



Universidade Federal do Ceará

Curso de Pós Graduação em Economia-CAEN

Doutorado em Economia

THIBERIO MOTA DA SILVA

**PREVISÃO DE VARIÁVEIS MACROECONÔMICAS
UTILIZANDO MODELOS FATORIAIS**

FORTALEZA

2017

THIBERIO MOTA DA SILVA

**PREVISÃO DE VARIÁVEIS MACROECONÔMICAS
UTILIZANDO MODELOS FATORIAIS**

Tese submetida à Coordenação do Curso de Pós-Graduação em Economia da Universidade Federal do Ceará-UFC, como requisito parcial à obtenção do grau de Doutor em Economia.

Orientador: Prof. Dr. Roberto Tatiwa Ferreira.

FORTALEZA

2017

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

S584p Silva, Thiberio Mota da Silva.
PREVISÃO DE VARIÁVEIS MACROECONÔMICAS UTILIZANDO MODELOS FATORIAIS :
Modelos Fatoriais de Previsão / Thiberio Mota da Silva Silva. – 2017.
90 f. : il.

Tese (doutorado) – Universidade Federal do Ceará, Faculdade de Economia, Administração, Atuária e Contabilidade, Programa de Pós-Graduação em Economia, Fortaleza, 2017.
Orientação: Prof. Dr. Roberto Tatiwa Ferreira.
Coorientação: Prof. Dr. Rafael B. Barbosa.

1. Previsão. 2. Modelos de Fatores. 3. Variáveis Macroeconômicas. I. Título.

CDD 330

THIBERIO MOTA DA SILVA

PREVISÃO DE VARIÁVEIS MACROECONÔMICAS UTILIZANDO MODELOS FATORIAIS

Tese submetida à Coordenação do Curso de Pós-Graduação em Economia da Universidade Federal do Ceará-UFC, como requisito parcial à obtenção do grau de Doutor em Economia.

Orientador: Prof. Dr. Roberto Tatiwa Ferreira.

Aprovada em: ____/____/____.

BANCA EXAMINADORA

Prof. Dr. Roberto Tatiwa Ferreira
(Orientador)

Prof. Dr. Rafael B. Barbosa
(Co-orientador)

Prof. Dr. Luiz Ivan de Melo Castelar
(Membro)

Prof. Dr. Fabrício Carneiro Linhares
(Membro)

Prof. Dr. Elano Ferreira Arruda
(Membro-Externo)

AGRADECIMENTOS

Aproveito a oportunidade para agradecer as pessoas e a instituição que de forma direta e indireta colaboraram para minha formação acadêmica, representada pelo meu doutorado. Nesse sentido, agradeço aos professores Ivan Castelar, Elano Ferreira, Fabrício Linhares e, em especial, aos professores, Roberto Tatiwa e Rafael Barros, minha eterna e sincera gratidão pela paciência, pelo estímulo científico e dedicação a mim dispensada na elaboração deste trabalho.

Estendo os agradecimentos à minha mãe, Dona Fátima, pela sua bravura, perseverança e abdicção de sua vida em prol de seus filhos, à minha namorada, Joana Idayanne, pelo carinho, paciência, companheirismo, seriedade e amizade dado a mim em prol desse trabalho. Ao meu irmão, Jonathan Mota, e ao meu primo, Lucas Matos, pelo carinho e amizade, fundamentais na consecução desse objetivo, e a todos os amigos que conheci durante a minha vida acadêmica por contribuírem para conclusão desta etapa da minha vida e pelos momentos inesquecíveis que vivemos juntos.

Lista de Figuras

1	Ponderação do Fatores de Esquecimento	27
2	Model Confidence Set (Esquema Recursivo)	42
3	Model Confidence Set (Janela Móvel)	45
4	Diferença no erro quadrático médio de previsão da média histórica vis-à-vis um modelo de previsão (DEQMF)	85

Lista de Tabelas

1	Modelos	31
2	Variáveis Financeiras	33
3	Raiz do Erro Quadrático Médio (Esquema Recursivo)	41
4	Raiz do Erro Quadrático Médio (Esquema Janela Móvel)	44
5	Comparando as previsões nos esquemas recursivos e janela móvel	46
6	Raiz do Erro Quadrático Médio (Esquema Recursivo).	57
7	Raiz do Erro Quadrático Médio (Esquema Janela Móvel).	58
8	Raiz do Erro Quadrático Médio (Esquema Recursivo).	59
9	Raiz do Erro Quadrático Médio (Esquema Janela Móvel)	60
10	Modelos Supervisionados versus Não Supervisionados	61
11	Horizonte de previsão h=1	75
12	Horizonte de previsão h=3	76
13	Horizonte de previsão h=6	77
14	Horizonte de previsão h=12	78
15	Variáveis Financeiras	80
16	Resultados Excesso de Retorno sobre o S&P 500	83

Sumário

INTRODUÇÃO	12
CAPÍTULO 1: PROCEDIMENTOS METODOLÓGICOS.	14
1 Índices de Difusão.	14
2 <i>Statistical Learning</i> e métodos de regressão penalizada	16
2.1 Métodos <i>Shrinkage</i> para Seleção de Variáveis e Fatores	16
2.2 Métodos de Ponderação de Previsões.	18
2.2.1 Modelos Bayesianos Ponderados (BMA).	18
2.2.2 Mallows Model Averaging (MMA).	19
2.2.3 Leave-h-Out and Jackknife Model Averaging.	19
3 Supervisão de Fatores.	20
3.1 CFPC.	20
3 Avaliação da previsão	21
4 Instabilidades nos parâmetros	22
5 Modelo com Fatores Comuns e Média dinâmica de modelos (DMA).	23
5.1 Modelos com Fatores Comuns.	23
5.2 Média dinâmica de modelos (DMA).	26
5.3 Extensões.	29
5 Base de dados	31

**CAPÍTULO 2: PREVISÃO DE VARIÁVEIS MACROECONÔMICAS BRASILEIRAS
USANDO MODELOS DE SÉRIES TEMPORAIS DE ALTA DIMENSÃO. 34**

1 Introdução	36
2 Resultados	39
2.1 Previsões usando o esquema recursivo	39
2.2 Previsão usando esquema de janela móvel	42
2.3 Comparando as previsões nos esquemas recursivos e janela móvel	45
3 Conclusões	47
REFERÊNCIAS BIBLIOGRÁFICAS.	48

**CAPÍTULO 3- PREVISÃO DE VARIÁVEIS MACROECONÔMICAS
BRASILEIRAS USANDO MODELOS FATORIAIS
SUPERVISIONADOS. 52**

1 Introdução	54
2 Resultados	57
2.1 Modelos Não supervisionados.	57
2.1.2 Previsão usando o esquema recursivo.	57
2.1.3 Previsão usando esquema de janela móvel.	58
3. Modelos Supervisionados.	58
3.1 Previsão usando o esquema recursivo.	59
3.2 Previsão usando o esquema janela móveis	59
3.3 Modelos supervisionados versus não supervisionados.	60
4 Conclusão.	61
REFERÊNCIAS BIBLIOGRÁFICAS.	63

**CAPÍTULO 4- PREVISÃO POR MEIO DA COMBINAÇÃO
DINÂMICA DE MODELOS FATORIAIS. 66**

1 Introdução	68
2 Resultados Empíricos	72
2.1 Variáveis Macroeconômicas dos E.U.A.	72
2.2 Prevendo excesso de retorno do índice S&P 500	74
4.3. Exercício Empírico.	79
5 Conclusões.	85
REFERÊNCIAS BIBLIOGRÁFICAS.	87

CONCLUSÃO GERAL 91

RESUMO

O primeiro artigo analisa o desempenho de vários modelos de fatores de alta dimensão para prever nove variáveis macroeconômicas brasileiras, incluindo a taxa de desemprego, o índice de produção industrial, seis índices de preços e taxa de câmbio. Os fatores são extraídos de um conjunto de dados composto por 117 variáveis macroeconômicas. Métodos de shrinkage e combinações de previsões são utilizadas para selecionar ou combinar o melhor modelo fatorial de previsão. Os métodos de shrinkage utilizados foram Least Angle Regressions e Elastic Net, Bagging, Non Negative Garrote. Por sua vez, os métodos de combinações de previsões aplicados foram baseados no critério de Mallows, métodos bayesianos ponderados, validação cruzada, Leave-h-Out e modelo de ponderação de Jackknife. Todas as previsões foram avaliadas por meio do Model Confidence Set (MCS) que estabelece os melhores modelos de previsão que satisfaçam algum intervalo de confiança para o erro de previsão. Os resultados sugerem que, em geral, os modelos fatoriais apresentam um erro quadrático médio de previsão (MSFE) menor que o benchmark AR (4). Os melhores modelos de fatores para prever variáveis macroeconômicas brasileiras foram o Elastic Net e o Least Angle Regressions. Em algumas situações os modelos de ponderações Non Negative Garrote e Bayesiano apresentaram previsões satisfatórias. O segundo artigo analisa o desempenho dos modelos fatoriais supervisionados frente aos não supervisionados para prever quatro variáveis macroeconômicas brasileiras, dentre elas, o índice de produção industrial, o índice de preço ao consumidor amplo, o índice de nacional de preços ao consumidor e a taxa de juros de longo de prazo. Os fatores são extraídos de um conjunto de dados composto por 117 variáveis macroeconômicas. Os fatores foram extraídos por meio de combinação de fatores, no caso, dos modelos supervisionados, e no caso dos modelos não supervisionados, foi utilizado o método de previsões combinadas, o mesmo utilizado por Tu e Lee (2012). Os modelos de previsão utilizados foram: Modelos Autoregressivos Aumentados (FAAR), Modelos Bayesianos Ponderados (BMA), modelo Mallows ponderado (MMA), modelo jackknife ponderado (JMA), validação cruzada leave-h-out (LHO) e o modelo de média simples (SMA). O melhor modelo será o que apresenta a menor raiz quadrada do erro quadrático médio (REQM). Os resultados mostram que o esquema de janela móvel foi mais eficaz e o modelo de previsão que obteve melhor performance foi o modelo ponderado BMA, para ambos os modelos fatoriais, supervisionados ou não, além disso, os modelos supervisionados são mais eficazes para realizar previsões no médio prazo, pois previu com mais acurácia três, dentre as quatro, variáveis alvos, apresentando um menor (REQM). O terceiro artigo propõe um método de ponderação dinâmica de previsores (DMA) aplicado a grandes bases de dados. As variáveis contidas na base de dados têm a dimensão reduzida a um número r de fatores que são combinados dinamicamente por meio de fatores de esquecimento. A extração dos fatores considera também uma janela exponencial de esquecimento que visa reduzir o impacto de observações muito antigas. Mostra-se que tal modelo, chamado de FDMA, converge assintoticamente para a combinação dinâmica de fatores observados quando o número de variáveis ou o tamanho da amostra de fora cresce. Além disso, o modelo FDMA é aplicado em dois diferentes exercícios empíricos. Primeiro para prever seis variáveis macroeconômicas americanas, incluindo variáveis reais e nominais, e após para prever o excesso de retorno do S&P 500. Os resultados empíricos mostram que o FDMA e suas variações apresentam-se como uma alternativa promissora para previsão.

Palavras chaves: Previsão, Modelos de Fatores, Modelos de Fatores Dinâmicos e Variáveis Macroeconômicas.

ABSTRACT

The first article analyzes the performance of several high quality models to predict nine Brazilian macroeconomic variables, including unemployment rate, industrial production index, six price indices and exchange rate. The factors are extracted from a set of data composed of 117 macroeconomic variables. Reduction methods and combinations of forecasts are used to select or combine the best predictor factorial model. Regressions of minimum angle and Elastic Net, Bagging, Non-Negative Garrote. In turn, the methods of prediction combinations applied were based on the criterion of Mallows, weighted Bayesian methods, cross-validation, Leave-h-Out and Jackknife weighting model. All predictions were evaluated using the Model Condition Set (MCS) that establish the best forecast models that satisfy a confidence interval for the forecast error. The results suggest that, in general, the factorial models present a mean square error of prediction (MSFE) lower than the AR benchmark (4). The best models of factors to predict Brazilian macroeconomic variables. In some positions of the Non-Negative Garrote and Bayesian weighting models presented satisfactory predictions. The second article analyzes the performance of supervised versus unsupervised factorial models to predict four Brazilian macroeconomic variables, including industrial production index, broad consumer price index, national consumer price index and Long-term interest. The factors are extracted from a set of data composed of 117 macroeconomic variables. The models were extracted by means of combination of factors, no case, supervised models, and there are no cases of unsupervised models, was used by the method of combined forecasts, or even used by Tu and Lee (2012). The Autoregressive Models Augmented (FAAR), Weighted Bayesian Models (BMA), weighted Mallows model (MMA), weighted jackknife model (JMA), cross validation (SMA). The best model is the one that presents a smaller mean square error quadrant (REQM). The results show that the mobile window model was more capable and predictive model that obtained better performance for the BMA weighted model, for both factorial models, supervised or not, in addition, supervised models are more efficient to perform non-average predictions Three-year-old, among the four, target variables, presenting a minor (REQM). The third article proposes a method of dynamic data weighting (DMA) applied to large databases. How variables contained in the database have a size reduced to a number of factors that are dynamically combined through forgetting factors. The extraction of factors also considers an exponential window of exhaustion that aims to reduce the impact of very old observations. It is shown that model, called FDMA, converge asymptotically to a dynamic combination of observed factors when the number of variables or the size of the sample from outside grows. In addition, the FDMA model is applied in two different empirical exercises. 1. The empirical results show that the FDMA and its variations are considered as an alternative to the forecast.

Keywords: Forecasting, Factor Models, Dynamic Factor Models and Macroeconomic Variables.

INTRODUÇÃO

A previsão sempre foi um dos objetivos das análises quantitativas em economia. Em conjunto com a teoria econômica, diversas técnicas de previsão auxiliam a tomada de decisões por parte dos agentes envolvidos em atividades que necessitam de planejamento, avaliação de políticas e redução da incerteza. Um dos objetivos das previsões econômicas, a redução da incerteza, é de especial importância dentro do ambiente macroeconômico. A fim de alcançar esse objetivo, várias técnicas de previsão foram desenvolvidas, dentre elas, modelos de fatores.

O presente estudo analisa a eficiência das previsões de importantes variáveis econômicas, realizadas por modelos de fatores já existentes na literatura, bem como de uma técnica proposta nesse estudo. Nesse sentido, o primeiro capítulo desse trabalho verifica se modelos de fatores, nos quais os fatores e outras variáveis de controle adicionais, selecionadas através de métodos *shrinkage* e por meio de técnicas de ponderação apropriados a modelos de previsão com variáveis geradas, produzem previsões mais eficientes para um conjunto de variáveis da economia brasileira. Dentre as variáveis previstas nesse trabalho estão a taxa de inflação medida a partir de seis índices diferentes: índice geral de preços (IGP_10), índice nacional de custo de construção (INCC), índice nacional de preço ao consumidor (INPC), índice de preço ao consumidor (IPC), o índice de preço ao consumidor amplo (IPCA) e o índice de preços ao atacado (IPA_10). Também serão previstas as variáveis taxa de desemprego (U_taxa), índice de produção industrial (IPI) e a taxa de câmbio Brasil-EUA (real_exr). A avaliação dessas previsões foram realizadas por meio da raiz quadrática do erro quadrático médio de previsão (REQM) e por meio do método do *Model Confidence Set* (MCS). Os resultados encontrados apontam que os as técnicas LAR e *Elastic Net* são os melhores modelos de previsão para todas as variáveis macroeconômicas em ambos os esquemas, recursivo e de janela móvel.

O segundo capítulo analisa o desempenho dos modelos fatoriais supervisionados, comparados aos modelos fatoriais não supervisionados, tanto no esquema recursivo e de janela móvel. As variáveis previstas nesse artigo serão, o índice nacional de preço ao consumidor (INPC), o índice de preço ao consumidor amplo (IPCA), índice de produção industrial (IPI) e a taxa de juros de longo prazo (TJLP). Para estimar os fatores supervisionados utiliza-se a técnica utilizada por Tu e Lee (2016), (CFPC), e para estimar os fatores não supervisionados emprega-se o método dos componentes principais. A base utilizada nessa comparação consiste em 117 variáveis com frequência mensal no período de 1996.5 à 2015.12. Os resultados mostram que o esquema de janela móvel foi mais eficaz, além disso, o modelo de previsão que obteve melhor performance foi o modelo ponderado BMA. Por fim, os modelos supervisionados são mais eficazes para realizar previsões no médio prazo, pois previu com mais acurácia três, dentre as quatro, variáveis alvos previstas, apresentando um menor (REQM).

De acordo com Kim e Swanson (2014) a eficiência preditiva dos modelos fatoriais é melhorada quando combinada com métodos de seleção. Cheng e Hansen (2015) realizam previsões utilizando o mesmo conjunto de dados macroeconômicos para os EUA de Stock e Watson (2012) e relatam que a melhor previsão é obtida por meio da ponderação de previsões de modelos fatoriais. Logo, o terceiro artigo propõe o uso do método de combinação dinâmica de variáveis (DMA), desenvolvido por Raftery et al (2010), para selecionar os fatores e verificar se essa nova combinação de técnicas é capaz de melhorar a eficiência preditiva desse tipo de modelo. A eficiência desse método é avaliada empiricamente através de dois exercícios empíricos. O primeiro verifica a eficiência preditiva desse

modelo na previsão de seis variáveis macroeconômicas dos EUA: índice de preços ao produtor (IPI) e ao consumidor (CPI); índice de produção industrial (IP), M2, house starts (Hous) e taxa de desemprego (Unrate). Os fatores serão estimados de uma base de dados contendo 115 variáveis macroeconômicas, com frequência mensal entre 1963.1 até 2015.12 . As previsões do modelo proposto serão comparadas a modelos que são reconhecidos na literatura empírica como tendo boa performance em prever variáveis macroeconômicas utilizando variáveis geradas. Estes modelos de comparação são os que utilizam os critérios de Mallows (MMA), Leave-h-out e Jakkinfe (CHENG and HANSEN, 2015) para a seleção de fatores, métodos bayesianos de combinação de previsões (KOOP and POTTER, 2004 e KIM and SWANSON, 2014), previsões com pesos iguais, modelos fatoriais aumentados (STOCK and WATSON, 2002, 2012). Será considerada também as extensões em que o DMA é aplicado a fatores supervisionado.

CAPÍTULO 1: PROCEDIMENTOS METODOLÓGICOS.

O objetivo desse capítulo é apresentar ao leitor os modelos e as técnicas que serão utilizados nessa trabalho. Dessa forma, iniciamos com a descrição da metodologia aplicável ao problema analisados nos capítulos 2, 3 e 4 desse trabalho.

1 Índices de Difusão.

Considere um grande conjunto de dados macroeconômicos representados na matriz por X_{it} , em que $i = 1, \dots, N$ correspondem aos dados em corte transversal e $t = 1, \dots, T$ as unidades temporais. Em um modelo fatorial supõe-se que um pequeno número fatores $r \ll N$ possa representar a maior parte da variabilidade em X_{it} , isto é:

$$X_{it} = \lambda_i' F_t + e_{it} \quad (1)$$

Em que F_t é um vetor de fatores comuns, λ_i é um vetor de fatores de carga e e_{it} é uma matriz de componentes idiossincráticos. É possível representar (1) na forma matricial ao considerar $X_i = (X_{i1}, \dots, X_{iT})'$, $X = (X_1, \dots, X_T)'$, $\Lambda = (\lambda_1, \dots, \lambda_N)'$, $F_t = (F_1, \dots, F_T)'$, $e_i = (e_{i1}, \dots, e_{iT})'$ e $e = (e_1, \dots, e_N)'$,

$$\underset{(T \times N)}{X} = \underset{(T \times r)}{F} \underset{(r \times N)}{\Lambda}' + \underset{(T \times N)}{e} \quad (2)$$

Métodos de previsão que utilizam fatores comuns como variáveis de controle, chamados de índice de difusão, buscam reduzir a dimensão das variáveis contidas em variáveis X_{it} para r variáveis. O método mais utilizado para a extração dos fatores¹ é o de componentes principais (PC). Bai (2003) mostra que mesmo diante de limitada dependência transversal entre os dados e o erro na presença de heteroscedasticidade, a estimação por PC é consistente.

Uma vez estimado os fatores, as previsões para a variável alvo y_t podem ser estimadas por meio da seguinte equação:

$$y_{t+h} = \beta(L) \widehat{F}_t + \gamma(L) W_t + \epsilon_{t+h} \quad (3)$$

Em que $\beta(L)$ e $\gamma(L)$ são defasagens polinomiais com ordens p e q , respectivamente, e $W_t = (X_t, y_{t-j})$ é um conjunto formado por variáveis exógenas da matriz X_t e por variáveis predeterminadas representadas por valores passados da variável a ser prevista, y_{t-j} , $t = 1, \dots, T$. O horizonte de previsão é h e indica quantos períodos à frente será realizada a previsão.

¹Veja Bai e Ng (2002, 2003, 2008, 2009), Stock e Watson (2002, 2005, 2011, 2012) para os detalhes do procedimento de estimação.

Note que em um modelo de fator, o número desses componentes é dado a priori. Esse número também pode ser selecionado através de critérios de informação, como o proposto por Bai e Ng (2002). Um outro problema refere-se ao número de variáveis utilizadas na estimação dos fatores. De acordo com Boivin e Ng (2006) nem sempre utilizar o maior número de variáveis disponíveis para estimar os fatores geram as previsões mais eficientes.

Entretanto, aqui os métodos de seleção serão aplicados após a estimação dos fatores. Autores como Stock e Watson (2002), Kim e Swanson (2014) e Cheng e Hansen (2015) mostram que as previsões de modelos de fatores utilizando tal procedimento melhoram o poder de previsão dos modelos fatoriais.

Considere novamente a equação (3). Assuma que exista um número máximo de defasagens, tal que: $0 < p < p_{max}$ e $0 < q < q_{max}$. Suponha todos os regressores de (3) possam ser incluídos numa única variável $z(\hat{F}_t, W_t) = (y_1, \dots, y_T, x_{11}, \dots, x_{NT}, \hat{F}_1, \dots, \hat{F}_T)$. Então, (3) pode ser reescrito como função de z ,

$$y_{t+h} = z_t' b + \epsilon_{t+h} \quad (4)$$

Em que: $b = (\beta(L), \gamma(L))$. Suponha que existam M modelos derivados da combinação de todas as variáveis contidas em z . Indexando cada modelo por $m = 1, \dots, M$ em que o índice m representa um subconjunto de z . Então,

$$y_{t+h}(m) = z_t(m)' b(m) + \epsilon_{t+h}(m) \quad (5)$$

A equação (5) é estimada por MQO considerando os regressores $\tilde{z}(m) = (\tilde{F}_t, W_t)$ e $\hat{b}(m)$ são os parâmetros estimados. A variável \tilde{F}_t indica que os fatores foram estimados por PC. Então, $\tilde{z}(m)$ indica que dentre as variáveis contidas em z existem algumas que foram anteriormente estimadas. A previsão para o período $t + h$ dado o conjunto de informação em T é:

$$\hat{y}_{t+h|T} = \tilde{z}(m)' \hat{b}(m) \quad (6)$$

Por sua vez, o erro de previsão é dado por $\hat{\epsilon}_{t+h|T} = \hat{y}_{t+h|T} - \tilde{z}(m)' \hat{b}(m)$. Uma combinação de previsões de (6) é obtida através da ponderação das previsões $\hat{y}_{t+h|T}$, a qual pode ser expressa como:

$$\hat{y}_{t+h|T}(w) = \sum_{m=1}^M w(m) \hat{y}_{t+h|T}(m) \quad (7)$$

Em que, $w(m) = (w(1), \dots, w(M))'$ denota o vetor de pesos. Assume-se que $0 < w(m) < 1$ e $\sum_{m=1}^M w(m) = 1$. Por fim, os resíduos da combinação de previsões são obtidos por $\hat{\epsilon}_{t+h}(w) = \sum_{m=1}^M w(m) \hat{\epsilon}_{t+h|T}(m)$. A escolha dos melhores modelos nas equações (5) e (6)

e dos pesos dos métodos de ponderação da equação (7) são baseadas na minimização de uma função perda. As próximas duas subseções, apresentam as técnicas de ponderação de previsões e de variáveis utilizadas nesse estudo.

2 *Statistical Learning* e métodos de regressão penalizada

2.1 Métodos *Shrinkage* para Seleção de Variáveis e Fatores

Na técnica de seleção de variáveis através de agregação via re-amostragem (*bootstrap aggregation* - bagging), a equação (6) ou uma similar é estimada várias vezes através de re-amostragem via *bootstrap*. Em cada passo, os parâmetros insignificantes são descartados. Após a seleção de B^* modelos cujas estimativas dos seus parâmetros são significativas, a previsão é realizada através da média das estimativas, isto é, $\hat{y}_{t+h}^{Bagging} = \frac{1}{B} \sum_{i=1}^B \hat{y}_i^*$. Similarmente a Stock e Watson (2012) e Kim e Swanson (2014), neste artigo será assumido o valor crítico de pré-teste para o Bagging de $c = \pm 1,96$. Vale ressaltar, que este procedimento tem sido aplicado em modelos de previsão com regressores gerados, como é o caso deste artigo, ou em regressões correlacionadas como em Inoue e Kilian (2008), Stock e Watson (2012), Kim e Swanson (2014).

Outra forma de seleção de variáveis é o método do *Least Angle Regression* (LAR) que foi proposto inicialmente por Efron *et al* (2004) como um método de seleção de variáveis similar ao *Forward Stepwise Regression*². Esta técnica, primeiramente, ordena as variáveis candidatas a participar do modelo de acordo com seu poder preditivo. Depois, seleciona, entre as variáveis melhor ranqueadas, um modelo parcimonioso e usa este para realizar a previsão. Em comparação com outros métodos de seleção, o LAR é considerado mais eficiente computacionalmente do que o LASSO, desenvolvido por Tibshirani (1996). O algoritmo do LAR é rápido e possui previsão acurada em diferentes estruturas de dados³.

Gelper e Croux (2008) estendem o algoritmo do LAR para um contexto em séries temporais. A ideia principal da abordagem é selecionar os previsores em blocos compostos por séries temporais defasadas e não defasadas. Portanto, ao invés de considerar apenas um conjunto de variáveis, a seleção é feita sobre tais blocos. O que implica que caso um bloco de variáveis seja selecionado, então todas as variáveis contidas nele serão utilizadas para a previsão. Kim and Swanson (2014, 2016) usam a metodologia de Gelper and Croux (2008) para regressores gerados, como os fatores. A equação (3) pode ser escrita como,

$$y_{t+h} = \alpha(L)y_t + \sum_{i=1}^r \beta_i(L)F_t^i + \sum_{j=1}^k \gamma_j(L)x_{jt} + \epsilon_{t+h} \quad (8)$$

²*Forward Stepwise Regression* constrói um modelo por meio da adição de novas variáveis a cada processo de estimação. Após, cada novo conjunto de variáveis é utilizado para se realizar a previsão.

³Ver Efron *et al* (2004), Bovesstad *et al* (2007), Saigo *et al* (2007), Gelper e Croux (2008).

Onde é assumido que existe r fatores e k covariadas. Para aplicar o LAR, primeiro se ordena os preditores incluindo um por vez. A cada inclusão, é computado o erro de estimação por MQO, isto é, ϵ_l , para $l = 0, 1, \dots, p + r \times q + k \times h$, onde p , q e h são as respectivas as defasagens da variável alvo, dos fatores e das covariadas. Cada variável é incluída no conjunto ativo, conjunto de variáveis utilizadas para realizar a previsão, de acordo com sua correlação com a variável a ser predita. Então, o modelo estimado é obtido por critérios de informação. Seguindo Gelper e Croux (2008) o número de defasagens é escolhido ao se minimizar o BIC. Aqui será adotado metodologia semelhante a Gelper e Croux (2008).

Uma outra alternativa para selecionar variáveis ou determinar os pesos das combinações é o método *Elastic Net* proposto por Zou e Hastie (2005). Esse método permite que variáveis fortemente correlacionadas estejam no conjunto de informação inicial e escolhe o modelo que minimiza a seguinte soma dos quadrados dos resíduos:

$$L(\eta_1, \eta_2, b) = |y_t - z_t' b|^2 + \eta_1 |b|_1 + \eta_2 |b|^2 \quad (9)$$

Em que: $|b|^2 = \sum_j^N (b_j)^2$ e $|b|_1 = \sum_j^N |b_j|$, são restrições normadas de ordem l_1 e l_2 . Os parâmetros η_1 e η_2 , conhecidos como *tunning parameters*, indicam o grau de penalização sobre o processo de minimização de (9). Este valores são obtidos, normalmente, por validação cruzada, a qual consiste em separar a amostra em k subamostras. Uma das subamostras é chamada de *training set* (k^*). Então, é realizado um *grid search* para valores dos *tunning parameters*. Os valores desses parâmetros que minimizam a equação(9) na amostra são reservados. O procedimento se repete para todas as subamostras. As estimativas finais desses coeficientes são os que geraram melhores resultados em todas as subamostras. Neste artigo é utilizado metodologia semelhante a proposta por Bai and Ng (2008), que foi a primeira versão do *Elastic Net* aplicada a séries temporais econômicas.

Breiman (1995) propôs o método *Non-negative Garotte* (NNG), como um método de regularização que realiza uma regressão com penalização sobre o estimador de mínimos quadrados. Considere o modelo em (4), então o estimador de NNG minimiza,

$$\frac{1}{2} \|y_t - Pv\|^2 + T\xi \sum_{j=1}^N v_j \quad (10)$$

sujeito a $v_j > 0$, $j = 1, \dots, N$

Onde $P = (P_1, \dots, P_N)'$, $P_j = Z_j \hat{b}_j^{LS}$, \hat{b}_j^{LS} é o estimador de MQO, $v(\xi) = (v_1(\xi), \dots, v_N(\xi))'$ é o fator regularização e ξ é o *tunning parameter*. O estimador NNG é definido por: $\hat{b}_j^{NNG}(\xi) = v_j(\xi) \hat{b}_j^{LS}$, $j = 1, \dots, p$. Tal estimador possui duas limitações. Primeiro, apresenta uma dependência explícita em relação ao estimador de MQO. Portanto, situações em que o estimador de MQO não funciona adequadamente, poderá afetar negativamente

as propriedades estatísticas do estimador de NNG. Por exemplo, em pequenas amostras o estimador de NNG pode ser prejudicado devido a inconsistência do MQO.

Segundo, similarmente a outros métodos de regularização, o NNG é computado usando técnicas de programação quadrática para dado *tunning parameter*. Isto pode ser um problema se a amplitude do *grid* para o *tunning parameter* for pequena. Entretanto, apesar dessas ressalvas, Yuan e Lin (2006) mostram que, em geral, o NNG produz estimativas consistentes dos parâmetros e produz adequada seleção de variáveis. Zhang et al (2014), usando técnicas de simulação, encontraram que o estimador de NNG funciona apropriadamente para previsão, entretanto, a sua habilidade de seleção de variáveis é prejudicada em modelos com alta dimensão. Zou (2006) mostra que o NNG é consistente, se $T \rightarrow \infty$ com N fixo. Kim e Swanson (2014) aplicam o NNG similarmente ao que será aplicado no presente artigo.

2.2 Métodos de Ponderação de Previsões.

2.2.1 Modelos Bayesianos Ponderados (BMA).

Seguindo Stock e Watson (2012) é possível descrever a representação generalizada de vários modelos que incluem métodos de encolhimento e bayesianos ponderados para um grande número de previsores por:

$$\hat{y}_{t+h|T} = \sum_{j=1}^r \psi(kt_j) \hat{b}_{F_j} \hat{F}_{t,j} + o_p(1) \quad (11)$$

Em que, \hat{b}_{F_j} é a estimativa de MQO e os resíduos são: $\epsilon_{t+h} = y_{t+h} - z_t' \hat{b}$ sobre $\hat{F}_{t,j}$ e t_j é a estatística do teste t com valor crítico c e variância estimada por método de Newy-West. O modelo bayesiano ponderado⁴ (BMA) constroi \hat{m} estimado a ponderação de $m = 1, \dots, M$ modelos candidatos.

Seja $p(b | Z)$ a probabilidade do parâmetro b condicional as informações contidas em Z , como segue:

$$P(b | Z) = \sum_{m=1}^M p(b | Z, \hat{m}) p(m | Z) \quad (12)$$

Os pesos são escolhidos de forma a maximizar $p(b | Z)$. Considere uma função de b , dada por $g(b)$. A expectativa condicional de g em relação ao conjunto de dados em Z é:

$$E[g(b) | Z, \hat{m}] p(\hat{m} | Z) \quad (13)$$

⁴Para uma concisa introdução dos BMA veja Hoeting et al (1999) e Chipman et al (2001). Para aplicações empíricas dos BMA em previsão ver Koop and Potter (2004), Wright (2008,2009) e Kim and Swanson (2014, 2013).

Seguindo Kim and Swanson (2014, 2016) e Koop e Portter (2004), duas funções são possíveis para g : $g = (1/T)$ e $g = (1/N^2)$. O modelo de previsão BMA pode ser expresso por:

$$\hat{y}_{t+h}^{*BMA} = \hat{b}\tilde{Z}_t \quad (14)$$

Em que, $\hat{y}_{t+h|T}^{*BMA} = [I_T - Z_t(Z_t'Z_t)^{-1}Z_t']\hat{y}_{t+h}^{*BMA}$.

2.2.2 Mallows Model Averaging (MMA).

Hansen (2007) propôs selecionar modelos utilizando o critério de Mallows (1973) e mostrou que estes são assintoticamente ótimos no sentido de produzirem erros de previsão pequenos. Hansen (2008) mostrou que os Modelos de Mallows ponderados (MMA) possuem relativo sucesso em previsão e em Hansen (2012) obteve-se resultados semelhantes para $h > 1$ e heteroscedasticidade condicional. Por m, Cheng e Hansen (2015) estudam o comportamento do MMA utilizando regressores gerados como os neste artigo. Assim, será adotado metodologia semelhante.

O critério de Mallows é definido como a média quadrática do erro de previsão mais um termo que depende do tamanho da amostra, da variância do erro de previsão e do número de regressores contidos no m -ésimo modelo.

$$C_t(m) = \frac{1}{T} \sum_{t=1}^T \hat{\epsilon}_t(m)^2 + \frac{2\hat{\sigma}_T^2}{T} k(m) \quad (15)$$

Em que $k(m)$ é o número de regressores em cada modelo $m = 1, \dots, M$; $\hat{\sigma}_T^2$ é a variância estimada do erro de previsão $\sigma^2 = E(\epsilon_t^2)$. Seguindo Cheng and Hansen (2015), o estimador da variância será dado por $\hat{\sigma}_T^2 = (T - k(m))^{-1} \sum_{t=1}^T \hat{\epsilon}_t(M)^2$, em que M é o maior modelo. $\hat{w} = \operatorname{argmin}_{w \in W^M} C_t(m)$, em que W^M é o simples unitário⁵ em \mathbb{R}^M e

$$C_t(m) = \frac{1}{T} \sum_{t=1}^T \left(\sum_{m=1}^M w(m) \hat{\epsilon}_t(m) \right)^2 + \frac{2\hat{\sigma}_T^2}{T} \sum_{m=1}^M w(m) k(m) \quad (16)$$

2.2.3 Leave-h-Out and Jackknife Model Averaging.

Modelos ponderados Jackknife (JMA) foram propostos por Hansen e Racine (2012) em que m modelos são estimados e a última observação é deixada fora da amostra em cada procedimento. Os pesos $\hat{w}(m)$ minimizam $\hat{w} = \operatorname{argmin}_{w \in W^M} CV_1(w(m))$, em que:

$$CV_1(m) = \frac{1}{T} \sum_{t=1}^T \tilde{\epsilon}_{t,1}^2(m) \quad (17)$$

⁵Se $w \in W^M$, então $0 \leq w(m) \leq 1$ e $\sum_{m=1}^M w(m) = 1$, $\forall m = 1, \dots, M$

Em que, $\tilde{e}_{t,1}$ é o resíduo obtido pela estimação de mínimos quadrados⁶ com a observação t omitida. Hansen and Racine (2012) mostram que JMA e MMA são aproximadamente equivalentes na presença de homocedasticidade. Entretanto, JMA tem melhor performance do que o MMA na presença de heteroscedasticidade.

O comportamento dessas combinações de previsão não são os melhores diante de horizontes $h > 1$. Assumindo que os erros seguem um processo $MA(h - 1)$, Hansen (2010) mostra que o erro de previsão médio quadráticos (MSFE) depende do horizonte de previsão (h), dos parâmetros de $MA(h - 1)$ e do tamanho da amostra. Para superar esse problema, Hansen (2010) propôs o método de validação cruzada leave-h-out. Os pesos são obtidos:

$$CV_h(m) = \frac{1}{T} \sum_{t=1}^T \tilde{e}_{t,h}^2(m) \quad (18)$$

Em que $\tilde{e}_{t,h}$ são os resíduos da estimação por MQO com a omissão de $t - h + 1, \dots, t + h - 1$ observações. No mesmo artigo, Hansen demonstrou que tal critério tem performance superior que o JMA, MMA e outros critérios clássicos.

3 Supervisão de Fatores.

3.1 CFPC.

O método do CFPC foi proposto por Tu e Lee (2012) e tem o objetivo de estimar os fatores de uma base de dados que considere a contribuição de cada variável para a previsão da variável alvo. Assim, o CFPC consiste em estimar em primeiro estágio a regressão de cada variável contida em X contra a variável alvo, isto é:

$$y_t = x_{it}b_i + e_{it} \quad (19)$$

Em que: $b_i = (x'_{it}x_{it})^{-1}x'_{it}y_{it}$ é a estimativa de MQO e x_{it} é o vetor formado pela i -ésima coluna de X . E a previsão de y_t h períodos à frente como: $y_{it+h} = x_{it}b_i$. Considere todas as previsões a partir de cada variável x_i e para o horizonte de previsão h inseridas na matriz $\hat{Y} = [\hat{y}_{1t+h}, \hat{y}_{2t+h}, \dots, \hat{y}_{Nt+h}] = x_{1t}b_1, \dots, x_{Nt}b_N = XB$.

Portanto, o método dos componentes principais (PCM) pode ser aplicado diretamente em \hat{Y} ao invés de X . Assim, o poder de previsão de cada variável sobre y_t é considerado durante a extração dos fatores.

⁶Hansen and Racine (2012) argumentam que a otimalidade assintótica do JMA requer estimadores lineares, não necessariamente MQO. Outros métodos que satisfazem a linearidade são: Ridge regression, Nadaraya-Watson, nearest neighbor estimators, series estimators, additive interaction models and spline estimators.

3 Avaliação da previsão

Seja função perda quadrática,

$$L_{mt} = (y_{t+h} - \hat{y}_{m,t+h})^2 \quad m = 1, \dots, M \quad (20)$$

Onde m representa o m –ésimo modelo de previsão. A raiz quadrática do erro quadrático médio de previsão (REQM) é definida por:

$$REQM = \sqrt{\sum_{j=t+h}^T (y_{t+h} - \hat{y}_{m,t+h})^2} \quad m = 1, \dots, M$$

Os melhores modelos de previsão são aqueles que apresentam menor REQM. Assim, uma das medidas do poder de previsão dos modelos utilizadas neste trabalho será o REQM.

Como existem quatorze modelos de previsão utilizados para prever as variáveis macroeconômicas, então, caso os modelos fossem avaliados por meio de testes *à la* Diebold-Mariano, seriam necessários $2^{14} = 16384$ testes. O método do *Model Confidence Set* (MCS), desenvolvido por Hansen et al (2011) consiste numa sequência de testes que permitem construir o conjunto contendo os modelos com habilidade superior de previsão (SSM) no qual a hipótese nula de igual poder de previsão não é rejeitada ao nível de significância α .

Seja y_t a variável alvo no tempo t e \hat{y}_{it} a previsão pontual do i –ésimo modelo no tempo t . Reescreva a função perda em (18) por:

$$L_{it} = L(\hat{e}_t)_i = L(y_t, \hat{y}_{it}) \quad (21)$$

Em que, \hat{e}_t é uma medida do erro de previsão. Considere o conjunto de \hat{M}_0 modelos de dimensão m . O procedimento do MCS produz, dado certo nível de confiança $1 - \alpha$ o menor conjunto SSM, $\hat{M}_{1-\alpha}^*$ de dimensão $m^* \leq m$. Denote $d_{ij,t}$ o diferencial de perda entre os modelos i e j no tempo t .

$$d_{ij,t} = L_{it} - L_{jt}, \quad i, j = 1, \dots, m \quad t = 1, \dots, T$$

Seja a perda média do modelo i relativo a outro modelo j dada por:

$$d_{i.,t} = (m - 1)^{-1} \sum_{j \in M} d_{ij,t}, \quad i = 1, \dots, m.$$

A hipótese de igual habilidade preditiva (EPA) para dado conjunto de modelos M é definida por $c_{ij} = E(d_{ij})$ e $c_i = E(d_i)$

$$H_{o,M} : c_{ij} = 0, \text{ para todo } i, j = 1, 2, \dots, m$$

$$H_{A,M} : c_{ij} = 0, \text{ para todo } i, j = 1, \dots, m.$$

ou

$$H_{0,M} : c_i = 0, \text{ para todo } i = 1, 2, \dots, m$$

$H_{A,M} : c_i = 0$, para todo $i = 1, \dots, m$.

Hansen et al (2011) propuseram duas estatísticas de teste para as hipóteses descritas acima, dadas por:

$$t_{ij} = \frac{\bar{d}_{ij}}{\sqrt{\hat{var}(\bar{d}_{ij})}} et_i = \frac{\bar{d}_i}{\sqrt{\hat{var}(\bar{d}_i)}}$$

Onde $\bar{d}_i = (m - 1)^{-1} \sum_{j \in M} \bar{d}_{ij,t}$ é a perda média do i -ésimo modelo relativo a média

de perda entre os modelos do conjunto M, e $\bar{d}_{ij} = m^{-1} \sum_{t=1}^m \bar{d}_{ij,t}$ mensura perda amostral relativa entre o i -ésimo e j -ésimo modelos, $\hat{var}(\bar{d}_{ij})$ e $\hat{var}(\bar{d}_i)$ são estimativas de bootstrap de $var(\bar{d}_{ij})$ e $var(\bar{d}_i)$, respectivamente.

O MCS consiste em um procedimento sequencial de testes em que a cada etapa, o pior modelo é eliminado até que a hipótese EPA não seja rejeitada para todos os modelos restantes. Sumarizando, os passos do MCS:

1. Considere o conjunto de modelos $M = M_0$.
2. Teste para a hipótese EPA: se EPA é não rejeitada, o algoritmo se encerra e $M_{1-\alpha}^* = M$, caso contrário usa a regra de eliminação do pior modelo.
3. Remove-se o pior modelo e retorne a etapa 2.

Assim, o modelo resultante deste algoritmo $M_{1-\alpha}^*$ é o conjunto cujos modelos possuem igual poder preditivo entre si e habilidade superior de previsão em relação aos modelos fora do conjunto final. Um modelo será considerado ideal para a previsão das variáveis macroeconômicas brasileiras se estiver na maioria (ou na totalidade) do MCS. Observe que um REQM baixo em certas situações não é suficiente para um modelo estar entre os melhores previsores para todas as variáveis. Ressalta-se que o MCS será utilizado apenas no primeiro capítulo.

4 Instabilidades nos parâmetros

O Brasil, assim como outros países emergentes, é sujeito a muitos choques macroeconômicos seja em virtude de choques externos⁷ ou por instabilidades internas. Resulta que as séries macroeconômicas apresentam elevados níveis de volatilidade e são sujeitas a quebras estruturais⁸.

⁷Gourinchas and Obstfeld (2012) apontam que as principais causas da maior vulnerabilidade dos países emergentes em comparação aos avançados são: instabilidades políticas e econômicas internas; mercado financeiro instável e pouco desenvolvido; dolarização, pecado original e *currency mismatches*; medo de flutuar; intolerância ao débito, *sudden stops* e sobrevalorização do setor não financeiro. Estudos que apontam essas peculiaridades dos países emergentes, ver referências dos autores.

⁸Existem diversos artigos que indicam as causas da volatilidade nas séries macroeconômicas de países emergentes. Ver Neumeyer and Ferri (2005), Aguiar and Gopinath (2007), García-Cicco et al (2010) e Shousha (2016).

Métodos de previsão baseados em séries temporais que apresentam instabilidade nos parâmetros ainda é uma questão que não foi devidamente resolvida na literatura, como apontam Elliot e Timmermann (2016). Todavia, existem procedimentos de estimação que permitem minimizar o efeito das instabilidades sobre as previsões.

Serão adotados dois esquemas de previsão que tentam mitigar o impacto das instabilidades, o esquema recursivo e o de janela móvel. Assuma que as observações temporais possam ser indexadas por $i = 1, \dots, T$. Chama-se a amostra sobre a qual será realizada a estimação de *in-sample*, cujo tamanho é $P \leq T$. Por sua vez, a amostra sobre a qual são realizadas as previsões é chamada de *out-of-sample*. Sendo a *out-of-sample* de tamanho R , então $T = P + R$.

Considere a estimação da previsão em h períodos à frente dado o conjunto de informação Θ , $y_{t+h|\Theta}$. No esquema recursivo serão estimados os fatores na amostra $t = 1, \dots, P$ e a previsão é realizada para $t + h$, assumindo que $\Theta = \{1, \dots, P\}$. Posteriormente, a primeira observação do *out-of-sample* é incluída na *in-sample* de forma que a *in-sample* constitua $t = 1, \dots, P + 1$, enquanto a *out-of-sample* $R - 1$. O processo de estimação é realizado novamente sobre o conjunto de informação $\Theta = \{1, \dots, P, P + 1\}$. Este procedimento continua até que sejam realizadas previsões para todas as observações da *out-of-sample*.

O outro esquema de previsão utilizado neste trabalho é o de janela móvel (*rolling window*). Neste esquema, a *in-sample* tem um tamanho fixo dado por P e o conjunto de informação, embora sempre com o mesmo tamanho, varia ao incluir e excluir observações ao longo do processo de previsão. Desta forma, para prever dado o conjunto de informação $\Theta = \{1, \dots, P\}$, $y_{t+h|\Theta}$, usa-se a amostra de tamanho P . Após este primeiro procedimento, a primeira observação da *in-sample* é eliminada e a primeira observação da *out-of-sample* é acrescentada a *in-sample*, de forma que o novo conjunto de informação é $\Theta = \{2, \dots, P, P + 1\}$. Novas previsões são realizadas considerando este novo conjunto de previsão e este procedimento persiste até que tenha sido realizado previsões para todas as observações contidas na *out-of-sample*.

5 Modelo com Fatores Comuns e Média dinâmica de modelos (DMA).

5.1 Modelos com Fatores Comuns.

Seja X_{it} um painel em que $i = 1, \dots, N$ corresponde a N variáveis macroeconômicas e $t = 1, \dots, T$ a variação temporal. Será assumido que X_{it} possui uma estrutura fatorial aproximada, isto é:

$$X_{it} = \lambda_i' F_t + e_{it} \quad (22)$$

Na qual F_t são os fatores comuns; λ_i são os pesos dos fatores e $e_{it} \sim N(0, \sigma^2)$ é o erro de aproximação de X_{it} por $\lambda_i' F_t$. A hipótese de que X_{it} possui uma estrutura fatorial implica que a dimensão de X_{it} poderá ser reduzida a $r \ll N$ fatores, em que cada fator

representa uma parcela da variabilidade de X_{it} . Os fatores serão estimados por meio do método de componentes principais (PCM).

Assuma que se objective prever a variável y_t , h períodos à frente, utilizando o conjunto de informação em t , $y_{t+h|t}$. Seja W_t um conjunto de variáveis exógenas, normalmente representadas pelas defasagens de y_t . Então a equação de previsão é dada por:

$$y_{t+h|t} = \beta(L)\hat{F}_t + \alpha(L)W_t + \varepsilon_t \quad (23)$$

Na equação acima, $\beta(L)$ e $\alpha(L)$ correspondem a polinômios das defasagens de \hat{F}_t e W_t , respectivamente. O DMA fará a escolha do melhor modelo de previsão a cada período considerando todas as possíveis combinações da variável $Z_t = (\hat{F}_t, W_t)$.

Em certo sentido, o procedimento proposto constitui uma nova forma de seleção do modelo de previsão, diferentemente da abordagem tradicional proposta por Stock e Watson (2002). Na abordagem tradicional a equação (23), após estimado os fatores, é escolhida utilizando critérios de informação, como o BIC e o AIC. Bai e Ng (2008) argumentam que tal procedimento é inadequado, pois ao contrário de modelos autorregressivos, os fatores não possuem uma ordenação natural entre si. Cheng e Hansen (2015) e Rossi e Carrasco (2016) propõe critérios alternativos, como o critério de Mallows e o critério de validação cruzada generalizada para selecionar as variáveis em (23).

O método proposto aqui acrescenta duas inovações a forma tradicional de seleção das variáveis que compõe (23). Por um lado, a utilização do DMA permite que os parâmetros da equação de previsão se modifiquem no tempo. Esta característica é bastante interessante tendo em vista que instabilidades nos parâmetros podem ser prejudiciais as previsões, como aponta Rossi (2012).

Por outro lado, a própria forma funcional do modelo poderá se alterar, possibilitando que diferentes modelos possam ser escolhidos em cada período. Ou seja, a cada período o modelo de previsão ideal pode mudar, permitindo com isso uma previsão mais acurada em longos períodos de tempo. De fato, existem evidências de que os melhores modelos de previsão mudam no tempo. Stock e Watson (1998), por exemplo, apontam que os modelos de previsão que eram bastante adequados antes do período da Grande Moderação⁹ não apresentavam acurácia significativa durante a Grande Moderação.

A utilização de índices de difusão ponderados pelo DMA foi inicialmente proposta por Koop and Korobilis (2011), que analisaram o poder de previsão de fatores extraídos de blocos de variáveis macroeconômicas. Cada fator \hat{F}_t era obtido a partir de um grupo de variáveis comuns, como índices de preços, variáveis reais, variáveis financeiras etc. O motivo para adoção desta estimação em blocos de variáveis, segundo os autores, deve-se a possibilidade da interpretação do comportamento dos fatores para a previsão y_t ao longo do tempo.

⁹A Grande Moderação corresponde a um período de baixa volatilidade nos ciclos de negócios que iniciou-se por volta de 1980 e estendendo-se até a Grande Crise Financeira de 2008.

Todavia, há algumas limitações nesta abordagem. Primeiro, Bai (2008) demonstrou que o erro de estimação dos fatores não seria prejudicial a previsão se $N, T \rightarrow \infty$ e $\sqrt{N}/T \rightarrow \infty$. O motivo para essa constatação, segundo Ng (2012), é que à medida que $N \rightarrow \infty$, a variação em ε_{it} passa a ser dominada pela estrutura $\lambda_i'F_t$, possibilitando a aproximação de X_{it} por meio de $\lambda_i'\hat{F}_t$. O procedimento adotado por Koop and Korobilis (2011) não parece adequado, pois reduz N para a extração de fatores, possivelmente gerando uma aproximação imprecisa. Portanto, não parece plausível confiar, por exemplo, que fatores estimados considerando apenas os índices de preço sejam uma boa aproximação dos próprios índices.

Uma limitação adicional decorre da não consideração da interdependência entre as variáveis contidas em X_{it} . A estrutura fatorial em (22) possui matriz de variância-covariância dada por:

$$\sum_X = \Lambda' \sum_F \Lambda + \sum_\varepsilon \quad (24)$$

Na equação acima, \sum_F é a matriz de variância de F. Bai (2002) comenta que a hipótese de independência entre as variáveis sobre as quais são extraídos os fatores não parece apropriada a modelos econômicos. De fato, na abordagem clássica da análise fatorial é assumido que \sum_ε é uma matriz diagonal. Economicamente tal hipótese não faz sentido, tendo em vista a possível correlação entre as variáveis contidas em X_{it} .

Chamberlain e Rothschild (1983) assumem uma hipótese mais realista de que os autovalores da matriz $\sum_\varepsilon = E(e_t e_t')$ são limitados, isto é, que existe uma correlação fraca, serial e transversalmente, nos erros e_{it} , passando tal estrutura fatorial a ser chamada de modelo fatoriais aproximados. A abordagem de Koop e Korobilis (2011), portanto, ao separar as variáveis em grupos comuns não parece interessante pois o efeito das correlações entre as variáveis pode não ser captados pelos fatores estimados.

Em diversos pontos o método proposto neste trabalho se distingue do método de Koop and Korobilis (2011). Primeiro, não será assumido nenhum tipo de bloco de variáveis. Ou seja, os fatores serão estimados a partir de um conjunto grande de dados sem que haja nenhum tipo de separação ad-hoc das variáveis.

Segundo, serão propostos algumas extensões a abordagem dos fatores ponderados pelo DMA cujo objetivo é buscar formas mais eficientes de estimar os fatores. Isso é feito ao estimar os fatores considerando quais variáveis serão previstas¹⁰. Três formas alternativas de estimação dos fatores serão propostas: fatores estimados em variáveis previamente selecionadas pelo método do Lasso (Least Absolute Selector and Shrinkage Operator) ou Lar (Least Angle Regression) como em Bai e Ng (2008) e o método de supervisão do CFPC (Combining Forecasts using Principal Components) desenvolvido

¹⁰Essa forma de estimação faz com que os fatores sejam chamados de fatores supervisionados, Ver Bair et al (2006), Tu e Lee (2013).

por Tu e Lee (2013). Estes métodos de pré-seleção de variáveis mostraram-se bastante promissores e serão aqui considerados como uma alternativa ao modelo principal.

5.2 Média dinâmica de modelos (DMA).

O DMA consiste em um método de combinar previsões na qual permite que tanto os parâmetros quanto os modelos se modifiquem no tempo. Foi desenvolvido originalmente por Raftery *et al* (2010) e posteriormente aplicado a dados econômicos como em Koop e Korobilis (2011, 2012), Beckmann e Schussler (2016), Caldeira *et al* (2015), entre outros.

Suponha que a variável que se queira prever seja y_t e que x_t represente um conjunto de regressores, isto é, x_t é uma matriz de tamanho $T \times N$, em que: T é o tamanho da série temporal e N é o número de variáveis utilizadas como previsores. Perceba que existem $2^N = k$ possíveis combinações de variáveis que geram previsões para y_t . Por isso, chame $y_{t+h|t}^{(k)}$ a previsão de y_t para h períodos a frente, considerando o conjunto de informação em t e o k -ésimo modelo.

Raftery *et al* (2010) propuseram o seguinte filtro de Kalman para este problema:

$$y_{t+h|t}^{(k)} = x_t^{(k)} \beta_t^{(k)} + \epsilon_t^{(k)} \quad (25)$$

$$\beta_t^{(k)} = \beta_{t-1}^{(k)} + \eta_t^{(k)} \quad (26)$$

A equação (25) representa a equação de espaço e a equação (26) é chamada de equação estado. $x_t^{(k)}$ é um subconjunto de x_t . Assim, para estimar a dinâmica dos parâmetros ao longo do tempo é preciso considerar a equação de transição do filtro de Kalman. Por outro lado, para combinar estes k modelos é preciso considerar a ponderação dinâmica de cada modelo ao longo do tempo. É mais didático se cada um destes dois aspectos for tratado separadamente.

Sem perda de generalidade, considere $k = 1$ e $h = 1$. Então, o filtro de Kalman é dado por:

$$y_{t+1|t} = \beta_t x_t + \epsilon_t \quad (27)$$

$$\beta_t = \beta_{t-1} + \eta_t \quad (28)$$

Onde, $\epsilon_t \sim N(0, \sigma_{\epsilon_t}^2)$, $\eta_t \sim N(0, \sigma_{\eta_t}^2)$ e ϵ_t e η_t são não correlacionados. Considere as seguintes equações de transição $P_{t|t-1} = E \left[(\epsilon_t - \epsilon_{t|t-1}) (\epsilon_t - \epsilon_{t|t-1})' \right]$, $P_{t|t} = E \left[(\epsilon_t - \epsilon_{t|t}) (\epsilon_t - \epsilon_{t|t})' \right]$ e $G_{t|t-1} = E \left[(y_t - y_{t|t-1}) (y_t - y_{t|t-1})' \right]$. De acordo com essa notação, $\epsilon_{t|t-1}$ é o erro de previsão em t cometido utilizando o conjunto de informação em $t - 1$. Equivalentemente, $y_{t|t-1}$ corresponde a previsão de y_t considerando o conjunto de informação em $t - 1$.

Supondo que ϵ_t tem distribuição normal, então $\beta_{t|t-1} \sim N(\hat{\beta}_{t-1}, P_{t|t-1})$ e :

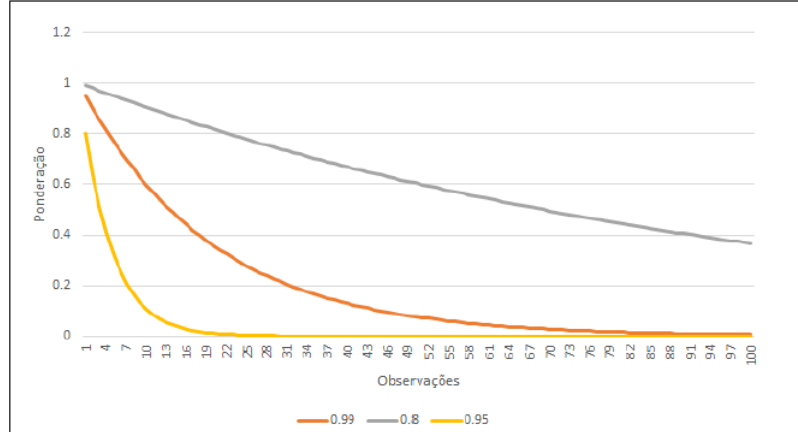
$$P_{t|t-1} = P_{t-1|t-1} + \sigma_{\eta t} \quad (29)$$

Raftery *et al* (2010) propuseram uma forma computacionalmente mais simples para calcular (29), sem recorrer a métodos do tipo *Markov Chain Monte Carlo* (MCMC) que são computacionalmente intensivos. A sugestão dos autores foi considerar $\sigma_{\eta t} = (\lambda^{-1} - 1) P_{t-1|t-1}$, na qual λ é chamado de fator de esquecimento. Dessa forma,

$$P_{t|t-1} = \frac{1}{\lambda} P_{t-1|t-1} \quad (30)$$

Se for assumido $0 < \lambda \leq 1$ então o fator de esquecimento sugere uma ponderação das observações que compõe $P_{t|t-1}$. Isto é, a j -ésima observação receberá peso λ^j . Por exemplo, supondo que a frequência da base de dados seja mensal, então a observação de cinco anos atrás, considerando um $\lambda = 0.99$, receberá apenas 55% do peso¹¹ sobre $P_{t|t-1}$. A Figura 1 mostra o efeito exponencial da ponderação sobre as variáveis considerando 100 observações e os valores do $\lambda = \{0.80; 0.95; 0.99\}$. Percebe-se que quanto menor o λ então mais peso é atribuído as observações mais recentes.

Figura 1: Ponderação do Fatores de Esquecimento



Nota: A figura 1 mostra o peso atribuído a observações passadas devido a escolha do fator de esquecimento λ . Três valores para o fator de esquecimento são considerados: $\lambda = \{0.80; 0.95, 0.99\}$. Percebe-se que quanto menor o fator de esquecimento, maior é o peso dado a observações mais distantes no tempo.

Esta forma de especificação de (29) foi originalmente analisada por Fagin (1964) e Jazwinsky (1970). Existem duas vantagens nesta abordagem. Primeiro, não é necessário recorrer a métodos computacionais pesados, como o MCMC. A segunda vantagem é que a ponderação exponencial das observações produz uma janela de previsão que atribuí mais peso a observações mais recentes. Desta forma, apenas as instabilidades mais recentes são incorporadas nas estimações.

¹¹Em cinco anos há 60 meses. Portanto, $0.99^{60} = 0.5471$.

A previsão é completada por meio da equação de atualização,

$$\beta_{t|t} \sim N(\hat{\beta}_t, P_{t|t}) \quad (31)$$

Na qual,

$$\hat{\beta}_t = \hat{\beta}_{t-1} + P_{t|t-1}x_tH_t^{-1}(y_t - x_t\hat{\beta}_{t-1}) \quad (32)$$

e

$$P_{t|t} = P_{t|t-1} - P_{t|t-1}x_tH_t^{-1}x_tP_{t|t-1} \quad (33)$$

Sendo $H_t = \sigma_{\epsilon t} + x_tP_{t|t-1}x_t'$. A previsão é feita recursivamente e atualizadas a cada período pelas equações (32) e (33). Portanto, $y_{t|t-1} \sim N(x_t\hat{\beta}_{t-1}, H_t)$. O mesmo procedimento é aplicado no caso da presença de muitos modelos. No caso do sistema formado por (27) e (28), seja Θ_t o vetor de todos os k possíveis coeficientes no tempo t , isto é: $\Theta_t = (\beta_t^{(1)}, \dots, \beta_t^{(k)})$. O DMA busca atribuir uma ponderação para cada previsão de $E(y_t^{(k)}|y_{t-1})$ no tempo t condicional a informação, já observada, sobre o comportamento da variável alvo no tempo $t - 1$.

Considere que cada previsão receba a probabilidade em cada período t , dada por $p(L_t = k|y_{t-s})$, isto é, a probabilidade do modelo no tempo t ser k ($L_t = k$), condicional a ocorrência de y_{t-s} ¹². O DMA, portanto, consiste na previsão ponderada de todos os possíveis modelos, com peso dado por $p(L_t = k|y_{t-s})$ e supondo um esquema recursivo de estimação.

$$E(y_t|y_{t-1}) = \sum_{i=1}^K p(M_{ti}|I_{t-1}) x_t^{(i)} \hat{\beta}_t^{(i)} \quad (34)$$

O problema consiste em determinar $p(M_{ti}|I_{t-1})$. Raftery *et al* (2010) propuseram um método para determinar $p(M_{ti}|I_{t-1})$ que depende de um fator de esquecimento α , cuja interpretação é semelhante a λ . Assim,

$$P(M_{it}|I_{t-1}) = \frac{P(M_{it}|I_{t-1})^\alpha}{\sum_{j=1}^J P(M_{jt}|I_{t-1})^\alpha} \quad (35)$$

A atualização da ponderação do modelo é realizada pela Regra de Bayes,

$$P(M_{it}|I_{t-1}) = \frac{p(y_t|M_{ti}, I_{t-1})P(M_{ti}|I_{t-1})}{\sum_{j=1}^J p(y_t|M_j, I_{t-1})P(M_{jt}|I_{t-1})}$$

Onde $p(y_t|M_{ti}, I_{t-1})$ é a densidade preditiva do modelo i . Pode-se escrever a equação (35) como:

¹²De uma forma mais geral, pode-se considerar que tal probabilidade refere-se a probabilidade de ocorrer o modelo M_{ti} , $i = 1, \dots, k$ e $t = 1, \dots, T$, condicional ao conjunto informação I_{t-1} . Portanto, $P(L_t = k|y_{t-s}) = P(M_{ti}|I_{t-s})$

$$P(M_{ti}|I_{t-1}) \propto [P(M_{ti}|I_{t-2})p(y_{t-2}|M_{ti}, I_{t-2})]^\alpha \quad (36)$$

$$= \prod_{s=1}^{t-1} p(y_s|M_{ti}, I_{t-s-1})^{\alpha^s}$$

A forma de ponderação proposta por Raftery *et al* (2010) possui a vantagem de não requerer uma matriz de transição entre cada um dos modelos ao longo do tempo, o que é computacionalmente pesado. Assim, modelos que tenham realizado boas previsões no passado recente, recebam maior peso para a previsão do período atual.

A noção de passado recente depende do parâmetro de esquecimento α . Este atribui ponderação exponencial ao passado, de maneira similar ao fator de esquecimento λ . Note que se $\alpha = 1$, então a forma de ponderação passa a obedecer apenas a atualização pela regra de Bayes, gerando, portanto, uma previsão bayesiana ponderada (BMA). Por outro lado, se $\alpha = 0$, então a ponderação atribui igual peso a todos os modelos, sendo então um modelo com igual ponderação.

O DMA por sua vez, assume que $0 < \alpha \leq 1$. A tamanho da janela exponencial dependerá da frequência dos dados e da escolha de α . Escolhas recomendadas por Koop e Korobilis (2012) e por Raftery *et al* (2010) são $\alpha = \{0.95; 0.99\}$. Estes serão os *benchmarks*.

No modelo proposto nesse trabalho o conjunto de previsores x_t é formado por fatores estimados por componentes principais f_t e um conjunto de variáveis exógenas, w_t , que no caso incluem as defasagens da variável alvo. Ou seja, $x_t = (f_t, w_t)$.

5.3 Extensões.

A estimação de modelos fatoriais reduz um grande conjunto de dados de tamanho $N \times T$ a alguns $r \ll N$ fatores. Uma questão importante é saber quais variáveis devem estar contidas nesse grande conjunto de dados para que as estimativas sejam eficientes. Boivin e Ng (2006) apresentaram evidências empíricas da existência de subconjuntos de N que proporcionam estimações mais eficientes. Os autores buscaram selecionar este subconjunto por meio da correlação entre as variáveis alvo da previsão e as variáveis contidas na base de dados. Bai e Ng (2008) propuseram métodos de seleção de variáveis através de *hard* e *soft thresholds*. No primeiro caso, as variáveis são selecionadas utilizando critérios de infomção como o de Akaike (AIC) e o Baysiano (BIC). No segundo caso, são aplicados métodos de regularização como o LASSO (*Least Absolute Selector and Shrinkage Operator*) e o LAR (*Least Angle Regression*)¹³ para selecionar previamente as variáveis.

¹³O método do Lasso foi proposto originalmente por Tibsharani (1996) e o método do Lar foi desenvolvido por Efron et al (2002)

De certa forma, esses métodos buscam corrigir uma dificuldade dos modelos fatoriais. Modelos fatoriais, quando aplicados em versão tradicionais, encontram estimativas que não levam em consideração as variáveis que serão previstas. Isto é, os fatores estimados serão os mesmos independentes se o objetivo é prever variáveis reais, como o PIB, ou variáveis monetárias, como a inflação, por exemplo. Na literatura estatística, tal aproximação para a estimação dos fatores é chamada de fatores não supervisionados como em Hastie et al (2012).

Bair et al (2006) foram os primeiros a propor métodos fatoriais supervisionados, isto é, modelos fatoriais cujos fatores levam em consideração no processo de estimação as variáveis que serão previstas. Outras técnicas de supervisão tem sido propostas. Tu e Lee (2014), por exemplo, propuseram selecionar os fatores de uma base de dados transformada. Esta base de dados é obtida pela estimação da influência de cada variável contida em X_{it} sobre a variável alvo da previsão. Este método foi chamado de Componentes Principais usando Combinação de Previsões (CFPC¹⁴). Boldrini e Hillenbrand (2015) e Giovannelli e Proietti (2014) também propuseram métodos alternativos de supervisão de fatores.

Para os propósitos deste trabalho, mais especificamente do capítulo 4, o DMA usando fatores será estendido para o caso supervisionado por meio ou da pré-seleção de variáveis através dos métodos do LASSO e do LAR ou por meio da supervisão pelo CFPC. O objetivo é verificar se o poder de previsão pode ser melhorado caso seja empregado este tipo de supervisão dos fatores, obtida ou pela prévia redução do número de observações contidas em X_{it} ou pela extração dos fatores de uma base de dados transformada.

No caso do método de seleção do Lasso as variáveis serão selecionadas buscando minimizar o erro quadrático médio condicionado a uma restrição normada L_1 , isto é:

$$\sum_{t=1}^T (y_t - \beta X_{it})^2 + \lambda |\beta|$$

Na qual, β é um vetor de coeficientes a ser estimado de tamanho $N \times 1$, λ é chamado de parâmetro de *tunning* e mensura a sensibilidade da seleção e $|\beta| = \sum_{i=1}^N \beta_i$. O parâmetro de *tunning* será selecionado por meio de validação cruzada. O método do Lasso promove a seleção de variáveis ao atribuir a variáveis irrelevantes o coeficiente estimado igual a zero.

Outra forma de seleção de variáveis é o método do *Least Angle Regression* (LAR) que foi proposto inicialmente por Efron *et al* (2004) como um método de seleção de variáveis similar ao *Foward Stepwise Regression*¹⁵. Esta técnica, primeiramente, ordena as variáveis candidatas a participar do modelo de acordo com seu poder preditivo. Depois, seleciona, entre as variáveis melhor ranqueadas, um modelo parcimonioso e usa este para

¹⁴Sigla original do inglês.

¹⁵*Foward Stepwise Regression* constrói um modelo por meio da adição de novas variáveis a cada processo de estimação. Após, cada novo conjunto de variáveis é utilizado para se realizar a previsão.

realizar a previsão. Em comparação com outros métodos de seleção, o LAR é considerado mais eficiente computacionalmente do que o LASSO. O algoritmo do LAR é rápido e possui previsão acurada em diferentes estruturas de dados¹⁶.

Por fim, o método do CFPC, desenvolvido por Tu e Lee (2014), primeiramente, regride cada uma das variáveis contidas em X_{it} sobre a variável alvo y_t . Após, para cada período de fora da amostra é realizada a previsão da variável alvo $\hat{y}_{i,t+h|t}$. Sobre a base de dados formada por todas essas previsões serão extraído os fatores.

Dessa forma, tem-se quatro possíveis especificações para o DMA usando fatores. A especificação não supervisionada, como a (FDMA); as especificações supervisionadas com pré-seleção utilizando o Lasso (Lasso-FDMA), o LAR (Lar-FDMA) e o CFPC (CFPC-FDMA). Segue abaixo um quadro resumo relacionando os principais modelos que serão utilizados nesse trabalho.

Tabela 1: Modelos

<i>FAAR</i>	<i>Modelos Autoregressivos Aumentados</i>
<i>BMA</i>	<i>Modelos Bayesianos Ponderados</i>
<i>MMA</i>	<i>Modelos Mallows Ponderados</i>
<i>JMA</i>	<i>Modelos Jacknife Ponderados</i>
<i>LHO</i>	<i>Modelos de Validação Cruzada Leave-h-Out</i>
<i>DMA</i>	<i>Modelos com Fatores Comuns e Média Dinâmica</i>
<i>SMA</i>	<i>Modelos Média Simples.</i>
<i>GSMA</i>	<i>Modelos de Média Simples Geral</i>

5 Base de dados

A base de dados utilizada no segundo e terceiro capítulos contém 117 variáveis macroeconômicas com frequência mensal incluindo: índices de preços setoriais, dívidas externas e governamentais, alguns componentes da balança de pagamentos, importação e exportação de bens setoriais, salário, taxa de juros de longo prazo, indicadores financeiros, índices da atividade econômica, taxa de desemprego, algumas variáveis de economia internacional como importação e exportação dos EUA, taxa de juros de longo prazo dos EUA, consumo de energia elétrica e combustíveis, índices de produção setorial e outros¹⁷.

Essa base de dados inicia-se em 1995.05 e termina em 2015.12, perfazendo 259 observações. Todavia, devido a necessidade de se computar variáveis defasadas, restaram 242 observações. A *amostra in-sample* contém 100 observações, de 1996.10 a 2004.02. Assim, a janela móvel terá tamanho $P = 100$. Por sua vez, a *amostra out-of-sample* inicia-se em 2004.03 e termina em 2015.12, totalizando 142 observações.

¹⁶Ver Efron *et al* (2004), Boveslud *et al* (2007), Saigo *et al* (2007), Gelper e Croux (2008).

¹⁷A base de dados completa e os códigos utilizados para realizar as estimativas estão disponíveis desde que requisitados por email aos autores. Em breve, será disponibilizado no site dos autores o *link* para o *download* de todo material necessário para a replicação dos resultados.

As variáveis alvos para o segundo capítulo incluem seis índices de preços, o índice de produção industrial, a taxa de desemprego e a taxa real de câmbio real-dólar. Os índices de preços são: índice geral de preços (IGP_10), índice nacional de custo da construção (INCC), índice nacional de preços ao consumidor (INPC), índice geral de preços ao consumidor (IPC), índice de preço ao consumidor amplo (IPCA) e o índice de preços no atacado (IPA). A Figura 1 apresenta o comportamento temporal das variáveis alvo. Entre os índices de preços, IGP_10 e IPA apresentaram maior volatilidade. Enquanto que as variáveis analisadas no terceiro capítulo são apenas o índice de produção industrial, índice nacional de preços ao consumidor (INPC), índice de preço ao consumidor amplo (IPCA) e taxa de juros de longo prazo.

Foram realizadas transformações nas variáveis para obter a estacionariedade das séries temporais antes da estimação dos fatores. O procedimento para alcançar a estacionariedade foi baseado em Stock e Watson (2002, 2012). Primeiramente, as variáveis são testadas utilizando o teste de Dickey-Fuller aumentado (ADF). Se a hipótese nula não é rejeitada, então é aplicada uma transformação sobre a série até o ADF rejeitar a hipótese de estacionariedade. Dentre as transformações aplicadas estão primeira ou segunda diferença, logaritmo e logaritmo da diferença. Ao final, a robustez da estacionariedade da série é verificada ao se aplicar o teste de estacionariedade KPSS.

O quarto capítulo possui dois exercícios empíricos. No primeiro, as variáveis alvos são seis variáveis macroeconômicas dos EUA: índice de preços ao produtor (IPI) e ao consumidor (CPI); índice de produção industrial (IP), M2, *house starts* (Hous) e taxa de desemprego (Unrate). Os fatores serão estimados de uma base de dados contendo 115 variáveis macroeconômicas, com frequência mensal entre 1963.1 até 2015.12. A amostra de dentro estende-se de 1963.1 até 1972.12. Por sua vez, a amostra de fora, ou amostra de avaliação, inicia-se em 1973.1 e finaliza em 2015.12. No segundo exercício, a variável a ser prevista é o excesso de retorno do S&P 500. São utilizadas quatorze variáveis financeiras como previsores. A base de dados corresponde a uma versão atualizada de Goyal e Welch (2008) e se estende de 1960.1 até 2011.12, com frequência mensal. As variáveis utilizadas encontram-se na Tabela 2.

Tabela 2: Variáveis Financeiras

Símbolo	Nome da variável	Definição
log_dp	Log da razão do dividendo-preço	log da soma móvel de 12 meses de dividendos pag menos log do índice S&P 500.
log_dy	Log do dividendo-yield	log da soma móvel de 12 meses de dividendos pag menos log do preço dos ativos
log_ep	log dos rendimentos-preço	log da soma móvel de 12 meses de dos ganhos sobre S&P menos log do preço dos ativos
log_de	Log da razão do pagamentos de dividendos	log da soma móvel de 12 meses de dividendos pagos menos log da soma móvel de 12 meses de dos ganhos sobre S&P
svar	Variância do índice	Soma mensal da variância dos retornos diários do S&P
bm	razão do Book-to-market	razão do book-to-market para DJIA
ntis	Expansão líquida de ativos	razão da soma móvel de 12 meses dos ativos líquidos pelo preço de capitalização de mercado no fim do exercício listados
T_bill	Treasury bill rate	interest rate sobre 3 meses do T-Bill (mercado secundário)
lty	Juros de longo prazo	Títulos do governo americano de longo prazo
ltr	Retorno de longo prazo	Retorno de longo prazo dos títulos do governo
ts	Estrutura a termo	taxa de juros de longo prazo menos taxa do T-bill
dfy	Default do spread da taxa de juros	Diferença entre corporate bond yields de empresas de baixa e alta qualidade (BAA- e AAA-)
dfr	Default spread do retorno	Retorno de longo prazo do corporate bond menos retorno de longo prazo do títulos do governo
inf	Inflação	Inflação calculada a partir dos índices de preços ao consumidor (CPI (all urban consumers))

Notas: A tabela 5 descreve as variáveis utilizada no exercício empírico de previsão do excesso de retorno sobre o S&P 500. Toda a base de dados, assim como as rotinas utilizadas, encontram-se disponíveis desde que solicitadas por email aos autores.

A amostra de dentro inicia-se em 1960.1 e termina em 1972.12 (109 observações). Por sua vez, a amostra de fora, sobre a qual será feita a avaliação, inicia-se em 1973.1 e termina em 2011.12, perfazendo 469 observações. A estimação é feita de forma recursiva, isto é, para cada nova previsão no tempo $t + 1$ a $t - \text{ésima}$ observação será incorporada e a amostra será de 1960.1 até t . Os detalhes dessa base serão explicados de forma detalhada no capítulo 4

CAPÍTULO 2: PREVISÃO DE VARIÁVEIS MACROECONÔMICAS BRASILEIRAS USANDO MODELOS DE SÉRIES TEMPORAIS DE ALTA DIMENSÃO.

RESUMO

Este artigo analisa o desempenho de vários modelos de fatores de alta dimensão para prever nove variáveis macroeconômicas brasileiras, incluindo a taxa de desemprego, o índice de produção industrial, seis índices de preços e taxa de câmbio. Os fatores são extraídos de um conjunto de dados composto por 117 variáveis macroeconômicas. Métodos de shrinkage e combinações de previsões são utilizadas para selecionar ou combinar o melhor modelo fatorial de previsão. Os métodos de shrinkage utilizados foram Least Angle Regressions e Elastic Net, Bagging, Non Negative Garrote. Por sua vez, os métodos de combinações de previsões aplicados foram baseados no critério de Mallows, métodos bayesianos ponderados, validação cruzada, Leave-h-Out e modelo de ponderação de Jacknife. Todas as previsões foram avaliadas por meio do Model Confidence Set (MCS) que estabelece os melhores modelos de previsão que satisfaçam algum intervalo de confiança para o erro de previsão. Os resultados sugerem que, em geral, os modelos fatoriais apresentam um erro quadrático médio de previsão (MSFE) menor que o benchmark AR (4). Os melhores modelos de fatores para prever variáveis macroeconômicas brasileiras foram o Elastic Net e o Least Angle Regressions. Em algumas situações os modelos de ponderações Non Negative Garrote e Bayesiano apresentaram previsões satisfatórias

Palavras chaves: Previsão, Modelos de Fatores, Métodos de Shrinkage, Combinação de Previsão, Variáveis Macroeconômicas Brasileiras.

ABSTRACT

This paper analyzes the performance of several high dimensional factor models to forecast nine Brazilian macroeconomics variables, including unemployment rate, industrial production index, six price indexes and exchange rate. The factors are extracted from a dataset containing 117 variables. Shrinkage methods and forecast combinations are used to select or combine the best forecast factor model. The shrinkage methods used are Least angle regression, Bagging, Elastic Net, Non-Negative Garrote and forecast combination methods that includes Mallows, Bayesian, Leave-h-Out Cross Validation and Jackknife Model Averaging. The forecasts are evaluated by Model Condence Set (MCS) that establish the best forecast models satisfying some forecast error condence interval. The results suggest that, in general, high dimensional models present smaller mean squared forecast error (MSFE) than AR model. Elastic Net and Least Angle Regression are the best factor models to predict the Brazilian macroeconomic variables. In some situations the Non-Negative Garrote and Bayesian Model Averaging have satisfactory accuracy. Como o objetivo é a previsibilidade sobre o excesso de retorno, então

Key words: Forecast, Diffusion Index, Shrinkage Methods, Forecast Combination, Brazilian Macroeconomics

1 Introdução

A disponibilidade de grandes conjuntos de dados que podem ser empregados na previsão de variáveis econômicas, introduz a necessidade de novas técnicas capazes de utilizar essas informações de forma eficiente. Os modelos de fatores extraem um pequeno número de fatores a partir de um grande conjunto de dados, os quais são utilizados posteriormente para gerar previsões.

Os trabalhos de Stock e Watson (1999,2002) mostram que modelos de fatores podem gerar previsões para a inflação e outras 7 importantes variáveis macroeconômicas dos E.U.A., quando comparadas a modelos AR, VAR e de curva de Phillips. Depois desses estudos, vários outros têm encontrado resultados que corroboram a eficiência preditiva dos modelos de fatores (por exemplo Marcellino et al 2008, Artis et al 2005 e Dias, Pinheiro e Rua, 2010).

Para o Brasil, Ferreira, Bierens e Castelar (2005) estimam modelos de fatores lineares e não lineares para prever a taxa de crescimento do PIB brasileiro e reportam que estes modelos geram previsões com menores erros quadráticos médios de previsão (EQMP) do que as dos modelos VAR e AR. De acordo com Figueiredo (2010), modelos de fatores produzem melhores previsões para a taxa de inflação brasileira, principalmente em horizontes de previsão mais longos.

Uma outra possibilidade para elaborar modelos de previsão parcimoniosos, a partir de um conjunto de informação grande, é o uso de técnicas que promovem uma redução no número de variáveis através de uma pré-seleção ou através da combinação das mesmas. Nesse sentido a metodologia do Least Angle Regression (LAR) proposta por (Efron et al, 2004) e Least Absolute Shrinkage and Selection Operator (LASSO) de Tibshirani (1996) podem ser úteis. Esses métodos podem ser utilizados para pré selecionar as variáveis que serão utilizadas na estimação dos fatores comuns como em Bai e Ng (2008) ou para selecionar o melhor modelo de previsão considerando os fatores estimados, como em Kim e Swanson (2014), Cheng and Hansen (2015), Stock e Watson (2012) e outros.

Stock e Watson (2012) investigam o poder de previsão de algumas técnicas de seleção de variáveis através de métodos do tipo *shrinkage* aplicados na estimação de fatores por componentes principais para prever 143 variáveis trimestrais dos EUA. Eles concluem que as previsões que utilizam as técnicas de seleção *shrinkage* possuem uma precisão inferior aos modelos de fatores dinâmicos (DFM), especialmente para prever as variáveis relacionadas à atividade real da economia.

Por outro lado, Kim e Swanson (2014) preveem onze variáveis macroeconômicas e financeiras usando métodos de combinação Baesianas, LASSO, elastic net, bagging, entre outros. Eles reportam que modelos fatoriais são substancialmente melhorados em termos de previsão quando combinados com métodos de seleção. Cheng e Hansen (2015) realizam previsões utilizando o mesmo conjunto de dados macroeconômicos para os EUA de Stock

e Watson (2012) e relatam que a melhor previsão é obtida por meio da ponderação de previsões de modelos fatoriais. Os resultados apontam que o critério de informação de Mallow pode ser usado para variáveis (fatores) que não são diretamente observadas e sim estimadas.

No Brasil, Medeiros et al (2016) aplicam os métodos de LASSO para prever a taxa de inflação brasileira (IPCA e IGP). Os autores mostram que este método gera melhores previsões que modelos autoregressivos. Garcia et al (2016) estudam o poder de previsão de alguns métodos de *shrinkage* como o LASSO, o ADA-LASSO, POST-LASSO em relação a previsão de especialistas para a inflação. Eles mostram que no curto prazo, os métodos de *shrinkage* não são melhores que os especialistas. Entretanto, no longo prazo sempre existe algum modelo que prevê melhor a inflação que os especialistas. Estes autores, entretanto, aplicam os métodos de *shrinkage* diretamente nas variáveis observadas.

Existem evidências empíricas de que a combinação de métodos de previsão fatorial com seleção robusta de fatores, geram previsões mais precisas (KIM e SWANSON, 2014). Entretanto, há poucas pesquisas que investigam a eficiência preditiva dessas técnicas de combinação e seleção através de métodos tipo *shrinkage* aplicados em modelos de fatores para dados de países emergentes. As séries temporais dessas economias tendem a apresentar maior variabilidade e mudanças estruturais do que as mesmas séries observadas em países desenvolvidos.

Desta forma, o presente artigo verifica se modelos de fatores, nos quais os fatores e outras variáveis de controle adicionais, são selecionadas através de métodos *shrinkage* e por meio de técnicas de ponderação apropriados a modelos de previsão com variáveis geradas, produzem previsões mais eficientes para um conjunto de variáveis da economia brasileira. Dentre as variáveis previstas nesse trabalho estão a taxa de inflação medida a partir de seis índices diferentes: índice geral de preços (IGP_10), índice nacional de custo de construção (INCC), índice nacional de preço ao consumidor (INPC), índice de preço ao consumidor (IPC), o índice de preço ao consumidor amplo (IPCA) e o índice de preços ao atacado (IPA_10). Também serão previstas as variáveis taxa de desemprego (U_taxa), índice de produção industrial (IPI) e a taxa de câmbio Brasil-EUA (real_exr).

O método utilizado nesse estudo investiga qual a melhor forma de combinar ou selecionar modelos cujos previsores são fatores anteriormente estimados para prever esse conjunto de variáveis macroeconômicas. A vantagem desta abordagem sobre as precedentes é que os fatores, por reduzirem a dimensão de grandes conjuntos de dados, podem ser mais informativos do que qualquer variável considerada isoladamente. Vale ressaltar, que até o momento os autores desconhecem outro trabalho que tenha utilizado as mesmas técnicas e as comparado com um amplo conjunto de modelos alternativos que consideram a problemática de previsão a partir de um grande conjunto de variáveis disponíveis.

Os modelos de previsão utilizados são o modelo autoregressivo de quarta ordem (AR(4)), modelo fatorial autoregressivo aumentado (FAAR), modelo fatorial estático

(PCM), bagging, dois tipos de modelos Bayesianos ponderados (BMA), *least angle regression* (LAR), *elastic net* (EN), *non-negative garrote* (NNG), modelo Mallows ponderado (MMA), modelo jackknife ponderado (JMA), validação cruzada leave-h-out (LHO), modelo de média simples (SMA) e o modelo de média simples geral (GSMA). As previsões são comparadas em termos de raiz quadrada dos erros quadráticos médios de previsão (REQM) e pelo *model confidence set* (MCS) proposto por Hansen et al (2011).

Os fatores são estimados pelo método dos componentes principais (PC) a partir de uma ampla base de dados com 117 variáveis macroeconômicas brasileiras, com frequência mensal no período de 1996.5 à 2015.12. Todas as variáveis são transformadas em estacionárias. Os testes para identificar a estacionariedade foram o *Augmented Dickey-Fuller* (ADF) e o KPSS. Como a instabilidade nos parâmetros é bastante comum em séries temporais econômicas para os países emergentes foi empregado duas formas de lidar com este problema. Primeiro, para cada período de tempo t , os fatores serão extraídos e as estimativas de previsão serão feitas para horizonte $t + h$, tal estratégia é chamada de método de previsão recursiva. Em seguida, será considerado um método de previsão com janelas móveis (*rolling window*), lidando, nesse sentido, com instabilidades estruturais, como discutido por Pesaran e Timmermann (2007) e Rossi (2012).

Os melhores modelos de previsão são avaliados usando o *model confidence set* (MCS). Tais conjuntos de confiança são construídos de uma forma que contenha o melhor modelo de previsão dentre todos os possíveis modelos. Uma vantagem do MCS é o fato de que não é assumido que haja um modelo verdadeiro de previsão. Logo, é possível comparar vários modelos ao mesmo tempo. Os melhores modelos de previsão para variáveis macroeconômicas brasileiras estarão no MCS para a maioria das variáveis e horizontes.

Os resultados encontrados apontam que os as técnicas LAR e *Elastic Net* são os melhores modelos de previsão para todas as variáveis macroeconômicas em ambos os esquemas, recursivo e de janela móvel. Esses modelos estão no MCS para todas as nove variáveis e horizontes. Portanto, tais métodos apresentam-se como uma alternativa segura para realizar previsão para um amplo conjunto de variáveis macroeconômicas de interesse, sejam elas reais ou nominais. Outros métodos, como BMA e GSMA mostraram relativo sucesso na previsão das variáveis, entretanto, para alguns horizontes ou variáveis eles não estão no MCS. Os métodos ponderação de previsões, exceto BMA, não demonstram um bom desempenho para os dados brasileiros.

O restante desse artigo está organizado da seguinte forma: seção 2 exhibe os principais resultados para de REQM relacionados a ambos os métodos, recursivo e janela móvel e os resultados de MCS e identifica os melhores modelos de previsão. Por fim, a seção 3 apresenta as conclusões.

2 Resultados

2.1 Previsões usando o esquema recursivo

Esta seção apresenta os resultados para a previsão das variáveis macroeconômicas brasileiras utilizando o esquema recursivo de previsão. Neste esquema, uma a uma as observações são incorporadas ao modelo após cada procedimento de previsão. O horizonte de previsão é de $h=1,3,6$ e 12 meses. A Tabela 3 mostra a REQM para cada variável e modelo. Em negrito estão marcados os modelos que apresentaram o menor REQM.

Ao analisar a Tabela 3 os resultados parecem indicar que os melhores modelos para cada variável macroeconômica dependem do horizonte que está sendo previsto. Para $h=1$, a taxa de desemprego e o INPC são melhores previsto pelo AR(4). O modelo PCR, que é o modelo que utiliza apenas os fatores sem nenhum tipo de seleção ou combinação de variáveis, possui menor REQM para a produção industrial. O modelo que pondera igualmente as previsões geradas pela por JMA, MMA e LHO apresenta-se como o melhor predictor para o ipc e o ipca. O modelo que melhor prevê a taxa de câmbio é o modelo de fatores aumentado, que inclui os fatores estimados por PCM e as variáveis defasadas da variável alvo. Aqui, a defasagem das variáveis contidas no FAAR foi escolhida pelo critério de informação bayesiano (BIC), similar a Stock e Watson (2002).

Para o horizonte de previsão de três períodos à frente, ou seja $h=3$, o modelo SMA apresenta menor REQM para os índices de preços INCC, INPC, IPC e IPCA e o MMA para IGP_10 e IPA_10. Taxa de desemprego e IPI são melhores previstos pelo FAAR e PCR. Por sua vez, o BMA é o melhor predictor para a taxa de câmbio. Considerando o médio prazo ($h=6$), SMA é o melhor modelo para INPC, IPC e IPCA e o BMA1 para o INCC. Os índices IGP_10 e IPA_10 foram melhores previstos pelo FAAR. O PCR continua a ter a melhor acurácia para o IPI. Por fim, no longo prazo ($h=12$) o modelo PCR é o melhor modelo para IGP_10, IPA_10, IPC e o IPI. FAAR apresentou menor REQM para o INPC e a taxa de câmbio. O SMA permanece o melhor predictor para o IPCA.

É possível obter algumas conclusões gerais. Primeiro, o modelo PCR e SMA foram os melhores previsores para o IPI e o IPCA, respectivamente, considerando todos os horizontes de previsão. O IPCA, que é o índice de preços oficial do governo brasileiro, foi previsto com maior acurácia pelo SMA em todos os horizontes considerados, utilizando o esquema recursivo. Por sua vez, o resultado de o IPI ser melhor previsto pelo PCR era esperado, pois este modelo apresenta elevado poder de previsão em variáveis reais da economia, como indicado por vários autores como Stock e Watson (2002, 2004, 2006 e 2012), Marcellino (2008) e Eickmeier e Ziegler (2008).

Para as demais variáveis o melhor modelo de previsão depende do horizonte escolhido. O IPC foi melhor previsto pelo SMA nos horizontes $h=1,3$ e 6. FAAR, por sua vez, foi o modelo que gerou previsões mais acuradas para a taxa de câmbio real considerando

$h=1,3$ e 12 . Uma observação adicional refere-se ao sucesso dos modelos fatoriais no longo prazo ($h \rightarrow \infty$) para prever os índices de preços. Entretanto, para o curto e médio prazo, os modelos que ponderam as previsões foram mais eficientes.

A limitação dos resultados apresentados na Tabela 3 decorre da impossibilidade de distinguir se os modelos possuem poder de previsão similar para as variáveis macroeconômicas. O procedimento de construção do MCS permite lidar com este problema ao estabelecer um intervalo de confiança para o erro de previsão, ao nível de significância de α . Em outras palavras os modelos incluídos dentro desse intervalo possuem igual poder de previsão entre si e superior habilidade preditiva em relação aos demais que não estão no conjunto. Assim, é possível identificar se algum modelo possui poder de previsão para todas as variáveis e para todos os horizontes.

A Figura 2 apresenta as proporções em que cada modelo aparece no MCS para cada horizonte de previsão. Por exemplo, caso determinado modelo apresente valor igual a 0.80, então, este modelo está presente em 80% dos MCS para todas as variáveis no horizonte de previsão h . Esse resultado indica que para 80% das variáveis consideradas, este modelo possui elevado poder de previsão. A análise do MCS apresentada na Figura 2 realiza uma análise particular para os índices de preços. Assim, em cinza estão as proporções para todas as variáveis alvo (Todas as Variáveis) e em preto apenas para os índices de preços (Índices de Preço).

Note que para o horizonte de previsão $h = 1$ os modelos *Bagging* (BAG), *Least Angle Regression* (LAR) e *Elastic Net* (E.N.) estão em todos os MCS para todas as variáveis alvo. Por sua vez, os modelos *Non Negative Garrote* (NNG), Combinação Simples Ponderada Geral (GSMA), E.N., LAR, BMA 1 e 2 e BAG estão no MCS para todos os índices de preços. Além disso, note que o SMA foi o melhor modelo previsor do IPCA. Entretanto, não está no MCS para todas as variáveis de inflação.

Tabela 3: Raiz do Erro Quadrático Médio (Esquema Recursivo)

Horizonte de Previsão h=1														
	AR(4)	FAAR	PCR	Bagging	BMA 1	BMA 2	LAR	EN	NNG	MMA	JMA	LHO	SMA	GSM
U_rate	0.0080	0.0081	0.0085	0.0081	0.0084	0.0083	0.0081	0.0081	0.0081	0.0083	0.0084	0.0084	0.0089	0.0081
igp_10	0.0060	0.0058	0.0069	0.0059	0.0059	0.0060	0.0060	0.0060	0.0059	0.0065	0.0065	0.0065	0.0068	0.0060
incc	0.5648	0.5796	0.5513	0.5655	0.5868	0.5899	0.5682	0.5682	0.5702	0.6176	0.6118	0.6118	0.5868	0.5868
inpc	0.1261	0.1302	0.1313	0.1267	0.1266	0.1272	0.1270	0.1270	0.1294	0.1371	0.1368	0.1368	0.1376	0.1368
ipa_10	0.5010	0.4923	0.6395	0.5041	0.4938	0.4951	0.4957	0.4957	0.4965	0.5805	0.5904	0.5904	0.6039	0.5513
ipc	0.1182	0.1184	0.1220	0.1177	0.1188	0.1192	0.1202	0.1202	0.1237	0.1150	0.1155	0.1155	0.1114	0.1114
ipca	2.3279	2.3896	2.2735	2.3453	2.3372	2.3416	2.3501	2.3498	2.3571	2.2718	2.2632	2.2632	2.2409	2.3712
ipi	0.2467	0.2198	0.2022	0.2360	0.2342	0.2311	0.2443	0.2443	0.2410	0.2350	0.2335	0.2335	0.2387	0.2513
real_exr	0.7080	0.5967	0.6399	0.6969	0.6057	0.5971	0.6998	0.6998	0.6963	0.7859	0.8111	0.8111	0.8549	0.7612
Horizonte de Previsão h=3														
	AR(4)	FAAR	PCR	Bagging	BMA 1	BMA 2	LAR	EN	NNG	MMA	JMA	LHO	SMA	GSM
U_rate	2.2024	0.0069	0.0069	0.0085	0.0071	0.0070	0.0096	0.0097	0.0091	0.0088	0.0088	0.0088	0.0088	0.0081
igp_10	0.0070	0.0070	0.0068	0.0069	0.0071	0.0071	0.0070	0.0070	0.0069	0.0064	0.0066	0.0069	0.0066	0.0060
incc	0.5391	0.6184	0.6076	0.5278	0.5635	0.5747	0.5465	0.5465	0.5630	0.5281	0.5232	0.5211	0.5109	0.5412
inpc	0.1355	0.1461	0.1519	0.1291	0.1456	0.1479	0.1403	0.1402	0.1446	0.1337	0.1324	0.1358	0.1281	0.1412
ipa_10	0.6183	0.6069	0.6019	0.6140	0.6139	0.6109	0.6036	0.6036	0.5994	0.5545	0.5625	0.5849	0.5826	0.6012
ipc	0.1182	0.1184	0.1220	0.1177	0.1188	0.1192	0.1202	0.1202	0.1237	0.1150	0.1155	0.1155	0.1114	0.1114
ipca	2.3630	2.4181	2.5499	2.2993	2.4498	2.4999	2.3945	2.3940	2.4633	2.2515	2.2491	2.2969	2.1321	2.4112
ipi	0.2292	0.2116	0.2053	0.2144	0.2085	0.2089	0.2272	0.2272	0.2263	0.2146	0.2172	0.2151	0.2142	0.2212
real_exr	0.8707	0.7829	0.7782	0.8871	0.7583	0.7550	0.8544	0.8545	0.8500	0.8901	0.9084	0.9223	0.9080	0.8812
Horizonte de Previsão h=6														
	AR(4)	FAAR	PCR	Bagging	BMA 1	BMA 2	LAR	EN	NNG	MMA	JMA	LHO	SMA	GSM
U_rate	0.0098	0.0093	0.0090	0.0095	0.0083	0.0092	0.0094	0.0094	0.0094	0.0090	0.0090	0.0092	0.0091	0.0090
igp_10	0.0062	0.0060	0.0060	0.0062	0.0061	0.0060	0.0061	0.0061	0.0060	0.0066	0.0066	0.0065	0.0063	0.0060
incc	0.5209	0.5166	0.5231	0.5321	0.5003	0.5018	0.5169	0.5169	0.5222	0.5181	0.5226	0.5293	0.5159	0.5012
inpc	0.1478	0.1510	0.1439	0.1485	0.1471	0.1481	0.1461	0.1461	0.1449	0.1345	0.1332	0.1324	0.1290	0.1412
ipa_10	0.5837	0.5323	0.5561	0.5872	0.5690	0.5616	0.5601	0.5602	0.5529	0.6092	0.6148	0.6067	0.5839	0.5712
ipc	0.1216	0.1346	0.1376	0.1228	0.1240	0.1255	0.1210	0.1210	0.1192	0.1208	0.1204	0.1205	0.1151	0.1212
ipca	2.4841	2.3712	2.2767	2.4342	2.4419	2.4312	2.4548	2.4547	2.4252	2.2386	2.2256	2.1861	2.1804	2.3812
ipi	0.2194	0.1985	0.1883	0.2201	0.2172	0.2145	0.2197	0.2197	0.2174	0.2131	0.2136	0.2081	0.2099	0.2212
real_exr	0.8804	0.7701	0.8363	0.8339	0.7799	0.7733	0.8513	0.8513	0.8735	0.9049	0.9233	0.9214	0.9233	0.8712
Horizonte de Previsão h=12														
	AR(4)	FAAR	PCR	Bagging	BMA 1	BMA 2	LAR	EN	NNG	MMA	JMA	LHO	SMA	GSM
U_rate	0.0089	0.0097	0.0099	0.0087	0.0089	0.0090	0.0089	0.0089	0.0090	0.0079	0.0078	0.0079	0.0083	0.0081
igp_10	0.0070	0.0070	0.0066	0.0068	0.0070	0.0070	0.0070	0.0070	0.0070	0.0071	0.0071	0.0070	0.0070	0.0060
incc	0.6099	0.5971	0.5512	0.6031	0.6015	0.6034	0.6048	0.6048	0.6006	0.5394	0.5442	0.5413	0.5693	0.5112
inpc	0.1435	0.1322	0.1361	0.1483	0.1409	0.1399	0.1417	0.1417	0.1345	0.1454	0.1438	0.1435	0.1344	0.1312
ipa_10	0.6536	0.6255	0.5743	0.6536	0.6462	0.6420	0.6528	0.6528	0.6449	0.6498	0.6522	0.6632	0.6501	0.6412
ipc	0.1238	0.1090	0.1078	0.1305	0.1164	0.1151	0.1175	0.1175	0.1111	0.1187	0.1197	0.1220	0.1115	0.1114
ipca	2.5157	2.4127	2.3796	2.5990	2.4832	2.4462	2.4901	2.4903	2.3761	2.5118	2.4617	2.4557	2.3297	2.4412
ipi	0.2121	0.1966	0.1955	0.2112	0.2101	0.2083	0.2085	0.2085	0.2012	0.2089	0.2101	0.2064	0.2066	0.2112
real_exr	0.6513	0.5474	0.6955	0.6511	0.6210	0.6048	0.6437	0.6438	0.6198	0.7533	0.7781	0.7770	0.8134	0.7612

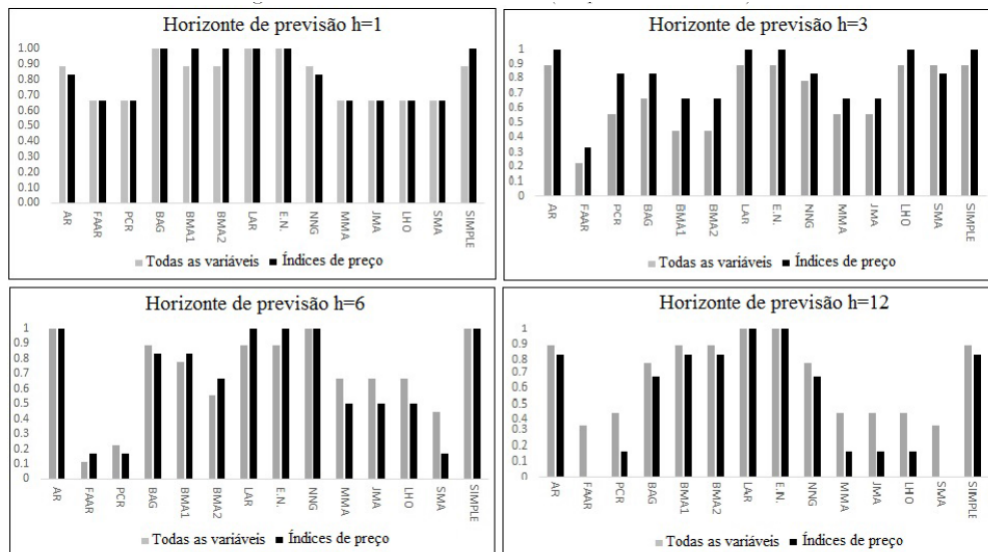
NOTA: A Tabela 1 apresenta a comparação entre os REQM para cada variável considerando diferentes horizontes de previsão e os quatorze modelos de previsão para o esquema recursivo. Em negrito, o modelo que apresentou o menor REQM

Para $h = 3$, a análise do MCS não fornece um único modelo que produza as melhores previsões, para todas as variáveis. Os modelos que tiveram melhor resultados foram E.N, AR, LAR, LHO, SMA e GSMA que estiveram em 87% dos MCS. Para os índices de preços os modelo mais acurados foram AR, LAR, E.N., LHO e GSMA.

Considerando o médio prazo de previsão ($h = 6$) os melhores modelos foram AR, NNG e GSMA para todas as variáveis. Todavia, para os índices de preços, e E.N., NNG, GSMA, LAR e AR estão contidos em todos os MCS. Para $h = 12$, E.N e LAR foram os melhores modelos para ambos todas as variáveis e índices de preços.

Os resultados indicam que o *Least Angle Regression* (LAR) e *Elastic Net* (E.N) possuem maior poder de previsão para todas as variáveis macroeconômicas consideradas neste trabalho, em especial, para as variáveis de preço, independente do horizonte de previsão escolhido. Desta forma, apesar de outros métodos possuírem superior poder de previsão para algumas variáveis, a estimação do MCS mostrou que LAR e E.N. possuem igual poder preditivo e ainda que podem ser utilizados em todas as variáveis com relativo sucesso.

Figura 2: Model Confidence Set (Esquema Recursivo)



NOTA: A figura 2 apresenta a frequência pela qual cada modelo está inserido no MCS em cada horizonte de previsão, considerando o esquema recursivo.

2.2 Previsão usando esquema de janela móvel

Esta seção, por sua vez, explora os resultados considerando o esquema de previsão de janela móvel. O esquema de janela móvel tem a vantagem de lidar com a instabilidade das séries temporais como observado por Pesaran e Timmerman (2007), Pesaran, Pettenuzzo e Timmerman (2006) e Rossi e Inoue (2012).

A Tabela 4 apresenta os resultados para a raiz do erro quadrático médio (REQM) considerando o esquema de janela móvel. Analogamente ao caso com o esquema recursivo, em negrito estão os modelos que apresentaram menor REQM. Comparando os resultados para ambos os esquemas percebe-se algumas similaridade e algumas diferenças.

Em termos gerais, o IPCA e o IPI continuam sendo melhor previstos pelos modelo SMA e PCR na maioria dos horizontes. Exceção ao IPCA considerando seis meses à frente que foi melhor previsto pelo *Bagging* e ao ipi que para seis e doze meses à frente foi melhor previsto pelo FAAR.

A taxa de desemprego também foi melhor prevista pelos modelos FAAR e PCR. Diferentemente do esquema recursivo, os modelos FAAR e PCR aparecem como os melhores previsores de curto prazo e longo prazo, ressaltando a perda de importância das combinações de previsões em maiores horizontes. Além disso, modelos como *Bagging*, LAR e *Elastic Net* foram os melhores previsores para alguns índices de preços.

Tabela 4: Raiz do Erro Quadrático Médio (Esquema Janela Móvel)

Horizonte de Previsão h=1														
	AR(4)	FAAR	PCR	Bagging	BMA 1	BMA 2	LAR	EN	NNG	MMA	JMA	LHO	SMA	GSD
U_rate	0.0081	0.0077	0.0066	0.0077	0.0078	0.0078	0.0083	0.0083	0.0083	0.0073	0.0074	0.0074	0.0084	0.0084
igp_10	0.0062	0.0059	0.0067	0.0061	0.0061	0.0061	0.0062	0.0062	0.0061	0.0071	0.0071	0.0071	0.0072	0.0072
incc	0.5677	0.5926	0.5947	0.5335	0.5816	0.5871	0.5781	0.5781	0.5879	0.6202	0.6271	0.6271	0.6155	0.5926
inpc	0.1243	0.1231	0.1490	0.1266	0.1241	0.1237	0.1242	0.1242	0.1242	0.1307	0.1308	0.1308	0.1287	0.1287
ipa_10	0.5255	0.4938	0.6139	0.5205	0.5121	0.5085	0.5209	0.5209	0.5109	0.6247	0.6238	0.6238	0.6273	0.5205
ipc	0.1189	0.1177	0.1340	0.1207	0.1179	0.1180	0.1168	0.1168	0.1144	0.1129	0.1129	0.1129	0.1089	0.1189
ipca	2.2497	2.3311	2.5329	2.2985	2.2788	2.2837	2.2645	2.2643	2.2678	2.1816	2.1878	2.1878	2.1411	2.2497
ipi	0.2449	0.1918	0.1834	0.2045	0.2077	0.2058	0.2433	0.2434	0.2434	0.2304	0.2317	0.2317	0.2392	0.2449
real_exr	0.7045	0.6060	0.6623	0.7477	0.6042	0.6031	0.6756	0.6756	0.6984	0.6379	0.6464	0.6464	0.8280	0.7045
Horizonte de Previsão h=3														
	AR(4)	FAAR	PCR	Bagging	BMA 1	BMA 2	LAR	EN	NNG	MMA	JMA	LHO	SMA	GSD
U_rate	0.0099	0.0097	0.0092	0.0096	1.9989	1.9786	2.1187	2.1188	2.1615	0.0092	0.0092	0.0093	0.0093	2.0099
igp_10	0.0070	0.0071	0.0066	0.0072	0.0072	0.0071	0.0069	0.0069	0.0072	0.0061	0.0062	0.0065	0.0064	0.0070
incc	0.5501	0.5730	0.5482	0.5620	0.6057	0.6031	0.5718	0.5717	0.5591	0.5041	0.5023	0.4837	0.5147	0.5501
inpc	0.1324	0.1328	0.1319	0.1285	0.1291	0.1298	0.1291	0.1291	0.1285	0.1372	0.1386	0.1405	0.1331	0.1324
ipa_10	0.6288	0.6214	0.6027	0.6469	0.6160	0.6169	0.6245	0.6245	0.6308	0.5445	0.5538	0.5709	0.5727	0.6288
ipc	0.1268	0.1204	0.1174	0.1118	0.1287	0.1283	0.1274	0.1274	0.1291	0.1330	0.1357	0.1365	0.1247	0.1268
ipca	2.3216	2.4055	2.3001	2.2052	2.3574	2.3703	2.3149	2.3149	2.3586	2.2571	2.2661	2.2676	2.1190	2.3216
ipi	0.2300	0.1970	0.1971	0.2001	0.2104	0.2091	0.2313	0.2313	0.2326	0.2258	0.2261	0.2278	0.2234	0.2300
real_exr	0.8604	0.8076	0.8684	0.9017	0.8230	0.8175	0.8338	0.8341	0.8180	0.9207	0.9415	0.9530	0.9143	0.8604
Horizonte de Previsão h=6														
	AR(4)	FAAR	PCR	Bagging	BMA 1	BMA 2	LAR	EN	NNG	MMA	JMA	LHO	SMA	GSD
U_rate	0.0097	0.0083	0.0086	0.0096	0.0095	0.0094	0.0095	0.0095	0.0090	0.0089	0.0089	0.0091	0.0089	0.0097
igp_10	0.0064	0.0070	0.0069	0.0066	0.0065	0.0066	0.0064	0.0064	0.0065	0.0065	0.0065	0.0065	0.0066	0.0064
incc	0.5624	0.6061	0.5878	0.5532	0.6410	0.6405	0.5618	0.5618	0.5615	0.6212	0.6145	0.6071	0.5647	0.5624
inpc	0.1276	0.1332	0.1370	0.1206	0.1347	0.1361	0.1275	0.1275	0.1280	0.1298	0.1303	0.1283	0.1261	0.1276
ipa_10	0.5813	0.6300	0.6601	0.5936	0.5952	0.5969	0.5843	0.5843	0.5886	0.6141	0.6193	0.6076	0.5970	0.5813
ipc	0.1136	0.1129	0.1205	0.1078	0.1157	0.1156	0.1134	0.1134	0.1132	0.1162	0.1148	0.1142	0.1119	0.1136
ipca	2.0112	2.1103	2.3426	1.8232	2.1921	2.2187	2.0223	2.0222	2.0586	1.9330	1.9173	1.9013	1.8679	2.0112
ipi	0.2171	0.2039	0.2050	0.2105	0.2190	0.2182	0.2208	0.2208	0.2230	0.2162	0.2152	0.2087	0.2183	0.2171
real_exr	0.8710	0.7592	0.8517	0.8022	0.8581	0.8511	0.8675	0.8672	0.8571	0.8905	0.8962	0.9002	0.8821	0.8710
Horizonte de Previsão h=12														
	AR(4)	FAAR	PCR	Bagging	BMA 1	BMA 2	LAR	EN	NNG	MMA	JMA	LHO	SMA	GSD
U_rate	0.0087	0.0086	0.0076	0.0084	0.0087	0.0087	0.0086	0.0086	0.0087	0.0080	0.0079	0.0082	0.0083	0.0087
igp_10	0.0074	0.0074	0.0072	0.0076	0.0073	0.0074	0.0073	0.0073	0.0071	0.0069	0.0070	0.0072	0.0071	0.0074
incc	0.6149	0.6227	0.5432	0.6371	0.6132	0.6119	0.6130	0.6130	0.6070	0.5561	0.5559	0.5603	0.5488	0.6149
inpc	0.1403	0.1314	0.1325	0.1439	0.1390	0.1379	0.1408	0.1408	0.1374	0.1381	0.1380	0.1436	0.1307	0.1403
ipa_10	0.6727	0.6514	0.6447	0.6898	0.6843	0.6847	0.6692	0.6692	0.6700	0.6238	0.6305	0.6666	0.6426	0.6727
ipc	0.1241	0.1113	0.1084	0.1287	0.1206	0.1185	0.1216	0.1215	0.1168	0.1233	0.1220	0.1180	0.1139	0.1241
ipca	2.4957	2.3005	2.2573	2.5539	2.3853	2.3680	2.4898	2.4899	2.4385	2.4265	2.4257	2.5057	2.2572	2.4957
ipi	0.2111	0.1830	0.2011	0.2079	0.2081	0.2059	0.2093	0.2093	0.2031	0.2224	0.2206	0.2096	0.2200	0.2111
real_exr	0.6189	0.6032	0.8870	0.6231	0.6181	0.6117	0.6126	0.6127	0.6000	0.7247	0.7327	0.7712	0.7822	0.6189

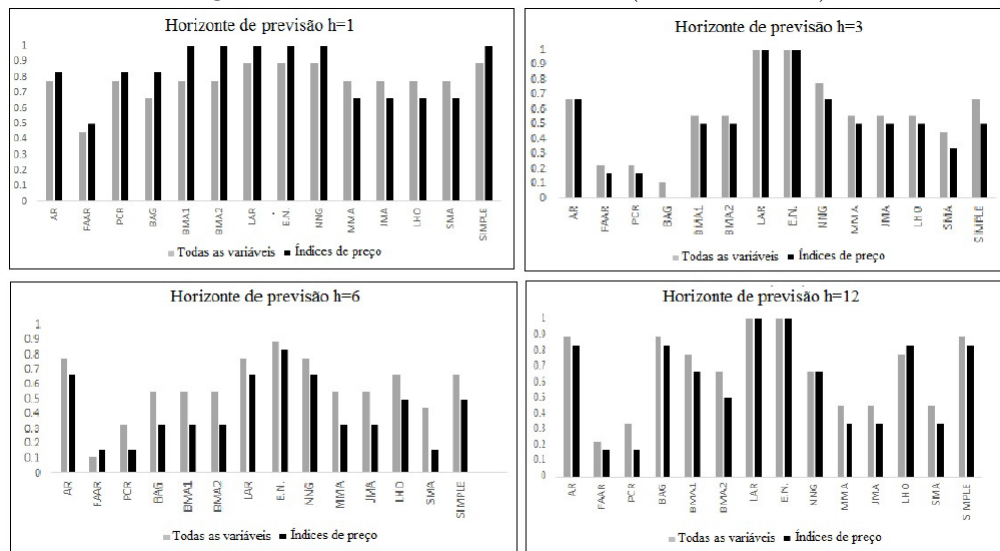
NOTA: A Tabela 2 apresenta a comparação entre os REQM para cada variável considerando diferentes horizontes de previsão e os quatorze modelos de previsão para o esquema de janela móvel. Em negrito, o modelo que apresentou o menor REQM.

A Figura 3 apresenta os resultados para o MCS considerando o esquema de janela móvel. Os resultados mostraram-se bastante similares ao caso recursivo. De fato, para o curto prazo ($h = 1$) os melhores modelos foram o NNG, EN, LAR, BMA e GSMA considerando apenas as variáveis de inflação. Já considerando todas as variáveis, os modelos mais acurados foram LAR, EN, NNG e GSMA que estão contidos em 87% dos MCS.

Para três meses à frente, apenas LAR e EN pertenceram a todos os MCS em ambas amostras, todas as variáveis e índices de preços. Para $h = 6$, EN foi o modelo mais frequente e para $h = 12$, novamente, LAR e EN foram os melhores.

Diante dos resultados é possível apontar que o LAR e o EN são os melhores modelos de previsão para as nove variáveis macroeconômicas considerando os horizontes de previsão de $h = 1, 3, 6$ e 12 . Assim, embora outros modelos tenham apresentado resultados superiores considerando apenas o REQM como medida, a análise subsequente mostrou que os modelos fatoriais selecionados pelo LAR e EN possuem igual poder preditivo aos melhores modelos para cada variável e para cada horizonte.

Figura 3: Model Confidence Set (Janela Móvel)



NOTA: A figura 3 apresenta a frequência pela qual cada modelo está inserido no MCS em cada horizonte de previsão, considerando o esquema de janela móvel.

2.3 Comparando as previsões nos esquemas recursivos e janela móvel

Nesta seção busca-se comparar as previsões nos esquemas recursivo e de janela móvel por meio do valor absoluto do REQM. A Tabela 5 apresenta os resultados do REQM para os melhores modelos selecionados em horizonte de previsão e para cada variável. Os termos em negrito indicam qual dos esquemas de previsão possui menor REQM.

Depreende-se da análise da Tabela 5 que o esquema de janela móvel possui menor REQM do que o esquema recursivo em quase todas as variáveis no curtíssimo prazo

($h = 1$). Isso indica que o esquema de janela móvel pode ser mais acurado no curto prazo. Exceção é a variável taxa de câmbio real-dólar (`real_exr`). Com relação ao horizonte $h = 3$ algumas variáveis passam a ter o valor absoluto do REQM menor no esquema recursivo que no esquema de janela móvel. São os casos das variáveis INPC e IPC. Por sua vez, as variáveis IGP_10, INCC, IPCA e IPI continuam a ser melhor previstas pelo esquema de janela móvel.

Tabela 5: Comparando as previsões nos esquemas recursivos e janela móvel

Variáveis	Horizonte de previsão							
	h=1		h=3		h=6		h=12	
	Janela Móvel	Recursivo	Janela Móvel	Recursivo	Janela Móvel	Recursivo	Janela Móvel	Recursivo
<code>U_rate</code>	0.0069	0.008	0.0092	0.0069	0.0083	0.0083	0.0076	0.0078
<code>igp_10</code>	0.0059	0.0059	0.0059	0.0064	0.0064	0.006	0.0069	0.0066
<code>incc</code>	0.5513	0.5335	0.4837	0.5109	0.5532	0.5003	0.5432	0.5394
<code>inpc</code>	0.1231	0.1261	0.1285	0.1281	0.1261	0.129	0.1307	0.1322
<code>ipa_10</code>	0.4938	0.4938	0.5545	0.5545	0.5843	0.5323	0.6238	0.5743
<code>ipc</code>	0.1089	0.1114	0.1118	0.1114	0.1078	0.1151	0.1084	0.1078
<code>ipca</code>	2.1411	2.2409	2.1190	2.1321	1.8232	2.1804	2.2572	2.3297
<code>ipi</code>	0.1834	0.2022	0.1971	0.2053	0.2039	0.1883	0.183	0.1955
<code>real_exr</code>	0.6031	0.5967	0.8684	0.755	0.7592	0.7701	0.6000	0.5474

NOTA: A Tabela 3 apresenta a comparação entre os REQM para cada variável considerando diferentes horizontes de previsão. Em negrito, o esquema que apresentou o menor REQM.

Para os horizontes $h = 6$ e $h = 12$ não houve um padrão que permita determinar qual o melhor esquema de previsão. Algumas variáveis foram melhor previstas pelo esquema recursivo e outras pelo esquema de janela móvel. Algumas observações são relevantes. Primeiro, as variáveis U-rate, INCC, INPC, IPC e IPI mostraram previsão mais acurada ao se utilizar o esquema de janela móvel em quase todos os horizontes, exceção ao horizonte de seis meses à frente. O IPCA apresentou menor REQM em todos os horizontes para o esquema de janela móvel. Por sua vez, a taxa de câmbio real-dólar foi melhor prevista pelo esquema recursivo, com exceção do horizonte $h=6$.

Em síntese, esses resultados apontam que o esquema de janela móvel apresentou melhores resultados em termos de REQM para a maioria das variáveis e em quase todos os horizontes de previsão. Este resultado pode derivar da forma mais adequada pelo qual o esquema de janela móvel lida com as instabilidades dos parâmetros.

3 Conclusões

Este artigo comparou o poder de previsão de quatorze modelos para nove variáveis macroeconômicas brasileiras, incluindo taxa de desemprego, índice de produção industrial, taxa real de câmbio real-dólar e seis índices de preços: IPCA, IGP_10, INCC, INPC, IPC e IPA_10.

Os métodos utilizados para realizar as previsões foram modelos que combinavam fatores estimados seja por meio da seleção de variáveis, ou seja, através da ponderação de previsões. Os métodos utilizados foram Modelo de Fator Autoregressivo Aumentado (FAAR), Modelo de Fator Estático (PCR), Bagging, dois tipos de Modelos Bayesianos Ponderado (BMA), *Least Angle Regression* (LAR), *Elastic Net* (EN), *Non-Negative Garrote* (NNG), Modelo Mallows Ponderado (MMA), Modelo Jackknife Ponderado (JMA), Validação Cruzada *Leave-h-Out* (LHO), Modelo de Média Simples (SMA) e Modelo de Média Simples Geral (GSMA). Apenas o modelo AR(4) não inclui os fatores estimados por PCM. E apenas o PCR não inclui variáveis defasadas.

A base de dados contém 117 variáveis macroeconômicas entre 1995.05 a 2015.12 com frequência mensal. Essas variáveis buscam descrever o comportamento da economia brasileira, por isso incluem dentre outras índices de preços setoriais, dívidas externas e governamentais, alguns componentes da balança de pagamentos, importação e exportação de bens setoriais, salário, taxa de juros de longo prazo, indicadores financeiros, índices da atividade econômica, taxa de desemprego, algumas variáveis de economia internacional como importação e exportação dos EUA, taxa de juros de longo prazo dos EUA, consumo de energia elétrica e combustíveis, índices de produção setorial.

Dois métodos diferentes de estimação foram utilizados: o esquema recursivo e o de janela móvel. Os horizontes de previsão considerados foram um, três, seis e doze meses à frente. As previsões foram avaliadas utilizando a raiz quadrada do erro quadrático médio (REQM) e o *Model Confidence Set* (MCS).

Os resultados indicam que os modelos fatoriais híbridos selecionados pelo LAR e EN foram os que apresentaram maior poder preditivo, considerando todas as variáveis, todos os horizontes de previsão e ambos os esquemas, recursivo e de janela móvel. Além disso, o esquema de janela móvel apresentou menor valor absoluto do REQM em quase o dobro de vezes que o esquema recursivo, indicando que este esquema possui melhor poder de previsão para a maioria das variáveis em quase todos horizontes. Exceção foi a variável taxa de câmbio que foi melhor prevista pelo esquema recursivo.

REFERÊNCIAS BIBLIOGRÁFICAS.

- AGUIAR, Mark e GOPINATH, G. Emerging Market Business Cycles: The Cycle is the Trend. **Journal of Political Economy** 115 (1): 69-102, 2007.
- ARTIS, Michael, MARCELLINO e PROIETTI. Business Cycles in the New EU Member Countries and their Conformity with the Euro Area. *Journal of Business Cycle Measurement and Analysis*, OECD Publishing, Centre for International Research on Economic Tendency Surveys, vol. 1, 7-41, 2005
- ARRUDA, E. F., FERREIRA, R. T. and CASTELAR, I. Modelos Lineares e não lineares da curva de phillips para previsão da taxa de inflação no Brasil. **Revista Brasileira de Economia** 65, 237-252, 2011.
- BAI, J. Inferential theory for factor models of large dimensions. **Econometrica** 71, 135–171, 2003
- BAI, J. and NG, S., Determining the number of factors in approximate factor models. **Econometrica** 70, 191–221, 2002.
- BAI, J. and NG, S., Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. **Econometrica** 74, 1133–1150, 2006.
- BAI, J. and NG, S., Forecasting economic time series using targeted predictors. **J. Econometrics** 146, 304–317, 2009.
- BAI, J., NG, S., Boosting diffusion indices. **J. Appl. Econometrics** 4, 607–629, 2008.
- BOIVIN, J., and NG, S., Are more data always better for factor analysis? **J. Econometrics** 132, 169–194, 2006.
- BOVELSTAD, H.; NYGARD, S.; STORVOLD, H.; ALDRIN, M.; BOGAN, O.; FRIGUESI, A. and LINGJAERDE, O. Predicting Survival from Microarray Data, a Comparative Study , **Bioinformatics**, 23, 2080–2087, 2007.
- BREIMAN, L. Better subset regression using the nonnegative garrote. **Technometrics**, 37(4):373-384, 1995.
- CHAMBERLAIN, G. and ROTHSCCHILD, M. Arbitrage, Factor structure and mean-variance analysis in large asset markets. **Econometrica**, 51, 1305-1324,1983.
- CHENG, X. and HANSEN, B. Forecasting with Factor-Augmented Regression: A frequentist model averaging approach. **Journal of Econometrics**, 186, 280-293, 2015.
- DIAS, F., PINHEIRO, M. e RUA, A. Forecasting using targeted diffusion indexes, **Journal of Forecasting**, , vol. 29(3), 341-352, 2010.
- DIEBOLD, F. X. and MARIANO, R. S. Comparing predictive accuracy. **Journal of Business & Economic Statistics**, 13:253-63, 1995.
- EFRON, B., HASTIE, T., JOHNSTONE, L., and TIBSHIRANI, R. Least angle regression. **Annals of Statistics**, 32:407-499, 2004.
- EICKMEIER, S. and ZIEGLER, C. How successful are dynamic factor models at forecasting output and inflation? A meta-analytic approach, **Journal of Forecasting**, 27(3), 237-265, 2008.
- ELLIOT, G. e TIMMERMANN, A. Economic Forecasting. **Princeton University Press**, New Jersey, 2016.
- FERREIRA, BIERENS, H. e CASTELAR, I. Forecasting Quarterly Brazilian GDP Growth Rate With Linear and NonLinear Diffusion Index Models, *Economia*, vol. 6(3), p 261-292, 2005.
- FERREIRA, D. ; PALMA, A. A. . Forecasting Inflation with the Phillips Curve: a dynamic model averaging approach for Brazil. **Revista Brasileira de Economia** (Impresso), v. 69, p. 331-354, 2015.

- FIGUEREIDO, F. M. R. Forecasting Brazilian inflation using a large data set. **Brazilian Central Bank**, Working Paper Series, 228, December, 2010.
- GARCIA-CICCO, J.; PANCRAZI, R e URIBE, M. Real Business Cycles in Emerging Countries? **American Economic Review**, 100 (5), pp. 2510–2531, 2010
- GARCIA, M. MEDEIROS, M. e VASCONCELOS, G. Real-Time inflation forecasting with high dimensional models: the case of Brazil. XVI Encontro de Finanças, Rio de Janeiro, 2016.
- GELPER, S. and CROUX, C. Least angle regression for time series forecasting with many predictors, working paper. Technical report, **Katholieke Universiteit Leuven**, 2008.
- GOURINCHAS, P e OBSTEFELD, M. Stories of the Twentieth Century for the Twenty-First. **American Economic Journal: Macroeconomics**, 4(1), pp 226-265, 2012.
- HANSEN, B. E. Least squares model averaging. **Econometrica** 75, 1175–1189, 2007.
- HANSEN, B. E. Least squares forecasting averaging. **J. Econometrics** 146, 342–350. 2008.
- HANSEN, B. E. and RACINE, J.S. Jackknife model averaging. **J. Econometrics** 167, 38–46, 2012.
- HANSEN, P. LUNDE, A. e NASON, J. M. The Model Confidence Set. *Econometrica* Vol. 79 pp. 453-497, 2011.
- HASTIE, T.; TIBSHIRANI, R. and FRIEMAN, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer. 2^a Ed., 2009.
- INOUE, A., and KILIAN, L. How useful is bagging in forecasting economics time series? A case study of US CPI inflation. **J. Amer. Statist. Assoc.** 103, 511–522, 2008.
- KELLY, B., and PRUITT, S. The three-pass regression filter: A new approach to forecasting using many predictors. **J. Econometrics**, 2014
- KIM, H., and SWANSON, N. Forecasting financial and macroeconomic variables using data reduction methods: new empirical evidence. **J. Econometrics** 178, 352–367, 2014.
- KIM, H., and SWANSON, N. Mining Big Data Using Parsimonious Factor Machine Learning, Variable Selection, and Shrinkage Methods, **Rutgers University**, Working Paper, 2016.
- KOOP, G. and POTTER, S. Forecasting in dynamic factor models using bayesian model averaging. **Econometrics Journal**, 7:550-565, 2004.
- LUDVIGSON, S., NG, S. A factor analysis of bond risk premia. In: Ullah, A., Giles, D. (Eds.), **Handbook of Empirical Economics and Finance**. Chapman and Hall, pp. 313–372, 2011.
- KWIATKOWSKI, D; PHILLIPS, P.; SCHIMDT, P. and SHIN, Y. Testing the null hypothesis of stationarity against the alternative of a unit root. **Journal of Econometrics**, 54-9, 159-178, 1992.
- MALLOWS, C.L. Some comments on Cp. **Technometrics** 15, 661–675, 1973.
- MARCELLINO, M. A comparison of time series model forecasting GDP growth and inflation. **Journal of Forecasting**, 27, 305-340, 2008.
- MEDEIROS, M. C.; VASCONCELOS, G. and FREITAS, E. Forecasting Brazilian Inflation with High-Dimensional Models. **Brazilian Econometric Review**, forthcoming, 2016.
- NEUMEYER, Pablo e PERRI, Fabrizio. Business Cycles in Emerging Economies: the Role of Interest Rates. **Journal of Monetary Economics** 52 (2), pp. 345–380, 2005.
- ONATSKI, A. Determining the number of factors from empirical distribution of eigenvalues, **Review of Economics and Statistics** 92 (4), pp. 1004-1016, 2010.

- PESARAN, H. and TIMMERMANN, A. Selection of Estimation Window in the Presence of Breaks, **Journal of Econometrics** 137(1), 134-161, 2007.
- PESARAN, H.; PETENUZZO, D. e TIMMERMANN, A. Forecasting Time Series Subject to Multiple Structural Breaks. **Review of Economic Studies**, vol. 73, pp. 1057-1084, 2006.
- ROSSI, B. Advances in forecasting under instability. In Elliott, G. and Timmermann, A., editors, **Handbook of Economic Forecasting**, volume 2B, Chapter 21, pp. 1203-1324, 2012.
- ROSSI, B. e INOUE, A. Out-of-sample Forecast Tests Robust to the Window Size Choice, with A. Inoue, **Journal of Business and Economic Statistics** 30(3), pp. 432-453, 2012.
- SAIGO, H.; UNO, T. and TSUDA, K. Mining Complex Genotypic Features for Predicting HIV-1 Drug Resistance, **Bioinformatics**, 23, 2455–2462, 2007.
- SHOUSHA, Samer. Macroeconomic effects of commodity booms and busts: The Role of Financial Frictions. Working Paper, Columbia University, 2016.
- STOCK, J. H. and WATSON, M. W. A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. In: Engle, R., White, H. (Eds.), **Cointegration, Causality and Forecasting: A Festschrift for Clive W.J. Granger**. Oxford University Press, 1999.
- STOCK, J. H. and WATSON, M. W. Forecasting using principal components from a large number of predictors. **Journal of the American Statistical Association**, 97:1167-1179, 2002
- STOCK, J. H. and WATSON, M. W. Combination forecasts of output growth in a seven country data set. **Journal of Forecasting**, 23:405-430, 2004.
- STOCK, J. H. and WATSON, M. W. Implications of dynamic factor models for var analysis. NBER Working Papers 11467, **National Bureau of Economic Research**, Inc, 2005
- STOCK, J. H. and WATSON, M. W. Forecasting with many predictors. In Elliott, G., Granger, C., and Timmermann, A., editors, **Handbook of Economic Forecasting**, volume 1, chapter 10, pages 515-554. Elsevier, 2006
- STOCK, J. H. and WATSON, M. W. Generalized shrinkage methods for forecasting using many predictors. **Journal of Business and Economic Statistics**. 30(4):481-493, 2012.
- STONE, M. Cross-validatory choice and assessment of statistical predictions. **J. R. Stat. Soc. Ser. B** 36, 111–147, 1974.
- TIBSHIRANI, R. . Regression shrinkage and selection via the lasso. **Journal of the Royal Statistical Society, Series B**, 58:267-288, 1996.
- TU, Y., and LEE, T.H. Forecasting Using Supervised Factor Models. Working paper, **University of California, Riverside**, 2012.
- WATSON, M. and AMENGUAL, D. Consistent Estimation of the Number of Dynamic Factors in a Large N And T Panel. **Journal of Business and Economic Statistics**, January 2007.
- YUAN, M. and LIN, Y. On the non-negative garrotte estimator. **Journal of the Royal Statistical Society**, 69(2):143-161, 2007.
- ZOU, H. The adaptive lasso and its oracle properties. **Journal of the American Statistical Association**, 101:1418-1429, 2006.
- ZOU, H. e HASTIE, T. Regularization and variable selection via the elastic net, **J. R. Statist. Soc. B** Vol. 67, Part 2, pp. 301–320, 2005.

ZHANG, Ke; YIN, Fan e XIONG, S. Comparisons of penalized least squares methods by simulations. Working paper, Chinese Academy of Sciences, 2014.

CAPÍTULO 3- PREVISÃO DE VARIÁVEIS MACROECONÔMICAS BRASILEIRAS USANDO MODELOS FATORIAIS SUPERVISIONADOS.

RESUMO

Este artigo analisa o desempenho dos modelos fatoriais supervisionados frente aos não supervisionados para prever quatro variáveis macroeconômicas brasileiras, dentre elas, o índice de produção industrial, o índice de preço ao consumidor amplo, o índice de nacional de preços ao consumidor e a taxa de juros de longo de prazo. Os fatores são extraídos de um conjunto de dados composto por 117 variáveis macroeconômicas. Os fatores foram extraídos por meio de combinação de fatores, no caso, dos modelos supervisionados, e no caso dos modelos não supervisionados, foi utilizado o método de previsões combinadas, o mesmo utilizado por Tu e Lee (2012). Os modelos de previsão utilizados foram: Modelos Autoregressivos Aumentados (FAAR), Modelos Bayesianos Ponderados (BMA), modelo Mallows ponderado (MMA), Modelo Jackknife ponderado (JMA), Validação Cruzada Leave-h-Out (LHO) e o Modelo de Média Simples (SMA). O melhor modelo será o que apresenta a menor raiz quadrada do erro quadrático médio (REQM). Os resultados mostram que o esquema de janela móvel foi mais eficaz e o modelo de previsão que obteve melhor performance foi o modelo ponderado BMA, para ambos os modelos fatoriais, supervisionados ou não, além disso, os modelos supervisionados são mais eficazes para realizar previsões no médio prazo, pois previu com mais acurácia três, dentre as quatro, variáveis alvos, apresentando um menor (REQM).

Palavras chaves: Previsão, Modelo de Fatores Supervisionados, Variáveis Macroeconômicas Brasileiras.

ABSTRACT

This paper analyzes the performance of supervised versus unsupervised factorial models to predict four Brazilian macroeconomic variables, including industrial production index, broad consumer price index, national consumer price index, and interest rate long term. The factors are extracted from a set of data composed of 117 macroeconomic variables. The factors were extracted by means of a combination of factors, in the case of supervised models, and in the case of the unsupervised models, the combined forecast method was used, the same used by Tu and Lee (2012). The models used were: Autoregressive Models (FAAR), Weighted Bayesian Models (BMA), weighted Mallows model (MMA), weighted jackknife model (JMA), leave-h-out cross validation (LHO) and the mean model (SMA). The best model will be the one that presents the smallest square root of the mean square error (REQM). The results show that the mobile window scheme was more efficient and the prediction model that obtained better performance was the BMA weighted model, for both factorial models, supervised or not, in addition, supervised models are more efficient to carry out predictions in the Medium, since it predicted with more accuracy three of the four target variables, presenting a lower one (REQM).

Key words: Forecast, Model of Supervised Factors, Brazilian Macroeconomic Variables.

1 Introdução

Quase a totalidade das decisões econômicas tomadas pelos agentes (famílias, firmas, governos) dependem do valor futuro de uma ou mais variáveis. Quando uma família decide o quanto poupar, quanto ofertar de mão obra, pressupõe uma certa capacidade de prever um fluxo de salários futuros e ganhos de poupanças. Similarmente, quando uma firma decide o quanto investir, depende das suas previsões de capacidade de investimentos, da taxa de juros e do preço das ações. Além disso, decisões do governo de quanto gastar no período corrente e futuro, dependem das suas previsões sobre receitas e despesas.

Tudo isso, mostra que os problemas de previsões estão presentes, praticamente, em todas as áreas da economia e finanças. Porém, para solucionar esses problemas de previsão, quais variáveis preditoras devem ser incluídas e como serão selecionadas, qual a especificação correta do modelo de previsão a ser utilizado e qual o processo de estimação dos parâmetros desse modelo, são questões fundamentais para solução de qualquer trabalho que envolva previsão.

Uma das dificuldades encontradas em problemas de previsão é grande conjunto dimensional de preditores. Uma alternativa para solucionar esse problema é extrair características comuns dos dados, conhecidas como conjuntos de fatores, e, em seguida, usá-las como base para o modelo de previsão, conhecidos como modelos fatoriais em previsão. Esses modelos obtiveram sucesso devido ao fato de que modelos fatoriais conseguem resumir um grande conjunto de dados, representativos de cenários macroeconômicos, em alguns poucos fatores. Esses fatores possuem elevado conteúdo informacional sobre o comportamento da economia e, portanto, previsões neles baseadas evitariam o problema da escolha inadequada de variáveis independentes.

Stock and Watson (2002) definiram a metodologia padrão para utilizar o modelo fatoriais para previsão. O procedimento é dividido em duas etapas¹⁸. Na primeira etapa os fatores são estimados por meio da técnica de principais componentes (PC). No segundo estágio, é estimado por MQO um modelo linear que inclui os fatores estimados, e suas defasagens, e algumas variáveis de controle. É a partir deste último modelo que as previsões são realizadas. Evidentemente, nem todas as defasagens e não necessariamente todos os fatores estimados serão utilizados no modelo para gerar a previsão. O modelo ideal é escolhido por meio de critérios de informação como o Bayesiano (BIC).

Esta abordagem padrão possui algumas limitações. Primeiro, a estimação dos fatores é realizada sem considerar a variável que será prevista. Isto é, os fatores são estimados a partir de um conjunto grande dados, porém, o procedimento de estimação não considera qual será a variável sobre o qual será realizada a previsão. Esse tipo de procedimento é chamado de estimação de fatores não supervisionados. Modelos fatoriais que levam em

¹⁸Aqui supõe-se que número de fatores a ser estimado é conhecido a priori. Entratanto, na prática o número de fatores é estimado usando, na maioria das vezes, critérios de informação denidos em Bai e Ng (2002)

consideração a variável a ser prevista são chamados de modelos fatoriais supervisionados¹⁹. Tu e Lee (2012) e Boldrini e Hillenbrand (2016) mostram que a estimação de modelos fatoriais supervisionados tem melhor acurácia que modelos fatoriais não supervisionados.

A segunda limitação deve-se a desconsideração da combinação de previsões. De fato, modelos de combinações de previsões são bastante conhecidos na literatura empírica quanto ao seu sucesso em realizar previsões²⁰. Alguns pesquisadores tem direcionado atenção para a combinação de previsões baseadas em modelos fatoriais. Swanson and Kim (2013), Cheng and Hansen (2015) e Stock and Watson (2012) são alguns exemplos. É importante notar que o uso de combinações de previsões de fatores requer um cuidado adicional, tendo em vista que os fatores são estimados e não observados a priori, como ressalta Cheng and Hansen (2015).

Portanto, existem significativas evidências de que modelos que envolvam combinação de previsões de modelos fatoriais supervisionados possam gerar resultados mais poderosos em termos de previsão do que modelos fatoriais supervisionados ou não. Assim, o objetivo deste trabalho consiste em estimar e testar a acurácia de combinação de modelos fatoriais supervisionados. Para tanto, será verificado se tais tipos de modelos geram melhores previsões que modelos fatoriais não supervisionados com ou sem combinação de previsão. A base de dados é obtida de Stock e Watson (2012) e contém dados com frequência trimestral para os EUA, entre 1959 a 2011.

Existem diversas formas de supervisionar modelos fatoriais. Primeiro, o método dos Principal Covariance Regression (PCovR) desenvolvida por DeJong and Kiers (1992). Este método estima os fatores considerando a variável a ser prevista, por meio de minimização da média ponderada entre os erros de previsão e os erros de compressão. O segundo método foi desenvolvido por Tu e Lee (2012) sendo chamado de previsões combinadas usando componentes principais (CFPC). Este método, primeiramente, realiza uma previsão para cada uma das variáveis da base de dados. Em um segundo estágio, os fatores são estimados a partir da nova base de dados contendo apenas as previsões. Como apontaram Tu e Lee (2012) este método possui significativo sucesso na previsão para a taxa de inflação nos EUA. Uma vez estimados os fatores serão utilizadas técnicas de combinação de previsão. Como ressaltado acima, a combinação de previsões utilizando fatores requer um cuidado adicional, pois os fatores são gerados por estimações e não observados diretamente, ou seja, fatores latentes. O fato de serem gerados faz com que a distribuição dos métodos estatísticos seja alterada, como aponta Pagan (1986).

As combinações de previsões de modelos fatoriais supervisionados (CPFS) serão comparadas a uma gama de outros modelos de previsão usando fatores supervisionados ou não. Seis métodos de combinação de previsões: Modelos Autoregressivos Aumentados (FAAR), Modelos Bayesianos Ponderados (BMA), modelo Mallows ponderado (MMA),

¹⁹Ver Bair et al (2006)

²⁰Ver Timmerman et al (2010) e Bates and Granger (1969)

modelo jackknife ponderado (JMA), validação cruzada leave-h-out (LHO) e o modelo Cheng and Hansen (2015) mostraram que os critérios de Mallows e Leave-h-out podem ser utilizados em modelos fatoriais, pois, o fato de os fatores serem gerados aumenta o valor tanto do critério quanto do erro quadrático médio (EQM) e com isso, os tais critérios continuam informativos. O critério de Mallows é mais indicado diante da homoscedasticidade e o critério do Leave-h-out para o caso da presença de heteroscedaticidade. O critério de Jackknife é um caso especial do Leave-h-out, sendo igual a este quando em horizontes de previsão unitários ($h=1$)²¹.

Desta forma, o objetivo do presente artigo é comparar as previsões que utilizam modelos de fatores supervisionados com as fornecidas pelos modelos de fatores não supervisionados. As variáveis previstas nesse artigo serão, o índice nacional de preço ao consumidor (INPC), o índice de preço ao consumidor amplo (IPCA), índice de produção industrial (IPI) e a taxa de juros de longo prazo (TJLP). Para estimar os fatores supervisionados iremos utilizar a técnica utilizada por Tu e Lee (2016), (CFPC), e para estimar os fatores não supervisionados iremos utilizar o método dos componentes principais (PC), ambos extraído de uma base com 117 variáveis e com frequência mensal no período de 1996.5 à 2015.12.

Todas as variáveis são transformadas em estacionárias. Foram utilizados os testes Augmented Dickey-Fuller (ADF) e o KPSS, a fim de tentar detectar possíveis não-estacionariedades ou quebras nas séries analisadas. Como a instabilidade nos parâmetros é bastante comum em séries temporais econômicas para os países emergentes foi empregado duas formas de lidar com este problema. O método de previsão recursiva, em que os fatores são estimados no tempo t e suas previsões são realizadas no tempo $t + h$. O outro método de previsão considerado foi o com janelas móveis (rolling window), lidando, nesse sentido, com instabilidades estruturais, como discutido por Pesaran e Timmermann (2007) e Rossi (2012).

Os resultados mostram que o esquema de janela móvel foi mais eficaz, além disso, o modelo de previsão que obteve melhor performance foi o modelo ponderado BMA. Por fim, os modelos supervisionados são mais eficazes para realizar previsões no médio prazo, pois previu com mais acurácia três, dentre as quatro, variáveis alvos previstas, apresentando um menor (REQM).

O restante desse artigo está organizado da seguinte forma: seção 2 exhibe os principais resultados para de REQM relacionados a ambos os métodos, recursivo e janela móvel e a seção 3 apresenta as conclusões.

²¹Ver Hansen e Racine (2012)

2 Resultados

2.1 Modelos Não supervisionados.

2.1.2 Previsão usando o esquema recursivo.

Esta subseção apresenta os resultados dos modelos não supervisionados utilizando o esquema recursivo de previsão. Foi utilizado quadro horizontes de previsão, $h=1, 3, 6$ e 12 meses. A Tabela 6, mostra a REQM para cada variável do modelo. Em negrito estão marcados os modelos que apresentam o menor REQM. De acordo com essa tabela, as previsões para o IPI foram as que apresentaram o menor REQM. O SMA apresentou um bom desempenho para $h=3, 6$ e 12 períodos, pois apresentou menor REQM para a maioria das variáveis previstas. Para para $h=1$, o BMA previu com mais acurácia três das quatro variáveis previstas.

Tabela 6: Raiz do Erro Quadrático Médio (Esquema Recursivo).

Horizonte de Previsão $h=1$						
	FAAR	BMA	MMA	JMA	LHO	SMA
IPI	0.06468	0.06343	0.06485	0.06428	0.06428	0.06448
IPCA	2.47359	2.12266	2.23664	2.21660	2.21660	2.21376
INPC	4.15625	3.92758	4.16911	4.15021	4.15021	4.17920
TJLP	0.09773	0.09788	0.10050	0.10056	0.10056	0.11008
Horizonte de Previsão $h=3$						
	FAAR	BMA	MMA	JMA	LHO	SMA
IPI	0.06661	0.06580	0.06634	0.06496	0.06473	0.06475
IPCA	2.21881	2.40136	2.20066	2.20615	2.22890	2.08404
INPC	4.14772	4.41261	4.00033	3.96288	4.02912	3.82514
TJLP	0.09544	0.09635	0.10128	0.10087	0.10286	0.08651
Horizonte de Previsão $h=6$						
	FAAR	BMA	MMA	JMA	LHO	SMA
IPI	0.06892	0.06826	0.06963	0.06781	0.06830	0.06726
IPCA	2.30531	2.19908	2.17114	2.15551	2.11308	2.06540
INPC	4.57175	4.29634	3.92120	3.89699	3.86020	3.76036
TJLP	0.08010	0.07827	0.09911	0.09958	0.10115	0.10187
Horizonte de Previsão $h=12$						
	FAAR	BMA	MMA	JMA	LHO	SMA
IPI	0.06500	0.06268	0.06510	0.06452	0.06417	0.06519
IPCA	2.26240	2.26569	2.33207	2.27535	2.25340	2.14408
INPC	3.90251	4.03402	4.16252	4.09929	4.12237	3.80524
TJLP	0.10291	0.10283	0.10039	0.10033	0.10308	0.09496

NOTA: A Tabela 1 apresenta a comparação entre os REQM para cada variável considerando diferentes horizontes de previsão e os quatorze modelos de previsão para o esquema recursivo. Em negrito, o modelo que apresentou o menor REQM

2.1.3 Previsão usando esquema de janela móvel.

Esta subseção apresenta os resultados utilizando o esquema de janelas móveis. Para todos os horizontes o IPI foi novamente a variável melhor prevista, no sentido de apresentar o menor REQM. Entretanto, não há um modelo que gere previsões mais eficientes para a maior parte das variáveis, como no processo de estimação anterior. A maior recorrência refere-se as previsões mais eficientes do IPCA proporcionadas pelo SMA para $h=1$, 3 e 6. Vale ressaltar também que esse método gera menores RQME para três das quatro variáveis para $h=1$, 6 e 12 e para duas variáveis no caso de $h=3$. As previsões para IPI, IPCA e INPC possuem menores REQM para $h=6$ e 12, quando comparadas com os REQM produzidos pelo método recursivo.

Tabela 7: Raiz do Erro Quadrático Médio (Esquema Janela Móvel).

Horizonte de Previsão $h=1$						
	FAAR	BMA	MMA	JMA	LHO	SMA
IPI	0.07036	0.06844	0.06743	0.06492	0.06492	0.06872
IPCA	2.31884	2.33528	2.14171	2.14867	2.14867	2.09364
INPC	3.68108	4.23093	3.88140	3.89170	3.89170	3.78777
TJLP	0.08729	0.08587	0.09399	0.09455	0.09455	0.09731
Horizonte de Previsão $h=3$						
	FAAR	BMA	MMA	JMA	LHO	SMA
IPI	0.06849	0.06748	0.06608	0.06507	0.06493	0.06576
IPCA	2.08146	2.19271	2.12869	2.14036	2.12361	2.02178
INPC	3.86042	3.94287	4.02353	4.02803	4.07420	3.86472
TJLP	0.07338	0.07075	0.10234	0.10156	0.09905	0.08696
Horizonte de Previsão $h=6$						
	FAAR	BMA	MMA	JMA	LHO	SMA
IPI	0.06869	0.06758	0.06438	0.06422	0.06408	0.06464
IPCA	2.08112	2.30313	1.77089	1.76202	1.75884	1.72700
INPC	4.12354	4.21497	3.60956	3.60118	3.53306	3.61238
TJLP	0.09512	0.09357	0.09825	0.10296	0.10117	0.09910
Horizonte de Previsão $h=12$						
	FAAR	BMA	MMA	JMA	LHO	SMA
IPI	0.06072	0.05770	0.06850	0.06435	0.06440	0.07063
IPCA	2.14867	2.09727	1.94744	2.20405	2.19706	2.22735
INPC	3.65305	3.61238	3.91392	3.90999	4.00293	3.71912
TJLP	0.09728	0.09827	0.09736	0.09798	0.10117	0.09689

NOTA: A Tabela 2 apresenta a comparação entre os REQM para cada variável considerando diferentes horizontes de previsão e os quatorze modelos de previsão para o esquema de janela móvel. Em negrito, o modelo que apresentou o menor REQM.

3. Modelos Supervisionados.

Como ressaltado, os modelos supervisionados ao estimar os fatores, diferente dos modelos não supervisionados, considera a variável que será prevista e não desconsidera a combi-

nação de previsão. As subseções 8.1 e 8.2. mostram os resultados de previsões realizadas por modelos supervisionados, nos esquemas recursivos e janelas móveis, respectivamente.

3.1 Previsão usando o esquema recursivo.

A Tabela 8 mostra os resultados de previsões que utilizaram modelos supervisionados e o esquema recursivo, para os horizontes $h=1, 3, 6$ e 12 meses à frente. Comparando os resultados das Tabelas 6, 7 e 8, verifica-se que esse novo esquema de previsão gera previsões mais eficientes para o IPI ($h=3, 6$); IPCA($h=6$); INPC($h=6$) e TJLP ($h=1, 12$).

Tabela 8: Raiz do Erro Quadrático Médio (Esquema Recursivo).

Horizonte de Previsão $h=1$						
	FAAR	BMA	MMA	JMA	LHO	SMA
IPI	0.07397	0.07202	0.08198	0.07570	0.07570	0.07125
IPCA	2.53305	2.25021	2.14234	2.16415	2.16415	2.44883
INPC	4.92397	4.24466	4.26248	4.26040	4.26040	4.42101
TJLP	0.08383	0.08338	0.08810	0.09071	0.09071	0.09322
Horizonte de Previsão $h=3$						
	FAAR	BMA	MMA	JMA	LHO	SMA
IPI	0.06623	0.06484	0.06503	0.06445	0.06448	0.06451
IPCA	2.31704	2.36676	2.13415	2.14678	2.14904	2.09203
INPC	4.17267	4.21362	4.11741	4.13373	4.12930	4.00041
TJLP	0.09535	0.09349	0.09867	0.10093	0.10263	0.09879
Horizonte de Previsão $h=6$						
	FAAR	BMA	MMA	JMA	LHO	SMA
IPI	0.06584	0.06484	0.06421	0.06422	0.06406	0.06425
IPCA	2.06734	1.99563	1.78431	1.80038	1.87553	1.80527
INPC	3.59288	3.54287	3.29356	3.30938	3.55635	3.33847
TJLP	0.09218	0.09074	0.10369	0.10143	0.10296	0.10625
Horizonte de Previsão $h=12$						
	FAAR	BMA	MMA	JMA	LHO	SMA
IPI	0.06524	0.06196	0.06491	0.06444	0.06418	0.06530
IPCA	1.99005	2.06729	2.33665	2.25705	2.34521	2.16782
INPC	3.89498	3.94291	3.94360	3.93039	4.11043	3.75285
TJLP	0.10452	0.10117	0.09969	0.09987	0.10038	0.09178

NOTA: A Tabela 3 apresenta a comparação entre os REQM para cada variável considerando diferentes horizontes de previsão e os quatorze modelos de previsão para o esquema recursivo. Em negrito, o modelo que apresentou o menor REQM

3.2 Previsão usando o esquema janela móveis

A Tabela 9 mostra os resultados de previsões que utilizaram modelos supervisionados estimados pelo método de janelas móveis. Para todos os horizontes, a variável que apresentou o menor REQM foi o IPI. Comparando com os resultados anteriores, esse procedimento gera previsões mais eficientes para o IPI ($h=1, 3, 6$); IPCA ($h=6$) e INPC ($h=3, 6, 12$)

Tabela 9: Raiz do Erro Quadrático Médio (Esquema Janela Móvel)

Horizonte de Previsão h=1						
	FAAR	BMA	MMA	JMA	LHO	SMA
IPI	0.05640	0.05638	0.07262	0.07266	0.07266	0.07918
IPCA	2.89083	2.83286	2.41875	2.41991	2.41991	2.20308
INPC	4.79160	4.76933	3.85178	3.85955	3.85955	3.81511
TJLP	0.10782	0.10243	0.10392	0.10657	0.10657	0.10510
Horizonte de Previsão h=3						
	FAAR	BMA	MMA	JMA	LHO	SMA
IPI	0.06715	0.06723	0.06599	0.06427	0.06422	0.06619
IPCA	2.03437	2.05413	2.10029	2.09290	2.10906	2.04378
INPC	4.06947	4.02932	3.99982	3.99004	3.96578	3.76907
TJLP	0.09450	0.09175	0.09982	0.09975	0.09917	0.09796
Horizonte de Previsão h=6						
	FAAR	BMA	MMA	JMA	LHO	SMA
IPI	0.06591	0.06404	0.06689	0.06517	0.06429	0.06428
IPCA	2.08935	1.96491	1.80436	1.81111	1.78205	1.64892
INPC	3.78168	3.54951	3.55740	3.60853	3.46663	3.25783
TJLP	0.10945	0.10517	0.10868	0.10831	0.10593	0.10851
Horizonte de Previsão h=12						
	FAAR	BMA	MMA	JMA	LHO	SMA
IPI	0.06950	0.06441	0.06578	0.06452	0.06458	0.06767
IPCA	1.99420	1.98416	2.02934	2.05244	2.24599	2.06230
INPC	3.53288	3.46914	3.77075	3.79250	3.98500	3.72080
TJLP	0.10981	0.11133	0.09911	0.09922	0.10161	0.09781

NOTA: A Tabela 4 apresenta a comparação entre os REQM para cada variável considerando diferentes horizontes de previsão e os quatorze modelos de previsão para o esquema de janela móvel. Em negrito, o modelo que apresentou o menor REQM.

3.3 Modelos supervisionados versus não supervisionados.

Nessa vamos realizar uma análise comparativa entre as previsões com modelos supervisionados e não supervisionados nos esquemas recursivo e de janela móvel por meio do valor absoluto do REQM. A Tabela 10 apresenta os resultados do REQM para os melhores modelos selecionados em horizonte de previsão e para cada variável. Os termos em negrito indicam qual dos esquemas de previsão possui menor REQM.

Ao observar a tabela 10, percebe-se que o esquema janela móvel apresenta um menor REQM em quase todos os horizontes e para a maior parte das variáveis, isso tanto para modelos supervisionados como para modelos não supervisionados. O esquema recursivo obteve um menor REQM no curtíssimo prazo para o IPI e no médio, h=6, e longo prazo, h=12, para a TJLP, modelos não supervisionados, já nos modelos supervisionados o esquema recursivo obteve um melhor desempenho preditivo para IPI em h=12, o IPCA para h=1, e para a TJLP para h=1, 6 e 12.

Tabela 10: Modelos Supervisionados versus Não Supervisionados

Horizonte de previsão.								
Variáveis	Modelos Não Supervisionados							
	h=1		h=3		h=6		h=12	
	Janela Móvel	Recursivo	Janela Móvel	Recursivo	Janela Móvel	Recursivo	Janela Móvel	Recursivo
IPI	0.06492	0.06343	0.06493	0.06473	0.06422	0.06726	0.05770	0.06268
IPCA	2.09364	2.12266	2.02178	2.08404	1.72700	2.06540	1.94744	2.14408
INPC	3.68108	3.92758	3.86042	3.96288	3.61238	3.76036	3.61238	3.80524
TJLP	0.08587	0.09773	0.07338	0.08651	0.09357	0.07827	0.09689	0.09496
Modelos Supervisionados								
Variáveis	h=1		h=3		h=6		h=12	
	Janela Móvel	Recursivo	Janela Móvel	Recursivo	Janela Móvel	Recursivo	Janela Móvel	Recursivo
IPI	0.05638	0.07125	0.06422	0.06445	0.06404	0.06421	0.06441	0.06196
IPCA	2.20308	2.14234	2.03437	2.09203	1.64892	1.78431	1.98416	1.99005
INPC	3.81511	4.42101	3.76907	4.00041	3.25783	3.29356	3.46914	3.75285
TJLP	0.10243	0.08338	0.09175	0.09349	0.10517	0.09074	0.09911	0.09178

Observando o menor REQM para cada variável e em cada horizonte de previsão ao analisar os esquemas recursivos e de janelas móveis e, logo após, confrontando esses resultados para cada modelo, supervisionado e não supervisionado, consta-se que o modelo supervisionado apresentou menor REQM no curtíssimo prazo, $h=1$, para o IPI e a TJLP; no curto prazo, $h=3$, para o IPI e IPCA; no médio prazo, para o IPI, IPCA e INPC, e por fim, no longo prazo, para o INPC e para a TJLP.

4 Conclusão.

Ao estimar fatores considerando a variável a ser prevista, queremos analisar o poder de previsão dos modelos de fatores supervisionados. Esse artigo comparou o poder de previsão de modelos de fatores supervisionados e não supervisionados para quatro variáveis, índice de produção industrial, índice de preços: IPCA e INPC; e para taxa de juros de longo prazo, tanto no esquema recursivo como no de janela móvel.

Foi utilizado os métodos de previsão FAAR, BMA, MMA, JMA, LHO e SMA, tanto para modelos de fatores supervisionados como para não supervisionados, Para estimar os fatores supervisionados foi utilizada a técnica proposta Tu e Lee (2016), (CFPC), e para estimar os fatores não supervisionados foi utilizado o método dos componentes principais (PC).

A base de dados utilizada contém 236 observações de periodicidade mensal de 1996.5 à 2015.12, utilizamos 117 variáveis, dentre elas, diversos índices de preços setoriais, dívidas externas e governamentais, taxa de juros de longo prazo, com o intuito de descrever o comportamento da economia brasileira.

Os métodos recursivos e de janela móvel foram utilizados no processo de estimação.

Além disso, a previsão foi realizada para um, três, seis e doze períodos à frente e avaliadas por meio da raiz quadrada do erro quadrático médio (REQM).

Os resultados mostram que o esquema de janela móvel foi mais eficaz, além disso, o modelo de previsão que obteve melhor performance foi o modelo ponderado BMA. Por fim, os modelos supervisionados são mais eficazes para realizar previsões no médio prazo, pois previu com mais acurácia três, dentre as quatro, variáveis alvos previstas, apresentando um menor (REQM).

REFERÊNCIAS BIBLIOGRÁFICAS.

- AGUIAR, Mark e GOPINATH, G. Emerging Market Business Cycles: The Cycle is the Trend. **Journal of Political Economy** 115 (1): 69-102, 2007.
- ARTIS, Michael, MARCELLINO e PROIETTI. Business Cycles in the New EU Member Countries and their Conformity with the Euro Area. *Journal of Business Cycle Measurement and Analysis*, OECD Publishing, Centre for International Research on Economic Tendency Surveys, vol. 1, 7-41, 2005
- ARRUDA, E. F., FERREIRA, R. T. and CASTELAR, I. Modelos Lineares e não lineares da curva de phillips para previsão da taxa de inflação no Brasil. **Revista Brasileira de Economia** 65, 237-252, 2011.
- BAI, J. Inferential theory for factor models of large dimensions. **Econometrica** 71, 135–171, 2003
- BAI, J. and NG, S., Determining the number of factors in approximate factor models. **Econometrica** 70, 191–221, 2002.
- BAI, J. and NG, S., Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. **Econometrica** 74, 1133–1150, 2006.
- BAI, J. and NG, S., Forecasting economic time series using targeted predictors. **J. Econometrics** 146, 304–317, 2009.
- BAI, J., NG, S., Boosting diffusion indices. **J. Appl. Econometrics** 4, 607–629, 2008.
- BOIVIN, J., and NG, S., Are more data always better for factor analysis? **J. Econometrics** 132, 169–194, 2006.
- BOVELSTAD, H.; NYGARD, S.; STORVOLD, H.; ALDRIN, M.; BOGAN, O.; FRIGUESI, A. and LINGJAERDE, O. Predicting Survival from Microarray Data, a Comparative Study , **Bioinformatics**, 23, 2080–2087, 2007.
- BREIMAN, L. Better subset regression using the nonnegative garrote. **Technometrics**, 37(4):373-384, 1995.
- CHAMBERLAIN, G. and ROTHSCCHILD, M. Arbitrage, Factor structure and mean-variance analysis in large asset markets. **Econometrica**, 51, 1305-1324, 1983.
- CHENG, X. and HANSEN, B. Forecasting with Factor-Augmented Regression: A frequentist model averaging approach. **Journal of Econometrics**, 186, 280-293, 2015.
- DIAS, F., PINHEIRO, M. e RUA, A. Forecasting using targeted diffusion indexes, **Journal of Forecasting**, , vol. 29(3), 341-352, 2010.
- DIEBOLD, F. X. and MARIANO, R. S. Comparing predictive accuracy. **Journal of Business & Economic Statistics**, 13:253-63, 1995.
- EFRON, B., HASTIE, T., JOHNSTONE, L., and TIBSHIRANI, R. Least angle regression. **Annals of Statistics**, 32:407-499, 2004.
- EICKMEIER, S. and ZIEGLER, C. How successful are dynamic factor models at forecasting output and inflation? A meta-analytic approach, **Journal of Forecasting**, 27(3),

237-265, 2008.

ELLIOT, G. e TIMMERMANN, A. Economic Forecasting. **Princeton University Press**, New Jersey, 2016.

FERREIRA, BIERENS, H. e CASTELAR, I. Forecasting Quarterly Brazilian GDP Growth Rate With Linear and NonLinear Diffusion Index Models, *Economia*, vol. 6(3), p 261-292, 2005.

FERREIRA, D. ; PALMA, A. A. . Forecasting Inflation with the Phillips Curve: a dynamic model averaging approach for Brazil. **Revista Brasileira de Economia** (Impresso), v. 69, p. 331-354, 2015.

FIGUEREIDO, F. M. R. Forecasting Brazilian inflation using a large data set. **Brazilian Central Bank**, Working Paper Series, 228, December, 2010.

GARCIA-CICCO, J.; PANCRRAZI, R e URIBE, M. Real Business Cycles in Emerging Countries? **American Economic Review**, 100 (5), pp. 2510–2531, 2010

GARCIA, M. MEDEIROS, M. e VASCONCELOS, G. Real-Time inflation forecasting with high dimensional models: the case of Brazil. XVI Encontro de Finanças, Rio de Janeiro, 2016.

GELPER, S. and CROUX, C. Least angle regression for time series forecasting with many predictors, working paper. Technical report, **Katholieke Universiteit Leuven**, 2008.

GOURINCHAS, P e OBSTEFELD, M. Stories of the Twentieth Century for the Twenty-First. **American Economic Journal: Macroeconomics**, 4(1), pp 226-265, 2012.

HANSEN, B. E. Least squares model averaging. **Econometrica** 75, 1175–1189, 2007.

HANSEN, B. E. Least squares forecasting averaging. **J. Econometrics** 146, 342–350. 2008.

HANSEN, B. E. and RACINE, J.S. Jackknife model averaging. **J. Econometrics** 167, 38–46, 2012.

HANSEN, P. LUNDE, A. e NASON, J. M. The Model Confidence Set. *Econometrica* Vol. 79 pp. 453-497, 2011.

HASTIE, T.; TIBSHIRANI, R. and FRIEMAN, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer. 2^a Ed., 2009.

INOUE, A., and KILIAN, L. How useful is bagging in forecasting economics time series? A case study of US CPI inflation. **J. Amer. Statist. Assoc.** 103, 511–522, 2008.

KELLY, B., and PRUITT, S. The three-pass regression filter: A new approach to forecasting using many predictors. **J. Econometrics**, 2014

KIM, H., and SWANSON, N. Forecasting financial and macroeconomic variables using data reduction methods: new empirical evidence. **J. Econometrics** 178, 352–367, 2014.

KIM, H., and SWANSON, N. Mining Big Data Using Parsimonious Factor Machine Learning, Variable Selection, and Shrinkage Methods, **Rutgers University**, Working Paper, 2016.

KOOP, G. and POTTER, S. Forecasting in dynamic factor models using bayesian model

- averaging. **Econometrics Journal**, 7:550-565, 2004.
- LUDVIGSON, S., NG, S. A factor analysis of bond risk premia. In: Ullah, A., Giles, D. (Eds.), **Handbook of Empirical Economics and Finance**. Chapman and Hall, pp. 313–372, 2011.
- KWIATKOWSKI, D; PHILLIPS, P.; SCHIMDT, P. and SHIN, Y. Testing the null hypothesis of stationarity against the alternative of a unit root. **Journal of Econometrics**, 54-9, 159-178, 1992.
- MALLOWS, C.L. Some comments on Cp. **Technometrics** 15, 661–675, 1973.
- MARCELLINO, M. A comparison of time series model forecasting GDP growth and inflation. **Journal of Forecasting**, 27, 305-340, 2008.
- MEDEIROS, M. C.; VASCONCELOS, G. and FREITAS, E. Forecasting Brazilian Inflation with High-Dimensional Models. **Brazilian Econometric Review**, forthcoming, 2016.
- NEUMEYER, Pablo e PERRI, Fabrizio. Business Cycles in Emerging Economies: the Role of Interest Rates. **Journal of Monetary Economics** 52 (2), pp. 345–380, 2005.
- ONATSKI, A. Determining the number of factors from empirical distribution of eigenvalues, **Review of Economics and Statistics** 92 (4), pp. 1004-1016, 2010.
- PESARAN, H. and TIMMERMANN, A. Selection of Estimation Window in the Presence of Breaks, **Journal of Econometrics** 137(1), 134-161, 2007.
- PESARAN, H.; PETENUZZO, D. e TIMMERMANN, A. Forecasting Time Series Subject to Multiple Structural Breaks. **Review of Economic Studies**, vol. 73, pp. 1057-1084, 2006.
- ROSSI, B. Advances in forecasting under instability. In Elliott, G. and Timmermann, A., editors, **Handbook of Economic Forecasting**, volume 2B, Chapter 21, pp. 1203-1324, 2012.
- ROSSI, B. e INOUE, A. Out-of-sample Forecast Tests Robust to the Window Size Choice, with A. Inoue, **Journal of Business and Economics Statistics** 30(3), pp. 432-453, 2012.
- SAIGO, H.; UNO, T. and TSUDA, K. Mining Complex Genotypic Features for Predicting HIV-1 Drug Resistance, **Bioinformatics**, 23, 2455–2462, 2007.
- SHOUSA, Samer. Macroeconomic effects of commodity booms and busts: The Role of Financial Frictions. Working Paper, Columbia University, 2016.
- STOCK, J. H. and WATSON, M. W. A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. In: Engle, R., White, H. (Eds.), **Cointegration, Causality**

CAPÍTULO 4- PREVISÃO POR MEIO DA COMBINAÇÃO DINÂMICA DE MODELOS FATORIAIS.

RESUMO

Este artigo propõe um método de ponderação dinâmica de previsores (DMA) aplicado a grandes bases de dados. As variáveis contidas na base de dados tem a dimensão reduzida a um número r de fatores que são combinados dinamicamente por meio de fatores de esquecimento. A extração dos fatores considera também uma janela exponencial de esquecimento que visa reduzir o impacto de observações muito antigas. Mostra-se que tal modelo, chamado de FDMA, converge assintoticamente para a combinação dinâmica de fatores observados quando o número de variáveis ou o tamanho da amostra de fora cresce. Além disso, o modelo FDMA é aplicado em dois diferentes exercícios empíricos. Primeiro para prever seis variáveis macroeconômicas americanas, incluindo variáveis reais e nominais, e após para prever o excesso de retorno do S&P 500. Os resultados empíricos mostram que o FDMA e suas variações apresentam-se como uma alternativa promissora para previsão.

Palavras chaves: Previsão, Modelos de Fatores, Combinação de Previsão.

ABSTRACT

This paper propose a method of dynamic models average (DMA) applied to large datasets. The dimension of the variables in the dataset is reduced to r factors that are dynamically combined using forgetting parameters. The factor estimation also consider a exponential forgetting window which the older observations are excluded from sample. The model, called FDMA, converge asymptotically for dynamic weight of observed factors, when the number of variables or the size of out-of-sample grow to infinity. The FDMA are applied in two empirical exercises. First it is used to forecast six US macroeconomic variables, including real and nominal variables. Second, it is applied to forecast the equity premium of S&P 500. In both exercises, the FDMA has satisfactory performance.

Key words: Forecast, Diffusion Index, Forecast Combination.

1 Introdução

O objetivo deste artigo é avaliar o poder de previsão do método de combinação dinâmica de variáveis (DMA), desenvolvido por Raftery et al (2010), considerando um grande número de previsores. O DMA consiste num método de combinar dinamicamente as previsões em que é permitido que tanto os parâmetros quanto os modelos de previsão se modifiquem ao longo do tempo.

O DMA é um método computacionalmente intensivo, portanto, a sua utilização em modelos contendo muitos previsores não é factível. Koop and Korobilis (2012), por exemplo, estimam uma curva de Phillips generalizada utilizando DMA em modelos com apenas sete variáveis exógenas, além das variáveis defasadas.

Contudo, em economia, existem grandes bases de dados disponíveis e a escolha de apenas algumas variáveis pode não ser adequada, especialmente quando se objetiva a construção de previsões ótimas. Algumas alternativas tem sido propostas como em Onorante e Raftery (2016) que analisam o método de janela de Occam para estimar o DMA usando grandes bases de dados.

Este artigo, por sua vez, propõe um método alternativo que permite realizar previsões usando o DMA em grandes bases de dados. A ideia básica é a utilização da análise fatorial para reduzir a dimensão do conjunto de dados a poucas variáveis, chamadas de fatores. Seja X_{it} um painel em que $i = 1, \dots, N$ corresponde a N variáveis macroeconômicas e $t = 1, \dots, T$ a variação temporal. Será assumido que X_{it} possui uma estrutura fatorial aproximada, isto é:

$$X_{it} = \lambda_i' F_t + e_{it} \quad (37)$$

Em que: F_t são chamados de fatores; λ_i são os fatores de carga e $e_{it} \sim N(0, \sigma^2)$ é o erro de aproximação de X_{it} por $\lambda_i' F_t$. A hipótese de que X_{it} possui uma estrutura fatorial implica que a dimensão de X_{it} poderá ser reduzida a $r \ll N$ fatores, em que cada fator representa uma parcela da variabilidade de X_{it} . Os fatores serão estimados por meio do método de componentes principais (PCM)²². Tais fatores estimados, em conjunto com algumas variáveis exógenas, constituirão a equação de previsão.

Assuma que se objetive prever a variável y_t , h períodos à frente, utilizando o conjunto de informação em t , $y_{t+h|t}$. Seja W_t um conjunto de variáveis exógenas, normalmente representadas pelas defasagens de y_t . Então a equação de previsão é dada por:

$$y_{t+h|t} = \beta(L)\hat{F}_t + \alpha(L)W_t + \varepsilon_t \quad (38)$$

Em que: $\beta(L)$ e $\alpha(L)$ correspondem a ordem das defasagens de \hat{F} e W , respectivamente. Esta forma de realizar previsão é chamada de índices de difusão, ver Stock

²²Para maiores detalhes sobre o método dos componentes principais, ver Sotck e Watson (2002).

and Watson (2002). O DMA fará a escolha do melhor modelo de previsão a cada período considerando todas as possíveis combinações da variável $Z_t = (\hat{F}_t, W_t)$.

Em certo sentido, o procedimento proposto constitui uma nova forma de seleção do modelo de previsão, diferentemente da abordagem tradicional proposta por Stock e Watson (2002). Na abordagem tradicional a equação (38), após estimado os fatores, é escolhida utilizando critérios de informação, como o BIC e o AIC. Bai e Ng (2008) argumentam que tal procedimento é inadequado, pois ao contrário de modelos autorregressivos, os fatores não possuem uma ordenação natural entre si. Cheng e Hansen (2015) e Rossi e Carrasco (2016) propõe critérios alternativos, como o critério de Mallows e o critério de validação cruzada generalizada²³ para selecionar as variáveis em (38).

O método proposto aqui acrescenta duas inovações a forma tradicional de seleção das variáveis que compõe (38). Por um lado, a utilização do DMA permite que os parâmetros da equação de previsão se modifiquem no tempo. Esta característica é bastante interessante tendo em vista que instabilidades nos parâmetros podem ser prejudiciais as previsões, como aponta Rossi (2012).

Por outro lado, a própria forma funcional do modelo poderá se alterar, possibilitando que diferentes modelos possam ser escolhidos em cada período. Ou seja, a cada período o modelo de previsão ideal pode mudar, permitindo com isso uma previsão mais acurada em longos períodos de tempo. De fato, existem evidências de que os melhores modelos de previsão mudam no tempo. Stock e Watson (1998), por exemplo, apontam que os modelos de previsão que eram bastante adequados antes do período da Grande Moderação²⁴ não apresentavam acurácia significativa durante a Grande Moderação.

A utilização de índices de difusão ponderados pelo DMA foi inicialmente proposta por Koop and Korobilis (2011), em que analisaram o poder de previsão de fatores extraídos de blocos de variáveis macroeconômicas. Cada fator \hat{F}_t era obtido a partir de um grupo de variáveis comuns, como índices de preços, variáveis reais, variáveis financeiras etc. O motivo para adoção desta estimação em blocos de variáveis, segundo os autores, deve-se a possibilidade da interpretação do comportamento dos fatores para a previsão y_t ao longo do tempo.

Todavia, existem algumas limitações nesta abordagem. Primeiro, Bai (2008) demonstrou que o erro de estimação dos fatores não seria prejudicial a previsão se $N, T \rightarrow \infty$ e $\sqrt{N}/T \rightarrow \infty$. O motivo para essa constatação, segundo Ng (2012), é que à medida que $N \rightarrow \infty$, a variação em ϵ_{it} passa a ser dominada pela estrutura $\lambda_i' F_t$, possibilitando a aproximação de X_{it} por meio de $\hat{\lambda}_i' \hat{F}_t$. O procedimento adotado por Koop and Korobilis

²³Cheng e Hansen (2015) demonstraram que sob algumas hipóteses o critério de Mallows permite selecionar adequadamente a forma funcional de (2), uma vez que diante de variáveis previamente estimadas o critério de Mallows aumenta o seu valor assim como a função perda, mantendo, portanto a característica de seleção. Já Rossi e Carrasco (2016) mostraram que o critério de validação cruzada generalizada é Asymptotic loss efficient.

²⁴A Grande Moderação corresponde a um período de baixa volatilidade nos ciclos de negócios que iniciou-se por volta de 1980 e estendendo-se até a Grande Crise Financeira de 2008.

(2011) não parece adequado, pois reduz N para a extração de fatores, possivelmente gerando uma aproximação imprecisa. Portanto, não parece plausível confiar, por exemplo, que fatores estimados considerando apenas os índices de preço sejam uma boa aproximação dos próprios índices.

Uma limitação adicional decorre da não consideração da interdependência entre as variáveis contidas em X_{it} . A estrutura fatorial em (37) possui matriz de variância-covariância dada por:

$$\Sigma_X = A' \Sigma_F A + \Sigma_e$$

Em que: Σ_F é a matriz de variância de F . Bai (2002) comenta que a hipótese de independência entre as variáveis sobre as quais são extraídos os fatores não parece apropriada a modelos econômicos. De fato, na abordagem clássica da análise fatorial é assumido que Σ_e é uma matriz diagonal. Economicamente tal hipótese não faz sentido, tendo em vista a possível correlação entre as variáveis contidas em X_{it} .

Chamberlain e Rothschild (1983) assumem uma hipótese mais realista de que os autovalores da matriz $\Sigma_e = E(e_t e_t')$ são limitados, isto é, que existe uma correlação fraca, serial e transversalmente, nos erros e_{it} , passando tal estrutura fatorial ser chamada de modelo fatoriais aproximados. A abordagem de Koop e Korobilis (2011), portanto, ao separar as variáveis em grupos comuns não parece interessante pois o efeito das correlações entre as variáveis pode não ser captados pelos fatores estimados²⁵.

Em diversos pontos o método proposto neste trabalho se distingue do método de Koop and Korobilis (2011). Primeiro, não será assumido nenhum tipo de bloco de variáveis. Ou seja, os fatores serão estimados a partir de um conjunto grande de dados sem que haja nenhum tipo de separação *ad-hoc* das variáveis.

Segundo, será demonstrado que o erro quadrático de previsão do modelo DMA usando fatores converge assintoticamente para o erro quadrático de um modelo cuja estrutura seja fatorial. Este resultado indica que caso a estrutura da base de dados possa ser representada fatorialmente, então, assintoticamente, a adoção da combinação dinâmica de fatores não será prejudicial a previsão.

Terceiro, serão propostos algumas extensões a abordagem dos fatores ponderados pelo DMA cujo objetivo é buscar formas mais eficientes de estimar os fatores. Isso é feito ao estimar os fatores considerando quais variáveis serão previstas²⁶. Três formas alternativas de estimação dos fatores serão propostas: fatores estimados em variáveis previamente selecionadas pelo método do Lasso (*Least Absolute Selector and Shrinkage*

²⁵Suponha que $X_{it} = (X_{it}^1, X_{it}^2)$, em que: X_{it}^k corresponde a um bloco de variáveis, $k = 1, 2$. Então: $\Sigma_X = \begin{pmatrix} \sum X^1 & \sum X^1 \sum X^2 \\ \sum X^2 \sum X^1 & \sum X^2 \end{pmatrix}$. Assim, ao estimar apenas para X_{it}^1 , por exemplo, a correlação entre X^1 e X^2 não é considerada, mesmo que seja relevante.

²⁶Essa forma de estimação faz com que os fatores sejam chamados de fatores supervisionados, Ver Bair et al (2006), Tu e Lee (2013).

Operator) ou Lar (*Least Angle Regression*) como em Bai e Ng (2008) e o método de supervisão do CFPC (*Combining Forecasts using Principal Components*) desenvolvido por Tu e Lee (2013). Estes métodos de pré-seleção de variáveis mostraram-se bastante promissores e serão aqui considerados como uma alternativa ao modelo principal.

Por fim, o método proposto será aplicado em dois exercícios empíricos. Primeiro, será realizada a previsão por meio do DMA usando fatores para seis variáveis macroeconômicas dos EUA: índice de preços ao produtor (IPI) e ao consumidor (CPI); índice de produção industrial (IP), M2, house starts (Hous) e taxa de desemprego (Unrate). Os fatores serão estimados de uma base de dados contendo 115 variáveis macroeconômicas, com frequência mensal entre 1963.1 até 2015.12²⁷.

As previsões serão comparadas a modelos que são reconhecidos na literatura empírica como tendo boa performance em prever variáveis macroeconômicas utilizando variáveis geradas. Estes modelos de comparação são os que utilizam os critérios de Mallows (MMA), Leave-h-out e Jakkinfe (CHENG and HANSEN, 2015) para a seleção de fatores, métodos bayesianos de combinação de previsões (KOOP and POTTER, 2004 e KIM and SWANSON, 2014), previsões com pesos iguais, modelos fatoriais aumentados (STOCK and WATSON, 2002, 2012). Será considerada também as extensões em que o DMA é aplicado a fatores supervisionados.

O segundo exercício empírico consiste em verificar o poder de previsão do DMA usando fatores sobre o excesso de retorno do índice S&P 500. A separação destes dois exercícios deve-se ao fato de que os modelos utilizados para fazer a comparação da previsão são diferentes em cada aplicação. Para prever o *equity premium* será testado o modelo fatorial combinado dinamicamente por meio do DMA (e suas extensões) contra alguns modelos bastante utilizados na literatura empírica, como: modelos com variáveis individuais²⁸, modelo *kitchen sink*, modelo de média histórica e modelos fatoriais tradicionais²⁹.

Este trabalho, portanto, tem o objetivo de contribuir para literatura sobre previsão de variáveis macroeconômicas e financeiras. Primeiro, ao propor selecionar os fatores contidos no modelo de previsão por meio do DMA, permite que esta seleção seja feita considerando tanto a variação dos parâmetros quanto a variação dos modelos ao longo do tempo. Segundo, por mostrar que o efeito sobre o erro de previsão causado por este método é assintoticamente negligenciado. Terceiro, ao realizar dois exercícios empíricos em variáveis tradicionalmente difíceis de prever.

²⁷Os dados foram obtidos no banco de dados do Federal Reserve of St. Louis (FRED). Ver apêndice para maiores detalhes.

²⁸Este modelo utiliza apenas uma variável para prever o *equity premium*.

²⁹Este modelo fatorial tradicional refere-se ao modelo cujos fatores são extraídos apenas do conjunto de dados financeiros.

2 Resultados Empíricos

2.1 Variáveis Macroeconômicas dos E.U.A.

Nesta seção são apresentados os resultados para o exercício empírico de previsão do DMA usando fatores para seis variáveis macroeconômicas dos EUA: índice de preços ao produtor (IPI) e ao consumidor (CPI); índice de produção industrial (IP), M2, *house starts* (Hous) e taxa de desemprego (Unrate). Os fatores serão estimados de uma base de dados contendo 115 variáveis macroeconômicas, com frequência mensal entre 1963.1 até 2015.12. A amostra de dentro estende-se de 1963.1 até 1972.12. Por sua vez, a amostra de fora, ou amostra de avaliação, inicia-se em 1973.1 e finaliza em 2015.12.

A base de dados foi extraída do FRED e conta com vários grupos de variáveis macroeconômicas que incluem: variáveis de preço, variáveis reais, variáveis de mercado de trabalho, variáveis de consumo, variáveis monetárias, taxa de juros e taxa de câmbio, preços e variáveis financeiras.

Ng and MacCracker (2016) propuseram um algoritmo que permite extrair a base de dados disponível no FRED para ser utilizada para estimação de modelos fatoriais. Neste mesmo trabalho os autores compararam a base de dados do FRED com a base de dados de Stock e Watson (2005,2006) que normalmente era utilizada para estimar modelos fatoriais para a economia americana. Não houve significativa diferença nas estimações dos fatores e nem nas comparações por meio do FAVAR.

Todas as variáveis foram estacionarizadas pelo próprio algoritmo de Ng e McCracker (2016). A estacionariedade é obtida por meio de uma transformação das variáveis que incluem: primeira diferença, segunda diferença, logaritmo, logaritmo da diferença, logaritmo da segunda diferença e $\Delta(x_t/x_{t-1} - 1)$. No apêndice é descrito a base de dados utilizada, bem como as transformações realizadas para alcançar a estacionariedade.

A equação de previsão a ser estimada é da forma da equação (2). As variáveis alvo foram anualizadas para que todas observações mensais possam ser utilizadas para previsão de outros horizontes. As variáveis reais (IP, house starts e Unrate) a anualização mensal foi da seguinte forma:

$$y_{t+h} = (1200/h) \ln(y_{t+h}^r/y_t^r)$$

Em que: y_{t+h}^r é a variável real no tempo $t + h$ e y_t^r é a variável alvo no tempo t

Por sua vez, para as variáveis nominais, como IPI, CPI e M2, a anualização mensal se deu do seguinte modo:

$$y_{t+h} = (1200/h) \ln(y_{t+h}^n/y_t^n) - 1200 \ln(y_{t+1}^n/y_t^n)$$

Em que: y_{t+h}^n é a variável nominal no tempo $t + h$. Por sua vez, para as variáveis nominais, como IPI, CPI e M2, a anualização mensal se deu do seguinte modo:

$$y_{t+h} = (1200/h) \ln(y_{t+h}^n/y_t^n) - 1200 \ln(y_{t+1}^n/y_t^n)$$

Na qual y_{t+h}^n é a variável nominal no tempo $t + h$. Foram realizadas previsões para quatro horizontes de previsão diferentes: 1, 3, 6 e 12 meses. Os modelos estimados consideram quatro combinações para os parâmetros de esquecimento: $\{\alpha, \lambda\} = \{(0.99, 0.99), (0.95, 0.95), (0.99, 0.95), (0.95, 0.99)\}$.

Os modelos DMA usando fatores serão comparados a outros modelos de previsão para as mesmas variáveis alvo. Os modelos alternativos de previsão são: AR(4), Fatores autorregressivos aumentados do tipo Stock e Watson (2002) (FAAR), *Mallows Model*

Average (MMA), *Jackknife Model Average* (JMA), *Leave-h-out Model Average* (LHO) e *equal weight model average* (SMA). Os modelos MMA, JMA e LHO foram aplicados recentemente para a ponderação de fatores por Cheng e Hansen (2015).

Serão apresentados os resultados para o DMA usando fatores e as três extensões propostas: Lar-FDMA, Lasso-FDMA e CFPC-FDMA. Não serão apresentados, por concisão, os resultados de todas as combinações dos parâmetros de esquecimento para as extensões do DMA usando fatores, apenas a melhor combinação será reportada, isto é, dada as quatro possíveis combinações para cada extensão, aquela que gerar a melhor previsão será apresentada. O critério para selecionar a melhor previsão será baseado no erro quadrático médio de previsão (EQMP) para dado horizonte e variável.

As Tabelas de 11-14 apresentam os resultados para a previsão das seis variáveis macroeconômicas consideradas. Duas estatísticas são apresentadas: i. a razão da raiz do erro quadrático de previsão em relação a previsão do modelo AR(4) (REQMP) e ii. a razão da erro erro absoluto de previsão em relação a previsão do modelo AR(4) (RMAEP). As fórmulas das estatísticas para o k –ésimo modelo de previsão no horizonte h :

$$REQMP_{h,k} = \sqrt{\frac{\sum_{t=P+1}^R (\hat{y}_{t+h,k} - y_{t+h})^2}{\sum_{t=P+1}^R (\hat{y}_{t+h,AR(4)} - y_{t+h})^2}}$$

$$RMAEP_{h,k} = \frac{\sum_{t=P+1}^R |\hat{y}_{t+h,k} - y_{t+h}|}{\sum_{t=P+1}^R |\hat{y}_{t+h,AR(4)} - y_{t+h}|}$$

Em que: $\hat{y}_{t+h,k}$ é a previsão usando o modelo k para o horizonte h . É importante notar que essa normalização em relação ao AR(4) permite comparar vários modelos tendo um benchmark como referência e adicionalmente, comparar cada modelo ao próprio benchmark, pois se tais estatísticas forem menores que a unidade, então o modelo tem maior poder para analisar o poder de previsão dos modelos de fatores supervisionados.

Em negrito estão os modelos, dispostos nas colunas, que apresentaram maior poder preditivo para cada variável. Para o horizonte de curto prazo ($h = 1$) é possível as variáveis reais terem poder de previsão semelhantes. Entretanto, para a produção industrial (IP) e para a taxa de desemprego (Unrate) o modelo CFPC-FDMA foi o que apresentou menor erro quadrático médio de previsão em relação ao AR(4). Já para a variável House Starts (Hous), as três combinações de fatores, JMA, MMA e LHO foram as mais acuradas.

A diferença de desempenho entre os modelos dinamicamente combinados DMA e os demais foi maior nas variáveis monetárias. O modelo LAR-FDMA foi o melhor predictor para os índices de preço IPI e CPI e o modelo CFPC-FDMA foi o que previu melhor o M2.

A Tabela 12 apresenta os resultados para um trimestre ($h = 3$). O modelo FDMA com $(\lambda, \alpha) = (0.99, 0.95)$ foi o melhor predictor para as variáveis reais. LAR-FDMA apresentou melhor desempenho para M2 e os índices de preços foram melhor previstos pelo modelo JMA.

Considerando o horizonte de previsão de seis meses ($h = 6$), apresentado na Tabela 13, percebe-se que o modelo FDMA com $(\lambda, \alpha) = (0.95, 0.95)$ foi o melhor predictor para a maioria das variáveis, incluindo produção industrial (IP), House Starts (Hous), M2, taxa de desemprego (Unrate) e índice de preço ao consumidor (CPI). A exceção foi a variável índice de preços ao produtor (IPI) que foi melhor prevista pelo modelo CFPC-FDMA.

Por fim, a Tabela 14 apresenta os resultados para o horizonte de 12 meses à frente ($h = 12$). Os resultados mostram que os melhores modelo de previsão foram o FDMA

com $(\lambda, \alpha) = (0.95, 0.95)$ e a extensão LASSO-FDMA. O primeiro teve maior poder de previsão para as variáveis produção industrial (IP), House Starts (Hous) e índice de preços ao consumidor (CPI). As demais variáveis foram melhor previstas pelo segundo modelo. Assim, como para $h = 3$, nota-se uma significativa diferença entre as estatísticas com ponderação dinâmica em relação aos demais modelos.

Constata-se, portanto, que para maioria das variáveis e para a todos os horizontes os modelos que utilizam ponderação dinâmica de fatores, inclusive a sua versão estendida, apresentaram maior poder de previsão que os demais modelos utilizados para comparação. De fato, considerando todas as variáveis e todos os horizontes os modelos DMA usando fatores foram os melhores previsores em 87,5% das vezes. Este resultado indica uma possível superioridade deste tipo de ponderação sobre as demais.

2.2 Prevendo excesso de retorno do índice S&P 500

Esta seção abordará o poder de previsão do DMA usando fatores para o excesso de retorno do S&P 500 (*equity premium*). A realização deste exercício separado do exercício empírico para as variáveis macroeconômicas deve-se a dois motivos. Primeiro, o *equity premium* possui alguns *benchmarks* específicos, diferente das variáveis macroeconômicas. De fato, como apontado por Rapach e Zhou (2012) poucos modelos conseguem ter maior poder de previsão que a média histórica do excesso de retorno do índice. Goyal e Welch (2008) mostram que a maioria dos modelos que apresentam bom comportamento preditivo do excesso do retorno no período de dentro da amostra, não apresentam boa previsão para fora da amostra do que a média histórica.

Segundo, os métodos de avaliação do equity premium são diferentes dos tradicionais métodos aplicados a outras variáveis econômicas. Isso decorre da observação de que investidores não estão interessados em saber se um modelo prevê melhor o excesso de retorno que outro modelo, mas sim, se determinado modelo gera, dado um nível de aversão ao risco, um retorno financeiro maior. Nesse caso, as previsões são utilizadas como uma ferramenta para avaliar o ganho de utilidade de um investidor que utiliza critérios de média-variância para maximizar seu portfólio³⁰.

Seja R_t o índice do S&P 500 no tempo t e r_f o retorno de um investimento livre de risco. Defina o excesso de retorno como: $r_t = R_t - r_f$. Assim, o excesso de retorno (ou equity premium), y_{t+1} é o retorno do índice entre o tempo t e $t + 1$ acima do retorno livre de risco. Ou seja, $y_{t+1} = (r_{t+1} - r_t)/r_t$.

A equação de previsão é dada por:

$$y_{t+1} = \alpha + \beta x_t + \epsilon_{t+1}$$

No qual: x_t é um conjunto de variáveis no tempo t que explicam o equity premium e ϵ_{t+1} é um termo de erro com média zero. Diferentes escolhas do conjunto de variáveis

³⁰Ver Elliott e Timmermann (2016), Rapach e Zhou (2012), Elliott, Timmermann e Gargano (2013) e outros.

Tabela 11: Horizonte de previsão h=1

	FAAR	MMA	JMA	LHO	SMA	FDMA (0.95,0.95)	FDMA (0.99,0.99)	FDMA (0.95,0.99)	FDMA (0.99,0.95)	LASSO-FDMA	LAR-FDMA	CFPC-FDMA
IP	1.0101	1.0004	1.0004	1.0004	1.0004	1.0591	1.0102	1.0640	1.0050	1.0086	0.9915	0.9842
	1.0088	0.9913	1.0082	1.0082	1.0082	1.0245	1.0213	1.0230	1.0217	1.0123	1.0026	0.9904
Hous	0.9899	0.9890	0.9890	0.9890	0.9890	1.0152	1.0194	1.0146	1.0147	1.0130	1.0154	1.0067
	0.9699	1.0311	0.9888	0.9888	0.9888	1.0247	1.0240	1.0244	1.0219	1.0291	1.0215	1.0109
Unemp	0.9821	0.9753	0.9753	0.9753	0.9753	0.9625	0.9495	0.9495	0.9529	0.9461	0.9430	0.9402
	0.9635	1.0379	0.9879	0.9879	0.9879	0.9549	0.9518	0.9438	0.9546	0.9486	0.9462	0.9413
M2	1.0093	1.0255	0.9777	0.9777	1.0267	0.8101	0.8066	0.8054	0.7986	0.7815	0.7811	0.7806
	1.0113	0.9888	1.0027	1.0027	1.0027	0.7900	0.7928	0.7834	0.7851	0.7691	0.7673	0.7759
IPI	1.0373	0.9586	0.9545	0.9545	0.9626	0.9866	0.9616	0.9737	0.9476	0.9525	0.9342	0.9615
	1.0285	0.9722	0.9511	0.9511	0.9511	0.9504	0.9481	0.9439	0.9360	0.9508	0.9312	0.9490
CPI	1.0336	0.9661	0.9529	0.9529	0.9676	0.8420	0.8469	0.8369	0.8461	0.8329	0.8247	0.8274
	1.0432	0.9586	0.9528	0.9528	0.9528	0.8263	0.8467	0.8197	0.8445	0.8262	0.8139	0.8095

Notas: A tabela 1 apresenta os resultados do i. a razão da raiz do erro quadrático de previsão em relação ao AR(4) (REQMP) e ii. a razão da erro absoluto de previsão em relação ao AR(4) (RMAEP) para o horizonte de previsão de um mês a frente. São considerados seis variáveis macroeconômicas americanas (índice de preços ao produtor (IPI) e ao consumidor (CPI); índice de produção industrial (IP), M2, house starts (Hous) e taxa de desemprego (Unrate)). Os resultados para o DMA usando fatores foram apresentados considerando a variação dos parâmetros de esquecimento $\{\alpha, \lambda\} = \{(0.99, 0.99), (0.95, 0.95), (0.99, 0.95), (0.95, 0.99)\}$. Por sua vez, nas extensões do DMA com fatores, foram apresentados apenas os resultados para o melhor modelo considerando a mesma variação dos fatores de esquecimento.

Tabela 12: Horizonte de previsão h=3

	FAAR	MMA	JMA	LHO	SMA	FDMA (0.95;0.95)	FDMA (0.99;0.99)	FDMA (0.95;0.99)	FDMA (0.99;0.95)	LASSO-FDMA	LAR-FDMA	CFPC-FDMA
IP	1.0648	0.9392	0.9521	0.9521	0.9521	0.8441	0.8441	0.8762	0.1415	0.8397	0.8439	0.8163
	1.0257	0.9749	0.9519	0.9519	0.9519	0.8546	0.8546	0.8879	0.1484	0.8516	0.8590	0.8277
Hous	1.0561	0.9469	0.9590	0.9590	0.9590	0.8935	0.8935	0.9246	0.1782	0.8820	0.8754	0.8650
	1.0407	0.9609	0.9625	0.9625	0.9625	0.9188	0.9188	0.9527	0.1891	0.8992	0.9059	0.8893
Unemp	1.0157	0.9846	0.9613	0.9671	0.9613	0.9539	0.9539	0.9732	0.2526	0.9420	0.9276	0.9273
	1.0077	0.9923	0.9583	0.9670	0.9583	0.9654	0.9654	0.9899	0.2571	0.9475	0.9371	0.9353
M2	0.9950	1.0050	0.9682	0.9973	0.9860	0.9828	0.9828	0.9929	0.9829	0.9647	0.9623	0.9650
	0.9996	1.0004	0.9664	1.0018	0.9906	0.9553	0.9553	0.9664	0.9553	0.9605	0.9635	0.9739
IPI	1.0125	0.9876	0.9784	0.9931	0.9912	1.0180	1.0180	1.0151	1.0180	1.0025	0.9988	0.9951
	1.0019	0.9981	0.9765	0.9954	0.9911	0.9979	0.9979	1.0086	0.9979	0.9796	1.0027	0.9923
CPI	1.0390	0.9625	0.9568	0.9689	0.9654	0.9660	0.9660	0.9669	0.9661	0.9671	0.9693	0.9844
	1.0512	0.9513	0.9510	0.9577	0.9523	0.9274	0.9274	0.9346	0.9275	0.9352	0.9612	0.9696

Notas: A tabela 1 apresenta os resultados do i. a razão da raiz do erro quadrático de previsão em relação ao AR(4) (REQMP) e ii. a razão da erro absoluto de previsão em relação ao AR(4) (RMAEP) para o horizonte de previsão de um mês a frente. São considerados seis variáveis macroeconômicas americanas (índice de preços ao produtor (IPI) e ao consumidor (CPI); índice de produção industrial (IP), M2, house starts (Hous) e taxa de desemprego (Unrate)). Os resultados para o DMA usando fatores foram apresentados considerando a variação dos parâmetros de esquecimento $\{\alpha, \lambda\} = \{(0.99, 0.99), (0.95, 0.95), (0.99, 0.95), (0.95, 0.99)\}$. Por sua vez, nas extensões do DMA com fatores, foram apresentados apenas os resultados para o melhor modelo considerando a mesma variação dos fatores de esquecimento.

Tabela 13: Horizonte de previsão h=6

	FAAR	MMA	JMA	LHO	SMA	FDMA (0.95;0.95)	FDMA (0.99;0.99)	FDMA (0.95;0.99)	FDMA (0.99;0.95)	LASSO-FDMA	LAR-FDMA	CFPC-FDMA
IP	1.2250	0.8163	0.8385	0.8468	0.8383	0.0432	0.8422	0.8632	0.6847	0.7066	0.6993	0.6768
	1.0710	0.9337	0.8634	0.8728	0.8634	0.0432	0.8502	0.8702	0.7172	0.7394	0.7243	0.7051
Hous	1.2635	0.7914	0.8529	0.8377	0.8098	0.0506	0.8350	0.8444	0.6936	0.7017	0.7122	0.6786
	1.2155	0.9337	0.8634	0.8706	0.8328	0.0536	0.8532	0.8558	0.7273	0.7323	0.7387	0.6979
Unemp	1.0832	0.9232	0.9102	0.8844	0.8946	0.0853	0.9186	0.9358	0.8074	0.8232	0.8428	0.8617
	1.0487	0.9337	0.9194	0.9160	0.9206	0.0874	0.9297	0.9348	0.8175	0.8274	0.8522	0.8736
M2	0.9793	1.0211	0.9726	1.0277	1.0159	0.8684	1.0248	0.9722	0.9625	0.9291	0.9639	0.9415
	0.9659	1.0353	0.9804	1.0406	1.0263	0.8525	1.0266	0.9528	0.9466	0.9393	0.9635	0.9477
IPI	1.0210	0.9794	0.9712	0.9864	0.9850	0.9968	0.9872	0.9958	0.9968	0.9782	0.8822	0.8650
	1.0246	0.9337	0.9653	0.9840	0.9801	0.9630	1.0032	0.9747	0.9786	0.9813	0.8978	0.8984
CPI	1.0310	0.9699	0.9476	0.9609	0.9594	0.8767	0.9810	0.9752	0.9772	0.9556	0.9691	0.9669
	1.0387	0.9337	0.9298	0.9306	0.9308	0.8454	0.9680	0.9518	0.9560	0.9502	0.9568	0.9549

Notas: A tabela 1 apresenta os resultados do i. a razão da raiz do erro quadrático de previsão em relação ao AR(4) (REQMP) e ii. a razão da erro absoluto de previsão em relação ao AR(4) (RMAEP) para o horizonte de previsão de um mês a frente. São considerados seis variáveis macroeconômicas americanas (índice de preços ao produtor (IPI) e ao consumidor (CPI); índice de produção industrial (IP), M2, house starts (Hous) e taxa de desemprego (Unrate)). Os resultados para o DMA usando fatores foram apresentados considerando a variação dos parâmetros de esquecimento $\{\alpha, \lambda\} = \{(0.99, 0.99), (0.95, 0.95), (0.99, 0.95), (0.95, 0.99)\}$. Por sua vez, nas extensões do DMA com fatores, foram apresentados apenas os resultados para o melhor modelo considerando a mesma variação dos fatores de esquecimento.

Tabela 14: Horizonte de previsão h=12

	FAAR	MMA	JMA	LHO	SMA	FDMA (0.95;0.95)	FDMA (0.99;0.99)	FDMA (0.95;0.99)	FDMA (0.99;0.95)	LASSO-FDMA	LAR-FDMA	CFPC-FDMA
IP	1.1653	0.8582	0.8572	0.9340	0.9047	0.7179	0.8849	0.8357	0.8387	0.7210	0.7809	0.8139
	1.0426	0.9591	0.8790	1.0482	1.0052	0.7392	0.9019	0.8511	0.8482	0.7434	0.8050	0.8383
Hous	1.2076	0.8281	0.8540	0.9576	0.9320	0.6880	0.8681	0.7982	0.8120	0.7210	0.7538	0.8040
	1.2616	0.7926	0.8085	0.9235	0.8772	0.6902	0.7972	0.8560	0.7934	0.6834	0.7648	0.8051
Unemp	1.1124	0.8990	0.9306	1.0342	1.0076	0.7586	0.9087	0.8539	0.8725	0.7529	0.7788	0.8089
	1.1143	0.8974	0.9106	1.0236	1.0030	0.7503	0.8438	0.9080	0.8695	0.7371	0.7672	0.7909
M2	0.9575	1.0444	0.9892	1.0882	1.0696	0.9768	1.0545	0.9822	1.0363	0.9480	0.9868	0.9713
	0.9485	1.0543	0.9940	1.1011	1.0806	0.9559	0.9629	1.0572	1.0456	0.9396	0.9725	0.9470
IPI	1.0183	0.9820	1.0043	1.0284	1.0236	0.9754	0.9933	0.9683	0.9857	0.9581	0.9674	0.9615
	1.0080	0.9920	1.0036	1.0460	1.0345	0.9679	0.9636	0.9754	0.9744	0.9480	0.9626	0.9471
CPI	1.0383	0.9631	0.9624	0.9975	0.9855	0.9026	1.0065	0.9147	0.9936	0.9040	0.9255	0.9134
	1.0463	0.9557	0.9519	1.0020	0.9848	0.8745	0.8832	0.9813	0.9675	0.8838	0.9001	0.8841

Notas: A tabela 1 apresenta os resultados do i. a razão da raiz do erro quadrático de previsão em relação ao AR(4) (REQMP) e ii. a razão do erro absoluto de previsão em relação ao AR(4) (RMAEP) para o horizonte de previsão de um mês a frente. São considerados seis variáveis macroeconômicas americanas (índice de preços ao produtor (IPI) e ao consumidor (CPI); índice de produção industrial (IP), M2, house starts (Hous) e taxa de desemprego (Unrate)). Os resultados para o DMA usando fatores foram apresentados considerando a variação dos parâmetros de esquecimento $\{\alpha, \lambda\} = \{(0.99, 0.99), (0.95, 0.95), (0.99, 0.95), (0.95, 0.99)\}$. Por sua vez, nas extensões do DMA com fatores, foram apresentados apenas os resultados para o melhor modelo considerando a mesma variação dos fatores de esquecimento.

x_t tem sido testadas para verificar o poder de previsão sobre y_{t+1} . Por exemplo, diversos previsores econômicos individuais foram propostos: razão dividendo-preço (Campbell e Shiller 1988a; Fama e French, 1988; Cochrane 2009 e outros), razão rendimento-preço (Campbell e Shiller 1988b) razão book-to-market (Pontiff e Schall, 1998), taxa de juros nominais (Ang e Beckaert, 2007), spread da taxa de juros (Campbell, 1987), inflação (Nelson, 1976; Campbell e Voulteenahe, 2004), razão de pagamentos de dividendos (Lamont, 1998), entre outros.

Outros tipos de modelos consideram restrições econômicas a equação de previsão buscando melhorar a acurácia na amostra de fora. Campbell e Thompson (2008), por exemplo, recomendaram impor restrições no sinal de $\hat{\beta}_t$ e \hat{y}_{t+1} para evitar retornos negativos, pois os investidores atribuem risco apenas a retornos positivos.

Ferreira e Santa Clara (2011) por sua vez, propõe o método da soma das partes em que decompõe o índice no tempo $t + 1$ em três componentes: GM_{t+1} , GE_{t+1} , DP_{t+1} , em que: $GM_{t+1} = (M_{t+1}/M_t)$, com $M_t = P_t/E_t$, isto é, a razão preço-rendimento. $GE_t = E_{t+1}/E_t$ e E_t rendimentos no tempo t . Por fim, $DP_t = D_t/P_t$ é a razão dividendo-preço. Em forma de log retorno, Ferreira e Santa-Clara (2012) aproximam o excesso de retorno do índice por:

$$y_{t+1} = gm_{t+1} + ge_{t+1} + dp_{t+1}$$

Autores como Rapach et al (2010) exploram algumas propriedades das combinação de previsões indicadas por Timmermann (2006) e Bates e Granger (1969). Estes propuseram modelos de ponderação com igual peso e com peso descontado, em que no último é atribuído maior peso a variáveis que tiveram bom desempenho preditivo.

Por fim, Ludvigson e Ng (2007), Kelly e Pruitt (2015) e Neely et al (2012) utilizam índices de difusão para prever o excesso de retorno. Na abordagem de Ludvigson e Ng (2007) é adotado um modelo fatorial aproximado semelhante a equação (2). Os fatores são estimados por componentes principais e são extraídos de bases de dados contendo variáveis macroeconômicas e financeiras³¹.

A abordagem proposta neste artigo é utilizar os modelos de DMA com fatores (FDMA) para prever o excesso de retorno do S&P 500. O objetivo é explorar algumas das vantagens dos modelos propostos na literatura. Seja pela adoção dos índices de difusão seja pela ponderação dinâmica dos fatores por meio do seu poder preditivo passado.

Além disso, os modelos DMA com fatores possuem a vantagem de ser variáveis nos parâmetros. Dangl e Halling (2012) adotam um modelo com parâmetros variáveis, estimados por métodos bayesianos, e encontram evidências de superior previsibilidade frente a média histórica. Henkel *et al* (2011) apontam que os modelos de parâmetros variáveis possuem maior relevância durante recessões.

4.3. Exercício Empírico.

Será utilizado quatorze variáveis financeiras como previsores. A base de dados corresponde a uma versão atualizada de Goyal e Welch (2008) e se estende de 1960.1 até 2011.12, com frequência mensal. O excesso de retorno é calculado como o log do retorno do S&P 500 menos o log de retorno da índice livre de risco. As variáveis utilizadas encontram-se na Tabela 15.

³¹Rapach e Zhou (2012) fazem uma revisão mais ampla dos métodos utilizados para prever o equity premium.

A amostra de dentro inicia-se em 1960.1 e termina em 1972.12 (109 observações). Por sua vez, a amostra de fora, sobre a qual será feita a avaliação, inicia-se em 1973.1 e termina em 2011.12, perfazendo 469 observações. Como o objetivo é a previsibilidade sobre o excesso de retorno, então a estimação será feita de forma recursiva, isto é, para cada nova previsão no tempo $t + 1$ a $t - \acute{e}sima$ observação será incorporada e a amostra será de 1960.1 até t .

Os fatores serão extraídos da base de dados contendo as quatorze variáveis financeiras. Diferentemente de Rapach e Zhou (2012) que utiliza apenas o primeiro fator estimado, aqui será suposto que o número máximo de fatores é seis e será permitido que o DMA selecione o melhor fator para prever o excesso de retorno³².

Tabela 15: Variáveis Financeiras

Símbolo	Nome da variável	Definição
log_dp	Log da razão do dividendo-preço	log da soma móvel de 12 meses de dividendos pagos menos log do índice S&P 500.
log_dy	Log do dividendo-yield	log da soma móvel de 12 meses de dividendos pagos menos log do preço dos ativos
log_ep	log dos rendimentos-preço	log da soma móvel de 12 meses de dos ganhos sobre S&P 500 menos log do preço dos ativos
log_de	Log da razão do pagamentos de dividendos	log da soma móvel de 12 meses de dividendos pagos menos log da soma móvel de 12 meses de dos ganhos sobre S&P 500
svar	Variância do índice	Soma mensal da variância dos retornos diários do S&P 500
bm	razão do Book-to-market	razão do book-to-market para DJIA
ntis	Expansão líquida de ativos	razão da soma móvel de 12 meses dos ativos líquidos pelo índices de capitalização de mercado no fim do exercício listados no NYSE
T_bill	Treasury bill rate	interest rate sobre 3 meses do T-Bill (mercado secundário)
lty	Juros de longo prazo	Títulos do governo americano de longo prazo
ltr	Retorno de longo prazo	Retorno de longo prazo dos títulos do governo
ts	Estrutura a termo	taxa de juros de longo prazo menos taxa do T-bill
dfy	Default do spread da taxa de juros	Diferença entre corporate bond yields de empresas ranqueadas BAA- e AAA-
dfr	Default spread do retorno	Retorno de longo prazo do corporate bond menos retorno de longo prazo do títulos do governo
inf	Inflação	Inflação calculada a partir dos índices de preços ao consumidor CPI (all urban consumers)

Notas: A tabela 5 descreve as variáveis utilizada no exercício empírico de previsão do excesso de retorno sobre o S&P 500. Toda a base de dados, assim como as rotinas utilizadas, encontram-se disponíveis desde que solicitadas por email aos autores.

Para avaliar a performance de fora da amostra serão utilizadas duas medidas. A primeira será o R^2 de fora da amostra (R_{OS}^2) que mensura a redução proporcional no

³²Como discutido por Ludvigson e Ng (2007) e Kelly e Pruitt (2015) existem outras formas de incorporar os fatores. Por exemplo, poderia incluir todos os fatores estimados depois de selecionado o número máximo de fatores por algum critério de informação, ou extrair fatores de base de dados macroeconômicas como fizeram Ludvigson e Ng (2007). Ou adotar o método do Three-Pass regression filter desenvolvido por Kelly e Pruitt (2015). Ou mesmo adotar a estimação em blocos de variáveis, como defendida por Koop e Korobilis (2011). Entretanto, um estudo mais amplo será deixado para o futuro.

erro quadrático médio de previsão de determinado modelo ($EQMF_i$)³³ em relação a um *benchmark*, no caso, em relação ao EQMF da média histórica ($EQMF_0$). Sua definição original deve-se a Campbell e Thompson (2008).

$$R_{OS,i}^2 = 1 - (EQMF_i/EQMF_0)$$

Perceba que se $R_{OS,i}^2 > 0$ então o modelo i tem maior poder de previsão que a média histórica. Para verificar a hipótese de que: $H_0 : R_{OS}^2 \leq 0$ contra $H_A : R_{OS}^2 > 0$ será calculado o p-valor segundo o teste de Clark e McCracken (2001) e Clark e West (2007) para modelos aninhados³⁴.

Entretanto, a medida do R_{OS}^2 apenas mensura o poder preditivo dos modelos sem considerar sua capacidade de gerar retornos financeiros. Investidores não estão diretamente interessado no poder preditivo dos modelos, mas sim, na possibilidade que aquele modelo gere lucros futuros. Leitch e Tanner (1991) foram os primeiros a observar que não existia uma relação direta entre modelos com maior poder preditivo e modelos economicamente lucrativos.

Assim, é comum em análises de previsão³⁵ do excesso de retorno utilizar métricas baseadas na lucratividade ou na utilidade dos modelos. Nesse caso, a previsão do excesso de retorno serve como um meio para obter resultados mais lucrativos, isto é, a previsão é relevante para a tomada de decisão econômica.

Considere um investidor que busca maximizar sua utilidade por critérios de média-variância. Assuma que ψ seja a aversão ao risco do investidor que aloca recursos entre o índice e um ativo livre de risco baseado na previsão realizada pelo modelo i do excesso de retorno para o tempo $t + 1$. ψ seja a aversão ao risco do investidor que aloca recursos entre o índice e um ativo livre de risco baseado na previsão realizada pelo modelo i do excesso de retorno para o tempo $t + 1$.

$$\alpha_{i,t} = \left(\frac{1}{\psi} \right) \left(\frac{\hat{y}_{i,t+1}}{\hat{\sigma}_{t+1}^2} \right) \quad (39)$$

Em que: $\hat{\sigma}_{t+1}^2$ é a estimativa da variância do índice. Existem diversas formas de estimar a variância. Andersen *et al* (2006) e Christoffersen (2012) apresentam boas revisões sobre o tema. Aqui será adotado o mesmo procedimento de Campbell e Thompson (2008) que estimaram $\hat{\sigma}_{t+1}^2$ usando uma amostra da variância computada numa janela de cinco anos dos retornos históricos. A utilidade do investidor estimada pelo modelo i é dada por:

³³O erro quadrático médio de previsão (EQMF) em sua versão absoluta é definido por:

$$EQMF_i = \frac{1}{P} \sum_{i=1}^P (y_{t+i} - \hat{y}_{t+i|t})^2$$

Em que: P é o tamanho da amostra de fora.

³⁴Clark e West (2007) adaptaram o clássico teste de Diebold e Mariano (1997) para modelos que aninhados. Essa adaptação deveu-se ao fato de que modelos aninhados a distribuição do teste de Diebold e Mariano não é padrão. Essa adaptação foi também utilizada em outros estudos que mensuram o poder de previsão sobre o excesso de retorno como Rapach *et al* (2010), Dangl e Halling (2012) e Neely *et al* (2012).

³⁵Ver Elliott, Timmermann e Gargano (2013), Rapach e Zhou (2012), Campbell e Thompson (2008), Elliott e Timmermann (2016), entre outros.

$$\hat{U}_i = \hat{\mu}_i - 0.5\psi\hat{\sigma}_i^2 \quad (40)$$

Em que: $\hat{\mu}_i$ e $\hat{\sigma}_i^2$ são a média e a variância, respectivamente, do portfólio formado sobre as previsões para $\hat{y}_{i,t+1}$ e $\hat{\sigma}_{t+1}^2$. Por sua vez, a utilidade ao se basear sua alocação de recursos por meio da média histórica é dada por:

$$\hat{U}_0 = \hat{\mu}_0 - 0.5\psi\hat{\sigma}_0^2 \quad (41)$$

Em que, novamente, $\hat{\mu}_0$ e $\hat{\sigma}_0^2$ são a média e a variância, respectivamente, do portfólio formado sobre as previsões para a média histórica e $\hat{\sigma}_{t+1}^2$. A medida que será reportada é o Retorno do Certo Equivalente (CER) que é a diferença entre \hat{U}_i e \hat{U}_0 . Será assumido que $\psi = 5$.

Os modelos para comparação dos resultados do DMA usando fatores, além das regressões individuais, serão: POOL-SMA - modelo de igual ponderação, DMSFE - modelo de ponderação dinâmica baseado em Rapach e Zhou (2010), SOP - modelo soma das partes, desenvolvido por Ferreira e Santa-Clara (2012), KS - modelo *Kitchen Sink*, que utiliza todas as quatorze variáveis em um modelo de regressão linear para prever o excesso de retorno.

A Tabela 16 apresenta os resultados para este exercício. O mesmo exercício foi realizado considerando três cenários distintos: toda amostra, período de recessão e período de expansão. Tais períodos são determinados a partir das classificação do *National Bureau of Economic Reserach* (NBER).

As primeiras três colunas refere-se a toda amostra. Note que considerando toda amostra nenhum modelo foi superior a média histórica, pois todos os R_{OS}^2 foram negativos. Em negrito estão marcados os cinco modelos que tiveram melhor desempenho. O melhor modelo de previsão do excesso de retorno foi o SOP desenvolvido por Ferreira e Santa-Clara (2012). Três dos melhores modelos foram o FDMA usando fatores ou alguma de suas extensões. Por fim, o modelo que usa a razão do *Book-to-market* também deve desempenho satisfatório.

Na primeira coluna de resultados está o retorno do equivalente certo (CER) considerando o caso para toda a amostra. Em negrito estão os cinco melhores modelos novamente. Isto é, aqueles que apresentaram retorno superior a média histórica, caso as previsões fossem utilizadas para selecionar o peso dos investimento entre S&P 500 e o ativo livre de risco. Todos os resultados estão em percentuais.

Os cinco melhores modelos, ordenados em relação ao retorno, foram: POOL-SMA, SOP, Lasso-FDMA (0.99,0.95), CPFC-FDMA (0.99,0.95) e CPFC-FDMA (0.99,0.99). O modelo que apresentou melhor resultado financeiro foi o Lasso-FDMA (0.99,0.95), que produziu retorno 2,34% superior a média histórica.

Nas colunas de 4 a 6 estão os resultados para os períodos de recessão. Novamente, nenhum modelo teve desempenho superior a média histórica em termos de poder de previsão. Os cinco melhores modelos estão em negrito. O modelo SOP foi o melhor modelo em termos de previsão. Os demais modelos dentre os cinco primeiros foram: Lasso-FDMA (0.99,0.95), Lasso-FDMA (0.99,0.99), FDMA (0.99,0.99) e LAR-FDMA (0.99,0.99).

Em relação ao retorno financeiro acima da média histórica, os modelos que apresentaram maior retorno foram: ltr, ts, dfy, dfr e SOP. Nenhum modelo DMA usando fatores apresentou resultado superior em termos de retorno financeiro que a média histórica, com exceção do Lasso-FDMA (0.99,0.99).

Tabela 16: Resultados Excesso de Retorno sobre o S&P 500

	Todo o período			Recessão			Expansão		
	CER	R_{OS}^2	p-valor	CER	R_{OS}^2	p-valor	CER	R_{OS}^2	p-valor
log_dp	-1,505	-23,920	0,271	-5,712	-3019,873	0,864	19,815	-1076,565	0,007
log_dy	-0,854	-29,761	0,229	-5,623	-3727,112	0,854	23,401	-1402,701	0,003
log_ep	-1,403	-2,001	0,451	-4,191	-221,632	0,730	12,581	-154,888	0,334
log_de	0,018	-0,128	0,444	-1,706	-13,317	0,304	8,516	-11,612	0,573
svar	-0,576	-468,660	0,547	-4,068	-43433,352	0,860	17,027	-54057,561	0,339
bm	-0,224	-0,020	0,439	-0,059	-2,384	0,373	-1,109	-1,135	0,526
ntis	-1,077	-40,006	0,606	-5,103	-5294,756	0,896	19,306	-1289,276	0,016
T_bill	-1,180	-36,137	0,493	-5,103	-4192,360	0,786	18,677	-2401,391	0,237
lty	-0,586	-17,082	0,446	-3,987	-1327,652	0,761	16,567	-2505,376	0,325
ltr	-0,107	-29,776	0,359	1,642	-3535,163	0,099	-9,030	-1809,504	0,843
ts	-0,107	-81,832	0,365	1,642	-9871,902	0,111	-9,030	-4645,456	0,839
dfy	-0,107	-601,801	0,386	1,642	-75859,427	0,107	-9,030	-27331,615	0,935
dfr	0,938	-14,220	0,154	0,716	-1739,611	0,121	1,998	-756,469	0,436
inf	0,109	-532,130	0,363	-3,922	-60042,982	0,783	20,512	-38904,012	0,144
POOL-SMA	2,994	-0,061	0,016	1,311	-5,386	0,019	11,320	-7,529	0,162
DMSFE	-0,893	-6,174	0,306	-4,530	-566,342	0,645	17,459	-724,508	0,232
KS	-1,180	-5,407	0,376	-5,103	-647,615	0,830	18,677	-316,739	0,086
SOP	1,466	-0,004	0,323	0,220	-0,746	0,421	7,669	0,416	0,305
FDMA (0.95,0.95)	-0,1707	-0,2006	0,6386	-2,3471	-15,4114	0,5362	10,6747	-29,7876	0,6496
Lasso-FDMA (0.95,0.95)	0,0476	-0,1853	0,5132	-1,1064	-13,5701	0,4504	5,7787	-28,9073	0,5473
LAR-FDMA (0.95,0.95)	-0,7970	-0,1927	0,5963	-2,6073	-17,7303	0,7364	8,2179	-22,4987	0,4035
CFPC-FDMA (0.95,0.95)	1,1167	-0,2590	0,2898	-3,3173	-31,4835	0,8434	23,5898	-14,1962	0,0833
FDMA (0.95,0.99)	-1,2512	-0,1405	0,5211	-3,1341	-11,7225	0,4976	8,1289	-18,9298	0,5284
Lasso-FDMA (0.95,0.99)	-1,4206	-0,1431	0,5283	-2,9228	-11,6851	0,5015	6,0662	-19,8134	0,5338
LAR-FDMA (0.95,0.99)	-1,5024	-0,1417	0,5393	-3,3364	-12,4074	0,5740	7,6478	-17,8567	0,49169
CFPC-FDMA (0.95,0.99)	0,0379	-0,1204	0,3965	-3,2867	-12,7694	0,7478	16,7990	-10,4991	0,2275
FDMA (0.99,0.95)	0,2689	-0,0528	0,5475	-0,6227	-3,6595	0,4516	4,6361	-8,6711	0,6043
Lasso-FDMA (0.99,0.95)	2,3984	-0,0334	0,3920	1,2870	-1,4038	0,1564	7,8907	-7,4110	0,5890
LAR-FDMA (0.99,0.95)	0,7409	-0,0238	0,2548	-0,9718	-3,3473	0,5953	9,2738	-0,3469	0,1640
CFPC-FDMA (0.99,0.95)	2,0379	-0,2168	0,2306	-2,7193	-30,9956	0,8291	26,1951	-2,1778	0,0403
FDMA (0.99,0.99)	0,3490	-0,0340	0,5170	-0,3377	-1,9493	0,3986	3,7062	-6,4483	0,6115
Lasso-FDMA (0.99,0.99)	1,1213	-0,0258	0,5002	0,1913	-1,1091	0,2642	5,7053	-5,6737	0,6377
LAR-FDMA (0.99,0.99)	-0,1660	-0,0199	0,4241	-1,9124	-2,3522	0,6639	8,5339	-1,2481	0,2933
CFPC-FDMA (0.99,0.99)	1,8763	-0,1360	0,1723	-2,6490	-23,1362	0,8493	24,8233	6,3697	0,0168

Notas: A tabela 6 apresenta os resultados dos exercícios empíricos para prever o excesso de retorno do S&P 500. A tabela é dividida em três distintos períodos: Toda a amostra (1973.1 até 2011.12), períodos de recessão e períodos de expansão. Os períodos de expansão e recessão são separados segundo o National Bureau of Economic Research (NBER). Para cada período, a primeira coluna apresenta o resultado do Retorno de Equivalente Certo (CER), considerando $\psi = 5$. A segunda coluna apresenta o R_{OS}^2 e a terceira o p-valor do teste de Clark e West (2007)

Por fim, os resultados do período de expansão encontram-se listados nas colunas de 8 a 10. Com relação ao poder de previsão, o melhor modelo em termos de previsão foi o CFPC-FDMA (0.99; 0.99), com desempenho bastante superior aos demais modelos. Nos períodos de expansão o retorno financeiro dos modelos FDMA foi também bastante o melhor dentre todos os modelos. O modelo CFPC-FDMA(0.99; 0.95) apresentou retorno superior a média histórica em mais de 26%. Outros modelos FDMA também tiveram destaque, como: CFPC-FDMA(0.99; 0.99) e CFPC-FDMA(0.95; 0.99).

Estes resultados indicam que os modelos de DMA usando fatores possuem bom desempenho em termos de poder preditivo, entretanto, sua performance está condicionada a uma amostra relativamente grande (toda a amostra). Em períodos de recessão o FDMA não produziu bons resultados, nem em termos de poder de previsão, nem em termos de retorno financeiro. Entretanto, em períodos de expansão o FDMA apresentou resultados bastante interessantes, sendo o modelo que apresentou o maior desempenho tanto em termos preditivos quanto em termos de retorno.

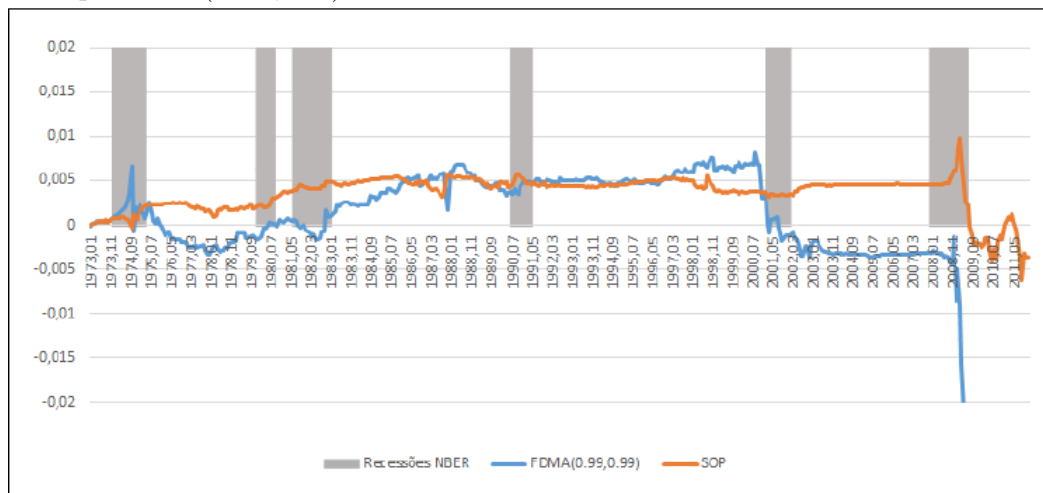
No exercício empírico o modelo SOP foi o que apresentou melhor resultado em termos de poder de previsão para toda a amostra e para o período de recessão. Para comparar o desempenho dos melhores modelos de previsão a Figura 2 apresenta o gráfico da diferença no erro quadrático médio de previsão da média histórica vis-à-vis um modelo de previsão (DEQMF). No caso, é comparado o melhor modelo de previsão do tipo DMA usando fatores, no caso o FDMA (0.99; 0.99), com o modelo SOP. A análise aqui é feita considerando toda a amostra.

$$DEQMF_{i,P} = \sum_{s=1}^R (y_{t+s} - \bar{y}_{t+s})^2 - \sum_{s=1}^R (y_{t+s} - \hat{y}_{i,t+s})^2$$

Em que: $i = \{FDMA(0,99; 0,99), SOP\}$. As marcações em cinza no gráfico representam os períodos de recessão segundo o *National Bureau of Economic Research* (NBER). Ambos os modelos são analisados considerando a amostra completa. Percebe-se que após os períodos de recessão 1973.11 até 1975.03 (1º Choque do Petróleo) e 2001.03 à 2001.11 (Ataque terrorista em New York) o comportamento do FDMA piorou tanto em comparação com o SOP, quanto em relação a média histórica. Todavia, períodos relativamente estáveis, como de 1983.02 até 2001.02 o comportamento do FDMA foi melhor que a média histórica e durante alguns anos superior ao SOP.

Estes resultados confirmam o que foi verificado na Tabela 6. De fato, o FDMA teve um desempenho em média pior que a média histórica e o SOP durante toda a amostra. Porém, durante períodos de expansão a performance do FDMA foi superior em relação aos *benchmarks*.

Figura 4: Diferença no erro quadrático médio de previsão da média histórica vis-à-vis um modelo de previsão (DEQMF)



Nota: A Figura 2 apresenta a comparação, para toda a amostra, da diferença no erro quadrático de previsão da média histórica vis-à-vis um modelo de previsão. Os modelos de previsão são o SOP e o FDMA (0,99; 0,99) que tiveram bom desempenho preditivo em toda a amostra. Em cinza estão marcados os períodos de recessão segundo *National Bureau of Economic Research* (NBER).

Por fim, vale ressaltar que o exercício empírico realizado aqui possui algumas limitações, não permitindo uma conclusão mais precisa do desempenho preditivo do DMA usando fatores. Dentre as principais limitações estão: i. O exercício foi realizado apenas para a previsão do excesso de retorno de um período a frente. Outras possibilidades podem ser prever para dados trimestrais e anuais. ii. O modelo DMA usando fatores estima os fatores sobre uma base de dados que contém apenas 14 variáveis financeiras. Bai e Ng (2008) mostram que o desempenho do métodos de componentes principais não é muito adequado quando N é pequeno, além disso, Ludvigson e Ng (2007) obtém boas estimações de modelos fatoriais ao estimar sobre 147 variáveis financeiras. iii. Não foi incluído fatores relativos a características macroeconômicas. Ludvigson e Ng (2007) mostram que variáveis macroeconômicas tem um papel importante na relação média-variância do S&P 500. Todas essas limitações serão averiguadas em um trabalho futuro. Entretanto, por agora, o DMA usando fatores apresenta-se como um modelo promissor.

5 Conclusões.

Este artigo propôs um novo método de previsão que utiliza uma combinação dinâmica de fatores estimados a partir de uma base de dados grande. Tal método possui a vantagem de permitir uma atualização das estimativas a cada período levando em consideração o desempenho preditivo do período anterior.

Além disso, foi apresentado três extensões do DMA usando fatores em que a priori são selecionados sobre quais variáveis serão extraídos os fatores. As extensões incluem a pré-seleção usando métodos do Lasso e Lar e a supervisão dos fatores por meio do CFPC.

Foi demonstrado que o erro quadrático médio de previsão do DMA usando fatores é assintoticamente equivalente ao erro quadrático médio de um modelo com estrutura fatorial. Este resultado mostra que se a estrutura de uma base de dados grande tiver representação fatorial, então, não haverá diferenças assintóticas ao se estimar o modelo

CHENG and HANSEN, 201 pelo DMA usando fatores.

Dois exercícios empíricos foram realizados visando identificar o poder preditivo do DMA usando fatores. Primeiro, foi realizada a previsão para seis variáveis macroeconômicas dos EUA: índice de produção industrial, M2, índice de preços ao consumidor e ao produtor, taxa de desemprego e house starts. O modelo DMA usando fatores teve um desempenho bastante satisfatório em comparação a modelos que tradicionalmente possuem uma boa performance como: modelo ponderado com pesos iguais, modelo ponderado pelo critério de Mallows e o modelo fatorial aumentado. De fato, as extensões do DMA usando fatores tiveram performance bastante superiores em alguns casos do que o benchmark, sendo este resultado robusto para a variação do horizonte de previsão.

Segundo, foi utilizado o DMA usando fatores para prever o excesso de retorno do S&P 500. Os fatores foram extraídos de uma base de dados contendo 14 variáveis financeiras. Os resultados encontrados, apesar de não ideais, foram bastante satisfatórios. O DMA usando fatores apresentou relativo desempenho em termos de poder de previsão e retorno financeiro quando comparados a outros benchmarks. Entretanto, os resultados forma melhores quando considerados uma amostra maior. Ao se separar a amostra em fases de recessão e de expansão, o DMA usando fatores apresentou desempenho pior.

Assim, conclui-se que o DMA usando fatores e suas extensões apresentam-se como métodos promissores para a realização de previsão seja de variáveis macroeconômicas seja de variáveis financeiras.

REFERÊNCIAS BIBLIOGRÁFICAS.

- ANG, Andrew; BEKAERT, Geert. Stock return predictability: Is it there?. **Review of Financial studies**, v. 20, n. 3, p. 651-707, 2007.
- BAI, Jushan; NG, Serena. Determining the number of primitive shocks in factor models. **Journal of Business & Economic Statistics**, 2012.
- BAI, Jushan; NG, Serena. Forecasting economic time series using targeted predictors. **Journal of Econometrics**, v. 146, n. 2, p. 304-317, 2008.
- BAIR, Eric et al. Prediction by supervised principal components. **Journal of the American Statistical Association**, 2012.
- BATES, John M.; GRANGER, Clive WJ. The combination of forecasts. **Journal of the Operational Research Society**, v. 20, n. 4, p. 451-468, 1969.
- BOIVIN, Jean; NG, Serena. Are more data always better for factor analysis?. **Journal of Econometrics**, v. 132, n. 1, p. 169-194, 2006.
- CALDEIRA, João F.; MOURA, Guilherme V.; SANTOS, André AP. Previsões Macroeconômicas Baseadas em Modelos TVP-VAR: Evidências Para o Brasil. **Revista Brasileira de Economia**, v. 69, n. 4, p. 407-428, 2015.
- CAMPBELL, John Y.; SHILLER, Robert J. Stock prices, earnings, and expected dividends. **The Journal of Finance**, v. 43, n. 3, p. 661-676, 1988.
- CAMPBELL, John Y.; THOMPSON, Samuel B. Predicting excess stock returns out of sample: Can anything beat the historical average?. **Review of Financial Studies**, v. 21, n. 4, p. 1509-1531, 2008.
- CAMPBELL, John Y.; THOMPSON, Samuel B. Predicting excess stock returns out of sample: Can anything beat the historical average?. **Review of Financial Studies**, v. 21, n. 4, p. 1509-1531, 2008.
- CAMPBELL, John Y.; VUOLTEENAHO, Tuomo. Inflation illusion and stock prices. **National bureau of economic research**, 2004.
- CARRASCO, Marine; ROSSI, Barbara. In-sample Inference and Forecasting in Misspecified Factor Models. **Journal of Business & Economic Statistics**, n. just-accepted, p. 1-72, 2016.
- CHENG, X. and HANSEN, B. Forecasting with Factor-Augmented Regression: A frequentist model averaging approach. **Journal of Econometrics**, 186, 280-293, 2015.
- CHENG, X. and HANSEN, B. Forecasting with Factor-Augmented Regression: A frequentist model averaging approach. **Journal of Econometrics**, 186, 280-293, 2015.
- CHRISTOFFERSEN, Peter F. Elements of financial risk management. **Academic Press**, 2012.
- CLARK, Todd E.; MCCracken, Michael W. Tests of equal forecast accuracy and encompassing for nested models. **Journal of econometrics**, v. 105, n. 1, p. 85-110, 2001.

CLARK, Todd E.; WEST, Kenneth D. Approximately normal tests for equal predictive accuracy in nested models. **Journal of econometrics**, v. 138, n. 1, p. 291-311, 2007.

COCHRANE, John H. Asset Pricing:(Revised Edition). **Princeton university press**, 2009.

DANGL, Thomas; HALLING, Michael. Predictive regressions with time-varying coefficients. **Journal of Financial Economics**, v. 106, n. 1, p. 157-181, 2012.

DANGL, Thomas; HALLING, Michael. Predictive regressions with time-varying coefficients. **Journal of Financial Economics**, v. 106, n. 1, p. 157-181, 2012.

EFRON, Bradley et al. Least angle regression. **The Annals of statistics**, v. 32, n. 2, p. 407- 499, 2004.

ELLIOT, G. e TIMMERMANN, A. Economic Forecasting. **Princeton University Press**, New Jersey, 2016.

ELLIOTT, Graham; GARGANO, Antonio; TIMMERMANN, Allan. Complete subset regressions. **Journal of Econometrics**, v. 177, n. 2, p. 357-373, 2013.

ELLIOTT, Graham; GARGANO, Antonio; TIMMERMANN, Allan. Complete subset regressions. **Journal of Econometrics**, v. 177, n. 2, p. 357-373, 2013.

FAGIN, Ronald; HALPERN, Joseph Y.; MEGIDDO, Nimrod. A logic for reasoning about probabilities. **Information and computation**, v. 87, n. 1, p. 78-128, 1990.

FAMA, Eugene F.; FRENCH, Kenneth R. Dividend yields and expected stock returns. **Journal of financial economics**, v. 22, n. 1, p. 3-25, 1988.

FAMA, Eugene F.; SCHWERT, G. William. Asset returns and inflation. **Journal of financial economics**, v. 5, n. 2, p. 115-146, 1977.

FERREIRA, BIERENS, H. e CASTELAR, I. Forecasting Quarterly Brazilian GDP Growth Rate With Linear and NonLinear Diusion Index Models, **Economia**, vol. 6(3), p 261-292, 2005.

FERREIRA, Miguel A.; SANTA-CLARA, Pedro. Forecasting stock market returns: The sum of the parts is more than the whole. **Journal of Financial Economics**, v. 100, n. 3, p. 514-537, 2011.

FIGUEREIDO, F. M. R. Forecasting Brazilian ination using a large data set. **Brazilian Central Bank**, Working Paper Series, 228, December, 2010.

GELPER, Sarah; CROUX, Christophe. Least angle regression for time series forecasting with many predictors. 2008.

GIOVANNELLI, Alessandro; PROIETTI, Tommaso. On the selection of common factors for macroeconomic forecasting. 2015.

HARVEY, David; LEYBOURNE, Stephen; NEWBOLD, Paul. Testing the equality of prediction mean squared errors. **International Journal of forecasting**, v. 13, n. 2, p. 281-291, 1997.

HENKEL, Sam James; MARTIN, J. Spencer; NARDARI, Federico. Time-varying short-horizon predictability. **Journal of Financial Economics**, v. 99, n. 3, p. 560-580,

2011.

HUANG, Huiyu; LEE, Tae-Hwy. Forecasting value-at-risk using high-frequency information. **Econometrics**, v. 1, n. 1, p. 127-140, 2013.

JACQUIER, Eric; POLSON, Nicholas G.; ROSSI, Peter E. Bayesian analysis of stochastic volatility models. **Journal of Business & Economic Statistics**, 2012.

JAZWINSKI, A. H. Mathematics in science and engineering. **Stochastic Processes and Filtering Theory**, v. 64, 1970.

KELLY, Bryan; PRUITT, Seth. The three-pass regression filter: A new approach to forecasting using many predictors. **Journal of Econometrics**, v. 186, n. 2, p. 294-316, 2015.

KIM, H., and SWANSON, N. Forecasting financial and macroeconomic variables using data reduction methods: new empirical evidence. **J. Econometrics** 178, 352367, 2014.

KIM, H., and SWANSON, N. Forecasting financial and macroeconomic variables using data reduction methods: new empirical evidence. **J. Econometrics** 178, 352367, 2014.

KOOP, G. and POTTER, S. Forecasting in dynamic factor models using bayesian model averaging. **Econometrics Journal**, 7:550-565, 2004.

KOOP, Gary; KOROBILIS, Dimitris. Forecasting inflation using dynamic model averaging. **International Economic Review**, v. 53, n. 3, p. 867-886, 2012.

KOOP, Gary; KOROBILIS, Dimitris. UK macroeconomic forecasting with many predictors: Which models forecast best and when do they do so?. **Economic Modelling**, v. 28, n. 5, p. 2307-2318, 2011.

LAMONT, Owen. Earnings and expected returns. **The journal of Finance**, v. 53, n. 5, p. 1563-1587, 1998.

LEE, Tae-Hwy; TU, Yundong; ULLAH, Aman. Nonparametric and semiparametric regressions subject to monotonicity constraints: Estimation and forecasting. **Journal of Econometrics**, v. 182, n. 1, p. 196-210, 2014.

LEHMANN, Bruce N.; MODEST, David M. The empirical foundations of the arbitrage pricing theory. **Journal of Financial Economics**, v. 21, n. 2, p. 213-254, 1988.

LEITCH, Gordon; TANNER, J. Ernest. Economic forecast evaluation: profits versus the conventional error measures. **The American Economic Review**, p. 580-590, 1991.

LUDVIGSON, Sydney C.; NG, Serena. The empirical risk–return relation: A factor analysis approach. **Journal of Financial Economics**, v. 83, n. 1, p. 171-222, 2007.

NEELY, C. J. et al. Forecasting the equity risk premium: The role of technical indicators. **Federal Reserve Bank of St. Louis working paper**, 2012.

ONORANTE, Luca; RAFTERY, Adrian E. Dynamic model averaging in large model spaces using dynamic Occam's window. **European economic review**, v. 81, p. 2-14, 2016.

PONTIFF, Jeffrey; SCHALL, Lawrence D. Book-to-market ratios as predictors of market returns. **Journal of Financial Economics**, v. 49, n. 2, p. 141-160, 1998.

- RAFTERY, Adrian E.; KÁRNÝ, Miroslav; ETTLER, Pavel. Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. **Technometrics**, v. 52, n. 1, p. 52-66, 2010.
- RAPACH, David E. et al. Forecasting stock returns. **Handbook of Economic Forecasting**, v. 2, n. Part A, p. 328-383, 2013.
- SHILLER, Robert J.; BELTRATTI, Andrea E. Stock prices and bond yields: Can their comovements be explained in terms of present value models?. **Journal of Monetary Economics**, v. 30, n. 1, p. 25-46, 1992.
- STOCK, J. H. and WATSON, M. W. Forecasting using principal components from a large number of predictors. **Journal of the American Statistical Association**, 97:1167-1179, 2002.
- STOCK, J. H. and WATSON, M. W. Testing for Common Trends. **Journal of the American Statistical Association**, Vol.83, No.404, December 1988.
- STOCK, J. H. and WATSON, M. W. Generalized shrinkage methods for forecasting using many predictors. **Journal of Business and Economic Statistics**. 30(4):481-493, 2012.
- SVENSSON, Anders et al. The Greenland ice core chronology 2005, 15–42ka. Part 2: comparison to other records. **Quaternary Science Reviews**, v. 25, n. 23, p. 3258-3267, 2006.
- TIMMERMANN, Allan. Forecast combinations. **Handbook of economic forecasting**, v. 1, p. 135-196, 2006.
- WELCH, Ivo; GOYAL, Amit. A comprehensive look at the empirical performance of equity premium prediction. **Review of Financial Studies**, v. 21, n. 4, p. 1455-1508, 2008.
- WELCH, Ivo; GOYAL, Amit. A comprehensive look at the empirical performance of equity premium prediction. **Review of Financial Studies**, v. 21, n. 4, p. 1455-1508, 2008.
- WEST, Kenneth D. Forecast evaluation. **Handbook of economic forecasting**, v. 1, p. 99-134, 2006.

CONCLUSÃO GERAL

O conjunto de metodologias de previsão disponível para o economista aplicado expandiu-se amplamente nas últimas décadas. Atualmente, não há uma abordagem única e a escolha do método de previsão é muitas vezes ditadas pela situação atual, como as necessidades específicas do usuário, a disponibilidade de dados, espaço de previsão (curtíssimo, curto, médio e/ou longo prazo) e da familiaridade do pesquisador com o modelo a ser estimado. Seguindo essa linha, este trabalho ousou prever algumas variáveis macroeconômicas por meio de modelos fatoriais supervisionados e não supervisionados e por meio de combinação dinâmica de modelos fatoriais.

O segundo capítulo comparou o poder de previsão de quatorze modelos para nove variáveis macroeconômicas brasileiras, incluindo taxa de desemprego, índice de produção industrial, taxa real de câmbio real-dólar e seis índices de preços: IPCA, IGP_10, INCC, INPC, IPC e IPA_10. A base de dados contém 117 variáveis macroeconômicas entre 1995.05 a 2015.12 com frequência mensal. Os métodos utilizados foram Modelo de Fator Autoregressivo Aumentado (FAAR), Modelo de Fator Estático (PCR), Bagging, dois tipos de Modelos Bayesianos Ponderado (BMA), *Least Angle Regression* (LAR), *Elastic Net* (EN), *Non-Negative Garrote* (NNG), Modelo Mallows Ponderado (MMA), Modelo Jackknife Ponderado (JMA), Validação Cruzada *Leave-h-Out* (LHO), Modelo de Média Simples (SMA) e Modelo de Média Simples Geral (GSMA). Apenas o modelo AR(4) não inclui os fatores estimados por PCM. E apenas o PCR não inclui variáveis defasadas.

Dois métodos diferentes de estimação foram utilizados: o esquema recursivo e o de janela móvel. Os horizontes de previsão considerados foram um, três, seis e doze meses à frente. As previsões foram avaliadas utilizando a raiz quadrada do erro quadrático médio (REQM) e o *Model Confidence Set* (MCS).

Os resultados indicam que os modelos fatoriais híbridos selecionados pelo LAR e EN foram os que apresentaram maior poder preditivo, considerando todas as variáveis, todos os horizontes de previsão e ambos os esquemas, recursivo e de janela móvel. Além disso, o esquema de janela móvel apresentou menor valor absoluto do REQM em quase o dobro de vezes que o esquema recursivo, indicando que este esquema possui melhor poder de previsão para a maioria das variáveis em quase todos horizontes. Exceção foi a variável taxa de câmbio que foi melhor prevista pelo esquema recursivo.

Quanto ao terceiro capítulo, a fim de analisar o poder de previsão dos modelos fatoriais supervisionados, comparou-se o poder de previsão desses modelos frente aos não supervisionados para quatro variáveis, índice de produção industrial, índice de preços: IPCA e INPC; e para taxa de juros de longo prazo, tanto no esquema recursivo como no de janela móvel.

Os métodos de previsão utilizados foram: FAAR, BMA, MMA, JMA, LHO e SMA, tanto para modelos de fatores supervisionados como para não supervisionados, Para esti-

mar os fatores supervisionados foi utilizada a técnica proposta Tu e Lee (2016), (CFPC), e para estimar os fatores não supervisionados foi utilizado o método dos componentes principais (PC).

A base de dados utilizada contém 236 observações de periodicidade mensal de 1996.5 à 2015.12, . Os métodos recursivos e de janela móvel foram utilizados no processo de estimação. Além disso, a previsão foi realizada para um, três, seis e doze períodos à frente e avaliadas por meio da raiz quadrada do erro quadrático médio (REQM).

Os resultados mostram que o esquema de janela móvel foi mais eficaz, além disso, o modelo de previsão que obteve melhor performance foi o modelo ponderado BMA. Por fim, os modelos supervisionados são mais eficazes para realizar previsões no médio prazo, pois previu com mais acurácia três, dentre as quatro, variáveis alvos previstas, apresentando um menor (REQM).

Por fim, o quarto capítulo propôs um novo método de previsão que utiliza uma combinação dinâmica de fatores estimados a partir de uma base de dados grande, mas também,apresentou três extensões do DMA usando fatores em que a priori são selecionados sobre quais variáveis serão extraídos os fatores. As extensões incluem a pré-seleção usando métodos do Lasso e Lar e a supervisão dos fatores por meio do CFPC.

Foi revelado que o erro quadrático médio de previsão do DMA usando fatores é assintoticamente equivalente ao erro quadrático médio de um modelo com estrutura fatorial. A fim de verificar a capacidade de previsão do DMA usando fatores, foi realizado dois exercícios empíricos. Primeiro, foi realizada a previsão para seis variáveis macroeconômicas dos EUA, em que o modelo DMA usando fatores teve um desempenho bastante satisfatório em comparação a modelos que tradicionalmente possuem uma boa performance.

Segundo, foi utilizado o DMA usando fatores para prever o excesso de retorno do S&P 500. Os fatores foram extraídos de uma base de dados contendo 14 variáveis financeiras. Os resultados encontrados, apesar de não ideais, foram bastante satisfatórios. O DMA usando fatores apresentou relativo desempenho em termos de poder de previsão e retorno financeiro quando comparados a outros benchmarks. Entretanto, os resultados forma melhores quando considerados uma amostra maior. Ao se separar a amostra em fases de recessão e de expansão, o DMA usando fatores apresentou desempenho pior.

Assim, conclui-se que o DMA usando fatores e suas extensões apresentam-se como métodos promissores para a realização de previsão seja de variáveis macroeconômicas seja de variáveis financeiras.