

Séries temporais e redes neurais: uma análise comparativa de técnicas na previsão de vendas do varejo brasileiro

Claudio Felisoni de Angelo[†]
FEA/USP

Ronaldo Zwicker^Ω
FEA/USP

Nuno Manoel Martins Dias Fouto[‡]
FEA/USP

Marcos Roberto Luppe[£]
FEA/USP

RESUMO: Uma importante atividade econômica em qualquer sociedade diz respeito à comercialização de bens. O varejo consiste exatamente no vínculo que se estabelece entre a indústria e o consumidor final. Prever as vendas é essencial para que se possa gerenciar de modo adequado os processos produtivos e de comercialização. No varejo esse aspecto reveste-se de importância ainda maior. Vender significa harmonizar os interesses dos que produzem com aqueles que compram. Portanto, o presente trabalho tem por propósito examinar comparativamente a aplicação de dois métodos de previsão de vendas de varejo no mercado brasileiro: as séries temporais e as redes neurais. A escolha dessas duas técnicas como objeto dessa comparação foi suscitada pela importância que essas duas concepções têm assumido na literatura. Embora a utilização de redes neurais tenha proporcionado a menor soma dos quadrados dos resíduos, pode-se dizer que os resultados empregando modelos do tipo ARIMA se mostraram praticamente equivalentes.

Palavras-chave: Séries temporais; redes neurais; previsão de vendas; varejo.

Recebido em 15/01/2009; revisado em 15/06/2010; aceito em 21/06/2010; disponível em 29/04/2011

Correspondência autores*:

[†] Doutor em Economia das Organizações pela FEA/USP.
Vinculação: FEA/USP.
Endereço: Avenida Luciano Gualberto 908, sala G-173 - Cidade Universitária, São Paulo – SP – Brasil
CEP: 05508-010.
E-mail: cfa@usp.br
Telefone: (11)3091 6045

^Ω Doutor em Economia das Organizações pela FEA/USP.
Vinculação: FEA/USP.
Endereço: Avenida Luciano Gualberto 908, sala G-173 - Cidade Universitária, São Paulo – SP- Brasil
CEP: 05508-010.
E-mail: rzwicker@usp.br
Telefone: (11)3091 6045

[‡] Doutor em Economia das Organizações pela FEA/USP.
Vinculação: FEA/USP.
Endereço: Avenida Luciano Gualberto 908, sala G-173 - Cidade Universitária, São Paulo – SP – Brasil
CEP: 05508-010.
E-mail: nfouto@usp.br
Telefone: (11)3091 6045

[£] Doutor em Administração pela FEA/USP.
Vinculação: FEA/USP.
Endereço: Avenida Luciano Gualberto 908, sala G-173 - Cidade Universitária, São Paulo – SP – Brasil
CEP: 05508-010.
E-mail: mluppe@usp.br
Telefone: (11)3091 6045

Nota do Editor: *Esse artigo foi aceito por Antonio Lopo Martinez*



Esta obra está licenciada sob a Licença Creative Commons – Atribuição-Use não-comercial-Compartilhamento pela mesma licença 3.0 Unported License

1. INTRODUÇÃO

Vender constituiu-se em uma atividade absolutamente essencial em qualquer tipo de empresa ou organização. As vendas geram o fluxo de recursos que financia as despesas correntes, os investimentos, e é claro, sustenta a própria rentabilidade. Essa atividade, entretanto, requer obviamente um trabalho prévio.

O sistema produtivo e a distribuição, dependendo do tipo de produto ou serviço, requerem a antecipação de ações preparatórias. A indústria precisa ultimar os pedidos e, uma vez processados, os lotes de bens ou a alocação dos serviços necessitam ser adequadamente direcionados. Essas transações demandam tempo, esforço e recursos.

O tempo diz respeito ao período completo da transação, isto é, desde o momento que a demanda se materializou no mercado até o instante em que a referida solicitação foi atendida. O esforço e recursos correspondem aos fatores de produção envolvidos e os respectivos níveis de utilização tanto de pessoas, como de materiais e equipamentos.

Quanto mais próximas as expectativas do volume de vendas estiverem das condições fornecidas de atendimento, mais eficiente será a operação. Altos volumes estimados sem a correspondente realização geram custos elevados de carregamento de estoques. Por outro lado, rupturas de fornecimento por escassez geram perdas reais de oportunidades e abrem espaço desnecessário para um melhor posicionamento da concorrência.

Se a previsão é uma função presente em todas as organizações, a avaliação das vendas futuras talvez seja uma das mais importantes e frequentes atividades nesse campo. Por essa razão, a literatura que trata do assunto é rica e extensa. Muitas e variadas técnicas vêm sendo aplicadas em um grande número de exemplos e casos relacionados aos mais diversos segmentos dos mercados de bens e serviços.

Genericamente, pode-se dividir as técnicas preditivas de vendas em dois grandes grupos. De um lado as técnicas qualitativas e de outro as quantitativas. As técnicas qualitativas procuram captar por meio de diferentes recursos analíticos a percepção dos indivíduos sobre os fluxos futuros. Os modelos quantitativos por outro lado, baseiam-se em estruturas conceituais bem definidas e objetivas para realizar suas previsões.

Pode-se dizer que se de um lado as técnicas qualitativas são mais versáteis e ricas no sentido de incorporar diferentes situações, de outro são mais limitadas exatamente pela falta de objetividade de seus processos. As técnicas quantitativas por seu turno são menos flexíveis, porém permitem uma discussão mais clara sobre os pressupostos em que se assentam seus

resultados. Em outras palavras pode-se argumentar que os modelos qualitativos procuram ampliar os horizontes perceptivos captando subjetivamente as conseqüências emanadas da estrutura mental dos indivíduos (BROCKHOFF, 1983). As concepções quantitativas por sua vez apóiam-se em estruturas analíticas formais e perfeitamente definidas.

Não se pode estabelecer preferências entre os diferentes métodos. A riqueza da literatura evidencia que as técnicas não competem, na verdade complementam-se. A aplicação conjunta dos dois arcabouços analíticos tem sido largamente empregada visando a enriquecer o rigor dos modelos quantitativos com a flexibilidade criativa das concepções qualitativas. Desse modo, hoje em dia, as áreas de planejamento das empresas devidamente organizadas combinam diferentes abordagens a fim de realizar a atividade de previsão considerando o maior volume de informações disponíveis, sejam elas qualitativas ou quantitativas.

A interação entre os indivíduos melhora a qualidade da previsão (SNIEZEK, 1989). A participação ativa de um número maior de indivíduos alarga os horizontes analíticos ampliando a capacidade crítica dos envolvidos. Esse processo pode ser ainda mais eficiente apoiando suas interpretações em parâmetros mais objetivos fornecidos pelos modelos quantitativos (ANG, O'CONNOR, 1991; FRANSES, 2008).

Reconhecendo a importância da combinação das técnicas quantitativas e qualitativas, este trabalho tem por propósito específico examinar comparativamente duas técnicas quantitativas aplicadas à previsão das vendas do varejo no Brasil: as séries temporais e as redes neurais. Fixou-se nestas duas abordagens em razão da ampla aplicação dos dois métodos em muitos problemas semelhantes registrados na literatura sobre esse assunto.

A comparação entre técnicas de previsão também tem sido largamente contemplada nos trabalhos acadêmicos. Diversos estudos têm sido realizados comparando resultados fornecidos por diferentes concepções teóricas. Este trabalho inspira-se nessa linha da produção científica procurando oferecer subsídios ao processo preditivo das vendas no varejo brasileiro.

O artigo está organizado em cinco itens. No item dois é delineado o panorama da economia brasileira em que se baseiam as projeções das vendas de varejo. O item 3 apresenta uma revisão da literatura sobre os modelos de previsão associados às séries temporais e as redes neurais. O item quatro estabelece os modelos, sintetiza e compara os resultados obtidos. Finalmente, o item cinco explicita as principais conclusões e assinala as possíveis extensões do artigo.

2. A ECONOMIA BRASILEIRA E O MERCADO DE CONSUMO

Um dos principais componentes da demanda agregada é sem dúvida o consumo: os gastos das famílias na compra de produtos ou na contratação de serviços. A evolução dessa variável depende das condições gerais da economia delineadas principalmente pela evolução da renda, do nível geral de preços e da taxa de juros. O crescimento da renda real favorece o consumo do mesmo modo que as condições mais favoráveis de crédito, expressas em termos de menores taxas de juros e/ou prazos maiores.

Um das possibilidades para expressar a evolução do consumo é considerar as vendas de produtos destinadas ao consumidor final, ou seja, o varejo de bens. Embora o consumo seja mais abrangente incluindo todas as despesas realizadas, neste trabalho a perspectiva analítica cinge-se exclusivamente no que diz respeito à comercialização de bens.

O comércio de bens em decorrência da elevada inflação e da conseqüente perda do poder aquisitivo que assolou o País desde principalmente o final da década de 70 evoluiu muito lentamente durante todos os anos 80 e parte da década de 90. No período de 1960 a 1990 a concentração de renda diminuiu sistematicamente. Os 50% mais pobres que representavam em 1960 17,7 % da população caiu para pouco menos de 14% em 1980 e para 12% no início dos anos 90 (IPEADATA, 2008).

Diversas tentativas foram empreendidas com o propósito de controlar o ritmo de expansão dos preços e a conseqüente desorganização da economia. Entretanto, todas as iniciativas, como o Plano Cruzado, Plano Bresser e Plano Collor, por uma ou outra razão, ou ainda, mais propriamente, por uma combinação de fatores econômicos e políticos fracassaram.

Em 1994, entretanto, foi lançado o Plano Real. De fato, pela primeira vez, depois de diversas iniciativas a inflação foi efetivamente reduzida e mantida em patamares similares aos registrados nas economias dos países desenvolvidos. O índice geral de preços declinou de aproximadamente 5.150,00 % em junho de 1994, para aproximadamente 10% em dezembro de 2001. Na seqüência, a evolução dos preços foi sempre muito inferior aos níveis verificados antes da edição do Plano (GIAMBIASI, 2005).

As mudanças promovidas pela redução dos patamares inflacionários foram muito amplas. A exemplo de outros mercados, o comércio varejista no Brasil passou também por grandes e profundas transformações. Até então, ou seja até 1994, os aspectos operacionais do comércio, ou seja a função de ligação entre a indústria e o consumidor final foram

negligenciadas em razão dos ganhos oferecidos pelas aplicações financeiras. De fato, o lucro proveniente da atividade precípua do comércio, ou seja, a compra e venda de produtos ficava muito aquém dos resultados que podiam ser conseguidos no mercado financeiro (SESSO FILHO, 2001).

Ao reduzir a inflação e obrigando as empresas a atentarem pelas respectivas operações a competitividade no comércio de bens cresceu significativamente. A variância dos preços relativos contraiu-se naturalmente expressando de forma mais adequada e perene os valores intrinsecamente comunicados pelo mercado. Essa maior visibilidade dos valores explicitados nos preços implicou em uma redução das margens forçada pela maior intensidade da rivalidade entre as empresas.

O impacto do Plano Real vê-se na desconcentração de renda constatada entre os anos 1996 e 2006. Ao mesmo tempo em que a economia cresceu, a porção dos 50% mais pobres aumentou seu peso relativo na renda, de 12,09% para 14,47% (IPEADATA, 2008). Apesar da reordenação econômica promovida pelo plano de estabilização, a expansão do consumo somente se deu de forma contínua a partir de meados de 2003.

As limitações à expansão do consumo nos primeiros anos do Real podem ser atribuídas ao baixo crescimento da economia. Ao longo dos primeiros cinco anos do Plano Real, o mercado financeiro passou por três importantes crises financeiras: México (1994), Ásia (1997) e Rússia (1998). Nestes anos, embora a inflação tenha mantido-se em níveis bastante baixos, o crescimento da economia foi muito discreto, apenas 2,8% ao ano (FERRARI-FILHO, DE PAULA, 2003).

Como se vê no Gráfico 1, a melhora do quadro internacional permitiu uma expansão mais acelerada da renda a partir do início de 2003, onde também fica evidenciado o caráter sazonal das vendas de varejo em ciclos de doze meses. Tal expansão associou-se, ou ainda sustentou, um processo gradual, mas contínuo, de queda das taxas de juros internas e um alongamento sistemáticos dos prazos médios de pagamentos. O aumento da renda real pode ser ilustrado, por exemplo, pela situação do grupo constituído pelos trabalhadores do setor privado com carteira assinada. De meados de 2003 a maio de 2008 o referido grupo registrou um ganho real de oito pontos percentuais. Por sua vez a taxa de juros no financiamento ao consumo declinou substancialmente desde maio de 2003, passando de 98,09% ao ano para 48,88% no final de 2007. Ao mesmo tempo, o prazo médio das operações de financiamento dilatou-se de 249 dias para 441 dias (BANCO CENTRAL, 2008).

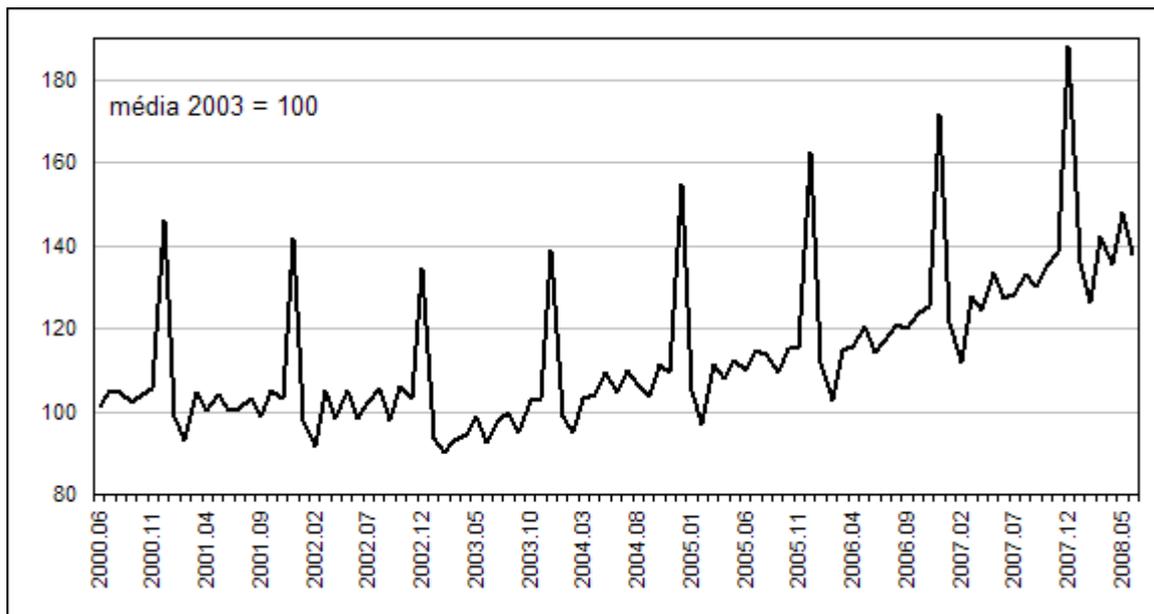


Gráfico 1 - Vendas reais do varejo brasileiro
Fonte: IPEADATA, 2008

Mais uma vez essas condições favoráveis implicaram também em uma desconcentração da renda. Em recente pesquisa encomendada pela financeira do grupo francês BNP Paribas, a Cetelem, em parceria com o Instituto de Pesquisa Ipsos, sobre a distribuição da população brasileira por classe de consumo, a classe C, que representava 36% em 2006, passou para 46% em 2007, chegando a 86 milhões de pessoas. Já as classes D/E, que até 2006 tinham uma proporção maior que a C, apresentaram uma queda de 46% para 39%, caindo para 73 milhões de pessoas, em 2007. A pesquisa demonstra também que houve diminuição na desigualdade de renda, com uma ligeira queda da renda média das classes A/B, ascensão de um grande contingente para a classe C e um pequeno aumento da renda média das classes D/E (DE CHIARA, 2008).

É no ambiente econômico vivenciado no Brasil a partir de junho de 2000 que se pretende examinar o desempenho das vendas no varejo brasileiro, investigando comparativamente a utilização das séries temporais e das redes neurais na previsão dos volumes comercializados. Com este propósito, na próxima seção é feita uma revisão dessas duas técnicas, que serão posteriormente aplicadas aos dados disponíveis. O período de estimação inicia-se em junho de 2000 e termina em junho de 2007. Os dados referentes ao período que vai de julho de 2007 a junho de 2008 serão usados para comparar o grau de ajuste das duas técnicas de previsão examinadas neste artigo.

3. SÉRIES TEMPORAIS E REDES NEURAS

Este item foi dividido em dois tópicos distintos. No primeiro, faz-se um retrospecto do desenvolvimento da econometria das séries temporais enfatizando os métodos de alisamento e os modelos estocásticos. No segundo, é feita uma síntese do argumento das redes neurais e seu recente desenvolvimento. Procura-se explicar a natureza desse método ressaltando sua aplicabilidade nos problemas de predição.

3.1 Econometria das séries temporais

As séries temporais constituem-se em um capítulo absolutamente a parte da econometria. Diferentemente dos modelos econométricos, o estudo das séries temporais têm o objetivo precípua de realizar previsões. Não há preocupação em estabelecer os mecanismos causais, pretende-se unicamente realizar previsões acuradas ao máximo.

A abordagem mais tradicional consiste na aplicação das técnicas de alisamento. Holt em 1957 e posteriormente Winters em 1960 conceberam o modelo que veio a ser conhecido na literatura como Holt-Winters, no qual se identificam três parâmetros: componente permanente, tendencial e sazonal. Em 1969, Pegel amplia a contribuição dos autores anteriores considerando especificações aditivas e multiplicativas (DE GOOIJER, HYNDMAN, 2006), explicitadas a seguir:

$$\hat{Y}_{t+k} = \alpha + b \cdot K + c_{t+k} \quad (1)$$

$$\hat{Y}_{t+k} = (a + b \cdot K) \cdot c_{t+k} \quad (2)$$

O valor de “a” corresponde ao componente permanente, “b” ao parâmetro representativo da tendência e “c_t” ao fator associado ao comportamento sazonal aditivo ou multiplicativo. Os valores desses parâmetros são dados pelas seguintes expressões:

$$a(t) = a \cdot [y_t/c_t(t-s)] + (1-\alpha) \cdot [c \cdot (t-1) + b \cdot (t-1)] \quad (i)$$

$$b(t) = \beta \cdot [a(t) - a \cdot (t-1)] + (1 - \beta) \cdot b \cdot (t - 1) \quad (ii)$$

$$c(t) = \gamma \cdot [y_t/a(t)] + (1 - \gamma) \cdot c_t \cdot (t-s) \quad (iii)$$

Para que a suavização seja obtida é preciso, portanto, considerar a componente permanente, a tendência e a sazonalidade. Os parâmetros de amortecimento são atribuídos de modo a realizar o melhor ajuste possível aos dados. Os programas estatísticos calculam esses parâmetros interativamente e automaticamente de modo a minimizar os erros. Um exemplo desse procedimento pode ser encontrado em Segura e Vercher (2001). Os autores utilizam-se do Solver para estabelecer os parâmetros mais adequados dada uma determinada série.

Um grande avanço na econometria das séries temporais deve-se à utilização de modelos estocásticos. Tal modelagem parte do pressuposto que as séries são estacionárias. Caso as séries não sejam estacionárias em nível, deve-se tomar as primeiras diferenças. Se a série continuar manifestando comportamento não estacionário, conforme evidenciado por testes específicos é necessário proceder às segundas diferenças. Obviamente esse processo continua até que a série torne-se estacionária. Entretanto, a prática tem demonstrado que as variáveis econômicas tornam-se estacionárias no máximo na segunda diferença.

Uma série estacionária pode ser modelada de diversas formas considerando basicamente dois processos: o auto-regressivo e o de médias móveis. A modelagem visa a reproduzir os valores da variável de interesse “Y” a partir dos dois processos, isoladamente ou de forma combinada. O modelo auto-regressivo expressa o valor corrente de “Y” em função dos valores da variável registrados no passado. A estrutura de médias móveis realiza a representação em termos dos erros incorridos nos períodos anteriores. A expressão (3) abaixo expressa um modelo de geração de “Y” auto-regressivo:

$$(Y_t - \delta) = \alpha_1 (Y_{t-1} - \delta) + \alpha_2 (Y_{t-2} - \delta) + \dots + \alpha_p (Y_{t-p} - \delta) + u_t \quad (3)$$

Onde δ é a média da variável Y. Diz-se que (3) descreve um processo auto-regressivo de ordem “p”, ou em outras palavras, um processo AR(p). De outro modo, como se assinalou no parágrafo anterior, os valores de Y podem também ser gerados por combinações lineares de termos de erro (ruído branco), ou seja:

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2} + \dots + \beta_q u_{t-q} \quad (4)$$

A expressão (4) mostra um modelo de médias móveis de ordem “q”, ou ainda um processo MA(q). Pode-se finalmente supor um processo de geração combinando termos auto-regressivos e de média móveis. Neste caso diz-se que o modelo é ARMA (p,q). Acrescentando-se o grau de diferenciação para tornar a série estacionária obtêm-se os modelos ARIMA (CHU, ZHANG, 2003).

O trabalho de Box e Jenkins (1976) é, fundamentalmente, um critério de estruturação e composição do modelo. O método divide-se em quatro etapas distintas. A primeira parte consiste na especificação adequada para as séries estacionárias dos termos auto-regressivos e os representativos das médias móveis. A segunda etapa diz respeito ao processo de estimação. Com base nos valores obtidos para os parâmetros passa-se para a terceira etapa que é o exame dos resultados, principalmente no que concerne aos resíduos. Uma indicação da adequação do modelo é que tais resíduos precisam ser qualificados como ruído branco. Finalmente na quarta e última fase se realiza a previsão.

De Gooijer e Hyndman (2006) apresentam um excelente resumo a respeito da utilização das técnicas de séries temporais em modelos de previsão nos últimos 25 anos, até 2006.

3.2 Redes Neurais Artificiais (RNAs)

As redes neurais artificiais são inspiradas no comportamento de redes neurais biológicas. Elas são compostas de nós denominados neurônios ligados entre si por conexões que guardam relação com as ligações formadas por axônios e dendritos componentes de células neurais biológicas. Em geral uma RNA é constituída de um grande número de nós organizados em camadas e que se conectam a outros nós por meio de conexões nas quais fluem sinais, da mesma forma que entre neurônios biológicos fluem sinais de natureza elétrica. Cada conexão de um nó a outro nó tipicamente tem associado um peso que, guardadas as devidas diferenças, pode ser entendido como representando o grau de acoplamento existente entre dois neurônios biológicos.

A tarefa realizada por um nó é tipicamente simples. Inicialmente consiste no recebimento dos sinais, oriundos de outros nós pelas conexões de entrada, ponderados pelos correspondentes pesos das conexões e cuja somatória irá corresponder ao total do sinal de entrada do nó. Em seguida o limiar de ativação do neurônio, representado na figura 1 por a_j , é adicionado a este total. Este limiar é tipicamente um valor negativo que funciona como ponto a partir do qual o (total do) sinal de entrada irá determinar um sinal de saída. Sinais fracos de entrada serão inibidos por este limiar. Em seguida é aplicada uma função ao sinal de entrada que computa um valor y_j correspondente ao sinal de saída do nó e que é transmitido, como entrada, aos outros nós aos quais o primeiro está conectado. A função é denominada por função de ativação e em geral é de natureza não linear.

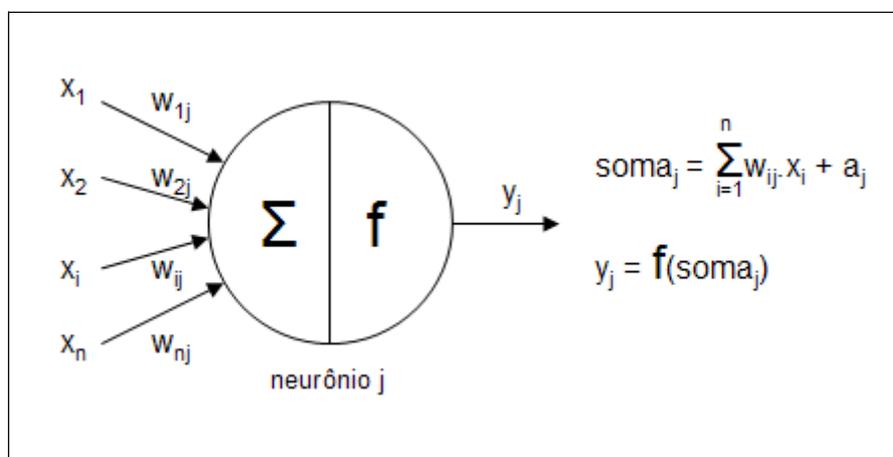


Figura 1 – Funcionamento do neurônio artificial
Fonte: Elaborado pelos autores

O modelo mais difundido de rede é o modelo MLP (*Multi Layer Perceptron*) e que também é o modelo mais utilizado na previsão de séries temporais. Redes deste tipo são constituídas por uma camada de nós de entrada, uma ou mais camadas intermediárias e uma camada de saída. Conforme a figura 2, os sinais de entrada são transmitidos da camada de entrada para a camada de saída através das camadas intermediárias e não existe retroalimentação dos sinais intermediários ou de saída (a alimentação é dita *feedforward*). Do ponto de vista estatístico cada nó de entrada representa uma variável independente e cada nó de saída representa uma variável dependente que está sendo prevista por meio da rede (PALMER *et al.*, 2006).

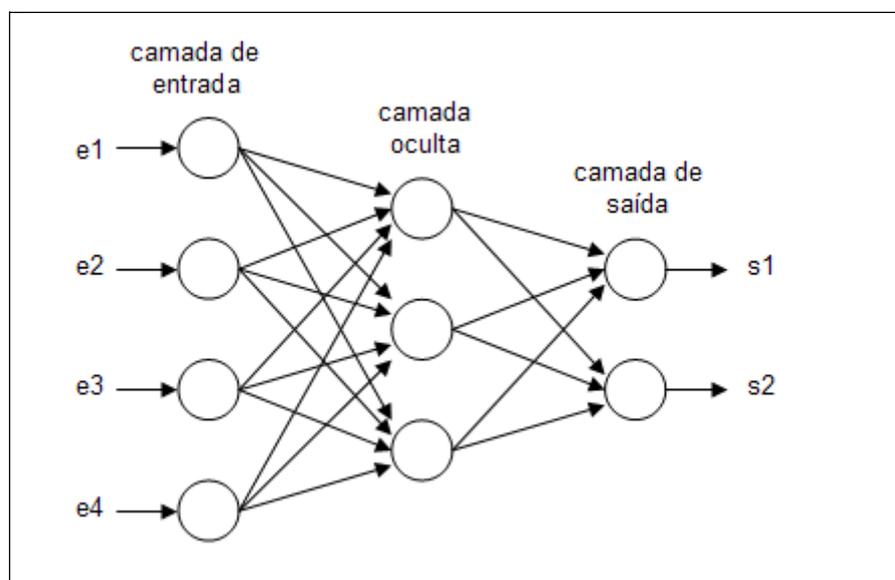


Figura 2 - Exemplo de rede neural
Fonte: Elaborado pelos autores

Em geral uma rede MLP passa por uma etapa inicial dita de treinamento para a seguir ser executada o que, no caso da previsão, corresponde a tarefa de efetivamente computar previsões. Na etapa de treinamento são computados os pesos das conexões a partir de um conjunto de pares entrada-saída, pertinentes aos valores conhecidos das variáveis independentes e correspondentes valores das variáveis dependentes. Nessa fase, a partir dos valores conhecidos dos sinais de entrada, são computados os correspondentes valores dos sinais de saída, que por sua vez são comparados com os valores conhecidos das variáveis dependentes. Em síntese a diferença entre os valores conhecidos e os computados determina um ajuste iterativo nos pesos das conexões de forma a minimizar esta diferença ou erro. Na etapa de execução são apresentados à rede treinada os valores das variáveis de entrada e ela computa os valores das variáveis de saída ou valores previstos. Nas duas etapas a computação de saídas a partir de entradas em cada nó é realizada de acordo com o funcionamento anteriormente especificado do neurônio artificial.

A obtenção da diferença entre os valores conhecidos e os computados mantém analogia com a aprendizagem supervisionada por um professor, onde o resultado computado pelo aprendiz é sistematicamente avaliado pelo professor e a diferença com o resultado desejado determina mudanças no comportamento do aprendiz. Por outro lado, o ajuste dos pesos mantém relação com a fisiologia do aprendizado animal que, de forma extremamente simplificada, remete a uma aproximação ou afastamento das ligações (sinapses) entre neurônios biológicos. Quanto maior o peso tanto mais estreita é a ligação. A analogia entre o aprendizado de RNAs e animais pode ser vislumbrada considerando o condicionamento de um cão. O condicionamento se vale de um procedimento muito simples: recompensar o comportamento desejado e penalizar o indesejado. Na medida em que o cão reage corretamente a uma ordem de sentar devemos recompensá-lo, por exemplo, com um simples afago. Podemos entender que este afago faz com que as conexões neurais, que vão desde os neurônios responsáveis por captar o som da ordem até os terminais neurais que acionam os músculos e dobram suas pernas traseiras, têm suas ligações aproximadas de forma a facilitar o fluxo do sinal de entrada (som) até os terminais acionadores dos músculos. Dessa forma, ao receber uma nova ordem de sentar ele tende a sentar com mais facilidade. Inversamente, se ao reagir indevidamente a ordem de sentar (por exemplo, deitando) o penalizamos (por exemplo, com um leve tranco na coleira) e assim estaremos aumentando a distância entre as conexões neurais que conduzem o sinal de entrada até os músculos indevidos. Dessa forma, ao receber uma nova ordem de sentar o fluxo do sinal para os músculos indevidos é dificultado fazendo com que o comportamento indesejado seja suprimido.

A designação “artificial” na denominação das redes se refere ao fato destas construções analíticas apenas se inspirarem nos sistemas biológicos, particularmente no estudo do cérebro humano. A pesquisa em RNAs tem mostrado que elas possuem capacidades significativas de classificação e reconhecimento de padrões. Guardadas as devidas diferenças, RNAs têm capacidade de aprendizado e de generalização a partir da experiência análoga aos seres humanos. Conforme Widrow *et al.* (1994), RNAs estão sendo usadas em variadas aplicações nas áreas de negócios, indústria e ciência com muito sucesso.

Uma das aplicações de RNAs diz respeito a procedimentos de previsão em geral. Sem querer desqualificar procedimentos já bem estabelecidos de previsão, as RNAs oferecem eventualmente alternativas interessantes e atrativas para quem estuda e realiza previsões.

Conforme Zhang *et al.* (1998), várias características específicas de RNAs as tornam atrativas no âmbito de procedimentos de previsão. Em primeiro lugar, em oposição aos

métodos tradicionais, RNAs constituem métodos direcionados por dados e auto-adaptativos no sentido de que requerem poucas premissas relativas aos modelos que representam os problemas em estudo. Isso significa que elas podem aprender a partir de exemplos capturando relacionamentos sutis entre os dados disponíveis, mesmo que estes relacionamentos sejam a priori desconhecidos ou difíceis de descrever. Posto de outra forma, usa-se uma RNA quando não se conhece a exata natureza do relacionamento entre entradas e saídas, caso se conhecesse o relacionamento então se modelaria o mesmo diretamente. Em síntese RNAs são indicadas para auxiliar na resolução de problemas que necessitam de conhecimento difícil de especificar, desde que haja suficientes dados ou observações.

Pode-se dizer que a capacidade de aprendizagem das RNAs, a partir da experiência, constituiu-se em uma forma muito útil de abordar problemas para os quais há dados sem que, entretanto, se disponha de maior conhecimento a respeito dos processos geradores das referidas informações. Claramente existem muitas situações onde é mais fácil obter dados do que obter bons modelos teóricos relativos ao problema em estudo. Esta situação é similar a apontada anteriormente quando se enunciou os modelos econométricos de séries temporais. Também a exemplo destas concepções analíticas, a modelagem baseada nas redes neurais pode ser muito limitada quando se dispõe de poucas observações.

Uma outra característica registrada na literatura sobre as RNAs é que, uma vez estabelecida uma rede estável, pode-se realizar inferências ou generalizações. Redes assim estruturadas geradas pelo processo de aprendizagem são, em princípio, capazes de inferir corretamente resultados associados a dados não usados explicitamente no seu treinamento inclusive quando afetados de ruído. Essa capacidade de inferência tem particular interesse em previsão onde a associação de dados futuros observados com dados passados constitui um treinamento para em seguida poder realizar generalizações por meio da associação de dados observados com resultados futuros não observados.

Tais associações constituem aproximações universais de funções. Conforme Haykin (2001) foi demonstrado que determinados tipos de RNAs podem aproximar qualquer função contínua a qualquer nível de precisão. De modo geral, RNAs constituem formas funcionais mais gerais e flexíveis do que as formas com as quais os métodos estatísticos convencionais conseguem trabalhar. Em geral assume-se que existe algum relacionamento, conhecido ou não, entre entradas (variáveis ou valores passados na situação de previsão) e saídas (variáveis ou valores futuros). Tipicamente os métodos estatísticos convencionais de previsão possuem

limitações para estimar este relacionamento ou função e, em princípio, as RNAs podem contribuir para ultrapassar estas limitações.

Desta forma as RNAs são intrinsecamente capazes de capturar não linearidades enquanto os métodos tradicionais de previsão tipicamente têm por detrás modelos lineares. Por exemplo, os modelos gerados pelo método de Box-Jenkins, como se viu, assumem que a série de tempo analisada seja gerada a partir de processos lineares. É fato reconhecido que modelos lineares têm a grande vantagem de poderem ser entendidos e analisados em profundidade e detalhe, além de serem fáceis de explicar e implementar. Por outro lado, pode ser amplamente questionável assumir que o modelo por detrás dos dados seja linear, de fato observa-se que isso é totalmente inapropriado quando os dados decorrem de um processo não linear, o que não é incomum na prática.

Apesar de existirem métodos estatísticos convencionais não lineares que assumem um modelo não linear pré-definido, eles são intrinsecamente restritos quando se assume tal modelo sem conhecimento maior dos mecanismos que geraram os dados em questão. Ao formular um modelo pelos métodos convencionais estamos tipicamente limitando os possíveis mecanismos geradores, claramente este modelo pode ser insuficientemente para capturar todas as características não lineares dos dados. Em princípio pode-se argumentar que RNAs, pelo fato de serem abordagens não lineares sustentadas exclusivamente pelos dados, são capazes de modelagem não linear sem qualquer conhecimento a priori dos mecanismos que ligam as variáveis de entrada às variáveis de saída. Como argumentam Zhang *et al.* (1998), RNAs constituem um recurso de modelagem mais geral e flexível para as tarefas de previsão.

A utilização de RNAs em previsão se difundiu a partir da introdução do algoritmo de retro-propagação para treinamento de redes de múltiplas camadas por volta de 1986. Em síntese foi este algoritmo que estabeleceu inicialmente a capacidade não linear das RNAs. Desde esta época muitos autores realizaram testes comparativos entre a efetividade das redes e recursos estatísticos. Um extenso estudo recente que utilizou uma base de dados padronizada e de uso público para testes pode ser encontrado em Zhang e Kline (2007).

Deste modo, no capítulo seguinte são discutidos os modelos de estimação utilizados no presente texto. Apresenta-se os modelos resultantes das aplicações das técnicas de alisamento, o modelo suscitado pelo método Box-Jenkins e o decorrente da utilização das redes neurais.

4. MODELOS DE ESTIMAÇÃO E RESULTADOS

Esta parte do trabalho é reservada à apresentação dos resultados. Inicialmente são apontados os modelos e as respectivas previsões. Utilizou-se como período para definição dos modelos os dados de junho de 2000 a junho de 2007. O período de julho de 2007 a junho de 2008 foi considerado para examinar o ajustamento comparativo dos diferentes métodos. Os dados referem-se ao índice das vendas reais do varejo brasileiro (junho de 2000 =100) publicadas no *site* www.ipeadata.gov.br (IPEADATA, 2008). Ao final é realizada uma comparação dos métodos usando como critério de ajuste o somatório do quadrado dos resíduos, isto é, as diferenças quadráticas entre os valores reais e os previstos segundo as formulações alternativas.

4.1 Alisamento exponencial aditivo e multiplicativo

Nesta parte do trabalho são apresentados os procedimentos, os modelos e os correspondentes resultados. Iniciando pelos métodos de alisamento estimou-se o modelo Holt-Winters na forma aditiva e na multiplicativa. Os resultados referentes aos parâmetros dos dois modelos encontram-se na Tabela 1. Tomando como base as correspondentes funções realizaram-se as projeções para o exercício julho de 2007 a junho de 2008. Os resultados são apresentados na tabela 4 que registra os somatórios dos quadrados dos resíduos de todos os modelos.

<i>Parâmetros</i>	<i>Aditivo</i>	<i>Multiplicativo</i>
A	0.4100	0.4600
B	0.1000	0.1400
γ	0.0000	0.0000

Fonte: Elaborado pelos autores

4.2 Modelo ARIMA

O segundo modelo estimado baseou-se na técnica Box-Jenkins para obtenção de funções preditivas do tipo ARIMA. A sazonalidade dos dados de vendas gera uma série não estacionária. Como mostra Enders (1995, p.227), as séries podem exibir raízes unitárias sazonais. Essa foi exatamente a situação dos dados confirmada pelo teste Dickey – Fuller. Visando a tornar a série estacionária tomou-se a diferença entre o valor corrente (logaritmo do índice) e as vendas defasadas de doze meses ($DY_{12} = Y_t - Y_{t-12}$). Aplicando-se novamente o teste Dickey – Fuller verificou-se que a série assim considerada exibiu padrão estacionário.

Tendo sido obtida a estacionariedade da variável de interesse passou-se para a etapa de estimação, considerando as defasagens das vendas, a distribuição dos erros no tempo, o critério da economicidade na representação dos processos estocásticos (Akaike-Schwarz) e é claro a significância dos parâmetros estimados. Os melhores resultados foram obtidos trabalhando-se com um modelo conforme apresentado na Tabela 2. Verificou-se adicionalmente a estacionariedade dos resíduos do modelo, rejeitando-se a hipótese de existência de raiz unitária.

TABELA 2 - RESULTADOS DO MODELO ARIMA

<i>Variáveis</i>	<i>Coefficientes</i>	<i>Erro padrão</i>	<i>P</i>
C	0.034189	1.061767	0.2922
AR(1)	0.275941	2.330029	0.0229
AR(2)	0.350662	3.104815	0.0028
AR(3)	0.306025	2.609613	0.0112
R²	72,26%		
R² ajustado	71%		
Akaike	-5.830982		
Schwarz	-5.702497		
Durbin-Watson	1.863138		
F	57.30172		0.0000
Jarque-Bera	6,260		5%

Fonte: Elaborado pelos autores

4.3 Redes neurais

A aplicação de RNAs em problemas de previsão não é tarefa trivial, pois muitos aspectos precisam ser considerados e muitas decisões precisam ser tomadas. É muito importante a escolha apropriada da arquitetura da RNA o que envolve a escolha do número de camadas intermediárias, do número de nós em cada camada e a interconexão destes nós. Também precisam ser considerados o algoritmo de treinamento, as funções de ativação, a normalização e pré-processamento dos dados, a escolha dos conjuntos de dados para treinamento, verificação e teste e as medidas de qualidade do ajuste. Zhang *et al.* (1998) fazem uma síntese destas questões e da bibliografia que lhes dá tratamento empírico.

Não existe método definido para a determinação dos múltiplos parâmetros que compõem no problema de aplicação de RNAs a problemas de previsão. Não é possível estabelecer uma solução ótima, entretanto, existem diretrizes e regras práticas que em geral conduzem a soluções satisfatórias. Para problemas de previsão são tipicamente adotadas redes com apenas uma camada intermediária e o número de nós desta camada é determinada por experimentação com a sugestão de que esse número fique em torno do número de nós de

entrada. Em problemas de previsão de séries temporais o número de nós de entrada em geral corresponde ao número de observações defasadas que se consideram necessárias para desvendar o padrão de comportamento e previsão da série. Por exemplo, para séries mensais anuais univariadas é prática adotar o ciclo sazonal em meses como número de nós de entrada e o número de nós de saída é tipicamente o horizonte de previsão (em geral igual a um). Além disso, a interconexão entre os nós é completa, isto é, os nós de uma camada são conectados a todos os nós da camada seguinte num esquema de rede tipo *feedforward*.

Em problemas de previsão costuma ser usada como função de ativação a função logística para os nós intermediários e a função logística ou linear para os nós de saída. Zhang *et al.* (1998) alertam que a função linear limita a capacidade da rede modelar séries temporais que possuam tendência e sugerem o processamento prévio dos dados para eliminar esse efeito. Com relação ao algoritmo de treinamento o método mais difundido é o algoritmo de retro-propagação com a especificação de seus parâmetros de taxa de aprendizagem e momento de forma empírica. Pode ser indicada a utilização de métodos mais eficientes como o algoritmo de Levenberg-Marquardt ou Quickprop que tendem a convergir mais rapidamente e evitam ficar presos a mínimos locais.

Quanto a normalização dos dados deve ser observado que, quando são usadas funções de ativação não lineares nos nós de saída, os valores desejados de saída precisam ser transformados para o intervalo de saídas que a rede provê. De modo geral recomenda-se a normalização de todos os dados no caso de séries temporais inclusive se a função de ativação de saída for linear. A questão da escolha dos conjuntos de dados para treinamento, verificação e teste é peculiar de RNAs. Na medida do possível devem ser conjuntos disjuntos onde o conjunto de treinamento é usado exclusivamente para a determinação dos pesos e o conjunto de verificação é usado para verificar se o erro de saída (para este conjunto) voltou a aumentar indicando que a rede está “memorizando” os dados de entrada e perdendo sua capacidade de generalização. Já o conjunto para teste é usado para avaliar a capacidade de previsão da rede e permitir a comparação entre redes e métodos diversos. A necessidade do uso de conjuntos de verificação e teste em geral afeta a performance da rede, particularmente quando a série é constituída de poucas observações. A mensuração da performance em geral é feita com base na diferença entre o valor previsto pela rede e o valor desejado. Vários métodos são utilizados, por exemplo, a soma do quadrado das diferenças computadas para o conjunto de teste.

Finalmente, segundo Zhang e Qi (2005) particular atenção deve ser dispensada a questão da presença de efeitos sazonais e de tendência em séries temporais submetidas a redes neurais. Enquanto alguns autores consideram que RNAs são plenamente capazes de modelar estes efeitos diretamente, outros concluem exatamente o contrário. A análise empírica realizada pelos dois autores citados conclui pela efetividade do pré-processamento da série temporal com vistas à eliminação prévia e conjunta do efeito sazonal e de tendência.

A abordagem por redes neurais à série temporal considerada no presente texto não visou à busca de soluções consideradas ótimas. Foi utilizado o software Statistica Neural Networks 4.0B usando o recurso *Intelligent Problem Solver* (IPS). O período sazonal considerado foi de 12 observações, inclusive para a série com a sazonalidade removida, o que fixa o número de nós de entrada também em 12. Com auxílio deste recurso foram testadas várias arquiteturas de redes variando o número de nós intermediários e identificadas as redes candidatas com melhor desempenho com base no critério do aumento do erro do conjunto de verificação. Da mesma forma que relatado por Faraway e Chatfield (1998), o efeito da captura por mínimos locais é recorrente ao longo do procedimento. Isso conduz a múltiplas redes com mesmo número de nós intermediários, porém com erros de verificação diferentes em função dos pesos iniciais da rede.

Nos processamentos realizados o conjunto de verificação consistiu de 12 observações distribuídas ao longo do intervalo dos dados efetivamente usados no treinamento e o conjunto de teste foi constituído pelas últimas 12 observações da série. Os dados destes dois conjuntos não participam do treinamento. O emprego destes dois conjuntos implica que menos dados são usados no treinamento e, principalmente, o conjunto de verificação influi na determinação da “melhor” rede treinada. Além disso, essa “melhor” rede não é necessariamente a rede que fornece o melhor ajuste ao conjunto de teste. Para auxiliar no processo de escolha também foi analisada visualmente a distribuição dos erros de previsão ao longo de toda a série considerando que esses erros não deveriam estar concentrados em determinados trechos da série.

A tabela 3 relata os melhores resultados identificados. A coluna *transfor* indica a transformação prévia dos dados da série: dados originais (*ori*), eliminação de variabilidade e tendência (*varten*) e eliminação de variabilidade, tendência e sazonalidade (*vartensaz*). A coluna *nós* indica o número de nós intermediários da rede observando que todas as redes possuíam 12 nós de entrada e um nó de saída. A coluna correlação indica a correlação entre os dados originais e a previsão para o conjunto de treinamento (*tre*), verificação (*ver*) e teste

(*tes*). A coluna *REQM* indica a raiz do erro quadrático médio entre os dados originais e a previsão para os mesmos conjuntos. A coluna seguinte apresenta a somatória do quadrado dos desvios entre os dados originais do conjunto de teste e a previsão, servindo de comparativo com os outros métodos de previsão do presente texto. A coluna *n* relata o número de observações (melhor seria dizer dados de saída) efetivamente usadas no treinamento considerando a perda determinada pelos conjuntos de verificação e treinamento, a perda decorrente da transformação prévia dos dados e os dados “ignorados” do início da série (pois não constituem saídas).

TABELA 3 - RESULTADOS DOS MODELOS DE REDES NEURAIIS

<i>transfor</i>	<i>nós</i>	<i>correlação</i>			<i>REQM</i>			$\Sigma desv^2$	<i>n</i>
		<i>tre</i>	<i>ver</i>	<i>tes</i>	<i>Ter</i>	<i>ver</i>	<i>tes</i>		
ori	12	0.99	0.94	0.99	2.41	4.32	2.45	71.8	61
	12	0.96	0.99	-	4.39	1.37	-	22.5	73
varten	10	0.99	0.99	0.99	2.68	3.16	1.73	36.0	60
	12	0.99	0.99	-	2.73	1.17	-	16.4	72
vartensaz	12	0.96	0.98	0.99	8.33	2.90	2.15	55.2	48
	13	0.97	0.99	-	3.63	1.59	-	30.5	60

Fonte: Elaborado pelos autores

A tabela 3 relata as três transformações sugeridas pela bibliografia. O ajuste aos dados originais é considerada como satisfatória por vários autores o que se confirma no presente caso. A eliminação de tendência é considerada importante por outros e, aparentemente, ela produziu alguma melhoria. Finalmente outros autores sugerem a eliminação da variabilidade, tendência e sazonalidade o que aparentemente também se confirma em menor escala. Para cada transformação foram relacionadas duas redes: (1) a rede com o menor REQM de teste observada ao longo das tentativas; e (2) a rede com o menor EQM de verificação onde o conjunto de teste foi substituído pelo conjunto de verificação, ao final da série, de forma a prover mais 12 observações para o treinamento.

Da breve análise realizada observa-se que o uso da série original pode ser satisfatório, mas a eliminação de tendência parece ser a mais indicada, confirmando a bibliografia. A transformação mais ampla dos dados possivelmente também é indicada observando que é a situação na qual ocorre a maior perda de dados, o que é considerado relevante no caso, pois a série original já é curta. Em síntese todas as redes relatadas, e muitas outras identificadas ao longo dos testes, apontam para modelos comparativamente tão bons ou brevemente melhores que os métodos tradicionais.

TABELA 4 - COMPARAÇÃO ENTRE OS MODELOS

<i>Modelos</i>	$\Sigma desv^2$	<i>Classificação</i>
Alisamento exponencial aditivo	123.46	4
Alisamento exponencial multiplicativo	116.87	3
ARIMA	37.45	2
RNAs	16.40	1

Fonte: Elaborado pelos autores

5. CONCLUSÕES

Prever é uma das atividades mais importantes e desafiadores em todas as áreas do conhecimento humano. É importante porque previsões bem feitas e críveis permitem que se antecipem situações futuras de modo a preparar antecipadamente os sistemas para as respostas mais adequadas.

Na gestão das empresas e organizações os administradores ou os responsáveis pela elaboração e implantação de políticas estão permanentemente, tácita ou explicitamente, tentando perscrutar o futuro. Baseados em suas conjecturas ou certas decisões são tomadas com desdobramentos amplos ou restritos dependendo da natureza e da abrangência do que se prevê.

Por certo uma das áreas onde freqüentemente se realizam esforços no sentido de identificar o perfil da evolução futura é o conjunto de iniciativas abrangidas pela previsão de vendas. No varejo essa atividade é absolutamente essencial. Por se tratar de uma atividade de ligação entre a indústria e o consumidor final, a eficiência e a eficácia das operações comerciais dependem diretamente da capacidade do varejo ajustar suas compras e seus estoques em função do ritmo e perfil dos desejos de consumo.

Este artigo teve como objetivo precípua discutir comparativamente dois métodos quantitativos de previsão utilizando dados de vendas agregadas do mercado varejista brasileiro. De um lado os modelos decorrentes da econometria das séries temporais. De outro, as estruturas preditivas assentadas nas idéias suscitadas pelas redes neurais. A escolha dessas duas abordagens fundamentou-se em uma revisão da literatura que evidenciou a presença marcante dessas duas abordagens nos estudos quantitativos de previsão de vendas.

Tomando-se como medida comparativa a soma do quadrado dos resíduos gerados pelas diferenças entre as vendas previstas e as vendas realmente ocorridas, pode-se concluir que os modelos de séries chamados de alisamento mostraram-se substancialmente menos precisos que os modelos de séries temporais ARIMA e os modelos de Redes Neurais para

tratar a série agregada recente das vendas de varejo no Brasil. O mesmo não foi possível concluir em relação aos modelos ARIMA e Redes Neurais uma vez que a diferença entre esses modelos pode ser considerada pouco expressiva. Apesar das RNAs apresentarem resultados melhores, há de se observar que seu uso incorpora dificuldades, sendo a principal delas a necessidade de uso de conjuntos de verificação para interromper o processo de cálculo. Posto de forma mais geral, a questão é se a localização de boas RNAs pode prescindir de conhecimento prévio dos resultados de modelos ARIMA. Conforme salientam Faraway e Chatfield (1998), não há procedimento geral para tanto.

Este trabalho valeu-se de dados agregados. Seria interessante examinar o comportamento dos métodos aqui utilizados com informações desagregadas. Também seria interessante examinar séries que muito se desviam da linearidade, séries com maior volume de observações e o uso de outras estruturas de defasagem para as quais, possivelmente, o uso de RNAs pode envolver menos dificuldades e apresentar resultados melhores.

REFERÊNCIAS BIBLIOGRÁFICAS

- ANG, S.; O'CONNOR, M. The effect of group interaction processes on performance in time series extrapolation. **International Journal of Forecasting**, v.7, n.2, p. 141-149, 1991.
- BOX, G. E. P.; JENKINS, G. M. **Time Series Analysis: forecasting and control**. San Francisco: Holden Day, 1976.
- BANCO CENTRAL. **Séries compiladas pelo Banco Central do Brasil**. Disponível em: <<http://www.bcb.gov.br>>. Acesso em: 15 ago. 2008.
- BROCKHOFF, K. Group processes for forecasting. **European Journal of Operational Research**, v.13, n.2, p.115-127, 1983.
- CHU, C.; ZHANG, G. P. A comparative study of linear and nonlinear models for aggregate retail sales forecasting. **International Journal of Production Economics**, v.86, n.3, p.217-231, 2003.
- DE CHIARA, M. Com 86 milhões de pessoas, classe C já é maioria da população brasileira. **O Estado de São Paulo**, São Paulo, 27-03-2008, Caderno de Economia e Negócios. Disponível em: <<http://www.estado.com.br/editorias/2008/03/27/eco-1.93.4.20080327.31.1.xml>>. Acesso em: 30 abr. 2008.
- DE GOOIJER, J. G.; HYNDMAN, R. J. 25 years of time series forecasting. **International Journal of Forecasting**, v.22, n.3, p. 443-473, 2006.
- ENDERS, W. **Applied econometric time series**. New York: John Wiley & Sons, Inc., 1995.
- FARAWAY, J.; CHATFIELD, C. Time series forecasting with neural networks: a comparative study using the airline data. **Applied Statistics**, v.47, p.231-250, 1998.

FERRARI-FILHO, F.; DE PAULA, L. F. The legacy of the real plan and an alternative agenda for the Brazilian economy. **Investigación Económica**, v. 63, n. 244, p. 57-92, 2003.

FRANSES, P. H. Merging models and experts. **International Journal of Forecasting**, v.24, n.1, p.31-33, 2008.

GIAMBIASI, F. Estabilização, reformas e desequilíbrios macroeconômicos: os anos FHC (1995-2002). In: GIAMBIASI, F. e outros (Org.). **Economia brasileira contemporânea (1945-2004)**. Rio de Janeiro: Editora Campus, 2005, p.166-195.

HAYKIN, S. **Redes neurais: princípios e prática** (tradução). 2ª ed., Porto Alegre, Bookman, 900 p., 2001.

IPEADATA. **Séries compiladas pelo Instituto de Pesquisa Econômica e Aplicada**. Disponível em: <<http://www.ipeadata.gov.br>>. Acesso em: 14 ago. 2008.

PALMER, A.; MONTAÑO, J. J.; SESÉ, A. Designing an artificial neural network for forecasting tourism time series. **Tourism Management**, v.27, p.781-790, 2006.

SEGURA, J. V.; VERCHER, E. A spreadsheet modeling approach to the Holt-Winters optimal forecasting. **European Journal of Operational Research**, v.131, n.2, p.375-388, 2001.

SESSO FILHO, U. A. Crescimento e desempenho de redes de supermercado na década de 90. **Varejo Competitivo**, São Paulo, v.6, p.21-44, 2001.

SNIEZEK, J. A. An examination of group process in judgmental forecasting. **International Journal of Forecasting**, v.5, n.2, p.171-178, 1989.

WIDROW, B.; RUMELHART, D. E.; LEHR, M. A. Neural networks: applications in industry, business and science. **Communications of the ACM**, v.37, n.3, p.93-105, 1994.

ZHANG, G. P.; KLINE, D. M. Quarterly time series forecasting with neural networks. **IEEE Transactions on Neural Networks**, v.18, n.6, p.1800-1814, 2007.

ZHANG, G. P.; PATUWO, E.; HU, M. Y. Forecasting with artificial neural networks: the state of the art. **International Journal of Forecasting**, v.14, p.35-62, 1998.

ZHANG, G. P.; QI, M. Neural network forecasting for seasonal and trend time series. **European Journal of Operational Research**, v.160, p.501-514, 2005.