

5

Aplicação

Neste capítulo será apresentada a parte empírica do estudo no qual serão avaliados os principais regressores, um Modelo de Índice de Difusão com o resultado dos melhores regressores (aqui chamado de modelo proposto) e o Modelo de Índice de Difusão tradicional (aqui chamado de modelo tradicional). As previsões foram realizadas usando três horizontes de previsão, quais sejam: 1, 6 e 12 passos à frente. As estatísticas de Erro Quadrático Médio, Diebold-Mariano e Direction-of-change serão apresentadas.

Em todos os casos foi utilizada para previsão a abordagem de *Rolling Window approach with invariant specification* (Janela Rolante com especificação invariante). De acordo com Hollauer *et. al.* (2008), essa abordagem é um processo de estimação no qual se estima o modelo e se realiza previsões k passos à frente a cada observação seguinte na amostra; ou seja, o modelo é sempre atualizado, mas mantendo o tamanho da amostra fixa. Logo, os coeficientes estimados, seus testes individuais e coletivos e as estatísticas de adequação do modelo não serão apresentados, uma vez que os modelos serão reestimados a cada nova observação da amostra.

Assim, os modelos foram estimados de Agosto de 1994 a Dezembro de 2003, tendo a janela amostral de 113 observações e 64 observações fora da amostra a serem previstas.

5.1

Seleção dos melhores regressores univariados (modelo de Índice de Difusão univariado)

O objetivo fundamental nesta etapa inicial da modelagem foi de entender a natureza preditiva de cada variável. Para isso, utilizou-se da arquitetura do Modelo de Índice de Difusão Linear para escolher a melhor defasagem de cada preditor univariado do hiato do produto industrial, como abaixo:

$$\hat{y}_{t+h} = \hat{\beta}_0 + \sum_{i=1}^{11} \hat{\beta}_i y_{t-i} + \sum_{j=0}^{11} \hat{\beta}_j x_{t-j}$$

(34)

onde \hat{y}_{t+h} é o hiato do produto industrial e x_{t-j} cada uma das 74 variáveis presentes no banco de dados.

Assim, foram realizadas 9.768 regressões (74 variáveis * 11 defasagens de y * 12 defasagens de x), sendo 132 modelos para cada variável em cada um dos 3 horizontes de previsão. A escolha do melhor modelo univariado segue o critério de classificação pelas maiores significâncias das estatísticas de Direction-of-change e Diebold-Mariano e, em seguida, pelo menor Erro Quadrático Médio, respectivamente. Os gráficos a seguir mostram as previsões que não rejeitaram a hipótese nula do teste de Direction-of-change. Ou seja, aquelas que acertaram na mudança de comportamento do hiato do produto industrial. A Figura 3 mostra os resultados para as projeções 1 passo à frente ($h=1$) e a tabela pode ser vista no Anexo E.

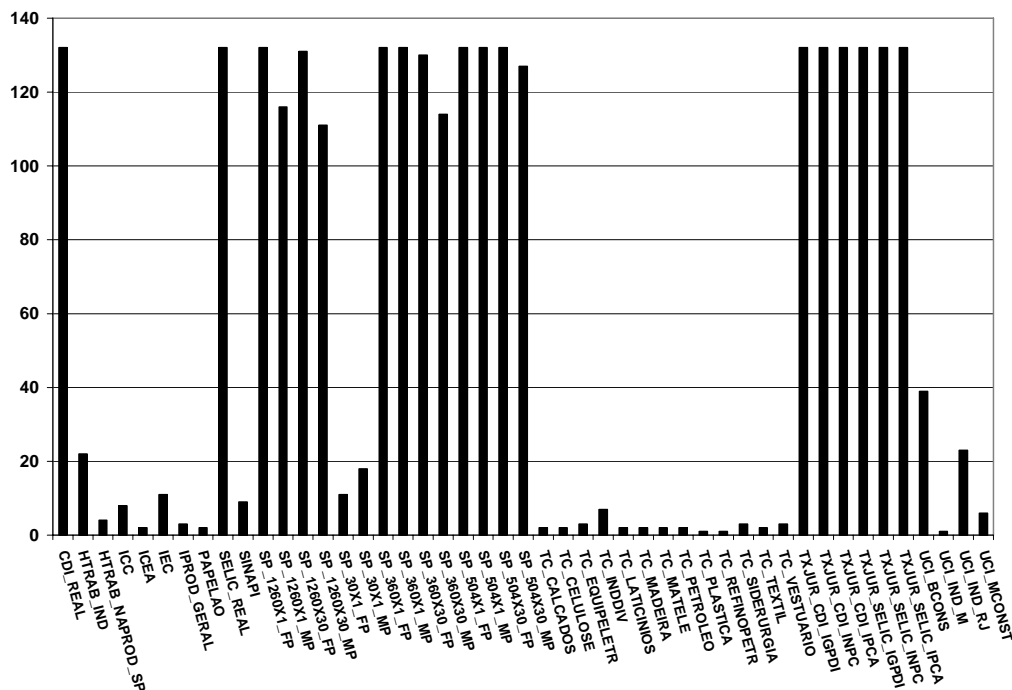


Figura 3: Frequência dos modelos que acertam na mudança de direção ($h=1$).

A Figura 3 mostra 47 variáveis com modelos que acertam na projeção do hiato do produto industrial, totalizando 2.768 regressões. Com relação às frequências, pode-se perceber que as variáveis CDI_REAL, SELIC_REAL, SP_1260x1_FP, SP_360x1_FP, SP_360x1_MP, SP_504x1_FP, SP_504x1_MP, SP_504x30_MP, TXJUR_CDI_IGPDI, TXJUR_CDI_INPC, TXJUR_CDI_IPCA, TXJUR_SELIC_IGPDI, TXJUR_SELIC_INPC e TXJUR_SELIC_IPCA são as que apresentam maior frequência nas projeções empatadas com 132. Em seguida, aparecem SP_504X30_MP, SP_1260X1_MP, SP_360X30_MP, além da SP_1260X30_MP, com 127, 116, 114 e 111 repetições, respectivamente. Analisando as séries, verifica-se sua natureza financeira com as variáveis de Taxa de Juros de Curto Prazo, *Spreads* e Taxa de Juros Real.

Analisando as demais variáveis, seus aparecimentos são considerados baixos quando comparados com as primeiras, sendo menores do que a metade. As variáveis de Taxa de Câmbio, por exemplo, somam 32, e as de Utilização da Capacidade Instalada, 69. Ou seja, a agregação da frequência de ambos os temas não alcança as mais frequentes.

Para as projeções 6 passos à frente ($h=6$), foram realizados os mesmos exercícios. A Figura 4 e o anexo F mostram os resultados.

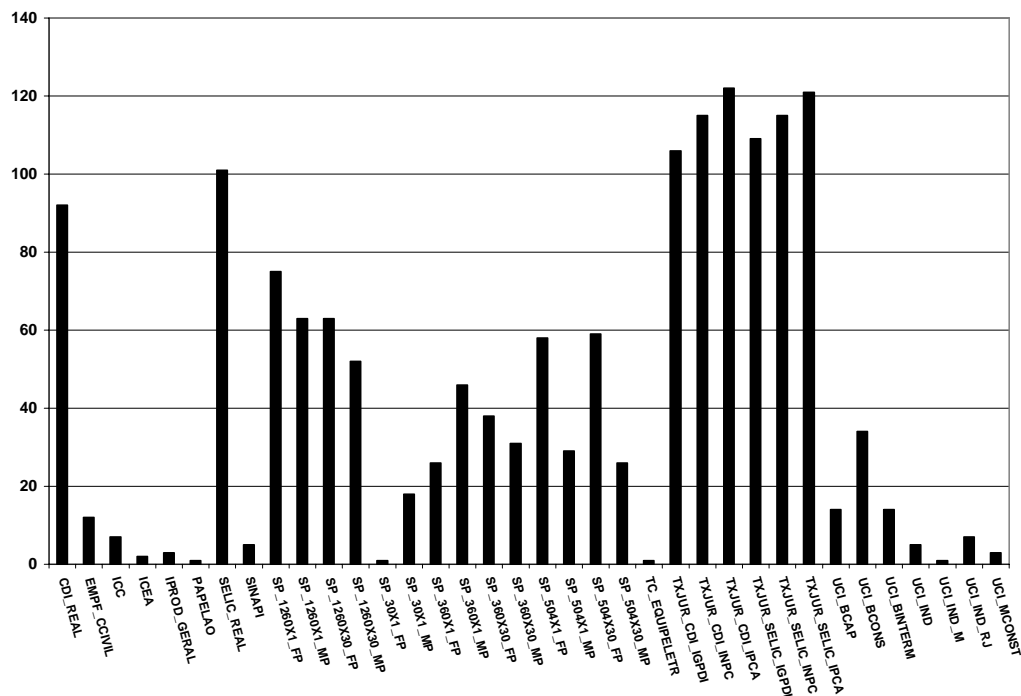


Figura 4: Frequência dos modelos que acertam na mudança de direção ($h=6$).

A Figura 4 mostra 36 variáveis com modelos que acertam estatisticamente na mudança de direção da variável dependente, com 1.575 regressões. Novamente as variáveis financeiras mostraram-se com maior frequência, tendo as variáveis relativas à Taxa de Câmbio Real sobrepondo-se às outras. Em seguida, aparecem as variáveis de Taxa de Juros de Curto Prazo e os *Spreads*, este último variando de 1 a 75 aparições.

As demais apresentaram frequência menor quando comparada com a dos primeiros. Cabe afirmar que a variável UCI_BCONS apresentou 34 repetições superando 6 das 14 referentes aos *Spreads*. O restante dos modelos não mostrou mais do que 15 aparecimentos.

Já a Figura 5 e o Anexo G mostram as 30 variáveis com modelos que acertam na mudança de direção, com 810 regressões para projeções 12 passos à frente. Mais uma vez as variáveis financeiras mostraram-se como as mais frequentes, sendo 24 delas. As demais apresentaram modelos com menos do que 5 repetições.

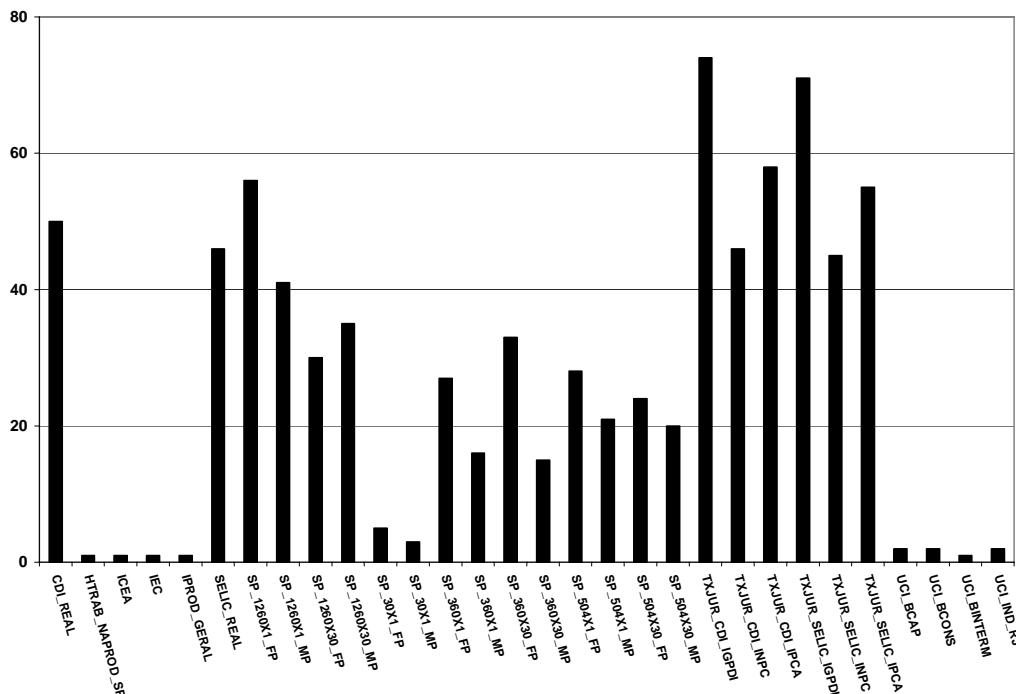


Figura 5: Frequência dos modelos que acertam na mudança de direção ($h=12$).

Assim, pode-se perceber que as variáveis referentes às taxas de juros e *spreads* são aquelas com maior capacidade preditiva nos três horizontes de previsão. Logo, antes de apresentar o Modelo de Índice de Difusão tradicional, foi realizado um modelo com as mesmas características, mas com somente essas variáveis.

5.2

O modelo de Índice de Difusão apenas com as melhores variáveis preditoras

O objetivo do uso de apenas as variáveis preditoras para o modelo de fatores é avaliar o desempenho preditivo deste quando comparado com o modelo tradicional. Para isto, foi utilizada a análise de fatores com as 22 variáveis representantes dos três temas. A Tabela 1 mostra as componentes com os autovalores acima da unidade.

Tabela 1: Extração dos fatores das melhores variáveis preditoras

Componentes	Total da Variância Explicada		
	Autovalores		
	Total	% da Variância	% acumulado da Variância
1	11.41425	51.88295	51.88295
2	5.26176	23.91708	75.80003
3	2.92978	13.31719	89.11722

A tabela sugere a extração dos 3 primeiros componentes, pois estes apresentam seus autovalores maiores do que a unidade e explicam 89% do total da variância.

Com isso, foi realizado o mesmo exercício anterior, realizando 396 regressões (3 fatores * 11 defasagens de y * 12 defasagens de x), como a seguir:

$$\hat{y}_{t+h} = \hat{\beta}_0 + \sum_{i=1}^{11} \hat{\beta}_i y_{t-i} + \sum_{i=1}^3 \sum_{j=0}^{11} \hat{\beta}_i F_{i,t-j}$$

(35)

onde \hat{y}_{t+h} é o hiato do produto industrial, e F_{t-j} , cada um dos 3 fatores extraídos das melhores preditoras.

A seleção do melhor modelo seguiu o mesmo critério da seção anterior e a tabela com os primeiros classificados é apresentada na Tabela 2.

Tabela 2: Melhores modelos com os fatores das variáveis financeiras

h	Nº de Fatores	Lag(y)	Lag(x)	Diebold-Mariano	d_signal	EQM	MAPE
1	1	7	6	3.854390	4.913538	0.000775	96.018870
6	2	6	8	1.975435	3.401680	0.001189	130.570200
12	3	7	5	2.293454	3.401680	0.001098	112.681100

Pode-se constatar a superação dos modelos em relação ao passeio aleatório e que acertam consideravelmente na mudança de direção do Hiato do Produto Industrial. Além disso, verifica-se um alto número de defasagens em y e x para se alcançar um bom ajuste na previsão. Cabe afirmar que quando previsto 6 passos à frente, existe uma menor rejeição da hipótese nula da estatística de Diebold-Mariano.

5.3

O modelo de Índice de Difusão para a produção industrial brasileira

Nesta seção serão realizadas as previsões dos Modelos de Índice de Difusão. Para isso foram extraídas as primeiras componentes da base de dados com as 74 variáveis tratadas inicialmente, seguindo o estudo de Stock e Watson (2002 a e b). A Tabela 3 mostra os autovalores com valores acima da unidade.

Tabela 3: Extração dos fatores das variáveis

Componentes	Autovalores		
	Total	% da Variância	% acumulado da Variância
1	28.14264	38.03059	38.03059
2	11.23023	15.17598	53.20657
3	8.654102	11.69473	64.90131
4	4.102724	5.54422	70.44553
5	3.334764	4.50644	74.95197
6	2.688076	3.63254	78.58450
7	2.639105	3.56636	82.15086
8	2.480588	3.35215	85.50300
9	1.424199	1.92459	87.42760
10	1.27129	1.71796	89.14556

Pode-se perceber que os dez primeiros fatores tiveram um poder de explicação significativo das 74 variáveis. No entanto, seguindo a teoria de Johnson e Wichern (1998), deve-se extrair os fatores cujo autovalor apresentasse acima da unidade, assim, decidiu-se fazer o mesmo exercício anterior com as 10 componentes. Com isso, foram estimados 1320 modelos (10 fatores * 11 defasagens de y * 12 defasagens de x), como a seguir:

$$\hat{y}_{t+h} = \hat{\beta}_0 + \sum_{i=1}^{11} \hat{\beta}_i y_{t-i} + \sum_{l=1}^{10} \sum_{j=0}^{11} \hat{\beta}_l F_{l,t-j} \quad (36)$$

Assim como antes, a seleção do melhor modelo obedeceu ao critério de classificação das regressões pelas estatísticas de Direction-of-change, Diebold-Mariano e pelo EQM, respectivamente. A tabela abaixo apresenta os resultados.

Tabela 4: Melhores modelos com os fatores das variáveis

h	Nº de Fatores	Lag(y)	Lag(x)	Diebold-Mariano	d_signal	EQM	MAPE
1	2	10	2	4.15941	5.66947	0.00074	94.10176
6	2	3	8	1.65860	3.90563	0.00114	130.06620
12	5	10	4	3.77917	4.15761	0.00111	113.22970

Apesar de necessitar de uma grande quantidade de defasagens do Hiato do Produto para prevê-lo, pode-se perceber que os modelos apresentam um elevado acerto na mudança de direção da variável dependente. No entanto, verifica-se que o melhor modelo para 6 passos à frente rejeita a hipótese nula de que o erro quadrático médio de suas previsões é diferente daquele do passeio aleatório. Ademais, para os outros horizontes de previsão, essa estatística é altamente significativa.

Para comparação dessas diferentes variações do Modelo de Índice de Difusão Linear, a próxima seção discutirá os resultados dos modelos por horizonte de previsão.

5.4

Resultados

O objetivo da comparação é verificar se os fatores com os melhores preditores considerados neste estudo apresentam melhor capacidade de prever o Hiato do Produto Industrial. Para isto, as tabelas a seguir utilizam das estatísticas de Diebold-Mariano e das medidas de Erro Quadrático Médio e MAPE.

Tabela 5: Comparação dos resultados para ($h=1$)

Modelo de Índice de Difusão	Nº de Fatores	Lag(y)	Lag(x)	Diebold-Mariano	d signal	EQM	MAPE
Fatores das Variáveis Financeiras	1	7	6	3.854390	4.91354	0.000775	96.01887
Fatores das Variáveis	2	10	2	4.15941	5.66947	0.00074	94.10176

Pode-se perceber na Tabela 5 o melhor ajuste do Modelo de Índice de Difusão tradicional em relação àquele somente com as variáveis financeiras. Verifica-se seu melhor ajuste pela maior significância das estatísticas de robustez de previsão e pelas medidas de desempenho.

A Tabela 6 também confirma a superioridade do modelo proposto por Stock e Watson (2002 a e b) para um horizonte de previsão de seis meses. Apesar da maior significância da estatística de Diebold-Mariano para o modelo proposto, todas as outras métricas indicam o modelo tradicional como superior.

Tabela 6: Comparação dos resultados para ($h=6$)

Modelo de Índice de Difusão	Nº de Fatores	Lag(y)	Lag(x)	Diebold-Mariano	d signal	EQM	MAPE
Fatores das Variáveis Financeiras	2	6	8	1.975435	3.40168	0.001189	130.57020
Fatores das Variáveis	2	3	8	1.65860	3.90563	0.00114	130.06620

Ademais, quando analisados os modelos projetando um horizonte de 12 meses, o tradicional mostra-se mais significativo nas estatísticas, porém, apresenta-se menos eficiente do que o modelo proposto.

Tabela 7: Comparação dos resultados para ($h=12$)

Modelo de Índice de Difusão	Nº de Fatores	Lag(y)	Lag(x)	Diebold-Mariano	d signal	EQM	MAPE
Fatores das Variáveis Financeiras	3	7	5	2.293454	3.40168	0.001098	112.68110
Fatores das Variáveis	5	10	4	3.77917	4.15761	0.00111	113.22970

Verifica-se, portanto, que existe uma pequena superioridade do modelo tradicional de Stock e Watson (2002 a e b) em relação ao modelo proposto. No entanto, seu custo computacional é bastante superior uma vez que utiliza 74 variáveis, enquanto o modelo proposto, apenas 22.

Com isto, pode-se concluir que as variáveis financeiras são suficientes para projetar o Hiato do Produto Industrial. A justificativa deve-se a elas serem as melhores antecedentes do Hiato do Produto Industrial e apresentarem uma capacidade preditiva comparável à projeção usando antecedentes da economia como um todo. Além disso, seu custo computacional para o tratamento dos dados (imputação de dados e transformações) é consideravelmente menor.