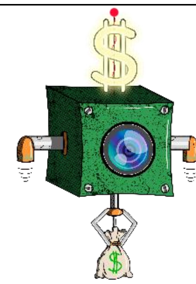
quantamental challenge

	
Nome do Robô	Leda
Explicação do Nome	Homenagem à gestora Leda Braga do Systematica Investments
Explicação da Lógica da Estratégia	VAR (Vector autoregression) é um modelo autorregressivo ideal para trabalhar com ativos cujos retornos estão correlacionados. O modelo foi aplicado a empresas do setor do varejo de roupas e calçados de acordo com critérios pré-estabelecidos (liquidez e correlação com consumo das famílias) e foi diariamente atualizado com uma janela de 484 observações para gerar os coeficientes da influência de cada série na predição do preço do dia seguinte das ações analisadas. A partir do resultado foram determinados pesos em reais para compra e venda de cada ação com base nos retornos estimados para cada ação diariamente.
Tipo de Estratégia	Portfólio
Classe de Ativos	Ações
Universo	Ações de varejo de roupas e calçados
Média de Trades por Mês	170
Holding Period	1 dia
Qual Plataforma Testou a Estratégia	Python
Objetivo da Estratégia	Superar o CDI

Modelo

O modelo de *Vector Autoregression*[1] consiste em um método preditivo utilizado para múltiplas séries temporais que influenciam umas às outras (relações bidirecionais). Esse modelo foi implementado conforme orientado pelo livro *Machine trading: Deploying Computer Algorithms to Conquer the Markets*.

Em resumo, esse modelo prediz o valor futuro de cada uma das séries com base em valores passados de todas as séries envolvidas na análise. Apesar de intuitivamente simples, sua implementação requer uma série de testes estatísticos tanto para escolha de parâmetros quanto para determinar em quais momentos o modelo pode ser capaz de realizar predições.

Após escolher as ações que serão negociadas segundo o modelo, os preços de fechamentos diários de cada ação são processados de modo que criem-se estruturas com os retornos percentuais diários de cada ação. Criadas essas estruturas, é realizado o *Augmented Dickey-Fuller Test* [2] (ADF) com nível de significância de 95% e número de *lags* (defasagens) escolhido a partir do *Akaike Information Criterion* [3] (AIC), esse teste consiste em verificar, para cada uma das séries de retornos, a presença de raiz unitária por meio da análise da regressão (I):

$$\Delta y_t = \beta_1 + \beta_2 t + \delta y_{t-1} + \sum_{i=1}^{m} \alpha_i \Delta y_{t-i} + \varepsilon_t \qquad (I)$$

No ADF, temos que os coeficientes β_1 e β_2 são, respectivamente, o *drift* da série e o coeficiente de tendência enquanto δ é o coeficiente de presença de raiz unitária e m é o número de lags da série. O teste tem como hipótese nula H_0: $\delta = 0$ e, após realizar a regressão, calcula-se a estatística T por meio da equação (II):

$$T = \frac{\hat{\delta}}{se(\hat{\delta})} \quad (II)$$

No cálculo da estatística T, têm-se que $\hat{\delta}$ é um estimador de δ e $se(\hat{\delta})$ é um estimador para o desvio padrão do erro de δ.

O AIC, por sua vez, consiste em uma maneira de aferir a qualidade estatística de um modelo (quão adequado seu modelo está aos dados sem apresentar problemas de *over-fitting*). Esse critério funciona premiando a adequação do modelo aos dados e penalizando a complexidade (o que diminui a probabilidade da ocorrência de *over-fitting*).

Para que faça sentido utilizar o modelo VAP(p) as séries analisadas deve ser cointegradas. Para verificar se esse é o caso realiza-se o $teste\ de\ cointegração\ de\ Johansen$[4] com 95% de significância, que consiste em realizar a autorregressão vetorizada das séries conforme a equação (I) com 1 *lag* e, com base nela, realizam-se dois diferentes testes da razão da verossimilhança dados pelas equações (III) e (IV):

$$J_{trace} = -N \sum_{i=r+1}^{n} \ln\left(1 - \widehat{\lambda_i}\right) \quad (III)$$

$$J_{max} = -N\, ln\left(1 - \widehat{\lambda_{r+1}}\right) \quad (IV)$$

Em que N é o tamanho da amostra e $\widehat{\lambda_\iota}$ é a i-ésima maior correlação canônica entre $\Delta y_t\ e\ Z_{t-1}$ e as hipóteses testadas são $H_0\colon \lambda_i = 0;\ i = r+1, \dots, n\ \ e\ \ H_1\colon \lambda_i = 0;\ i = 1{,}2, \dots, n.$

Por fim, realiza-se o $teste\ de\ causalidade - Granger^5$ para verificar a cointegração das séries por meio da função *grangers_causation_matrix()* da biblioteca *statsmodels* na linguagem python e é feita a regressão (I) com o número de lags obtido pelo *ADF* para a previsão do retorno de cada ativo no dia seguinte.

O VAR permite fazer uma previsão dos próximos termos de um conjunto de variáveis baseado em um determinado número de termos anteriores e um termo de erro.

Estratégia

Foi escolhido para a análise da estratégia o universo das ações de varejo de roupas e calçados presentes no índice IBrX 100 (os 100 ativos mais negociados e representativos no mercado). Tal escolha de critérios se deu porque as ações de varejo de roupas e calçados têm grande correlação com o consumo das famílias e outro fato de estarem no IBrX 100 evita possíveis problemas de liquidez nas negociações dos ativos.

O mecanismo de trade consiste em três etapas, realizadas diariamente. Primeiramente, treinamos o modelo VAR(p) com a série de retornos ao longo dos últimos 484 dias úteis. Testamos períodos variados entre 200 e 600 dias, e todos produziram resultados semelhantes, de forma que escolhemos 484 por ser o valor que maximiza o retorno. Em seguida, utilizamos o AIC para encontrar o número de *lags* (p) mais adequado sem que ocorra *over-fitting*. Em geral, os valores de p retornados variaram entre 0 e 3. Se o teste retorna p nulo, não é possível garantir que o modelo forneça uma boa previsão, logo não entramos em nenhuma posição, e consideramos o retorno diário do CDI. Caso contrário, tomamos os *forecasts* do modelo com p *lags* e definimos pesos para a posição em cada ação a partir da equação V:

$$P_i = \frac{f_i - f}{\sum \left| f_i - \bar{f} \right|} \quad (V)$$

Onde P_i e f_i são, respectivamente, o peso e o forecast da i-ésima ação. Note que o somatório dos pesos é igual a zero, indicando que o modelo fica 50% comprado e 50% vendido. A razão dessa escolha de pesos é que buscamos ficar comprados nas ações para as quais o modelo prevê uma valorização maior (ou perda menor) e vendidos nas que possuem expectativa de valorização menor (ou perda maior).

Por fim, entramos em cada posição com o respectivo peso considerando o preço de fechamento do dia, e liquidamos as posições no fechamento do dia seguinte. Fazemos isso ao longo de todos os dias do período de *backtest*. Estamos considerando que não é acrescentado dinheiro ao longo desse período, e que a metade comprada é utilizada como margem para a metade vendida. Além disso, custos de transação e de aluguel de ações foram desconsiderados.

Backtest

Foram obtidos os dados de preços de fechamento das ações de *tickers* LREN3, AMAR3, ALPA4, GUAR3, HGTX3 e ARZZ3 para o período de 01/03/2008 até 31/12/2019 por meio do *software factset* conforme indicado na figura 1:

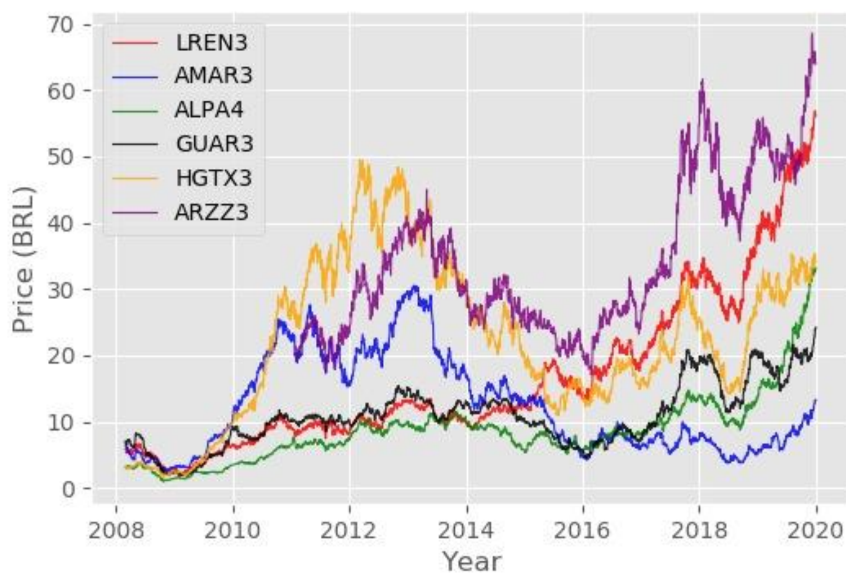


Figura 1: Séries de preços de fechamento de todos os ativos escolhidos de 01/01/2008 a 31/12/2019

A partir das séries de preços de fechamento de LREN3, AMAR3, ALP4, GUAR3 e HGTX3 no período de 2008-2020, foram obtidas as séries de variações percentuais nos preços de fechamento (usaremos o termo série de retornos). Em seguida, foram feitos testes de estacionariedade utilizando o ADF com nível de significância de 95% e número de *lags* escolhido a partir do AIC no período completo. Nas séries de retornos, os resultados foram todos positivos com p-valor igual a 0.0, conforme retornado pela biblioteca *statsmodels* do *python*, indicando que elas são estacionárias.

Realizando o *Johansen Cointegration Test* 4 com 1 *lag* e nível de significância de 95%, o resultado também foi positivo para todas as séries. Por fim, realizamos o *Granger's Causality Test* com número de *lags* máximo igual a 30, obtendo uma matriz com todos os coeficientes fora da diagonal principal inferiores a 0.07, confirmando a hipótese de causalidade para todos os pares com nível de significância maior ou igual a 93%. Esse conjunto de resultados positivos é importante para assegurar a eficácia do VAR nos dados utilizados.

Name	::	Test Stat	> C(95%)	=>	Signif
LREN3	::	6638.11	> 60.0627	=>	True
AMAR3	::	5002.82	> 40.1749	=>	True
ALPA4	::	3628.32	> 24.2761	=>	True
GUAR3	::	2340.68	> 12.3212	=>	True
HGTX3	::	1078.49	> 4.1296	=>	True

Tabela 1: ADF test das séries de preços

	LREN3_x	AMAR3_x	ALPA4_x	GUAR3_x	HGTX3_x
LREN3_y	1.0000	0.0739	0.0000	0.0000	0.0115
AMAR3_y	0.0005	1.0000	0.0472	0.0041	0.0080
ALPA4_y	0.0000	0.0019	1.0000	0.0001	0.0001
GUAR3_y	0.0000	0.0000	0.0000	1.0000	0.0000
HGTX3_y	0.0054	0.0004	0.0023	0.0000	1.0000

Tabela 2: Johansen Cointegration Test das séries de preços

Os testes incluindo ARZZ3 no período de 2011-2020 obtiveram resultados semelhantes.

O gráfico do patrimônio líquido da estratégia (retornando CDI quando p=0) está indicado na figura 2:

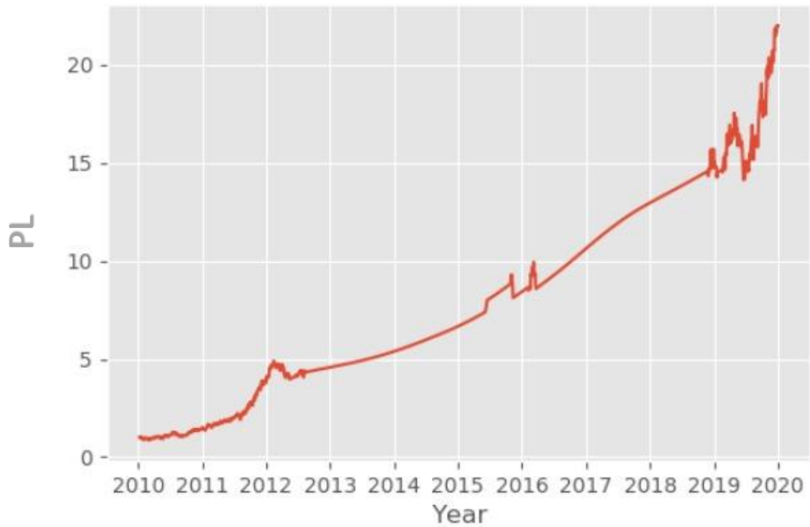


Figura 2: Patrimônio líquido da estratégia da estratégia no período 2010-2020

O modelo passou um longo período retornando p = 0 (0 lags). Alternativamente, apenas a fim de curiosidade, avaliamos os resultados do modelo (com ARZZ3) utilizando o segundo melhor p sempre que o teste AIC retorna p nulo, conforme indicado na figura 3.

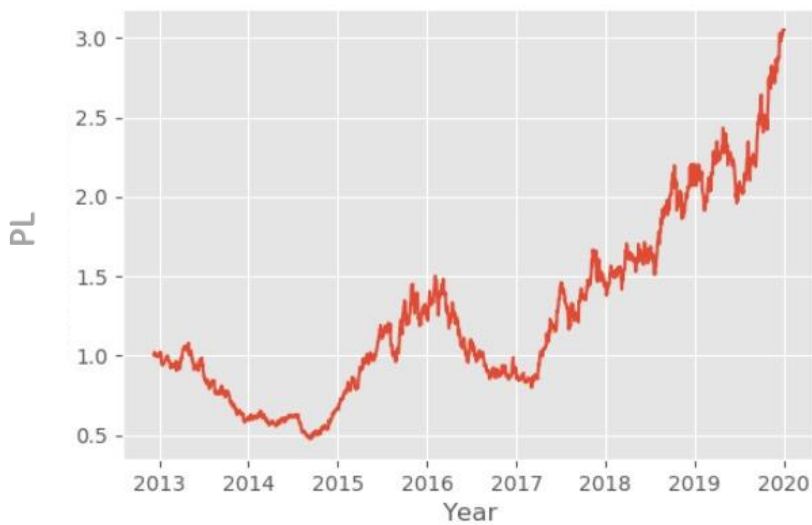


Figura 3: Patrimônio Líquido da estratégia no período 2013-2020 tomando o segundo melhor p quando p=0

Observa-se uma performance positiva, porém com *drawdowns* elevados no período em que o teste AIC retorna p nulo, de forma que preferimos utilizar a metodologia descrita previamente.

Considerando o período completo, obtivemos um *sharpe ratio* de 1,31. Além disso, $\beta = 0{,}016$, indicando que a estratégia é neutra em relação ao ibovespa. Também é possível ver os maiores *drawdowns* na tabela 3 e na figura 4 podemos observar todos os *drawdowns*, e o tempo de recuperação. Com exceção do *drawdown* que ocorre logo antes do período com p nulo, cuja recuperação é demorada devido ao menor retorno do CDI comparado ao do modelo, os períodos de recuperação são relativamente rápidos.

Worst drawdown periods	Net drawdown in %	Peak date	Valley date	Recovery date	Duration
0	19.45	2019-04-26	2019-06-21	2019-09-17	103
1	17.95	2010-07-23	2010-08-31	2010-10-25	67
2	17.48	2012-02-13	2012-04-20	2013-12-02	471
3	17.20	2010-01-14	2010-03-02	2010-04-13	64
4	14.01	2011-07-19	2011-08-05	2011-08-16	21

Tabela 3: Análise de *drawdowns*

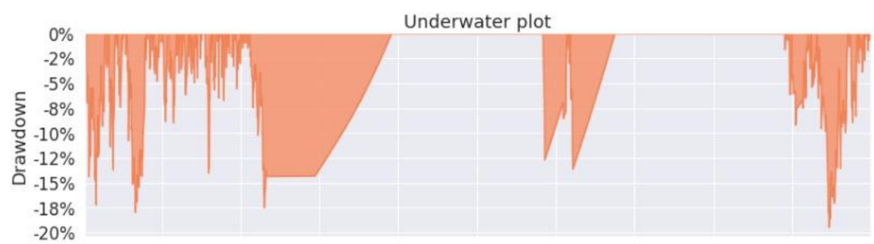


Figura 4: Análise de *drawdowns* por meio do *underwater plot*

Para ter uma noção mais precisa dos resultados do modelo VAR, fizemos os mesmos cálculos do período inicial até 01/05/2012 (observado na figura 5), antes do modelo começar a retornar continuamente p nulo, e obtivemos $\beta = 0{,}047$ *e sharpe ratio de 2,09.*

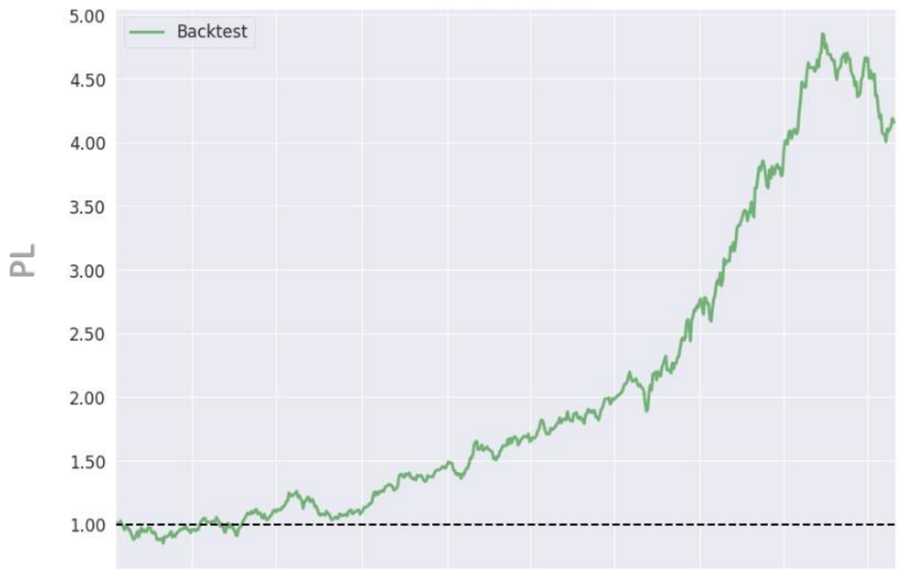


Figura 5: Patrimônio líquido da estratégia no período 01/01/2010-01/05/2012

Erros de Backtest

O *backtest* realizado apresenta como principal defeito assumir que todos os trades conseguem ser efetuados nos preços de fechamento dos ativos. Esse erro foi cometido não por falta de conhecimento, mas, por falta de acesso aos demais dados de interesse.

Outro viés que poderia atingir o modelo seria o *over-fitting* por um grande número de *lags*, porém tal erro foi mitigado pelo uso do *AIC* para a determinação do número de *lags* (no maior caso teve-se p=2). Poderia também haver o $Data - Snooping\ Bias$[6], que é o fato de escolher o tamanho da janela de observação com base nos resultados do modelo, mas, para garantir que esse viés não estava presente, variou-se o tamanho da janela de observação 20% (em ambos os sentidos) para verificar a consistência dos resultados (figuras 6 e 7):

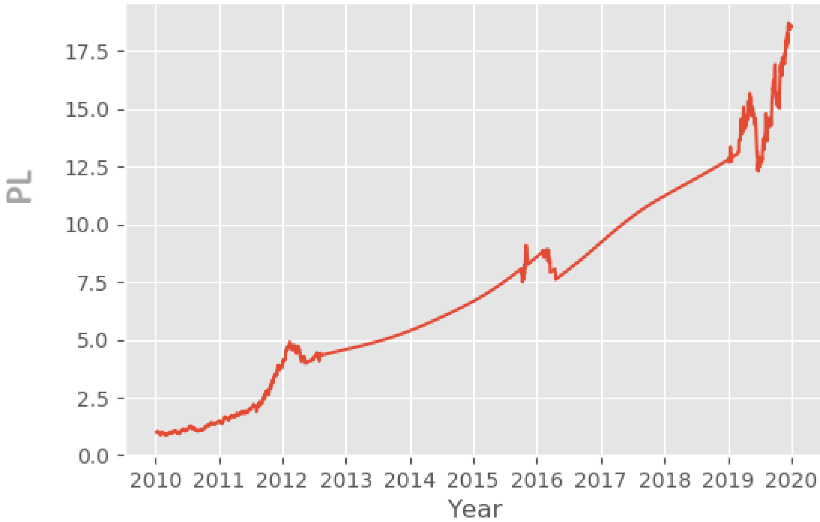


Figura 6: Patrimônio líquido da estratégia no período 2010-2020 com janela de 580 observações

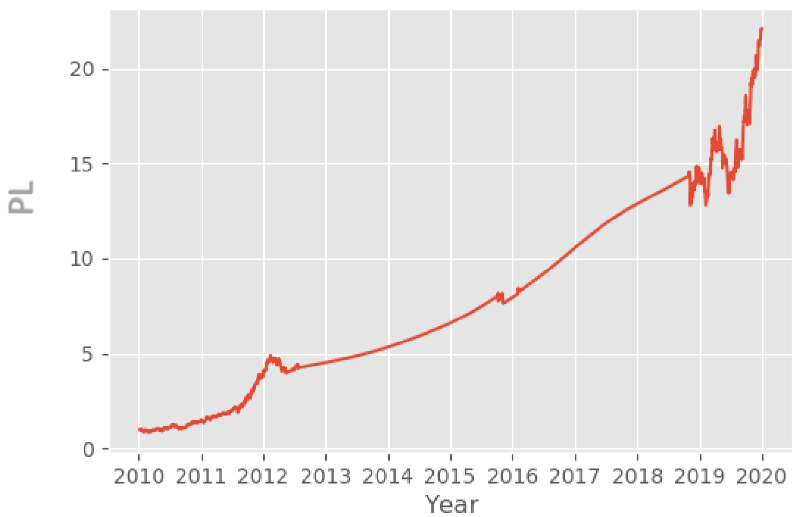


Figura 7: Patrimônio líquido da estratégia no período 2010-2020 com janela de 387 observações

Conforme observado, a escolha da janela de observação não está viesada pelo resultado uma vez que sua alteração em 20% não implica em variações significativas na performance.

Conclusão

Dos resultados obtidos, podemos concluir que a estratégia conseguiu com êxito superar o objetivo inicial de bater o CDI como *benchmark*. A escolha deste benchmark foi feita levando em consideração o comportamento da estratégia, que assume posições *long / short* em ativos possuem uma relação estatística entre si. Por conta disso, não era de se esperar uma correlação significativa do resultado da estratégia com certos índices representativos da classe de ativos operadas, como o Ibovespa.

Usando a métrica do índice beta, pudemos, de fato, constar uma correlação bem baixa com o índice de ações:

$$\beta = 0{,}084$$

A principal métrica para avaliar o desempenho da estratégia foi o *Sharpe Ratio*, que mensura o quão grande é o retorno em excesso da estratégia para o risco, parametrizado pelo desvio padrão dos retornos, ao qual ela está submetida. Usando como referência para o cálculo do retorno em excesso, o CDI acumulado no período, obtivemos:

$$Sharpe\ Ratio\ =\ 1{,}806$$

Por fim, ressalta-se que possível que o modelo funcione melhor para intervalos intradiários, sem retornar p nulo por períodos prolongados. Outro fator que pode ser utilizado para aperfeiçoar o modelo são as diferenças do retorno de acordo com o dia da semana. Espera-se que as previsões sejam mais distantes da realidade nas segundas-feiras, devido a eventos ou notícias durante os fins de semana.

Referências Bibliográficas

[1] Chan, E. P. (2017). Machine Trading: Deploying Computer Algorithms to Conquer the Markets. Wiley.

[2] MUSHTAQ, Rizwan. Testing Time Series Data For Stationarity. Université Paris, Paris, 2011.

[3] Portet S. A primer on model selection using the Akaike Information Criterion. Infect Dis Model. 2020;5:111-128. Published 2020 Jan 7. doi:10.1016/j.idm.2019.12.010.

[4] HJALMARSSON, Erik. Testing for Cointegration Using the Johansen Methodology when Variables are Near-Integrated. Published jun, 2007.

[5] M. Gregorova, A. Kalousis, S. Marchandmaillet, and J. Wang, “Learning vector autoregressive models with focalised granger-causality graphs,” Computer Science, 2015.

[6] Chan, E. P. (2013). Algorithmic Trading: Winning Strategies and Their Rationale. Wiley.