

PREVISÃO DE VENDAS POR TÉCNICAS DE CORRELAÇÃO

ROBERT FERBER

"A análise de correlação é um dos instrumentos de medição mais poderosos de que podem dispor os pesquisadores em ciências aplicadas; ademais, ocupa uma posição primordial no conjunto das técnicas analíticas empregadas pelo pesquisador em Economia Aplicada à Administração." — M. H. SPENCER e L. SIEGELMAN

O emprego de métodos de correlação na previsão das vendas de uma empresa ou de um ramo de atividade econômica vem obtendo crescente aceitação nos últimos anos. Resumidamente, o método implica no ajustamento de uma equação para explicar as flutuações nas vendas em termos de variáveis relacionadas e presumivelmente causais; substituem-se essas variáveis pelos seus valores de provável ocorrência no período a ser previsto e resolve-se a equação para obter o valor das vendas. O método, embora apresentando limitações amplamente reconhecidas,¹ pode

ROBERT FERBER — Professor de Economia da Universidade de Illinois. O autor deseja expressar sua gratidão a C. Ashley Wright e V. Lewis Bassie pela leitura dos originais deste artigo e por muitas sugestões úteis.

Nota da Redação: Este artigo é reproduzido com autorização da revista *Journal of Marketing*, publicação quadrimestral de circulação nacional da *American Marketing Association*, onde foi publicado em janeiro de 1954, volume XVIII, n.º 2, sob o título original "*Sales Forecasting by Correlation Techniques*". Traduzido do inglês por Frediano Quilici.

- 1) Em especial: (a) a suposta estabilidade das relações históricas futuras, (b) o perigo de que a causalidade possa dirigir-se tanto num como noutro sentido, e (c) a dificuldade de prever os futuros valores das variáveis a serem usadas na previsão de vendas, especialmente quando essas variáveis não são defasadas.

ser de considerável valia como auxiliar na previsão². Deve-se, também, levar em conta sua utilidade no contrôlo da pertinência dos vários fatores determinantes das vendas, bem como na avaliação da importância relativa de cada um desses fatores.

A técnica de determinar a equação relacionando-se as vendas com as variáveis explicativas — a equação de regressão — é tratada com pormenores em várias fontes³ e é fácil aprender. Contudo, não é finalidade deste artigo tratar dessas minúcias técnicas, mas oferecer sugestões e tecer considerações quanto à especificação e ao emprêgo dessas equações na previsão de vendas. Percebe-se que o valor de uma equação não pode ser maior do que o das hipóteses que lhe servem de base — escolha das variáveis explicativas, das unidades de medida, do período de observação etc. — embora sua importância tenda algumas vezes a ser diminuída ou tratada de maneira rotineira, na ânsia de obter coeficientes matematicamente exatos, com seis, sete ou oito algarismos significativos.

A finalidade deste artigo é dar mais ênfase à importância de algumas dessas considerações básicas pelo oferecimento de sugestões — algumas novas, outras não — quanto à sua solução. Essas sugestões baseiam-se, em grande parte, nos estudos, realizados pelo autor, das relações entre agregados⁴ que, embora não diretamente relacionados com a previsão de vendas, apresentaram dificuldades semelhantes.

Quase toda esta apresentação está relacionada com a especificação da equação de regressão ou previsão. Entretanto, alguma atenção é também dada no sentido de ava-

2) Como é apresentado por exemplo, por C. M. CRAWFORD, em *An Analysis of Sales Forecasting Policies, Procedures and Techniques*, tese de doutoramento não-publicada, Universidade de Illinois.

3) Por exemplo; M. EZEKIEL, *Methods of Correlation Analysis*; R. FERBER, *Statistical Methods in Market Research*; A. E. WAUGH, *Elements of Statistical Methods*.

4) Principalmente "A Study of Aggregate Consumption Functions", *National Bureau of Economic Research, Technical Paper 8*, 1953; "The Railroad Shippers' Forecasts", monografia de uma série de estudos sobre as expectativas dos homens-de-empresa, *Bureau of Economic and Business Research*, Universidade de Illinois, 1953.

liar a adequação das previsões determinadas pela equação de regressão, bem como sobre a possibilidade de melhorar a exatidão de previsões futuras, com base nas mesmas equações. Deve-se assinalar que nenhuma tentativa está sendo feita para dar cobertura completa às questões levantadas, ou mesmo para apresentar uma lista geral dos casos relacionados com o emprêgo de técnicas de correlação na previsão de vendas.

ELABORAÇÃO E EMPRÊGO DA EQUAÇÃO DE PREVISÃO

Supõe-se a seguinte situação inicial: tentativa vai ser feita para prever as vendas de uma empresa ou de um ramo de atividade econômica — inicialmente não importando qual — pela determinação de uma equação de regressão entre a variável dependente (as vendas) e uma ou mais presumíveis variáveis independentes que, segundo se acredita, ocasionam a flutuação das vendas. Embora não seja esse, certamente, o esboço preciso de uma situação, assim o traçamos deliberadamente, uma vez que muitas das questões relativas à especificação da equação e ao método de determinar as previsões serão discutidos a seguir. Em especial, questões relativas aos seguintes problemas serão apreciadas: o método de ajustamento, o método de determinação da previsão, a escolha do período de observação, as unidades de medida para as variáveis e a escolha das variáveis a serem incluídas na equação. A seqüência na apresentação desses tópicos não é a que na prática se verifica: ela é ditada, principalmente, pela provável ordem de influência de uns sobre outros.

A) *Método de Ajustamento* — Armada uma equação de previsão — digamos que vendas (V) seja uma função aritmética linear da renda nacional disponível (R) e das despesas de propaganda (P)⁵, isto é, $V = a + bR + cP$ —, apresenta-se o problema de estimar os valores dos parâmetros da equação, a , b e c . Isso é quase sempre feito pelo chamado método dos mínimos quadrados, que se baseia na determinação da equação que reduzirá ao míni-

5) Supondo-se que tais despesas sejam flexíveis.

mo a soma dos quadrados dos desvios das observações em relação à linha ajustada para aquele tipo de função. Implica, pois, na construção e na solução de um conjunto de equações "normais", pelo método-padrão apresentado nos livros de estatística para esse tipo de trabalho.

Últimamente esse método de ajustamento tem sido criticado por produzir estimativas viciadas dos parâmetros. Funda-se essa crítica na suposição implícita no método dos mínimos quadrados, de que as variáveis independentes são realmente independentes das variáveis dependentes. Quando a variável dependente e uma ou mais das variáveis independentes se referem à mesma unidade de tempo essa suposição pode ser refutada. Assim, se as vendas de aço forem relacionadas à renda nacional, pode-se argumentar que tanto essas vendas constituem *determinante* da renda nacional, como a renda nacional será *determinante* das vendas de aço. Neste caso pode-se demonstrar que a verdadeira regressão das vendas sobre a renda e outras variáveis é determinada, não pelo método dos mínimos quadrados, mas por meio de um sistema de equações contendo tantas equações quantas as variáveis inter-relacionadas ⁶.

Exemplo simples de um sistema de equações para esse caso seria o seguinte:

Equação 1:

$$\text{Vendas de aço} = a + b (\text{renda nacional}).$$

Equação 2:

$$\text{Renda nacional} = c + d (\text{despesas governamentais}) + e (\text{investimento privado}),$$

onde se supõe que as despesas governamentais e o investimento privado são determinados independentemente, isto é, conhecidos de antemão.

6) O raciocínio estatístico em que se baseia essa afirmação é, talvez, mais claramente explicado por E. G. BENNION, em "The Cowles Commission 'Simultaneous Equation Approach': A Simplified Explanation", *The Review of Economics and Statistics*, vol. 34, n.º 1 (1952), págs. 49 a 57.

Nesse caso temos duas equações em duas variáveis inter-relacionadas, vendas e renda nacional, sendo, portanto, possível obter estimativas justas dos parâmetros a e b ⁷.

A solução desses sistemas de equações pode tornar-se bastante complicada e, talvez por essa razão, não tenha merecido a preferência dos que se ocupam com previsão de vendas. Em razão dos acontecimentos essa relutância em determinar relações pelo emprêgo de sistemas de equações parece justificar-se.

Duas razões podem ser apontadas para explicar esse fato: (a) os sistemas de equações que têm sido empregados na previsão de tendências econômicas em base nacional não têm produzido bons resultados⁸, e (b) nos casos das relações entre as vendas de uma empresa ou de um ramo de atividade econômica com os indicadores nacionais, como renda, produção e população, freqüentemente utilizados na previsão de vendas, a direção da causalidade dificilmente pode ser imaginada como igualmente forte em ambas as direções. Mesmo no caso de uma indústria básica, como a do aço, pode-se perfeitamente dizer que a renda nacional afeta a produção do aço muito mais do que a produção do aço afeta a renda nacional.

Por essas razões justifica-se o uso continuado do método da equação única com mínimos quadrados para a determinação das equações de previsão de vendas. Entretanto, os que se dedicam à previsão de vendas estariam agindo inteligentemente se se mantivessem a par dos desenvolvimentos havidos no emprêgo dos sistemas de equações.

B) Determinação da Previsão — Discutir esse assunto pode, à primeira vista, parecer um tanto estranho. Com efeito, tendo-se a equação e os valores das variáveis independentes, que mais pode haver na determinação da previsão, além das necessárias substituições e da resolução,

7) O método de solução é tratado por G. TINTNER, *Econometrics*, capítulo 7.

8) Talvez o teste mais desenvolvido dessa espécie seja o de C. CHRIST, "A Test of an Econometric Model for the United States, 1921-1947", acompanhado de uma discussão em *Conference on Business Cycles, National Bureau of Economic Research*, Nova Iorque: 1951, págs. 35 a 131.

para encontrar o valor da variável (vendas)? É que a função pode ser utilizada, não para estimar o nível das poupanças, como se faz habitualmente, mas, sim, para estimar a *modificação* nas poupanças. Isso é realizado da seguinte maneira.

Suponhamos que a função de previsão seja a seguinte:

$$V_t^e = a + bR_t + cP_t$$

onde o subscrito t refere-se ao tempo, indicando que tôdas as variáveis se referem ao mesmo intervalo de tempo, e o superscrito e indica o valor que está sendo estimado.

Para prever a modificação nas vendas calcula-se $V_t - V_{t-1}$. As vendas do último período V_{t-1} podem ser expressas da mesma maneira que as vendas atuais, mas com um subscrito de tempo defasado. Assim:

$$V_{t-1}^e = a + bR_{t-1} + cP_{t-1}$$

Portanto,

$$(V_t - V_{t-1})^e = a + bR_t + cP_t - (a + bR_{t-1} + cP_{t-1})$$

de modo que a modificação (estimada) nas vendas é igual a: $b(R_t - R_{t-1}) + c(P_t - P_{t-1})$.

A previsão da modificação nas vendas é feita substituindo-se os valores de R e P na última equação e resolvendo-se essa equação para encontrar o valor de $V_t - V_{t-1}$. A previsão do nível de vendas no tempo t é, portanto, o resultado da soma das vendas do período anterior, à estimativa das modificações nas vendas do tempo $t-1$ ao tempo t . O resultado não é o mesmo que se obtém pela previsão do nível de vendas diretamente da equação original.

Quando um método de determinação da previsão é superior ao outro? É difícil responder, dependendo cada caso da comparação de ambos durante certo período. Pode-se dizer, entretanto, que, se as previsões pelo método usual apresentarem vício sistemático num sentido particular — se, digamos, os dados reais vierem sendo sistematicamente superestimados — o emprêgo das primeiras diferenças irá,

provavelmente, melhorar a exatidão das previsões, uma vez que o nível precedente de vendas se tornará automaticamente o dado de referência para a previsão. Deve-se preferir êsse método ao expediente, por vêzes adotado, de adicionar ou subtrair à equação uma constante arbitrariamente escolhida para corrigir o vício observado. Por outro lado, se as previsões vierem zigzagueando ao redor dos dados reais, ou se êstes estiverem sujeitos a grandes variações aleatórias, os cálculos, tendo por base a modificação nas vendas, irão, provavelmente, produzir resultados inferiores.

C) *O Período de Observação* — Dois impulsos antagônicos irão provavelmente apresentar-se na escolha do período cujos dados devam servir de base para o ajustamento da equação de previsão. Por um lado, existe o impulso de utilizar todos os dados disponíveis, com isso elevando ao máximo a fidedignidade dos coeficientes, pelo menos sob aspecto puramente estatístico. Por outro lado, é desejável deixar de lado alguns anos ou observações, de modo que a exatidão da previsão da função possa ser antecipadamente testada. Tendo em vista êsse conflito, qual a solução?

Podemos começar a responder a essa pergunta citando qual não deve ser a solução: isto é, glosar, arbitrariamente, dois ou mais anos em um dos extremos dos dados, ajustar a equação para as outras observações e testar as previsões sobre os anos remanescentes. Essa não é a solução, porque a escolha de um período de observação é consideração de ordem essencialmente lógica. Em outras palavras: o período de observação deve incluir todos os anos ou observações para os quais a natureza das relações dadas não aparente ter sido alterada, feita essa verificação mediante estudo cuidadoso de tôdas as informações disponíveis.

Acreditando-se que nenhum acontecimento tenha ocorrido no período em estudo para alterar a função de previsão de vendas, todo êle deve ser incluído no período de observação, porque não há, então, base lógica nenhuma para ex-

cluír quaisquer observações particulares. A determinação antecipada da precisão da previsão da função ajustada torna-se, então, matéria de julgamento e estimativas aproximadas das variáveis para os anos que estejam fora do período de observação.

Se, entretanto, se acredita que os parâmetros da função de previsão de vendas tenham em alguma ocasião sido alterados durante êsse período — digamos, pela repentina descoberta de uma nova utilidade para um produto básico, julgada essa utilidade menos dependente da renda nacional do que outras — o tratamento racional é a divisão dêsse período em dois subperíodos, usando o momento em que se acredite que a modificação tenha ocorrido como ponto de separação. Um subperíodo será, então, o período de observação, e os dados para o outro subperíodo serão utilizados como base para a verificação da precisão da previsão da função ajustada.

Se a precisão das previsões não diferir apreciavelmente da apresentada pelas estimativas da função no período de observação, existirá razão para combinar ambos os subperíodos num único período de observação e recalcular a função de previsão de vendas, utilizando-se todos os dados disponíveis. Se existirem grandes diferenças entre as estimativas para os dois subperíodos, a utilização de todos os dados disponíveis para o ajustamento da equação não é justificada e, dependendo das circunstâncias, a exatidão das previsões para períodos futuros pode ser duvidosa.

Como tudo isso poderá parecer um tanto complicado, vamos ilustrar as várias possibilidades com alguns exemplos. A situação básica, que vamos supor, é a de que as vendas do ramo de atividade econômica Z sejam conhecidas em base anual para 1923-41 e 1947-53, sendo que a produção, de 1942 a 1946, se destinou unicamente a fins militares. Deseja-se uma função de previsão de vendas como auxiliar na previsão das vendas dêsse ramo de atividade para 1954 em diante. A forma da função é a seguinte:

Vendas = $a + b$ (renda disponível) + c (despesas de propaganda)

A maneira de como o período de observação é escolhido para as estimativas dos valores a , b e c sob as diferentes circunstâncias é descrita abaixo.

| <i>Situação</i> | <i>Período de observação</i> |
|--|---|
| 1) Não há razão para acreditar que tenha ocorrido qualquer mudança nas relações. | 1923-21, 1947-53 |
| 2) A relação poderá ter-se alterado nos anos de após-guerra. | <p>a) Ajuste a função com os dados de 1923-41; teste a exatidão das previsões sobre 1947-53.</p> <p>i) Se os testes não indicarem apreciável perda de exatidão (baseada em estatística e em outras considerações), recalcule as estimativas para a, b e c utilizando tôdas as informações.</p> <p>ii) Se tiverem sido insatisfatórias as previsões para 1947-53, calculadas com base nos dados de 1923-41, e as relações dos anos de após-guerra forem claramente insatáveis, despreze a função.</p> <p style="text-align: center;">Ou,</p> <p>b) Se parecerem estáveis as relações dos anos de após-guerra, ajuste a função com os dados de 1947-53, e teste a exatidão sobre os dados de 1923-41.</p> <p>i) Se não resultar nenhuma diferença apreciável de exatidão, proceda como em (a) (i) acima.</p> <p>ii) Se forem insatisfatórias as previsões para 1923-41, feitas com base nos dados de 1947-53, fique com a função ajustada com os dados de 1947-53.</p> |

Obs. A segunda série de alternativas não é muito desejável, pois constitui trabalho muito arriscado julgar a estabilidade de uma relação com base em sete observações.

Os seguintes comentários podem ser apresentados quanto à escolha do período de observação:

1) Estabelecida a adequação de uma função de previsão, a máxima fidedignidade é obtida recalculando-se os coeficientes cada vez que nova observação tornar-se disponível, desde que não seja evidente u'a modificação na natureza das relações. Assim, no exemplo dado, feita a previsão para 1954 e tornando-se disponíveis informações referentes a 1954, as estimativas dos parâmetros da função para a previsão de 1955 seriam recalculadas, baseadas no maior período de observação, 1923-41 e 1947-54.

2) Dadas as mesmas condições, o emprêgo de um período de observação maior dará, provàvelmente, como resultado, previsões mais precisas do que um período menor. Foi essa, pelo menos, a experiência com as funções agregadas de consumo⁹.

3) A prática, por vêzes advogada, de escolher como período de observação sòmente os anos que se acredite sejam mais parecidos com o ano a ser previsto — como seja, utilizando os dados de 1923-30, 1935-40, para prever as condições para 1948, por imaginar restritas as possibilidades de 1948 tornar-se um ano de depressão — é muito perigosa. Essencialmente, êsse procedimento é uma petição de princípio, porque omitir os anos chamados atípicos é supor, implicitamente, que suas condições não mais hajam de prevalecer no futuro¹⁰. Além disso, restringe-se a aplicabilidade das funções derivadas dessa maneira no que toca à avaliação da importância relativa dos vários fatores sòbre as vendas. Na medida em que considerações de ciclos entrarem na escolha de um período de observação dever-se-á tentar incluir a experiência de pelo menos um ciclo completo, com diferenças substanciais entre os picos e os pontos mínimos.

D) *Unidades de Medida* — Fundamentados em testes levados a efeito sòbre uma série de funções agregadas de

9) R. FERBER, *S Study in Aggregate Consumption Functions*, pág. 45.

10) Essa crítica perde muito de sua fôrça quando a função de previsão não contém variáveis defasadas, porque é possível que, então, o estágio de flutuação dos negócios no período previsto seja implicitamente determinado pelos valores atribuídos às variáveis independentes.

consumo ¹¹, podemos dizer que as previsões de vendas baseadas em funções de regressão são, provavelmente, mais precisas quando as variáveis são ajustadas tendo-se em vista modificações nos preços e quando expressas em unidades *per capita*. Convém não esquecer, entretanto, a possibilidade de enganos resultantes de previsões erradas de preços e de movimentos populacionais quando a previsão é desejada em termos agregados e aos preços atuais.

Os testes sôbre a exatidão da previsão, aplicados a algumas funções de consumo, cujas variáveis foram alternadamente ajustadas para modificações nos preços, para modificações populacionais, para ambas e para nenhuma, indicaram, geralmente, melhor exatidão quando feitos os ajustamentos nos preços e na população. Êsses melhoramentos foram mais pronunciados nas previsões de nível (método habitual), mas foram também evidentes nas previsões de mudanças, quando empregadas as primeiras diferenças da função. Deve-se mencionar que pequena diferença na precisão da previsão foi observada tanto entre funções cujas variáveis foram expressas em unidades *per capita*, como entre funções nas quais a população foi usada como variável separada.

E) *Escolha das Variáveis* — Não pretendemos recomendar quais variáveis (independentes) devam figurar na função de previsão de vendas: constitui assunto a ser analisado em cada caso particular. Ao contrário, propomos apresentar algumas considerações que possam ser úteis na orientação da escolha das variáveis. Especificamente, existem cinco.

UM CRITÉRIO ADICIONAL DE ESCOLHA

O principal critério na escolha das variáveis independentes, na análise de regressão, é o de que sejam relacionadas à variável dependente, isto é, à variável que esteja sendo estudada. Desde que as variáveis independentes resultan-

11) R. FERBER, *op. cit.*, págs. 56 e 57.

tes se referem, geralmente, ao mesmo período de tempo que a variável dependente, a previsão desta última não pode ser obtida antes de que sejam previstos os valores das variáveis independentes. Esta é a fraqueza fundamental da técnica de correlação quando aplicada à previsão, especialmente devido à pouca atenção dada à previsão das variáveis independentes. Assim, a previsão baseada na seguinte função:

$$\text{Vendas} = a + b (\text{renda})$$

pode estar errada, não somente porque a relação no próximo ano pode divergir, ligeiramente, da que houve no passado, como, também, porque a estimativa da renda para o próximo ano, substituída na equação, é consideravelmente errada. Em muitos casos o erro na previsão de vendas pode ser devido mais ao último fator do que ao fato de a verdadeira relação afastar-se daquela que foi estimada.

Qual a resposta? É claro que se deve fazer o possível para reduzir o erro nas previsões das variáveis independentes. O emprêgo de variáveis defasadas é a solução ideal, mas, como a maioria dos ideais, êste é também raramente possível. Assim sendo, pode-se utilizar outro critério, que consiste em dar preferência às variáveis independentes que mais facilmente possam ser previstas. A finalidade dêsse critério é mais a de *complementar* do que propriamente *substituir* o critério básico citado acima. Quando é possível a escolha das variáveis independentes seu emprêgo pode contribuir enormemente para reduzir os erros de previsão. Torna-se, talvez, desnecessário acrescentar que, de qualquer modo, a previsão dos valores das variáveis independentes merece tanto cuidado e atenção quanto ao dado à própria derivação da função de previsão.

DEPENDÊNCIA DO MÉTODO DE ESTIMATIVA

Em geral, grandes diferenças de precisão nas previsões resultam da modificação das variáveis empregadas na função de regressão. Assim, a precisão da previsão: vendas =

= $a + b$ (renda) pode diferir, consideravelmente, de: vendas = $a + b$ (renda) + c (propaganda). Embora pareça estranho, não é êsse, entretanto, o caso quando as funções são transformadas em suas formas de primeiras diferenças e as previsões são feitas em termos de mudança. Uma vez que a mais importante variável tenha sido levada em consideração, o emprêgo de variáveis adicionais provavelmente irá fazer pouca diferença à precisão da previsão¹².

Êsse fato encontra explicação na maneira pela qual são elaboradas as previsões por êsse método que, como foi assinalado anteriormente, compreende a adição da mudança prevista às vendas do último ano. Mas, êste último dado é comum a tôdas as funções. Portanto, com essa referência comum, e com a principal variável determinante já levada em conta (por suposição), refinamentos subsequentes produzirão apenas ligeiras modificações na previsão final.

Quererá isso dizer que, para todos os fins práticos, não importam quais variáveis sejam empregadas, desde que a principal variável determinante seja incluída? Não; a menos que estejamos completamente convencidos da superioridade dêsse método de estimativa. Desprezar outros fatores relevantes — especialmente as possíveis modificações na composição dos agregados empregados como variáveis na equação de regressão — é ignorar a possibilidade de uma função que inclua maior número de fatores relevantes oferecer maior precisão pelo método usual de previsão (como igualmente, melhor idéia de relativa importância dos vários fatores). A vantagem de saber que o método das primeiras diferenças geralmente proporciona a mesma precisão para as várias funções é a de que, em determinada situação, será provavelmente suficiente testar a superioridade dêsse método aplicando-o a uma única função, em lugar de fazê-lo a várias delas.

12) *Accuracy of Savings Functions in the Postwar Period*. Como assinala ASHLEY WRIGHT, êsse fenômeno pode muito bem ser devido às tendências relativamente estáveis apresentados por muitas variáveis econômicas.

AJUSTAMENTO PARA OS EFEITOS CÍCLICOS

Problemas dos mais difíceis na previsão empresarial e econômica é o da obtenção, pelos métodos de regressão, de uma relação que se mantenha constante durante todo o ciclo econômico. O que geralmente se consegue é uma função bem aproximada das flutuações cíclicas na variável *vendas* durante o período de observação, mas que falha, depois, no traçamento da configuração cíclica das vendas durante o período a ser previsto. Por outras palavras, a relação tende a desfazer-se quando utilizada para fins de previsão.

Uma solução geral para esse problema parece estar ainda muito longe. Pode-se dizer, entretanto, que uma ruptura na relação parece de mais provável ocorrência quando só se empregam variáveis relacionadas com a mesma unidade de tempo, como, por exemplo, quando as vendas atuais são relacionadas à renda e às despesas atuais de propaganda. Em tal caso a inclusão de uma variável defasada que reflita as influências passadas sobre as vendas atuais pode melhorar, consideravelmente, a precisão das previsões. Afirmamo-lo como base em duas descobertas referentes às relações agregadas poupanças — renda, a saber ¹³:

1. A precisão da previsão de funções que relacionam poupanças atuais à renda atual, ou poupanças atuais à renda atual e uma tendência no tempo, foi muito melhorada quando nessas relações se incluiu uma variável de renda defasada.

2. As únicas funções cuja precisão foi tão boa nos anos de após-guerra, 1947-52, como quando ajustadas conforme os dados de 1923-40, ou com os de 1923-30 e 1935-40, foram as que incorporavam uma variável de renda defasada. De igual modo, a precisão das previsões de funções ajustadas conforme os dados do último período de observação, que exclui os principais anos de depressão, foi consideravelmente melhor. Assim, a renda defasada

13) R. FERBER, *op. cit.*, págs. 45 a 50.

neste caso parece ter efetivamente contrabalançado a influência do ciclo.

É duvidoso que a renda defasada seja tão efetiva em outras situações. O importante é obter uma variável que se possa supor reaja com uma defasagem sobre a variável que esteja prevista. No caso citado essa variável foi a renda defasada; em outra situação poderá ser outro fator.

Deve-se observar que uma vez escolhido o fator — neste caso, renda — ele poderá ser empregado de diferentes maneiras. No estudo mencionado duas formas foram empregadas, ambas quase igualmente eficientes. Uma das formas foi a renda defasada em uma unidade de tempo (ano), e a outra foi o pico cíclico passado da renda. A justificação para o emprêgo desta última é a de que, uma vez que as pessoas tenham atingido certo padrão de vida, os níveis de suas despesas e poupanças futuras, quando sua renda diminuir, tenderão a ser influenciados pelos níveis estabelecidos anteriormente, refletindo, em parte, o desejo de atingir novamente seu alto padrão de vida anterior¹⁴.

É desnecessário acrescentar que não é indesejável o emprêgo de mais de uma variável defasada. De fato, o ideal numa função de previsão seria ter tôdas as variáveis independentes defasadas, evitando-se, assim, a necessidade de estimar, por adivinhação, os valores do próximo período para essas variáveis.

EMPRÊGO DE CORRELAÇÃO SOCIAL

Dados mercadológicos e econômicos são, reconhecidamente, correlacionados em série (ou, mais simplesmente, autocorrelacionados). Ou, por outras palavras, as vendas de

14) Referências básicas quanto a êsse tratamento: J. DUESENBERY, *Income, Saving, and the Theory of Consumer Behavior*, Boston: Harvard University Press, 1949; F. MODIGLIANI, *Fluctuations in the Saving-Income Ratio: Problem in Economic Forecasting*, *Studies in Income and Wealth*, vol. 11: National Bureau of Economic Research, 1949 págs. 371 a 443.

um ano de uma empresa não são independentes das do ano anterior. As vendas do próximo ano serão um tanto aproximadas das presentes, com certeza quase dentro dos 15%. As vendas não irão dobrar em um ano e depois cair 50% no ano seguinte (exceção feita, talvez, às pequenas empresas).

Esse fenômeno da autocorrelação há muito vem sendo reconhecido e tem sido apontado como uma das mais importantes limitações dos meios atuais disponíveis para julgar da adequação das relações de regressão. Uma íntima relação entre vendas e renda pode, por vezes, ser devida tanto à autocorrelação dentro de cada série, como ao efeito causal de uma série sobre outra.

Um dos meios de contornar esse problema tem sido o emprêgo das primeiras diferenças das variáveis, em lugar das próprias variáveis¹⁵. Assim, pode-se correlacionar a mudança nas vendas com a mudança na renda e nas despesas de propaganda.

Outra possibilidade, entretanto, que parece ter merecido pouca atenção, é a de incorporar a autocorrelação diretamente na relação. Experimente utilizar as vendas do último período como uma variável independente extra, como:

$$\text{Vendas atuais} = a + b (\text{renda atual}) + c (\text{despesas atuais de propaganda}) + d (\text{vendas do último período}) :$$

Devido ao fator de autocorrelação, o emprêgo como foi feito das vendas do período passado pode servir como meio para refletir o efeito de u'a porção de influências passadas sobre o nível futuro de vendas, cada uma das quais pode ser muito pequena para ser incluída separadamente, ou, talvez, nem possa ser medida.

15) Uma exposição um tanto técnica dos problemas causados pela autocorrelação pode ser encontrada no artigo de D. COCHRANE e G. H. ORCUTT, "Application of Least Squares Regression to Relationships Containing Autocorrelated Error Terms", *Journal of the American Statistical Association*, vol. 44, 1949, págs. 32 e 61.

A investigação das propriedades auto-regressivas das vendas (a correlação serial dentro da série) é desejável, principalmente quando se procura uma função de previsão expressa inteiramente em termos de variáveis defasadas. Nesses casos uma função do tipo ¹⁶:

$$\text{Vendas atuais} = a + b (\text{vendas do último período}) + c (\text{diferença das vendas do período precedente e do último período})$$

pode produzir previsões de grande precisão se as vendas são diretamente autocorrelacionadas e pode mesmo identificar pontos de reversão.

Como regra geral, a utilidade da autocorrelação aumenta com a diminuição da unidade de tempo. Uma função, um pouco mais complicada, foi determinada pelo autor para prever, com ótimos resultados, a tonelage esperada de despachos ferroviários trimestrais, empregando unicamente as propriedades das correlações em série dos despachos atuais ¹⁷. Em outro estudo foi descoberto que as despesas trimestrais de consumo eram tão próximamente relacionadas com as despesas do trimestre anterior, como com qualquer outra variável defasada, inclusive renda ¹⁸.

VALOR DE UMA TENDÊNCIA NO TEMPO

Costuma-se incluir uma variável de tempo em muitas funções de regressão para refletir a mudança em gostos, tendências tecnológicas e outros fatores diversos. É passível de discussão o valor desse procedimento para fins de previsão. A razão disso, essencialmente, é que o tempo, considerado isoladamente, não é *causa* de transformações eco-

16) O fato de que todas as equações citadas neste artigo sejam lineares — a forma mais simples — não deve ser interpretado como de negação ao valor de formas mais complexas não lineares. Numerosos exemplos de tais funções, ajustadas a dados de consumo e emprego, podem ser encontrados na publicação do *National Resources Committee, Patterns of Resource Use*, U. S. Government Printing Office, Washington, D. C., 1938.

17) *The Railroad Shippers' Forecasts*, capítulos IV e V.

18) Trabalho não publicado do autor.

nômicas ou de negócios. O tempo é o *meio* no qual os acontecimentos ocorrem, mas não é a causa desses acontecimentos. Assim, ao ser um bife preparado ao fogo, o cozimento leva certo tempo, mas a razão de o bife ficar cozido não é a passagem do tempo, mas a presença do fogo. Sem o fogo o cozimento não se daria, não importando quanto tempo viesse a decorrer.

Numa função de previsão um coeficiente estatisticamente significativo para uma variável de tempo indica, de fato, que determinadas variáveis causais relevantes são omitidas da relação e que a tendência no tempo está agindo em substituição a elas. O emprêgo de tal relação para a previsão significa, portanto, a suposição implícita de que as relações entre o tempo e aquelas variáveis irão continuar da mesma maneira que antes. Se não se conhecerem aquelas variáveis, como geralmente acontece, será uma suposição perigosa. Assim, a população dos EUA aumentou, mais ou menos numa configuração aritmética linear, durante a década de 1930, e então elevou-se, mais rapidamente, durante a década de 1940. Qualquer função ajustada conforme os dados da década de 1930, que tivesse incluído o tempo, em vez da população, como uma variável independente, teria resultado em completo fracasso se utilizada para fins de previsão da década de 1940.

Resulta dessa discussão que, sempre que se considera o tempo variável significativa numa função de previsão de vendas, cria-se uma relação espúria pela omissão de uma ou mais das variáveis causais. Caso essas variáveis omitidas sejam identificadas, por elas deverá ser substituída a variável de tempo. Em muitos casos não são identificáveis, não são passíveis de medição ou deixam de poder ser incluídas na função por serem muito numerosas. Mantém-se, nesses casos, a variável de tempo, mas levando-se em conta o perigo de fazer previsões com essa função ¹⁹.

19) Nôvo emprêgo de uma variável de tempo, proposta por ASHLEY WRIGHT, é o de utilizá-la como contrôle independente nas previsões de vendas, através de uma relação linear entre as vendas e o tempo. A vantagem desse procedimento é a de não haver dificuldade em prever o valor da variável independente (tempo). Uma previsão baseada nessa relação,

AVALIAÇÃO DA PRECISÃO DAS PREVISÕES

A avaliação e o registro da precisão das previsões é tão importante como a própria feitura da previsão, pois é somente por esse meio que se pode obter indicações para melhorar a precisão de previsões futuras. Isso é verdade, qualquer que seja o método de previsão adotado, e é, entretanto, uma consideração que parece ter sido desprezada em muitíssimos casos²⁰.

Existem muitas maneiras de medir a precisão, das quais três serão tratadas aqui. Antes, entretanto, vejamos como não se deve avaliar a precisão das previsões. Este aviso poderá parecer um tanto estranho, mas é o seguinte: *não confie no coeficiente de correlação*.

O coeficiente de correlação é a medida geralmente aceita da aderência de uma função às observações. Tem sido mais ou menos implicitamente suposto que a mesma estatística indicaria, também, a adequação de uma função para fins de previsão. Com efeito, se a função proporciona bom ajustamento para os dados observados, não é lógico supor que ela haja de proporcionar melhores previsões do que outra que proporcione um ajustamento deficiente?

Isso poderá parecer lógico, mas, como no caso da ópera *Porgy and Bess*, "não é necessariamente assim". De fato, no caso do trabalho do autor sobre as funções agregadas de poupança, nenhuma relação foi encontrada entre o coeficiente de correlação de uma função para o período de

é claro, nada mais é do que uma extrapolação de tendência. Entretanto, no caso de produtos que possuam tendências razoavelmente estáveis tal previsão é um controle valioso nas previsões desenvolvidas por quaisquer outros meios, porque, em tais casos, as modificações anuais das vendas pouca probabilidade apresentarão de se desviarem, apreciavelmente, da tendência de tempo projetada. Como consequência, a previsão de uma companhia que apresente, digamos, quatro desvios-padrões fora da projeção da tendência prevista seria imediatamente posta em dúvida.

- 20) C. M. CRAWFORD, *op. cit.*. Obtém-se, muitas vezes, melhor aproximação do desvio-padrão da previsão mais pela experiência anterior do que pela substituição direta em fórmulas estatísticas.

observação, e a precisão das previsões da função ²¹. Quando se tratou de diferentes períodos de observação, a relação entre o coeficiente de correlação e a precisão da previsão foi, se qualquer coisa, negativa: das funções estudadas a que aderiu deficientemente aos dados observados tendia a produzir previsões mais exatas do que outra com melhor ajustamento.

Deve êsse resultado levar-nos à fantástica conclusão de que as funções cujas previsões sejam mais precisas são as que apresentam menores coeficientes de correlação? Não, porque pode ser demonstrado que fatores especiais podem estar presentes em diferentes períodos de observação, distorcendo, completamente, o valor do coeficiente de correlação como indicador da precisão da previsão ²². Mesmo com idêntico período de observação e com a mesma variável a ser estimada, entretanto, parece que pouca confiança pode ser depositada, *a priori*, no coeficiente de correlação como medida da precisão das previsões. O mesmo é verdade quanto a outro dado estatístico relativo ao período de observação que foi testado: o valor absoluto do erro relativo.

A razão para o funcionamento deficiente dessas medidas pode muito bem ser a tendência de dar muita ênfase à inclusão, numa função, das variáveis que devam proporcionar bom ajustamento em determinadas situações, e muito pouca atenção aos fatores básicos que influenciam a variável durante longo período de tempo. Isso, certamente, não passa de hipótese, e muito trabalho ainda de-

21) *A Study of Aggregate Consumption Functions*, págs. 50 a 56. A mesma ausência de relação foi encontrada nas funções explicativas das previsões dos despachos ferroviários, que depois foram empregadas para fazer previsões.

22) É que o menor período no estudo das poupanças, 1929-40, aumentou a relativa importância dos anos de depressão, 1931-34. Isso elevou o coeficiente de correlação devido à maior amplitude da variável dependente e, ao mesmo tempo, reduziu a exatidão das previsões dos anos (prósperos) de após-guerra. Por outro lado, o emprêgo de um período maior de observação (1923-40), pela redução da importância dos anos de depressão, tendeu a reduzir o coeficiente de correlação, mas, igualmente, aumentou a exatidão da previsão das funções de após-guerra.

verá ser feito para determinar sua validade e dar-lhe significação operacional.

Encerramos esta discussão sobre o coeficiente de correlação, com o que é, provavelmente, outro comentário não decisivo sobre o seu valor, a saber: uma comparação dos coeficientes de correlação de duas funções, baseadas em dois períodos diferentes de observação, ou incluindo duas diferentes variáveis dependentes, não oferece nenhuma indicação sobre a precisão relativa das funções, *nem mesmo no período de observação*. Por outras palavras: a grandeza relativa dos coeficientes de correlação não pode fornecer nenhuma indicação quanto ao tamanho dos resíduos ou quanto aos erros relativos das estimativas no período de observação.

Esse ponto é ilustrado no gráfico anexo, tomado em parte do estudo do autor sobre poupanças, que mostra os coeficientes de correlação, os resíduos e os períodos de observação para seis diferentes formas de função agregadas de poupança. O gráfico indica de maneira bem significativa (a) como a mesma função ajustada para diferentes períodos de observação pode ter bem diferenciados coeficientes de correlação ainda que os resíduos para os anos superpostos, do período de observação, sejam quase idênticos, e (b) como a precisão das previsões de duas formas da mesma função, cada qual ajustada a um período diferente de observação, difere acentuadamente e apresenta uma relação negativa com o coeficiente de correlação.

A explicação para ambos os fenômenos está nas diferenças de amplitude, durante o período de observação, da variável que esteja sendo estimada (poupanças, no caso). Quando diferir a amplitude das variáveis dependentes de duas funções — seja em razão de diferenças no período de observação, nas unidades de medida ou nas variáveis empregadas, seja em razão de diferenças nas definições — nenhuma confiança pode ser depositada na comparação dos coeficientes de correlação, quer sob o prisma da adequação das funções, quer sob o aspecto de sua precisão de previsão.

TRÊS PADRÕES DE PRECISÃO

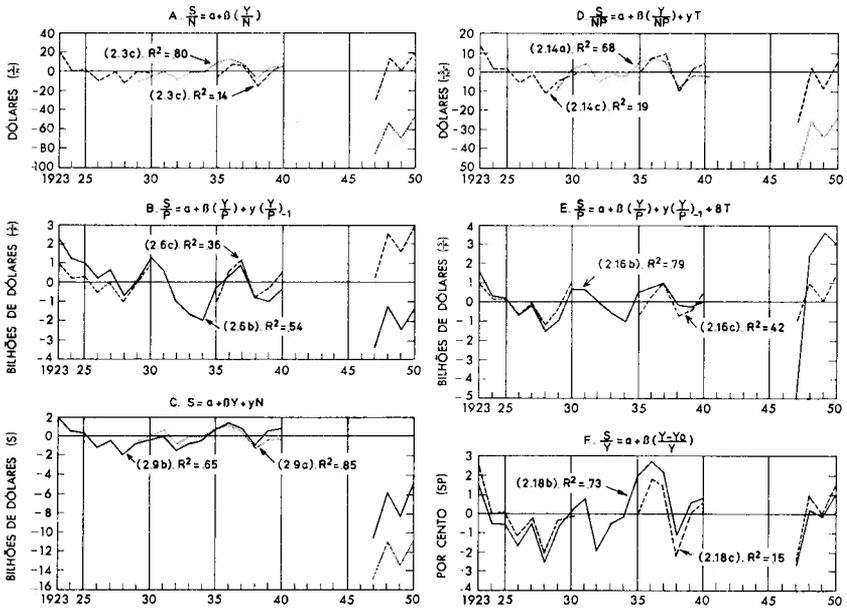
Existem pelo menos três maneiras pelas quais a precisão de um conjunto de previsões pode ser prontamente avaliada, cada qual com suas vantagens e desvantagens. Talvez, a mais evidente seja a do desvio absoluto médio percentual, isto é, a média percentual dos desvios das previsões em relação aos valores atuais considerados sem o sinal. Assim, no exemplo que a seguir apresentaremos é de 8,8% o desvio absoluto médio percentual, conforme demonstração a seguir²³.

| Período | Previsão | Valor Atual | Desvio do Valor Atual | Erro absoluto Percentual |
|---------|----------|-------------|-----------------------|--------------------------|
| 1 | 70 | 80 | — 10 | 12,5% |
| 2 | 105 | 100 | 5 | 5,0 |
| 3 | 98 | 90 | 8 | 8,9 |
| Média = | | | | 8,8% |

Embora essa medida ofereça uma idéia geral sobre a precisão de um conjunto de previsões, é de pouca importância quando se trata de julgar o valor relativo das previsões. Neste caso o que se deseja saber é: como pode ser comparado esse conjunto de previsões às previsões que poderiam ser obtidas por alguma técnica mais simples? Um conjunto de previsões pode produzir bons resultados em termos de precisão absoluta, como foi medido acima, mas não será de grande valor prático se se puder demonstrar que uma técnica mais simples poderia conduzir a uma precisão ainda maior. Igualmente, podem determinadas previsões desviar-se substancialmente, mas se não existir outro método que proporcione melhores resultados, essas previsões ainda poderão possuir valor prático.

23) A prática, algumas vezes adotada, de promediar a percentagem de erro em relação ao sinal pode ser muito enganadora, porque permite que grandes erros de sinais contrários se cancelem mutuamente. No caso citado, por exemplo, essa prática produziria um erro médio de 1,4%, que está longe de representar a verdadeira precisão. Isso não quer dizer, entretanto, que se deva ignorar os sinais dos desvios em outros casos, porque é freqüentemente pelo estudo dos sinais que se descobrem vícios das previsões.

GRÁFICO: Resíduos de Funções de Poupanças Selecionados no Período de Observação e no Período de Após-Guerra



OS NÚMEROS ENTRE PARÊNTESES QUE PRECEDEM OS SÍMBOLOS R^2 SÃO OS NÚMEROS DE IDENTIFICAÇÃO DAS FUNÇÕES NO ESTUDO ORIGINAL.
 FONTE: REPRODUZIDO, COM PERMISSÃO DO NATIONAL BUREAU OF ECONOMIC RESEARCH, DE A STUDY OF AGGREGATE CONSUMPTION FUNCTIONE. PÁGS. 52 E 53.

SÃO OS SEGUINTE OS PERÍODOS DE OBSERVAÇÃO:

LINHA SÓLIDA — 1923-40

LINHA INTERROMPIDA — 1923-30, 1935-40

LINHA PONTILHEADA — 1929-40

SÍMBOLOS EMPREGADOS: S — POUPANÇAS, Y — RENDA DISPONÍVEL, Y_0 — PICO PASSADO DA RENDA DISPONÍVEL, N — POPULAÇÃO, P — ÍNDICE DOS PREÇOS DE CONSUMO, AS LETRAS GREGAS (α, β, γ) = ALFA, BETA E GAMA) SÃO PARÂMETROS.

Por essa razão é desejável avaliar a precisão de um conjunto de previsões em relação a algum padrão simples de precisão. Esse padrão é fornecido pelo chamado “modelo prático”, que sustenta que o futuro representará uma extensão dos presentes níveis ou tendências. O mais simples “modelo prático” para dados anuais seria prever as vendas do próximo ano ao mesmo nível das do ano atual. (Se a unidade de tempo fôr menor do que *um* ano, deverá ser ajustada, em vista da variação estacional).

Essa previsão nada mais é do que a extensão dos níveis presentes. Para empresas ou ramos ainda em estágio de crescimento as projeções, pelo “modelo prático”, das tendências atuais seriam mais realistas, como, por exemplo, prevendo-se que:

$$\begin{aligned} & \text{Vendas do próximo ano} = \\ & = \text{Vendas deste ano} \times \frac{\text{Vendas deste ano}}{\text{Vendas do ano passado}}. \end{aligned}$$

Alternadamente, pode-se prever que o aumento nas vendas no próximo ano corresponderá somente a metade do aumento havido no ano passado ou a qualquer outra fração que pareça plausível em cada caso particular.

Comparando-se a precisão da previsão das funções de regressão ou dos métodos estudados, com os resultados obtidos pelo emprêgo dos “modelos práticos”, pode-se obter uma boa idéia do valor prático daqueles métodos mais elaborados, sob o modo de ver da empresa, e saber se vale a pena empregar os métodos mais elementares. Em relação a isto deve-se observar que um resultado em favor do “modelo prático” não oferece segurança quanto à sua utilização de maneira continuada (a menos que se tenha prova de que esse método particular não seja tão “elementar” assim). Ao contrário, esse fato indicará a necessidade de maiores estudos a fim de se obterem métodos de previsão que possam proporcionar melhores resultados em bases mais sólidas.

O terceiro padrão de precisão pretende remediar uma falha do primeiro, devido ao fato de que qualquer comparação dos níveis de previsão com as vendas realizadas provavelmente se apresentará favorável, em razão da autocorrelação das vendas. As vendas não mudam muito de um ano (ou período, depois do ajustamento para as variações estacionais) para o próximo, especialmente à medida que o nível de agregação se eleva. Não somente as vendas do próximo ano se relacionarão com os níveis deste ano, como também a previsão para o próximo ano será invariável e diretamente relacionada com os níveis deste ano. Introduce-se, portanto, uma relação espúria entre a previsão e o dado verificado que, em qualquer comparação de níveis, tende a produzir alta e injustificada impressão de precisão.

A maneira mais simples de remover esse fator espúrio (pelo menos, em muitos casos) é comparar as direções de mudança em vez de comparar os níveis. Dêsse modo é, em grande parte, afastada qualquer associação, devida aos níveis, entre as previsões e os dados atuais. Esse procedimento tem, também, a vantagem de indicar o sucesso do método de previsão na medição da direção do movimento das vendas, um dos problemas capitais na previsão empresarial.

O efeito da mudança do padrão de comparação, entre níveis, para outro de direções de mudança foi trazido à luz, eloqüentemente, em recente estudo que avaliava a precisão de um conjunto de previsões sobre a tonelage de despachos ferroviários trimestrais. Durante cerca de 25 anos esses despachos, em base agregada, aproximaram-se 9%, em média, dos dados verdadeiros. Ao mesmo tempo, entretanto, nenhuma relação foi encontrada entre a *mudança* prevista nos despachos e a mudança verificada²⁴. Surpreendente como possa parecer esse resultado, êle pode muito bem ser representativo das previsões empresariais e

24) *The Railroad Shippers' Forecasts*, Bureau of Economic and Business Research, Universidade de Illinois, 1953, capítulo III.

econômicas: fornecem boa aproximação do nível geral dos negócios, mas, com muita probabilidade, falham quanto à direção de mudança.

Finalmente, algumas palavras seriam úteis quanto à conveniência de manter registros das variações de precisão provocadas pelas diversas funções ou métodos de previsão. Se duas funções apresentarem erros médios similares, a que apresentar menor variabilidade deverá ser preferida, supondo-se as mesmas condições, pois é nela que se pode depositar maior confiança para fins de planejamento.

CONSIDERAÇÕES FINAIS

Este artigo, embora tenha tratado do emprêgo das técnicas de correlação na previsão de vendas, não pretende advogar o uso generalizado dessa técnica. As funções de correlação e de regressão constituem apenas um entre vários métodos disponíveis de previsão de vendas, e o método a empregar num caso determinado depende das circunstâncias que o cercam. Muitas vêzes é desejável o emprêgo de mais de um método a fim de testar uma previsão por meios mais ou menos independentes. Este artigo tentou pôr em relêvo várias maneiras pelas quais pode ser melhorada a eficiência das técnicas de correlação.