

FACULDADE DE ECONOMIA DA UNIVERSIDADE DE COIMBRA

MESTRADO EM GESTÃO - CIÊNCIA APLICADA À DECISÃO

*Algoritmos genéticos aplicados a séries temporais em  
mercados cambiais*

*Dissertação de Mestrado*

Luis Pedro do Vale Mendes

Ano de 2008

FACULDADE DE ECONOMIA DA UNIVERSIDADE DE COIMBRA

MESTRADO EM GESTÃO - CIÊNCIA APLICADA À DECISÃO

*Algoritmos genéticos aplicados a séries temporais em  
mercados cambiais*

*Dissertação de Mestrado*

Orientadores:

Doutora Joana Maria Pina Cabral Matos Dias

Professora Auxiliar na Faculdade de Economia da Universidade de Coimbra

Doutor Pedro Manuel Cortesão Godinho

Professor Auxiliar na Faculdade de Economia da Universidade de Coimbra

Mestrando: Luís Pedro do Vale Mendes

Ano de 2008

## **Abstract**

This paper reports the work done to build a trading system for the EUR/USD currency pair, in the forex market, that could beat the market. That is, it should be profitable when used in out of sample series. If this endeavour were to be achieved, it would be in disagreement with what is established in the Efficient Market Theory.

The simulation system was built in a way to be as close as possible to the real market conditions. Spread was taken into consideration as the cost to do the business. Time series data of four different frequencies were used. The system is supposed to take advantage of price trends and should allow a maximum of one open position at any given time. Five rules were designed to enter a position and five others to exit. A total of 31 parameters were considered. Due to the enormous search space considered, a genetic algorithm was built in order to achieve good solutions for the problem. The genetic algorithm was conceived with a constant set of ten individuals evolving in 50 generations. Fitness was conceptualized as the ration between return and maximum drawdown achieved by an individual.

Experiments have shown that when a group of individuals is trading in parallel, in the same time frequency data, positive fitness and profits should be expected. The best individuals can be interpreted as following short term trends and entering the market on certain rebounds. Exits, which were defined when the new position was taken, can be readjusted if the algorithm realizes there was a halt on the trend.

Key words: Genetic Algorithm, Foreign Exchange Market, High Frequency Data, Trading System

## Resumo

Com este trabalho pretendeu-se construir um sistema de negociação para o par de divisas EUR/USD, no mercado de Forex, que permitisse a obtenção de rentabilidades supra-normais, isto é, que se mostrasse consistentemente lucrativo nas diversas séries temporais. Tal feito mostrar-se-ia em desacordo com a Teoria da Eficiência dos Mercados.

O sistema de simulação foi construído de forma a aproximar-se, o mais possível, das condições reais de negociação no mercado, incorporando custos de transacção. Os dados utilizados dizem respeito a quatro séries base temporais, correspondendo às frequências de 1, 5, 15 e 60 minutos. O sistema procura tirar partido de tendências de evolução dos preços, permitindo, no máximo, a existência de uma posição aberta a cada momento. Foram previstas cinco regras de entrada em posição e outras cinco de saída. No total, foram considerados 31 parâmetros diferentes.

Dado o vasto espaço de soluções possíveis, foi desenvolvido um algoritmo genético responsável pela escolha da estrutura e dos valores dos parâmetros de modo a possibilitar a obtenção de boas soluções. O algoritmo genético labora na evolução de populações de dez indivíduos durante 50 gerações. O conceito de fitness, segundo o qual cada indivíduo é avaliado, resulta do quociente entre o lucro (ou prejuízo) obtido e a maior quebra a que esteve sujeito.

As conclusões apontam para a existência de soluções que permitem esperar a obtenção de um desempenho positivo nas séries de testes. Os melhores indivíduos têm uma estrutura que tira partido de tendências de curto prazo, com o aproveitamento de retrocessos, na evolução dos preços. A saída de posição, inicialmente prevista, pode ser alterada se for constatada uma atenuação na tendência.

Palavras chave: Algoritmo genético, Mercado cambial, Dados de grande frequência, Sistema de negociação

*Lista de abreviaturas*

DD	drawdown
EMA	média móvel exponencial
EMAc	valor da média móvel exponencial de curto prazo no período
EMAl	valor da média móvel exponencial de longo prazo no período
MDD	maximum drawdown
P	preço médio do período
pip	Percentage in point
PL	posição longa
PC	posição curta
sl	stop loss
tp	take profit
ANN	artificial neural network
GA	genetic algorithm
AUD	Australian dolar
DEM	Deutsche Mark
EUR	Euro
GBP	Great Britain Pound
USD	United States Dolar

# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>9</b>
<b>2</b>	<b>Revisão bibliográfica</b>	<b>12</b>
2.1	Indicadores, padrões técnicos e sistemas desenvolvidos . . . . .	14
2.1.1	Indicadores e padrões técnicos . . . . .	14
2.1.2	Sistemas desenvolvidos . . . . .	18
2.1.3	Séries temporais . . . . .	20
2.2	Algoritmos genéticos . . . . .	22
2.3	Aplicação de algoritmos genéticos aos mercados de divisas . . . . .	25
2.4	Medição do desempenho . . . . .	27
2.5	Evolução histórica dos resultados obtidos . . . . .	32
<b>3</b>	<b>O Mercado e as séries temporais</b>	<b>34</b>
3.1	Forex . . . . .	34
3.2	Séries temporais . . . . .	38
3.2.1	Séries usadas para evolução . . . . .	40
3.2.2	Séries usadas em testes . . . . .	41
<b>4</b>	<b>Metodologia</b>	<b>46</b>
4.1	Objectivo e pressupostos iniciais . . . . .	46
4.2	Estrutura do algoritmo genético . . . . .	48
4.3	Estratégias de entrada e saída de posição . . . . .	50
4.3.1	Regras de entrada de posição . . . . .	52

4.3.2	Regras de saída de posição . . . . .	58
4.4	Caracterização do algoritmo genético . . . . .	65
4.4.1	Comportamento do indivíduo . . . . .	66
4.4.2	Avaliação do desempenho . . . . .	74
4.4.3	Geração seguinte . . . . .	76
<b>5</b>	<b>Resultados obtidos</b>	<b>84</b>
5.1	Evolução da população . . . . .	84
5.2	Desempenho em séries de evolução . . . . .	85
5.3	Desempenho em séries de testes . . . . .	91
5.3.1	Indivíduos treinados em séries de 1 minuto . . . . .	91
5.3.2	Indivíduos treinados em séries de 5 minutos . . . . .	93
5.3.3	Indivíduos treinados em séries de 15 minutos . . . . .	94
5.3.4	Indivíduos treinados em séries de 60 minutos . . . . .	96
5.4	Estatística geral e aferição de relações . . .	97
5.4.1	Relação entre o desempenho em sé- ries de evolução e em séries de teste	98
5.4.2	Estatística geral de desempenho nos testes efectuados, por tipo de série .	99
5.4.3	Estatística de desempenho nos tes- tes efectuados, por frequência e tipo de série . . . . .	99
5.4.4	Desempenho médio em testes, por frequência . . . . .	100
5.4.5	Comparação entre o desempenho em testes de séries futuras e o de séries passadas . . . . .	101

5.4.6	Relação entre o desempenho em séries de testes anteriores e posteriores ao período de evolução . . . . .	102
5.4.7	Relação entre desempenho e número de negócios em séries de testes . . .	103
5.4.8	Relação entre desempenho numa série futura e numa sua subsérie . . .	103
5.4.9	Adequação de séries às operações de evolução e teste . . . . .	105
5.5	Características dos melhores indivíduos . .	107
5.6	Testes computacionais . . . . .	110
<b>6</b>	<b>Conclusões</b>	<b>113</b>



# Lista de Figuras

2.1	Reprodução através do cruzamento de genes entre dois progenitores . . . . .	25
3.1	Evolução das cotações nas séries de evolução primária . . . . .	40
3.2	Evolução das cotações nas séries de evolução secundária . . . . .	41
3.3	Séries de testes com periodicidade de 1m .	42
3.4	Séries de testes com periodicidade de 5m .	42
3.5	Séries de testes com periodicidade de 15m .	43
3.6	Séries de testes com periodicidade de 60m .	43
4.1	Cromossoma utilizado . . . . .	49
4.2	Vector de 31 parâmetros que constitui cada estratégia de transacção . . . . .	51
5.1	Exemplo de evolução da população . . . . .	85
5.2	Desempenhos obtidos em séries de treino .	87
5.3	Relação entre n. <sup>o</sup> de negócios e desempenho obtido . . . . .	90
5.4	Séries 1 a 4 treinadas com base no mês de Novembro de 2007; série 5 treinada com base no mês de Outubro do mesmo ano. . .	91
5.5	Séries 1 a 4 treinadas com base no 2. <sup>o</sup> quadrimestre de 2007; série 5 treinada com base no 1. <sup>o</sup> quadrimestre do mesmo ano. . . . .	93

5.6	Séries 1 a 4 treinadas com base no ano de 2006; série 5 treinada com base no ano de 2005. . . . .	95
5.7	Desempenho em séries de teste por indivíduo treinado em série de 60 min . . . . .	97
5.8	Relação entre desempenho obtido em séries de evolução e de testes . . . . .	98
5.9	Relação entre desempenho obtido em séries de testes imediatamente anteriores e posteriores ao período de evolução . . . . .	102
5.10	Relação entre n. <sup>o</sup> de negócios e desempenho obtido em séries de testes . . . . .	104
5.11	Relação entre desempenho obtido em séries de testes longas e suas subséries iniciais . . . . .	105
5.12	Negociação do indivíduo da 3. <sup>a</sup> série da frequência de 60 minutos nos primeiros 400 períodos do 1. <sup>o</sup> trimestre de 2007 . . . . .	111

# Lista de Quadros

3.1	Exemplo com os dados relativos à cotação de um período de 1 minuto . . . . .	39
3.2	Séries utilizadas pela evolução primária . .	40
3.3	Séries utilizadas pela evolução secundária .	41
3.4	Séries utilizadas em testes . . . . .	42
3.5	Séries usadas em testes <i>out-of-sample</i> agrupadas por período de evolução e período <i>out-of-sample</i> . . . . .	45
4.1	Representação da regra 1 . . . . .	54
4.2	Representação da regra 2 . . . . .	55
4.3	Representação da regra 3 . . . . .	56
4.4	Representação da regra 4 . . . . .	58
4.5	Representação da regra 5 . . . . .	58
4.6	Representação da regra 6 . . . . .	60
4.7	Representação da regra 7 . . . . .	61
4.8	Representação da regra 8 . . . . .	63
4.9	Representação da regra 9 . . . . .	64
4.10	Representação da regra 10 . . . . .	65
5.1	Desempenhos, resultados em <i>pips</i> e n.º de negócios, em séries de treino . . . . .	86
5.2	Estatística do número de negócios efectuados em séries de evolução . . . . .	89

5.3	Desempenho, resultado em <i>pips</i> e número de negócios, em séries de teste <i>out-of-sample</i> dos melhores indivíduos treinados em séries de 1 minuto . . . . .	92
5.4	Desempenho em séries de teste <i>out-of-sample</i> dos melhores indivíduos treinados em séries de 5 minutos . . . . .	93
5.5	Desempenho em séries de teste <i>out-of-sample</i> dos melhores indivíduos treinados em séries de 15 minutos . . . . .	95
5.6	Desempenho em séries de teste <i>out-of-sample</i> dos melhores indivíduos treinados em séries de 60 minutos . . . . .	96
5.7	Estatística geral de desempenho nos testes, por tipo de séries . . . . .	99
5.8	Estatística de desempenho nos testes, por frequência e tipo de séries . . . . .	100
5.9	Desempenho médio em séries de teste, por frequência . . . . .	101
5.10	Desempenho médio em séries de teste anteriores e posteriores ao respectivo período de evolução, por frequência . . . . .	102
5.11	Os melhores indivíduos . . . . .	112

# Capítulo 1

## Introdução

Nos últimos anos, tem-se assistido a um enorme interesse pela área financeira, por parte da população em geral. O desejo de obtenção de rendibilidades elevadas em investimentos, conjugado com uma abundante e diversa oferta propiciaram as condições de base para o florescimento da indústria financeira. A Internet e a sua onnipresença, em qualquer ponto do planeta, proporcionaram o meio de comunicação por excelência entre a procura e a oferta deste tipo de serviços.

Para um completo envolvimento da sociedade faltava ainda eliminar as barreiras de entrada ao sector. Tal foi conseguido com a profusão de plataformas electrónicas de negociação, acessíveis a partir de dispositivos computadorizados, e com a possibilidade de abertura de contas de investimento com pequenos valores, por parte dos novos investidores. No passado, eram poucos os que abraçavam a profissão de negociador (*trader*), nos mercados, e que com ela ganhavam a vida. Recentemente, essa actividade ficou ao dispor imediato de todos os interessados.

A ferramenta de eleição utilizada pelos negociadores para gerar lucros é a análise técnica. Por análise técnica entende-se uma forma de prever movimentos futuros de activos financeiros com base na observação do comporta-

mento passado das suas cotações.

A comunidade científica não esteve imune a todo este processo. Desde logo, a Teoria da Eficiência dos Mercados afirma que os preços do mercado reflectem o conhecimento e as expectativas de todos os investidores. Qualquer novo acontecimento é de imediato incorporado no preço de um activo, pelo que é impossível bater o mercado consistentemente, isto é, conseguir resultados positivos da prática de negociação contínua.

Muitos trabalhos científicos tentam comprovar ou repudiá-la esta teoria, sendo que ambos os lados tentam apresentar provas das suas conclusões.

Neste trabalho, pretende-se construir um sistema de negociação baseado em regras objectivas, que permita obter lucros supra-normais através do aproveitamento de tendências de evolução das cotações, e que seja tão próximo das condições reais quanto possível.

Como contrato a transaccionar, foi escolhido o par do Euro com o Dólar norte-americano, de entre os disponíveis no mercado de divisas. Este é um dos ramos com maior crescimento de entre as áreas financeiras e um que tem grande participação popular.

Dada a multiplicidade de estratégias de transacção possíveis e conseqüente impossibilidade física de as obter todas, houve o recurso à construção de um algoritmo genético. Neste algoritmo evolutivo, cada estratégia de transacção é uma solução admissível ou, na terminologia deste tipo de procedimentos, um indivíduo. Pretende-se desenvolver um procedimento expedito que percorra o espaço das soluções em busca de indivíduos com bom desempenho, isto é, de boas soluções, recorrendo à simulação, por computador, de processos de evolução e selecção natural, entre indivíduos de uma população.

Foram escolhidas, para treino e testes, séries temporais, de diferentes frequências: 1, 5, 15 e 60 minutos.

O desempenho de cada indivíduo é calculado com base no resultado obtido e no risco inerente à solução, tendo em conta que os custos de transacção são considerados na formação do resultado.

Com as simulações efectuadas pretende-se fazer luz sobre as seguintes hipóteses, entre outras: Será possível, com o algoritmo genético, identificar estratégias que permitam a obtenção de resultados positivos nos períodos de treino? Esse resultado manter-se-á quando a mesma estratégias for aplicada às séries de teste *out-of-sample*? O desempenho diminuirá com o afastamento temporal das séries em relação ao período de evolução? As soluções obtidas, em séries de uma certa frequência temporal, continuarão a permitir obter resultados positivos quando aplicadas a séries de teste de outras frequências?

No capítulo 2 será efectuado o enquadramento do trabalho em termos do que foi publicado por diversos autores na comunidade científica. O capítulo 3 dá a conhecer o mercado de divisas utilizado e apresenta as séries temporais utilizadas neste trabalho. No capítulo 4 é descrita a metodologia utilizada. Os resultados obtidos são descritos no capítulo 5. O capítulo 6 é composto pelas conclusões.

## Capítulo 2

# Revisão bibliográfica

*“In theory, theory and practice are the same.  
In practice, they are not.”*  
(Albert Einstein)

A comunidade científica não ficou imune a todo o interesse despertado pelos mercados financeiros, nas últimas décadas. Os trabalhos dedicados ao estudo de activos como as acções, obrigações, divisas, instrumentos financeiros derivados e outros abundam, tendo sido possível enriquecer esta área de saber com o contributo de muitos investigadores, por todo o mundo. Na sua generalidade, os estudos efectuados utilizam vários tipos de “indicadores técnicos”.

Da revisão bibliográfica efectuada, ressalta, desde logo, uma separação de caminhos de entre os trabalhos realizados na área deste. Qual o ponto de partida? Poder-se-á enveredar por uma tentativa de adivinhação do futuro, focando o problema na determinação dos valores da cotação seguinte, dos picos ou dos fundos do mercado e tentar maximizar os resultados que poderão resultar de tal perícia. Ou pode assumir-se, de antemão, que a previsão do futuro é um caminho tortuoso e que, para identificar uma boa estratégia de transacção, “basta” aproveitar tendências



de evolução dos mercados.

Como exemplo da primeira lógica pode ser referido o trabalho de Kamruzzaman e Sarker [13]. Os autores usam redes neuronais artificiais (ANN<sup>1</sup>), de três tipos, para prever taxas de câmbios futuras de seis divisas em relação ao dólar australiano (AUD). A investigação é baseada em cerca de 500 preços de fecho semanais. Depois de treinados, com o uso de indicadores técnicos simples, como as médias móveis, as ANN são aplicadas a dois períodos: um período de 35 e outro de 30 semanas. Atingiram 80% de correcta previsão da mudança direccional da tendência das cotações. Também Kaboudan, em [12], pretende obter projecções para o preço de acções no dia seguinte, com o uso de programação genética e de modelos de regressão.

Esta estratégia de previsão de picos e fundos do mercado não é, contudo, bem vista por quem se dedica profissionalmente à negociação nos mercados, para os quais, a segunda opção acima indicada é a correcta. No meio científico, vários autores optam por este segundo caminho. Um tal exemplo é o trabalho de Schulmeister [25], em que o autor considera a possibilidade de se obterem lucros através do aproveitamento de tendências persistentes. Para o efeito, estudou o comportamento do par de divisas Marco alemão (e posteriormente o Euro) em relação ao dólar norte-americano (USD), através da construção de 1024 sistemas de negociação. O período de treino ocorreu em séries diárias de 1973 a 1999, tendo os testes *out-of-sample*<sup>2</sup> tido lugar para as séries entre 2000 e 2004. Schulmeister conclui que: a) Os melhores 25 modelos no período de treino foram, quase sempre, lucrativos quando testados nas séries *out-of-sample*. b) O aproveitamento das tendências implica, segundo o mesmo, um número de

---

<sup>1</sup>Em inglês, Artificial Neural Networks.

<sup>2</sup>Testes efectuados com dados não utilizados para o treino do sistema utilizado.

negócios lucrativos menor do que o de negócios com prejuízo. c) As posições lucrativas duram 3 a 5 vezes mais do que as perdedoras. d) Deve minimizar-se a duração das posições perdedoras.

Tan e Dihadjo, em [27], testam sistemas de negociação para o AUD usando médias móveis das cotações, para além do modelo ARMA<sup>3</sup>. Consideram que existe uma tendência nos preços quando uma média móvel de curta duração cruza uma média móvel de longa duração. Tentam tirar partido da existências das tendências nos preços para gerarem lucros na negociação do AUD/USD.

Estes últimos modelos, entre muitos outros desenvolvidos pelos diversos investigadores, nesta área, usam ferramentas de análise técnica.

## **2.1 Indicadores, padrões técnicos e sistemas desenvolvidos**

### **2.1.1 Indicadores e padrões técnicos**

A “análise técnica” representa uma das formas utilizadas para tentar prever a evolução futura das cotações de um activo financeiro. Pressupõe que as cotações reflectem tudo o que é possível saber sobre um determinado activo, em cada momento. Os seus defensores afirmam que da observação, identificação e medida de alguns padrões de evolução passada das cotações é possível prever tendências futuras, pelo menos, no curto prazo. Os “indicadores técnicos” são uma das ferramentas que fazem parte da “análise técnica”.

Na literatura, o uso de “indicadores técnicos” para o desenvolvimento de algoritmos de negociação em mercados financeiros é algo de bastante comum. Por “indicador téc-

---

<sup>3</sup>Auto regressive moving average.

nico” entende-se uma série numérica que é determinada com a aplicação de uma fórmula matemática a uma ou mais séries base, seja esta de cotações, volumes de transacção ou de outros dados. A ideia subjacente ao uso dos indicadores técnicos é a de facilitar as decisões para uma entrada ou saída de posição, através da suposta antecipação, maior precisão ou confirmação de “sinais” dados por estes. Vários indicadores e de vários tipos podem ser utilizados simultaneamente numa dada estratégia de negociação.

Os indicadores técnicos podem agrupar-se em várias categorias: tendenciais, de *momentum*, de volatilidade, entre outros.

A média móvel é considerada como o mais simples dos indicadores tendenciais. Com a sua aplicação, pretende-se fazer sobressair a tendência de evolução das cotações. As médias móveis podem ser de vários tipos, sendo as simples e as exponenciais as mais utilizadas. No primeiro caso, são consideradas as médias aritméticas simples dos valores das cotações dos últimos “n” períodos. Ao passar ao período seguinte, às cotações consideradas no cálculo anterior, é acrescentado o valor da cotação do momento presente e desprezado o mais antigo. As médias móveis exponenciais (EMA)<sup>4</sup> têm como vantagem, em relação às simples, um maior ajustamento às variações que vão ocorrendo nas cotações, assinalando, com alguma antecedência, pontos de inflexão nas tendências. Tal resultado é conseguido por se atribuir um peso relativo superior às cotações mais recentes. As médias móveis servem, ainda, de base à construção de outros indicadores.

Os índices de *momentum* tentam medir o estado actual das cotações de um activo financeiro, em relação ao seu

---

<sup>4</sup>Em inglês, Exponencial Moving Average.

passado recente. Talvez o mais célebre dos indicadores de *momentum* seja o RSI, “Relative Strength Index”. Este índice compara a magnitude dos ganhos recentes, numa cotação, às suas perdas e retorna um valor compreendido entre zero e cem. Por exemplo, em valores abaixo de um certo patamar, normalmente 20 ou 30, é considerado que um activo está em modo sobre-vendido, o que poderá indicar uma subida de cotação em breve, ou apenas que a força vendedora se poderá esbater. Em valores acima de 70 ou 80, é considerado que o activo está em modo sobre-comprado. Tanto neste como em outros indicadores, são igualmente analisados os momentos de divergência que podem indiciar, para um momento próximo, uma inversão da tendência das cotações. Estes momentos acontecem quando, por exemplo, as cotações atingem um novo mínimo, em relação a um fundo recente, mas o indicador permanece a um nível superior ao que registava no mínimo anterior das cotações.

Os activos financeiros estão sujeitos a períodos alternados de alta e baixa volatilidade nas suas cotações. Por volatilidade, deve entender-se, neste sentido, uma medida da amplitude de variação dos preços, num determinado período de tempo. Geralmente, está associada ao cálculo do desvio-padrão das cotações. Um dos indicadores mais conhecidos que tenta medir a volatilidade, sem, no entanto, fazer uso do desvio-padrão, é o ATR, “Average True Range”. O “True Range” diz respeito à maior das seguintes diferenças: a) entre o máximo e o mínimo do período actual; b) entre o máximo actual e o valor de fecho anterior; c) diferença entre o mínimo actual e o fecho anterior. A média dos valores assim calculados para os últimos “n” períodos resulta no ATR. Este valor indica quanto um activo se apreciou ou depreciou por período, em média, durante

os últimos “n” períodos. Baixos valores são sinónimos de que as cotações se mantêm relativamente constantes. Valores altos do ATR significam, por seu turno, que os preços variam bastante, em média, durante cada período.

As áreas de suporte e resistência na evolução das cotações dos variados activos financeiros, embora não sendo indicadores técnicos, são considerados por muitos dos agentes a operar nos mercados. A subida nas cotações é normalmente originada por um excesso de procura do activo face à sua oferta, num determinado período. A descida é provocada pela existência de mais pressão do lado vendedor. No entanto, as duas forças tendem a equiparar-se cerca de alguns níveis de cotação. O suporte pode ser encontrado em áreas onde os preços parecem perder a força de queda que vinham a registar. As áreas de resistência, por outro lado, referem-se às zonas de cotações em que a oferta parece suficientemente forte para travar mais subidas. É tido como certo, por quem utiliza a análise técnica, de que estas zonas, quando são rompidas pelo valor das cotações, se convertem uma na outra. Quando as cotações baixam de uma zona até aí considerada de suporte, é assumido que ela se transforma em resistência. Zonas outrora de resistência são convertidas em zonas de suporte, se as cotações as ultrapassam. Estas áreas de cotações são consideradas importantes, pois podem servir como meta para a tomada de lucros ou como início de novas tendências.

A análise técnica é muitas vezes referida como uma análise gráfica das cotações. A evolução dos preços origina, por vezes, a ocorrência de determinados padrões que são objecto de uma atenção especial. Entre eles está o “*channel breakout*”. Para formar um canal (*channel*), as cotações devem evoluir, ao longo dos últimos períodos, com uma amplitude constante entre alternados máximos e mínimos.

O padrão só é estabelecido, no entanto, quando há um rompimento do canal de cotações, no sentido da subida ou da descida. Este padrão tem sido utilizado, quer na prática da negociação, quer em trabalhos científicos.

### 2.1.2 Sistemas desenvolvidos

São variados os indicadores técnicos usados pelos investigadores, nesta área. Desde a entrada ou saída de posição dada pelo simples cruzamento de médias móveis, com a ajuda de zonas de suporte e resistência, como em [27], até à criação de um indicador, Dimbeta<sup>5</sup>, em [23], inúmeras possibilidades existem. Para efeitos de comparabilidade de valores ao longo de um horizonte temporal extenso, Neely e Weller, em [19], normalizaram os indicadores técnicos utilizados. Dividiram os valores assumidos pelos indicadores pelas respectivas médias móveis de longa duração.

Há indicadores técnicos que não são calculados a partir do preço e que, por via disso, podem dar relevo a uma nova perspectiva sobre a realidade. Bates et al., em [3], usam, como indicadores, variáveis relativas a ordens de transacção. O *order book* corresponde ao conjunto de ordens, dadas ao limite, presentes no sistema de negociação de uma corretora, num determinado momento, que serão executadas à medida que a cotação do par de divisas, em causa, for atingindo os preços predeterminados. Foram consideradas as seguintes grandezas: 1a - apenas ordens *take profit*<sup>6</sup>; 1b - todas as ordens<sup>7</sup>; 2a - todas as ordens existentes no sistema, no momento de recolha da informação; 2b - apenas ordens colocadas no último dia; 3a - ordens distantes entre 0,0% e 0,5% do preço actual; 3b -

---

<sup>5</sup>Este indicador mede a distância relativa das cotações de um activo a uma média das cotações.

<sup>6</sup>Utilizadas para o fecho de posições a um preço limite que proporcione lucro.

<sup>7</sup>Inclui as ordens *stop loss*, as quais são utilizadas para limitar o prejuízo incorrido ou salvar o lucro existente.

ordens distantes do preço actual entre 0,5% e 1%; 3c - o conjunto de 3a e 3b. Ao todo, os autores construíram doze indicadores com base no *order book*. *Order flow* é definido como a diferença entre ordens originárias de compradores e ordens originárias de vendedores, o que representará a “procura líquida” em todo o mercado de Forex<sup>8</sup>. O *order flow* serve de base à construção de três conjuntos de indicadores. Os autores concluem que a coadjuvação de indicadores técnicos “tradicionais” com informação proveniente de *order book* e *order flow* dá indícios de melhorar o desempenho de negociação, no mercado de divisas.

O volume de transacções é considerado um indicador a ter em conta nos mercados financeiros. Em [10], Frino et al. testam 7.846 regras de negociação em mercados de futuros da Austrália, das quais 2.040 dizem respeito ao volume de transacções.

Dado que a análise técnica faz uso igualmente de padrões de evolução de cotações identificáveis pelo olho humano, vários investigadores tentaram estudar este campo de análise. Dempster e Jones, em [6], estudam a possibilidade de identificação e posterior utilização de canais de evolução de preços. Usam séries de dados do par GBP/USD com frequências de um minuto a diárias. Os outros tipos mais comuns de indicadores técnicos servem apenas como filtros, para tomada de decisões. Omrane e Van Oppens, em [22], põem igualmente a ênfase na previsão da evolução das cotações tendo por base a existência de determinados padrões gráficos. Para o efeito, estudam o comportamento do EUR/USD com base em períodos de cinco minutos. Concluem que mais de metade dos padrões tem uma previsibilidade significativa, embora o lucro não chegue para pagar os custos de transacção.

---

<sup>8</sup>Foreign Exchange - expressão inglesa por que é conhecido o mercado cambial.

Para além do uso de indicadores técnicos e correcta identificação de padrões de evolução das cotações, um sistema de negociação necessita da definição de regras precisas para a tomada de uma posição, de compra ou de venda, e posterior saída. Em [25], Schulmeister faz depender a entrada numa posição da emissão de um sinal por parte dos seus indicadores técnicos. No entanto, numa tentativa de eliminação de falsos sinais, aos indicadores é aplicado um filtro: de o sinal continuar activo “n” dias depois do primeiro sinal. Frino, em [10], usa regras de cinco tipos diferentes, no sistema por si desenvolvido: médias móveis, suporte e resistência, channel breakout, regras com base no volume e filtros.

Quanto às estratégias de saída, Meagher et al., em [17], desenvolvem três tipos de saída: 1) 30 minutos depois da entrada de posição; ou 2) passados 30 minutos da entrada é verificado se se mantém o sinal para a tomada de posição de entrada; se sim, estende o período por mais 30 minutos; 3) passados os 30 minutos iniciais ou possivelmente também os 30 minutos adicionais, haverá uma saída forçada da posição, com lucro ou prejuízo. Várias estratégias de saída são igualmente consideradas por Dempster e Jones, em [6]: 1) quando o preço atinge uma banda paralela à do canal de evolução do preço; 2) se o negócio está activo há um determinado período de tempo; 3) se os valores de *stop loss* são atingidos; ou 4) se uma *trailing stop*<sup>9</sup> é accionada.

### 2.1.3 Séries temporais

As séries temporais utilizadas pelos diversos investigadores variam, entre outros factores, quanto à dimensão, à frequência e ao tratamento que lhes é dado.

---

<sup>9</sup>Ordem de fecho de uma posição que acompanha a evolução do preço do activo. É normalmente usada como forma de um investidor garantir uma parte dos lucros entretanto obtida, quando a tendência de preços se inverte.



No que respeita à dimensão e frequência, para além da utilização de dados diários e semanais, de 32 anos e 65 semanas, respectivamente, nos estudos [25] e [13] acima citados, outros trabalhos podem ser referidos que comprovam a diversidade das escolhas efectuadas. Dempster e Jones, em [6], relativamente ao estudo do par de divisas GBP/USD, consideraram frequências de um minuto a um dia nos seus dados. Os mesmos autores, em [7], desenvolveram um sistema de negociação do par USD/GBP, para o período compreendido entre os anos de 1994 e 1997. Foram utilizadas séries temporais de várias frequências, desde intra-diária, de um minuto por período, a dados diários. Meagher et al., em [17], usam séries de cerca de 10.000 cotações, com periodicidade de cinco minutos para avaliar a negociação de acções da Ford e da IBM. Dados de frequência horária de oito anos serviram de base à aplicação de estratégias de negociação nos mercados de capitais, em Pictet et al. [24].

Apesar de estarem disponíveis séries completas de dados de negociação, alguns investigadores optaram por utilizar apenas uma parte destas. Em [19], Neely e Weller estudam a possibilidade de a estratégia, por eles desenvolvida, ser aplicada à negociação no mercado de divisas usando todos os dados de cada dia ou apenas os dados das doze horas de maior actividade. Dunis et al., em [8], consideram a existência de dois modos de operação: 24 horas por dia ou só no horário Europeu, para a negociação dos pares de moedas USD/DEM e DEM/JPY. A questão do horário de negociação assume igualmente relevância em [7]. Dempster e Jones, consideram a existência de três horários de negociação: o de Londres, de Nova Iorque e o Asiático, com implicações ao nível dos custos de transacção. Neely et al. [18] foram ainda mais longe. Os autores deixaram de

parte um determinado período de tempo, quer para a fase de treino, quer para a de selecção do seu algoritmo, em que o USD sofreu uma apreciação significativa, de forma, segundo eles, a evitar enviesamentos nos resultados.

Os dados das séries temporais podem ainda ser tratados de forma a obter construções diferentes. Em [8], Dunis et al. compararam as séries temporais tradicionais com séries desvolatilizadas, em dados intra-diários. A acção de desvolatilização das séries surge da observação efectuada, pelos autores, de que o procedimento de dividir cotações de grande frequência em intervalos regulares de tempo não seria o mais correcto, pois provocaria uma sub-representação dos períodos de maior actividade e uma sobre-representação dos de menor. É, assim, utilizada uma metodologia em que cada período é considerado em termos de intervalos de volatilidade e, não, de tempo. Um período acaba quando a amplitude de cotações entre a presente e a inicial é superior ou igual a uma constante predefinida. Os resultados não foram inequívocos quanto aos méritos da nova perspectiva.

## 2.2 Algoritmos genéticos

Os algoritmos genéticos, GA<sup>10</sup>, são uma meta-heurística de busca e optimização de soluções. Pertencem à família dos algoritmos evolucionários, os quais aplicam, em modelos computacionais, por analogia, técnicas inspiradas na biologia. Consideram uma população inicial de indivíduos, correspondendo, cada um destes, a uma solução candidata à resolução do problema em causa. Os indivíduos, de cada geração, competem entre si, de forma a terem lugar na geração seguinte, directamente ou através de descendência.

---

<sup>10</sup>Genetic algorithm, em inglês.

As operações de selecção, cruzamento entre indivíduos e consequente reprodução observadas na Natureza, encontram paralelo nos GA. Aqueles indivíduos que se mostram mais fortes, i.e., cuja função de avaliação de desempenho, conhecido por *fitness*, apresentar melhores valores, têm uma maior probabilidade de serem escolhidos para progenitores da geração seguinte.

A maior parte dos investigadores opta por pedir emprestada à Biologia as noções e definições que se referem à codificação de soluções tais como *cromossomas*, *genes* (elementos individuais que constituem um cromossoma), e *alelos* (os possíveis valores que um determinado gene pode tomar). Cada gene está localizado numa dada posição (*locus*) do cromossoma. Se a solução é codificada recorrendo a um único cromossoma, então cada indivíduo da população é *haploide*. Se a solução é codificada usando dois cromossomas (como acontece com o ser humano) é chamado *diploide*. A maioria das aplicações de algoritmos genéticos utiliza indivíduos haploides. Geralmente os cromossomas têm um número fixo de genes e cada gene tem uma localização fixa dentro do cromossoma. Por este motivo, a representação em que um cromossoma corresponde a um vector, sendo os genes os elementos desse vector, é muitas vezes utilizada.

Nos algoritmos genéticos, o cruzamento entre soluções pode ser levado a cabo utilizando diferentes procedimentos que dependem, por exemplo, da representação cromossomática escolhida, do número de progenitores, do número de descendentes e, em última análise, do espírito criador do autor do algoritmo. Um dos operadores de *crossover* mais utilizados é conhecido como *one-point crossover*. Duas soluções progenitoras serão recombinadas, dando origem a dois descendentes. Considerando, por exemplo, que cada

cromossoma é representado através de um vector com  $n$  elementos, um valor  $k$  entre 1 e  $n$  é aleatoriamente escolhido. O primeiro filho terá todos os genes nas posições de 1 a  $k$  iguais ao primeiro progenitor e todos os outros genes iguais ao segundo progenitor. O oposto acontece com o segundo filho. Esta operação pode ser visualizada na figura 2.1.

O operador de mutação é responsável por alterações aleatórias na composição genética de um indivíduo. A mutação funciona, basicamente, como forma de tentar evitar perda de material genético.

Dada uma certa probabilidade predefinida, os parâmetros de cada indivíduo poderão sofrer alterações, o que poderá traduzir-se na fuga de um candidato a um extremo local da função de avaliação de desempenho. Cada sequência de operações de selecção dos mais fortes, cruzamento e mutação dá origem a uma nova geração, seguida de nova avaliação de desempenho dos indivíduos que a constituem. O ciclo de evolução é interrompido quando uma condição predefinida de paragem é satisfeita. Esta condição poderá verificar-se por o GA ter obtido a solução óptima do problema, por os desempenhos dos vários indivíduos, geração após geração, convergirem para um determinado valor, por terem ocorrido um número suficiente de gerações, ou por outra razão estabelecida.

As séries de dados onde as populações de indivíduos evoluem são chamadas de *in-sample*. Alguns dos indivíduos treinados, i.e. evoluídos, nestas séries, são postos à prova em novas séries de dados, chamadas de *out-of-sample*.

Os GA podem incorrer em situações de *overfitting* no que respeita a séries temporais financeiras. Por *overfitting* entende-se um ajustamento excessivo de um sistema ao conjunto de dados em que foi treinado, conduzindo a bons

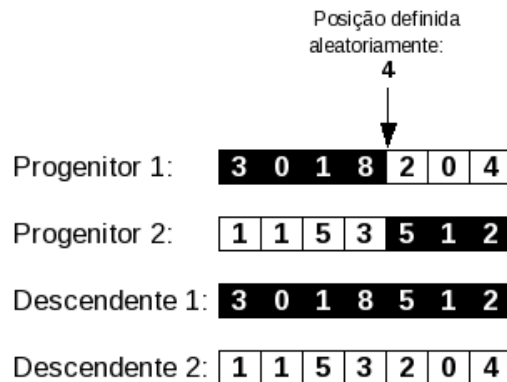


Figura 2.1: Reprodução através do cruzamento de genes entre dois progenitores

resultados que não se estendem à aplicação da solução a outros dados.

Uma solução que desafia o *overfitting* ou que tenta tirar partido dele foi a preconizada em [28], por Yao e Tan. Os algoritmos de evolução vão sendo executados em tempo real para manterem, nos períodos *out-of-sample* mais próximos, uma muito boa performance, conseguindo assim um resultado excepcional, no longo prazo.

### 2.3 Aplicação de algoritmos genéticos aos mercados de divisas

Os algoritmos genéticos têm vindo a ser utilizados na área dos mercados financeiros, por diversos autores, pelo que interessa conhecer algumas destas práticas.

Em [23], Papadamou e Stephanides, concluem que o número de indivíduos da população é um factor com grande impacto ao nível da evolução do desempenho do sistema. Pelo contrário, a probabilidade de haver cruzamento entre indivíduos não é um factor chave quanto àquela evolução.

Uma discussão sobre probabilidades de cruzamento e

mutação a adoptar num algoritmo genético pode ser encontrada em [20]. Neste artigo, Nuñez-Letamendia et al. analisam a influência dos parâmetros de controlo na robustez de um GA. Entre estes parâmetros, são estudadas as probabilidades de cruzamento e mutação. Ainda segundo Nuñez-Letamendia et al., os melhores resultados, em termos de robustez de um GA aplicado a séries temporais de dados financeiros, foram obtidos com a combinação de probabilidades altas de cruzamentos com probabilidades baixas de mutação, ou com probabilidades baixas de cruzamentos e probabilidades altas de mutação, o que faz sobressair uma complementaridade entre as operações de cruzamento e de mutação.

Lin et al., em [16], utilizam um algoritmo genético para descobrir a melhor relação de parâmetros que permitam tomar decisões de negociação, no mercado de acções. Por seu turno, em [12], Kaboudan dá conta do seu trabalho destinado a obter previsões de cotações para o dia seguinte, em acções, com o uso de programação genética. Para o efeito, usou igualmente modelos de regressão. Em [7], Dempster e Jones desenvolveram um sistema de negociação para o par USD/GBP que consiste num conjunto de regras, baseadas em combinações de diferentes indicadores, aplicadas a várias séries temporais de frequência intra-diária. Em [2], Alvarez-Diaz e Alvarez, usam um GA para encontrar uma função analítica que melhore a previsão das cotações no mercado de divisas, em relação aos resultados expectáveis pela teoria do “*random walk*”<sup>11</sup>. Em [4], Brabazon e O’Neill usaram um GA para a descoberta de regras de negociação com base em indicadores técnicos relativos a valores de fecho diário, no mercado de

---

<sup>11</sup>É uma teoria financeira para a qual as cotações dos mercados de acções evoluem de acordo com um processo estocástico, conhecido como “passeio aleatório”. Uma das consequências desta teoria é que os preços futuros não podem ser previstos.

Forex.

Allen e Karjalaine, em [1], fazem evoluir uma população de soluções usando árvores de decisão com um máximo de 100 nós e dez níveis de variáveis e indicadores técnicos, no sentido de descobrir regras de negociação e valores de parâmetros para negociação do índice S&P 500,<sup>12</sup> no período compreendido entre 1928 e 1995. A partir de uma população de 500 indivíduos, é dada a possibilidade ao sistema de construir uma árvore de regras sendo cada uma destas avaliada como se de uma função se tratasse. Outros trabalhos, como, por exemplo, [18] de Neely et al., usam estrutura semelhante.

Para além destes últimos modelos em árvore, outros autores optaram por uma estrutura representada por *strings* de *bits*<sup>13</sup>. Meagher et al., em [17], usam uma representação linear de uma *string* de *bits* para desenvolver a estrutura do modelo e para a determinação dos parâmetros. Brabazon e O'Neill, em [4], usam uma estrutura de modelo de negociação semelhante. Dempster e Jones, em [7], utilizam *strings* binárias de tamanho variável. Estes autores deixam ainda a ideia de usar os GA de uma forma pró-activa, isto é, reoptimizando periodicamente o modelo por forma a que este se vá adaptando à evolução do mercado.

## 2.4 Medição do desempenho

O que medir, em termos de desempenho de um sistema, depende do objectivo final de cada trabalho.

Kamruzzaman e Sarker, em [13], afirmam ter obtido uma correcta previsão da mudança direcciona dos preços em 80% das vezes, no seu trabalho que faz uso das ANN

---

<sup>12</sup>Índice que representa a evolução nas cotações de acções de 500 das maiores empresas a nível mundial.

<sup>13</sup>Por *string* de *bits* deverá entender-se um vector em que a cada coluna (*bit*) corresponde um parâmetro a ser determinado

no mercado de divisas. Contudo, em trabalhos futuros, ponderam a hipótese de terem em consideração o lucro como factor de avaliação do desempenho.

O lucro obtido, ou outra grandeza que tenha aquele como base de cálculo, tem sido usado nas investigações científicas, nesta área. Dunis et al., em [8] adoptam como medida de desempenho do modelo construído o ganho obtido. Li et al., em [15], dão conta de terem obtido uma rendibilidade bem maior do que a prevista pela hipótese de *random walk*. A negociação implica o uso de *spreads*<sup>14</sup> em três índices de futuros da bolsa de Taiwan com as posições a serem abertas e fechadas intra-diariamente.

A comparação dos resultados de uma estratégia *buy and hold*<sup>15</sup> do índice S&P 500 com outras baseadas em regras de negociação, foi alvo de estudo, por parte de Dempsey et al., em [5]. O algoritmo desenvolvido produz sinais, com base em regras de negociação, que podem resultar em uma de três acções a serem empreendidas pelo modelo: comprar, vender ou nada fazer. O valor do investimento a realizar em cada nova posição é variável, dependendo da potência dos sinais de entrada obtidos. O resultado final é 27% superior ao resultante da estratégia *buy and hold* comparável.

Apesar da consideração dos custos de transacção estar excluída das análises de desempenho de alguns modelos construídos por investigadores nesta área, outros há que os levaram em linha de conta. Allen e Karjalainen, em [1], concluem que o modelo encontrado não gera um excesso de rendimento consistente em relação às estratégias de *buy and hold*, depois de considerados os custos de transacção. A conclusão obtida por Neely e Weller, em [19],

---

<sup>14</sup>Nesta acepção, *spread* diz respeito a um tipo de negociação que procura tirar partido de diferenças de valor temporárias entre dois ou mais activos para a obtenção de lucros.

<sup>15</sup>Estratégia de investimento, segundo a qual, um activo financeiro é adquirido para ser mantido em carteira por um período de tempo muito alargado.



é de que os resultados obtidos são sensíveis aos custos de transacção. Quando os custos de transacção são baixos, o GA e a previsão baseada em modelos lineares conseguem resultados significativos. Quando aqueles são altos, os autores são forçados a não excluir a hipótese da eficiência dos mercados.

São, portanto, várias as formas estudadas, em trabalhos científicos, de tentar medir e comparar o desempenho do investimento em activos financeiros. A mais utilizada e aceite na comunidade científica será, contudo, a de Sharpe, que na sua obra [26] de 1966, avaliou o desempenho de fundos de investimento de acordo com uma medida, *Reward-to-variability ratio*<sup>16</sup> que mais tarde ficou conhecida como o rácio de Sharpe. Embora simples, esta medida teria as vantagens de ter um corpo teórico a sustentá-la; de possibilitar a avaliação de desempenho conhecendo-se a rendibilidade média do investimento e uma taxa de juro de referência, considerada sem risco, com a qual aquela se pudesse comparar; e de uma correcta comparação de investimentos entre diversas alternativas. No entanto, na época em que o rácio de Sharpe surgiu, os instrumentos financeiros derivados, que são hoje em dia negociados em grande volume, não existiam. Estes pressupõem, por norma, um grau de alavancagem<sup>17</sup> superior aos dos activos subjacentes, como o caso de acções.

Como a uma alavancagem superior dos investimentos está associado um risco também maior, surgiu a necessidade de implementar outras medidas que se pretenderam mais adequadas para avaliar o binómio lucro - risco. Surge assim o conceito de *maximum drawdown* (MDD) que cor-

---

<sup>16</sup>Este rácio mede o excesso de rentabilidade de uma estratégia de investimento sobre uma taxa de juro considerada sem risco, dividido pelo risco do activo, durante um determinado período.

<sup>17</sup>É uma estratégia de gestão de fundos que consiste em investir mais recursos do que os disponíveis para aumentar seu lucro, o que amplia os riscos da operação.

responde à maior quebra acumulada no lucro.<sup>18</sup> Este (e alguns rácios que dele fazem uso) é um indicador bastante popular nos meios financeiros, para aquilatar do risco associado a uma estratégia de investimento. Na comunidade científica, a sua importância tem vindo a notar-se, igualmente. Em [24], Pictet et al. fazem notar que o sucesso dos GA aplicados ao mundo financeiro está dependente da forma como é calculada a medida de *fitness* de cada indivíduo. Ainda segundo estes autores, os testes efectuados indicam que deverão tomar em consideração os períodos de *drawdown* (DD) na elaboração do modelo de negociação. Em [4], Brabazon e O'Neill adoptam uma metodologia para avaliar o desempenho que tem em conta o risco de vir a incorrer numa perda muito significativa. Comparam a rentabilidade ao maior valor acumulado de perda, potencial ou efectiva, quando estão tomadas posições. Incorporam três variáveis na função de avaliação de desempenho (*fitness*) que dirime sobre o mérito de cada sistema testado:

$$Fitness = ExcessReturn^{19} - x \times (MDD)$$

.  $x$  é considerado um factor do grau de aversão ao risco de cada investidor que deseja evitar perdas muito significativas, no seu investimento. Em [7], Dempster e Jones têm como primeiro objectivo desenvolver um sistema que permita a obtenção consistente de lucros supra-normais. Para a selecção das melhores estratégias, é usado o conceito de *drawdown*, no rácio modificado de Stirling.

O rácio de Stirling representa o quociente entre o lucro e o MDD, ou seja:  $raciodeStirling^{20} = \frac{lucro}{MDD}$ . O

---

<sup>18</sup>Um *drawdown* corresponde à diferença entre um pico de lucro atingido e o valor mais baixo que se lhe segue. O *maximum drawdown* corresponde à maior dessas diferenças, medidas a partir de cada pico intermédio de lucro alcançado, numa determinada estratégia de investimento.

<sup>19</sup>Lucro obtido.

<sup>20</sup>Apesar de o rácio de Stirling aqui indicado ser idêntico ao quociente usado nesta disser-

rácio modificado de Stirling (S) usado pelos autores é:  $S = \frac{Return}{1+ModDD}$ , com  $ModDD = \max(DD, 2\% \text{ da posição corrente})$ . Com o rácio modificado de Stirling pretende-se evitar que pequenas variações na variável DD, quando esta apresenta valores pequenos, tenha grandes reflexos no valor final do rácio. No entanto, quando os resultados de um sistema são negativos, são apenas usados os valores do prejuízo, na função de avaliação de desempenho, pois, consideram os autores, não teria sentido usar aquele rácio.

Diversos autores criticam a adopção do MDD como medida de risco. As suas posições poderão sintetizar-se nas de Harding et al., em [11]. Neste artigo, os autores tentam mostrar algumas das desvantagens do uso de MDD como medida do risco associado a investimentos financeiros e como aferição da qualidade histórica de estratégias ou gestores. Como desvantagens, apontam três: a) O MDD representa um único número resultante de um conjunto de dados ao qual estará associado um grande erro estatístico. A extrapolação do desempenho futuro usando o MDD histórico estará, assim, bastante susceptível a erros, pelo que não será, por isso, um indicador útil. b) Os MDD serão tão maiores quanto maior for a frequência de medição dos mesmos, uma vez que os autores consideram que a rentabilidade dos investimentos financeiros é avaliada, apenas, ao fim de um período. Assim, o MDD será maior numa série de dados diários do que numa série de dados semanais. O MDD de uma série de dados semanais será, por sua vez, maior do que uma série de dados mensais. c) O MDD será maior para séries mais longas de dados, pelo que gestores [ou sistemas] com registos de desempenho mais longos seriam penalizados. Como ponto positivo, no entanto, sa-

---

tação para avaliar o desempenho do algoritmo, outros autores, em outros trabalhos fazem referência ao mesmo rácio alterando, no entanto, a sua composição. Por este motivo, não se tomará pelo nome de rácio de Stirling o quociente de avaliação de desempenho utilizado neste trabalho.

lientam que o MDD se refere a uma realidade física, em comparação com realidades mais abstractas usadas noutro tipo de medidas, tais como a volatilidade.

## 2.5 Evolução histórica dos resultados obtidos

Em [14], LeBaron apresenta indícios fortes de mudanças no desempenho de sistemas que utilizam regras técnicas para a negociação no mercado de divisas, durante os anos 90. O bom desempenho que se fazia notar em tempos anteriores esbateu-se nos anos mais recentes. Há duas hipóteses que o autor estuda que podem, segundo ele, explicar esta evolução: serem causas relacionadas com *data snooping*<sup>21</sup> ou serem de cariz económico mais profundo. O autor acaba por concluir que o seu trabalho dá suporte à ideia de que se verificou, de facto, uma mudança nas séries temporais, entre as mais antigas e as mais recentes. No entanto, adverte que são necessários estudos mais aprofundados, com outras séries.

Ohlson, em [21], faz notar que muitos estudos reportavam resultados fora do normal quando aplicados a dados até meados dos anos 80, enquanto que estudos mais recentes admitem um menor sucesso das regras de negociação com base em indicadores técnicos. Neste trabalho, o autor constata que o lucro resultantes da utilização de regras de negociação, ajustado em termos de risco, decresceu durante o período de 1971 a 2000 e lança a hipótese de as ineficiências do mercado, que antes permitiam melhores resultados, poderem ter sido, apenas, temporárias.

Tal como Elton e Gruber afirmam em [9], “nunca poderemos provar a afirmação de que resultados do passado

---

<sup>21</sup>O Data Snooping consiste na selecção de regras óptimas de negociação *in-sample*, que poderão fazer concluir pela sua aplicabilidade geral, mas que falham quando aplicadas em períodos *out-of-sample*.

não podem ser usados para prever resultados futuros, uma vez que existe um número infinito de maneiras de usar as sequências de preços antigas para serem usadas na previsão de preços futuros. Tudo o que podemos fazer é testar modos particulares de combinar os dados dos preços passados para prever resultados no futuro”.

## Capítulo 3

# O Mercado e as séries temporais

### 3.1 Forex

As séries temporais que foram utilizadas neste trabalho provêm do mercado “Forex”. Forex, ou simplesmente FX, são acrónimos para “FOReign EXchange”, e representam o mercado global de negociação de divisas. Os seus participantes são variados, desde os investidores particulares, a bancos centrais, hedge funds, bancos comerciais, empresas que necessitam de cobrir a sua exposição cambial e fundos de investimento.

A negociação no mercado Forex tem por base as taxas de câmbio. Uma taxa de câmbio refere-se ao número de unidades de uma moeda necessárias para a compra de uma unidade de outra, ou o valor de uma moeda em termos de outra. Actualmente, as taxas de câmbio são de regime flutuante, mudando permanentemente. As taxas de câmbio estão cotadas em termos de pares de moedas. A moeda indicada em primeiro lugar é conhecida como sendo a moeda base e a segunda como a moeda de contrapartida. No caso do EUR/USD, a taxa de câmbio deste par indicará, para o momento, a quantos USD equivalerá uma unidade da moeda base, EUR. Numa operação de compra, uma

taxa de câmbio mostra quanto se terá de pagar na moeda de contrapartida para obter uma unidade da moeda base. No caso de uma venda, a taxa de câmbio indica quanto se obterá na moeda de contrapartida pela venda de uma unidade da moeda base. A cotação do EUR/USD, no mercado de Forex, faz uso de quatro casas decimais. Cada  $\frac{1}{100}$  de 1% da cotação do Euro face ao USD é comumente designado por um *pip* - “Percentage in point”, termo que é igualmente aplicado à negociação de outros pares de divisas. Por exemplo, dos 1,3000 aos 1,3025, diz-se que o par EUR/USD registou uma subida de 25 *pips*.

Segundo dados do Bank for International Settlements<sup>1</sup> (BIS), por dia, globalmente, são negociados, apenas no mercado à vista (*spot*), o equivalente a mais de um bilião<sup>2</sup> de USD. O crescimento do mercado tem sido forte nos últimos anos, não só na vertente *spot*, mas igualmente em contratos *forward*<sup>3</sup>, *swaps*<sup>4</sup> e opções<sup>5</sup>. Segundo os autores do relatório, a este crescimento não é alheia a generalização de plataformas electrónicas de negociação de divisas. A negociação nos mercados cambiais está, agora, ao dispor de qualquer investidor individual.

Neste trabalho, apesar da diversidade de moedas transaccionadas internacionalmente, optou-se por estudar o par EUR/USD, por ser o mais transaccionado a nível mundial e também por dizer respeito à moeda em curso no território nacional. Segundo dados do mesmo relatório do BIS, em Abril de 2007, o volume de negociação do par

---

<sup>1</sup>Relatório publicado em <http://www.bis.org/publ/rpfx07t.pdf>.

<sup>2</sup>Medida europeia.

<sup>3</sup>Um contrato *forward* representa um acordo feito entre duas partes para a compra ou venda de um bem numa data futura. O preço é decidido no momento da celebração do contrato.

<sup>4</sup>Um contrato entre duas partes, segundo o qual, cada parte deverá fazer e receber pagamentos da outra. Os respectivos montantes são determinados com o recurso a diferentes critérios ou fórmulas.

<sup>5</sup>Um instrumento financeiro que proporciona a quem o compra o direito, mas não a obrigação, de comprar (opção de compra) ou de vender (opção de venda) o bem subjacente, a um preço determinado, durante (ou apenas no final de) um período definido.

EUR/USD correspondia a 27% do total do mercado.

O enorme valor de transacção do EUR/USD no mercado Forex é bastante superior ao da negociação de contratos de futuros do EUR/USD em bolsas regulamentadas de produtos derivados. No entanto, convém frisar que a vasta maioria do volume enunciado resulta da negociação directa entre bancos, pelo que não corresponde a volume de negociação a nível do retalho, disponível para o público em geral.

Face aos mercados accionistas e seus derivados, o mercado Forex tem as vantagens de uma enorme liquidez, de ser negociável a qualquer hora do dia<sup>6</sup> e de ser possível alavancar os investimentos em algumas centenas de vezes. Todas estas vantagens tornaram muito popular este mercado descentralizado, em que a negociação realizada por pequenos investidores é feita, geralmente, com um *market maker*, seja um banco ou uma corretora.

O mercado Forex pressupõe um espaço virtual que não corresponde a um mercado centralizado, que não é regulamentado ou é-o fracamente. A transparência do mercado de Forex é posta em dúvida, para além destes factores, por nem todos os intervenientes estarem em posição de igualdade de oportunidades. Uma transacção no mercado de Forex envolve directamente duas partes. Cada parte compra uma determinada moeda por contrapartida da venda de outra divisa, assumindo cada parte o interesse oposto da outra. No entanto, os investidores particulares não têm acesso ao melhor preço do mercado, em cada momento. A negociação que efectuam é feita com intermediários financeiros, brokers, entre eles próprios e os bancos, responsáveis pelo maior volume de transacções no mercado. É dado como certo, que alguns destes intermediários, funcionam

---

<sup>6</sup>Excluindo o fim-de-semana, embora neste caso possa ainda ser possível negociar com alguns *market makers*, embora com custos mais elevados.



como *market maker*, assumindo, por vezes, eles próprios, a posição contrária à do investidor, lucrando com as perdas deste e vice-versa. A esta situação, talvez não seja alheio o facto, não confirmado em estatísticas publicadas, mas dado como certo, de que cerca de 80 a 90% dos investidores perdem ao negociar neste ou noutros mercados.

O *spread* disponível para a negociação dos investidores particulares concorre também para o lucro dos intermediários financeiros, como forma de compensação por estes proporcionarem liquidez ao mercado em cada momento. O *spread* corresponde à diferença entre o preço de compra e o de venda, em cada instante, podendo sofrer variações de valor em alturas de menor liquidez, como as primeiras horas de abertura semanal do mercado, e, também, em breves momentos de intensa negociação que se segue à divulgação de acontecimentos com mais impacto. O preço de compra (*bid*) indica o valor mais alto a que um determinado comprador está disposto a pagar o activo, naquele momento. O preço de venda (*ask*) é relativo ao valor mais baixo a que um vendedor está disposto a vendê-lo.

Dadas as características acima mencionadas, não é de estranhar, no Forex, diferenças na cotação do mesmo par de divisas entre bancos e brokers diferentes a cada momento. O preço verdadeiro da transação de um par de moedas nunca é conhecido, visto não haver uma bolsa centralizada em que aquele seja determinado. Não é invulgar haver desvios significativos dos valores de cotação de um mesmo par, entre várias instituições financeiras diferentes.

Devido ao carácter descentralizado deste mercado, é impossível obter dados certos do volume transaccionado.

## 3.2 Séries temporais

Os dados disponíveis, relativos aos anos de 2004, 2005, 2006 e 2007, foram fornecidos pela corretora<sup>7</sup> “Gain Capital”<sup>8</sup>.

Pelas razões apontadas na secção 3.1, os resultados e conclusões apresentados apenas são válidos para as cotações do EUR/USD tal como disponibilizadas pela corretora acima identificada. A sua validação com a utilização de cotações fornecidas por outra corretora ou relativa ao mercado de futuros do EUR/USD é uma incógnita.

As cotações obtidas pressupõem um *spread* de 0,0003 EUR/USD, ou seja 3 *pips*, que é o valor mais comum de entre os disponíveis para o pequeno investidor. Neste trabalho, o *spread* é considerado como sendo sempre constante.

É de duvidar que haja séries com comportamentos mais bem formados do que outras, ou seja, mais indicadas para o treino de um sistema. A aparente regularidade em excesso poderá indicar, apenas, que as séries temporais são curtas demais para abarcar uma diversidade de situações, que permita a obtenção de resultados satisfatórios em testes *out-of-sample*. Desta forma, as séries temporais não foram alvo de uma escolha selectiva de acordo com a sua forma aparente, nem foram excluídos quaisquer períodos que pudessem corresponder a picos ou fundos significativos.

Uma decisão importante a tomar, relativamente às séries temporais a usar, diz respeito à frequência de dados a utilizar. Neste trabalho, optou-se por escolher séries

---

<sup>7</sup>Apesar de a empresa mencionada ser aqui intitulada de empresa de corretagem, tal designação não é a mais correcta. É público que empresas que possibilitam a negociação de divisas no Forex a clientes, em geral, não actuam como intermediárias nas transacções entre estes e um mercado. Nestes casos, estas empresas podem assumir a parte contrária das transacções efectuadas pelos seus clientes, com todos os riscos daí inerentes.

<sup>8</sup>Sítio em [www.gaincapital.com](http://www.gaincapital.com).

temporais de base entre um minuto e uma hora, mais concretamente, séries de 1, 5, 15 e 60 minutos.

Depois de obtidos os dados *pip* a *pip* estes foram objecto de tratamento, que consistiu num teste à existência de picos ou fundos estranhos, nos valores das cotações, tendo-se fixado um limite de 1% de variação entre um *pip* e o imediatamente anterior. Não se verificou qualquer anomalia. Posteriormente, os valores dos *pips* foram agrupados em períodos de 1, 5, 15 e 60 minutos, tendo sido construídas as quatro séries temporais correspondentes.

Por exemplo, para um determinado período de uma série temporal, teremos:

data	hora	max	min
2007-11-30	14:57:00	1,4645	1,4639

Quadro 3.1: Exemplo com os dados relativos à cotação de um período de 1 minuto

O valor correspondente ao campo “max” representa o valor máximo a que foi transaccionado o par EUR/USD, no dia 2007-11-30, no minuto entre as 14h 57m 00s inclusive e as 14h e 58m 00s exclusive. O campo “min” corresponde ao mínimo transaccionado naquele mesmo intervalo de tempo.

Para cada uma das quatro séries base consideradas, foram calculados os valores das médias móveis exponenciais das cotações, período a período. A antecipação deste cálculo, em relação ao tempo de execução do algoritmo construído, visa incrementar a sua rapidez.

Para o estudo das diferentes séries temporais, procurou-se achar períodos de tamanho equivalentes, embora de dimensão significativa, quer para a parte da evolução da população, quer para a dos testes.

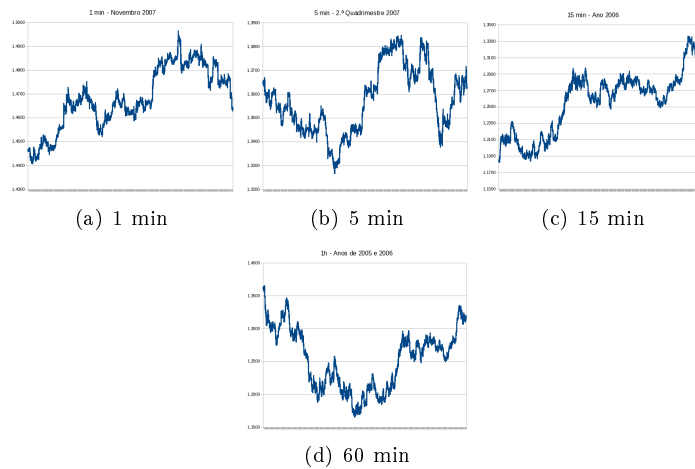


Figura 3.1: Evolução das cotações nas séries de evolução primária

### 3.2.1 Séries usadas para evolução

No que respeita à evolução das populações, foram usadas, por quatro vezes cada qual, as séries indicadas na tabela 3.2. Estas séries serão chamadas de evolução primária<sup>9</sup>.

Frequência	Tabela na base de dados	Espaço Temporal Correspondente	N.º períodos
1 min	eurusd_1m_2007_11mes	Mês de Novembro de 2007	28460
5 min	eurusd_5m_2007_2quad	2.º quadrimestre de 2007	25041
15 min	eurusd_15m_2006	Ano de 2006	24664
60 min	eurusd_1h_2005_2006	Anos de 2005 e 2006	12559

Quadro 3.2: Séries utilizadas pela evolução primária

Na figura 3.1 pode visualizar-se o comportamento das cotações em cada série. As séries de 1, 5 e 60 minutos possuem um maior equilíbrio entre momentos de subida e descida das cotações, alternando subidas com descidas fortes. A série de 15 minutos apresenta um padrão de subida alternado com períodos de indefinição de tendência.

O GA foi ainda utilizado para fazer evoluir as populações nas séries indicadas na tabela 3.3, por uma vez, cada qual. Esta evolução será designada de secundária.

Na figura 3.2 pode observar-se o comportamento des-

<sup>9</sup>Por motivo de escassez de dados, não foi possível conseguir, para a frequência de 60 minutos, um número de períodos equivalente aos das restantes séries.

Frequência	Tabela na base de dados	Espaço Temporal Correspondente	N.º períodos
1 min	eurusd_1m_2007_10mes	Mês de Outubro de 2007	27858
5 min	eurusd_5m_2007_1quad	1.º quadrimestre de 2007	24242
15 min	eurusd_15m_2005	Ano de 2005	24713
60 min	eurusd_1h_2004_2005	Anos de 2004 e 2005	12417

Quadro 3.3: Séries utilizadas pela evolução secundária

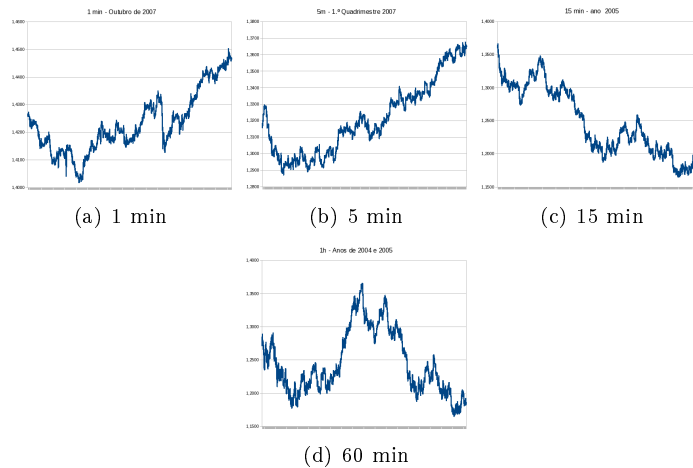


Figura 3.2: Evolução das cotações nas séries de evolução secundária

tas séries. O gráfico relativo à série de 15 minutos revela um comportamento de quebra nos preços, ao passo que as de 1 e 5 minutos apresentam uma predominância de uma tendência de crescimento de preços. O gráfico relativo às cotações horárias mostra subidas e descidas acentuadas nas cotações.

### 3.2.2 Séries usadas em testes

Na tabela 3.4 apresentam-se as séries utilizadas para testar as soluções encontradas.

Em termos gráficos, interessa conhecer como se comportaram estas séries temporais, agrupadas em termos de periodicidade.

As séries de 1 min, ver figura 3.3, têm a forma mais irregular de todas, à semelhança da série de treino primária com a mesma periodicidade. Não será alheio a este facto

Frequência	Tabela na base de dados	Espaço Temporal	N. de períodos
1m	eurusd_1m_2007_12mes	Dezembro 2007	22653
1m	eurusd_1m_2007_49sem	49.º Semana 2007	6397
5m	eurusd_5m_2007_3quad	3.º Quadrimestre 2007	23940
5m	eurusd_5m_2007_9mes	Setembro 2007	5876
15m	eurusd_15m_2007	Ano 2007	24764
15m	eurusd_15m_2007_1trim	1.º Trimestre 2007	6218
15m	eurusd_15m_2007_1mes	Janeiro 2007	2222
1h	eurusd_1h_2004	Ano 2004 <sup>10</sup>	6224
1h	eurusd_1h_2006	Ano 2006 <sup>11</sup>	6266
1h	eurusd_1h_2007	Ano 2007	6278
1h	eurusd_1h_2007_1trim	1.º Trimestre 2007	1633
1h	eurusd_1h_2007_1mes	Janeiro 2007	632

Quadro 3.4: Séries utilizadas em testes

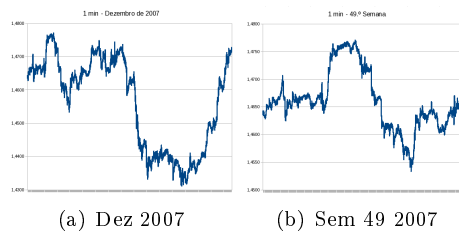


Figura 3.3: Séries de testes com periodicidade de 1m

o “ruído” existente na negociação de periodicidade muito baixa, tal como antes referido.

As séries de testes de 5 min, presentes na figura 3.4 são mais parecidas, na forma, entre si, denotando uma tendência de subida dos preços que contrasta, de sobremaneira, com a série de evolução primária respectiva, sendo, no entanto, similar à série de evolução secundária.

As séries de testes de 15 minutos, ver figura 3.5, possuem igualmente comportamentos diversos entre si. O ano de

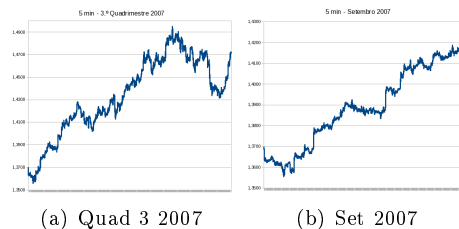


Figura 3.4: Séries de testes com periodicidade de 5m



Figura 3.5: Séries de testes com periodicidade de 15m

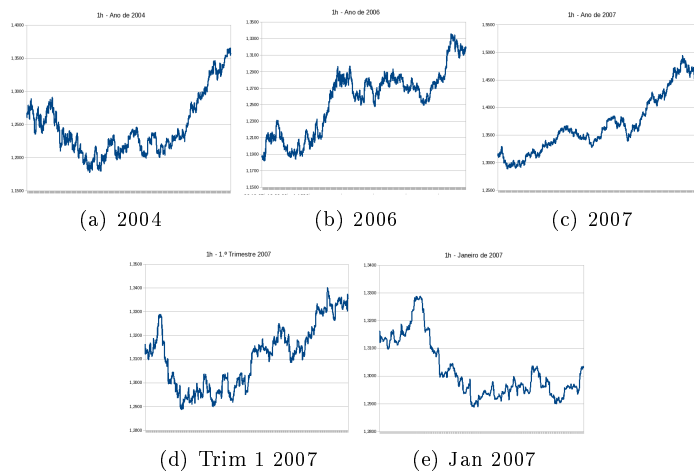


Figura 3.6: Séries de testes com periodicidade de 60m

2007 apresenta um comportamento tendencialmente crescente. O primeiro trimestre do mesmo ano caracteriza-se por um período inicial de baixa seguido de um intervalo constante de variação das cotações, sendo estas crescentes na parte final. Em Janeiro de 2007, sobressai uma descida inicial acentuada, seguida de um período de indefinição de uma tendência.

Das séries de dados de 1h presentes na figura 3.6 pode constatar-se que a forma das cotações do ano de 2007 é a que apresenta um padrão mais monótono, de crescimento. As relativas a 2004 e 1.º trimestre de 2007 variam entre períodos de quebra, indefinição de tendência e subida nos preços. O ano de 2006 caracteriza-se por um padrão de crescimento de preços alternado com períodos de indefini-

ção. Janeiro de 2007 revela descida dos preços alternada com períodos de indefinição de tendência.

A tabela 3.5 mostra todo o conjunto de séries usadas em testes *out-of-sample*, agrupadas por frequência de dados e por período de evolução. A coluna mais à direita representa o tipo da série testada, que será útil para a apresentação de resultados. Assim, a interpretação dos códigos ali inseridos é a seguinte:

- A** - série temporal anterior ao período de evolução respectivo - apenas disponível para os quatro indivíduos respeitantes à evolução primária.
- B** - série temporal imediatamente posterior ao final da série de evolução respectiva e que é equivalente a esta, em termos de extensão temporal, ressalva feita no que respeita à frequência de 60 minutos. Disponível tanto para os indivíduos treinados em séries de evolução primária como secundária.
- C** - série temporal imediatamente posterior ao final da série de evolução respectiva, mas de dimensão inferior a esta. No caso das frequências de 15 e 60 minutos, é apenas considerada a mais extensa série de testes das duas menores da mesma frequência. As séries deste tipo têm uma extensão de cerca de  $\frac{1}{4}$  das respectivas séries de testes consideradas no tipo B. Disponível apenas para indivíduos respeitantes a séries de evolução primária.



Frequência evolução	Tipo evolução	Período evolução	Frequência <i>out-of-sample</i>	Período <i>out-of-sample</i>	tipo
1m	Primária (4 vezes)	Novembro 2007	1m	Outubro 2007 Dezembro 2007 49.º Semana 2007	A B C
			5m	Setembro 2007	
			15m	Janeiro 2007	
	Secundária (1 vez)	Outubro 2007	1h	1.º Trimestre 2007	
			1m	Novembro 2007 Dezembro 2007 49.º Semana 2007	B
			5m	Setembro 2007	
5m	Primária (4 vezes)	2.º Quadrimestre 2007	5m	1.º Quadrimestre 2007 3.º Quadrimestre 2007 Setembro 2007	A B C
			1m	49.º Semana 2007	
			15m	Janeiro 2007	
	Secundária (1 vez)	1.º Quadrimestre 2007	1h	1.º Trimestre 2007	
			5m	2.º Quadrimestre 2007 3.º Quadrimestre 2007 Setembro 2007	B
			1m	49.º Semana 2007	
15m	Primária (4 vezes)	Ano 2006	15m	Ano 2005 Ano 2007 1.º Trimestre 2007 Janeiro 2007	A B C
			1m	49.º Semana 2007	
			5m	Setembro 2007	
	Secundária (1 vez)	Ano 2005	1h	1.º Trimestre 2007	
			15m	Ano 2006 Ano 2007 1.º Trimestre 2007 Janeiro 2007	B
			1m	49.º Semana 2007	
1h	Primária (4 vezes)	Anos 2005-2006	5m	Setembro 2007	
			1m	49.º Semana 2007	
			1h	Ano 2004 Ano 2007 1.º Trimestre 2007 Janeiro 2007	A B C
	Secundária (1 vez)	Anos 2004-2005	1h	Ano 2006 Ano 2007 1.º Trimestre 2007 Janeiro 2007	B
			15m	Janeiro 2007	
			5m	Setembro 2007	
1m	49.º Semana 2007				

Quadro 3.5: Séries usadas em testes *out-of-sample* agrupadas por período de evolução e período *out-of-sample*

## Capítulo 4

# Metodologia

*“The correctness of decisions is measured by two major criteria: adequacy of achieving the desired objective; and the efficiency with which the result was obtained.”*

(Herbert Simon)

### 4.1 Objectivo e pressupostos iniciais

O sistema construído visou determinar a possibilidade de obtenção consistente de lucros em períodos de teste, através de transacções com o par de divisas EUR/USD, no mercado de Forex. Procurou-se que o sistema aproveitasse tendências identificáveis de evolução de preços, com base em padrões existentes nos períodos anteriores aos de tomada de posições. Para esse fim, foi desenvolvido um algoritmo genético. O desenvolvimento do algoritmo obedeceu aos seguintes princípios: 1- Os dados das séries serem utilizados na sua totalidade, sem pôr de parte períodos específicos; 2- Os indicadores técnicos escolhidos serem construídos com base no valor das cotações; 3- As regras para entrada ou saída de posição serem simples e facilmente perceptíveis; 4- Em cada instante, haver, no máximo, uma

posição aberta; 5- O sistema desenvolvido ter em conta os custos de transacção, na forma do *spread*; 6- O desempenho de cada estratégia de negociação ser calculado com base na relação entre resultado obtido e risco incorrido, medidos em *pips*, cada um.

Pretendeu-se que o modelo de negociação construído fosse robusto. Com este objectivo em vista e o princípio da prudência como salvaguarda, a simulação decorreu, tanto quanto possível, de acordo com o que seria de esperar numa situação de negociação real. Em caso de incerteza, a opção tomada foi a mais penalizadora para o resultado final.

A escolha das séries temporais, com frequências de 1, 5, 15 e 60 minutos, foi efectuada em consonância com o objectivo de estudar o comportamento do algoritmo, quer em séries de dados em que o “ruído” e o “pulsar” do mercado imperam - caso das de 1, 5 e 15 minutos, quer numa em que a tendência dominante no par terá predominância - a de 60 minutos. O teste de cada estratégia de negociação foi efectuada em séries *out-of-sample* para averiguar da consistência da solução encontrada. A adição de testes em séries mais curtas do que as de treino, embora imediatamente subsequentes a estas, pretendeu averiguar a hipótese de o desempenho *out-of-sample* decair com a extensão do período de teste. As soluções encontradas foram, ainda, testadas em séries de outras frequências.

Neste trabalho, não foi ponderada a problemática relativa ao valor do investimento inicial ou ao montante correcto de investimento em cada posição aberta, para cada estratégia de negociação, o que é conhecido por *position sizing*. Em consequência, considerações sobre margens de negociação necessárias, impostas em situação de negociação real por uma corretora, também não foram levadas em

linha de conta.

## 4.2 Estrutura do algoritmo genético

Como foi exposto na introdução, o objectivo deste trabalho consistia no desenvolvimento de um sistema de negociação que fosse tão rentável quanto possível. O desafio subjacente a este desiderato implicou uma escolha cuidada quanto a uma heurística que se mostrasse particularmente adequada às características do problema. Tendo estes factos em consideração, foi escolhido um algoritmo genético como base de trabalho.

A população inicial, isto é, da primeira geração, é constituída por dez indivíduos criados de forma aleatória. Cada indivíduo constitui-se como uma estratégia de transacção incluindo em si todos os parâmetros relativos às dez regras de negociação consideradas no sistema de negociação, num total de trinta e um parâmetros. Os parâmetros relativos a cada uma das dez regras mostradas na figura 4.1 são aglutinados num vector de trinta e um elementos, sendo que, em termos biológicos, a cada elemento corresponde um gene e à designação de vector corresponde a de cromossoma. A figura 4.2 ilustra a composição de cada indivíduo, que se poderá designar igualmente por solução candidata à resolução do problema.

Cada indivíduo é sujeito a uma operação de medição do seu desempenho que consiste na determinação do lucro ou prejuízo finais e do risco associado à solução candidata. Para o efeito, a estratégia de transacção é submetida à negociação na série temporal considerada, período a período. A tomada de decisões quanto a ordens a emitir ou a modificar e a negócios a abrir ou a fechar só é possível, no entanto, se a informação presente em cada cromossoma

Regra 1	emac_periodos				
Regra 2	emal_bool	emal_periodos			
Regra 3	tr1_bool	tr1_periodos	tr1_percent	tr1_sinal	
Regra 4	tr2_bool	tr2_sinal	tr2_vezes	tr2_percent	tr2_periodos
Regra 5	p1				
Regra 6	sl1	tp1			
Regra 7	rap_bool	rap_percent	rap_tp	rap_sl	rap_periodos
Regra 8	len_bool	len_percent	len_tp	len_sl	len_periodos
Regra 9	mel_bool	mel_sl	mel_tp		
Regra 10	pio_bool	pio_sl	pio_tp		

Figura 4.1: Cromossoma utilizado

for decodificada. Assim, considerando uma leitura da esquerda para a direita no vector da figura 4.2 teremos que o primeiro gene corresponde ao parâmetro relativo ao número de períodos a considerar na EMA de curto prazo do sistema de negociação - regra 1. O segundo e o terceiro parâmetros dizem respeito à regra 2. O segundo é decodificado como um parâmetro do tipo lógico que assinala a utilização ou não da regra 2 na negociação. O terceiro representa o número inteiro de períodos a considerar no cálculo da EMA de longo prazo. O conjunto do quarto ao sétimo genes diz respeito à regra 3. O quarto gene presta uma função idêntica, nesta regra, ao segundo gene na regra 2. Os quinto e sexto genes representam números inteiros utilizados nos cálculos requeridos pela regra. O sétimo elemento tem um universo de apenas dois valores possíveis, constituído pelo conjunto  $\{<, >\}$ . Do oitavo ao trigésimo primeiro gene, é definido, para cada estratégia de transacção, se as regras em causa (4, 7, 8, 9 e 10) são de aplicar na negociação e os valores dos vários parâmetros pertencentes a cada uma das regras (da 4 à 10).

Os indivíduos da geração inicial competem entre si por

lugares na geração seguinte. A avaliação do desempenho efectuada resulta numa hierarquização dos indivíduos, cabendo aos mais aptos mais probabilidades de cruzamento com parceiros e a conseqüente mais numerosa geração de descendentes. O cruzamento é efectuado tendo por base os cromossomas de dois progenitores e uma posição determinada aleatoriamente. Esta operação dá origem apenas a um descendente aproveitando a primeira parte do vector do primeiro progenitor e a segunda do segundo. No entanto, se as várias estratégias de transacção apresentarem resultados que convirjam significativamente, os 50% de indivíduos com piores desempenhos são descartados e já não participam na operação de cruzamento. Neste caso são adicionadas outras tantas soluções candidatas geradas de forma aleatória, designadas por imigrantes, que farão parte directamente da geração seguinte. A melhor estratégia é sempre escolhida para a geração seguinte, uma característica habitualmente designada por elitismo. Os indivíduos que provêm da operação de cruzamento são objecto da operação de mutação. Cada gene de cada cromossoma é sujeito a um teste probabilístico que determina a mutação, ou não, do gene em causa.

Após esta série de operações, a nova população formada é sujeita a nova avaliação de desempenho, fechando um ciclo de evolução que decorre durante 50 gerações.

### **4.3 Estratégias de entrada e saída de posição**

Cada indivíduo assume, para cada período, um de três estados: a) estar em posição neutra, pronto a emitir uma ordem de compra ou de venda; b) estar comprado, i.e., ter efectuado a compra da moeda base e a correspondente venda da moeda de cotação - pode designar-se por estar

REGRAS	1	emac_periodos	VECTOR (CROMOSSOMA)
	2	emal_bool	
		emal_periodos	
		tr1_bool	
	3	tr1_periodos	
		tr1_percent	
		tr1_sinal	
		tr2_bool	
		tr2_sinal	
	4	tr2_vezes	
		tr2_percent	
		tr2_periodos	
	5	p1	
	6	sl1	
		tp1	
		rap_bool	
		rap_percent	
	7	rap_tp	
		rap_sl	
		rap_periodos	
		len_bool	
		len_percent	
	8	len_tp	
		len_sl	
		len_periodos	
		mel_bool	
	9	mel_sl	
		mel_tp	
		pio_bool	
	10	pio_sl	
		pio_tp	

Figura 4.2: Vector de 31 parâmetros que constitui cada estratégia de transacção

numa posição “longa” (PL); c) estar com uma posição vendida, também chamada de “curta” (PC), na qual é vendida a moeda base e comprada, por contrapartida, a moeda de cotação.

Quando o sistema se encontra em posição neutra, o indivíduo determina, período a período, se se verificam as condições de emissão de uma ordem de entrada numa posição, seja ela de compra ou de venda. A ordem é dada de acordo com a tendência detectada no mercado, tal como aferida pela posição relativa entre o preço médio do período e o valor da média móvel de curto prazo respectiva. O indivíduo determina o uso de três regras adicionais que procuram identificar padrões na evolução das cotações. É ainda usado um filtro, que consiste na definição de um preço fixo para a abertura de uma posição.

Depois de assumida uma posição, a ênfase é posta na gestão da saída da posição. A saída de uma posição, anteriormente tomada, é efectuada em valores limites, previamente estabelecidos, quer no caso de um ganho, quer no de uma perda. No entanto, se as características do indivíduo assim o previrem e se a evolução das cotações assim o determinarem, os valores de saída são alterados. Esta decisão é tomada com o recurso a quatro indicadores adicionais.

#### **4.3.1 Regras de entrada de posição**

Num período em que o sistema de negociação se encontre em posição neutra, são avaliados os indicadores activos que dizem respeito à abertura de posição para um determinado indivíduo. No caso de todos estes retornarem um estado de “Verdadeiro”, é introduzida uma ordem de compra ou de venda no sistema a um preço fixo estabelecido. No decorrer do período seguinte, o preço de entrada da



posição é comparado com os valores das cotações. Se os valores entre si forem compatíveis, o sistema assume a entrada numa posição. Se até ao fim do período seguinte, tal compatibilidade não é possível, isto é, se o intervalo entre o máximo e o mínimo das cotações não contiver o preço fixado, a ordem anteriormente dada é cancelada. Neste caso, os indicadores são calculados uma vez mais, no período seguinte, e as condições para emissão de uma nova ordem de entrada numa posição são reavaliadas.

De modo a ser tomada uma decisão sobre a entrada numa posição comprada ou vendida, foi construído um conjunto de cinco regras. As regras 1 e 2 comparam o preço médio do período actual com o valor de médias móveis exponenciais de modo a aproveitar tendências de evolução dos preços. As regras 3 e 4 são baseadas em indicadores construídos e, tanto quanto se sabe, originais. A regra 3 compara o preço médio actual com o máximo e o mínimo do conjunto dos últimos períodos. A regra 4 analisa se o conjunto das últimas cotações forma uma zona de congestionamento de preços. A regra 5 faz depender a entrada numa posição da evolução do preço do momento seguinte em relação ao actual.

Os indicadores construídos retornam valores de “1” ou “0”, isto é, “Verdadeiro” ou “Falso”, respectivamente.

#### **Regra 1 - Média de Curto Prazo**

A primeira regra utilizada por cada indivíduo, para uma possível tomada de posição, pretende tirar partido das tendências de evolução dos preços. Assim, pretende-se que o sistema apenas tome posições compradoras quando o seu preço ( $P$ ) estiver acima do valor de uma média móvel exponencial de preços (EMA). Posições vendedoras só podem ser assumidas se o preço estiver abaixo da EMA escolhida.

Para um dado período, o preço é calculado como sendo o valor médio entre o máximo e o mínimo de um período. Tomando o exemplo do quadro 3.1, teríamos:

$$P = \frac{P_{max} + P_{min}}{2} = 1,4642$$

A fórmula utilizada para o cálculo da EMA, num determinado período, é a seguinte:

$EMA = P \times K + EMA_{-1} \times (1 - K)$ , sendo  $K = \frac{2}{N+1}$  e  $N$  o número de períodos considerado para calcular a EMA.

O parâmetro  $N$  é determinado pelo GA, sob a designação *emac\_periodos*, nesta regra. É um valor inteiro e tem como limite mínimo o valor de 3 e máximo o de 50. Nesta primeira regra, a EMA a considerar é referida como *EMAc*. O quadro apresenta o parâmetro usado pela regra e as condições necessárias à emissão de ordens longas e curtas.

Posição	Condição	Tipo condição	Parâmetro	Domínio	Tipo
longa	$P > EMAc$	necessária	emac_periodos	3-50	inteiros
curta	$P < EMAc$	necessária			

Quadro 4.1: Representação da regra 1

## Regra 2 - Média de Longo Prazo

A segunda regra compara igualmente  $P$  com o valor de uma EMA, no caso da *EMAl*. *EMAl* é por definição, uma média de periodicidade mais longa do que *EMAc*. A ideia subjacente à constituição desta regra é a de o GA só poder tomar posições quando o preço estiver em consonância com uma tendência de mais longo prazo das cotações. A utilização ou não desta regra é determinada por cada indivíduo. O parâmetro *emal\_bool* toma os valores de “Verdadeiro” ou “Falso”, correspondendo, respectivamente, à utilização, ou não, da regra para a decisão de tomada de

uma posição. O parâmetro *emal\_periodos*, que corresponde ao *N* referido na fórmula da regra anterior, define o número de períodos considerados para o cálculo de *EMAl*. O quadro apresenta, em resumo, os parâmetros usados por esta regra e as condições para que ela assuma o valor lógico “Verdadeiro”.

Posição	Condição	Tipo condição	Parâmetro	Domínio	Tipo	Restrição
longa	$P > EMAl$	facultativa	<i>emal_bool</i>	Verdadeiro;Falso	lógicos	
curta	$P < EMAl$	facultativa	<i>emal_periodos</i>	10-200	inteiros	$emal\_periodos > emac\_periodos$

Quadro 4.2: Representação da regra 2

### Regra 3 - Posição num Trading Range

Esta é uma regra de carácter opcional, sendo a sua utilização, ou não, definida, em cada indivíduo, pelo parâmetro *tr1\_bool*. Com esta regra, pretende-se que o GA tire partido da posição relativa do preço actual em relação ao máximo e mínimo dos últimos períodos. O número de períodos a considerar é determinado pelo próprio indivíduo, através do parâmetro *tr1\_periodos*. São igualmente determinados um valor de percentagem *tr1\_percent* e o sinal da desigualdade *tr1\_sinal*. O máximo, *tr\_max*, e o mínimo, *tr\_min*, dos últimos *tr1\_periodos* são calculados. A amplitude entre esses máximo e mínimo é dividida em duas partes, de acordo com o cálculo dos parâmetros *tr\_l* no que respeita a PL e *tr\_s* relativamente a PC.

$$tr\_l = tr1\_percent \times (tr\_max - tr\_min) + tr\_min$$

$$tr\_s = tr\_max - tr1\_percent \times (tr\_max - tr\_min)$$

Tanto na situação de PL, como de PC, há duas possibilidades a considerar em cada, conforme o valor de *tr1\_sinal*. Em PL: a) se o sinal *tr1\_sinal* é “>”, o indicador tomará o valor de “Verdadeiro” se  $P > tr\_l$ , e “Falso”, caso contrário; b) se o sinal *tr1\_sinal* for “<”, a condição assumirá

o valor de “Verdadeiro” se  $P < tr\_l$ , e “Falso” se não. Em PC: a) se o sinal  $tr1\_sinal$  é “>”, o indicador tomará o valor de “Verdadeiro” se  $P < tr\_s$ , e “Falso”, caso contrário; b) se o sinal  $tr1\_sinal$  for “<”, a condição assumirá o valor de “Verdadeiro” se  $P > tr\_s$ , e “Falso” se não.

O sinal de  $tr1\_sinal$ , quando for “>”, faz depender o valor lógico da condição da proximidade relativa do preço actual à extremidade mais favorável do intervalo das últimas  $tr1\_periodos$  cotações, de acordo com a tendência de preços evidenciada. Por seu turno, o sinal “<”, avalia a proximidade relativa do preço à extremidade mais desfavorável. O quadro apresenta, em resumo, os parâmetros usados por esta regra e as condições para que ela assuma o valor lógico “Verdadeiro”.

Posição	Parâmetro	Condição	Tipo condição	Parâmetro	Domínio	Tipo
longa	$tr1\_sinal :>$	$P > tr\_l$	facultativa	$tr1\_bool$	Verdadeiro;Falso	lógicos
longa	$tr1\_sinal :<$	$P < tr\_l$	facultativa	$tr1\_periodos$	4-25	inteiros
curta	$tr1\_sinal :>$	$P < tr\_s$	facultativa	$tr1\_percent$	1-99	inteiros
curta	$tr1\_sinal :<$	$P > tr\_s$	facultativa	$tr1\_sinal$	<;>	

Quadro 4.3: Representação da regra 3

#### Regra 4 - Repetição de cotações

É uma regra igualmente de carácter opcional, sendo a sua utilização ou não determinada pelo próprio indivíduo, utilizando para tal o parâmetro  $tr2\_bool$ . Esta regra visa dar relevância à possibilidade de tomada de decisões de entrada de posição em situações de congestionamento ou fluidez de preços. Para tal, é contado o número de períodos, de entre os últimos  $tr2\_periodos$ , cujo preço médio pertence a um intervalo de cotações definido com ajuda do parâmetro  $tr2\_valor$ . Essa contagem é comparada com o parâmetro  $tr2\_vezes$ , através do operador  $tr2\_sinal$ , resultando no valor lógico “Verdadeiro” ou “Falso”.

Para a aplicação desta regra, em primeiro lugar, é calcu-

lado o valor da amplitude máxima do intervalo de tempo constituído pelos últimos  $tr2\_periodos$  períodos, que corresponde à diferença entre o máximo dos máximos ( $MaxMax$ ) e o mínimo dos mínimos ( $MinMin$ ), daqueles períodos. a) no que respeita a PL, o procedimento é o seguinte: a.1) calcular um novo mínimo ( $novo\_m$ ), que corresponde a  $MinMin$  adicionado do produto da amplitude máxima pelo parâmetro  $tr2\_valor$ , em percentagem; a.2) em relação aos últimos  $tr2\_periodos$ , é calculado o número de vezes  $Nvezes$  em que o preço médio ( $P$ ) de cada período se situa no novo intervalo que tem como máximo  $MaxMax$  e como mínimo  $novo\_m$ ; a.3) se  $Nvezes > tr2\_vezes$  e  $tr2\_sinal = '>'$  ou se  $Nvezes < tr2\_vezes$  e  $tr2\_sinal = '<'$ , o indicador adquire o valor lógico “Verdadeiro” para a tomada de PL, caso contrário, assume o valor lógico “Falso”.

b) no que respeita a PC, o procedimento é o seguinte: b.1) calcular um novo máximo ( $novo\_M$ ), que corresponde a  $MaxMax$  subtraído do produto da amplitude máxima pelo parâmetro  $tr2\_valor$ , em percentagem; b.2) em relação aos últimos  $tr2\_periodos$ , é calculado o número de vezes  $Nvezes$  em que o preço médio ( $P$ ) de cada período se situa no novo intervalo que tem como mínimo  $MinMin$  e como máximo  $novo\_M$ ; b.3) se  $Nvezes > tr2\_vezes$  e  $tr2\_sinal = '>'$  ou se  $Nvezes < tr2\_vezes$  e  $tr2\_sinal = '<'$ , o indicador adquire o valor lógico “Verdadeiro” para a tomada de PC, caso contrário, assume o valor lógico “Falso”. O quadro apresenta, em resumo, os parâmetros usados por esta regra e as condições para que ela assumo o valor de “Verdadeiro”.

Posição	Parâmetro	Condição	Tipo condição	Parâmetro	Domínio	Tipo	Restrição
longa	$tr2\_sinal >$	$Nvezes > tr2\_vezes$	facultativa	$tr2\_bool$	Verdadeiro/Falso	lógicos	
longa	$tr2\_sinal <$	$Nvezes < tr2\_vezes$	facultativa	$tr2\_periodos$	4-25	inteiros	
curta	$tr2\_sinal >$	$Nvezes > tr2\_vezes$	facultativa	$tr2\_vezes$	3-24	inteiros	$\leq tr2\_periodos$
curta	$tr2\_sinal <$	$Nvezes < tr2\_vezes$	facultativa	$tr2\_percent$	50-99	inteiros	
				$tr2\_sinal$	< >		

Quadro 4.4: Representação da regra 4

### Regra 5 - Preço de entrada fixo

Esta regra pretende garantir a entrada numa posição a um preço certo. O GA começa por determinar o valor do parâmetro  $p1$ .  $p1$  é inicialmente calculado pelo GA como um valor inteiro compreendido entre os  $-20$  e os  $20$ . É dividido por  $10.000$ , representando centésimas de percentagem de aumento ou diminuição do preço. Para determinar o preço a que o sistema aceita abrir uma posição no período seguinte ( $E$ ), a percentagem  $p1$ , positiva ou negativa, é adicionada ao valor do máximo do período anterior, no caso de uma PL, ou seja,  $E = P_{-1}^{max} \times (1 + p1)$ . No caso de uma PC, a percentagem  $p1$  é subtraída ao mínimo do período anterior:  $E = P_{-1}^{min} \times (1 - p1)$ .

Uma ordem de entrada de posição só é válida para o período seguinte ao da emissão. O quadro apresenta, em resumo, o parâmetro usado por esta regra e as fórmulas de cálculo do preço fixado para a abertura de uma nova posição.

Posição	Condição	Tipo condição	Parâmetro	Domínio	Tipo
longa	$P = P_{-1}^{max} \times (1 + p1)$	necessária	p1	(-20)-20	inteiros
curta	$P = P_{-1}^{min} \times (1 - p1)$	necessária			

Quadro 4.5: Representação da regra 5

### 4.3.2 Regras de saída de posição

Na altura em que é emitida uma ordem para uma possível entrada numa posição, são definidos, igualmente, os valores a que a posição pode ser fechada com lucro ou com

prejuízo. Dada que a variação nas cotações do EUR/USD pode ser significativa, optou-se por fixar os eventuais preços de saída de uma posição em termos de ganho ou perda relativa, em relação ao valor de entrada. Esta tarefa é atribuída à regra 6.

De acordo com o concluído por Schulmeister, em [25], e enunciado no início do capítulo 2, pretende-se que cada indivíduo não deixe aumentar desmesuradamente situações de perda, ao mesmo tempo que se procura estudar a hipótese de o sistema ir acumulando ganhos para além do que tinha sido inicialmente previsto. Assim, ficam definidas quatro regras adicionais, cuja utilização, ou não, será decidida pelo próprio indivíduo, e que poderão redefinir os valores de saída de uma posição.

A regra 7 - “Evolução Rápida” - pretende maximizar lucros quando uma determinada posição evolui favoravelmente num curto espaço de tempo, ao mesmo tempo que procura resguardar lucros ou diminuir a perda potencial pelo ajustamento da ordem de *stop loss*.

Com a regra 8 - “Evolução Lenta” - procura-se forçar a saída de uma posição que demore a evoluir favoravelmente. Dado que, como se disse na secção 4.1, o sistema só permite a existência de uma posição aberta de cada vez, o GA decidirá sobre o mérito de encerrar posições deste tipo de modo a permitir a abertura de outras eventualmente mais lucrativas.

O objectivo da regra 9 - “Melhoria de Cenário” é semelhante ao da 7 e o da regra 10 - “Pioria de Cenário” é semelhante ao da regra 8. No entanto, estas regras fazem depender a sua actuação da verificação de alteração do posicionamento relativo entre o preço médio do período e a EMA1 no momento de tomada de posição, e nos momentos subsequentes.

### Regra 6 - Take Profit e Stop Loss

Na altura em que é formada uma ordem de entrada numa posição, o GA define igualmente os valores da sua saída, de acordo com os genes de cada indivíduo. Numa PL, o parâmetro  $tp1$  corresponde às centésimas percentuais de acréscimo sobre a cotação de entrada na posição, de forma a estabelecer um preço de recolha,  $tp$ , de um possível lucro. Numa PC, a percentagem  $tp1$  é subtraída ao preço de entrada de forma a achar o valor de  $tp$ .  $sl1$  corresponde à perda relativa máxima, em centésimas percentuais, a que a posição esta sujeita, de início. Numa PL, a percentagem  $sl1$  é subtraída ao preço de entrada na posição, enquanto que numa PC, é adicionada. O quadro apresenta, em resumo, os parâmetros usados por esta regra e as fórmulas de cálculo do preço inicialmente fixado para recolha de ganho ou assunção de perda.

Posição	Condição	Tipo condição	Parâmetro	Domínio	Tipo
longa	$tp = precodeentrada \times (1 + tp1)$	necessária			
longa	$sl = precodeentrada \times (1 - sl1)$	necessária	tp1	1-100	inteiros
curta	$tp = precodeentrada \times (1 - tp1)$	necessária	sl1	1-100	inteiros
curta	$sl = precodeentrada \times (1 + sl1)$	necessária			

Quadro 4.6: Representação da regra 6

### Regra 7 - Evolução Rápida

Com esta regra, pretende-se tirar partido da possibilidade de o indivíduo ser capaz de garantir uma parte dos lucros entretanto obtidos numa posição, por um lado, ou da de poder expandir a perspectiva de lucro inicialmente proposta, por outro, no caso de uma evolução favorável nas cotações que tenha ocorrido de uma forma rápida. Entre os valores das cotações do período actual e de  $rap\_periodos$  períodos anteriores, é verificado se o preço médio se moveu favoravelmente mais de  $rap\_percent$  de centésimas de



ponto percentual. Se isso não aconteceu, a regra é abandonada. Se aquela condição foi verificada, o valor anterior de  $tp$  será afastado  $rap\_tp$  centésimos percentuais do ponto anterior de tomada de lucro. Da mesma forma, o valor anterior definido de perda máxima é reduzido, sendo o novo valor para a saída de posição, em caso de movimento adverso de cotações, aproximado do preço médio do período actual, em  $rap\_sl$  centésimas percentuais sobre o valor definido para a *stop loss*.

No caso de tomada de uma PL, o novo valor de  $tp$  é calculado como  $tp = tp_{anterior} \times (1 + rap\_tp)$  e o novo valor de  $sl$  é  $sl = sl_{anterior} \times (1 + rap\_sl)$ . No entanto, se  $sl$ , assim calculado, for maior do que o mínimo do período actual, o novo valor de  $sl$  é igual àquele diminuído de um  $pip$ . Em PC,  $tp = tp_{anterior} \times (1 - rap\_tp)$  e  $sl = sl_{anterior} \times (1 - rap\_sl)$ . Se o valor assim calculado para  $sl$  for inferior ao máximo do período actual,  $sl$  toma o valor deste acrescido de um  $pip$ .

O GA determinará, para cada indivíduo, a aplicação ou não desta regra com a definição de “Verdadeiro” ou “Falso” para o parâmetro  $rap\_bool$ , sendo todos os parâmetros referidos definidos por aquele.

Esta regra poderá ser aplicada uma única vez por cada posição aberta. O quadro apresenta, em resumo, os parâmetros usados por esta regra e as fórmulas de cálculo nos novos valores previstos para a obtenção de um ganho ou assunção de uma perda.

Posição	Condição	Tipo condição	Parâmetro	Domínio	Tipo
longa	$tp = tp_{anterior} \times (1 + rap\_tp)$	facultativa	$rap\_bool$	Verdadeiro;Falso	lógicos
longa	$sl = sl_{anterior} \times (1 + rap\_sl)$	facultativa	$rap\_periodos$	1-40	inteiros
curta	$tp = tp_{anterior} \times (1 - rap\_tp)$	facultativa	$rap\_percent$	1-100	inteiros
curta	$sl = sl_{anterior} \times (1 - rap\_sl)$	facultativa	$rap\_tp$	1-100	inteiros
			$rap\_sl$	1-100	inteiros

Quadro 4.7: Representação da regra 7

### Regra 8 - Evolução Lenta

A lógica subjacente a esta regra é a de se forçar a saída de uma posição que parece ter dificuldades em evoluir positivamente, com o estreitamento do intervalo entre o valor das ordens de *take profit* e *stop loss*. O indivíduo só poderá usar esta regra se o parâmetro *len\_bool* tiver o valor lógico de “Verdadeiro”. Se o fizer, ela apenas pode ser utilizada uma vez em cada posição tomada.

É medido o crescimento no valor das cotações, entre a média do máximo e mínimo do período actual, e a média de há *len\_periodos* períodos atrás. O valor calculado é comparado com o do parâmetro *len\_percent*. Se, em termos absolutos, o valor do crescimento de preços é menor do que *len\_percent*, os valores de *tp* e *sl* são actualizados.

No caso de uma PL, o valor de *tp* é alterado para  $tp = tp_{anterior} \times (1 - len\_tp)$  e *sl* toma o valor  $sl = sl_{anterior} \times (1 + len\_sl)$ . No caso do valor de *tp*, assim calculado, ser inferior ao máximo do período actual, o valor da ordem de *tp* toma o valor deste acrescido de um *pip*. Se o mínimo do período actual for inferior ao cálculo efectuado para *sl*, *sl* toma o valor daquele subtraído de um *pip*.

No caso de uma PC, *tp* é alterado para  $tp = tp_{anterior} \times (1 + len\_tp)$  ou para o mínimo do período actual subtraído de um *pip*, se este valor for inferior àquele. *sl* tomará o valor de  $sl = sl_{anterior} \times (1 - len\_sl)$  ou o máximo do período actual acrescido de um *pip*, se este for superior àquele. O quadro apresenta, em resumo, os parâmetros usados por esta regra e as fórmulas de cálculo nos novos valores previstos para as ordens *take profit* e *stop loss*.

Posição	Condição	Tipo condição	Parâmetro	Domínio	Tipo
longa	$tp = tp_{anterior} \times (1 - len\_tp)$	facultativa	$len\_bool$	Verdadeiro;Falso	lógicos
longa	$sl = sl_{anterior} \times (1 + len\_sl)$	facultativa	$len\_periodos$	1-40	inteiros
curta	$tp = tp_{anterior} \times (1 + len\_tp)$	facultativa	$len\_percent$	1-100	inteiros
curta	$sl = sl_{anterior} \times (1 - len\_sl)$	facultativa	$len\_tp$	1-100	inteiros
			$len\_sl$	1-100	inteiros

Quadro 4.8: Representação da regra 8

### Regra 9 - Melhoria de cenário

O uso, pelo indivíduo, desta regra está dependente da atribuição, pela evolução do GA, do valor “Verdadeiro” ao parâmetro  $mel\_bool$ . Por cada posição tomada, esta regra só pode ser aplicada uma vez. No caso de ter sido iniciada uma PL em que  $P < EMAl$  ou uma PC em que  $P > EMAl$ , na altura da emissão da ordem de entrada, o algoritmo irá verificar, nos períodos seguintes à abertura da mesma, se o preço médio actual das cotações já ultrapassou o valor de  $EMAl$ , isto é, se  $P > EMAl$ , numa PL, ou  $P < EMAl$ , numa PC. Se sim, os valores de  $tp$  e  $sl$  são modificados, de modo a tentar aproveitar uma tendência de evolução dos preços, que possibilite, eventualmente, a obtenção de um lucro maior, ou a redução de perdas numa posição.

Numa PL, o novo valor de  $tp$  é  $tp = tp_{anterior} \times (1 + mel\_tp)$  e  $sl$  é  $sl = sl_{anterior} \times (1 + mel\_sl)$ . No entanto, se o novo valor de  $sl$  for superior ao valor mínimo do período actual,  $sl$  toma aquele valor diminuído de um  $pip$ . Numa PC, o novo valor de  $tp$  é  $tp = tp_{anterior} \times (1 - mel\_tp)$ .  $sl$  toma um valor  $sl = sl_{anterior} \times (1 - mel\_sl)$  ou, se este valor for inferior ao valor máximo do período actual,  $sl$  toma este valor acrescido de um  $pip$ .

O quadro apresenta, em resumo, os parâmetros usados por esta regra e as fórmulas de cálculo nos novos valores previstos para a obtenção de um ganho ou assunção de uma perda.

Posição	Condição	Tipo condição	Parâmetro	Domínio	Tipo
longa	$tp = tp_{anterior} \times (1 + mel\_tp)$	facultativa	$mel\_bool$	Verdadeiro;Falso	lógicos
longa	$sl = sl_{anterior} \times (1 + mel\_sl)$	facultativa	$mel\_tp$	1-100	inteiros
curta	$tp = tp_{anterior} \times (1 - mel\_tp)$	facultativa	$mel\_sl$	1-100	inteiros
curta	$sl = sl_{anterior} \times (1 - mel\_sl)$	facultativa			

Quadro 4.9: Representação da regra 9

### Regra 10 - Pioria de cenário

A última regra a ser incorporada na estrutura do indivíduo diz respeito à face contrária da regra anterior. Pode ser activada, no máximo de uma vez por cada posição, se, com o decorrer dos períodos, o preço médio actual da cotação de uma posição aberta se deteriorar face ao valor actual de  $EMAl$ . A sua utilização só é, contudo, possível se o parâmetro  $pior\_bool$  assumir o valor de “Verdadeiro”.

Começa-se por guardar a situação do preço médio de entrada em relação ao valor de  $EMAl$  do período em que foi emitida a ordem de entrada que deu origem à posição aberta. No caso de uma PL, se a situação inicial era de  $P > EMAl$ , o indivíduo verifica, para cada um dos períodos subsequentes, enquanto a posição se mantiver aberta, se o preço médio actual de cada período continua, ou não, a ser superior a  $EMAl$  actual. No primeiro período em que o preço médio seja inferior ao valor da  $EMAl$ , as ordens de *take profit* e *stop loss* anteriormente inseridas no sistema são alteradas segundo uma lógica de forçar a saída da posição. Os novos valores de *take profit* e *stop loss* são:  $tp = tp_{anterior} \times (1 - pior\_tp)$  e  $sl = sl_{anterior} \times (1 + pior\_sl)$ . No entanto, caso o valor de  $tp$  assim determinado seja inferior ao máximo do período actual,  $tp$  assume o valor desse máximo acrescido de um  $pip$ . Da mesma forma, se  $sl$  for superior ao mínimo do período actual, o novo valor de  $sl$  é igual a esse mínimo diminuído de um  $pip$ . Desta forma garante-se a existência de valores possíveis para  $tp$  e  $sl$ . No caso de PC, o raciocínio é idêntico. Se a re-

gra estiver activa e  $P < EMAl$ , no momento da emissão da ordem de entrada, o sistema verifica se, em algum período,  $P > EMAl$ . Se esta condição for verdadeira, os novos valores de  $tp$  e  $sl$  são:  $tp = tp_{anterior} \times (1 + pio\_tp)$  e  $sl = sl_{anterior} \times (1 - pio\_sl)$ . Estes valores são alterados para o mínimo do período diminuído de um  $pip$  e máximo acrescido de um  $pip$ , caso sejam superiores ao mínimo ou inferiores ao máximo do período actual, respectivamente.

O quadro apresenta, em resumo, os parâmetros usados por esta regra e as fórmulas de cálculo nos novos valores previstos para as ordens *take profit* e *stop loss*.

Posição	Condição	Tipo condição	Parâmetro	Domínio	Tipo
longa	$tp = tp_{anterior} \times (1 - pio\_tp)$	facultativa	$pio\_bool$	Verdadeiro;Falso	lógicos
longa	$sl = sl_{anterior} \times (1 + pio\_sl)$	facultativa	$pio\_tp$	1-100	inteiros
curta	$tp = tp_{anterior} \times (1 + pio\_tp)$	facultativa	$pio\_sl$	1-100	inteiros
curta	$sl = sl_{anterior} \times (1 - pio\_sl)$	facultativa			

Quadro 4.10: Representação da regra 10

#### 4.4 Caracterização do algoritmo genético

O algoritmo começa por carregar para memória as configurações de que necessita e que são relativas a universos dos valores possíveis para cada parâmetro identificado na secção 4.3 e outros relativos à sua implementação computacional. Depois de adquirir a informação necessária à sua execução, o GA inicia os seus procedimentos com a geração de dez indivíduos de forma aleatória, o que constitui a população inicial. No entanto, dado que existem algumas restrições a respeitar, a geração de valores aleatórios para os parâmetros em causa é repetida até que haja compatibilidade entre os valores atribuídos àqueles: A atribuição de valores a  $emal\_periodos$  será repetida, se necessário, até que  $emal\_periodos > emac\_periodos$ ; E  $tr2\_vezes$  só será aceite quando  $tr2\_vezes \leq tr2\_periodos$ .

Em seguida, o GA obtém os dados relativos à série temporal que servirá de base à evolução da população. Esta informação compreende os seguintes campos: índice da base de dados onde a série é guardada, data, hora, máximo e mínimo de cada período, bem como os relativos aos valores de todas as médias móveis exponenciais de periodicidade entre 3 e 200.

#### 4.4.1 Comportamento do indivíduo

Depois de finalizados os procedimentos de inicialização do GA, é principiado um ciclo que avalia o desempenho de cada um dos indivíduos constantes da população inicial. Com o primeiro indivíduo, é iniciado um outro ciclo interior que tem como objectivo o cálculo do seu valor de desempenho e que percorre os períodos da série temporal considerada, do mais antigo ao mais recente. A série de operações que o ciclo interior comporta é a seguinte: a) evidenciar os dados relativos à cotação do período presente; b) analisar se há evolução rápida das cotações; c) analisar se há evolução lenta das cotações; d) verificar se houve uma melhoria no cenário; e) verificar se houve pioria no cenário; f) verificar se houve a abertura de uma posição; g) verificar a emissão de uma ordem de entrada de posição; h) verificar se houve o fecho de uma posição; i) no final deste ciclo é possível calcular a estatística do indivíduo.

Em a), é inicializada a instância *actual* do objecto *Cotacao*, com os atributos *indice*, *data*, *hora*, *max*, *min*, *ema\_c* e *ema\_l*, sendo que *ema\_c* e *ema\_l* correspondem aos parâmetros *emac\_periodos* e *emal\_periodos* do indivíduo escolhido. Em a), são considerados os dados relativos às cotações do período em causa, ou seja, o índice, data, hora, máximo, mínimo, valor da EMA de curto prazo respectiva

e valor da EMA de longo prazo respectiva.

A operação referida em b) só é executada nas seguintes condições: 1) se há um negócio aberto; 2) se *rap\_bool* tem valor lógico “Verdadeiro”; 3) Se já decorreram, pelo menos, *rap\_periodos* desde o momento de abertura da posição. No caso de estas condições serem satisfeitas o GA executa a regra 7 e avalia a possibilidade de redefinição do preços estabelecidos para saída, seja com a realização de um ganho ou de uma perda.

Para a execução da operação referida em c) é necessário que: 1) haja um negócio aberto; 2) *len\_bool* tenha o valor lógico “Verdadeiro”; 3) a regra 7 não tenha alterado *tp* e *sl*, neste período; 4) já tiverem decorrido pelo menos *len\_periodos* períodos desde a tomada de posição. Se todas estas condições retornarem o valor lógico “Verdadeiro”, a regra 8 é executada. Da avaliação efectuada pela regra, é decidida a alteração, ou não, dos valores definidos para *tp* e *sl*.

A operação a que d) se refere, que corresponde à regra 9, é apenas executada se as seguintes condições prévias forem verdadeiras: 1) existir um negócio aberto; 2) *mel\_bool* possuir valor lógico “Verdadeiro”; 3) as regras 7 e 8 não terem alterado os valores fixados para saída de posição, neste período; 4) *emal\_bool* do período de emissão da ordem de entrada ter o valor lógico “Falso”; Se tiver ocorrido uma “Melhoria de cenário” tal como prevista pela regra, os valores anteriormente considerados para *take profit* e *stop loss* são alterados.

Para que a operação referida em e), concretizada na regra 10, seja executada, são verificadas as seguintes condições: 1) haver um negócio aberto; 2) *pio\_bool* possuir valor lógico “Verdadeiro”; 3) os métodos referidos em b) e c) não terem alterado os valores fixados para saída de

posição: *tp* e *sl* neste período; Se a regra concluir por uma “Pioria de cenário”, há lugar à modificação dos valores de *take profit* e *stop loss*.

O passo seguinte, referido na alínea f), consiste em verificar se existem, para o período, ordens de abertura de posição, inseridas no sistema no período anterior, de acordo com o definido pela regra 5. Se se verificar a existência de uma ordem de entrada, há lugar à abertura de uma posição se o preço máximo do período for superior ou igual ao preço fixado na ordem acrescido do valor do *spread* e se o preço mínimo for inferior ou igual ao estabelecido diminuído do valor do *spread*. Esta operação só é, no entanto, executada se não houver qualquer posição aberta, no período.

A operação seguinte, referida na alínea g), diz respeito à possibilidade de emissão de uma nova ordem de entrada numa posição. Se houver algum negócio aberto no presente período, a execução da operação é abandonada. Se não for, são verificadas as diversas condições descritas na secção 4.3.1. Para cada regra mencionada, são definidos os valores lógicos de dois atributos: *valida\_longas* e *valida\_curtas*, ou seja há sempre a validação da emissão de uma ordem, quer para uma PL, quer para uma PC. A validação da emissão de uma ordem de compra resulta da conjunção dos vários atributos *valida\_longas* relativos a cada uma das quatro regras: 1, 2, 3 e 4, mencionadas na secção 4.3.1. Se as quatro regras resultarem todas no valor lógico “Verdadeiro”, será emitida uma ordem de compra. Por seu turno, haverá a emissão de uma ordem de venda se os atributos *valida\_curtas* relativos às quatro regras assumirem o valor “Verdadeiro”. Dado que a regra 1 é mutuamente exclusiva permitindo a tomada, ou de PL, ou de PC, não haverá a possibilidade de serem introduzidas duas



ordens, uma longa e uma curta, simultâneamente.

Como foi descrito na secção 4.3.1, a utilização das regras 2, 3 e 4, depende da estrutura de cada indivíduo, nomeadamente pela atribuição do valor lógico “Verdadeiro” ou “Falso” aos parâmetros *emal\_bool*, *tr1\_bool* e *tr2\_bool*. Quando o indivíduo não possui algum destes genes activos, os atributos *valida\_longas* e *valida\_curtas* das regras correspondentes são considerados como tendo o valor “Verdadeiro” por forma a neutralizar a sua acção na conjunção referida no parágrafo anterior. Assim, na hipótese de *emal\_bool*, *tr1\_bool* e *tr2\_bool* terem todos o valor lógico “Falso”, será emitida uma ordem de compra se  $P > EMAC$  ou um ordem de venda se  $P < EMAC$ .

Nesta simulação, houve sempre a preocupação de ajustar todos os procedimentos às condições do mercado real, de forma a conseguir resultados fidedignos. Se houver emissão de uma ordem de entrada, é definido o preço de entrada, bem como os de saída com lucro: *take profit*, e de saída com prejuízo: *stop loss*. A ordem de entrada emitida é do tipo “*stop limit*”. Uma ordem do tipo “*stop limit*” é accionada quando a cotação do activo subjacente atinge o preço escolhido para a ordem. Quando ocorre tal momento, a ordem transforma-se numa ordem “ao limite”. Por uma ordem “ao limite” entende-se uma ordem para comprar um activo a um preço não superior a (ou vender a um preço não inferior a) um valor específico. O tipo de ordem “*stop limit*” confere, pois, um grande grau de segurança sobre o preço a que a ordem irá ser executada, em situações de negociação simulada e real. Se a ordem de entrada emitida fosse do tipo “ao limite” apenas, perder-se-ia precisão quanto ao valor de entrada, em qualquer das duas situações. Na hipótese de ser utilizada uma ordem “ao limite”, a falta de precisão em situação de simulação

seria originada por o sistema apenas ter acesso ao resumo das cotações de cada período, isto é, ao máximo e mínimo. A informação relativa à ordem em que as cotações são obtidas, em cada período, é ignorada pelo simulador. Considere-se o exemplo de ter sido emitida uma ordem de compra “ao limite” de 1,4000 e de o período seguinte ter tido um máximo de 1,4100 e um mínimo de 1,3900. Em situação real, se os preços, no período seguinte começaram em 1,3950, desceram até 1,3900 e subiram a 1,4100, teria sido iniciada uma posição ao preço de 1,3950. Na simulação, seria considerada uma entrada (compra) a 1,4000, menos vantajosa do que a real. Por outro lado, se as cotações começaram por atingir os 1,4100 e desceram, depois, até 1,3900, tanto no caso real, como na simulação, o preço de entrada seria 1,4000. Situação semelhante ocorre na tomada de posições curtas, sempre que o máximo do período seguinte for superior ao preço de entrada fixado.

A última operação do ciclo que percorre todas as cotações, referida na alínea h), trata do encerramento de posições eventualmente abertas. É designada por “Fechar negócio”. A sua inclusão depois do método de abertura de um novo negócio dá-lhe a possibilidade de fechar imediatamente uma posição aberta no próprio período, se a ordem de *stop loss* for activada. Possibilidade semelhante não é permitida a ordens *take profit* resultantes de posições abertas no período actual. Dado que não se estuda o comportamento do sistema *pip a pip*, mas apenas de uma forma resumida, aplicar-se-á o princípio da prudência. No caso de uma PL, por exemplo, em que o mínimo do período seja inferior ao valor definido para *stop loss*, ao não se saber se, na realidade, esse mínimo teria sido atingido antes ou depois da entrada na posição, opta-se por fechar a posição com perda, imediatamente. Por ou-

tro lado, se uma acção idêntica fosse tomada em relação a ordens *take profit*, poder-se-iam fechar negócios com lucro que, se a transacção fosse real, nunca aconteceriam. As ordens de *take profit* só têm efeito quando tenha decorrido, no mínimo, um período desde o momento em que houve abertura de posição. Deste modo, quando o preço de *take profit* é atingido, é “seguro” que o sistema geraria um lucro em condições reais de negociação.

Este método começa por testar se a ordem *stop loss* foi activada e, só depois, faz verificação idêntica para a ordem *take profit*, mais uma vez por prudência. Assim, assumindo o exemplo de uma PL, se, por hipótese, num período, o máximo ultrapassou o preço de *take profit* estabelecido e o mínimo foi menor do que o preço predefinido para a activação da ordem *stop loss*, o sistema irá considerar esta última e desprezar a primeira.

No fecho de posições, foi ainda adoptado outro mecanismo que visa salvaguardar os resultados obtidos, mais uma vez, tendo presente o objectivo de realizar a simulação o mais próxima possível das condições reais de negociação. Para uma ordem *stop loss* ser activada, numa PL, é necessário que o mínimo do período seja inferior ou igual ao valor definido para *sl*. Numa PC, basta que o máximo seja superior ou igual a *sl*. No que às ordens *take profit* diz respeito, ela só será activada se o extremo da cotação exceder o valor predefinido para aquela em, pelo menos, o valor do *spread*. Assim, se for uma PL, haverá lugar ao *take profit* se o máximo do período for igual ou superior ao valor de *tp* adicionado do valor do *spread*. No caso de uma PC, haverá recolha do lucro se o mínimo do período é igual ou inferior ao valor de *tp* diminuído do valor do *spread*. Esta diferenciação tem por base a seguinte justificação. Para haver a execução de uma ordem ao limite, do tipo *take*

*profit*, é necessário que o preço ultrapasse, de facto, o limite estabelecido em, pelo menos o valor do *spread*. Em condições reais, mesmo que a cotação seja igual ao valor do *tp* adicionado do valor do *spread*, a execução da ordem não é garantida, uma vez que a ordem antes introduzida estará sujeita a um mecanismo de fila de espera, sendo que as ordens que chegam primeiro à corretora serão servidas à frente. Para além disso, as ordens dadas “ao melhor” têm prioridade sobre as restantes. Ainda assim, dados estes constrangimentos, decidiu-se sobre a execução das ordens de *take profit* da forma acima descrita. Quanto às ordens de *stop loss*, são ordens que se transformam em ordens “ao melhor”, logo que o preço definido seja atingido. Coloca-se aqui também uma dúvida em termos de negociação real: Qual é o preço que foi atingido? O preço do último negócio efectuado na corretora? O preço da última oferta? O preço da última procura? O preço médio entre a melhor oferta e a melhor procura? Dadas estas várias possibilidades e dado também que a resposta certa dependerá da corretora, optou-se por fazer accionar a ordem de *stop loss* logo que haja uma cotação que iguale o preço preestabelecido.

Depois de percorridos todos os períodos da série temporal com a efectuação de negócios, é possível saber o resultado do desempenho do indivíduo. A operação referida na alínea i), começa por verificar se o último negócio efectuado ficou por fechar por a série temporal ter esgotado os períodos disponíveis sem que, quer a ordem de *stop loss*, quer a ordem de *take profit*, tivessem sido executadas.. Se tal ocorreu, o preço de fecho considera-se igual ao mínimo transaccionado no último período, se for uma PL, ou igual ao máximo, no caso de uma PC. Esta opção pauta-se igualmente por um critério prudente. É usada a cotação que mais penaliza o desempenho do indivíduo.

Todos os negócios efectuados pelo indivíduo são escrutinados e são calculados: i) número de negócios longos ganhos; ii) de longos perdidos; iii) de curtos ganhos; iv) de curtos perdidos; v) *pips* ganhos em negócios longos; vi) *pips* perdidos em longos; vii) *pips* ganhos em curtos; viii) *pips* perdidos em curtos; ix) maior duração de um negócio longo ganho; x) menor duração de um longo perdido; xi) maior duração de um curto ganho; xii) menor duração de um curto perdido; xiii) resultado final; xiv) pico absoluto de resultado; xv) pico relativo; xvi) fundo absoluto; xvii) quebra máxima.

De i) a iv), é feita a contabilidade dos negócios ganhadores e perdedores. Um negócio é ganho se o preço de fecho é superior ao preço de entrada, se for longo, ou se o preço de entrada é inferior ao preço de fecho, se for curto; nos casos contrários, o negócio é considerado como perdido. De v) a viii) contabiliza-se o número de *pips* ganhos e perdidos em negócios longos e curtos. Esta informação torna-se pertinente devido, sobretudo, à mobilidade dos preços estabelecidos como *take profit* e *stop loss*. Um negócio longo com preço de entrada, por exemplo, a 1,2005 e fecho a 1,2049, acumulará na variável definida em v) 44 *pips* (1,2049-1,2005).

Com as variáveis que medem a duração dos negócios, pretende-se que seja possível um estudo mais pormenorizado do perfil de negociação de cada indivíduo. Os negócios arrastam-se por muito tempo e são em baixo número por causa disso? Ou são pouco numerosos apenas por as condições de entrada de posição serem demasiado selectivas? Evoluem de forma rápida e tiram partido da regra 4.3.2? São algumas interrogações que esta informação poderá ajudar a revelar, numa forma sumariada.

O resultado final é igual à soma de todos os *pips* ganhos

e perdidos, nos negócios efectuados por um indivíduo.

Em xiv), é guardado o pico absoluto do resultado, ou seja, o valor máximo alcançado pelo resultado. De forma similar, mas em termos negativos, é calculado o fundo absoluto, que representa o mínimo, em termos de *pips*, que resultou da negociação efectuada pelo indivíduo. Estes dois valores servem apenas como uma “fraca” indicação do desempenho do algoritmo.

De muito mais utilidade revestem-se os cálculos do pico relativo e da quebra máxima. Em xv), é calculado o pico relativo que corresponde ao valor máximo de incremento de resultado desde o mais recente fundo absoluto. Este valor representa a melhor evolução que se pode esperar do algoritmo. Mais importante, em termos deste algoritmo, é o cálculo da quebra máxima, indicado em xvi), tratado na subsecção seguinte.

#### 4.4.2 Avaliação do desempenho

Na secção 2.4, foram descritos alguns obstáculos à utilização do MDD como forma de medir o desempenho de investimentos. As críticas efectuadas não se mostram, no entanto, adequadas ao presente trabalho. Quanto aos fundos de investimentos, poderá fazer sentido, como afirma Harding et al. [11], não penalizar aqueles que apresentam séries de rendibilidades mais longas, com a utilização do MDD como critério de avaliação do desempenho. Esta noção será tão mais correcta quanto menor for a alavancagem definida por quem os gere. À negociação nos mercados financeiros derivados e seus similares, como o Forex, são proporcionados, no entanto, níveis altos de alavancagem. Como foi afirmado no início do presente capítulo, neste trabalho, o tema do *position sizing* não é discutido. Contudo, e tal como considerado por diversos autores, a que se fez

referência na secção 2.4, julga-se por apropriado medir o grau de risco de investimentos em instrumentos financeiros derivados e similares, pelo MDD, por estes incorporarem um risco elevado de perda total do investimento inicial.

Harding et al. [11] expõem uma segunda desvantagem, a de que os MDD serão tanto maiores quanto maior for a frequência de medição dos mesmos. Essa conclusão é produto da comparação efectuada do grau de risco reportado por investimentos em várias escalas temporais: diária, semanal, mensal, trimestral e anual. A forma de medição do risco utilizada neste trabalho anula este tipo de observação, pois a medição é efectuada período a período quando existem posições abertas e não, apenas, no momento em que as posições são encerradas. A quebra máxima é sempre exposta, na metodologia escolhida, não ficando encoberta por motivo de medições interpoladas no tempo. Por exemplo, um negócio que tenha definido um valor de *take profit* a 50 *pips* do ponto de entrada e um *stop loss* a 100, poderá, a certa altura, ter uma perda potencial de 90 *pips* e ainda assim fechar com lucro, nunca activando a ordem de *stop loss*. Estes 90 *pips* serão tomados em consideração para o cálculo da quebra máxima. Esta opção justifica-se, também, pela tentativa de aproximação das condições de simulação à realidade. Em situação real, é efectuada a verificação da manutenção de uma margem de segurança, em termos monetários, que possibilite a continuação da negociação, a cada momento, para uma dada estratégia.

Quanto à outra desvantagem referida na secção 2.4, por Harding et al. [11], de uma possível distorção causada pela extrapolação de rendibilidades futuras baseadas no cálculo do MDD, esta não parece também poder aplicar-se ao presente trabalho, dado que as realidades a comparar não são díspares entre si, pois dizem todas respeito a solu-

ções do algoritmo desenvolvido. Para além deste facto, a negociação, em instrumentos financeiros que possibilitam mecanismos de alavancagem, carece primordialmente, que seja garantida a não perda total do investimento inicial. A consideração da quebra máxima incorrida por cada indivíduo tenta satisfazer aquela necessidade.

Dado que as desvantagens enunciadas por Harding et al. [11] quanto à utilização do MDD não se mostram adequadas em relação à realidade do presente trabalho, será usado o rácio entre o resultado final e a quebra máxima (MDD) como medida de avaliação do desempenho dos indivíduos. Para cada indivíduo, o MDD indica, como foi dito, a quebra máxima acumulada de *pips*, desde o último pico absoluto, ou seja, representa a perda máxima que um sistema de negociação teve, em qualquer momento, em termos do seu resultado. Este tipo de avaliação de desempenho dá, portanto, relevo a uma medida relativa. O numerador, constituído pelo valor do resultado obtido, corresponde ao número de *pips* ganhos ou perdidos. Quanto maior for o seu valor, maior será o desempenho do indivíduo. No entanto, o valor do resultado, por si só, revela-se insuficiente como base para a selecção de uma estratégia de transacção, pois o risco de diferentes soluções de negociação pode ser muito diverso. A quebra máxima, usado no denominador do rácio, é a variável que indica o risco associado ao indivíduo, permitindo assim seleccionar estratégias de transacção tendo em conta rentabilidade e risco.

#### 4.4.3 Geração seguinte

Na secção anterior, explicitou-se de que forma se avalia o desempenho de cada indivíduo em relação a uma dada série temporal. Depois de avaliado o desempenho de todos os indivíduos da geração actual, é necessário proceder à



evolução da população. Nesta secção, são expostas as operações que farão evoluir a população à geração seguinte.

Em primeiro lugar, os indivíduos que tiverem realizado um número insuficiente de negócios são excluídos das operações de reprodução. Por número insuficiente entende-se um número inferior a dois negócios por cada mil períodos, em média. Este limite foi fixado à priori. Deste modo, pretende garantir-se que não se verificam desempenhos excepcionais de indivíduos devido à ocorrência de situações excepcionais, na série temporal utilizada. Por absurdo, poder-se-ia considerar um exemplo, dum série temporal tendencialmente crescente, em que um indivíduo poderia abrir uma PL num período inicial e fechá-la apenas num período próximo do final. Esta única posição poderia obter um ganho muito significativo com um risco associado baixo. Numa situação *out-of-sample*, esta estratégia de negociação teria uma forte probabilidade de revelar-se impraticável ou de gerar perdas.

Se não tiver havido qualquer exclusão de um indivíduo por número insuficiente de negócios, é verificada a convergência de desempenho dos indivíduos. O desempenho de cada indivíduo é medido, como antes enunciado, pelo rácio  $\frac{\text{resultado}}{\text{quebramaxima}}$ . Para o conjunto dos dez indivíduos que compõem a população actual, são calculados os valores máximo e mínimo de desempenho. Em seguida, é calculada a convergência dos valores de desempenho da seguinte forma:  $conv = \frac{\text{maximo} - \text{minimo}}{\text{minimo}}$ . Se o valor de convergência for inferior ou igual ao predefinido, 8%, isto é,  $|conv| \leq 8\%$ , são aproveitados os 50% de indivíduos com melhor desempenho e descartados os restantes. A percentagem de 50% foi definida à priori.

Apesar desta triagem, os melhores indivíduos mantêm-se na população e estão aptos a ser submetidos às opera-

ções de reprodução e mutação. Permite-se, assim, que a população mantenha os indivíduos com os melhores genes, salvaguardando o máximo local obtido. Os piores serão abandonados e entrarão, para o seu lugar, imigrantes, que correspondem a novos indivíduos gerados aleatoriamente pelo sistema, enriquecendo a diversidade genética da nova população, o que permitirá, ao algoritmo, evitar situações de convergência prematura e ir em busca de novos e melhores máximos locais. A opção pela chamada de imigrantes justifica-se, também, pela pequena dimensão da população, pois as possibilidades de cruzamento, entre diferentes indivíduos, são menores do que seriam numa população mais alargada. Esta solução favorece, igualmente, o tempo computacional requerido para fazer evoluir a população.

No passo seguinte, o algoritmo irá escolher os parceiros para a reprodução, utilizando o método da roleta aplicado ao desempenho de cada indivíduo. De forma a criarem-se intervalos de probabilidade para o uso daquele método, optou-se por converter os valores de desempenho dos indivíduos que ainda fazem parte da população (não incluindo, nem os que foram afastados devido à verificação de convergência, nem os imigrantes que serão chamados), no caso em que haja pelo menos um indivíduo com um rácio de desempenho negativo. Se tal acontecer, será adicionado ao desempenho de todos os indivíduos o dobro do valor negativo, em termos absolutos, do desempenho do pior indivíduo. Deste modo é mantida a hierarquia entre os indivíduos, embora seja alterada a proporcionalidade entre os seus valores de desempenho. Consegue-se assim manter o conceito de relativização entre resultado e risco, ao contrário de em [7], pois mesmo entre indivíduos que não consigam obter lucros, o risco que lhes está associado poderá fazer diferença entre os seus perfis de resultados. Esta

diferença poderá ter consequências ao nível da evolução genética da população, pelo que foi tida em conta neste trabalho.

Os indivíduos reprodutores são escolhidos, segundo o método da roleta, da seguinte forma: a) É criada a variável “desempenho” na qual são adicionados os valores individuais de desempenho, tenham sido eles modificados conforme descrito no parágrafo anterior ou não; b) É calculada a proporção do desempenho individual em relação ao total da população, sendo cada valor multiplicado por 1000, por questões de legibilidade; c) com base nos indivíduos ordenados por ordem crescente de desempenho, são criados intervalos tendo por limite inferior o valor acumulado de desempenho (inicia em 0), provindo dos indivíduos anteriores, e por limite superior o resultado da adição do valor acumulado anterior à proporção de desempenho individual (multiplicada de 1000); d) durante um número de vezes igual ao número de indivíduos aptos para reprodução, são obtidos números aleatórios no intervalo entre 0 e 1000 e será escolhido para reprodutor o indivíduo cujo intervalo de desempenho compreenda o valor aleatório produzido. Deste processo, resulta um número de reprodutores igual ao número de indivíduos ainda disponíveis na população, podendo haver a repetição de indivíduos na lista de reprodutores.

A escolha dos parceiros que são objecto do cruzamento de genes, entre eles, é feita com base no conjunto dos reprodutores, de uma forma aleatória. É possível que um par de progenitores seja constituído pelo mesmo indivíduo ou que pares idênticos de indivíduos seleccionados pela mesma ordem se repitam. O número de pares de parceiros é igual ao número de indivíduos seleccionados como reprodutores.

Dada a opção tomada de excluir metade do número de

indivíduos, no caso de haver convergência dos valores de desempenho, optou-se por garantir que o melhor indivíduo da geração presente fizesse parte da geração seguinte. Escolhidos os pares de parceiros, é verificado se o indivíduo com mais alto desempenho, mais concretamente, se um par constituído pelo indivíduo e por si mesmo, faz parte daqueles. Se isso acontecer, a presença do indivíduo na geração seguinte está confirmada. Se o melhor indivíduo não estiver presente nos parceiros escolhidos para cruzamento, é iniciada uma variável que albergará os indivíduos a juntar aos descendentes da actual geração, na seguinte. Nesta hipótese, dado que a cada par de parceiros corresponderá um descendente, depois do cruzamento de genes, há um par dos parceiros originais que precisa de ser descartado, para que o número de indivíduos da população se mantenha constante de geração em geração. O algoritmo começa por tentar determinar se um par constituído duplamente pelo indivíduo com o menor desempenho da população existe no conjunto de parceiros escolhidos. Se existir, é este o escolhido a sair do conjunto. Se não existir, averigua se o indivíduo com o desempenho mínimo faz parte de algum dos pares. Se tal acontecer, o primeiro par achado é eliminado do conjunto dos parceiros. Se não, o algoritmo escolhe o indivíduo com o nível de desempenho seguinte e volta a realizar as mesmas operações.

No fim de estabelecer, ao certo, o número de pares de parceiros e de saber se o melhor indivíduo é repescado para a geração seguinte, o algoritmo está em condições de determinar quantos imigrantes precisa para completar a população de dez elementos. Os indivíduos assim criados, se os houver, são adicionados ao conjunto a acrescentar, o qual poderá já conter o melhor indivíduo da geração.

Na operação seguinte, o algoritmo faz o cruzamento dos

parceiros escolhidos anteriormente. Cada indivíduo corresponde a um vector com “n” elementos, sendo que “n” representa o número de parâmetros que o caracteriza. O número de parâmetros utilizados é de 31. Para cada par de indivíduos é calculado aleatoriamente um número que corresponderá a um elemento do vector entre o segundo e o último, inclusive. O indivíduo descendente é formado pelos parâmetros do pai à esquerda dessa posição, no vector correspondente, e pelos parâmetros da mãe, no vector desta, para a direita, a partir, inclusive, do elemento cuja posição foi calculada aleatoriamente. O filho, assim nascido, vai fazer parte da nova geração. O procedimento de gerar dois descendentes por cada par de progenitores, tal como exposto na secção 2.2, não foi seguido. Esta decisão teve em linha de conta o número reduzido de indivíduos que constitui a população em evolução. Optou-se por aproveitar apenas um descendente de um número maior de pares, em detrimento de dois de um número mais reduzido de parceiros, por se presumir que daquele modo se possibilitaria uma maior diversidade genética na geração seguinte, devido a uma possível maior diversidade nos progenitores.

De acordo com o trabalho de Nuñez-Letamendia et al. [20], referido na secção 2.3, neste algoritmo, a probabilidade de cruzamento de indivíduos foi estabelecida num valor alto, no caso, em 100%, ou seja, todos os pares escolhidos serão submetidos a esta operação. As razões para tal decisão são: a) A existência da característica do elitismo, com a presença certa do indivíduo com melhor desempenho na geração seguinte. Isto implica que o máximo local antes encontrado fica salvaguardado. Os outros pares poderão assim ser todos sujeitos à operação de cruzamento de genes de forma a agilizar o processo de evolução. b) O peso computacional requerido para avaliar o desempenho

de cada indivíduo implica que se tente dinamizar o mais possível a evolução da população com a troca de genes.

Depois da operação do cruzamento de genes, os indivíduos da nova geração que resultam do cruzamento dos progenitores são submetidos a uma operação de mutação, um a um. Ainda de acordo com Nuñez-Letamendia et al. [20], foi escolhida uma probabilidade de mutação baixa, neste caso, de 3‰. Num algoritmo em que a população não vá sendo alterada com a inclusão de código genético novo de imigrantes, a mutação é a única forma de o GA introduzir informação genética não representada na população inicial e de fugir a ótimos locais. Numa situação deste tipo, poderia optar-se por atribuir à operação de mutação uma probabilidade bem maior do que aquela que foi atribuída.

De acordo com o explanado anteriormente, neste GA, foi decidido consagrar à figura do imigrante o principal papel de introdução de código genético novo na população. Para além das razões anteriormente apontadas, esta decisão é reforçada pela especificidade da estrutura construída. Uma vez que as características de cada indivíduo vão sendo determinadas pela evolução da população, há a probabilidade de várias das regras estabelecidas estarem inactivas. No caso de uma regra estar inactiva e de haver uma mutação num parâmetro que lhe pertença, tal mutação não terá qualquer efeito no desempenho do indivíduo. Optou-se assim por fixar a probabilidade afectada à operação de mutação num valor considerado baixo, em termos de algoritmos genéticos, tal como concluído por Nuñez-Letamendia et al. [20].

O procedimento de mutação consiste no seguinte. Foram usados valores entre 0 e 1000 para a geração de números aleatórios de forma a utilizarem-se números inteiros.

Para cada um dos indivíduos e para cada um dos seus genes, o sistema acha um valor aleatório entre 0 e 1000. Se aquele for inferior ou igual a 3, o gene em causa é objecto de mutação, isto é, se  $numeroaleatorio \leq 3\%_0 \times 1000$ . O algoritmo assegura que o novo valor para o parâmetro em causa seja diferente do anterior e verifica, no caso dos atributos *ema\_c*, *ema\_l*, *tr2\_vezes* e *tr2\_periodos* se os novos valores são compatíveis com os existentes que lhes estão relacionados.

Depois de a operação de mutação ter terminado, os indivíduos descendentes da geração anterior, o indivíduo que teve na geração anterior melhor desempenho (no caso de não ter pertencido aos pares de parceiros escolhidos para cruzamento) e os novos imigrantes constituirão a nova geração.

A nova geração está apta a que o sistema avalie o desempenho dos seus membros. Desta segunda vez, e a partir daqui, a avaliação de desempenho pode não ser efectuada de forma extensiva, pois o GA verifica se o código genético de cada indivíduo já foi estudado pelo sistema. Se o tiver sido, o indivíduo irá obter a informação relativa ao resultado obtido, à quebra máxima a que esteve sujeito e ao número de negócios, a partir do seu idêntico.

O valor do número de gerações que o algoritmo vai fazer evoluir ficou pré-estabelecido em 50.

## Capítulo 5

# Resultados obtidos

*“No amount of experimentation can ever prove me right; a single experiment can prove me wrong.”*  
(Albert Einstein)

Neste capítulo, serão indicados os resultados das várias execuções do GA e será feita uma análise aos valores obtidos.

Em cada quadro apresentado que se refira a cinco execuções do GA, as quatro primeiras linhas de dados referem-se às execuções do algoritmo tendo por base indivíduos gerados em séries de evolução primária. As quintas linhas de dados referem-se a indivíduos referentes às evoluções secundárias.

### 5.1 Evolução da população

A título de exemplo, é mostrado, na figura 5.1, o comportamento genérico do GA na evolução do desempenho de uma das populações consideradas neste trabalho.

A linha que une quadrados (“maximo”) representa a evolução, ao longo das 50 gerações, do desempenho do melhor indivíduo de cada geração. A linha “minimo” que utiliza círculos para representar valores, corresponde ao pior valor



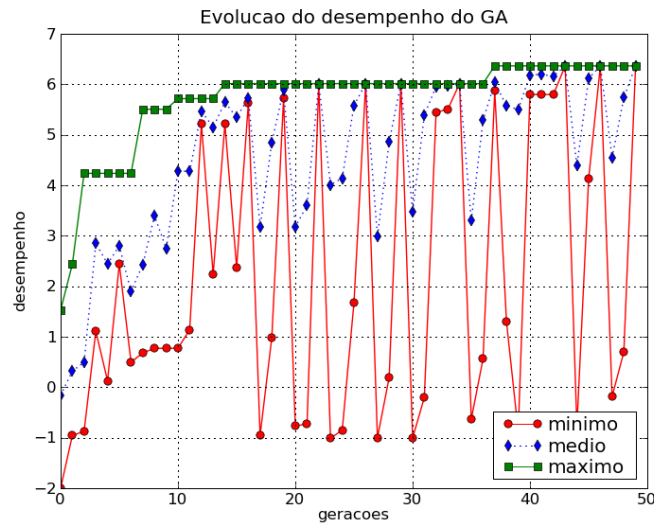


Figura 5.1: Exemplo de evolução da população

de desempenho, em cada geração. A linha “medio” representa a evolução no desempenho médio da população.

O desempenho do melhor indivíduo na geração 0 (a primeira) situa-se num valor pouco acima de 1,5 e sofre subidas acentuadas com a sucessão das primeiras gerações. A subida marginal no desempenho torna-se nula ou muito baixa a partir, de cerca, do final do terço inicial do período de evolução.

Os momentos de convergência podem ser inferidos pela observação da linha “minimo”, quando esta linha toca ou se aproxima bastante da linha de desempenho “maximo”. A entrada de imigrantes origina quebras acentuadas na linha “minimo”.

## 5.2 Desempenho em séries de evolução

De acordo com o mencionado na secção 3.2.1, o algoritmo foi executado quatro vezes para cada uma das séries mencionadas no quadro 3.2 e uma vez para as séries constantes

do quadro 3.3.

O quadro 5.1 apresenta os valores obtidos para as variáveis desempenho, resultado (em *pips*) e número de negócios, para os vinte indivíduos seleccionados, de cada série de evolução, relativamente a cada série temporal base. A coluna referente à “5.<sup>a</sup> evolução” refere-se à execução do GA tendo por base as séries de evolução secundária. As outras quatro colunas com valores referem-se a evoluções primárias.

A figura 5.2 mostra, graficamente, os valores de desempenho obtidos, agrupados em termos de série temporal de evolução.

		1 <sup>a</sup> evolução	2 <sup>a</sup> evolução	3 <sup>a</sup> evolução	4 <sup>a</sup> evolução	5 <sup>a</sup> evolução
1m	desempenho	5,14	4,68	3,34	5,27	1,29
	resultado ( <i>pips</i> )	663	192	464	849	150
	n. <sup>o</sup> negócios	85	63	78	286	421
5m	desempenho	4,34	2,63	5,65	5,75	9,97
	resultado ( <i>pips</i> )	729	583	774	592	598
	n. <sup>o</sup> negócios	80	165	58	177	71
15m	desempenho	2,15	9,87	4,44	4,54	6,36
	resultado ( <i>pips</i> )	571	1559	1828	1616	942
	n. <sup>o</sup> negócios	116	237	67	96	108
60m	desempenho	0,57	0,97	0,56	0,79	5,66
	resultado ( <i>pips</i> )	869	1714	1009	1485	3059
	n. <sup>o</sup> negócios	60	574	166	204	398

Quadro 5.1: Desempenhos, resultados em *pips* e n.<sup>o</sup> de negócios, em séries de treino

Da observação dos valores de desempenho obtidos pelo GA em séries de treino, verifica-se que são todos positivos. Variam entre um mínimo de 0,56 e um máximo de 9,97, sendo a média de 4,2 e a mediana de 4,49.

Em termos de evolução primária, as séries de 1, 5 e 15 minutos tiveram médias de desempenho próximas entre si, de 4,61, 4,59 e 5,25, respectivamente. A série de 60 minutos teve uma média de desempenho de 0,72.

Na secção 3.2 foi dado conta da escolha das séries para efeitos de treino. Afirmou-se que não haveria séries com formas mais “certas” que devessem ser usadas para a evo-

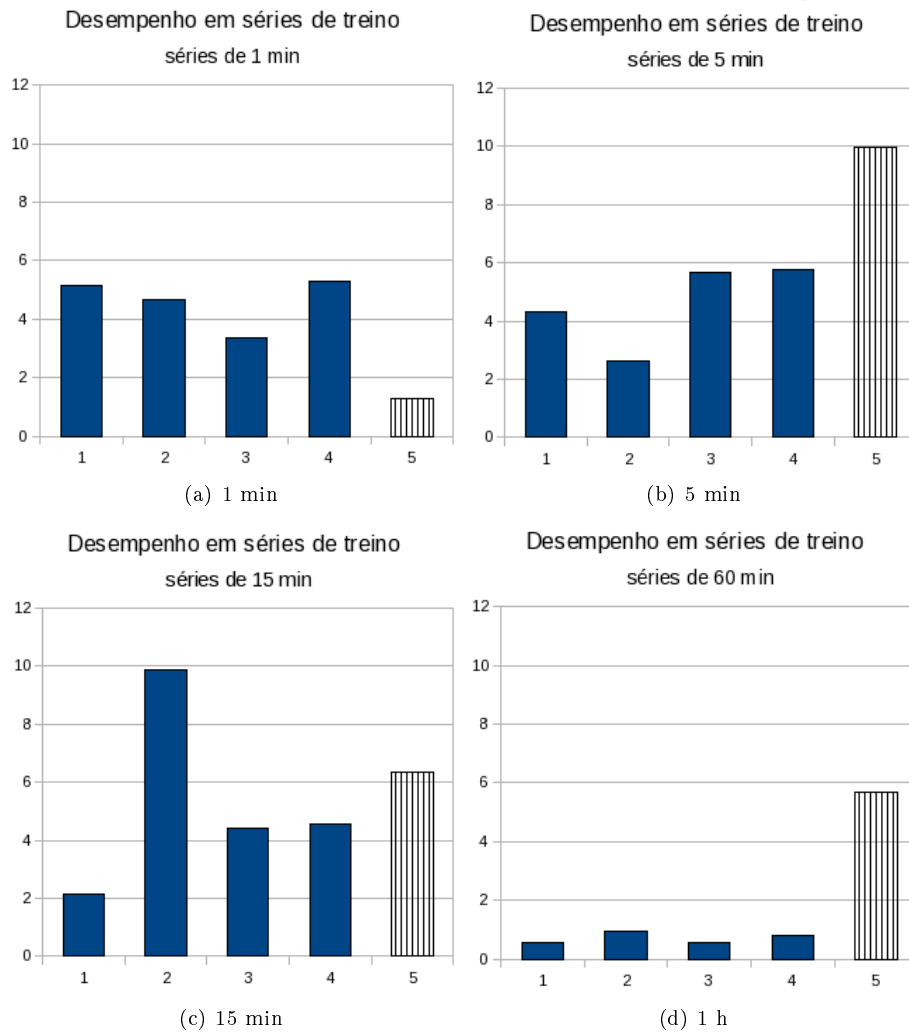


Figura 5.2: Desempenhos obtidos em séries de treino

lução das populações, em detrimento de outras. A introdução do conceito de evolução secundária, apesar de esta só ter sido realizada por uma vez, para cada série temporal base, pretende comparar o resultado obtidos quando são usadas séries diferentes para o treino de populações. Constata-se que a média de desempenho das dezasseis execuções do algoritmo referentes à evolução primária é de 3,79. A média das quatro execuções do GA referentes à evolução secundária é de 5,82. Estes números parecem sugerir que as séries temporais utilizadas para efeitos de evolução primária não tinham características mais adequadas a este processo do que as séries de evolução secundária.

Como observação curiosa, verifica-se que os valores de desempenho referentes à evolução secundária nas séries de 5 e 60 minutos são superiores a qualquer um dos referentes à evolução primária respectiva.

Os resultados do quadro 5.1, que estão expressos em *pips*, indicam que os melhores indivíduos, de cada uma das execuções do GA, obtiveram lucro com o sistema de negociação. Deve ser salientado que a média de resultados das cinco execuções do algoritmo teve o seu valor menor, em termos absolutos, na série temporal de 1 minuto. O resultado médio, por série, cresceu à medida que foram considerados períodos base mais longos de cada série temporal.

No que respeita ao número de negócios efectuados em cada execução do GA, foi construída a estatística constante do quadro 5.2.

No que respeita à evolução primária, verifica-se que o valor do quociente  $\frac{\text{periodos}}{\text{negocio}}$ , com valores arredondados à unidade, diminui à medida que aumenta a duração de cada período da série considerada. Esta tendência mostra que na série de 60m, por exemplo, o GA abre e encerra negócios

série	evolução	mínimo	máximo	média	n. <sup>o</sup> períodos	períodos/negócio	per/neg ponderada
1m	primária	63	286	128	28460	222	152
	secundária			421	27858	66	
5m	primária	58	177	120	25041	209	226
	secundária			71	24242	341	
15m	primária	67	237	129	24664	191	198
	secundária			108	24713	229	
60m	primária	60	574	251	12559	50	45
	secundária			398	12417	31	

Quadro 5.2: Estatística do número de negócios efectuados em séries de evolução

em maior número relativo, por período, do que nas séries de 1, 5 e 15 minutos. Uma possível explicação poderá residir no facto de as séries relativas a períodos de tempo maiores abrangerem horizontes temporais mais alargados, os quais albergam variações maiores, em termos absolutos, dos preços dos activos. Dado que o universo dos valores possíveis a escolher pelo GA para os diversos parâmetros é igual para cada uma das quatro séries temporais base, a assunção antes enunciada poderá significar uma maior adequação dos valores dos parâmetros às séries de período maior.

Os dados relativos à evolução secundária, que resultam apenas de uma execução do algoritmo, não contrariam a tendência descrita no parágrafo anterior, no que respeita à evolução primária, tendo apenas um valor dissonante, o relativo ao número de negócios efectuado na série de um minuto.

A coluna mais à direita, com o título “per/neg ponderada”, resulta da média ponderada, arredondada para as unidades, do número de períodos por negócio nas séries primária e secundária relativas a cada frequência. Para a série de 1 minuto, por exemplo,  $152 = \frac{4 \times 28460 + 27858}{4 \times 128 + 421}$ .

Um maior número de negócios efectuados deverá presupor um algoritmo mais fiável para uma aplicação *out-of-sample*, pois é testado mais amiudamente. O número de possíveis surpresas negativas, em termos de comportamento do sistema, deverá ser assim menor do que naqueles

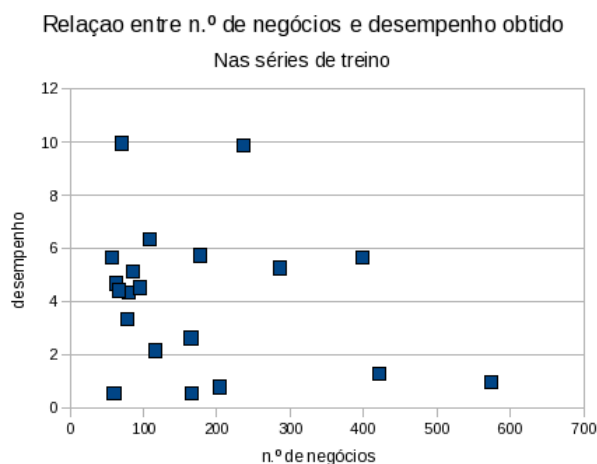


Figura 5.3: Relação entre n.º de negócios e desempenho obtido

sistemas que fazem depender o desempenho final de um número reduzido de negócios, devido à possível existência de condições particulares de negociação, nestes últimos. Para evitar este problema, foi estabelecido, tal como descrito na secção 4.4.3, um mínimo obrigatório de dois negócios por cada mil períodos. Será então de analisar se haverá uma relação directa entre o nível de desempenho obtido e o número de negócios efectuado por cada execução do GA. De forma a analisar a possível relação linear entre os dois factores, foi calculado o valor do coeficiente de correlação, “r”. Para cada uma das variáveis, foi considerada a série de vinte valores obtidos pela execução do GA e constantes do quadro 5.1. Foi obtido um valor de  $r = -0,23$  e  $r^2 = 0,05$ , o que sugere a não existência de uma relação linear estatisticamente significativa entre as duas variáveis, dado o valor de  $r^2$  próximo de zero. A figura 5.3 mostra a distribuição dos pares de valores, em gráfico.

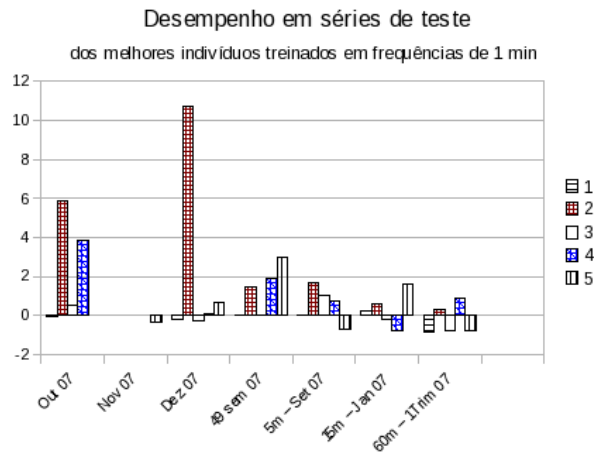


Figura 5.4: Séries 1 a 4 treinadas com base no mês de Novembro de 2007; série 5 treinada com base no mês de Outubro do mesmo ano.

### 5.3 Desempenho em séries de testes

Depois de apresentados os resultados das execuções do GA referentes à evolução das populações, interessa conhecer como se comportaram os melhores indivíduos de cada treino, em séries *out-of-sample*. Os testes efectuados foram os indicados no quadro 3.5.

#### 5.3.1 Indivíduos treinados em séries de 1 minuto

Cada indivíduo treinado em séries temporais com frequência de um minuto foi testado, *out-of-sample*, em seis outras séries, das quais três são séries de um minuto. As outras séries correspondem a frequências diferentes: de cinco, quinze e 60 minutos.

Os valores de desempenho obtidos são os constantes do quadro 5.3. A figura 5.4 mostra graficamente os mesmos valores.

Na série de tipo A, Outubro de 2007, o GA teve um desempenho maioritariamente positivo, com destaque para duas execuções, com desempenhos próximos de quatro e

seis unidades. Numa das execuções o resultado obtido foi ligeiramente negativo.

A quinta execução do GA, de um indivíduo treinado em série diferente dos demais, mostra um resultado negativo para a série de tipo B, Novembro de 2007. Nas outras quatro execuções primárias, relativas à série de Dezembro de 2007 (tipo B), o GA teve um comportamento irregular. Dois dos valores obtidos são negativos, um é ligeiramente positivo e o outro é largamente positivo. O indivíduo referente à quinta execução teve um comportamento positivo.

Quanto ao comportamento dos cinco indivíduos na série da 49.<sup>a</sup> semana de 2007 (tipo C), verifica-se que os resultados são positivos em três execuções, mas há lugar a dois valores tangencialmente negativos.

No que respeita às séries de outras frequências, apenas a segunda execução proporcionou valores positivos de desempenho em todas elas.

série	medida	Out 07	Nov 07	Dez 07	49 sem 07	5m - Set 07	15m - Jan 07	60m - 1Trim 07
1	desemp	-0,0139		-0,2309	-0,0024	0,0435	0,2457	-0,8475
	result	-7		-109	-24	8	43	-339
	n.º neg	64		69	21	35	51	95
2	desemp	5,8462		10,6875	1,5000	1,6585	0,6234	0,2787
	result	152		171	18	68	48	34
	n.º neg	35		38	8	26	18	22
3	desemp	0,5059		-0,2520	-0,0039	1,0581	-0,1893	-0,7905
	result	86		-62	-39	164	-81	-498
	n.º neg	42		54	17	22	25	32
4	desemp	3,8304		0,1233	1,8873	0,7659	-0,7769	0,8947
	result	429		37	134	157	-202	204
	n.º neg	196		219	60	90	48	47
5	desemp		-0,3711	0,6789	3,0172	-0,7143	1,5844	-0,8044
	result		-144	129	175	-185	244	-699
	n.º neg		587	413	106	196	97	99

Quadro 5.3: Desempenho, resultado em *pips* e número de negócios, em séries de teste *out-of-sample* dos melhores indivíduos treinados em séries de 1 minuto

Na quinta execução foram gerados sempre mais negócios do que nas outras quatro.



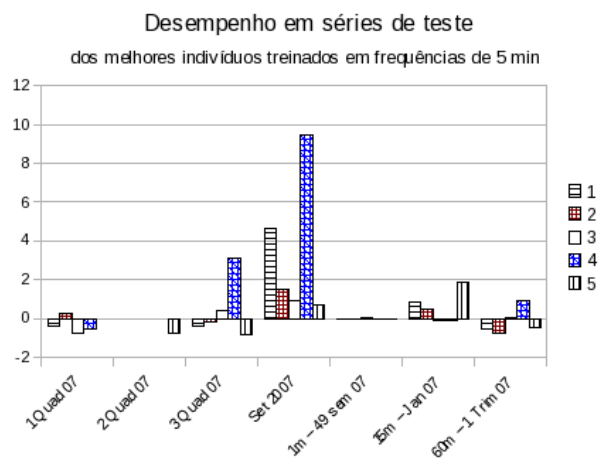


Figura 5.5: Séries 1 a 4 treinadas com base no 2.<sup>o</sup> quadrimestre de 2007; série 5 treinada com base no 1.<sup>o</sup> quadrimestre do mesmo ano.

### 5.3.2 Indivíduos treinados em séries de 5 minutos

Cada indivíduo treinado em séries temporais com frequência de cinco minutos foi testado, *out-of-sample*, em seis outras séries, das quais três são séries de cinco minutos. As outras séries dizem respeito a frequências diferentes: de um, quinze e 60 minutos.

Os valores de desempenho obtidos são os constantes do quadro 5.4. A figura 5.5 mostra graficamente os mesmos valores.

série	medida	1 Quad 07	2 Quad 07	3 Quad 07	Set 2007	1m - 49 sem 07	15m - Jan 07	60m - 1 Trim 07
1	desemp	-0,4031		-0,4067	4,6471	0,0010	0,8400	-0,5545
	result	-308		-484	474	10	210	-356
	n. <sup>o</sup> neg	135		189	24	16	35	85
2	desemp	0,2486		-0,1818	1,4839	-0,0153	0,4769	-0,7668
	result	136		-116	230	-153	93	-572
	n. <sup>o</sup> neg	175		298	50	22	53	106
3	desemp	-0,7679		0,3906	0,8958	0,0209	-0,0767	0,0625
	result	-493		298	215	209	-25	24
	n. <sup>o</sup> neg	49		92	19	5	19	48
4	desemp	-0,5271		3,1025	9,4894	-0,0113	-0,1128	0,9091
	result	-301		757	446	-113	-29	140
	n. <sup>o</sup> neg	185		267	45	18	45	65
5	desemp		-0,7439	-0,8479	0,7059	0,0022	1,8841	-0,4609
	result		-366	-524	60	22	260	-212
	n. <sup>o</sup> neg		57	104	15	2	32	73

Quadro 5.4: Desempenho em séries de teste *out-of-sample* dos melhores indivíduos treinados em séries de 5 minutos

As quatro execuções do GA relativas a testes no 1.<sup>o</sup> qua-

drimestre de 2007 (tipo A) denotam resultados predominantemente negativos. Quanto às séries de tipo B, respeitantes a execuções primárias, dois indivíduos alcançaram valores de desempenho positivos. O indivíduo referente à quinta execução do GA não consegue obter um desempenho positivo na série referente ao segundo quadrimestre de 2007 (tipo B), nem na referente ao terceiro quadrimestre de 2007.

No primeiro mês posterior ao quadrimestre de evolução, todos os indivíduos produzem um resultado positivo, tendo o indivíduo referente à quarta execução do algoritmo um desempenho muito alto.

Nas séries referentes às outras frequências, o desempenho dos cinco indivíduos é misto, oscilando entre o positivo e o negativo.

### 5.3.3 Indivíduos treinados em séries de 15 minutos

Os melhores indivíduos, de cada execução do GA, que evoluíram em séries de quinze minutos foram testados em sete séries *out-of-sample*, das quais quatro são da mesma periodicidade. As restantes três séries dizem respeito a periodicidades de um, cinco e 60 minutos.

O quadro 5.5 contém os valores relativos ao desempenho obtido pelos melhores indivíduos relativos a cinco evoluções do GA, em séries *out-of-sample*.

Os mesmos valores podem ser visualizados graficamente na figura 5.6

O ano de 2005, série de tipo A, registou três valores de desempenho negativos e apenas um positivo.

No que às séries de tipo B diz respeito, apenas a primeira execução do sistema se revelou negativa, tendo as outras três relativas a evoluções primárias registado valores positivos e relativamente altos de desempenho. O indivíduo

série	medida	Ano 05	Ano 06	Ano 07	1 Trim 07	Jan 07	1m - 49 sem 07	5m - Set 07	60m - 1 Trim 07
1	desemp	-0,4179		-0,0641	0,1502	-0,6798	0,0000	1,8485	-0,4439
	result	-369		-37	38	-172	0	61	-186
	n.º neg	163		70	12	5	0	5	25
2	desemp	0,7583		3,6995	0,3210	0,3462	-0,5738	3,2353	0,0270
	result	436		751	52	36	-35	165	6
	n.º neg	241		219	55	22	12	34	28
3	desemp	-0,8905		2,5717	0,4680	-0,4512	0,2737	5,2475	1,3425
	result	-1415		1471	139	-134	26	530	243
	n.º neg	97		51	10	5	2	7	5
4	desemp	-0,9408		2,1609	0,6429	-0,3762	-0,6722	7,4938	0,5592
	result	-1573		1303	270	-158	-121	607	137
	n.º neg	134		72	15	7	2	7	13
5	desemp		3,0778	0,1444	1,0000	-0,0513	0,0000	2,8000	-0,0766
	result		554	55	110	-4	0	98	-16
	n.º neg		85	59	12	4	0	5	19

Quadro 5.5: Desempenho em séries de teste *out-of-sample* dos melhores indivíduos treinados em séries de 15 minutos

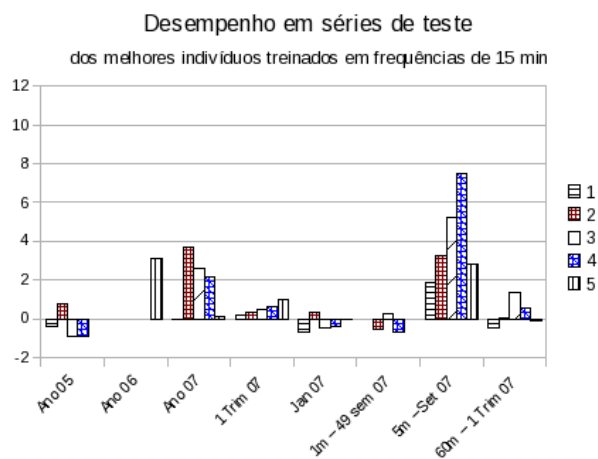


Figura 5.6: Séries 1 a 4 treinadas com base no ano de 2006; série 5 treinada com base no ano de 2005.

relativo à evolução secundária registou um comportamento semelhante aos três últimos, embora a sua série de tipo B fosse diferente. Este indivíduo produziu no Ano de 2007 um resultado ligeiramente positivo.

No primeiro trimestre de 2007, todos os cinco indivíduos obtiveram valores de desempenho positivos, embora modestos.

Tomado em consideração apenas o mês de Janeiro de 2007, o resultado foi maioritariamente negativo, tendo apenas um indivíduos alcançado resultado positivo.

Relativamente às séries de frequências diferentes, destaca-

se a série de cinco minutos, com um desempenho positivo por parte de todos os indivíduos, sendo que apenas um valor é relativamente baixo. As séries de um e sessenta minutos registam um comportamento misto, ora com resultados positivos, ora com negativos. Convém, no entanto, ter em consideração que nas séries de um e cinco minutos, o número de negócios gerados por quatro dos cinco indivíduos foi relativamente baixo.

### 5.3.4 Indivíduos treinados em séries de 60 minutos

O melhor indivíduo de cada uma das cinco evoluções realizadas pelo GA em séries de 60 minutos foi testado em sete séries *out-of-sample*, sendo quatro das quais relativas a periodicidades de 60 minutos e as outras três relativas a frequências de um, cinco e quinze minutos.

Os valores de desempenho obtidos nestas 35 simulações constam do quadro 5.6.

série	medida	Ano 2004	Ano 2006	Ano 2007	1 trim 2007	1 mês 2007	15m - 1 mês	5m - 9 mês	1m - 49 semana
1	desemp	-0.9197		-0.3927	-0.2885	-1,000	-1,0000	0,0000	0,0000
	result	-916		-130	-30	-101	-45	0	0
	n.º neg	47		15	4	2	1	0	0
2	desemp	4.4828		4.3446	0.1723	-0.1489	-0.0043	-0.6964	0.5424
	result	1690		1412	56	-46	-1	-172	64
	n.º neg	284		254	69	28	62	89	37
3	desemp	6.8535		0.3991	2.0541	1.9189	0.0278	0.6429	0.0000
	result	1357		93	228	213	1	9	0
	n.º neg	73		61	21	12	8	2	0
4	desemp	-0.4626		2.8108	-0.6409	-0.3320	0.6395	1.6053	1.4500
	result	-1034		1307	-298	-81	110	183	29
	n.º neg	150		68	17	8	8	5	1
5	desemp		3.2589	-0.5890	-0.1551	0.0558	-0.5362	0.2139	-0.1574
	result		1095	-708	-47	12	-163	37	-17
	n.º neg		201	172	46	20	50	83	51

Quadro 5.6: Desempenho em séries de teste *out-of-sample* dos melhores indivíduos treinados em séries de 60 minutos

Os mesmos valores podem ser visualizados graficamente na figura 5.7

Na série de tipo A, ano de 2004, dois indivíduos tiveram desempenho negativo e, outros dois, desempenho positivo, sendo que estes últimos valores são relativamente grandes.

Quanto às séries de tipo B, o resultado é positivo em todas, com excepção da primeira execução. A segunda e

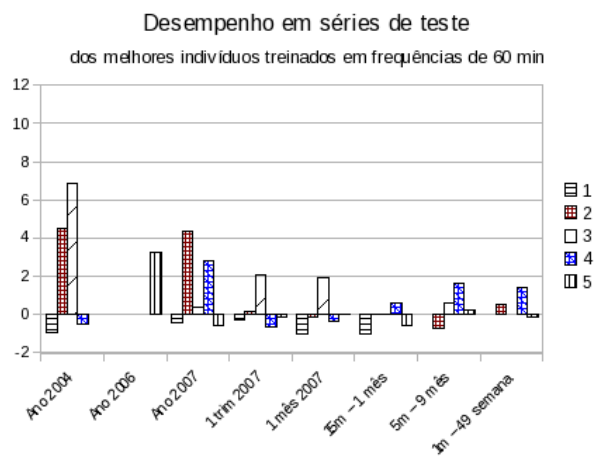


Figura 5.7: Desempenho em séries de teste por indivíduo treinado em série de 60 min

quarta execuções alcançam resultados relativamente fortes. No que à quinta execução diz respeito, embora numa série diferente, ano de 2006, é conseguido um desempenho idêntico aos melhores dos indivíduos treinados na evolução primária. No entanto, este indivíduo obtém um desempenho negativo quando testado no ano de 2007.

O desempenho relativo aos testes efectuados no primeiro trimestre de 2007 é irregular, oscilando entre o positivo e o negativo. O mesmo sucede quando é analisado somente o mês de Janeiro de 2007.

Quanto às três outras séries de frequências diferentes da de evolução, os valores de desempenho obtido alternam entre o positivo e o negativo, havendo a ressaltar a escassez de negócios efectuados em vários dos testes.

#### 5.4 Estatística geral e aferição de relações

Depois de apresentados os resultados relativos aos testes efectuados, dever-se-ão analisar os mesmos por um prisma que ponha em evidência algum significado mais geral e

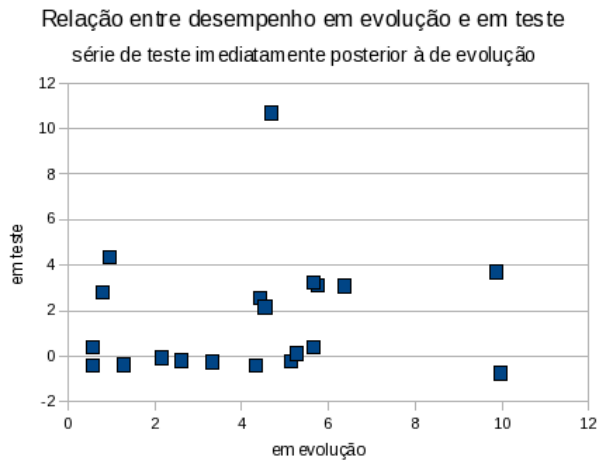


Figura 5.8: Relação entre desempenho obtido em séries de evolução e de testes

verifique a existência de relações entre diversos factores.

#### 5.4.1 Relação entre o desempenho em séries de evolução e em séries de teste

Poder-se-á colocar, desde logo, a hipótese de aferição da existência de uma relação linear entre o desempenho numa série de teste futura e o correspondente desempenho na série de evolução. Foram escolhidas como séries de teste as de tipo B relativas a execuções primárias.

O gráfico 5.8 apresenta graficamente os pares de valores de desempenho obtidos.

Foi calculado o valor do coeficiente de correlação entre as duas variáveis: desempenho em série de evolução e na respectiva série de teste. Foram obtidos resultados de  $r = 0,13$  e  $r^2 = 0,02$ , que sugerem a não existência de uma relação linear entre as duas variáveis, pelo que resultados futuros não serão previsíveis com base nos desempenhos individuais nas séries de evolução. No entanto, em 60% das observações, ao valor positivo na série de evolução, sucedeu um valor igualmente positivo no desempenho da série de teste respectiva.

#### 5.4.2 Estatística geral de desempenho nos testes efectuados, por tipo de série

	tipo A	tipo B	tipo C
média	1,0739	1,6992	1,4235
mínimo	-0,9408	-0,7439	-0,6409
máximo	6,8535	10,6875	9,4894
desvio padrão	2,5286	2,6404	2,4179
n.º dados	16	20	16
n.º positivos	7	12	12
% positivos	43,75	60,00	75,00
n.º observações	4	5	4

Quadro 5.7: Estatística geral de desempenho nos testes, por tipo de séries

A série do tipo A apresentou os piores resultados dos três. Teve a pior média de desempenho, o indivíduo com resultado mais negativo e o menor dos desempenhos máximos. A percentagem de indivíduos que conseguiram resultados positivos é a mais baixa de todas.

A série do tipo C revela os melhores resultados globais. A média de desempenho, embora menor do que a obtida pelos indivíduos do tipo B, superioriza-se a esta em termos relativos, por dizer respeito a períodos de tempo com cerca de um quarto da duração dos do tipo B. Em termos de desempenho mínimo, foi a que apresentou o resultado menos negativo, ficando o resultado máximo pouco aquém do conseguido pelo melhor indivíduo do tipo B. A percentagem de indivíduos que obtiveram desempenho positivo, na série, é a mais alta dos três tipos considerados, com 75%.

#### 5.4.3 Estatística de desempenho nos testes efectuados, por frequência e tipo de série

O quadro 5.8 corresponde ao anterior 5.7 detalhado por frequência. Pode observar-se que as médias de desempenho nas séries de tipo B e C são positivas qualquer que

	tipo	1 min	5 min	15 min	60 min
média	A	2,5422	-0,3624	-0,3727	2,4885
média	B	1,9914	0,4321	2,2892	2,0841
média	C	0,8453	4,1291	0,3955	0,3243
mínimo	A	-0,0139	-0,7679	-0,9408	-0,9197
mínimo	B	-0,3711	-0,7439	-0,0641	-0,3927
mínimo	C	-0,0039	0,8958	0,1502	-0,6409
máximo	A	5,8462	0,2486	0,7583	6,8535
máximo	B	10,6875	3,1025	3,6995	4,3446
máximo	C	1,8873	9,4894	0,6429	2,0541
desvio padrão	A	2,4112	0,3763	0,6841	3,2922
desvio padrão	B	4,3512	1,3854	1,2841	1,7884
desvio padrão	C	0,8594	3,4078	0,1818	1,0395
% positivos	A	75,00	25,00	25,00	50,00
% positivos	B	40,00	40,00	80,00	80,00
% positivos	C	50,00	100,00	100,00	50,00
n. <sup>o</sup> observações	A	4	4	4	4
n. <sup>o</sup> observações	B	5	5	5	5
n. <sup>o</sup> observações	C	4	4	4	4

Quadro 5.8: Estatística de desempenho nos testes, por frequência e tipo de séries

seja a frequência de dados considerada. As séries de tipo C apenas apresentam dois valores negativos de desempenho, para as frequências de 1 e 60 minutos, em contraste com a totalidade de valores negativos para as séries de tipos A e B. A maior consistência de resultados máximos é conseguida pelas séries de tipo B, com todos os valores acima de 3, secundada pelas de tipo C. Apenas nas séries de tipo C, a totalidade das execuções do GA resultou num resultado positivo, no caso das frequências de 5 e 15 minutos.

#### 5.4.4 Desempenho médio em testes, por frequência

Para cada frequência, foi calculado o desempenho médio global *out-of-sample* dos cinco indivíduos respectivos, cada qual nas séries de teste de dimensão equivalente às de treino. Assim, dos quatro indivíduos respeitantes à evolução primária, foram utilizados os valores de desempenho



nas séries anterior e posterior ao período de evolução, enquanto que, no que respeita aos indivíduos respeitantes à evolução secundária, foram aproveitados os valores respeitantes aos testes nas duas séries seguintes, da mesma frequência.

1 min	5 min	15 min	60 min
2,0804	-0,0137	1,0009	1,9786

Quadro 5.9: Desempenho médio em séries de teste, por frequência

O quadro 5.9 sintetiza os valores obtidos das médias. Tendo presente as quatro observações registadas, não é evidente a existência de qualquer relação entre o desempenho médio e a frequência utilizada em testes. Três dos valores obtidos são positivos. Apenas a série relativa à frequência de 5 minutos apresenta um valor negativo.

#### 5.4.5 Comparação entre o desempenho em testes de séries futuras e o de séries passadas

Na secção 2.2, foi referido que Yao e Tan [28] tentaram obter um desempenho excepcional do seu algoritmo, no longo prazo, com a realização de contínuas séries de evoluções, de forma a evitar um degradamento nos resultados *out-of-sample* com o aumento da distância temporal face ao período de treino. Não se pretendendo simular situação idêntica neste trabalho, optou-se por comparar o desempenho médio obtido pelos indivíduos provindos de evoluções primárias nas séries antes (tipo A) e após o período em que evoluíram (tipo B).

Dos resultados espelhados no quadro 5.10, constata-se que os indivíduos em teste obtiveram um desempenho médio superior nas séries posteriores aos respectivos períodos de evolução, em relação ao das séries anteriores. É ainda de notar que o desempenho médio nos períodos posteriores

séries teste	1 min	5 min	15 min	60 min
anteriores	2,5422	-0,3624	-0,3727	2,4885
posteriores	2,5820	0,7262	2,0920	1,7905

Quadro 5.10: Desempenho médio em séries de teste anteriores e posteriores ao respectivo período de evolução, por frequência

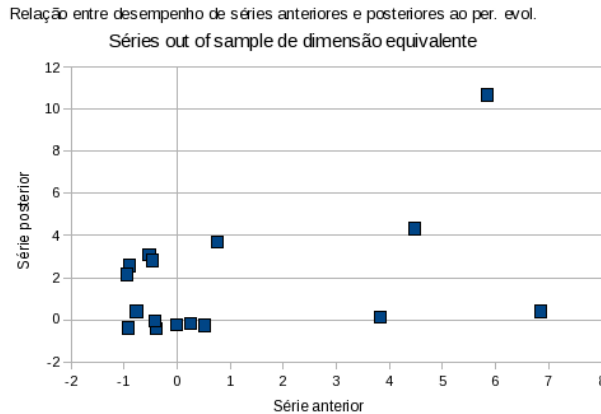


Figura 5.9: Relação entre desempenho obtido em séries de testes imediatamente anteriores e posteriores ao período de evolução

foi positivo para todas as frequências analisadas.

#### 5.4.6 Relação entre o desempenho em séries de testes anteriores e posteriores ao período de evolução

Dado que o comportamento de um sistema de negociação, no futuro, será sempre uma incógnita, será útil analisar se haverá uma relação linear entre desempenhos em séries de teste futuras e em séries de teste passadas. Por exemplo, para a frequência de 15 minutos, o desempenho no teste à série relativa ao ano de 2004 poderá servir de bom indicador ao desempenho no teste ao ano de 2007? As séries escolhidas são as mesmas da secção anterior.

A figura 5.9 mostra, em termos gráficos, a distribuição dos pares de valores considerados, tendo em conta que apenas foram utilizados os valores de desempenho relativos a indivíduos gerados em evoluções primárias

Foi calculado o valor do coeficiente de correlação entre as duas variáveis: desempenho em série anterior (tipo A) e posterior (tipo B) ao de evolução. Os resultados obtidos de  $r = 0,44$  e  $r^2 = 0,2$ , sugerem a não existência de uma relação linear entre as duas variáveis, pelo que desempenhos em períodos de testes anteriores não terão utilidade em ajudar a prever desempenhos em períodos futuros.

#### **5.4.7 Relação entre desempenho e número de negócios em séries de testes**

À semelhança do realizado para as séries de evolução, dado conta na secção 5.2, foi verificada a existência de uma relação linear entre o desempenho obtido em séries de testes *out-of-sample* e o número de negócios efectuados. Esta verificação ficou restringida às séries de testes de dimensão equivalente às de evolução. Nestas foram considerados os períodos anteriores e imediatamente posteriores aos períodos de evolução, para as séries de evolução primária, e os dois períodos seguintes à série de evolução, no caso de evolução secundária. Foram assim obtidos vinte pares de valores, a que corresponde um coeficiente de correlação  $r = -0,05$  e  $r^2 = 0,003$ , o que sugere a ausência de correlação linear entre as duas variáveis.

A figura 5.10 mostra a distribuição dos pares de valores, em gráfico.

#### **5.4.8 Relação entre desempenho numa série futura e numa sua subsérie**

Outra relação que se poderia considerar como vantajosa para um investidor consistiria em aquilatar do desempenho de um indivíduo num período de teste alargado, tendo por base o seu desempenho num subperíodo inicial. Caso existisse uma relação linear entre os dois valores de desem-

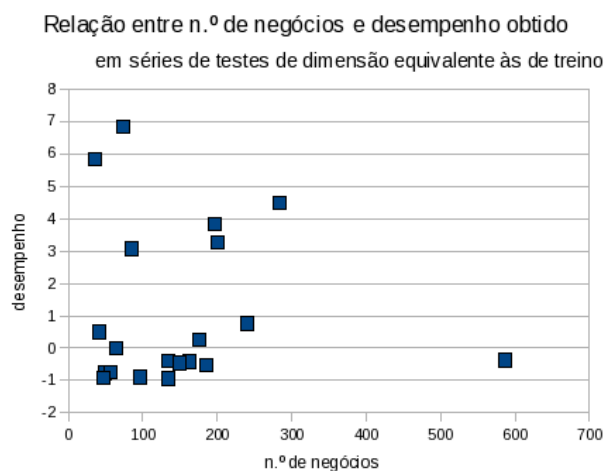


Figura 5.10: Relação entre n.º de negócios e desempenho obtido em séries de testes

penho, poder-se-ia decidir sobre a continuação, ou não, da execução de um sistema de negociação ao fim de um período inicial de experiência. Por exemplo, no caso da negociação em séries de frequência de cinco minutos, o desempenho obtido durante a negociação no mês de Setembro de 2007, avaliado no final desse subperíodo, serviria para decidir pela continuação, ou não, da utilização do mesmo sistema de negociação durante os restantes três meses do terceiro quadrimestre de 2007.

Foram utilizados os valores de desempenho dos indivíduos referentes à evolução primária, nas séries de tipo B e em todas as séries de menor extensão de cada frequência. Os 24 pares de valores de desempenho obtidos distribuem-se graficamente tal como apresentado na figura 5.11.

Em termos estatísticos, foi calculado um coeficiente de correlação  $r = 0,0452$  e  $r^2 = 0,002$  o que sugere a ausência de uma relação linear entre os valores de desempenho das duas variáveis. É, no entanto, de notar que em 2/3 das observações, o valor de desempenho na série longa teve o mesmo sinal do que na subsérie inicial respectiva. Se,

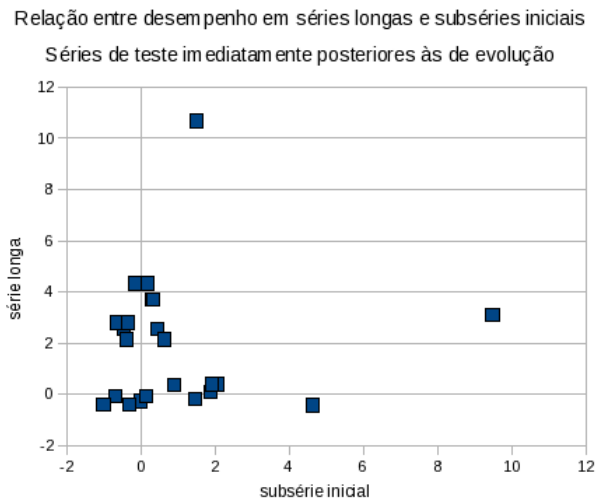


Figura 5.11: Relação entre desempenho obtido em séries de testes longas e suas subséries iniciais

das observações antes consideradas, forem retiradas aquelas que dizem respeito aos menores subperíodos iniciais para as séries de 15 e 60 minutos, que dizem respeito ao mês de Janeiro de 2007, a proporção de  $2/3$  referida no parágrafo anterior sobe para  $3/4$ .

#### 5.4.9 Adequação de séries às operações de evolução e teste

Na secção 2.1.2, referiu-se que Neely et al. [18], prescindiram de um determinado período de tempo, nas fases de evolução e selecção do seu GA, de forma a evitar enviesamentos nos resultados devido a uma apreciação muito significativa do USD, nesse espaço temporal. Neste texto, houve a intenção deliberada de não pôr de parte quaisquer períodos, tal como mencionado na secção 4.1. Esta decisão foi tomada por não se aceitar o argumento de que há séries temporais com formas mais ou menos indicadas para a utilização de sistemas de negociação, tal como expresso na secção 3.2.

Resta averiguar se o pressuposto assumido se verificou

correcto, comparando os resultados obtidos com a descrição antes efectuada da forma de cada uma das séries. A análise é efectuada em três vertentes.

Em termos de treino das populações, há a destacar o modesto desempenho das quatro execuções do GA referentes à evolução primária na frequência de 60 minutos, tanto comparando com as evoluções primárias nas séries de outras frequências, como em relação à série de evolução secundária da própria frequência. Tal como descrito na secção 3.2.1, a série de evolução referente aos anos de 2005 e 2006, na frequência de 60 minutos, não suscita, visualmente, características que a distingam das suas congéneres de outras frequências, por uma expressiva acentuação de regularidade ou irregularidade ou pela existência de picos ou fundos ostensivos. Também não se destaca quando comparada com a série de evolução secundária da mesma frequência, dando, aliás, a impressão de o comportamento de uma ser o inverso do da outra. O desempenho registado na evolução da população não parece, pois, poder atribuir-se à forma excepcional de uma série. Dado que apenas foi realizada a evolução de quatro populações, o nível de desempenho poderá ser fruto de uma casualidade estatística.

Outro factor em relevo é a não existência de um indivíduo evoluído na frequência de 5 minutos que apresente desempenho positivo nas séries de testes anterior e posterior à de evolução e que seja de dimensão equivalente a esta. Tanto a série referente ao primeiro quadrimestre de 2007 (figura 3.2 b)), como a referente ao terceiro (figura 3.4 a)) apresentam comportamentos maioritariamente monótonos de subida dos preços da moeda base, pelo que a forma das duas séries não parece ser determinante para a não existência de indivíduos com o duplo desempenho

positivo acima mencionado.

Um terceiro facto a destacar refere-se ao desempenho positivo conseguido pelos cinco indivíduos que foram testados na série do mês de Setembro de 2007, na frequência de 5 minutos. A forma da série não difere muito, no entanto, a olho nu, das séries de testes do primeiro e terceiro quadrimestres de 2007. Este comportamento positivo pode ficar a dever-se ao prazo mais curto desta série em relação às demais, devido ao aproveitamento de situações mais particulares de negociação.

## 5.5 Características dos melhores indivíduos

Depois de apresentados os resultados obtidos, quer em termos de evolução da população, quer em testes *out-of-sample* ao melhor indivíduo de cada execução do GA, interessa conhecer as características dos indivíduos que proporcionaram desempenhos mais estáveis. De entre os dezasseis indivíduos de evolução primária, foram escolhidos aqueles que apresentam desempenhos positivos nas respectivas séries de tipo A e B. Foram apurados dois indivíduos cada referentes à frequência de 1 e 60 minutos e um da frequência de 15 minutos. Nenhum indivíduo evoluído em séries de cinco minutos preencheu os requisitos para figurar no quadro 5.11 dos melhores indivíduos.

Quanto aos critérios de entrada em posição, verifica-se que nenhum dos cinco indivíduos usa a média móvel de longa duração - regra2, mas apenas a de curto prazo (obrigatória). O parâmetro relativo ao número de períodos da regra 1 assume valores distribuídos por todo o universo de valores possíveis, sem qualquer concentração aparente.

A regra 3 - “Posição num Trading Range” - é, por seu turno, sempre utilizada pelos cinco melhores indivíduos.

O parâmetro *tr1\_periodos* assume valores entre o terceiro mais pequeno e o maior dos possíveis. O sinal “<”, relativo ao parâmetro *tr1\_sinal* foi sempre escolhido. Quanto a *tr1\_percent*, que tinha como valores possíveis, todos os inteiros entre 0 e 100, verifica-se que a escolha de valores se resumiu ao intervalo inter-quartis.

A regra 4 - “Repetição de cotações” - foi usada por apenas dois indivíduos. Nestas duas soluções, o parâmetro *tr2\_sinal* assumiu o valor “<”.

No que respeita à situação de entrada em posição apenas resta a observação do comportamento de *p1*. O valor mais alto assumido por este parâmetro foi de 0 e o mais baixo de -7. Para quatro dos indivíduos, os valores situam-se, no entanto, no intervalo fechado entre 0 e -2.

A saída de posição é, em primeira mão, definida pelos parâmetros *tp1* e *sl1*. A distribuição dos valores tomados por estes parâmetros revela que o GA escolheu, para estes cinco indivíduos, valores relativamente altos. Tanto quanto a *tp1* como a *sl1*, só um indivíduo apresentou valores abaixo de 50, para um universo possível entre 0 e 100.

A regra 7 - “Evolução rápida” - foi utilizada por apenas dois dos indivíduos para ajudar na decisão de saída de posição. Os valores escolhidos por estas duas soluções para o parâmetro *rap\_sl* foram muito próximos do limite superior dos valores possíveis: 100.

Todos os indivíduos utilizaram a regra de “Evolução lenta” - regra 8. *len\_percent* assumiu valores entre 16 e 71, de entre os possíveis 0 a 100. O parâmetro *len\_periodos* foi escolhido com valores relativamente baixos, entre 3 e 23. *len\_tp* assumiu valores na metade superior da distribuição de valores possíveis, com uma exceção: valor 15. Um indivíduo possui para o parâmetro *len\_sl* um valor



próximo do limite inferior do seu domínio. No entanto, três outros indivíduos utilizaram o parâmetro com valores situados no quarto quartil.

A regra 9 - “Melhoria de cenário” - apenas foi utilizada por um dos cinco melhores indivíduos.

A última regra considerada como decisora para a saída de posição - “Pioria de cenário” - foi activada por quatro indivíduos. *pio\_tp* registou valores ao longo de todo o seu domínio. Quanto a *pio\_sl*, nenhum dos quatro indivíduos utilizou o parâmetro com valores pertencentes ao primeiro quartil dos valores possíveis.

Para além do conhecimento detalhado dos cinco melhores indivíduos, poder-se-á interpretar os padrões genéticos obtidos. Assim, quanto à entrada em posição: a) são preferidas as médias de períodos mais curtos para a definição de tendências de evolução dos preços; b) as ordens de entrada em posição surgem quando o preço actual retrocede de um máximo recente e se aproxima do mínimo do trading range constituído pelas últimas cotações; c) se houver abertura de posição, o preço de entrada deverá ser ao máximo (mínimo) do período anterior para PL (PC) ou a um preço ligeiramente inferior (superior); No que respeita à saída de posição: d) são estabelecidos valores iniciais relativamente afastados do preço de entrada, para *take profit* e *stop loss*; e) é verificado se a expectativa de evolução de preços corresponde ao esperado, num curto período após o momento de entrada em posição e, se aquela não se confirmar, é forçada uma saída antecipada de posição.

A figura 5.12 mostra a negociação efectuada pelo indivíduo evoluído na terceira execução do GA referente à frequência de 60 minutos, nos primeiros quatrocentos períodos referentes ao primeiro trimestre de 2007. Os triângulos virados para cima indicam a tomada de PL e os

virados para baixo mostram a abertura de uma PC. Os quadrados são sinónimo de que uma ordem *stop loss* foi activada. Quando é a ordem *take profit* a responsável pelo fecho de uma posição, o símbolo usado é o círculo.

## 5.6 Testes computacionais

As evoluções e testes das diversas séries temporais foram realizadas num computador portátil com processador Intel Core 2 Duo T5500 @ 1,66GHz e memória RAM disponível de 2GB.

As operações foram executadas, na totalidade, com recurso a software open-source. O sistema operativo utilizado foi o Slackware<sup>1</sup> Linux 12.1. Como linguagem de programação foi eleita a Python<sup>2</sup>, versão 2.5.2, pela sua rapidez de desenvolvimento. Para diminuir o tempo de execução dos algoritmos, foi usado o módulo Psycho<sup>3</sup> 1.6.0. A base de dados utilizada foi a PostgreSQL<sup>4</sup>, versão 8.3.3 e o módulo de acesso a esta, pela linguagem de programação, foi o psycopg2<sup>5</sup>, versão 2.07. Para a construção do gráfico de evolução do desempenho da população houve recurso ao módulo matplotlib, versão 0.98.1<sup>6</sup>.

As operações de treino, testes *out-of-sample* e registo na base de dados relativas às evoluções das séries mais curtas, de 60 minutos, demoraram entre um mínimo de 1h e 15min e um máximo de 3h cada.

As séries mais longas, tiveram tempos de execução que variaram entre uma e catorze horas, cada uma.

---

<sup>1</sup>Sítio em <http://www.slackware.org>

<sup>2</sup>Sítio em <http://www.python.org>

<sup>3</sup>Sítio em <http://psyco.sourceforge.net>

<sup>4</sup>Sítio em <http://www.postgresql.org>

<sup>5</sup>Disponível em <http://www.initd.org/pub/software/psycopg>

<sup>6</sup>Sítio em <http://matplotlib.sourceforge.net>

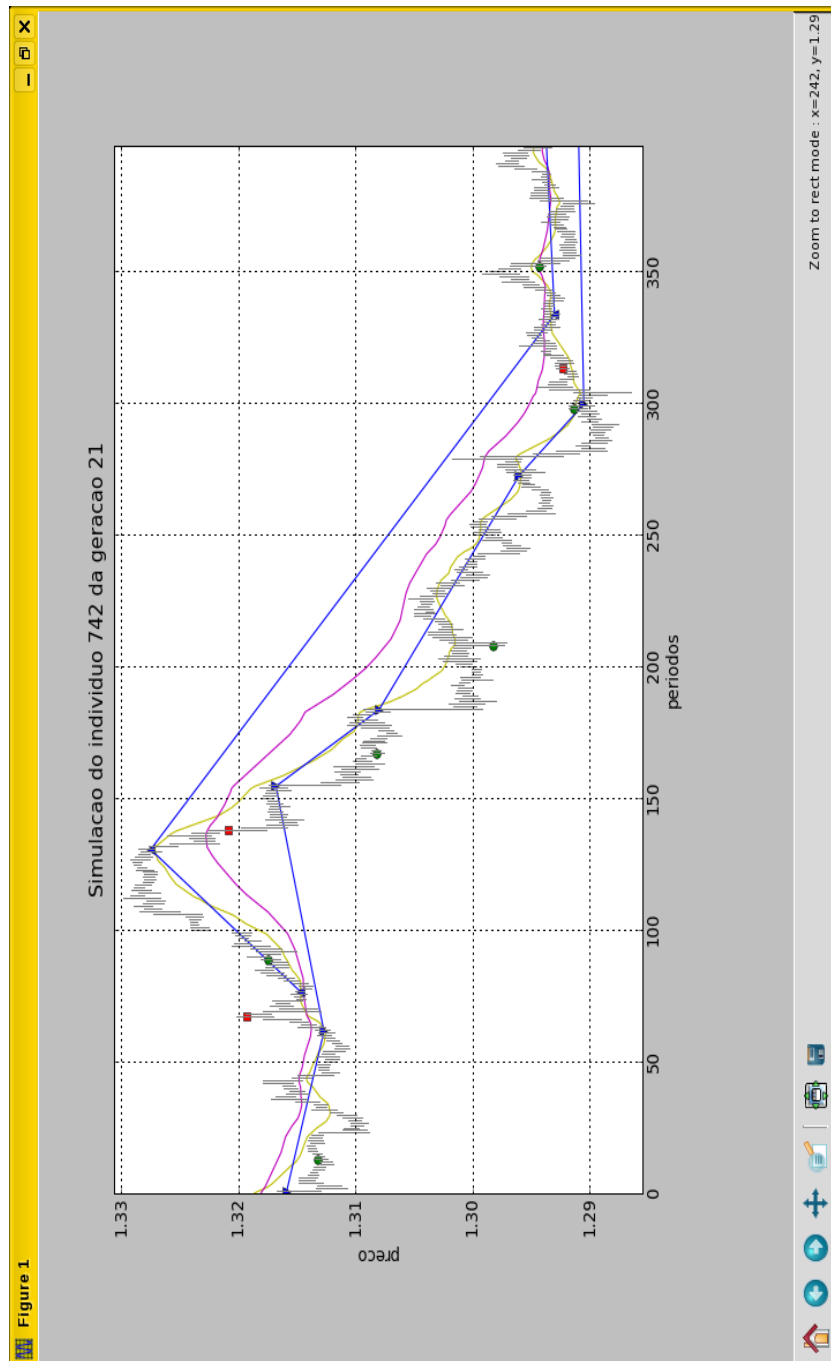


Figura 5.12: Negociação do indivíduo da 3.<sup>a</sup> série da frequência de 60 minutos nos primeiros 400 períodos do 1.<sup>o</sup> trimestre de 2007

frequência	1 min	1 min	15 min	60 min	60 min
série	2	4	2	2	3
emac_periodos	26	3	18	39	23
emal_bool	f	f	f	f	f
emal_periodos	91	31	151	176	73
tr1_bool	t	t	t	t	t
tr1_periodos	10	6	24	25	16
tr1_sinal	<	<	<	<	<
tr1_percent	0,27	0,58	0,41	0,68	0,35
tr2_bool	f	t	f	t	f
tr2_periodos	20	22	20	14	9
tr2_sinal	>	<	>	<	<
tr2_percent	0,79	0,85	0,74	0,63	0,56
tr2_vezes	11	17	10	4	3
p1	0	0	-1	-2	-7
tp1	39	69	58	72	89
sl1	61	62	66	42	50
rap_bool	t	f	f	f	t
rap_percent	10	83	46	85	34
rap_periodos	17	8	6	9	4
rap_tp	36	12	26	66	55
rap_sl	99	3	23	53	95
len_bool	t	t	t	t	t
len_percent	40	16	60	42	71
len_periodos	6	23	14	3	11
len_tp	15	89	84	68	73
len_sl	96	41	75	7	87
mel_bool	f	f	f	f	t
mel_tp	69	78	56	79	89
mel_sl	78	34	63	91	27
pio_bool	t	t	t	f	t
pio_tp	70	42	91	57	15
pio_sl	60	27	44	36	84

Quadro 5.11: Os melhores indivíduos

## Capítulo 6

### Conclusões

O objectivo que presidiu à realização deste trabalho foi o da obtenção de estratégias de transacção que permitissem a obtenção de lucros sistemáticos na negociação do EUR/USD, no mercado de Forex, com a utilização de regras objectivas, baseadas na análise técnica das cotações passadas daquele activo. O sucesso em tal demanda pressuporia uma discordância com as conclusões emanadas da Teoria da Eficiência dos Mercados.

Foi, pois, construído um sistema de negociação que aproveitasse tendências de evolução dos preços nas séries temporais consideradas, de quatro frequências distintas: 1, 5, 15 e 60 minutos. Para a entrada numa posição, o sistema de negociação tem em conta um máximo de cinco regras e o mínimo de duas. A saída de posição, com lucro ou prejuízo, é definida no momento em que uma posição é aberta. No entanto, considerou-se a hipótese de o sistema poder adequar os valores de saída, antes estabelecidos, de acordo com as condições do mercado, em cada período, tal como reveladas por quatro outras regras de utilização facultativa. Houve sempre a preocupação de adequar o sistema às condições reais de negociação, o que se traduziu por considerar os custos de transacção, na forma do *spread* do par de divisas, por um lado, e, por outro, na tomada

de decisões baseadas em critérios de precaução.

Dada esta multiplicidade de regras, cada qual fazendo uso de um ou mais parâmetros, obteve-se um espaço máximo de combinações possíveis, para os valores dos diversos indicadores técnicos e variações na estrutura de cada solução, que ascende a acerca de  $1,69 \times 10^{43}$ . Perante a impossibilidade de estudar a totalidade do universo de possíveis soluções, houve o recurso a um algoritmo genético que, de forma eficaz e eficiente, pudesse encontrar boas soluções finais. Cada execução do GA resumiu-se à determinação de um máximo de  $500 = 50 \times 10$  soluções, cada qual expressa como um indivíduo, resultante da evolução de uma população inicial e constante de dez indivíduos ao longo de 50 gerações. Por questões de demora no tempo de execução do algoritmo, para cada uma das frequências consideradas, apenas houve lugar à evolução de cinco populações iniciais, sendo uma delas em série temporal distinta das restantes. Nas séries de evolução, os melhores indivíduos de cada execução do GA obtiveram sempre resultados positivos. Embora tendo em conta que o GA apenas foi executado uma vez, por frequência, em série diferente, não foram notórias diferenças substanciais ao nível de desempenho, em relação às evoluções primárias ocorridas.

Cada melhor indivíduo resultante do processo de evolução foi testado em séries *out-of-sample*, de frequência idêntica à de treino, em número de três ou quatro, e, por uma vez cada, para as restantes três frequências. Ao todo, foram efectuados 130 testes.

No que respeita aos testes efectuados em séries da mesma frequência da de evolução, as séries de tipo A registaram os piores valores. As séries de tipo B e C partilharam os méritos dos melhores desempenhos, embora deva ser tido em conta que a extensão das séries de tipo C representam

cerca da quarta parte das séries de tipo B. Numa análise detalhada por frequência de dados, constata-se que as séries dos tipos B e C apresentam valores médios de desempenho sempre positivos. A proporção do número de execuções do GA com desempenho positivo é mais elevada nas séries de tipo C, seguida das de tipo B. Destes factos, constata-se a tendência de o sistema de negociação ser mais forte quando testado em séries posteriores aos períodos de evolução. A negociação nestas séries não é, contudo, coroada com um desempenho positivo para todas as soluções encontradas. No entanto, se for considerado um cabaz com um conjunto de soluções em negociação paralela, para cada frequência, é de esperar um desempenho positivo e relativamente elevado em séries *out-of-sample* posteriores, mas adjacentes, ao período de evolução.

Tendo em consideração os níveis de desempenho atingidos, nos diversos testes, e a extensão das séries adstritas àqueles valores, parece poder inferir-se que será mais favorável a negociação das séries de 1 minuto, apesar de o custo de transacção, na forma do *spread*, ter nesta frequência, um impacto negativo superior. Se se considerar, por exemplo, o desempenho médio dos cinco indivíduos de cada série quando testados nas séries de tipo B, a valor atingido na série de um minuto (1,9914) é pouco menor em relação ao melhor de todos, alcançado na frequência de 15 minutos (2,2892). A relevância é dada pela extensão relativa das séries de testes, consideradas, das duas frequências: as séries de um minuto têm apenas cerca de um doze-avos da extensão das de 15 minutos. Uma vez que o rácio de desempenho é constituído, no numerador, pelo resultado obtido, em termos absolutos, as séries de tipo B, por serem mais extensas, oferecem a possibilidade de obtenção de valores mais elevados para esta variável. Quanto ao denomi-

nador, ele expressa a quebra máxima incorrida por cada solução obtida. Dada a condição pré-estabelecida de considerar um número mínimo de negócios em cada sistema, o valor da quebra máxima pode considerar-se como estando menos sujeito a significativos agravamentos, quando são consideradas séries de dados mais extensas, em relação a uma situação em que esta restrição não existisse. De todo este conjunto de razões, pode concluir-se, pois, pelo melhor desempenho relativo do sistema em testes *out-of-sample* do tipo C. Um desempenho excepcional no longo prazo pode estar dependente, portanto, da re-execução do GA em intervalos regulares, de forma a encontrar novas soluções, mais adequadas aos novos períodos.

No que respeita a testes efectuados em séries de frequências diferentes da da respectiva evolução, estes só foram efectuados em séries de extensão equivalente às do tipo C. Num grande número de vezes, o sistema acabou por não gerar qualquer posição de compra ou venda do activo financeiro, ou gerou posições, mas em número muito reduzido. Os valores de desempenho alcançados oscilam entre o positivo e o negativo, sendo, no entanto, positivos na maioria dos testes, em 55%. Este conjunto de testes tinha como intuito o de analisar a aplicabilidade das soluções obtidas em séries de frequência diversa, de forma a concluir pela existência, ou não, de uma solução geral. Tal não foi possível, na sua plenitude, por não se terem considerado, à partida, séries de uma extensão equivalente às do tipo A e B.

Foi analisada um conjunto de possíveis relações lineares entre diversas variáveis.

Tanto na fase de evolução das populações, como na de testes, não houve indícios da existência de qualquer relação linear, medida pelo coeficiente de correlação “r” entre



o desempenho obtido e o número de negócios efectuado, dados os baixos valores de  $r^2$ .

Poder-se-ia colocar a hipótese de existir uma relação linear entre o desempenho em séries futuras e o desempenho na correspondente série de evolução. O baixo valor do coeficiente de correlação não sugere uma confirmação de tal hipótese. Dado que a operação de evolução admite um maior ou menor grau de ajustamento da solução obtida ao conjunto de dados utilizados, foi estudada a possibilidade de existir uma relação linear entre desempenhos em séries de testes passadas e futuras, em relação ao momento de evolução. Conclui-se igualmente pela falta de confirmação de existência de correlação.

Analisou-se se o desempenho num período poderia ser extrapolado a partir do desempenho num seu subperíodo, o que poderia ter vantagens em termos da tomada de decisões sobre as melhores soluções a adoptar. O coeficiente de correlação obteve um valor muito baixo o que sugere a inexistência de relação linear entre os valores obtidos. No entanto, verifica-se que, em dois terços das observações, o valor de desempenho nas duas séries tem o mesmo sinal, o que poderá indiciar uma não completa falta de relação entre as duas variáveis.

Poder-se-á, no entanto, colocar a hipótese de os valores alcançados e das conclusões obtidas estarem fortemente condicionadas pela escolha arbitrária das séries utilizadas pelo algoritmo. De entre as séries utilizadas, quer para evolução, quer para testes, não se verificou a existência de padrões de evolução dos preços que demonstrassem ter um comportamento excepcional, numa apreciação humana. Os valores obtidos para o desempenho, resultado da negociação e número de negócios efectuados não parece estar distorcido nas comparações efectuadas. Conclusões

mais seguras só poderão ser obtidas, no entanto, com a utilização de outras séries temporais e com a execução do algoritmo num número de vezes significativamente superior ao ocorrido.

Foram considerados cinco indivíduos como sendo os melhores, por obterem resultados positivos nas séries de evolução, anteriores e posteriores àquela, de dimensão equivalente, o que traduz uma certa imutabilidade das soluções encontradas. Estas vão ao encontro do pretendido, na definição inicial do problema: soluções que façam sentido em termos humanos e que sejam tão simples quanto possível. O modelo “típico” escolhido pelo GA resultou numa opção de acompanhamento de tendências de evolução de preços em prazos mais curtos, com o aproveitamento de alguns retrocessos. A evolução de uma posição aberta é acompanhada, após um período inicial de carência, e sempre que o GA detecta que a tendência de evolução se esbateu, é forçada a saída de posição.

Tendo presente que a solução óptima para cada execução do algoritmo é uma incógnita, o GA obteve, ainda assim, valores médios de desempenho positivos e relativamente grandes, nas séries de evolução. Verificou-se que uma grande parte dos incrementos significativos resultou da introdução de imigrantes, ou seja, de soluções aleatórias. Tal situação poderá significar que o algoritmo construído terá uma margem de progressão com a mudança de alguns pilares em que foi assente. A consideração de uma população inicial de dimensão bastante superior à considerada poderia obviar a necessidade sentida de inserção de soluções aleatórias quando a convergência entre os indivíduos era notória. A probabilidade de mutação pode ter sido considerada num nível baixo demais, face às características do problema. Estas suspeitas não puderam

ser comprovadas, pois o GA não foi objecto, ele próprio de uma evolução, no seu algoritmo, que o optimizasse para este tipo de problema, por insuficiência de tempo.

Terminado este trabalho, dadas as conclusões e tendo em conta novas hipóteses entretanto surgidas foram identificados novos caminhos que poderão ajudar à investigação de um futuro sistema de negociação.

1. Os períodos de evolução deverão ser mais extensos, de forma a abarcar padrões de evolução de preços, o mais diversos possíveis.
2. Os testes deverão ser repetidos, para cada frequência ou situação, um número de vezes mais elevado;
3. Deverá ser estudado um sistema de em que o período de evolução se vá deslocando com o decorrer do tempo, de modo a poder gerar soluções para o momento imediato.
4. O GA a utilizar deverá contar com uma população inicial maior que garanta uma maior diversificação.
5. Um factor não incluído neste estudo prendeu-se com a volatilidade das cotações ao longo dos períodos. Tal consideração deverá ser incluída num estudo futuro.

*“The real nature of things, we shall never know,  
never.”*

(Albert Einstein)

# Bibliografia

- [1] F. Allen and R. Karjalainen. Using genetic algorithms to find technical trading rules. *Journal of Financial Economics*, (51):245–271, 1999.
- [2] M. Alvarez-Diaz and A. Alvarez. Forecasting exchange rates using genetic algorithms. *Applied Economics Letters*, 10(6):319–322, 2003.
- [3] RG Bates, MAH Dempster, and YS Romahi. Evolutionary reinforcement learning in FX order book and order flow analysis. *Computational Intelligence for Financial Engineering, 2003. Proceedings. 2003 IEEE International Conference on*, pages 355–362, 2003.
- [4] A. Brabazon and M. O’Neill. Trading Foreign Exchange Markets Using Evolutionary Automatic Programming. *GECCO*, pages 133–136, 2002.
- [5] I. Dempsey, M. O’Neill, and A. Brabazon. Live trading with grammatical evolution. *Proceedings of the Grammatical Evolution Workshop 2004*, 2004.
- [6] MAH Dempster and CM Jones. Can Technical Pattern Trading Be Profitably Automated? 1. The Channel. *Centre for Financial Research, Judge Institute of Management Studies, University of Cambridge, working paper.(1998a):?Can technical pattern trading be profitably automated*, 2, 1998.

- [7] MAH Dempster and CM Jones. A real-time adaptive trading system using genetic programming. *Quantitative Finance*, 1(4):397–413, 2001.
- [8] C. Dunis, A. Harris, and S. Leong. Optimising Intraday Trading Models with Genetic Algorithms. *Neural Netw. World (Czech Republic)*, 9(3):193–223, 1999.
- [9] E.J. Elton and M.J. Gruber. Modern portfolio theory and investment analysis. 1984.
- [10] A. Frino, M. Peat, and B. Wong. An Examination of the profitability of technical trading rules in Australian futures markets. 2006.
- [11] D. Harding, G. Nakou, A. Nejjar, et al. The Pros and Cons of Drawdown as a Statistical Measure of Risk for Investments. *AIMA Journal*, pages 16–17, 2003.
- [12] MA Kaboudan. Genetic Programming Prediction of Stock Prices. *Computational Economics*, 16(3):207–236, 2000.
- [13] J. Kamruzzaman and R.A. Sarker. ANN-Based Forecasting of Foreign Currency Exchange Rates. *Neural Information Processing*, 3(2):49–58, 2004.
- [14] B. LeBaron. Trading Rule Profitability in the 1990s. 2002.
- [15] J.B. Li, S.C. Fu, and A.P. Chen. Applying XCS Model to Spread Trading of Taiwan Stock Index Futures.
- [16] L. Lin, L. Cao, J. Wang, and C. Zhang. The applications of genetic algorithms in stock market data mining optimization. *Proceedings of Fifth International Conference on Data Mining, Text Mining and their Business Applications*, pages 273–280, 2004.

- [17] K. Meagher, E. Carty, A. Brabazon, M. O’Neill, and P. Keenan. Intra-day Financial Asset Trading Using Grammatical Evolution.
- [18] C. Neely, P. Weller, and R. Dittmar. Is Technical Analysis in the Foreign Exchange Market Profitable? A Genetic Programming Approach. *The Journal of Financial and Quantitative Analysis*, 32(4):405–426, 1997.
- [19] CJ Neely and PA Weller. Intraday technical trading in the foreign exchange market. *Journal of International Money and Finance*, 22(2):223–237, 2003.
- [20] L. Núñez-Letamendia and M. de Molina. Fitting the control parameters of a Genetic Algorithm to optimise technical trading rules. *European Journal of Operational Research*, 179(3):847–868, 2007.
- [21] D. Ohlson. Have Trading Rule Profits in the Currency Markets Declined Over Time? 2004.
- [22] W.B. Omrane and H. Van Oppens. The Predictive Success and Profitability of Chart Patterns in the Euro/Dollar Foreign Exchange Market.
- [23] S. Papadamou and G. Stephanides. Improving technical trading systems by using a new MATLAB-based genetic algorithm procedure. *Mathematical and Computer Modelling*, 46(1-2):189–197, 2007.
- [24] O.V. Pictet, M.M. Dacorogna, B. Chopard, M. Ousaidene, R. Schirru, and M. Tomassini. Using genetic algorithms for robust optimization in financial applications. *Neural Network World*, 5(4):573–587, 1995.
- [25] S. Schulmeister. Components of the profitability of technical currency trading. 2005.

- [26] W.F. SHARPE. Mutual Fund Performance. *The Journal of Business, Supplement on Security Prices*, 39(1, part 2):119–138, 1966.
- [27] C.N.W. Tan and H. Dihadjo. Trading system for australian dollar using multiple moving averages and auto-regressive models.
- [28] J. Yao and C.L. Tan. A case study on using neural networks to perform technical forecasting of forex. *Neurocomputing*, 34(1-4):79–98, 2000.