ARCTAN RESEARCH

Como tentamos não enganar a nós mesmos — nem a você

A disciplina por trás da nossa pesquisa: um guia em linguagem simples

Tony Volpon · Arctan Research · arctanresearch.net

> *"O primeiro princípio é que você não deve enganar a si mesmo — e você é a pessoa mais fácil de enganar."*
>
> — Richard Feynman

O problema: com tentativas e tempo suficientes, sorte parece competência

Imagine dez mil pessoas, cada uma jogando uma moeda dez vezes. Por puro acaso, cerca de dez delas tirarão dez caras seguidas. Se essas dez escreverem newsletters sobre sua "técnica" de jogar moedas, parecerão gênios.

A pesquisa quantitativa de investimentos enfrenta exatamente essa armadilha, em escala industrial. Os mercados nos dão uma única história — uma única sequência de preços — e os computadores modernos permitem testar milhões de ideias contra ela. Teste ideias suficientes e algumas se ajustarão lindamente ao passado por acidente. Os acadêmicos chamam o resultado de "zoológico de fatores": centenas de padrões publicados que "previam" retornos nos dados antigos, muitos dos quais silenciosamente pararam de funcionar assim que dinheiro de verdade os seguiu. A velha piada diz que, se você torturar os dados por tempo suficiente, eles confessam alguma coisa. A confissão não vale nada.

Tudo o que fazemos na Arctan está organizado em torno de um objetivo: **tornar o mais difícil possível que enganemos a nós mesmos**. Esta nota explica, em linguagem simples, as regras específicas que seguimos. Nenhuma delas é invenção nossa — vêm das melhores práticas da ciência experimental e da pesquisa moderna em finanças. O que é incomum é aplicar todas elas, o tempo todo e, **em alguns casos, publicamente**.

Regra 1 — Usar apenas o que era conhecível na época

Suponha que uma empresa divulgou seu resultado anual em fevereiro, mas três meses depois corrigiu um erro contábil e republicou os números. Um teste histórico descuidado usa os números corrigidos como se os investidores os tivessem em fevereiro. Uma estratégia construída assim está colando na prova: ela "sabia" em fevereiro o que o mundo só aprendeu em maio.

O remédio se chama **dado point-in-time**. Cada informação em nossa pesquisa carrega um carimbo de quando se tornou conhecível — a data em que o balanço foi de fato publicado, não o período que ele descreve — e cada decisão simulada só pode usar o que tinha carimbo anterior. Parece óbvio. É também, na prática, **a forma mais comum de inflação silenciosa de resultados históricos** — e por isso tratamos a procedência dos dados com a mesma seriedade das próprias estratégias. Quando recentemente reconstruímos um estudo em uma nova base de dados, **a primeira semana foi gasta não em resultados, mas em verificar carimbos de tempo, cruzar fontes independentes e confirmar que empresas que faliram décadas atrás continuam nos dados, com seus verdadeiros e dolorosos retornos finais**. Bases de dados têm o costume de esquecer discretamente os perdedores; os mercados não esqueceram, e nós também não podemos.

Regra 2 — Escrever o teste antes de rodá-lo

A maneira mais sutil de se enganar é decidir o que conta como sucesso **depois** de ver os resultados. Mova um limiar aqui, exclua um período "atípico" ali, e quase qualquer ideia pode ficar com boa aparência — cada pequena escolha é defensável; a soma delas é fatal.

Nossa resposta é o **pré-registro**, emprestado da medicina clínica, onde se tornou obrigatório depois que muitos testes de medicamentos mudaram discretamente sua definição de sucesso no meio do caminho. Antes de testar uma hipótese, **escrevemos um memorando que fixa tudo: os dados exatos, o procedimento exato, os limiares exatos que um resultado precisa superar — nós os chamamos de gates** — e, tão importante quanto, o que esperamos que aconteça e o que concluiremos em cada cenário, inclusive no fracasso. Depois, o memorando é congelado. Qualquer mudança posterior precisa ser uma emenda datada e por escrito, feita antes de a etapa afetada rodar — nunca um ajuste silencioso depois de espiar.

Para nossos estudos mais importantes, damos agora um passo além e **registramos o plano de análise no Open Science Framework**, um registro científico público, antes da execução. O registro é carimbado no tempo e não pode ser editado por ninguém — inclusive nós. Quando eventualmente afirmarmos "previmos isso de antemão", qualquer leitor pode clicar em um link e verificar contra um relógio independente. Até onde sabemos, estamos entre um número muito pequeno de gestoras que fazem isso.

Regra 3 — Uma única olhada no gabarito

Qualquer modelo flexível pode ser calibrado até se ajustar ao passado. A única pergunta honesta é como ele se comporta em dados que nunca viu. **Por isso dividimos a história: um período antigo de treino, onde o modelo é construído e calibrado, e um período recente reservado, trancado e intocado até o desenho estar congelado**.

Então vem a parte em que somos mais rígidos: os dados reservados podem ser examinados uma única vez. Nossos sistemas literalmente impõem isso — o script de avaliação exige uma declaração explícita para rodar e se recusa a rodar duas vezes. *Se um teste falha nos dados reservados, a resposta é não*. Não ajustamos o modelo para tentar de novo, porque cem novas tentativas contra os mesmos dados são apenas a história das moedas com etapas extras. Uma pergunta, uma resposta — e a resposta vale.

Regra 4 — Perguntar o que o puro ruído teria encontrado

O ferramental estatístico moderno é poderoso o bastante para encontrar padrões em qualquer coisa — inclusive no acaso. Por isso, antes de confiar no que nossas ferramentas encontram nos dados reais, **rodamos as mesmas ferramentas em dados sintéticos: centenas de histórias artificiais, construídas com a mesma textura estatística dos dados verdadeiros, mas contendo, por construção, nenhum sinal genuíno.**

Esse **teste placebo** nos diz o que nossas ferramentas "descobririam" em um mundo onde não há nada a descobrir. Só merecem atenção resultados que superem claramente esse padrão de referência. Em um estudo recente, o teste cumpriu papel duplo: confirmou que um padrão nos dados reais era genuíno — mais forte do que qualquer coisa em trezentos mundos de ruído — e, ao mesmo tempo, revelou que os detalhes finos do modelo ajustado eram indistinguíveis do que o ruído produz. As duas metades importaram. Saber precisamente em quais partes de um resultado confiar, e quais descartar, é a diferença entre pesquisa e contar histórias.

Regra 5 — Desconfiar dos próprios dados

Dados ruins produzem bobagens confiantes. Por isso, antes de qualquer hipótese ser testada, os dados precisam passar pelo próprio exame — **o que chamamos de gate de integridade**. Cruzamos cada série com fontes independentes. Procuramos valores impossíveis, séries congeladas e lacunas. E tiramos a "impressão digital" de cada série contra a história conhecida: uma série de momentum que não desaba na primavera de 2009, ou uma série de volatilidade que dorme na explosão de volatilidade de março de 2020, está quebrada, diga o que disser sua documentação. É assim, também, que erros de sinal — uma série acidentalmente invertida — são pegos antes de causar dano.

Essas checagens parecem prosaicas. Elas encontram problemas reais com regularidade: falências ausentes, históricos revisados em silêncio, convenções trocadas. Cada problema encontrado cedo é uma conclusão falsa que nunca chegou a existir.

Regra 6 — Quando a resposta é não, a resposta é não

A maioria dos nossos testes, cuidadosamente desenhados, falha. Isso não é uma confissão; é o sistema funcionando. Os mercados são ferozmente competitivos, e a maioria das ideias plausíveis já está refletida nos preços. Um processo de pesquisa que diz sim na maior parte do tempo não está descobrindo um mundo excepcional — está se enganando.

Por isso tratamos um "não" bem estabelecido como resultado de primeira classe. Testes que falham são documentados com o mesmo cuidado dos que passam, arquivados com o relato completo do porquê, e encerrados — **o que protege capital real de ideias que só funcionaram no papel**, e nos impede de re-testar em silêncio, a cada dois anos, a mesma ideia sedutora. Quando um teste falha por uma razão técnica identificável, corrigimos a falha e dizemos isso; quando falha de forma limpa, atualizamos nossas crenças — e dizemos isso também. As expectativas declaradas da Regra 2 dão dentes a essa disciplina: pontuamos nossas próprias previsões contra os resultados, por escrito, incluindo as que erramos.

O que isso lhe garante — e o que não pode garantir

Nada disso garante retornos. Nada garante, e você deve desconfiar de quem insinue o contrário. O que essa disciplina compra é algo mais estreito e, acreditamos, mais valioso: **quando um resultado sobrevive a essa prova de fogo — dados carimbados no tempo, um plano congelado e registrado publicamente, uma única olhada na história não vista, um padrão placebo, insumos auditados e a disposição permanente de aceitar não como resposta — ele conquistou o direito de ser levado a sério**. E quando dissermos que algo não funcionou, você pode confiar nisso também.

A pessoa mais fácil de enganar é você mesmo. Tentamos construir um processo de pesquisa em que essa pessoa é vigiada, o tempo todo, pelo auditor mais desconfiado que conseguimos construir: nossos próprios compromissos escritos, congelados e públicos. Preferimos reportar menos achados que sejam reais a muitos que sejam histórias. No longo prazo, acreditamos que essa é também a melhor estratégia de investimento.

—

Esta nota descreve metodologia de pesquisa. Não é recomendação de investimento e não descreve nenhuma estratégia, posição ou desempenho específicos. © Arctan Research · arctanresearch.net