

Universidade de São Paulo
Faculdade de Economia, Administração e Contabilidade de
Ribeirão Preto
Departamento de Economia
Programa de Pós-graduação em Economia
Área: Economia Aplicada

Estudo da suficiência dos fatores no apreçamento de ativos

Rafaela Dezidério dos Santos Rocha

Orientador: Prof. Dr. Márcio Poletti Laurini

Ribeirão Preto

2021

Prof. Dr. Vahan Agopyan

Reitor da Universidade de São Paulo

Prof. Dr. André Lucirton Costa

Diretor da Faculdade de Economia, Administração e Contabilidade de
Ribeirão Preto

Prof. Dr. Sérgio Kannebley Júnior

Chefe do Departamento de Economia

Prof. Dr. Luciano Nakabashi

Coordenador do Programa de Pós-Graduação em Economia Aplicada

Rafaela Dezidério dos Santos Rocha

Estudo da suficiência dos fatores no apreçamento de ativos

Dissertação apresentada ao Programa de Pós-Graduação em Economia – Área de Concentração: Economia Aplicada da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo para obtenção do título de Mestra em Ciências. Versão Corrigida. A original encontra-se disponível na FEA-RP/USP

Universidade de São Paulo - USP

Orientador: Prof. Dr. Márcio Poletti Laurini

Ribeirão Preto

2021

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Rocha, Rafaela Dezidério dos Santos
Estudo da suficiência dos fatores no apreçamento de ativos – Ribeirão Preto, 2021
47 p. : il.; 30 cm.
Orientador: Prof. Dr. Márcio Poletti Laurini

Dissertação de Mestrado – Universidade of São Paulo – Faculdade de Economia, Administração e Contabilidade – FEA-RP/USP – Campus Ribeirão Preto; Departamento de Economia
Programa de Pós-Graduação em Economia; Área de Concentração: Economia Aplicada, 2021.

1. Modelos Multifatoriais. 2. Fatores Omitidos. 3. Estimador Mean Group 4. Estimador Common Correlated Effects. 5. Prêmio de Risco.
I. Orientador: Prof. Dr. Márcio Poletti Laurini.
II. Universidade of São Paulo – USP – Campus Ribeirão Preto. III. Faculdade de Economia, Administração e Contabilidade. IV. Estudo da Suficiência dos Fatores no Apreçamento de Ativos.

Rafaela Dezidério dos Santos Rocha

Estudo da suficiência dos fatores no apreçamento de ativos

Dissertação apresentada ao Programa de Pós-Graduação em Economia – Área de Concentração: Economia Aplicada da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo para obtenção do título de Mestra em Ciências. Versão Corrigida. A original encontra-se disponível na FEA-RP/USP

Ribeirão Preto

2021

Agradecimentos

Primeiramente, gostaria de agradecer a Deus por ter me dado sabedoria, saúde e força.

Aos meus pais, Cristiane e Tertuliano, por depositarem em mim confiança, apoio, incentivo, e por todos os sacrifícios que fizeram para me dar oportunidade e condições para percorrer esse caminho. Essa conquista não seria possível sem eles e, portanto, posso dizer que conquistamos isso juntos.

Ao meu orientador Márcio Laurini, que me aceitou, ensinou e incentivou por todo o mestrado. Seus conhecimentos transmitidos por todas as disciplinas foi crucial para minha dissertação e seu apoio, principalmente nessa reta final que foi extremamente importante para a continuação do trabalho.

A Universidade de São Paulo por ter me proporcionado todo conhecimento e aprendizado desde minha graduação. Estudar na USP foi meu sonho desde os oito anos de idade. Não imaginava que chegaria tão longe dentro desta universidade, e esta realização não tem preço.

Ao Professor Rodrigo De Losso por ter me orientado na graduação e me incentivado a vir para o programa de mestrado em economia aplicada na FEARP-USP. Sem o seu apoio, orientação e dedicação não teria conseguido chegar até aqui.

A todos os professores do PPGE/FEA-RP/USP, em especial os professores: Sergio Sakurai por toda ajuda, orientação que me deu antes mesmo de eu entrar no programa de mestrado; os professores Jefferson Bertolai e Sérgio Kannebley por todo conhecimento transmitido nas matérias, e também por terem aceitado participar da minha qualificação e pré-defesa e pelas contribuições importantíssimas para o desenvolvimento da dissertação; e a professora Roseli Silva, por todo conhecimento, apoio, incentivo e compreensão que teve comigo nesse último semestre que foi especialmente difícil.

Aos meus colegas, que me ajudaram e ensinaram muito nesses dois anos.

A minha amiga de infância, Bruna Costa, que não me deixou desistir nos momentos mais difíceis e solitários. Especialmente no começo do mestrado.

À CAPES pelo apoio financeiro concedido ao longo do mestrado. O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Resumo

ROCHA, R. D. S. **Estudo da suficiência dos fatores no apreçamento de ativos.** Dissertação (Mestrado) - Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto 2021.

O modelo de cinco fatores de apreçamento de ativos derivado da abordagem de Fama-French é utilizado extensivamente em procedimentos de estimação de prêmio de risco de ativos. Mesmo incluindo um número considerável de fatores, ainda é possível que fatores omitidos afetem a estimação deste modelo. Neste trabalho utilizamos abordagens econométricas alternativas robustas à presença de fatores omitidos na estimação do prêmio de risco no mercado brasileiro. Inicialmente, aplicamos os estimadores de dados em painel de Mean Group e Common Correlated Effects para detectar a presença de fatores omitidos. Então comparamos os resultados com os obtidos pelo estimador proposto por Giglio e Xiu (2021), que usa uma abordagem de componentes principais para fatores latentes. Com esse trabalho concluímos que há evidência de fatores omitidos e o melhor estimador é o estimador Common Correlated Effects.

Palavras-chave: Modelos Multifatoriais; Fatores Omitidos; Estimador Mean Group; Estimador Common Correlated Effects; Prêmio de Risco.

Abstract

ROCHA, R. D. S. **Study of the sufficiency of factors in asset pricing.** Dissertation (Master Degree) - School of Economics, Business and Accounting at Ribeirão Preto, University of São Paulo, Ribeirão Preto, 2021.

The five-factors asset pricing model derived from the Fama-French approach is used extensively in asset risk premium estimation procedures. Even including a considerable number of factors, it is still possible that omitted factors affect the estimation of this model. In this work, we use robust alternative econometric approaches to the presence of omitted factors in the estimation of the risk premium in the Brazilian market. We initially apply the estimators of data in panel of Mean Group and Common Correlated Effects to detect the presence of omitted factors and we compare the results with the obtained ones for the estimator proposed by Giglio and Xiu (2021), Who uses an approach of principal component to estimate risk premium of the observed factors corrected to latent factors. With this work we conclude that there is evidence of omitted factors and the best estimator is the Common Correlated Effects estimator.

Key-words: Multifactorial Models; Omitted Factors; Mean Group Estimator, Common Correlated Effects Estimator, Risk Premium.

Lista de ilustrações

Figura 1 – Gráfico dos Retornos do Fator de Mercado	27
Figura 2 – Gráfico dos Retornos do Fator de Tamanho	28
Figura 3 – Gráfico dos Retornos do Fator de Book-to-Market	28
Figura 4 – Gráfico dos Retornos do Fator relacionado aos Retornos Passados	29
Figura 5 – Gráfico dos Retornos do Fator de Iliquidez	29
Figura 6 – Gráfico dos Retornos das Carteiras Classificadas pelo Tamanho	30
Figura 7 – Gráfico dos Retornos das Carteiras Classificadas pelo Book-to-Market	30
Figura 8 – Gráfico dos Retornos das Carteiras Classificadas pelo Momentum	31
Figura 9 – Gráfico dos Retornos das Carteiras Classificadas pela Iliquidez	31
Figura 10 – Doze primeiros autovalores obtidos pelo PCA	36
Figura 11 – Comparação dos Prêmios de riscos para carteira classificadas pelo book-to-market	37
Figura 12 – Comparação dos Prêmios de riscos para carteira classificadas pela Iliquidez	38
Figura 13 – Comparação dos Prêmios de riscos para carteira classificadas pelo Momentum	38
Figura 14 – Comparação dos Prêmios de riscos para carteira classificadas pelo Tamanho	39
Figura 15 – Gráfico dos Resíduos das Carteiras Classificadas pelo Book-to-Market	40
Figura 16 – Gráfico dos Resíduos das Carteiras Classificadas pela Iliquidez	40
Figura 17 – Gráfico dos Resíduos das Carteiras Classificadas pelo Momentum	41
Figura 18 – Gráfico dos Resíduos das Carteiras Classificadas pelo Tamanho	42
Figura 19 – Gráfico dos Prêmios de Risco Acumulado dos Fatores	43

Lista de tabelas

Tabela 1 – Estimador MG do modelo (3.1)	32
Tabela 2 – Estimador CCE do modelo (3.1)	32
Tabela 3 – Testes de Wald	33
Tabela 4 – Média e EQM dos estimadores para o número de fatores $f = 1$	34
Tabela 5 – Média dos estimadores para o número de fatores $f = 3$	35
Tabela 6 – Média dos estimadores para o número de fatores $f = 5$	35
Tabela 7 – Resíduos para Carteiras BM do NEFIN	40
Tabela 8 – Resíduos para Carteiras ILLIQ do NEFIN	41
Tabela 9 – Resíduos para Carteiras MOM do NEFIN	42
Tabela 10 – Resíduos para Carteiras SIZE do NEFIN	42

Sumário

1	INTRODUÇÃO	11
2	REFERENCIAL BIBLIOGRÁFICO	13
2.1	O estimador consistente para o prêmio de risco na presença de especificação incorreta proposto por Giglio e Xiu	16
3	METODOLOGIA	20
3.1	Suficiência dos Fatores	21
3.1.1	Estimador Mean Group	21
3.1.2	Estimador de Efeitos Comuns Correlacionados	21
3.1.3	Teste de Wald	22
3.2	Método de Giglio e Xiu	23
3.3	Previsões	26
4	BASE DE DADOS	27
5	RESULTADOS	32
5.1	Teste de Wald para os estimadores MG e CCE	32
5.2	Resultado da Simulação do estimador \hat{p}	33
5.3	Estimando modelo Giglio e Xiu para as carteiras do Nefin	36
5.4	Comparação dos três modelos	37
5.5	Fatores Corrigidos	43
6	CONCLUSÃO	44
	BIBLIOGRAFIA	46

1

Em 1993 Fama e French propuseram o modelo de apreçamento de três fatores [10]. Nesse artigo eles identificaram três fatores de risco sistemático nos retornos de ações. A partir desses fatores eles construíram carteiras de forma empírica para reproduzir esses fatores: mercado, tamanho e book-to-market. Porém, evidências dos estudos de Novy-Marx [22] e Titman, Wei, e Xie [28], mostram que o modelo de três fatores é insuficiente, pois perdem grande parte da variação nos retornos médios relacionados à lucratividade e ao investimento. Pensando nisso, em 2015, Fama e French adicionam ao modelo de três fatores os fatores de lucratividade e investimento [12], o então Modelo de Cinco Fatores. Eles concluem que o modelo de Cinco Fatores explicam de 71% a 94% da variação das cross-section dos retornos esperados para o tamanho, book-to-market, lucratividade e investimento.

Em 2017, Fama e French testaram o modelo de cinco fatores para 4 regiões [13] - América do Norte, Europa, Japão e Pacífico Asiático - e apesar do modelo global não ter resultado satisfatório, os modelos locais, isto é, construídos com os dados locais de cada região, explicaram a maior parte da variância dos retornos, como esperado. Existe uma relevante literatura sobre novos fatores para apreçamento de ativos, e o grande número de possíveis fatores ficou conhecido como factor zoo, indicando que a estimação dos modelos tradicionais de três, quatro e cinco fatores derivada da abordagem de Fama-French pode estar sujeita aos problemas causados pela ausência de fatores omitidos na estimação, levando a estimadores inconsistentes do prêmio de risco dos fatores incluídos no modelo. Um dos primeiros a mencionar o factor zoo foi Cochrane em 2011 [6]. Harvey, Liu e Campbell em 2015 [15], McLean and Pontiff em 2016 [20] e mais recentemente Hou, Xue, and Zhang em 2017 [16] falam sobre o factor zoo e como esses fatores podem influenciar nos modelos.

Reconhecendo a importância dos fatores omitidos na estimação de prêmio de risco, Giglio e Xiu [14] propõem um método em três etapas para estimar o prêmio de risco de um fator observável que é válido mesmo na presença de fatores no modelo não especificados ou observados, além de também controlar para um possível erro de medida nos fatores observáveis e também detectar se um tal fator é espúrio ou "inútil", no sentido de que ele não influencia a estimação.

Com isso em mente, estudamos duas formas de utilizar o modelo de cinco fatores para precificar ações com dados brasileiros, assumindo que a relevância dos fatores é constante no tempo. Primeiro focamos o nosso estudo para possível omissão de variável, e assim a existência de um possível viés na estimação. Para isso escolhemos dois estimadores para dados em painel para estimar o prêmio de risco dos fatores: o Mean Group (MG) [23]

e o estimador de efeitos comuns correlacionados (CCE) [24]. O estimador MG é definido como uma média de estimadores OLS, enquanto o estimador CCE é uma extensão do estimador MG que assume uma estrutura de fatores comuns não observados para os erros. A nossa nosso objetivo era que, ao estimá-los, se houver fatores omitidos no modelo de cinco fatores de Fama French, o estimador CCE deveria capturá-los, e portanto, seus coeficientes seriam diferentes dos coeficientes estimados para o estimador MG. A segunda forma utilizada para o estudo da relevância dos cinco fatores foi testar se eles são suficientes no sentido de apreçar corretamente os ativos, ou seja, se esses fatores conseguem estimar um preço aproximadamente correto para o conjunto de ativos em questão. Para isso, estimamos o prêmio de risco com o método do Giglio e Xiu [14], que teoricamente corrige a estimação para possível omissão de variáveis e presença de erro de medida, e usamos essa estimativa para prever os retornos. Por fim, comparamos qual modelo ajusta melhor os retornos observados, o modelo de cinco fatores de Fama-French ou a previsão feita com o prêmio de risco obtido pelo método do Giglio e Xiu.

Na realização da primeira etapa do nosso estudos notamos que existe um forte indício de que os coeficientes do estimador CCE é diferente dos coeficientes do estimador MG e, portanto, há a possibilidade de omissão de variável. Já na segunda etapa observamos que, apesar do estimador proposto por Giglio e Xiu supostamente corrigir a estimação para omissão de variáveis, o estimador CCE apresenta os melhores resultados.

Esse trabalho tem a seguinte estrutura: o referencial bibliográfico será apresentado na Seção 2; a metodologia será apresentada na Seção 3 dividida em 3 subseções: a primeira subseção será sobre a suficiência dos Fatores, descrevendo os estimadores Mean Group e CCE e o teste de Wald feito para compará-los. A segunda subseção descreve o método do Giglio e Xiu [14] para obter o estimador dos prêmios de risco dos nossos 5 fatores. Na terceira subseção apresentamos como faremos nossas estimações com as estimativas obtidas nas subseções anteriores. Na Seção 4 apresentaremos os dados utilizados. Na Seção 5 são apresentados os principais resultados obtidos. Conclusões finais são apresentadas na Seção 6.

2 Referencial Bibliográfico

O estudo de apreçamento de ativos tem sido estudado há muito tempo, e nesses últimos anos diversos fatores e modelos de fatores surgiram. A ideia desses fatores é capturar características que, em teoria, afetam e influenciam o preço dos ativos, como por exemplo tamanho (ME - valor de mercado), Book Equity (BE - valor contábil do patrimônio líquido), Book-to-Market-Equity (a razão BE/ME), Alavancagem (Leverage), Earning/Price (E/P), Dividend/Price (D/P), entre outros. E a muitos dos modelos sugeridos são baseados na combinação desses diversos fatores, de forma que sua combinação precifique melhor os ativos.

Alguns dos principais trabalhos de apreçamento de ativos, tais como os de Sharpe, Lintner, Black e muitos outros, se basearam nos trabalhos de Markowitz. Em 1952, Markowitz [18] estudou o problema de seleção de carteiras. Segundo ele, o processo de seleção de uma carteira ótima pode ser dividido em dois estágios. O primeiro estágio começa com a observação e experiência, terminando com suposições sobre o desempenho futuro dos ativos disponíveis. Já o segundo estágio começa com as suposições relevantes dos desempenhos futuros e termina com a escolha da carteira. Markowitz focou seu estudo no segundo estágio, baseando-se na regra Média-Variância para construir o conjunto de combinações eficientes de média-variância. Em 1959 Markowitz [19] expande seu estudo seguindo Von-Neumann e Morgenstern [21], desenvolvendo uma análise baseada na maximização da utilidade esperada e propõe uma solução para o problema de seleção de carteira.

Baseado nos estudos de Markowitz, Sharpe (1964) [27] propõe uma teoria de equilíbrio de mercado dos preços dos ativos sob condições de risco. Ele conclui que em equilíbrio existe uma relação linear simples entre os retornos esperados e o desvio padrão dos retornos para combinações eficientes de ativos arriscados. Além disso, também existe uma relação consistente entre os retornos esperados e o seus "riscos sistemático" (que pode ser medido pelo β de mercado, que é uma medida da volatilidade da ação em relação ao mercado).

Assim como Sharpe, Lintner (1965) [17] e Black (1972) [4] também estudaram a relação entre retorno médio e risco, tendo como previsão central a eficiência em termos de variação média da carteira (a eficiência proposta por Markowitz). E assim como o modelo de Sharpe, os modelos de Lintner e de Black concluem que retornos esperados são funções lineares positivas dos seus β s de mercado. Também concluíram que os β s de mercado absorvem o efeito da alavancagem sob os preços e são suficientes para descrever a cross-section dos retornos esperados.

Em 1988, Fama e French estudaram a relação dos Dividends Yields e os retornos esperados das ações [8]. Com as evidências coletadas neste trabalho, eles concluíram

que Dividend Yields explica menos do que 5% das variâncias dos retornos mensais ou trimestrais, mas, em contrapartida, geralmente explica mais do que 25% da variância de retornos de dois a quatro anos. Ou seja, Dividend Yields influenciam mais os retornos de longo prazo.

Em 1992, Fama e French publicaram outro artigo sobre retornos esperados das ações [9]. Neste artigo eles avaliam a relação entre retornos esperados e os β s de mercado (a inclinação na regressão do retorno de um título no retorno do mercado), tamanho (ME - valor de mercado medido pelo preço da ação vezes as ações em circulação), alavancagem, Book-to-Market equity (BE/ME) e Earning/Price (E/P). Em contradição com as previsões de Sharpe, Lintner e Black, neste trabalho eles não encontraram nenhuma relação confiável entre os β s de mercado e os retornos esperados. Além disso, também concluíram que alavancagem é bem capturado pelo Book-to-Market equity e a combinação de tamanho (ME) com Book-to-Market equity absorve bem a relação entre E/P e retornos esperados.

No seu artigo de "Common risk factors in the returns on stocks and bonds" de 1993 [10], Fama e French estendem o estudo feito no artigo anterior. Eles usam a abordagem de regressão de séries temporais de Black, Jensen, and Scholes (1972) [5], e por isso, construíram dois fatores de risco relacionados ao tamanho e BE/ME para as ações e dois fatores de estrutura a termo para títulos. Os fatores relacionados ao tamanho e ao BE/ME são, respectivamente, SMB e HML. Para construí-los, eles classificaram as ações por tamanho (*Big e Small*) e BE/ME (*low, medium e high*). Essa classificação por BE/ME é feita pela divisão da população das ações em 3 grupos, baseado-se no índice BE/ME da seguinte forma: 30% inferiores (*low*), 40% intermediário (*medium*) e 30% superior (*high*). A partir dessa classificação cria-se 6 carteiras baseadas nas interseções entre as classificações de tamanho e BE/ME: *Small/low (S/l)*, *Small/medium (S/m)*, *Small/high (S/h)*, *Big/low (B/l)*, *Big/medium (B/m)* e *Big/high (B/h)*. Com essas seis carteiras temos os seguintes retornos na carteira de tamanho grande (B) e pequeno (S):

$$R_B = \frac{1}{3}(R_{B/l} + R_{B/m} + R_{B/h})$$

$$R_S = \frac{1}{3}(R_{S/l} + R_{S/m} + R_{S/h})$$

A partir desses dois retornos mostrados acima são construídos os retornos dos fatores de investimento líquido zero SMB (*small minus big*, ou seja, posição comprada em ações de baixa capitalização e vendida em ações de alta capitalização):

$$R_{SMB} = R_S - R_B$$

De forma análoga os retornos das carteiras alto (H) e baixo (L) são:

$$R_H = \frac{1}{2}(R_{S/h} + R_{B/h})$$

$$R_L = \frac{1}{2}(R_{S/l} + R_{B/l})$$

A partir desses dois retornos acima é criado o fator de investimento líquido zero HML (*high minus low*, ou seja, posição comprada em BE/ME alto e vendida em BE/ME baixo):

$$R_{HML} = R_H - R_L.$$

Já para títulos, construíram duas carteiras para medir o risco comum relacionado à mudança inesperada na taxa de juros. As carteiras construídas são TERM (a diferença entre os retornos dos títulos do governo de longo prazo e a taxa mensal das letras do tesouro medidas no final do mês anterior) e DEF (a diferença entre os retornos na carteira do mercado de títulos corporativos de longo prazo e os retornos dos títulos de longo prazo do governo). As evidências deste trabalho mostraram que esses cinco fatores explicam bem a variação comum nos retornos de títulos e de ações.

Fama e French (1995) [11] também tentaram encontrar fundamentos econômicos para os resultados empíricos dos seus papers anteriores, e se a precificação é racional. Duas hipóteses guiam esse trabalho: 1- deve haver fatores de risco comuns nos retornos associados ao tamanho e BE/ME; 2- o tamanho e os padrões BE/ME nos retornos devem ser explicados pelo comportamento dos lucros (*earnings*). Porém eles não conseguem encontrar evidências de que os retornos respondam ao fator BE/ME nos lucros. Com isso eles concluem que esse artigo deixou importantes questões em aberto, as quais são: Quais são as variáveis de estado econômico subjacentes que produzem variação nos ganhos e retornos relacionados ao tamanho e BE/ME? Essas variáveis de estado sem nome produzem variação no consumo e na riqueza que não é capturada por um fator geral de mercado e, portanto, podem explicar os prêmios de risco nos retornos associados ao tamanho e BE/ME?

Após a publicação do modelo de três fatores, Titman, Wei, e Xie se basearam nesse modelo para estudar os retornos de ações e investimento capital em 2008. E em 2013, Novy-Marx se baseia nos estudos de Fama e French para estudar lucratividade. Ambos os estudos mostram que o modelo de três fatores é insuficiente pois perdem grande parte da variação no retornos médios relacionados à lucratividade e ao investimento.

Após referente sobre a insuficiência do modelo de três fatores em capturar a as retornos relacionadas à lucratividade e ao investimento, Fama e French publicam um novo artigo em 2015 com o modelo de apreçamento de ativos de cinco fatores [12]. Neste novo modelo eles acrescentam ao modelo de três fatores os fatores de rentabilidade e investimento. As carteiras que representam a rentabilidade e o investimento são, respectivamente, RMW (a diferença entre os retornos das carteiras diversificadas com rentabilidade robusta e fraca) e CMA (a diferença entre os retornos das carteiras diversificadas de ações de empresas de alto e baixo investimento, que chamamos de conservadoras e agressivas). Este modelo fornece uma descrição melhor dos retornos médios do que o modelo de três fatores.

Entretanto, esses métodos podem ter problemas causados por possível omissão de

variável e presença de erros de medida. Tais métodos podem causar um viés na estimativa e, portanto, os estimadores não seriam consistentes e possivelmente causariam uma piora na previsão do preço correto do ativo. Atualmente existe uma relevante literatura sobre novos fatores para apreçamento de ativos, e o grande número de possíveis fatores, que é conhecido como Factor Zoo. Um dos primeiros a mencionar o factor zoo foi Cochrane em 2011 [6]. Após essa menção diversos artigos foram publicados sobre o factor zoo e como eles podem influenciar nos modelos foram realizados, como por exemplo o artigo de Harvey, Liu e Campbell [15], o artigo de McLean an Pontiff [20], e o artigo de Hou Xue e Zhang [16]. Pensando nestes problemas Giglio e Xiu [14] propõe uma solução geral com uma metodologia dividida em três etapas. Essa metodologia se baseia em um resultado de invariância de rotação que vale para o prêmio de risco em modelos de fatores lineares, e ao combinar esse resultado com a Análise de Componentes Principais (Principal Component Analysis - PCA), fornece estimativas consistentes de prêmio de risco para qualquer fator observado.

2.1 O estimador consistente para o prêmio de risco na presença de especificação incorreta proposto por Giglio e Xiu

Descrevemos os elementos fundamentais do estimador em três etapas proposto por Giglio e Xiu [14] nesta seção. Para realizar a primeira etapa, é necessário um estimador consistente do número dos fatores. O estimador utilizado por eles possui a mesma ideia dos estimadores de fatores propostos por Bai e Ng [2] e Bai [3].

Bai e Ng demonstram que a penalidade para o overfitting deve ser função tanto de N , a dimensão das cross-section, quanto de T , a dimensão de tempo, para estimar consistentemente o número de fatores. De forma que, os usuais AIC e BIC, não funcionam bem quando ambas dimensões são grandes. Então, considerando o modelo

$$\underset{(T \times N)}{R} = \underset{(T \times p)}{v} \underset{(p \times N)}{\beta'} + \underset{(T \times N)}{e}$$

em que $R = (\underline{R}_1, \dots, \underline{R}_N)$, $\underline{R}_i = (R_{i1}, \dots, R_{iT})$ para $i = 1, \dots, N$, $v = (v_1, \dots, v_T)$, $e = (\underline{e}_1, \dots, \underline{e}_N)$, $\underline{e}_i = (e_{i1}, \dots, e_{iT})$ para $i = 1, \dots, N$, $\beta = (\beta_1, \dots, \beta_N)$ e supõem quatro hipóteses. A primeira hipótese se relaciona com o quarto momento dos fatores, que convergem para uma matriz definida positiva. A segunda hipótese é que a norma dos vetores que constituem em hipóteses referente as propriedades da matriz de carga fatoriais. terceira hipótese é referente a dependência nas Cross-Section, a dependência temporal e a heterocedasticidade. e a quarta e última hipótese se refere a dependência fraca entre os fatores e os erros idiossincráticos.

Bai e Ng também assumem que os p fatores são estimados por componentes prin-

cipais, eles mostram que o estimador

$$\hat{p} = \arg \min_{0 \leq p \leq p_{max}} (NT)^{-1} \sum_{i=1}^N \sum_{t=1}^T (R_{it} - \bar{\beta}_i^{p'} \hat{v}_t^p) + p\phi(N, T)$$

em que $\bar{\beta}^p$ é construído como \sqrt{N} vezes os autovetores correspondentes aos p maiores autovalores da matriz $N \times N$ $R'R$, $\bar{v}^p = R\bar{\beta}^p/N$ e $\hat{v}^p = \bar{v}^p(\bar{v}^{p'}\bar{v}^p/T)^{1/2}$, tem a seguinte propriedade:

$$\lim_{N, T \rightarrow \infty} Prob[\hat{p} = p] = 1$$

se (i) $\phi(N, T) \rightarrow 0$ e (ii) $(\min\{\sqrt{N}, \sqrt{T}\})^2 \cdot \phi(N, T) \rightarrow \infty$ quando $N, T \rightarrow \infty$.

Baseado nos estudos de Bai e Ng, Giglio e Xiu assumiram as seguintes hipóteses tanto para o desenvolvimento do estimador em três etapas quanto para construir um estimador consistente para p :

- I. f_t é um vetor de fatores de precificação de ativos, em que R_t denota um vetor de excesso de retornos $N \times 1$ de ativos de teste. O modelo de apreçamento satisfaz:

$$R_t = \beta\gamma + \beta v_t + u_t, \quad (2.1)$$

$$f_t = f + v_t, \quad (2.2)$$

$$E(v_t) = E(u_t) = 0, \quad e \quad (2.3)$$

$$Cov(u_t, v_t) = 0, \quad (2.4)$$

em que v_t é um vetor $p \times 1$ de inovações de f_t , u_t é um vetor $N \times 1$ de componentes idiossincráticos, β é uma matriz $N \times 1$ de cargas fatoriais, e γ é o prêmio de risco $p \times 1$.

- II. Existe um vetor $d \times 1$ observável, g_t , de fatores, o qual satisfaz:

$$g_t = \delta + \eta v_t + z_t, \quad (2.5)$$

$$E(z_t) = 0, \quad e \quad (2.6)$$

$$Cov(z_t, v_t) = 0, \quad (2.7)$$

onde as cargas de g em v η é uma matriz $d \times p$, δ é uma constante $d \times 1$, e z_t é um vetor de erro de medida $d \times 1$.

- III. Existe uma constante positiva K , tal que para todo N e T ,

$$(i) \quad T^{-1} \sum_{t=1}^T \sum_{t'=1}^T \left| E \left(N^{-1} \sum_{i=1}^N u_{it} u_{it'} \right) \right| \leq K, \quad \max_{1 \leq t \leq T} E \left(N^{-1} \sum_{i=1}^N u_{it}^2 \right) \leq K.$$

$$(ii) \quad T^{-2} \sum_{s=1}^T \sum_{t=1}^T E \left(\sum_{j=1}^N (u_{js} u_{jt} - E(u_{js} u_{jt})) \right)^2 \leq KN.$$

IV. As inovações fatoriais V satisfazem:

$$\begin{aligned}\|\bar{V}\|_{MAX} &= O_p(T^{-1/2}), \\ \|T^{-1}VV' - \Sigma^v\|_{MAX} &= O_p(T^{-1/2}),\end{aligned}$$

em que Σ^v é uma matriz positiva definida $p \times p$ e $0 < K_1 < \lambda_{\min}(\Sigma^v) \leq \lambda_{\max}(\Sigma^v) < K_2 < \infty$.

V. A matriz de cargas fatoriais β satisfaz

$$\|N^{-1}\beta'\beta - \Sigma^\beta\| = o_p(1), \text{ quando } N \rightarrow \infty,$$

onde Σ^β é uma matriz definida positiva $p \times p$ e $0 < K_1 < \lambda_{\min}(\Sigma^\beta) \leq \lambda_{\max}(\Sigma^\beta) < K_2 < \infty$.

VI. A matriz de cargas fatoriais β e os erros idiossincráticos u_t satisfazem as seguintes condições de momentos, para todo $1 \leq j \leq p$ e para todo N e T :

$$\begin{aligned}(i) \quad E \sum_{t=1}^T \left(\sum_{i=1}^N \beta_{ij} u_{it} \right)^2 &\leq KNT. \\ (ii) \quad E \left(\sum_{t=1}^T \sum_{i=1}^N \beta_{ij} u_{it} \right)^2 &\leq KNT.\end{aligned}$$

O estimador proposto por eles é

$$\hat{p} = \arg \min_{1 \leq j \leq p_{max}} \left(N^{-1}T^{-1} \lambda_j(\bar{R}'\bar{R}) + j \times \phi(N, T) \right) - 1 \quad (2.8)$$

onde p_{max} é algum limite superior de p , $\phi(N, T)$ é uma função de penalidade e $\lambda_j(\bar{R}'\bar{R})$ é o j -ésimo maior autovalor da matriz $\bar{R}'\bar{R}$. Eles mostram que se $\phi(N, T) \rightarrow 0$ quando $N, T \rightarrow \infty$, então temos que $Prob(\hat{p} \geq p) \rightarrow 1$. E se, além disso, $\phi(N, T)/(N^{-1/2} + T^{-1/2}) \rightarrow \infty$, então $\hat{p} \xrightarrow{P} p$.

Esse estimador é utilizado para construir o estimador dos fatores e das cargas fatoriais na primeira etapa conduzindo o PCA da matriz $N^{-1}T^{-1}\bar{R}'\bar{R}$, definindo o seguintes estimadores para os fatores e para as cargas fatoriais:

$$\hat{V} = T^{1/2}(\xi_1 : \xi_2 : \dots : \xi_{\hat{p}})', \text{ e} \quad (2.9)$$

$$\hat{\beta} = T^{-1}\bar{R}\hat{V}' \quad (2.10)$$

em que $\xi_1, \dots, \xi_{\hat{p}}$ são os autovetores correspondentes aos \hat{p} maiores autovalores do PCA da matriz $N^{-1}T^{-1}\bar{R}'\bar{R}$, e $(\xi_1 : \xi_2 : \dots : \xi_{\hat{p}})$ é a concatenação horizontal de matrizes, coluna a coluna, em que as colunas são equivalentes aos vetores ξ_i , para $i \in \{1, \dots, \hat{p}\}$.

A segunda etapa consiste em fazer uma regressão de mínimos quadrados ordinários

cross-sectional (OLS) das médias dos retornos contra as cargas fatoriais estimadas $\hat{\beta}$ para obter o prêmio de risco dos fatores latentes estimados

$$\hat{\gamma} = (\hat{\beta}'\hat{\beta})^{-1} \hat{\beta}'\bar{R}. \quad (2.11)$$

A última etapa consiste em fazer uma regressão de g_t nos fatores extraídos pelo PCA, \hat{V} , para obtermos o estimador $\hat{\eta}$ e o valor corrigido do fator observado:

$$\hat{\eta} = \bar{G}\hat{V}'(\hat{V}\hat{V}')^{-1}, \quad (2.12)$$

$$\hat{G} = \hat{\eta}\hat{V} \quad (2.13)$$

em que \bar{G} é a média da matriz $G = (g_1, g_2, \dots, g_T)$.

Por fim, o estimador do prêmio de risco de g_t é obtido por

$$\hat{\gamma}_g = \hat{\eta}\hat{\gamma} \quad (2.14)$$

$$= \bar{G}\hat{V}'(\hat{V}\hat{V}')^{-1}(\hat{\beta}'\hat{\beta})^{-1}\hat{\beta}'\bar{R} \quad (2.15)$$

3 Metodologia

Neste trabalho iremos assumir que a evolução na cross-section de ativos e no tempo pode ser sumarizada através de uma estrutura de painel, com a especificação geral:

$$R_{it} = \beta_i' d_t + e_{it}, \quad i = 1, \dots, N \quad (3.1)$$

e especificamente utilizamos:

$$\beta_i = \begin{bmatrix} \alpha_i \\ \beta_{iM} \\ \beta_{iSMB} \\ \beta_{iHML} \\ \beta_{iIML} \\ \beta_{iWML} \end{bmatrix}, \quad d_t = \begin{bmatrix} 1 \\ R_{Mt} - R_{ft} \\ SMB_t \\ HML_t \\ IML_t \\ WML_t \end{bmatrix} \quad (3.2)$$

- R_{Mt} é o retorno do mercado no período t,
- R_{ft} é o risk free no período t,
- SMB_t é o fator relacionado ao tamanho no período t.
- HML_t é o fator relacionado ao BE/ME no período t.
- IML_t é o fator relacionado a liquidez no período t.
- WML_t é o fator relacionado aos retornos passados no período t.

Esse modelo é baseado no *modelo de cinco fatores de Fama-French* [12]. O que o difere do modelo do Fama-French são os fatores de liquidez (IML) e dos retornos passados (WML) que substituem os fatores de rentabilidade (RMW) e de investimento (CMA). Essa substituição dos fatores foi necessária pois nosso objetivo era realizar o estudo com os fatores disponíveis pelo NEFIN. Contudo, os fatores CMA e RMW não estavam disponíveis. O nosso trabalho estudou esse modelo em duas etapas. Primeiro estudamos se esses cinco fatores são significantes, e portanto úteis para explicar os retorno das carteiras. Então por último, fizemos previsões com os fatores da forma usual e as previsões com o método do Giglio e Xiu [14], e comparamos qual das previsões nos deu um melhor resultado.

3.1 Suficiência dos Fatores

A escolha dos estimadores foi feita pensando na necessidade da comparação dos modelos. Nosso objetivo é tentar identificar a possibilidade de omissão de fatores no modelo (3.1), utilizando um diagnóstico simples comparando a estimação de um modelo de dados em painel não robusto a presença de fatores omitidos (estimador Mean Group - MG) com um estimador de robusto para dados em painel na presença de fatores latentes, dado pelo estimador de Common Correlated Effects (CCE). Note que o uso de painel na estimação de prêmio de risco é um procedimento comum na estimação do prêmio de risco de fatores, e pode ser pensado como uma forma alternativa de estimação em relação ao procedimento de Fama-Macbeth [7], e o uso de modelos em painel para a estimação de modelos multifatoriais é discutida em Petersen (2009) [26].

O estimador MG é uma média simples dos estimadores OLS de cada grupo, enquanto o estimador CCE é extensão do estimador MG assumindo fatores comuns correlacionados não observados nos erros. Logo, a ideia por trás do estimador CCE é a mesma que queremos testar. Por este motivo, escolhemos comparar o Mean Group com o CCE. Nas próximas subseções detalharemos mais esses estimadores.

3.1.1 Estimador Mean Group

Para obter o estimador Mean Group para o modelo de dados em painel heterogêneo (3.1) consideramos as seguintes matrizes:

$$D = \begin{bmatrix} d_1 & d_2 & \dots & d_T \end{bmatrix}' \quad (3.3)$$

$$R_i = \begin{bmatrix} R_{i1} & R_{i2} & \dots & R_{iT} \end{bmatrix}' \quad (3.4)$$

Então, o primeiro passo foi calcular os estimadores OLS de cada β_i , conforme a equação abaixo:

$$\hat{\beta}_i = (D'D)^{-1}D'R_i \quad (3.5)$$

Por fim, obtemos o estimador MG conforme a equação abaixo:

$$\hat{\beta}^F = \frac{1}{N} \sum_{i=1}^N \hat{\beta}_i \quad (3.6)$$

3.1.2 Estimador de Efeitos Comuns Correlacionados

Para calcular o estimador de efeitos comum correlacionados, consideramos o modelo de dados em painel heterogêneo (3.1) e assumimos que o erro e_{it} tem a seguinte estrutura fatorial comum

$$e_{it} = \sum_{j=1}^m \gamma_{ij} f_{jt} + \varepsilon_{it} = \boldsymbol{\gamma}'_i \mathbf{f}_t + \varepsilon_{it} \quad (3.7)$$

em que $\mathbf{f}_t = (f_{1t}, \dots, f_{mt})'$ é um vetor de fatores comuns não observados e $\boldsymbol{\gamma}_i = (\gamma_{i1}, \dots, \gamma_{im})'$ é o vetor das cargas fatoriais. Assumimos que o número de fatores, m , é fixo e $m \ll N$, em que N é o número de ativos.

Então, substituindo (3.7) em (3.1), o nosso modelo possui a seguinte forma:

$$R_{it} = \beta_i' d_t + \boldsymbol{\gamma}_i' \mathbf{f}_t + \varepsilon_{it} \quad (3.8)$$

O estimador de efeitos comuns correlacionados (CCE) consiste em aproximar a combinação linear dos fatores não observados pelas médias das cross-section das variáveis dependentes e explicativas, e então calcular a regressão para o painel padrão aumentado com as médias das cross-section.

Para calcular as médias consideramos um vetor de pesos não estocástico $\mathbf{w}_t = (w_{1t}, w_{2t}, \dots, w_{Nt})'$, para $t \in \mathcal{T} \subset \mathbb{Z}$, em que \mathcal{T} é o nosso horizonte de tempo. O vetor \mathbf{w}_t foi escolhido para satisfazer as duas hipóteses abaixo:

$$|\mathbf{w}_t| = (\mathbf{w}_t' \mathbf{w}_t)^{\frac{1}{2}} = O(N^{-\frac{1}{2}}), \quad (3.9)$$

$$\frac{w_{jt}}{|\mathbf{w}_t|} = O(N^{-\frac{1}{2}}) \text{ uniformemente em } j \in \mathbb{N}. \quad (3.10)$$

Assim, as médias foram calculadas da seguinte forma

$$\bar{R}_{wt} = \bar{\beta}_w' d_t + \bar{\boldsymbol{\gamma}}_w' \mathbf{f}_t + \bar{\varepsilon}_{wt} \quad (3.11)$$

onde

$$\bar{R}_{wt} = \sum_{i=1}^N w_i R_{it}, \quad \bar{\beta}_w = \sum_{i=1}^N w_i \beta_i, \quad (3.12)$$

$$\bar{\boldsymbol{\gamma}}_w = \sum_{i=1}^N w_i \boldsymbol{\gamma}_i, \quad \bar{\varepsilon}_{wt} = \sum_{i=1}^N w_i \varepsilon_{it} \quad (3.13)$$

E a partir da regressão do modelo (3.11) calculamos f_t e $\hat{\beta}_i^P$.

3.1.3 Teste de Wald

A ideia de compararmos os estimadores MG e CCE é identificar a possível presença de fatores omitidos na estimação, uma vez que o estimador CCE assume uma estrutura de fatores comuns não observados para os erros. Ao estimar o MG e o CCE obtivemos seus coeficientes e as covariâncias dos seus parâmetros para o modelo (3.1). Escolhemos fazer o teste de Wald, já que ele consiste em avaliar as restrições nos parâmetros estatísticos baseados na distância ponderada entre a estimativa irrestrita e seu valor hipotético sob a hipótese nula.

O primeiro teste feito foi um teste de Wald tradicional assumindo o modelo proposto pelo estimador MG e, portanto, sem fatores comuns nos erros. Assumimos como hipótese nula que os coeficientes são iguais aos coeficientes estimados pelo estimador MG, e testamos se o estimador CCE é igual.

De forma, análoga, no segundo teste, assumimos que o modelo estimado é o modelo proposto pelo estimador CCE e, portanto, tem fatores comuns nos erros. Nossa hipótese nula foi que os coeficientes são iguais aos coeficientes estimados pelo estimador CCE, e testamos se o estimador MG é igual.

Por fim, o último teste se resumiu a usar as estimativas de covariâncias dos parâmetros dos dois modelos no teste de Wald. Esse teste é similar ao teste F que é feito em análise de variância (ANOVA).

3.2 Método de Giglio e Xiu

O nosso modelo (3.1) só possui fatores observados. Aplicamos o método do Giglio e Xiu para calcular o prêmio de risco desses cinco fatores, controlando para a presença de possíveis fatores omitidos e erros de medida.

Para aplicarmos esse método definimos $R_t = (R_{1t}, \dots, R_{Nt})'$ e assumimos que

$$R_t = \beta\gamma + \beta v_t + u_t, \quad (3.14)$$

$$f_t = f + v_t, \quad (3.15)$$

$$E(v_t) = E(u_t) = 0, \quad e \quad (3.16)$$

$$Cov(u_t, v_t) = 0, \quad (3.17)$$

Definimos o vetor $g_t = (R_{Mt} - R_{ft}, SMB_t, HML_t, IML_t, WML_t)'$ (5×1). Observe que $d_t = (1, g_t)'$.

Nosso objetivo é estimar o prêmio de risco de g_t corrigido para os fatores latentes e usar esse prêmio de risco para obter os parâmetros do modelo (3.1). Para isso assumimos que

$$g_t = \delta + \eta v_t + z_t, \quad (3.18)$$

$$E(z_t) = 0, \quad e \quad (3.19)$$

$$Cov(z_t, v_t) = 0, \quad (3.20)$$

Denotamos por R , V , G , U e Z as seguintes matrizes

$$\underset{(N \times T)}{R} = \begin{bmatrix} R_1 & R_2 & \dots & R_T \end{bmatrix}, \quad (3.21)$$

$$\underset{(p \times T)}{V} = \begin{bmatrix} v_1 & v_2 & \dots & v_T \end{bmatrix}, \quad (3.22)$$

$$\underset{(5 \times T)}{G} = \begin{bmatrix} R_{M1} - R_{f1} & R_{M2} - R_{f2} & \dots & R_{MT} - R_{fT} \\ SMB_1 & SMB_2 & \dots & SMMB_T \\ HML_1 & HML_2 & \dots & HML_T \\ IML_1 & IML_2 & \dots & IML_T \\ WML_1 & WML_2 & \dots & WML_T \end{bmatrix}, \quad (3.23)$$

$$\underset{(N \times T)}{U} = \begin{bmatrix} e_1 & e_2 & \dots & e_T \end{bmatrix}, \quad (3.24)$$

$$\underset{(5 \times T)}{Z} = \begin{bmatrix} z_1 & z_2 & \dots & z_T \end{bmatrix}. \quad (3.25)$$

E com essas matrizes reescrevemos o modelo usado por Giglio e Xiu da seguinte forma

$$R = \beta\gamma + \beta V + U \quad (3.26)$$

$$G = \xi + \eta V + Z \quad (3.27)$$

Denotamos por $(\bar{R}, \bar{V}, \bar{G}, \bar{U}, \bar{Z})$ as matrizes das médias das respectivas variáveis. E portanto, temos que as equações acima se tornam

$$\bar{R} = \beta\bar{V} + \bar{U}, \quad (3.28)$$

$$\bar{G} = \eta\bar{V} + \bar{Z}. \quad (3.29)$$

Segundo Bai e Ng [2], o número de fatores estimados pelo método assintótico de componentes principais é $\min\{N, T\}$. Como utilizamos componentes principais nos passos futuros, adotamos $p_{max} = \min\{N, T\}$. Analisamos dois estimadores \hat{p}_j , $j = 1, 2$:

$$\hat{p}_1 = \arg \min_{1 \leq j \leq p_{max}} \left(N^{-1}T^{-1} \lambda_j(\bar{R}'\bar{R}) + j \times \phi(N, T) \right) - 1 \quad (3.30)$$

$$\hat{p}_2 = \arg \min_{1 \leq j \leq p_{max}} \left(N^{-1}T^{-1} \lambda_j(\bar{R}'\bar{R}) + j \times \phi(N, T) \right) \quad (3.31)$$

. O estimador \hat{p}_1 é o mesmo estimador proposto por Giglio e Xiu, já o \hat{p}_2 é baseado no estimador proposto por Bi e NG

Para cada \hat{p}_i , $i = 1, 2$, testamos 4 funções $\phi_k(N, T)$ distintas, $k = 1, 2, 3, 4$:

$$\phi_1 = \left(\log \left(\left(N^{-1/4} + T^{-1/4} \right)^{-1} \right) \right) \times \left(N^{-1/4} + T^{-1/4} \right) \quad (3.32)$$

$$\phi_2 = \left(\log \left(\frac{N \times T}{N + T} \right) \right) \times \left(\frac{N + T}{N \times T} \right) \quad (3.33)$$

$$\phi_3 = \left(\log \left(\min \{N, T\}^2 \right) \right) \times \left(\frac{N + T}{N \times T} \right) \quad (3.34)$$

$$\phi_4 = \frac{\log \left(\min \{N, T\}^2 \right)}{\min \{N, T\}^2} \quad (3.35)$$

e escolhemos o estimador que obteve melhores resultado. Ao todo, testamos 8 estimadores definidos pela seguinte equação:

$$\hat{p}_j^k = \begin{cases} \arg \min_{1 \leq l \leq p_{max}} \left((NT)^{-1} \lambda_l (\bar{R}' \bar{R}) + l \times \phi_k(N, T) \right) - 1 & , \text{ se } j = 1 \\ \arg \min_{1 \leq l \leq p_{max}} \left((NT)^{-1} \lambda_j (\bar{R}' \bar{R}) + l \times \phi_k(N, T) \right) & , \text{ se } j = 2 \end{cases} \quad (3.36)$$

Temos que $\phi_k(N, T) \rightarrow 0$ quando $N, T \rightarrow \infty$, para $k \in \{1, 2, 3, 4\}$. Porém, somente a função ϕ_1 possui a seguinte propriedade: $\phi_1(N, T)/(N^{-1/2} + T^{-1/2}) \rightarrow \infty$, quando $N, T \rightarrow \infty$.

Ao obter a estimativa \hat{p} do números de fatores, realizamos a primeira etapa do método do Giglio e Xiu, calculando o estimador dos fatores \hat{V} e o estimador das cargas fatoriais $\hat{\beta}$ foi calculado conforme

$$\hat{V} = T^{1/2} (\xi_1 : \xi_2 : \dots : \xi_{\hat{p}})', \text{ e} \quad (3.37)$$

$$\hat{\beta} = T^{-1} \bar{R} \hat{V}' \quad (3.38)$$

Na segunda etapa do método calculamos, através de um OLS na média dos retornos \bar{R} , o estimador do prêmio de risco dos fatores latentes $\hat{\gamma}$ conforme

$$\hat{\gamma} = \left(\hat{\beta}' \hat{\beta} \right)^{-1} \hat{\beta}' \bar{R}. \quad (3.39)$$

Por fim, com a última etapa obtivemos os estimadores $\hat{\eta}$ e \hat{G} das cargas fatoriais de g em v e do valor corrigido dos fatores observados após remoção de erros de medida, respectivamente. O estimador $\hat{\eta}$ e o estimador \hat{G} foram obtidos conforme

$$\hat{\eta} = \bar{G} \hat{V}' \left(\hat{V} \hat{V}' \right)^{-1}, \quad (3.40)$$

$$\hat{G} = \hat{\eta} \hat{V} \quad (3.41)$$

Então, utilizando os estimadores anteriores para estimar o prêmio de risco de g_t , que são

os cinco fatores observados, conforme

$$\hat{\gamma}_g = \hat{\eta} \hat{\gamma} \tag{3.42}$$

$$= \bar{G} \hat{V}' (\hat{V} \hat{V}')^{-1} (\hat{\beta}' \hat{\beta})^{-1} \hat{\beta}' \bar{R} \tag{3.43}$$

.

3.3 Previsões

A última parte do nosso trabalho foi a comparação das previsões usuais do modelo Fama-French com a corrigida pelo método de Giglio e Xiu.

Utilizamos o vetor de prêmio de risco $\hat{\gamma}_g$ para recuperar as cargas fatoriais do g_t obtendo assim um estimador $\hat{\beta}^G$. Então aplicamos esse estimador no modelo (3.1) para fazer as previsões dos retornos dos N ativos comparados com as previsões $\hat{\beta}^F$ obtidas pela regressão usual do modelo do (3.1).

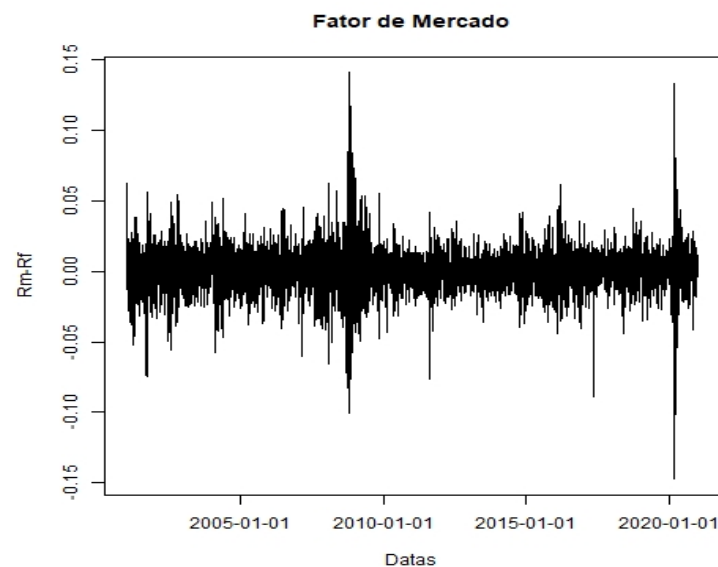
4 Base de Dados

Todos os dados desse trabalhos foram construídos pelo NEFIN da USP e seu período é de janeiro de 2001 a dezembro de 2020, para dados com frequência diária. Abaixo temos a descrição da construção realizada pelo NEFIN dos fatores e das 12 carteiras que selecionamos.

O fator livre de risco de um ano (R_f - risk free) foi calculado a partir do DI Swap de 360 dias, deflacionado pela inflação esperada medida pelo índice IPCA (dados disponíveis no site do Banco Central do Brasil).

O fator de Mercado ($R_M - R_f$) é diferença entre o retorno diário a valores ponderados da carteira de mercado e a taxa livre de risco diária, a qual é calculada do 30-day DI Swap.

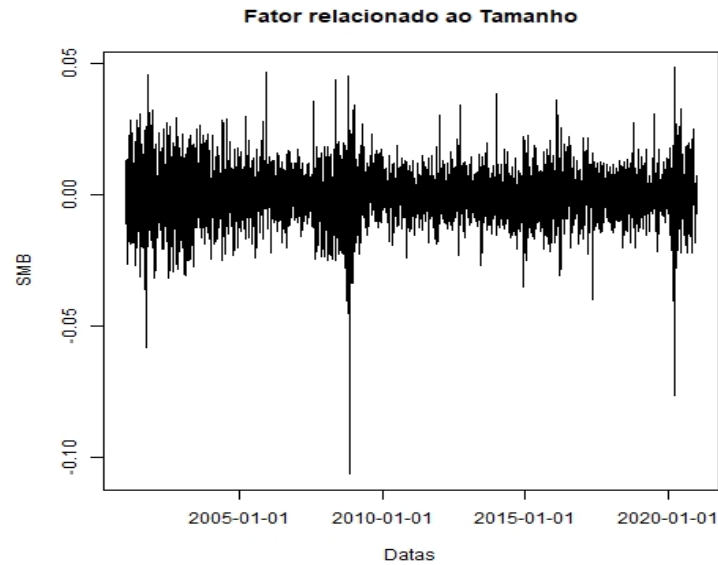
Figura 1 – Gráfico dos Retornos do Fator de Mercado



O fator relacionado ao tamanho *SMB* (*Small Minus Big*) é o retorno de uma carteira comprada em ações com baixa capitalização de mercado ("*Small*") e ações vendidas com alta capitalização de mercado ("*Big*"). Todo mês de janeiro do ano t as ações são classificadas elegíveis de acordo com a capitalização de mercado de dezembro do ano $t - 1$, e são separadas em 3 quantis (carteiras). Então, calculamos com ponderação igual os retornos da primeira carteira ("*Small*") e da terceira carteira ("*Big*"). O fator *SMB* é o retorno da carteira "*Small*" menos o retorno da carteira "*Big*".

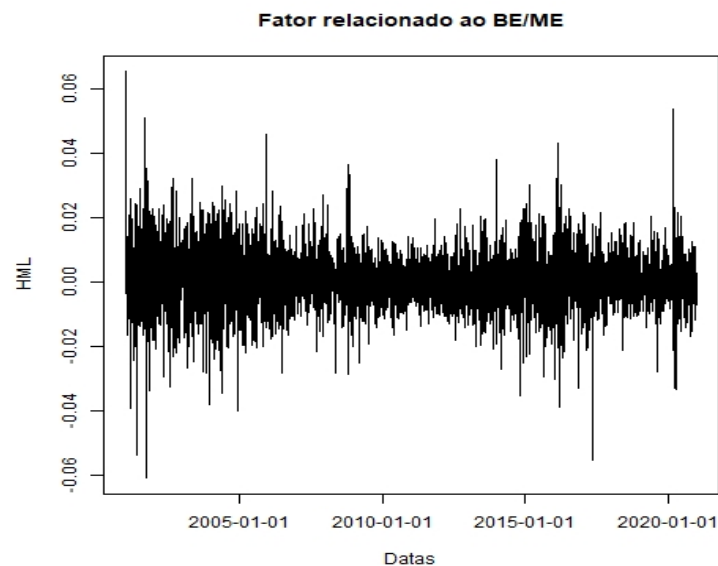
O fator relacionado ao BE/ME é o fator *HML* (*High Minus Low*). Esse retorno é o retorno de uma carteira comprada em ações com alta razão book-to-market ("*High*") e vendidas com baixa razão book-to market ("*Low*"). Todo mês de Janeiro do ano t as

Figura 2 – Gráfico dos Retornos do Fator de Tamanho



ações são classificadas elegíveis (de forma crescente) e divididas em 3 quantis (carteiras) conforme a razão book-to-market das firmas em Junho do ano $t - 1$. Então calculamos com ponderação igual os retornos da carteira "High" menos os retornos da carteira "Low".

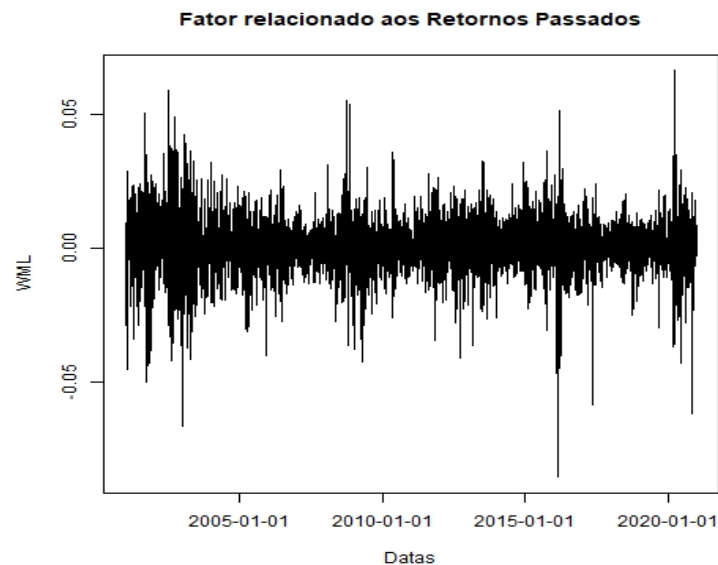
Figura 3 – Gráfico dos Retornos do Fator de Book-to-Market



O fator WML (*Winners Minus Losers*) é o retorno de uma carteira comprada em ações com retornos passados altos ("Winners") e vendidas com retornos passados baixos ("Losers"). Todos mês t as ações são classificadas elegíveis (de forma crescente) e divididas em 3 quantis (carteiras) conforme seus retornos acumulados entre os meses $t - 12$ e $t - 2$. Então calculamos com ponderação igual os retornos da primeira carteira ("Losers") e a terceira carteira ("Winners"). O fator WML é o retorno da carteira "Winners" menos o

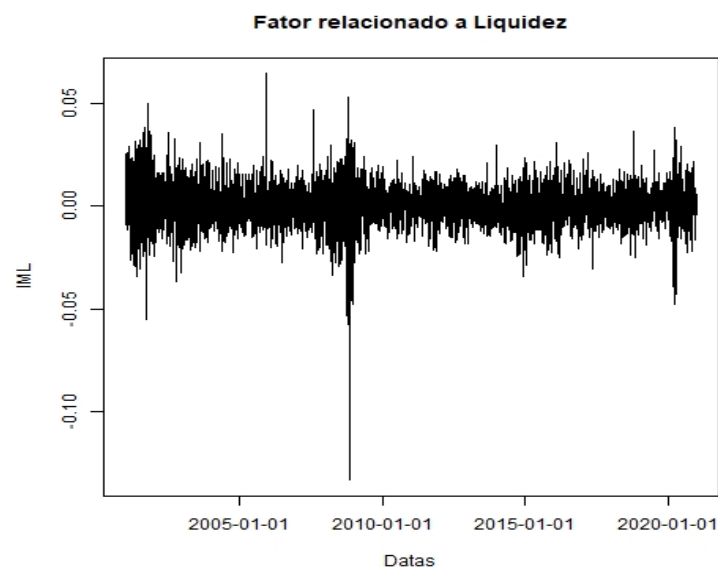
retorno da carteira "Losers".

Figura 4 – Gráfico dos Retornos do Fator relacionado aos Retornos Passados



O fator *IML* ("Illiquid Minus Liquid") é o retorno de uma carteira comprada em ações com alta iliquidez ("Illiquid") e vendidas com baixa iliquidez ("Liquid"). Todo mês t , classificamos as ações elegíveis (de forma crescente) em 3 quantis (carteiras) conforme a média móvel da iliquidez dos doze meses anteriores (a iliquidez das ações é calculada conforme Acharya e Pedersen [1]). Então calculamos com ponderação igual os retornos da primeira carteira ("Liquid") e da terceira carteira ("Illiquid"). O fator *IML* é o retorno da carteira "Illiquid" menos o retorno da carteira "Liquid".

Figura 5 – Gráfico dos Retornos do Fator de Iliquidez

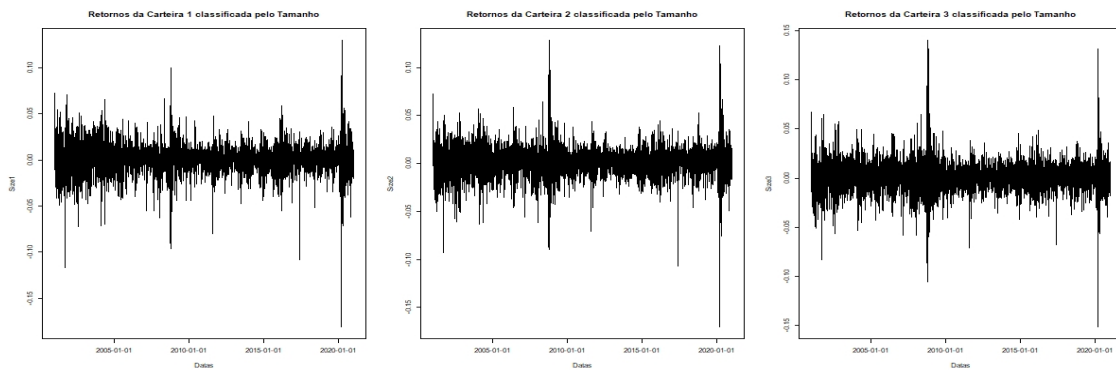


As 12 carteiras são divididas em quatro grupos:

- (i) 3 carteiras classificadas pelo tamanho.
- (ii) 3 carteiras classificadas pelo book-to-market.
- (iii) 3 carteiras classificadas pelo momentum.
- (iv) 3 carteiras classificadas pela iliquidez.

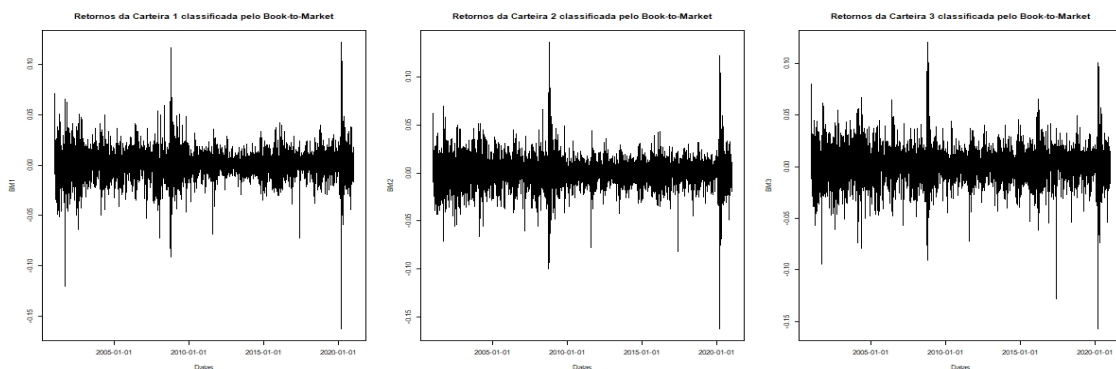
As carteiras classificadas pelo tamanho são obtidas da seguinte forma: Todo mês de janeiro do ano t , as ações elegíveis são classificadas de forma crescente em tercís de acordo com a sua capitalização de mercado em dezembro do ano $t - 1$. Em seguida, as carteiras são mantidas durante o ano t .

Figura 6 – Gráfico dos Retornos das Carteiras Classificadas pelo Tamanho



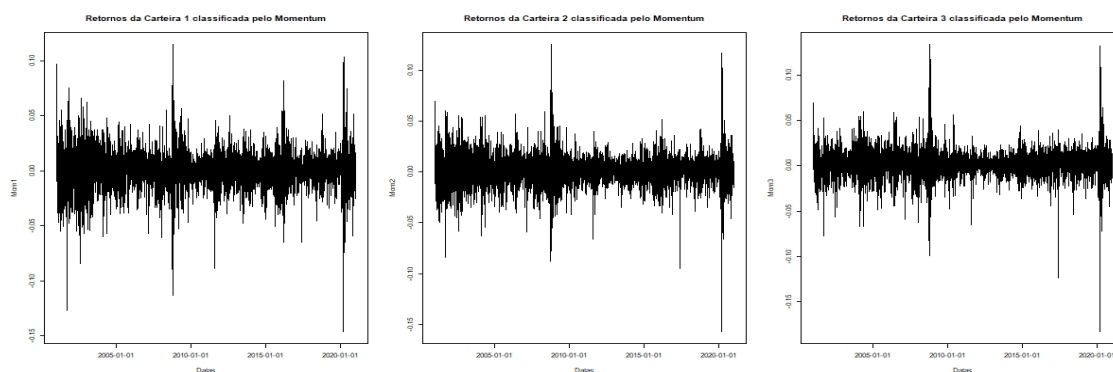
As carteiras classificadas pelo book-to-market são obtidas da seguinte forma: Todo mês de janeiro do ano t , as ações elegíveis são classificadas de forma crescente em tercís, de acordo com a razão entre o valor contábil e o valor de mercado em junho do ano $t - 1$. Em seguida, as carteiras são mantidas durante o ano t .

Figura 7 – Gráfico dos Retornos das Carteiras Classificadas pelo Book-to-Market



As carteiras classificadas por momentum são obtidas da seguinte forma: Todo mês t , as ações elegíveis são classificadas de forma crescente em tercís de acordo com seus retornos acumulados do mês $t - 12$ e do mês $t - 2$. Em seguida, as carteiras são mantidas

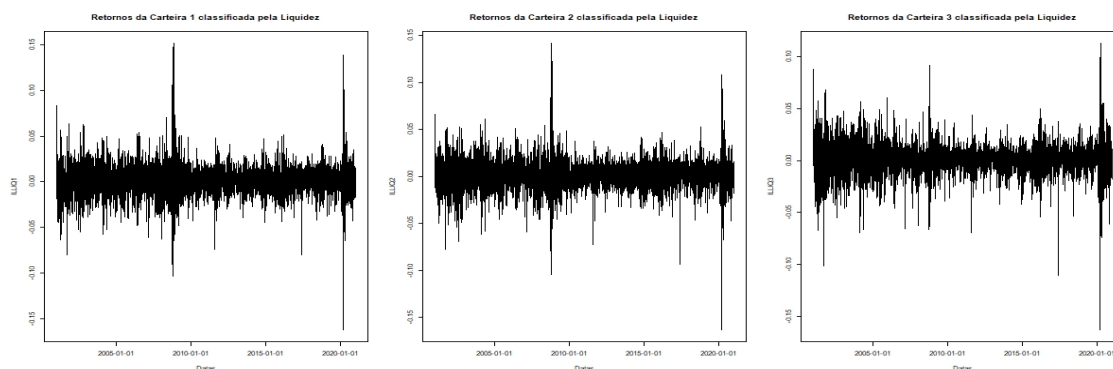
Figura 8 – Gráfico dos Retornos das Carteiras Classificadas pelo Momentum



durante o mês t .

As carteiras classificadas por iliquidez são obtidas da seguinte forma: todo mês t , as ações elegíveis são classificadas de forma crescente em terceis de acordo com a média móvel de iliquidez de doze meses anteriores, como em Amihud (2002) **Amihud**. Em seguida, mantemos as carteiras durante o mês t .

Figura 9 – Gráfico dos Retornos das Carteiras Classificadas pela Iliquidez



As ações negociadas na BOVESPA consideradas elegíveis atendem três critérios: A ação é a ação mais negociada da empresa (ou seja, aquela com o maior volume negociado durante o ano passado); As ações foram negociadas em mais de 80% dos dias do ano $t - 1$, com volume superior a R\$ 500,000.00 por dia, e caso a ação tenha sido listada no ano $t - 1$, o período considerado vai do dia da listagem até o último dia do ano; As ações foram listadas inicialmente antes de dezembro do ano $t - 1$.

5 Resultados

5.1 Teste de Wald para os estimadores MG e CCE

O primeiro passo para realizar os testes de Wald foi calcular os coeficientes de cada um dos estimadores para o modelo (3.1). Na tabela 1 apresentamos os resultados obtidos para o estimador Mean Group para o modelo (3.1). Observamos que para esse estimador o Estatística t do fator de Mercado e o de tamanho é maior do que dois. Além disso, o R-quadrado estimado para o modelo foi de aproximadamente 0.91133, o que nos leva a acreditar que ele explica bem a variação observada nos retornos e os fatores que mais influenciam são o de mercado e o de tamanho.

Tabela 1 – Estimador MG do modelo (3.1)

	$R_m - R_f$	SMB	HML	IML	WML
Estimativa	0.9548	0.2987	0.0121	0.0653	-0.0511
Desvio Padrão	0.0041	0.0644	0.0623	0.0641	0.0617
Estatística t	236.5998	4.6361	0.1934	1.0192	-0.8270

Nota: O $R - Quadrado$ estimado para o modelo foi aproximadamente 0.91133. Os resultados acima foram obtidos calculando o estimador MG para um painel com $n \times T$, em que $n = 12$ e $T = 4950$, que resulta em um total de 59400 dados.

Na tabela 2 apresentamos os resultados do estimador CCE para o modelo (3.1). Na tabela 3 apresentamos os resultados do teste de Wald com a hipótese nula de que os coeficientes são iguais aos coeficientes obtidos pelo estimador Mean Group $\hat{\beta}^F$. Na tabela 4 apresentamos os resultados do teste de Wald de que os coeficientes são iguais aos coeficientes obtidos pelo estimador CCE $\hat{\beta}^P$.

Tabela 2 – Estimador CCE do modelo (3.1)

	$R_m - R_f$	SMB	HML	IML	WML
Estimativa	0.0918969	1.2449408	0.1621465	0.0074829	-0.1936320
Desvio Padrão	0.1189371	0.8526953	0.1179169	0.3221000	0.1493925
Estatística t	0.7727	1.4600	1.3751	0.0232	-1.2961

Nota: O $R - Quadrado$ do modelo foi aproximadamente 0.97151. Os resultados acima foram obtidos calculando o estimador CCE-MG para um painel com $n \times T$, em que $n = 12$ e $T = 4950$, que resulta em um total de 59400 dados.

Na tabela 3 apresentamos os testes de Wald realizados. Como explicado na seção 3.1.3, realizamos três testes de Wald. No **Teste 1** foi realizamos um teste de Wald simples, assumindo o modelo (3.1) e que seus coeficientes eram iguais ao $\hat{\beta}^F$ e a hipótese nula era que $\beta = \hat{\beta}^P$. Podemos observar que o **Teste 1** rejeita a hipótese nula. Para o **Teste 2**, assumimos que o modelo (3.1) possui a estrutura de erros do estimador CCE e os coeficientes eram iguais ao $\hat{\beta}^P$ e a hipótese nula era que $\beta = \hat{\beta}^F$. Notamos na tabela

3 que o **Teste 2** rejeita a hipótese nula. O **Teste 3** é o teste de Wald comparando as estimativas de covariâncias dos dois modelos. Nesse teste testamos a nossa hipótese nula é que as covariâncias são iguais. Ele também rejeita a hipótese nula. Com isso, concluímos que têm fortes indícios de que o modelo possui fatores omitidos.

Tabela 3 – Testes de Wald

	Chisq	Pr(>Chisq)
Teste 1	434.53	0.00
Teste 2	49066.00	0.00
Teste 3	601.00	0.00

Nota: Esse teste consiste em um Teste de Hipótese Linear que calcula uma estatística F comparando o modelo e os resultados obtidos pelo estimador CCE-MG com os coeficientes obtidos pelo modelo MG.

5.2 Resultado da Simulação do estimador \hat{p}

O estimador de Giglio e Xiu é realizado em três etapas. Para finalizar a primeira etapa é necessário um estimador de número de fatores. No artigo que nos baseamos, foi proposto o estimador da equação (3.30). Contudo é necessário encontrar uma função penalidade $\phi(n, t)$ que possua as propriedades necessárias para a convergência e dê bons resultados de estimação. Escolhemos quatro funções penalidades e nos baseamos no paper do Bai e Ng [2] para realizar as simulações. A nossa ideia, foi realizar um teste semelhante às simulações do modelo homocedástico que eles adotaram.

Para cada estimador \hat{p}_j^k , em que $k \in \{1, \dots, 4\}$ e $j \in \{1, 2\}$, definido por

$$\hat{p}_j^k = \begin{cases} \arg \min_{1 \leq l \leq p_{max}} \left((NT)^{-1} \lambda_l(\bar{R}' \bar{R}) + l \times \phi_k(N, T) \right) - 1 & , \text{ se } j = 1 \\ \arg \min_{1 \leq l \leq p_{max}} \left((NT)^{-1} \lambda_j(\bar{R}' \bar{R}) + l \times \phi_k(N, T) \right) & , \text{ se } j = 2 \end{cases} \quad (5.1)$$

Escolhemos 19 pares (N, T) . Para cada par (N, T) , vamos gerar dados X que depende de uma quantidade f de fatores, $f \in \{1, 3, 4\}$. Ou seja, X será gerada a partir de um fator, ou de três fatores ou de 5 fatores. Abaixo temos a equação que representa o processo:

$$\underset{(N \times T)}{X} = \underset{(N \times f)}{C} \underset{(f \times T)}{F} + \underset{(T \times N)}{E}' \quad (5.2)$$

Todas as nossas matrizes foram geradas a partir de um processo normal multivariado: C é a matriz de cargas $(N \times f)$ gerado por um uma variável aleatória que segue uma $\mathcal{N}(\mu_f, \Sigma_f)$ de tamanho N , F é a matriz de fatores $(f \times T)$ gerado por uma variável aleatória que segue uma $\mathcal{N}(\mu_T, \Sigma_T)$ de tamanho f e E é a matriz de erros $(N \times T)$ gerada por uma

variável aleatória que segue $\mathcal{N}(\mu_N, \Sigma_N)$ de tamanho T , em que

$$\mu_r = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \text{ para } r \in \{f, t_i, n_i\} \quad (5.3)$$

$$\Sigma_r = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}, \text{ para } r \in \{f, t_i, n_i\} \quad (5.4)$$

Foram realizadas 1000 simulações e em cada simulação foi estimado o número de fatores de X utilizando cada um dos estimadores \hat{p}_j^k . Por fim, foi obtido um estimador $\bar{p}_{j,(c_i,f)}^k$, que é a média dos estimadores obtidos nas 1000 simulações.

Nas tabelas 4 a 6, mostramos os resultados obtidos pelos estimadores p_j^k para cada par (N, T) . Também reportamos o erro quadrático médio de cada estimador nas 1000 simulações.

Tabela 4 – Média e EQM dos estimadores para o número de fatores $f = 1$.

N	T	p_1^1	p_1^2	p_1^3	p_1^4	p_2^1	p_2^2	p_2^3	p_2^4
40	100	1.000	1.000	1.000	1.911	2.000	2.000	2.000	2.919
60	100	1.000	1.000	1.000	3.358	2.000	2.000	2.000	4.307
60	200	1.000	1.000	1.000	1.568	2.000	2.000	2.000	2.590
60	500	1.000	1.000	1.000	1.057	2.000	2.000	2.000	2.041
60	2000	1.000	1.000	1.000	1.000	2.000	2.000	2.000	2.000
100	40	1.000	1.000	1.000	1.917	2.000	2.000	2.000	2.890
100	60	1.000	1.000	1.000	3.335	2.000	2.000	2.000	4.297
100	100	1.000	1.000	1.000	8.841	2.000	2.000	2.000	9.902
200	60	1.000	1.000	1.000	1.554	2.000	2.000	2.000	2.554
200	100	1.000	1.000	1.000	2.956	2.000	2.000	2.000	4.049
500	60	1.000	1.000	1.000	1.059	2.000	2.000	2.000	2.059
500	100	1.000	1.000	1.000	1.253	2.000	2.000	2.000	2.268
1000	60	1.000	1.000	1.000	1.001	2.000	2.000	2.000	2.000
1000	100	1.000	1.000	1.000	1.037	2.000	2.000	2.000	2.036
2000	60	1.000	1.000	1.000	1.000	2.000	2.000	2.000	2.000
2000	100	1.000	1.000	1.000	1.001	2.000	2.000	2.000	2.000
4000	60	1.000	1.000	1.000	1.000	2.000	2.000	2.000	2.000
4000	100	1.000	1.000	1.000	1.000	2.000	2.000	2.000	2.000
12	4950	0.991	0.998	0.946	1.000	1.989	1.996	1.959	2.000
EQM		0.0000043	0.0000002	0.0002	4.1345	0.9988	0.9996	0.9958	7.0754

Tabela 5 – Média dos estimadores para o número de fatores $f = 3$.

N	T	\bar{p}_1^1	\bar{p}_1^2	\bar{p}_1^3	\bar{p}_1^4	\bar{p}_2^1	\bar{p}_2^2	\bar{p}_2^3	\bar{p}_2^4
40	100	2.970	2.999	2.950	14.914	3.971	4.000	3.956	15.886
60	100	2.971	3.000	2.997	24.844	3.981	4.000	4.000	26.062
60	200	2.941	3.000	3.000	13.514	3.943	4.000	4.000	14.497
60	500	2.907	3.000	3.000	5.266	3.891	4.000	4.000	6.286
60	2000	2.742	3.000	3.000	3.162	3.747	4.000	4.000	4.168
100	40	2.959	3.000	2.945	14.801	3.983	4.000	3.965	16.094
100	60	2.956	3.000	2.997	24.933	3.985	4.000	3.998	25.895
100	100	2.971	3.000	3.000	46.294	3.963	4.000	4.000	47.287
200	60	2.944	3.000	3.000	13.548	3.952	4.000	4.000	14.544
200	100	2.879	3.000	3.000	29.873	3.923	4.000	4.000	30.779
500	60	2.880	3.000	3.000	5.240	3.895	4.000	4.000	6.319
500	100	2.782	3.000	3.000	10.012	3.790	4.000	4.000	11.066
1000	60	2.835	3.000	3.000	3.607	3.826	4.000	4.000	4.630
1000	100	2.658	3.000	3.000	4.898	3.644	4.000	4.000	5.955
2000	60	2.776	3.000	3.000	3.156	3.788	4.000	4.000	4.141
2000	100	2.502	3.000	3.000	3.510	3.564	4.000	4.000	4.526
4000	60	2.675	3.000	3.000	3.017	3.720	4.000	4.000	4.018
4000	100	2.436	3.000	3.000	3.113	3.54	4.000	4.000	4.146
12	4950	2.662	2.887	1.836	2.999	3.651	3.877	2.815	4.000
EQM		0.0603	0.0007	0.0716	216.9143	0.7073	0.9878	0.9408	236.7685

Tabela 6 – Média dos estimadores para o número de fatores $f = 5$.

N	T	\bar{p}_1^1	\bar{p}_1^2	\bar{p}_1^3	\bar{p}_1^4	\bar{p}_2^1	\bar{p}_2^2	\bar{p}_2^3	\bar{p}_2^4
40	100	3.456	4.993	2.824	40.000	4.461	5.99	3.945	41.000
60	100	2.680	5.000	4.467	41.944	3.587	6.000	5.429	42.8100
60	200	1.513	5.000	4.987	60.000	2.494	6.000	6.000	61.000
60	500	0.709	5.000	5.000	60.000	1.678	6.000	6.000	61.000
60	2000	0.370	5.000	5.000	5.840	1.416	6.000	6.000	6.823
100	40	3.485	4.995	2.883	30.275	4.545	5.991	4.052	31.391
100	60	2.578	5.000	4.448	41.699	3.649	6.000	5.505	42.716
100	100	1.446	5.000	4.981	64.981	2.387	6.000	5.979	66.002
200	60	1.426	5.000	4.996	33.890	2.457	6.000	6.000	34.917
200	100	0.547	5.000	5.000	59.760	1.508	6.000	6.000	60.943
500	60	0.693	5.000	5.000	14.502	1.736	6.000	6.000	15.647
500	100	0.246	5.000	5.000	30.310	1.222	6.000	6.000	31.619
1000	60	0.501	5.000	5.000	7.901	1.461	6.000	6.000	8.913
1000	100	0.146	5.000	5.000	13.515	1.157	6.000	6.000	14.562
2000	60	0.395	5.000	5.000	5.848	1.388	6.000	6.000	6.842
2000	100	0.131	5.000	5.000	7.480	1.114	6.000	6.000	8.474
4000	60	0.365	5.000	5.000	5.236	1.347	6.000	6.000	6.251
4000	100	0.101	5.000	5.000	5.719	1.087	6.000	6.000	6.693
12	4950	3.084	3.952	1.744	12.000	4.016	4.925	2.723	13.000
EQM		15.3767	0.0578	1.1267	996.0771	8.7872	0.8930	1.1365	1046.2293

Com os resultados apresentados nas tabelas 4 - 6, observamos que o estimador

$$\hat{p}_1^2 = \arg \min_{1 \leq l \leq p_{max}} \left[(NT)^{-1} \lambda_j (\bar{R}' \bar{R}) + l \times \left(\log \left(\frac{N \times T}{N + T} \right) \right) \times \left(\frac{N + T}{N \times T} \right) \right] - 1 \quad (5.5)$$

apresenta o menor erro quadrático médio para todos os fatores e, portanto, concluímos que ele é o melhor estimador dentre os estimadores escolhidos. Por esse motivo ele será usado para estimar o número de fatores na próxima etapa.

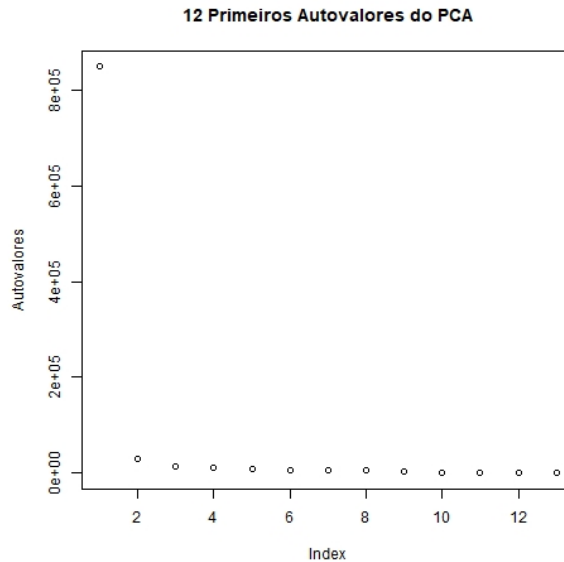
Também acreditamos que o fato da função ϕ_4 convergir mais lentamente que as anteriores no limite, pode ter causado este comportamento errático dos \hat{p}^4 , principalmente para N e T baixo e maiores números de fatores.

5.3 Estimando modelo Giglio e Xiu para as carteiras do Nefin

Os resultados da seção anterior nos auxiliou na escolha do estimador \hat{p}_1^2 . Após essa escolha o primeiro passo foi calcular o PCA da matriz $(NT)^{-1} \bar{R}' \bar{R}$. Com os autovalores obtidos pelo PCA, calculamos o estimador \hat{p}_1^2 . Para as nossas carteiras $\hat{p}_1^2 = 2$, definido usando o critério discutido na seção anterior, isto é, dois fatores latentes omitidos influenciam a nossa estimação.

Com \hat{p}_1^2 , estimamos $\hat{\beta}$ e \hat{V} . Ambos são utilizados para obter o prêmio de risco $\hat{\gamma}$.

Figura 10 – Doze primeiros autovalores obtidos pelo PCA



O prêmio de risco $\hat{\gamma}$ foi obtido para cada uma das doze carteiras e comparado com os valores originais.

Podemos observar na Figura 11 que para as carteiras classificadas pelo book-to-market os histogramas estimados possuem uma distribuição mais concentrada do que os originais. Já quando olhamos o prêmio de risco acumulado, observamos que, apesar deles

terem o mesmo movimento, o prêmio de risco acumulado estimado se distancia da série original.

Para as carteiras 1 e 2 classificadas pela iliquidez, os resultados mostrados na Figura 11 – Comparação dos Prêmios de riscos para carteira classificadas pelo book-to-market

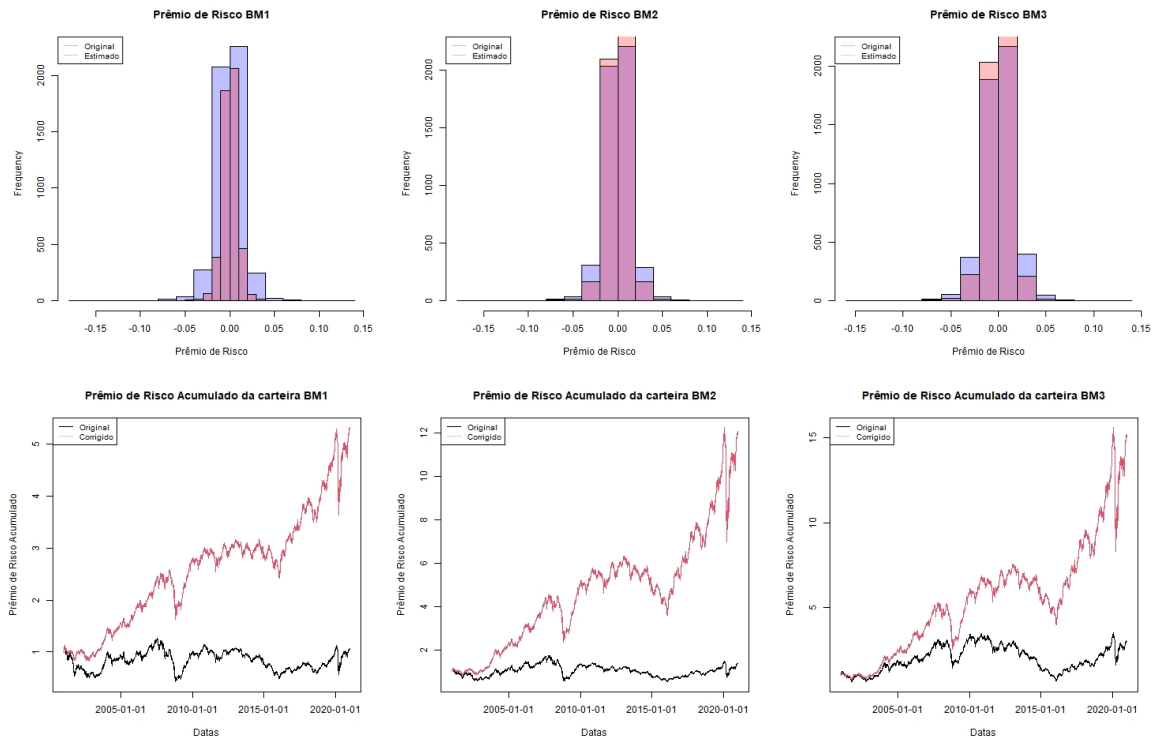


Figura 12 são semelhantes aos que obtivemos com as carteiras classificadas pelo book-to-market. Porém, para a carteira 3, os prêmios de risco estimados estavam muito concentrados no zero, causando uma diferença muito grande entre a distribuição do histograma dos prêmios de risco originais com a distribuição do histograma dos prêmios de risco estimados.

Tanto para as carteiras classificadas pelo momentum quanto para as carteiras classificadas pelo tamanho observamos nas Figuras 13 e 14 que os histogramas estimados possuem uma distribuição mais concentrada do que os originais. Já quando olhamos o prêmio de risco acumulado, podemos observar que, apesar deles terem o mesmo movimento, o prêmio de risco acumulado estimado se distancia da série original.

5.4 Comparação dos três modelos

A comparação do apreçamento dos modelos foi realizada por meio dos resíduos. Nossa ideia foi para cada um dos três modelos calcular a série de resíduos definida por

$$\hat{\epsilon}_{it} = R_{it} - \hat{R}_{it} \quad (5.6)$$

Figura 12 – Comparação dos Prêmios de riscos para carteira classificadas pela Iliquidez

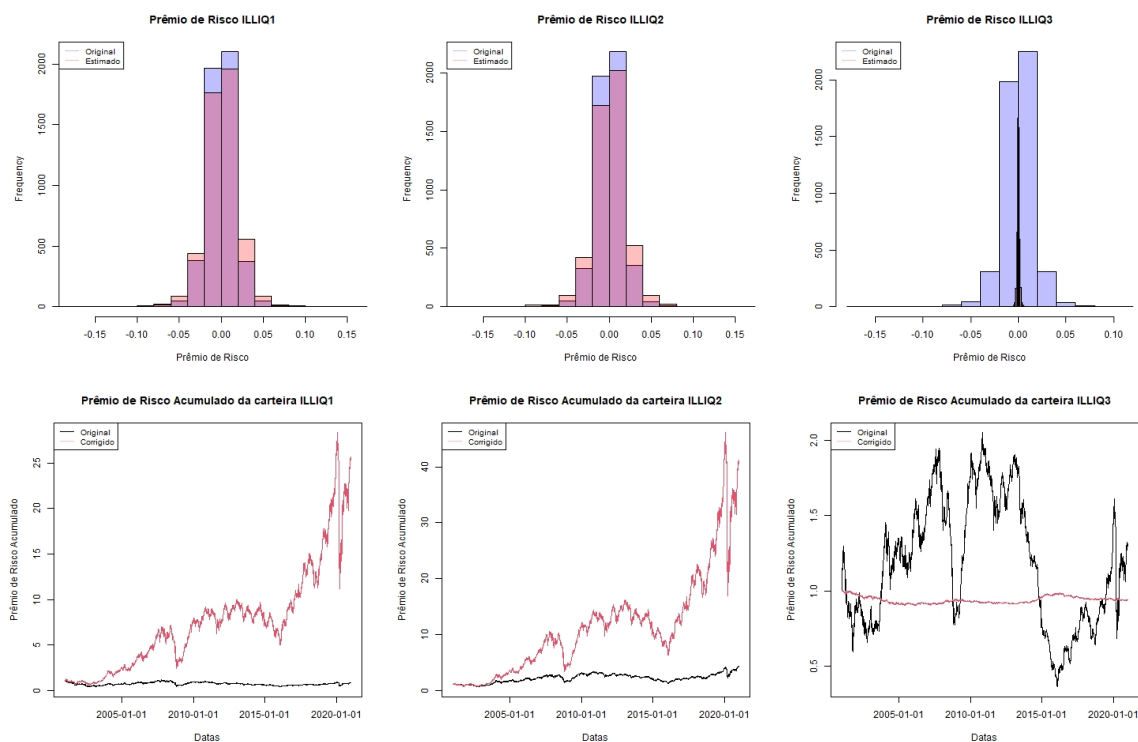
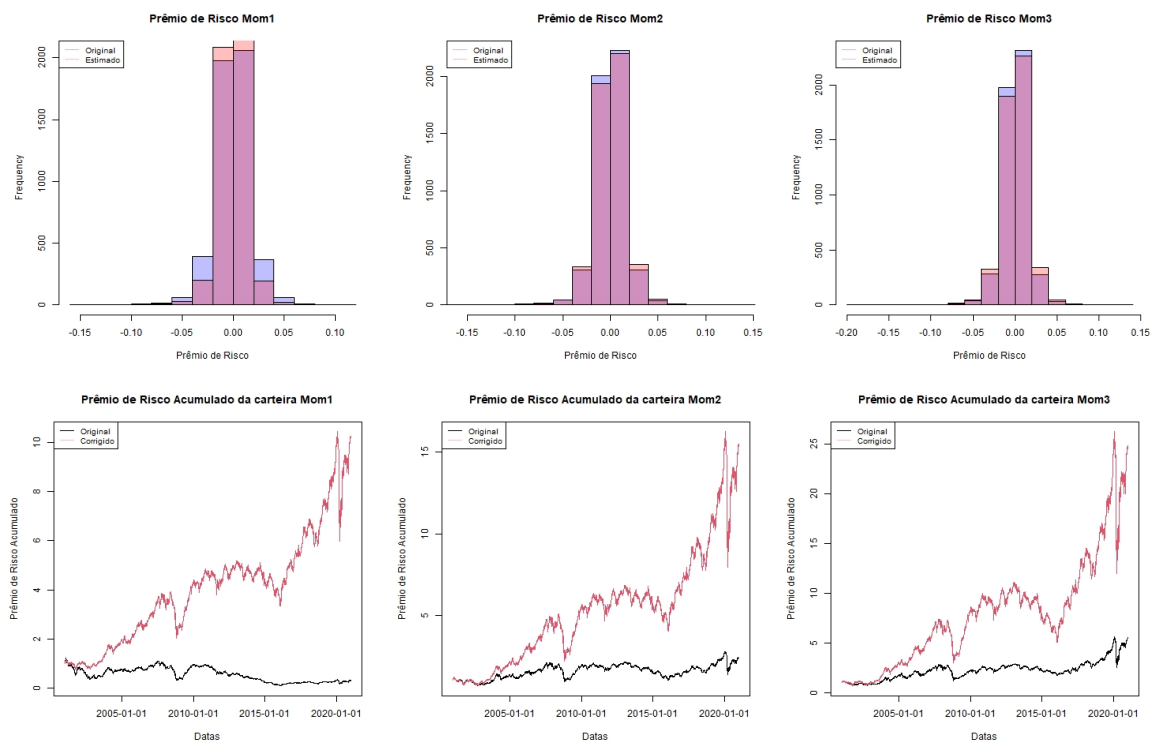


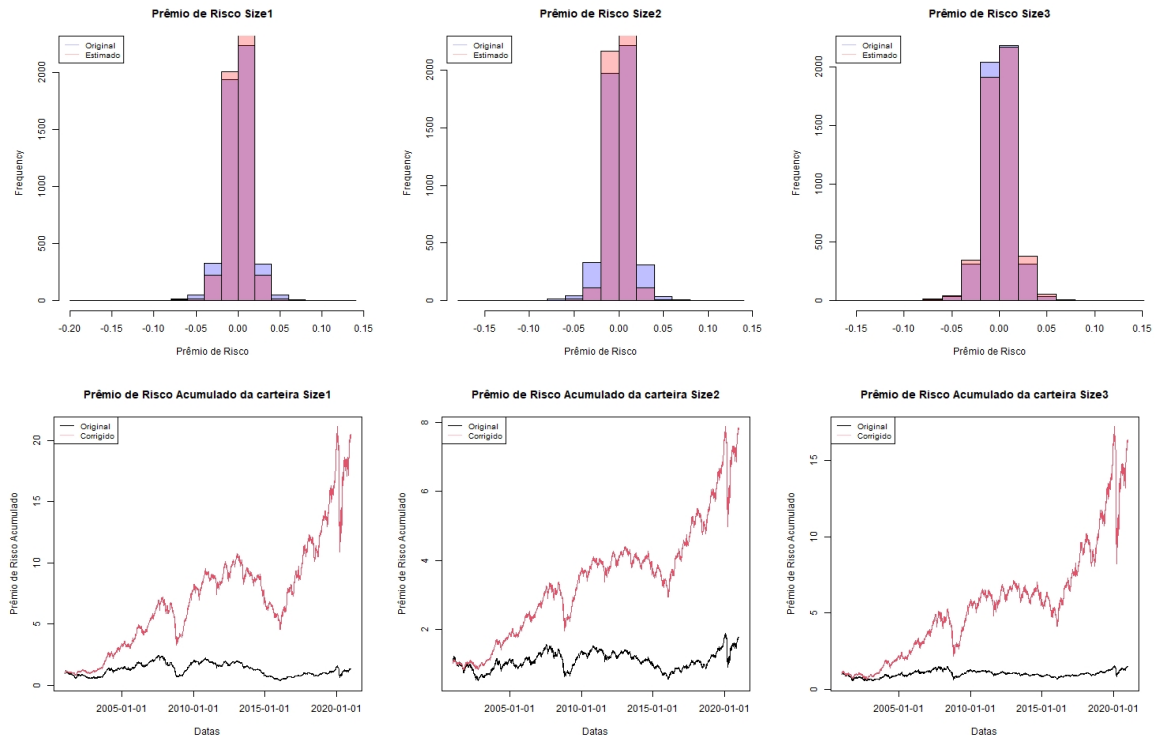
Figura 13 – Comparação dos Prêmios de riscos para carteira classificadas pelo Momentum



em que R_{it} é o retorno da carteira i para o período t e \hat{R}_{it} é o retorno estimado da carteira i para o período t .

As séries de resíduos foram utilizadas para calcular as seguintes métricas:

Figura 14 – Comparação dos Prêmios de riscos para carteira classificadas pelo Tamanho



- Média:

$$\bar{e}_i = \frac{1}{N} \sum_{t=1}^T \hat{e}_{it} \tag{5.7}$$

- Desvio Padrão:

$$\sigma_i = \left(\sum_{t=1}^T \frac{(\hat{e}_{it} - \bar{e}_i)^2}{N} \right)^{1/2} \tag{5.8}$$

- Erro Quadrático Médio (EQM):

$$EQM_i = \frac{1}{N} \sum_{t=1}^T \hat{e}_{it}^2 \tag{5.9}$$

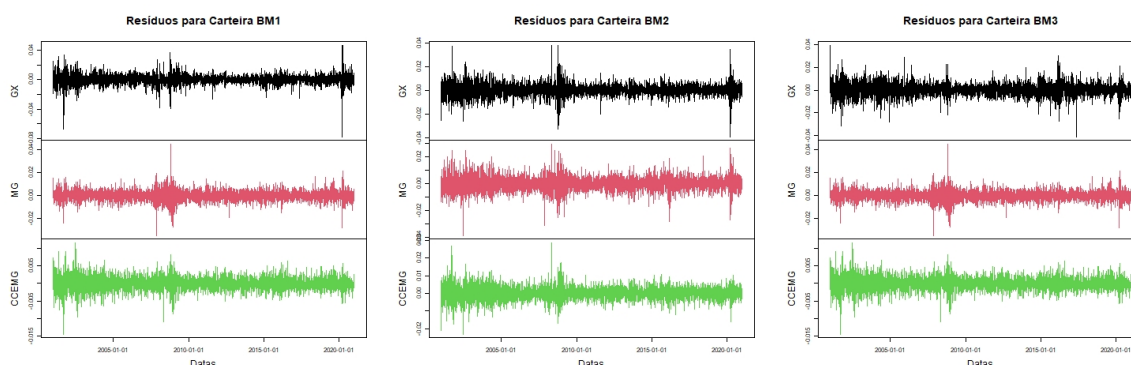
em que, $N = 12$, isto é, o número de carteiras e $T = 4950$ que é o número de períodos.

Na Tabela 7 apresentamos os resultados obtidos para as carteiras classificadas pelo book-to-market. Notamos que em todos os casos os estimadores MG e CCE possuem média dos resíduos zero, como esperado. Logo, podemos concluir que o melhor deles é aquele que possui um menor erro quadrático médio, que é o CCE. Já o estimador GX possui média diferente de zero, o que contraria a hipótese (3) do modelo do Giglio e Xiu. De qualquer forma, ele apresenta os maiores valores de EQM para pelo menos duas das três carteiras classificadas pelo book-to-market.

Tabela 7 – Resíduos para Carteiras BM do NEFIN

Carteira	Modelo	Mínimo	Máximo	Média	Desvio Padrão	EQM
BM1	GX	-0.0778	0.0472	0.0002	0.0068	0.000047
BM1	MG	-0.0346	0.0448	0.0000	0.0044	0.000019
BM1	CCE	-0.0145	0.0116	0.0000	0.0019	0.000004
BM2	GX	-0.0392	0.0381	0.0001	0.0051	0.000026
BM2	MG	-0.0385	0.0294	0.0000	0.0053	0.000028
BM2	CCE	-0.0232	0.0286	0.0000	0.0037	0.000014
BM3	GX	-0.0410	0.0396	0.0002	0.0055	0.000030
BM3	MG	-0.0346	0.0448	0.0000	0.0044	0.000019
BM3	CCE	-0.0145	0.0116	0.0000	0.0019	0.000004

Figura 15 – Gráfico dos Resíduos das Carteiras Classificadas pelo Book-to-Market



Na Tabela 8 apresentamos os resultados obtidos para as carteiras classificadas pela iliquidez. Notamos que em todos os casos os estimadores MG e CCE possuem média dos resíduos zero, como esperado. Logo, podemos concluir que o melhor deles é aquele que possui um menor erro quadrático médio, que é o CCE. Já o estimador GX possui média diferente de zero, o que contraria a hipótese (3) do modelo do Giglio e Xiu. De qualquer forma, ele apresenta os maiores valores de EQM para pelo menos duas das três carteiras classificadas pela iliquidez.

Figura 16 – Gráfico dos Resíduos das Carteiras Classificadas pela Iliquidez

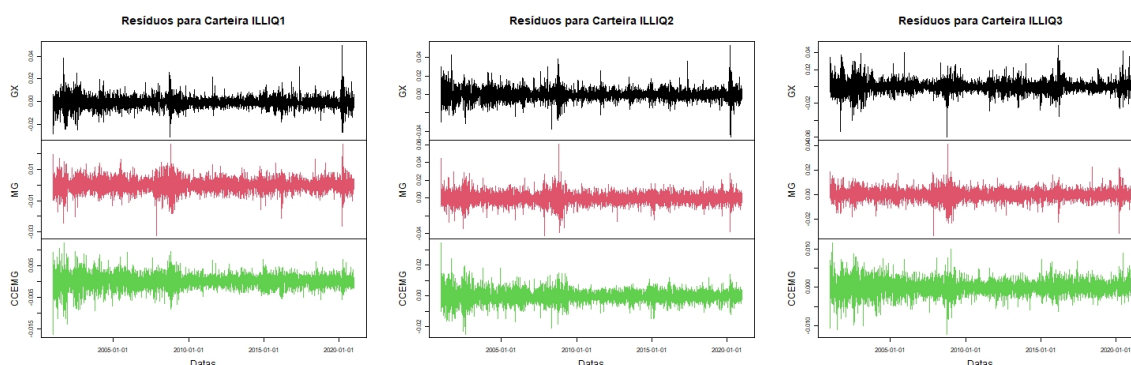


Tabela 8 – Resíduos para Carteiras ILLIQ do NEFIN

Carteira	Modelo	Mínimo	Máximo	Média	Desvio Padrão	EQM
ILLIQ1	GX	-0.0305	0.0496	-0.0003	0.0051	0.000026
ILLIQ1	MG	-0.0322	0.0261	0.0000	0.0042	0.000018
ILLIQ1	CCE	-0.0168	0.0122	0.0000	0.0022	0.000005
ILLIQ2	GX	-0.0457	0.0533	-0.0001	0.0064	0.000041
ILLIQ2	MG	-0.0420	0.0609	0.0000	0.0064	0.000042
ILLIQ2	CCE	-0.0248	0.0346	0.0000	0.0041	0.000017
ILLIQ3	GX	-0.1681	0.1139	0.0007	0.0156	0.000244
ILLIQ3	MG	-0.0322	0.0261	0.0000	0.0042	0.000018
ILLIQ3	CCE	-0.0168	0.0122	0.0000	0.0022	0.000005

Observamos os resultados apresentados na Tabela 9 para a carteira 1 classificada pelo momentum, observamos que o estimador GX possui média diferente de zero, o que contraria a hipótese dos modelos. Já os estimadores MG e CCE, possuem média zero, entretanto o estimador CCE é o que possui menor desvio padrão e EQM e, portanto, o que apresenta melhores resultados pra essa carteira. Para a carteira 2 classificada pelo mometum, observamos que o estimador MG possui média diferente de zero, o que contraria as hipóteses do modelo. Já os estimadores CCE e GX possuem média zero, porém o estimador CCE continua apresentando melhores resultados tendo menor desvio padrão e EQM. Os resultados da carteira 3 classificada pelo momentum é análogo aos resultados da carteira 1 classificada pelo momentum.

Figura 17 – Gráfico dos Resíduos das Carteiras Classificadas pelo Momentum

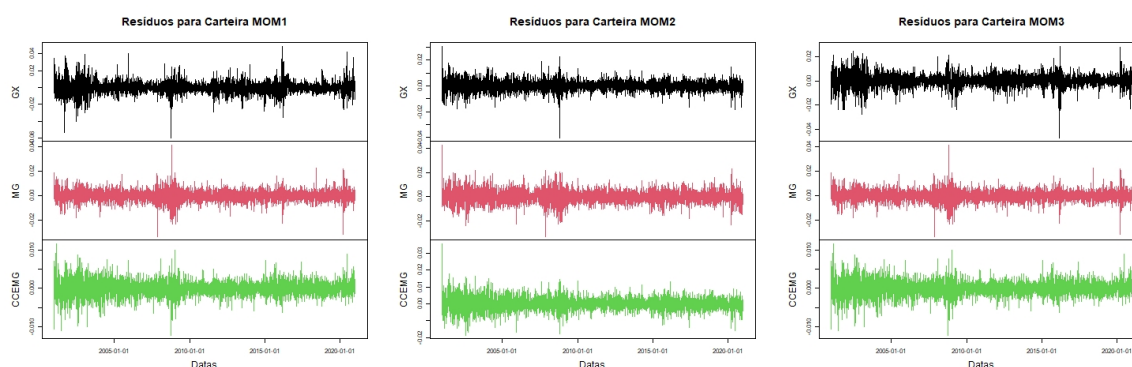


Tabela 9 – Resíduos para Carteiras MOM do NEFIN

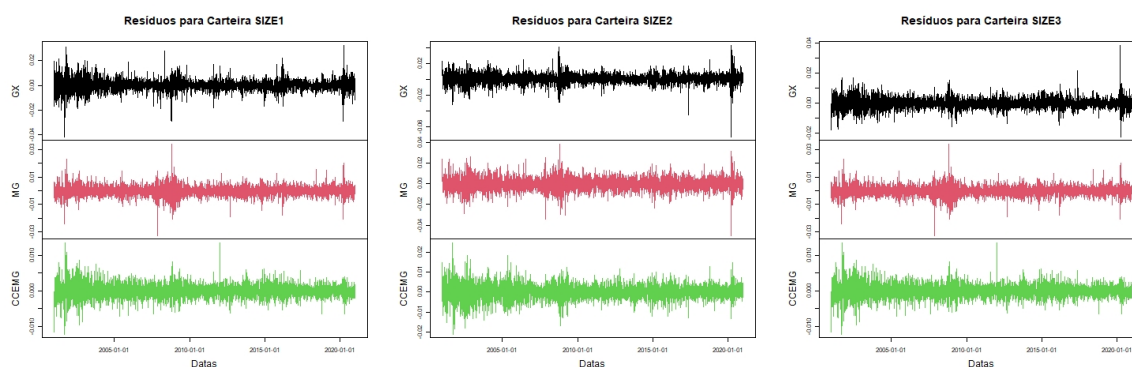
Carteira	Modelo	Mínimo	Máximo	Média	Desvio Padrão	EQM
MOM1	GX	-0.0597	0.0491	-0.0002	0.0073	0.000053
MOM1	MG	-0.0332	0.0412	0.0000	0.0045	0.000021
MOM1	CCE	-0.0123	0.0117	0.0000	0.0020	0.000004
MOM2	GX	-0.0330	0.0427	0.0000	0.0053	0.000028
MOM2	MG	-0.0400	0.0308	0.0001	0.0040	0.000016
MOM2	CCE	-0.0186	0.0355	0.0000	0.0037	0.000014
MOM3	GX	-0.0471	0.0287	0.0001	0.0053	0.000028
MOM3	MG	-0.0332	0.0412	0.0000	0.0045	0.000021
MOM3	CCE	-0.0123	0.0117	0.0000	0.0020	0.000004

Na Tabela 10 apresentamos os resultados obtidos para as carteiras classificadas pelo tamanho. Notamos que em todos os casos os estimadores MG e CCE possuem média dos resíduos zero, como esperado. Logo, podemos concluir que o melhor deles é aquele que possui um menor erro quadrático médio, que é o CCE. Já o estimador GX possui média diferente de zero, o que contraria a hipótese (3) do modelo do Giglio e Xiu. De qualquer forma, ele apresenta os maiores valores de EQM para pelo menos duas das três carteiras classificadas pelo tamanho.

Tabela 10 – Resíduos para Carteiras SIZE do NEFIN

Carteira	Modelo	Mínimo	Máximo	Média	Desvio Padrão	EQM
SIZE1	GX	-0.0416	0.0325	0.0000	0.0047	0.000023
SIZE1	MG	-0.0327	0.0342	0.0000	0.0038	0.000015
SIZE1	CCE	-0.0122	0.0135	0.0000	0.0019	0.000004
SIZE2	GX	-0.0726	0.0429	0.0002	0.0065	0.000042
SIZE2	MG	-0.0501	0.0388	0.0000	0.0062	0.000039
SIZE2	CCE	-0.0212	0.0247	0.0000	0.0037	0.000014
SIZE3	GX	-0.0225	0.0387	-0.0001	0.0036	0.000013
SIZE3	MG	-0.0327	0.0342	0.0000	0.0038	0.000015
SIZE3	CCE	-0.0122	0.0135	0.0000	0.0019	0.000004

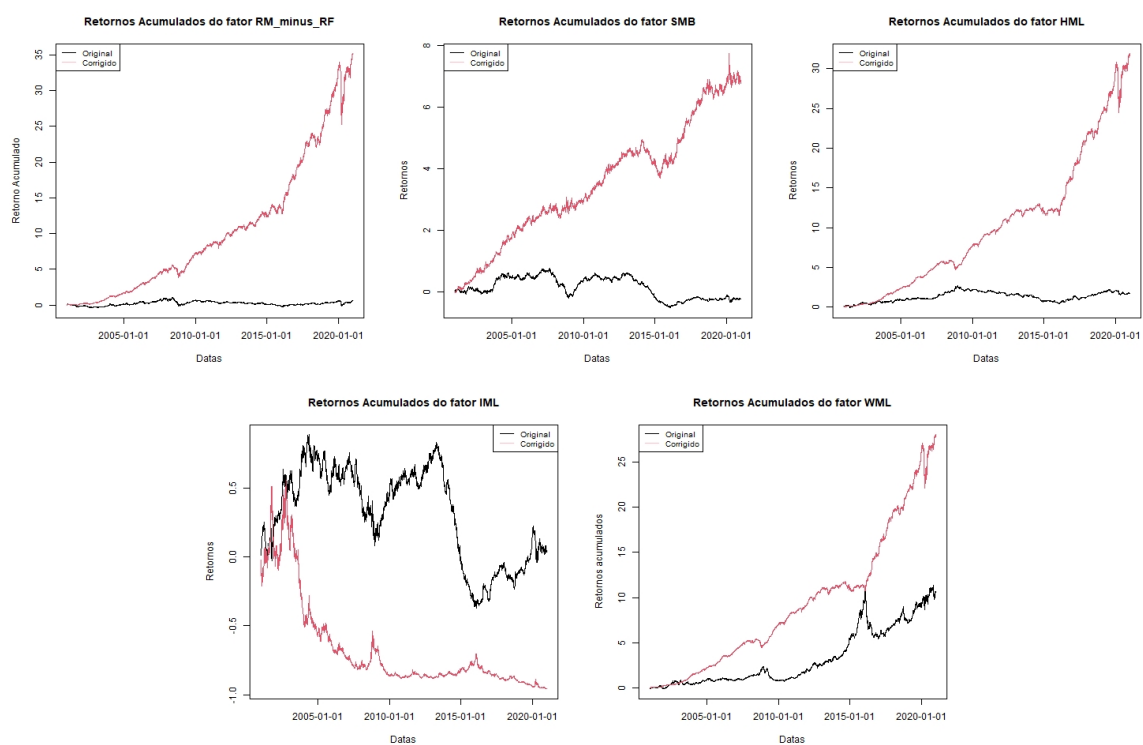
Figura 18 – Gráfico dos Resíduos das Carteiras Classificadas pelo Tamanho



5.5 Fatores Corrigidos

Um dos objetivos do estimador do Giglio e Xiu é recuperar o prêmio de risco dos fatores observados corrigido para possíveis erros de medida e fatores omitidos. Abaixo temos os gráficos da comparação dos retornos acumulados desses fatores:

Figura 19 – Gráfico dos Prêmios de Risco Acumulado dos Fatores



6 Conclusão

Neste trabalho fizemos dois tipos de estudos do modelo de cinco fatores do Fama-French para dados brasileiros: Primeiro estudamos duas formas de estimá-lo. Uma utilizando o estimador Mean Group (MG), e outra utilizando o estimador de Efeitos Comuns Correlacionados (CCE). O objetivo é comparar os dois estimadores e assim verificar a possível omissão de fatores; Segunda forma que estudamos o modelo é comparando as previsões dos três modelos, MG, CCE e o modelo do Giglio e Xiu (2021). Ou seja, utilizamos os modelos para calcular os retornos estimados e comparamos os resíduos gerados pelos modelos. Além dos estudos relacionados ao modelo de cinco fatores do Fama-French, também estudamos os novos estimadores propostos pelo Giglio e Xiu, o estimador de número de fatores e o estimador em três etapas.

Na primeira parte dos estudos, os resultados indicam que os coeficientes estimados para o modelo (3.1) pelo estimador MG é diferente dos obtidos pelo estimador CCE, o que nos indica a possibilidade de omissão de fatores no modelo (3.1).

O estimador para número de fatores proposto por Giglio e Xiu foi testado para quatro funções penalidades. Todas as funções penalidade convergem para zero quando N e T vão para infinito, e portanto para todas as funções penalidade temos que $Prob(p \hat{\geq} p \rightarrow 1)$. Contudo, não conseguimos encontrar uma função penalidade que satisfaça a segunda condição e estime bem os fatores nas simulações. Mesmo não podendo garantir a convergência em probabilidade para o verdadeiro valor, os estimador do Giglio e Xiu apresentou bons resultados para 3 funções penalidades, especialmente para simulações com 1 e 3 fatores.

Com relação ao estimador em três etapas proposto pelo Giglio e Xiu, estimamos o prêmio de risco das carteiras do NEFIN na segunda etapa e na terceira etapa obtivemos o prêmio de risco dos fatores observados corrigidos para possível omissão de fatores e erro de medida. pode-se notar que, apesar da série estimada na maioria dos casos possuir os mesmos movimentos, ela se distancia da série original significativamente no último ano da amostra. Porém, para a carteira 3 classificada pela iliquidez e para o fator construído baseado na iliquidez notamos uma característica inversa da série estimada, quando comparada com as outras.

Após esses estudos comparamos os resíduos gerados pelo modelo do Fama-French estimado pelo estimado Mean-Group, estimado pelo estimador CCE e estimado pelo modelo proposto por Giglio e Xiu. Notamos que o estimador em três etapas não obteve os resultados que esperávamos. Nossas hipóteses é que isto ocorreu por conta do baixo número de carteiras e por não termos conseguido encontrar uma função penalidade que garantisse a convergência em probabilidade. Portanto, concluímos que o Estimador CCE é o que apresenta melhor estimação para os retornos das carteiras do NEFIN.

Bibliografia

- [1] ACHARYA, V. V.; PEDERSEN, L. H. Asset pricing with liquidity risk. *Journal of Financial Markets*, 77(2), p. 31-56, 2002.
- [2] BAI, J.; NG, S. Determining the Number of Factors in Approximate Factors Models. *Econometrica*, 70(1), p. 191-221, 2002.
- [3] BAI, J. Inferential Theory for Factors Models of Large Dimensions. *Econometrica*, 71(1), p. 135-171, 2003.
- [4] BLACK, F. Capital Market Equilibrium with Restricted Borrowing. *Journal of Business*, 45(3), p. 444-455, 1972.
- [5] BLACK, F.; JENSEN, M. C.; SCHOLES, M. The Capital Asset Pricing Model: Some Empirical Tests. *Studies in the Theory of Capital Markets* Praeger Publishers Inc, 1972.
- [6] COCHRANE. J. H., Presidential Address: Discount Rates. *Journal of Finance* 66, 1047-110, 2011.
- [7] FAMA, E. F.,. MACBETH, J. D. Risk, Return, and Equilibrium: Empirical Tests. *Journal of Political Economy*, 81(3), p. 607-636, 1973.
- [8] FAMA, E. F.; FRENCH, K. R. Dividend Yields and Expected Stocks Returns. *Journal of Financial Economics*, 22(1), p. 3-25, 1988.
- [9] FAMA, E. F.; FRENCH, K. R. The Cross-Section of Expected Stock Returns. *The Journal of Finance*, 47(2), p. 427-465, 1992.
- [10] FAMA, E. F.; FRENCH, K. R. Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*, 33(1), p. 3-56, 1993.
- [11] FAMA, E. F.; FRENCH, K. R. Size and Book-to-Market Factors in Earnings and Returns. *The Journal of Finance*, 50(1), p. 131-155, 1995.
- [12] FAMA, E. F.; FRENCH, K. R. A Five-Factors Asset Pricing Model. *The Journal of Financial Economics*, 116(1), p. 1-22, 2015.
- [13] FAMA, E. F.; FRENCH, K. R. International Tests of a Five-Factor Assets Pricing Model. *The Journal of Financial Economics*, 123(3), p. 441-463, 2017.
- [14] GIGLIO, S.; XIU, D. Asset Pricing with Omitted Factor. *Journal of Political Economy*, Accepted February, 2021.
- [15] HARVEY, C. R.; LIU, Y.; ZHU, H. ...And The Cross-Section of Expected Returns, *Review of Financial Studies*, 29(1), p. 5-68, 2015.

-
- [16] HOU, K.; XUE, C.; ZHANG, L., Replicating Anomalies: An investment Approach. *Review of Financial Studies*, 28, p. 650-705, 2017.
- [17] LINTNER, J. The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *Review of Economics and Statistics*, 47(1), p. 13–37, 1965.
- [18] MARKOWITZ, H. Portfolio Selection. *The Journal of Finance*, 7(1), p. 77-91, 1952.
- [19] MARKOWITZ, H. Portfolio Selection: Efficient Diversification of Investments. *Yale University Press*, 1959.
- [20] MCLEAN, R. D.; PONTIFF, J. Does Academic Research Destroy Stock Return Predictability? *Journal of Finance*, 71(1), 5-32, 2016.
- [21] MORGENSTERN, O.; VON NEUMANN, J. Theory of Games and Economic Behavior. Princeton University Press, third edition, 1953.
- [22] NOVY-MARX, R. The Other Side of Value: The Gross Profitability Premium, *Journal of Financial Economics* 108(1), p. 1-28, 2013
- [23] PESARAN, M. H.; SMITH, R. Estimating Long-Run Relationships from Dynamic Heterogeneous Panels. *Journal of Econometrics*, 68(1), p. 79-113, 1995.
- [24] PESARAN, M. H. Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure. *Econometrica*, 74(4), p. 967-1012, 2006.
- [25] PESARAN, M. H., Time Series and Panel Data Econometrics. Oxford University Press, first edition, 2015.
- [26] PETERSEN, M. A. Estimating standard errors in finance panel data sets: Comparing approaches. *The Review of Financial Studies*, 22(1) p. 435-480, 209.
- [27] SHARPE, W. F. Capital Asset Prices: a Theory of Market Equilibrium under Conditions of Risk. *Journal of Finance*, 19(3), p. 425-442, 1964.
- [28] TITMAN, S.; WEI, K.; XIE, F. Capital Investments and Stocks Returns, *Journal of Financial and Quantitative Analysis* 39(4), 677-700, 2004.