Robô M A R K

Originação do nome:

O nome **MARK** se origina nos modelos utilizados na construção do portfólio. Sendo eles: (1) *Hidden **MARK**ov Model*; (2) modelo **A**uto-**R**regressivo; (3) *Hierarchical **R**is**K** Parity*.

Explicação Lógica:

Três são os pontos principais quando pensa-se na construção de um portfólio global: (1) a alocação percentual em cada geografia; (2) quais ativos serão escolhidos; (3) qual o peso de cada ativo. Assim, com a união de três modelos, a estratégia criada responde a todos os pontos: (1) *Hidden Markov Model*; (2) modelo Auto-Regressivo e, por fim, (3) modelo *Hierarchical Risk Parity*.

Tipo de Estratégia: Portfólio
Classe de Ativos: Equities
Universo: Brasil(Ibov Index), Estados Unidos(SPX Index), Reino Unido(UKX Index), França(CAC Index) e Alemanha(DAX Index)
Média de Trades Mensais: 1500
Holding Period: 1 mês
Qual Plataforma Testou: R
Benchmark Estratégia: ACWI em reais

An old science joke says: "Theory is when you know everything but nothing works. Practice is when everything works but no one knows why. In our lab, theory and practice are combined: Nothing works and no one knows why." Finance is better. The consumption CAPM theory is well developed, but it doesn't work. Anomaly strategies work, but no one knows why.

Zhang (2020)

1 Introdução

O presente trabalho se propõe a desenvolver uma estratégia de investimento fundamentada na fronteira do conhecimento econômico-financeiro e que, ao mesmo tempo, apresente resultados robustos em *backtests*. Ademais, a intenção é de que haja o mínimo de interferência humana possível. Assim, é imperativo que elenquemos quais as principais questões que investidores se deparam quando desejam construir um portfólio diversificado. Com essa "lista", seremos capazes de apresentar nossa estratégia, mostrando como cada parte se encaixa no todo.

1. Quais classes de ativos o investidor deve incluir no portfólio?
2. Em quais mercados deverá haver exposição?
3. Deve ser realizado algum tipo de *hedge* cambial?
4. A quais ativos o investidor deve ter exposição?
5. Qual deve ser o peso de cada um desses ativos no portfólio?

Na tentativa de responder às perguntas acima, assumimos o ponto de vista de um investidor brasileiro hipotético que decide alocar capital em ações brasileiras e de grandes economias mundiais. Além disso, decide não proteger o portfólio contra as variações cambiais, dado que se sente confortável com exposição a moedas consideradas mais fortes que o Real. A partir desse ponto, as perguntas serão respondidas com base em modelos já consolidados, com algumas adições e alterações que julgarmos pertinentes.

Primeiramente, em relação à definição dos pesos de cada mercado dentro do portfólio, será utilizado o modelo *Hidden Markov Model*, no qual assumimos que nosso sistema é governado por um processo de Markov com quatro estados ocultos (Lewin and Campani (2020)), sendo que a variável observável é o índice de volatilidade daquele mercado e a variável não observável é o índice de mercado em si (Iquiapaza (2021)).

Em seguida, para definição de quais ações farão parte do nosso portfólio, utilizamos um modelo de fator de risco. Nele, ordenamos os ativos com base no *drift* de um modelo AR(1). Acreditamos que essa nova "espécie" do "gênero" *momentum* aumenta a compreensão da origem desse prêmio de risco e ajuda a responder uma das grandes perguntas em aberto na literatura financeira: se *momentum* possui uma explicação baseada no risco, qual é ela?; e, caso possua apenas uma explicação baseada no comportamento dos investidores, por que o prêmio não é arbitrado pelos participantes do mercado cientes dessa ineficiência?

Agora, nos resta definir qual será o critério para determinar o peso de cada ação dentro do portfólio. Para isso, utilizaremos o modelo *Hierarchical Risk Parity*(HRP) (Lopez De Prado (2016)), com uma otimização em alguns parâmetros do modelo. O HRP, segundo o autor, "aplica matemática moderna (teoria dos grafos e técnicas de *machine learning*) para construir um portfólio diversificado, baseado nas informações contidas na matriz de covariância". Ademais, o autor ressalta que algumas das vantagens desse modelo, frente a alternativas como o modelo de Mínima Variância (Markowitz (1952)), são que ele não requer a inversão da matriz de covariâncias e que ele tende a entregar portfólios menos arriscados em situações *out of sample*.

O relatório é estruturado da seguinte forma: (i) apresentamos os dados e metodologia e, também, nos aprofundamos em como cada modelo será utilizado em nossa estratégia; (ii) demonstramos como ela não se apoia em nenhum viés de *backtest* para entregar um bom resultado; (iii) mostramos os resultados do *backtest*; (iv) por fim, tecemos algumas conclusões e expomos avanços futuros possíveis.

2 Dados e Metodologia

Para o modelo *Hidden Markov Model*, utilizamos o retorno diário dos cinco principais índices dos países em análise (SP500, FTSE100, CAC, DAX e IBOV) e a variação diária do índice de volatilidade referente a cada um deles (VIX, IVIUK, VCAC, V1X e IVol-Br). Todos os dados, com exceção do IVol-Br foram extraídos da Bloomberg e compreendem o período de 2009-12-31 a 2020-12-31. O IVol-Br (Astorino et al. (2017)) foi extraído da Nefin e compreende o período de 2011-08-01 a 2020-12-31. Além disso, utilizamos a Biblioteca do Professor Kenneth French e a Nefin para os retornos dos 4 fatores de risco de Cahart (Carhart (1997)) para os países desenvolvidos e para o Brasil. Além disso, extraímos os dados da composição dos índices e dos preços dos ativos também da Bloomberg.

A primeira restrição de data para início do *backtest* fica evidente: o índice IVol-Br começa apenas em 2011-08-01. Além disso, acreditamos que pelo menos 100 observações são necessárias para treinar nosso *Hidden Markov Model*. Como os dados estão na frequência semanal, iniciamos o nosso portfólio em 01-01-2014 e finalizamos o *backtest* em 31-12-2020.

Ademais, criamos um *benchmark* alternativo, que representaria a melhor comparação possível com a nossa estratégia. Esse índice é um portfólio *equal weighted*, rebalanceado semanalmente, que fica comprado nos índices de ações dos países abordados em nossa análise. Como ele é um índice desenvolvido especialmente para comparação com nossa estratégia de investimento, o chamamos de "Tailor Made Index".

2.1 *Hidden Markov Model*

Quando trabalhamos com Cadeias de Markov, temos, tipicamente, dois objetos: a matriz de transição (T) e o vetor contendo a probabilidade de estarmos, no tempo t, em cada um dos n estados (P). Com base nisso e no vetor contendo a média do retorno em cada um dos n estados (M), somos capazes de calcular qual o retorno esperado para o período seguinte (ER_{t+1}). Por exemplo:

$$T = \begin{bmatrix} 0.7 & 0.3 \\ 0.75 & 0.25 \end{bmatrix} P = \begin{bmatrix} 0.9 & 0.1 \end{bmatrix} M = \begin{bmatrix} -0.03 & 0.02 \end{bmatrix}$$

$$ER_{t+1} = P \times T \times M' = -0.01525$$

Tendo o retorno esperado para $t+1$, gostaríamos de ter uma exposição maior àquele mercado se o retorno esperado for alto e uma exposição menor se o retorno esperado for baixo. Assim, analisamos em que ponto da distribuição de retornos históricos o retorno esperado estaria e nos posicionamos de acordo. Matematicamente, o peso do mercado m (w_m) será:

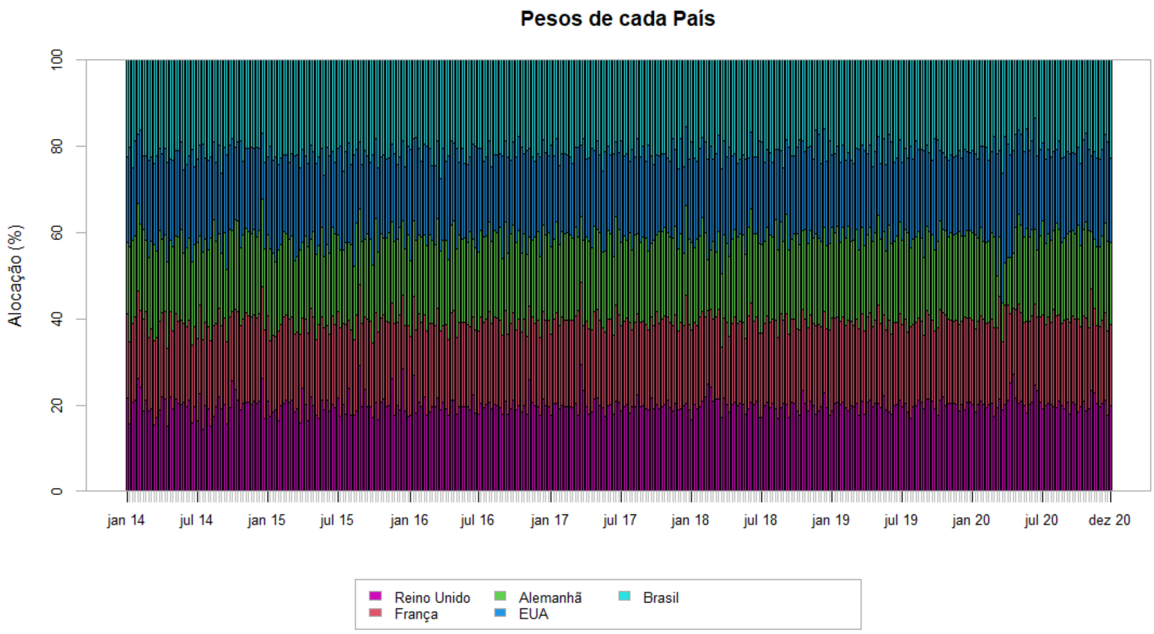
$$w_m = \frac{\sum_{n=1}^{N} I_n}{N}$$

Sendo N a quantidade de observações que temos na distribuição histórica e I_n:

$$I_n = \begin{cases} 1 & \text{se } ER_{t+1} > R_{t-n} \\ 0 & \text{se } ER_{t+1} \leq R_{t-n} \end{cases}$$

De modo que R_{t-n} é o retorno no tempo $t-n$, com $n = 0, 1, ..., N$.

Após esse processo, chegamos aos pesos apresentados na imagem abaixo. Evidentemente, alcançamos um portfólio bastante balanceado, no qual realizamos mudanças táticas nos pesos com base no retorno esperado para o período seguinte.



2.2 Uma nova forma de ver o fator *momentum*

A teoria financeira pressupõe que os retornos se comportam como uma v.a., iid, com distribuição $N(0, \sigma^2)$. Dessa forma, em uma equação auto regressiva de ordem 1 ($r_t = \mu + \beta r_{t-1} + \epsilon_t$), esperamos que $\mu = \beta = 0$. De fato, podemos considerar $\beta = 0$ - se não pudéssemos, esse seria o parâmetro mais importante para qualquer ativo financeiro.

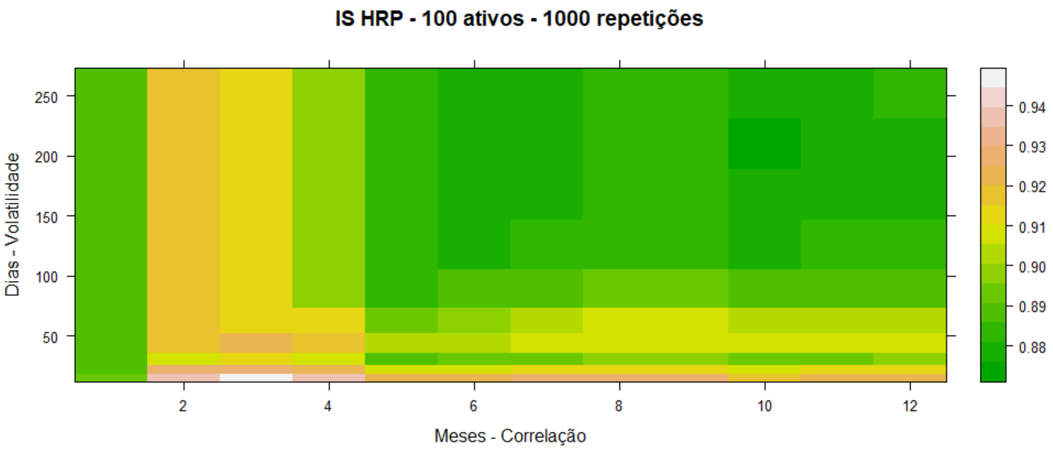
Entretanto, não podemos automaticamente considerar $\mu = 0$. De fato, quando ordenamos os ativos com base no *drift* de um modelo AR(1), podemos constatar que o *drift* médio do primeiro tercil é estatisticamente maior que o *drift* médio do último tercil (*Wilcox Test*) para todos os rebalanceamentos e mercados. Com base nessa constatação e na suposição de estacionaridade fraca de μ, podemos compreender de uma melhor forma a origem do prêmio de *momentum*: ativos com maior *drift* tendem a ter um termo (μ) que lhes confere uma "vantagem" em relação àqueles com um menor *drift* - se ϵ_t é muito negativo, o *drift* atenua esse impacto sob o retorno final; se é muito positivo, o *drift* amplifica esse impacto sob o retorno final. Desse modo, no fundo, a explicação de *momentum* talvez esteja mais relacionada ao Processo Gerador de Dados (PGD) do que explicações baseadas em risco ou vieses dos investidores.

Assim, visando testar essa hipótese, ordenamos os ativos de cada um dos 5 mercados em análise com base no *drift* de um modelo AR(1) - estimado com base em 1 ano de dados - e ficamos comprados naqueles 30% com maior *drift*. Como nesses casos é muito importante a robustez dos resultados, desenvolvemos um aplicativo web no qual o leitor pode testar essa ideia sob diferentes regimes de mercado e variações possíveis nos parâmetros do *backtest*.

2.3 *Hierarquical Risk Parity*

Em essência, utilizamos exatamente o mesmo algorítimo proposto por Lopez De Prado (2016) para encontrar o peso que cada ativo terá no portfólio. Entretanto, como sabemos que o período de dados utilizado para estimar a matriz de covariâncias é vital, realizamos uma otimização nesse parâmetro. Para tal, utilizamos o fato de $Cov(x, y) = Cor(x, y) \times \sqrt{Var(x)Var(y)}$.

Em uma amostra contendendo o retorno diário de mais de 500 ativos de 2009-12-31 a 2020-12-31, selecionamos aleatoriamente 100 ativos e montamos um portfólio com rebalanceamento mensal no qual o peso de cada ativo é definido utilizando HRP, com n meses de estimação para as correlações e m dias de estimação para a volatilidade. Tendo o retorno diário de todos esses portfólios resultantes das combinações de n e m, calculamos o Índice de Sharpe para cada um desses portfólios e analisamos qual região parece conter o maior *IS*. Para garantir a robustez desse resultado, repetimos esse processo 1000 vezes e, ao final, calculamos a média do IS de cada combinação de m e n. O resultado pode ser observado na imagem abaixo:



Como podemos observar, após 1000 repetições, o portfólio que gerou maior Sharpe foi aquele construído a partir da combinação de um período de estimação de 3 meses para as correlações e 15 dias para as volatilidades. Sendo assim, essa será a combinação utilizada para construção do nosso portfólio.

3 Vieses de *Backtest*

Visando garantir que nosso *backtest* seja verossímil, apresentamos alguns vieses de *backtest*, retirados do *white paper* escrito pela Newfound Research (Newfound (2013)), e demonstramos como o nosso processo se ancora nas melhores práticas para garantir a confiabilidade do resultado.

3.1 *Survivorship Bias*

Nosso *backtest*: Para a construção do nosso portfólio, utilizamos a composição mensal de cada índice. Dessa forma, trabalhamos tanto com os ativos que sobreviveram até os dias atuais como com aqueles que não sobreviveram.

3.2 *Look Ahead Bias*

Nosso *backtest*: Nosso modelo utiliza apenas informação de preço e há separação clara entre nosso período de estimação e nosso período de avaliação.

3.3 *Data Snooping Bias*

Nosso *backtest*: Não realizamos nenhum tipo de mineração de dados ou busca por padrões. Pelo contrário: utilizamos apenas métodos e processos já consagrados e discutidos.

3.4 *Selection Bias*

Nosso *backtest*: Não incorremos neste viés, dado que utilizamos uma amostra bastante representativa do mercado de capitais global.

4 Resultados

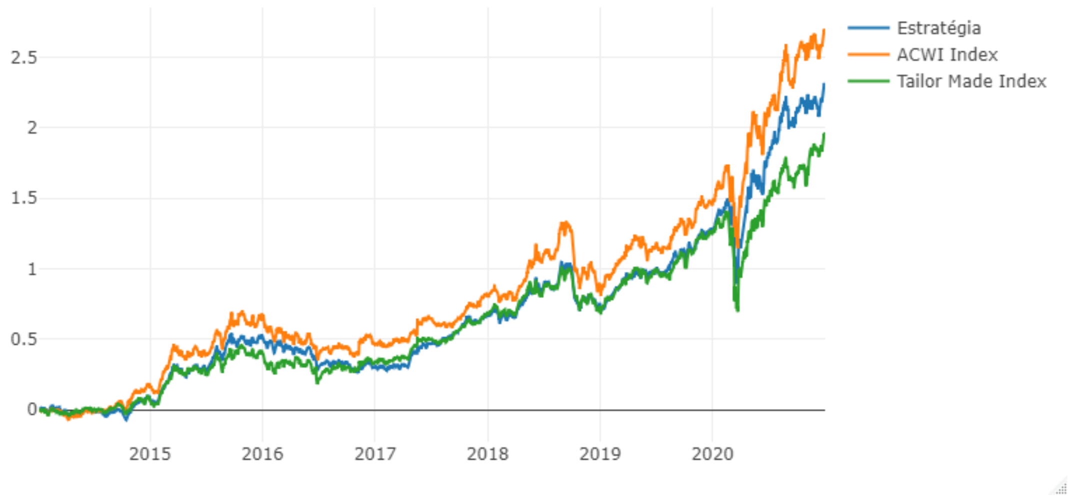
Em todas as geografias, a estratégia é capaz de gerar bons retornos ajustados ao risco, como exposto pelos indicadores *Information Ratio* e Índice M^2. Medidas de risco, como a volatilidade dos retornos, o *Beta* e o *CVaR* também são favoráveis. Portanto, é possível concluir que, de forma individual, por país, temos resultados bastante satisfatórios.

	França	Reino Unido	Brasil	Alemanha	EUA
Alfa	3.47%	5.91%	1.35%	1.58%	2.26%
P-valor Alfa	0.25	0.08	0.79	0.63	0.4
Beta	0.87	0.76	0.76	0.83	0.54
M^2	7.19%	5.91%	13.19%	6.46%	12.3%
Retorno Anual.	6.81%	5.64%	13.04%	5.87%	9.03%
Vol	18.78%	15.33%	24.06%	18.79%	11.88%
Sharpe	0.33	0.28	0.13	0.31	0.58
Treynor	0.08	0.07	0.17	0.07	0.17
CVaR	-3.77%	-2.27%	-6.95%	-3.61%	-0.47%
Inform. Ratio	0.76	0.44	0.22	0.61	0.64

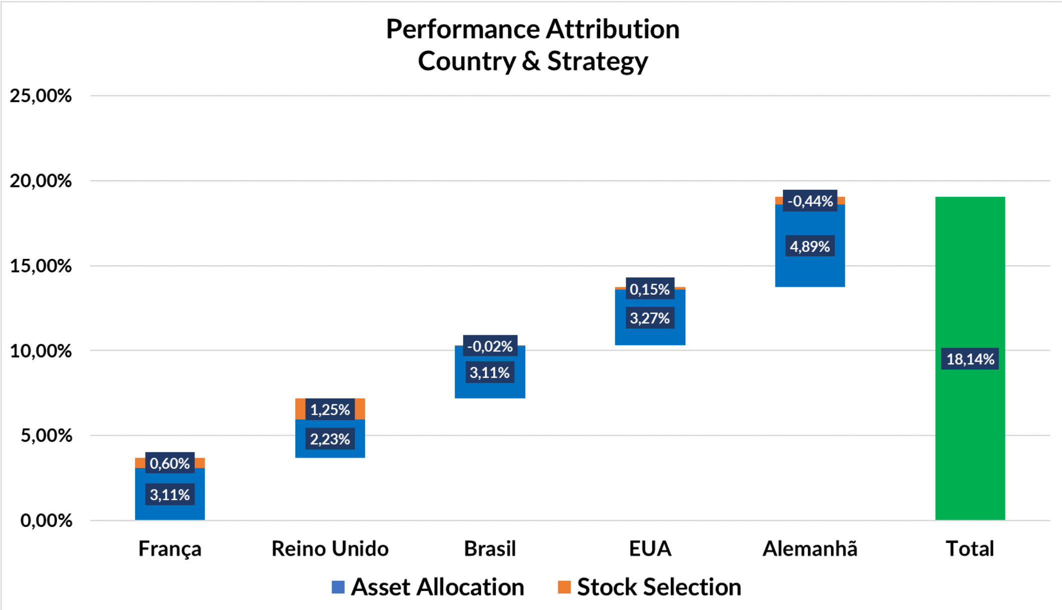
Quando analisamos a estratégia em si, os resultados são igualmente interessantes, com melhora, inclusive, dos indicadores M^2 e *Information Ratio* contra o índice global *ACWI*. A estratégia apresenta também melhores indicadores e rentabilidade absoluta do que o índice *Tailor Made*, explicado anteriormente no relatório. Apesar de um bom desempenho ajustado ao risco, o portfólio não é capaz de superar, em termos absolutos, a rentabilidade do *ACWI*.

	Estratégia	ACWI Index	Tailor Made Index
Alfa	0.64%		
P-valor Alfa	0.83		
Beta	0.76		
M^2	18.45%		
Retorno Anual.	18.14%	19.9%	16.24%
Vol	16.11%	18.32%	17.42%
Sharpe	0.48	0.51	0.34
Treynor	0.24		
CVaR	-3.44%	-2.87%	-3.51%
Inform. Ratio	0.85		

Rentabilidade:



Além disso, dividimos os retornos por país, em termos anualizados, entre *Stock Selection* & *Asset Allocation*, conforme a metodologia descrita em CFA (2018). Constatamos que o portfólio é capaz de gerar retornos positivos em todas os mercados através da seleção de ações, com exceção da Alemanha e Brasil, com contribuições negativas de -0,44% e -0,02%.



O principal fator de risco que afeta os retornos do portfólio é o fator Mercado, sendo estatisticamente significativo em todas as geografias. Quanto aos demais fatores, vemos uma influência relevante do fator *momentum* - esperado, dado que trabalhamos com uma estratégia desse "gênero" - e do fator valor, que possui um coeficiente significativo e negativo para quase todas as localidades. Essa relação negativa entre nossa estratégia e o fator valor pode ser um indicativo que compor um portfólio com esses dois fatores pode melhorar a relação risco retorno geral.

Regressão Long Only - Cahart

	Dependent variable:					
	Estratégia	França	Reino Unido	Brasil	Alemanhã	EUA
	(1)	(2)	(3)	(4)	(5)	(6)
Mercado	0.848 p = 0.000	0.933 p = 0.000	0.936 p = 0.000	0.839 p = 0.000	0.889 p = 0.000	0.623 p = 0.000
Size	0.084 p = 0.002	0.058* p = 0.238	0.247 p = 0.00000	0.182 p = 0.000	0.109 p = 0.036	-0.164 p = 0.000
Value	-0.086 p = 0.001	-0.054* p = 0.238	-0.185 p = 0.00001	-0.081 p = 0.001	-0.097 p = 0.043	0.176 p = 0.000
Momentum	0.172 p = 0.000	0.042* p = 0.127	0.050 p = 0.044	0.367 p = 0.000	0.087 p = 0.003	0.346 p = 0.000
Observations	1,806	1,807	1,807	1,807	1,807	1,807
R^2	0.808	0.615	0.677	0.768	0.589	0.832
Adjusted R^2	0.808	0.614	0.676	0.768	0.588	0.832
Note:	Intercepto estatisticamente igual a 0 — *Não significativo (5%)					

5 Conclusão

Concluindo, acreditamos termos sido capazes de expor o desenvolvimento de uma estratégia de investimento que possui robustez tanto no arcabouço teórico quanto em *backtests*. Elencamos os principais questionamentos que surgem quando um investidor se propõe a montar um portfólio e como, de forma sistemática, nossa estratégia se propõe a equacioná-los. Ademais, listamos os principais vieses que poderíamos estar incorrendo ao testar nosso modelo e como não dependemos de nenhum deles para entregar um bom desempenho.

Também demonstramos que os resultados foram bastante robustos, tanto em termos de retorno absoluto quanto em termos de retorno ajustado ao risco. Portanto, podemos concluir que nossa estratégia está pronta para compor o portfólio dos nossos investidores.

Dado que este trabalho não é a última palavra sobre o tema, outros analistas interessados podem ampliar o escopo deste. Um caminho natural seria adicionar mais países à amostra. Um outro caminho poderia envolver a inclusão de um universo maior de ações, focando em índices que possuam menos restrições para inclusão de ativos. Ademais, o aumento do período da análise também poderia agregar ao mostrar como a estratégia performa sobre outros regimes e condições de mercado. Por fim, poderia ser analisado com maior profundidade como nossa estratégia interage com outros fatores de risco e outras estratégias.

References

E. S. Astorino, F. Chague, B. Giovannetti, and M. E. d. Silva. Variance premium and implied volatility in a low-liquidity option market. *Revista Brasileira de Economia*, 71:3–28, 2017.

M. M. Carhart. On persistence in mutual fund performance. *The Journal of finance*, 52(1):57–82, 1997.

I. CFA. Cfa institute investment foundations® program, 3rd edition, 2018. URL https://www.cfainstitute.org/-/media/documents/support/programs/investment-foundations/19-performance-evaluation.pdf.

R. A. Iquiapaza. Apostila métodos computacionais no mercado de capitais, 2021. URL https://sites.google.com/site/robertiquiapaza/disciplinas/metodos-computacionais-financas?authuser=0.

M. Lewin and C. H. Campani. Gestão de carteiras sob múltiplos regimes: Estratégias que performam acima do mercado. *Revista de Administração Contemporânea*, 24:300–316, 2020.

M. Lopez De Prado. Building diversified portfolios that outperform out of sample. *The Journal of Portfolio Management*, 42(4):59–69, 2016.

H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952. ISSN 00221082, 15406261. URL http://www.jstor.org/stable/2975974.

R. Newfound. Backtesting with integrity, 10 2013. URL https://bit.ly/3q7mTFj.

L. Zhang. Q-factors and investment capm, 06 2020.