O backtesting, o que é, como se constrói e por que razão não demonstra o que promete.

De muitas das provas que são apresentadas ao trader para o convencer de que uma estratégia funciona, o backtesting é a mais usada e a que tem mais peso, porque vem com um gráfico, uma curva de resultados que sobe da esquerda para a direita, uma percentagem de acertos, um número de operações e uma frase que fecha qualquer "dúvida": testado em 10 anos de dados, e com isso vendem-se sistemas, indicadores, robots e cursos completos.

A ideia é simples: toma-se um conjunto de regras, aplica-se sobre o preço passado de um ativo e calcula-se o que teria ocorrido se essas regras tivessem sido seguidas durante todo esse tempo. O resultado é um histórico simulado de operações e dele saem os números que depois se mostram. E quero desenvolver este tema porque a conclusão que se tira desses números é exatamente a contrária daquilo que os números permitem tirar: um backtesting não demonstra que uma estratégia funcione, demonstra que umas regras encaixaram num tramo concreto do passado, e entre essas duas coisas há uma distância que a curva não mostra, sem contar que a dinâmica de mercado maioritária que comanda o preço se baseia na especulação atual e não em estruturas do passado.

Para entender porquê, há que ver como se constrói: o processo não é tomar umas regras e testá-las uma vez, é um ciclo; testam-se, olha-se para o resultado, vê-se onde perdeu, adiciona-se um filtro para evitar essas perdas, volta-se a testar, ajusta-se outro parâmetro, repete-se, e a cada volta a curva melhora, porque cada ajuste foi feito olhando para o resultado anterior, ou seja, as regras terminam escolhidas depois de ter visto os dados e não antes, e isso inverte por completo o valor do teste; quando uma regra é decidida antes de ver o resultado e depois é comprovada, a comprovação diz algo; quando a regra é moldada sobre o resultado até encaixar, a comprovação não diz nada, somente confirma que quem a construiu foi o bastante paciente para a fazer coincidir. E a questão de fundo é que qualquer série de preços se pode explicar para trás se se lhe impuserem condições suficientes: com os parâmetros necessários pode-se construir um sistema que ganhe em qualquer período passado que se lhe entregue, e essa facilidade é precisamente o sinal de que o exercício não está a medir o que diz medir, e ainda para mais com a tecnologia atual é mesmo fácil fazer encaixar bons resultados num gráfico passado.

Portanto, desse processo sai a armadilha central: cada filtro que se adiciona para eliminar perdas passadas torna o sistema mais específico e menos geral; um sistema com três regras descreve um comportamento amplo do mercado, um sistema com quinze regras descreve os detalhes de um tramo concreto, e esses detalhes não se repetem. O resultado é que a curva mais bonita corresponde ao sistema mais frágil, porque o que melhor se ajustou à história é o que mais depende de que essa história se repita exata, e por isso o padrão que se vê uma e outra vez é o mesmo: resultados históricos impecáveis, e no momento em que o sistema é posto a funcionar sobre preços que ainda não ocorreram, a curva quebra-se de maneira surpreendente em

qualquer momento, não porque o sistema esteja mal calculado, mas sim porque foi calculado até ao ponto em que já somente descrevia um passado específico, e o mercado que vem não é idêntico àquele que se usou para o construir; em poucas palavras, passou de algo que tenta prever um mercado especulativo para prever um gráfico já sucedido.

Há uma segunda parte do engano que não está no sistema, mas sim naquilo que não se mostra: quem constrói uma estratégia não testa uma, testa dezenas ou centenas de variantes, distintas combinações de regras, distintos parâmetros, distintos períodos, e de todas essas variantes fica com a que se saiu melhor e essa é a que mostra; as outras desaparecem. E este ponto tem consequências maiores do que parece, porque se se testarem 500 conjuntos de regras escolhidos ao acaso sobre os mesmos dados, por pura probabilidade vários deles vão mostrar resultados excelentes nesse passado sem ter nenhuma lógica por trás; a casualidade produz curvas vencedoras quando se testa o suficiente, logo a curva que se apresenta como evidência é na realidade o resultado de uma seleção: escolheu-se a que funcionou entre muitas que não, e uma seleção não é um teste, é o contrário, é o ocultamento de todos os testes que falharam, de modo que possivelmente no teste se descartaram variantes que podiam funcionar melhor no mercado especulativo em tempo real do que num gráfico que já sucedeu.

A tudo isso soma-se que os dados sobre os quais se faz o teste não são aqueles que o trader teria enfrentado: o histórico que se descarrega é uma série limpa de preços, e a operativa real não ocorre sobre uma série limpa, ocorre com lacunas na cotação, com execuções que não chegam ao preço exato, com velas que no histórico aparecem completas mas que em tempo real se formaram de outra maneira, e com informação que o sistema usa como se estivesse disponível no momento quando na realidade se conheceu depois: por exemplo, quando se toma o fecho de uma vela para decidir uma entrada que na prática teria tido de ser feita antes de essa vela fechar; esse último erro é dos mais comuns e dos menos visíveis, porque o sistema na simulação vê o futuro por uns segundos ou uns minutos, e com isso basta para que a curva melhore de forma artificial; além disso está a questão do ativo: a maioria destes históricos corresponde a derivados e a cotações de fornecedores, ou seja, a referências do preço e não ao preço formado no ativo real; além do mais, a execução não tem nada a ver com a que sucedeu em tempo real, e nem falar de toda a pressão que sente qualquer trader, seja especialista ou novato, a operar em tempo real, enquanto o backtesting não diferencia, no que toca ao aspeto qualitativo do trader, um cenário de muito nervosismo ou um cenário flexível normal.

E chego ao ponto que mais importa dentro daquilo que defendo sobre os mercados: um backtesting assume que as causas que produziram o comportamento passado continuarão presentes, e essa é a suposição que não se cumpre e que faz com que se estrague; o preço não se move por padrões que se repetem sozinhos, move-se pelo que o sistema financeiro executa sobre o ativo real em cada momento, e essa execução responde às condições desse momento, a quem está preso, a que contraparte há disponível, a que objetivos ficam pendentes — condições que mudam de ano para ano e de mês para mês (o mercado do ouro de 2019 e o de 2026 partilham o nome do ativo e pouco mais) —, a volatilidade, a dimensão dos movimentos, a maneira

como se limpa e se defende, tudo isso é distinto; portanto, uma regra que ganhou em 2019 não ganhou por ser correta, ganhou porque encaixou na dinâmica de 2019, e essa dinâmica já não existe.

E há algo mais que nenhum backtesting pode incorporar: quando uma regra se torna popular deixa de funcionar pela sua própria popularidade, porque uma quantidade previsível de capital a fazer o mesmo no mesmo sítio é exatamente o que a dinâmica de mercado utiliza como combustível e vai contra toda essa multidão; o sistema testado sobre o passado não pode medir o efeito que terá a sua própria difusão, e esse efeito é o que termina a anulá-lo.

Com tudo o anterior resta a pergunta de se o backtesting serve para alguma coisa, e sim, serve, mas para o contrário do que se lhe pede: serve para descartar; se um conjunto de regras perde de forma consistente sobre o passado, isso sim é informação, porque o que não funcionou em nenhum período dificilmente vai funcionar agora; serve também para conhecer o caráter de umas regras — quanto chegam a perder no seu pior tramo, quão longas são as suas sequências negativas, como se distribuem os seus resultados. Tudo isso descreve o comportamento do sistema e é útil como descrição; o que não pode fazer é confirmar, porque confirmar exige que o futuro se pareça com o passado, e essa é justamente a parte que ninguém pode garantir.

Na minha leitura, o backtesting ocupa o mesmo lugar que qualquer ferramenta numérica: mede o que já ocorreu e dimensiona, e quem o trata como um teste de rentabilidade futura está a pedir-lhe a única coisa que, por construção, não pode dar.

Resta dizer por que razão se vende tanto, e a resposta está naquilo que a curva produz em quem olha para ela: é visual, é limpa, parece objetiva porque tem números, e não exige que quem a mostra demonstre nada em tempo real; um vendedor com resultados ao vivo tem de os sustentar todas as semanas, um vendedor com uma curva histórica mostra-a uma única vez e fica fechada, e quando o comprador perde aplicando o sistema, a explicação já está preparada: não seguiu as regras, faltou-lhe disciplina, não respeitou a gestão, e todas essas tontices; ou seja, a curva traslada a responsabilidade do sistema para a pessoa, e essa é a razão pela qual se prefere em relação ao resultado real, porque o resultado real pode ser discutido e a curva não; portanto, quando alguém apresentar uma estratégia respaldada por 10 anos de dados, o que está a apresentar é uma regra moldada sobre esses 10 anos, selecionada entre variantes que falharam, testada sobre uma série que não inclui as condições reais e baseada em que as causas de então continuem a operar agora, e nenhuma dessas quatro coisas se vê na curva.

E para fechar deixo a minha posição própria sobre isto, porque o backtesting no trading profissional de alto nível tem sim um uso real e de valor, e eu próprio o usei em muitas estratégias ao longo do tempo, e com a variação quantitativa, com uma condição que muda tudo: que se faça com seriedade e com realismo dos factos, sem se deixar levar por vieses qualitativos nem por aquilo que se quer que saia, mas sim realmente sobre os factos, buscando precisamente que o resultado se sustente no presente sem conhecer o futuro, ou seja, testando as regras como se o gráfico passado não existisse, sem o ter visto, sem ajustar nada depois de

olhar para o que ocorreu — um backtesting quantitativo real, onde o teste é feito nas mesmas condições de ignorância nas quais se opera todos os dias, e onde o resultado vale porque não foi moldado.

Isso é o que eu ensino, isso é o que aplico e isso é o que emprego no dia a dia, e por isso a crítica deste documento não vai contra a ferramenta, vai contra o uso que se lhe dá para vender, porque bem feita é uma das poucas formas honestas de saber se uma leitura tem fundamento antes de expor capital a ela, e mal feita é somente uma curva bonita com um preço.

Reagam bastante para continuar com estes PDFs, e até à próxima.