. R/S é uma proxy da razão ruído-sinal (obtida pela razão entre RRQM e o desvio-padrão do hiato final). CORR é a correlação entre o hiato em tempo real e o hiato final. SIOP é a frequência relativa na qual o hiato em tempo real tem sinal oposto do hiato final. FRMA é a frequência relativa na qual a revisão do hiato é maior do que o hiato final, ambos em valores absolutos.

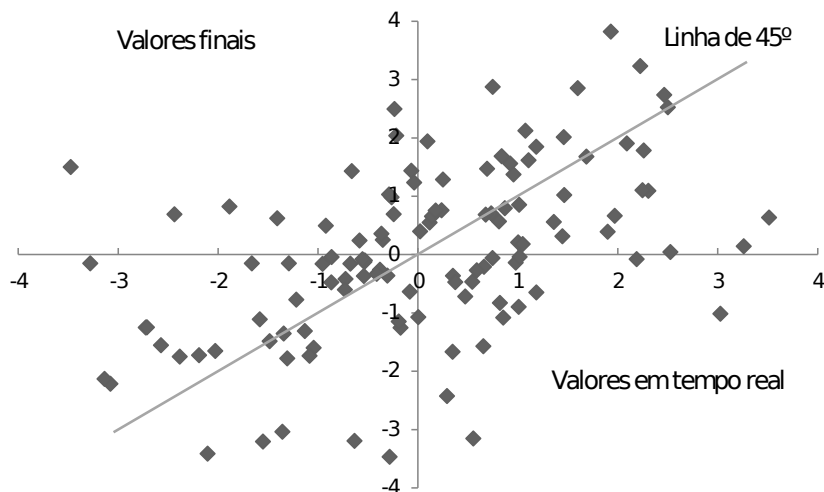


Figura 12: Valores dos hiatos HP (tempo real versus finais)

tendem por períodos prolongados. Nesse caso, estimativas de hiato da PI em tempo real podem levar os formuladores de política e outros agentes econômicos a percepções equivocadas persistentes sobre o estado do ciclo econômico. A razão ruído-sinal (R/S) mostra valores próximos ou superiores a 1 para dois dos três métodos. A correlação entre os hiatos finais e em tempo real calculados pelo método de tendência linear foi a mais elevada (0,87), enquanto que a mais baixa foi para o método HP (0,55). O indicador SIOF mostrou-se elevado para pelo menos dois dos três métodos. Por exemplo, 44% dos hiatos em tempo real calculados pelo método TQ apresentaram o sinal “errado”. A TL apresentou a SIOF mais baixa (0,19). Finalmente, o indicador FRMA, que é a frequência relativa na qual a revisão do hiato é maior do que o hiato final, ambos em valores absolutos, é elevado para todos os métodos.

4.3 Decomposição das revisões do hiato da produção industrial

Como ressaltado na seção 4.1, as revisões totais do hiato da PI podem ser decompostas em dois componentes — parcela associada a revisões da PI e parcela associada ao aumento da amostra. Os efeitos da revisão são isolados na Tabela 6, onde:

Revisão total	= série final – série em tempo real
Efeito da revisão dos dados	= série quase real – série em tempo real
Efeito do aumento da amostra	= série final – série quase real

No método HP, o viés positivo gerado pela revisão de dados (0,12 p.p.) é mais do que contrabalançado pelo viés negativo gerado pelo aumento da amostra (–0,16 p.p.). Os indicadores RAM, RRQM e R/S mostram que tanto a revisão de dados quanto o aumento da amostra são relevantes para explicar a magnitude das revisões totais. Por exemplo, a RAM proveniente da revisão de dados é 0,76 p.p., enquanto que a RAM decorrente do aumento da amostra é 1,19 p.p., totalizando uma RAM total de 1,26 p.p.

Tabela 6: Decomposição das Revisões dos Hiatos da Produção Industrial (%) 1998:4-2008:10

Método	Média	Desvio-Padrão	Valor Mínimo	Valor Máximo	RAM	RRQM	R/S	AR
Hodrick-Prescott (HP)								
Revisão total	-0,04	1,69	-6,06	4,98	1,26	1,69	0,97	0,68
Efeito da revisão dos dados	0,12	1,06	-3,93	2,74	0,76	1,06	0,61	0,29
Efeito do aumento da amostra	-0,16	1,46	-3,36	2,79	1,19	1,47	0,84	0,96
Tendência Linear (TL)								
Revisão total	-1,21	2,03	-5,90	4,72	2,02	2,36	0,66	0,72
Efeito da revisão dos dados	0,16	1,22	-5,17	2,52	0,88	1,22	0,34	0,32
Efeito do aumento da amostra	-1,37	1,72	-3,50	3,67	1,93	2,19	0,62	0,97
Tendência Quadrática (TQ)								
Revisão total	-2,78	2,31	-8,85	4,75	3,18	3,61	1,31	0,78
Efeito da revisão dos dados	0,19	1,18	-4,81	2,49	0,86	1,19	0,43	0,32
Efeito do aumento da amostra	-2,97	2,11	-6,13	3,70	3,34	3,64	1,33	0,96

Notas: A amostra de dados utilizada para as estimações dos hiatos inicia em 1991:1. RAM é a revisão absoluta média. RRQM é a raiz da revisão quadrática média. R/S é a razão ruído-sinal (obtida pela razão entre RRQM e o desvio-padrão do hiato final). AR é a correlação serial de primeira ordem da série de revisão.

Contudo, não é possível afirmar de maneira não-ambígua que o efeito do aumento da amostra é responsável pela maior parte da RAM total, uma vez que, sem esse efeito, a RAM total seria de 0,76 p.p. Em outras palavras, na margem, o efeito do aumento da amostra elevou a RAM em 0,50 p.p. Raciocínio semelhante aplica-se aos casos dos indicadores RRQM e R/S. Por outro lado, a importância relativa da revisão de dados em relação ao aumento da amostra não é ambígua para os métodos TL (segundo o indicador RAM) e TQ (segundo todos os indicadores de magnitude). Em ambos os casos, embora o efeito da revisão de dados seja importante, há uma predominância do efeito do aumento da amostra. Além disso, por meio do indicador AR e da revisão média, pode-se perceber que a maior parte da persistência e do viés das séries de revisão total em todos os métodos é proveniente do efeito do aumento da amostra.

5 Conclusões

A partir da elaboração de um conjunto de dados de produção industrial em tempo real para o Brasil, examinamos vários indicadores que visam capturar características das revisões da PI e do hiato da PI. Os indicadores sugerem que as revisões de crescimento da PI são substanciais. Em termos absolutos, o crescimento mensal da PI é revisado, em média, 0,91 p.p. acima ou abaixo do valor inicialmente divulgado, apresentando mudança de sinal em 19% das vezes. Porém, ao aumentar o período de agregação, todos os indicadores de revisão tornam-se mais favoráveis. Por exemplo, quando a PI do mês é comparada com a PI do mesmo mês do ano anterior, a RAM reduz-se para 0,07 p.p. por mês, em média (0,86 p.p. ao ano).

A análise do crescimento da média móvel trimestral da PI mostra indicadores de revisão mais favoráveis do que os indicadores do crescimento da série original (mês/mês anterior). Porém, as revisões permanecem importantes. Em termos absolutos, o crescimento da média móvel trimestral é revisado, em média, 0,39 p.p. acima ou abaixo do valor inicialmente divulgado. Os indicadores de revisão da aceleração da PI também sugerem revisões substanciais. Em termos absolutos, a aceleração da PI é revisada, em média, 1,48 p.p. acima ou abaixo do valor inicialmente divulgado.

Na análise das revisões de hiato da produção industrial, utilizamos três métodos de extração de tendência: o filtro de Hodrick-Prescott (HP), a tendência linear (TL) e a tendência quadrática (TQ). Todos os métodos apresentaram revisões de magnitude elevada. Os métodos TL e TQ apresentaram tanto as maiores magnitudes de revisão, quanto os maiores vieses. Em dois dos três métodos (HP e TQ), as revisões implicam mudança de sinal do hiato em 30% ou mais das vezes, e a magnitude da revisão é maior do que a própria magnitude do hiato em aproximadamente 50% ou mais das vezes.

Em geral, tanto o efeito da revisão de dados da PI quanto o efeito do aumento da amostra são importantes para explicar as revisões totais dos hiatos da PI. Entretanto, há certa predominância do efeito do aumento da amostra. Tanto o fato de ambos os efeitos serem importantes, assim como o da predominância do efeito do aumento da amostra, está de acordo com os achados de vários autores para revisões de hiatos estimados a partir do PIB: Orphanides & Van Norden (2002) para os Estados Unidos; Cayen & Van Norden (2005) para o Canadá; e Bernhardsen et al. (2004, 2005) para a Noruega.

Como os dados disponíveis de produção industrial são usualmente mais recentes do que os dados de PIB, uma prática comum é utilizar o último dado de PI ou a última estimativa de hiato da PI para analisar o estado da economia ou a posição da economia no ciclo econômico. Porém, nossos resultados recomendam muita cautela nesse procedimento. Em tempo real, tanto a medida de produção industrial quanto as medidas analisadas de hiatos da PI apresentam importantes limitações. Potencialmente, isso tem implicações importantes também para estudos que utilizem a PI ou o hiato da produção industrial para analisar o comportamento passado da política monetária ou para prever a inflação. No primeiro caso, dado que os formuladores de política tomam suas decisões em tempo real, nossos resultados sugerem que pode não ser adequado utilizar a última série de dados disponível para efetuar uma análise das decisões passadas. Uma alternativa seria utilizar um conjunto de dados em tempo real. No segundo caso, os resultados sugerem que não é razoável comparar um novo modelo de previsão (estimado com os últimos dados disponíveis) com o desempenho de previsões de inflação realizadas em tempo real. À medida que o último conjunto de dados disponível tende a ser mais preciso, a comparação pode não ser justa.

Além disso, os resultados deste trabalho apontam para a importância de se utilizar um conjunto maior de séries econômicas para a análise do estado do ciclo econômico. Em princípio, a utilização de um grande conjunto de variáveis pode reduzir os riscos associados a dados sujeitos a revisões. Essa é, na verdade, uma prática comum dos bancos centrais. Por exemplo, nas atas das reuniões do Comitê de Política Monetária (Copom) do Banco Central do Brasil é evidente a utilização de um grande conjunto de informações sobre a atividade econômica e suas perspectivas.

Referências Bibliográficas

- Bernhardsen, T., Eitrheim, O., Jore, S. A. & Roisland, O. (2004), Real time data for norway: challenges for monetary policy, Discussion Paper Series 1: studies of the economic research centre. Deutsche Bundesbank 26.
- Bernhardsen, T., Eitrheim, O., Jore, S. A. & Roisland, O. (2005), 'Real time data for norway: challenges for monetary policy', *North American Journal of Economics and Finance* **16**, 333–349.
- Cayen, J. P. & Van Norden, S. (2005), 'The reliability of canadian output-gap estimates', *North American Journal of Economics and Finance* **16**, 373–393.
- Croushore, D. (2008), 'Frontiers of real-time data analysis', Working Paper, n.08-4. Research Department, Federal Reserve Bank of Philadelphia.
- Croushore, D. & Stark, T. (2000), 'A funny thing happened on the way to the data bank: a real-time data set for macroeconomists', *Business Review*. Federal Reserve Bank of Philadelphia.
- Croushore, D. & Stark, T. (2001), 'A real-time data set for macroeconomists', *Journal of Econometrics* **105**, 111–130.
- Diebold, F. X. & Rudebusch, G. D. (1991), 'A real-time analysis', *Journal of the American Statistical Association* **86**(415), 603–610.

Granger, C. W. J. & Jeon, Y. (2004), 'Thick modeling', *Economic Modelling* **21**, 323–343.

Mckenzie, R. & Park, S. Y. (2006), 'Revisions analysis of the index of industrial production for oecd countries and major non-member economies', OECD Short-term Economic Statistics Working Party.

Orphanides, A. & Van Norden, S. (2002), 'The unreability of output-gap estimates in real time', *Review of Economics and Statistics* **84**(4), 569–583.

Stark, T. (2002), 'A summary of the conference on real-time data analysis', *Business Review* **Q1**. Federal Reserve Bank of Philadelphia.