

# **Um Ensaio de Aplicação de Redes Neurais na Estimativa de Elasticidades de Funções Econômicas**

Sidney de Castro Oliveira<sup>α</sup>

João Sicsú<sup>δ</sup>

Adriano Joaquim de Oliveira Cruz<sup>φ</sup>

## **Resumo**

Este artigo apresenta um estudo sobre a eficácia e as limitações das redes neurais na aproximação de funções. Seu objetivo é avaliar a possibilidade de se estimar derivadas parciais de funções econômicas a partir destas redes. São apresentados aspectos metodológicos importantes que devem ser observados na definição da topologia e dos parâmetros das redes neurais a fim de que se possa atuar sobre a significância dos resultados. Conhecendo as derivadas parciais das funções econômicas é possível conhecer suas elasticidades, permitindo que se investigue o papel e o peso de cada uma das variáveis explicativas na composição da variável explicada, possibilitando uma análise e crítica de fenômenos econômicos. Um ensaio hipotético sobre a aplicabilidade das redes neurais na economia é apresentado ao final.

**Palavras-chave:** redes neurais, elasticidades de funções econômicas, inteligência computacional.

---

<sup>α</sup> Mestre em sistemas de computação e pesquisador do Núcleo de Computação Eletrônica da UFRJ.

<sup>δ</sup> Doutor em economia e professor adjunto do Instituto de Economia da UFRJ.

<sup>φ</sup> PhD em sistemas de computação, professor adjunto do Instituto de Matemática da UFRJ e pesquisador do NCE/UFRJ.

## 1 - Introdução

A ferramenta clássica utilizada para a análise quantitativa de relações econômicas é a econometria. Contudo, existem alternativas que também podem apresentar bons resultados. Uma delas são as redes neurais artificiais. Por terem a vantagem de deixar a maior parte do complexo trabalho de modelagem à cargo de exaustivas iterações computacionais, estas redes revelam-se uma ferramenta importante de auxílio à análise quantitativa de relações econômicas.

As redes neurais artificiais são um modelo computacional cujo paradigma de processamento da informação é inspirado no funcionamento do sistema nervoso cerebral. O ponto chave deste paradigma diz respeito a estrutura organizacional de tais redes. Elas são formadas por um conjunto de elementos processadores simplórios, chamados de “neurônios”, altamente interconectados entre si e que operam simultânea e conjuntamente para resolver problemas específicos. Assim como o cérebro, que armazena conhecimento em função das sinapses entre os neurônios, o comportamento da rede neural é determinado pela sua organização<sup>1</sup>. Dependendo do número de elementos processadores, do tipo de processamento inserido neles, da forma com que eles são interconectados e da importância (peso) de cada uma de suas interconexões, a rede neural define o seu comportamento.

Em geral, as redes neurais se adequam bem a problemas em que não se conhece uma solução algorítmica para eles ou a solução algorítmica é muito complexa para ser encontrada. O ponto forte é a capacidade de manuseio de dados imprecisos ou de explicação complicada, sendo útil na extração de significados difíceis de serem notados humanamente ou com técnicas computacionais tradicionais<sup>2</sup>.

Este artigo procura explorar a capacidade das redes neurais de modelar estes tipos de problemas, em especial aqueles em que se procura explicar fenômenos econômicos<sup>3</sup>. A idéia central é apresentar mecanismos capazes de explicar relações de interdependência entre as variáveis. Sabendo apenas que uma variável é explicada em função de determinadas outras, chamadas explicativas, busca-se estimar o quanto a variável explicada é sensível às variações

---

<sup>1</sup> O comportamento da rede neural representa sua função de transferência, ou seja, a forma com que ela responde (saída) às diversas configurações da entrada.

<sup>2</sup> Técnicas tradicionais são técnicas algorítmicas, onde o resultado computacional é previsível e decorrente de uma sequência conhecida de procedimentos.

<sup>3</sup> Os experimentos deste artigo se restringiram a um tipo de rede neural artificial conhecido como MLP (*multilayer perceptron*) que, por suas características, são capazes de tratar dados não linearmente separáveis.

das explicativas. Em outras palavras, o objetivo é buscar, através do uso de redes neurais, estimativas para as sensibilidades das variáveis explicativas em relação à variável explicada nas relações econômicas, permitindo que se investigue o papel e o peso de cada uma delas na composição do problema. Espera-se, desta forma, contribuir para a análise quantitativa dos fenômenos econômicos, abrindo espaço para que se realizem ensaios e estudos que objetivam um maior entendimento da economia real.

Cabe ressaltar que as redes neurais são um recurso essencialmente computacional. Envolvem paradigmas e abstrações próprios da ciência da computação, cujo entendimento, apesar de ser condição importante para que se alcance bons resultados na sua utilização, não é fundamental para que se entendam os aspectos metodológicos aqui apresentados. Desta forma, não há, neste artigo, o compromisso de se apresentar os fundamentos das redes neurais, a descrição de seu funcionamento, suas configurações, variações típicas e, muito menos, a modelagem matemática associada a elas. São assuntos que encontram uma vasta abordagem na literatura (como por exemplo em BRAGA *et. al.*, 1998; HYKIN, 2000; ZURADA, 1992), inclusive com diversas aplicações em problemas econômicos (por exemplo em DIAZ & ARAÚJO, 1998; SILVA *et. al.*, 2001). Em relação às redes neurais, este artigo procura se ater apenas aos aspectos metodológicos envolvendo sua *sintonia*<sup>4</sup>, buscando expor suas potencialidades na tentativa de suplantiar os desafios e demandas envolvendo as áreas da economia e da computação.

A exposição das idéias no artigo está dividida em duas partes principais. A primeira delas apresenta o equacionamento do problema da busca de estimativas para as derivadas parciais de funções econômicas, adequando-o às características do modelo neural de computação. Através de alguns testes de validação são identificados certos aspectos metodológicos necessários ao ajustamento destas redes a fim de que elas respondam satisfatoriamente ao problema. Estes testes revelam tanto a potencialidade do modelo empregado, quanto as restrições e limitações que devem ser observadas para se alcançar os resultados.

Na segunda parte, procura-se exemplificar a aplicabilidade do modelo neural na economia, apresentando um problema hipotético e analisando o comportamento das variáveis econômicas envolvidas. Mais especificamente, procura-se, através das redes neurais,

---

<sup>4</sup> A sintonia da rede neural artificial é o ajuste de sua topologia e de todos os seus parâmetros a fim de que ela responda, satisfatoriamente, às demandas do problema.

exemplificar sua utilização na explicação de uma variável dependente em função de suas variáveis explicativas em um problema macroeconômico típico.

## **2 - Aproximação de Funções e o Peso das Variáveis Explicativas**

Em economia se busca, muitas vezes, analisar certos fenômenos procurando uma associação entre um conjunto de variáveis que se supõe, teoricamente, serem explicativas de determinada relação de interdependência com uma variável explicada. Mesmo quando a teoria econômica oferece muitos subsídios para a identificação destas variáveis (explicativas e explicada), dado um problema real, é difícil, ao mesmo tempo que é necessário, dimensionar o quanto cada uma das variáveis explica ou contribui para a relação sob análise, ou seja, qual o grau de influência de cada variável explicativa sobre a variável explicada.

Como se sabe, avaliar o papel e o peso das variáveis explicativas na composição da variável explicada é, na verdade, uma busca por uma função que relaciona as variáveis envolvidas, se abstraindo da relação temporal entre elas. Pode-se dizer que esta função é um mapeamento das variáveis explicativas na explicada, cuja representação gráfica, no espaço euclidiano, é uma superfície em  $\mathbf{R}^n$  para  $n-1$  variáveis explicativas. As características desta superfície representam o comportamento da variável explicada frente ao universo de combinações das variáveis explicativas.

Uma vez conhecida a superfície, é possível saber o quanto cada variável explicativa contribui para a variável explicada, já que a taxa de variação da superfície na direção de cada eixo do plano representa o quanto a variável explicada é sensível às variações da variável explicativa correspondente. Estas taxas de variação em cada ponto da superfície são conhecidas como sensibilidades da variável explicada em relação às variáveis explicativas, que equivalem, matematicamente, às derivadas parciais da função.

Mas como encontrar esta superfície? A maneira mais natural seria procurar observar um conjunto de pontos que supostamente pertence a superfície e buscar um aprendizado do comportamento do fenômeno econômico a fim de deduzir a relação de interdependência entre as variáveis. Mas esta não é uma tarefa fácil. Quanto mais variáveis estiverem envolvidas e quanto mais complexa e não linear for a relação de interdependência entre elas, mais difícil é a dedução de qualquer regra de formação entre os pontos. Uma forma de facilitar as ações é

admitir algum erro na dedução da interdependência, o que reduz a tarefa à busca de uma superfície aproximada que explique, satisfatoriamente, a relação entre as variáveis. Dependendo do grau desta aproximação, as derivadas parciais da função assumem uma maior ou menor significância. Quanto mais informação se tem da superfície, ou seja, quanto mais e melhor distribuídos são os pontos conhecidos, maior a representatividade da superfície expressa pelo problema e maior a confiabilidade das conclusões extraídas.

É fácil perceber que o caminho de busca da solução pela observação exaustiva do comportamento das variáveis envolvidas pode ser fortemente facilitado pela utilização de instrumentos automáticos de reconhecimento das interdependências entre as variáveis. E, as redes neurais, pelas suas características, se adequam bem às necessidades de tais instrumentos. Em um processo conhecido como *aprendizagem*<sup>5</sup> as redes neurais procuram extrair, a partir dos exemplos que lhes são apresentados, características generalizáveis que relacionam a variável explicada (saída da rede) ao comportamento das variáveis explicativas (entrada da rede). Com isso, adquirem uma capacidade de responder consistentemente a *estímulos*<sup>6</sup>, tanto os usados no seu treinamento quanto a outros, desde que estes outros sejam relativamente assemelhados àqueles utilizados no treinamento.

Esta capacidade de responder a estímulos desconhecidos revela uma característica importante das redes neurais, a predição, que encontra diversas aplicações. Embora a mais popular delas, na área econômica, seja a previsão de séries temporais, onde se procura descobrir a vizinhança futura de novas ocorrências de uma seqüência conhecida de pontos, este não é o foco do presente artigo. O objetivo é explorar a predição visando um outro tipo de aplicação, que é a aproximação de funções econômicas e a estimativa de suas sensibilidades.

A partir de um conjunto suficiente de pontos que se sabe pertencer à função, muitas vezes expresso por séries temporais, pode-se treinar uma rede neural para que ela seja representativa de uma superfície que expresse o seu comportamento, relacionando as variáveis explicativas com a explicada. Uma vez treinada a rede, é possível predizer os valores da função nas vizinhanças de cada ponto utilizado no treinamento, abrindo espaço para que se estime o valor das derivadas parciais da função nestes pontos. Através de métodos numéricos simples é

---

<sup>5</sup> A aprendizagem das redes neurais é uma abstração que está associada a um processo computacional iterativo, chamado treinamento, que resulta na convergência de seus parâmetros para valores que melhor se ajustam às condições de contorno impostas ao problema.

<sup>6</sup> Estímulos são padrões válidos de entrada da rede, ou seja, configurações válidas para as variáveis explicativas.

possível chegar a estas estimativas, usando o próprio conceito matemático de derivadas parciais no ponto, tal como expresso pela fórmula a seguir (LIMA, 1995):

$$\frac{\partial f}{\partial x_i}(a) = \lim_{t \rightarrow 0} \frac{f(a + t.e_i) - f(a)}{t}, \quad \text{<equação 1>}$$

onde o componente  $e_i$  é um vetor unitário representativo do eixo da variável  $x_i$ .

Cabe ressaltar que, diferentemente da econometria, o modelo neural não resulta em uma expressão analítica de aproximação da função, isto é, não se tem como resposta parâmetros que possam compor uma fórmula matemática que relacione as variáveis envolvidas. O que se obtém é uma estrutura comportamental das características do conjunto de pontos apresentados no treinamento. Desta forma, uma vez treinada, a rede neural pode ser encarada, do ponto de vista operacional, como sendo uma *caixa-preta* que tem como entrada as variáveis explicativas e como saída a variável explicada. Não sendo possível deduzir nenhuma relação de interdependência entre as variáveis sem que se exponha a rede a uma seqüência de estímulos provenientes das variáveis explicativas e verificando os valores obtidos como resposta para a variável explicada.

Também é importante frisar que os parâmetros da rede neural não têm qualquer relação com o papel e o peso de cada variável envolvida. São parâmetros estritamente computacionais, que determinam exclusivamente o comportamento da rede. A partir deles não se deduz qualquer relação entre a variável explicada e as explicativas.

A primeira questão que surge quando se procura utilizar redes neurais para aproximar funções econômicas a fim de se estimar suas sensibilidades é saber se elas são realmente apropriadas para isto, ou seja, se a superfície gerada pela rede, após o treinamento, carrega características que permitem estimar suas derivadas parciais pela equação 1, cujos resultados sejam consistentes. Mas a resposta esbarra logo numa dificuldade: como avaliar a qualidade dos resultados se não se conhece nem a função nem os corretos valores das derivadas?

De fato, o grau de desconhecimento a respeito dos aspectos analíticos do problema impediria uma análise sobre a qualidade das estimativas das derivadas parciais. Contudo, é possível encontrar indicativos que permitam inferir se os resultados apresentados pelas redes neurais, apesar de sujeitos a imprecisões, podem representar uma boa estimativa daquilo que seria o verdadeiro.

Um artifício que pode ser utilizado para se colher tais indicativos é testar a técnica oferecida pelas redes neurais em problemas que sejam inteiramente conhecidos. Conhecendo a função de interdependência e as variáveis envolvidas é possível o confronto dos resultados apresentados pelo modelo neural com os valores verdadeiros, permitindo, portanto, uma avaliação da qualidade dos resultados. Desta forma, estabeleceu-se uma função hipotética qualquer (a equação 2 apresentada a seguir) para definir o comportamento de uma variável explicada ( $z$ ) para um problema com duas variáveis explicativas ( $x$  e  $y$ ):

$$z = f(x, y) = \sin(x) + \frac{y^2}{3} - y + 3. \quad \text{<equação 2>}$$

Partindo de duas séries temporais também hipotéticas, escolhidas para serem as variáveis explicativas, gerou-se, segundo a função estabelecida, um conjunto de pontos para ser a variável explicada do problema, conforme ilustrado na figura 1. Como se conhece, de acordo com as hipóteses adotadas, a expressão analítica da função, também se conhece as expressões analíticas de suas derivadas parciais, o que permite saber o real valor destas derivadas nos pontos de treinamento da rede neural. Então, agora, é possível comparar as estimativas geradas pelas redes neurais para estas derivadas com seus valores verdadeiros, avaliando a capacidade do modelo neural na aproximação de funções e de suas respectivas sensibilidades.

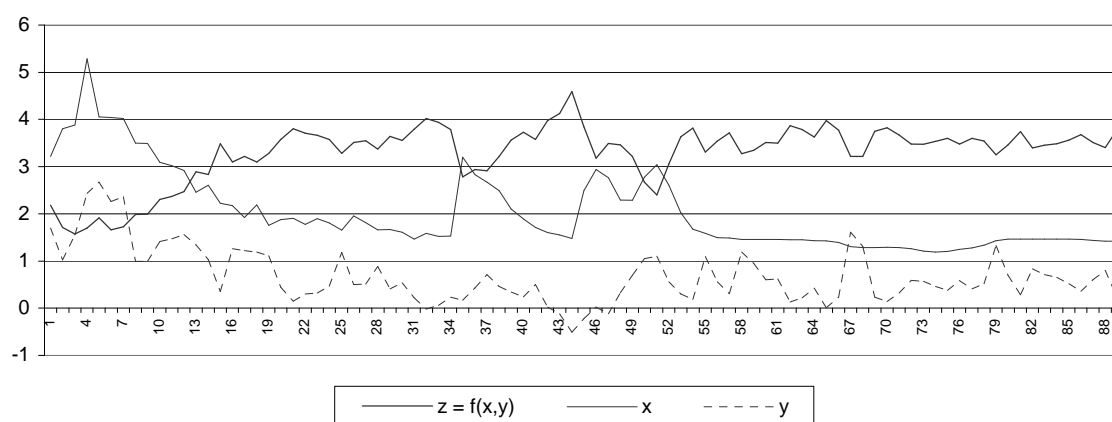


FIGURA 1: Séries temporais das variáveis da função estabelecida.

O ponto chave na busca de bons resultados na utilização de redes neurais é a sintonia da rede, que consiste na definição de sua topologia e de seus parâmetros. Como não há valores ótimos conhecidos para esta sintonia, eles precisam ser ajustados caso a caso em função das características do problema em questão, resultando em diversas possibilidades de modelagem.

Desta maneira, sua utilização em um problema inteiramente conhecido como forma de avaliar o modelo neural também tem a vantagem de explicitar aspectos metodológicos no processo de sintonia, já que passa a ser possível avaliar o impacto de alterações na organização da rede na busca de melhores resultados.

Partindo da figura 1, o objetivo é definir aspectos metodológicos para sintonizar uma rede neural que seja capaz de atender satisfatoriamente a duas condições fundamentais para o bom equacionamento do problema (que consiste em encontrar a relação que mapeia as variáveis explicativas na explicada). A primeira delas (condição 1) é que a rede neural, uma vez estimulada pelas séries das variáveis explicativas  $x$  e  $y$ , deve responder o mais próximo possível da série temporal da variável explicada  $z$ . A segunda condição (condição 2) é estabelecida em relação às derivadas parciais da equação 2. Ao se estimular a rede neural com valores na vizinhança de cada um dos pontos das séries temporais das variáveis  $x$  e  $y$ , ela deve fornecer respostas que, uma vez usadas segundo a equação 1, produzam resultados próximos dos verdadeiros, que podem ser calculados matematicamente. Entretanto, como as derivadas parciais em cada ponto não são informações passadas à rede na fase de treinamento, esta condição representa, na prática, um outro fator qualitativo para o cumprimento da condição 1.

A combinação destas duas condições define as restrições que devem ser obedecidas na busca de uma superfície que expresse a função que relaciona as variáveis explicativas e explicada. E, como esta superfície é estabelecida de forma aproximada, chamaremos de *aproximação de funções* o equacionamento do problema segundo as condições citadas.

Desta forma, a sintonia da rede neural deve ser guiada observando-se o ajuste tanto em relação aos pontos apresentados para treinamento (condição 1), quanto em relação às suas derivadas parciais nestes pontos (condição 2). Um bom cumprimento da primeira condição não significa, necessariamente, um bom cumprimento da segunda, ou seja, uma excelente aproximação em relação aos pontos usados no treinamento pode resultar em estimativas discrepantes em relação às derivadas. A princípio isto pode parecer contraditório, mas é preciso lembrar que as informações apresentadas à rede durante o treinamento são pontuais, esparsas no domínio da função e desassociadas das derivadas parciais em cada ponto. Sendo assim, a rede neural tem liberdade para moldar a superfície, de acordo com suas necessidades, observando apenas a condição 1 durante o treinamento. E, dependendo da forma com que a rede molda a superfície, pode-se obter estimativas diferentes para as derivadas parciais a cada treinamento distinto, o que é tão indesejável quanto inevitável, frente às propriedades da rede neural.



Apesar desta flutuação das respostas ser intrínseco ao modelo neural, é possível encontrar caminhos que podem minimizar esta questão. Basicamente, o que se deve buscar, dada a complexidade do problema, é compatibilizar a *densidade de informação* oferecida a rede durante o treinamento com sua estrutura organizacional. Esta densidade é a razão entre a quantidade de pontos utilizados no treinamento e a quantidade de parâmetros a serem determinados.

Quando a topologia da rede está super dimensionada, isto é, o número de “neurônios” na sua *camada escondida*<sup>7</sup> está excessivo frente a quantidade (e também complexidade) de pontos apresentados para treinamento, haverá muitos parâmetros a serem determinados com dados que podem ser insuficientes para tal. Vale lembrar que cada “neurônio” da rede se interconecta aos demais com ligações cujos pesos precisam ser determinados a partir dos dados. Esta baixa densidade de informações pode fazer com que o processo de treinamento tenha pouca qualidade, conduzindo a resultados frágeis e inconstantes para o comportamento da rede. Como resultado tem-se uma superfície mais volátil, mais nervosa, com maiores flutuações para a função aproximada, o que resulta em estimativas irregulares e inconstantes para as derivadas parciais.

Por outro lado, aumentar a quantidade de informação sobre a superfície, através de um maior número de pontos, aumenta os compromissos do treinamento, tornando-o mais complexo. Se essa maior complexidade não for acompanhada por uma topologia com mais “neurônios” e, conseqüentemente, mais poderosa, a superfície resultante tende a ser mais suave, mais alisada, acompanhando mais as tendências do que os pontos. Este menor grau de aproximação em relação aos pontos de treinamento resulta em um maior *erro médio quadrático*<sup>8</sup> para o ajuste.

Não há formulas conhecidas para se definir a melhor densidade de informações a ser utilizada no treinamento das redes neurais. Para cada situação específica deve-se buscar um equilíbrio entre a quantidade de pontos utilizada no treinamento, o erro médio aceitável e a topologia da rede (especialmente a quantidade de camadas intermediárias e o número de

---

<sup>7</sup> A rede neural tem uma camada de entrada, que recebe as variáveis explicativas; uma camada de saída, por onde as variáveis explicadas são lidas e uma (ou mais) camada intermediária, chamada escondida, responsável pelo caráter não linear do modelo. Teoricamente, redes com uma camada intermediária podem implementar qualquer função contínua, e a utilização de duas camadas intermediárias permite a aproximação de qualquer função (BRAGA *et. al.*, 1998).

<sup>8</sup> O erro médio quadrático do ajuste é medido pela média do quadrado da diferença entre os valores apresentados como resposta pela rede neural e os valores alvo quando a rede é estimulada pelos pontos de treinamento das variáveis explicativas.

“neurônios” nestas camadas). Este equilíbrio é o principal aspecto metodológico na sintonia das redes neurais para uma boa aproximação de funções.

Apesar do processo de busca deste equilíbrio ter um forte componente empírico, é possível guiá-lo, fazendo da qualidade da estimativa da derivada uma importante condição de contorno a ser perseguida para o equacionamento do problema. Entretanto, ao contrário da função, onde ao menos se conhecem alguns pontos pertencentes à superfície que a representa, nada se sabe a respeito das derivadas. A única informação geralmente disponível é que elas existem e são únicas para cada ponto da superfície. É justamente no sentido de se convergir para esta propriedade da unicidade das derivadas que se deve balizar a sintonia da rede. Equilibrando, de um lado, a acurácia dos resultados estimados para as derivadas parciais da função, e do outro, um erro médio quadrático aceitável para a aproximação da superfície em relação ao conjunto de pontos usados no treinamento.

De modo geral, deve-se buscar este equilíbrio utilizando o menor número possível de “neurônios” na topologia da rede. Uma consequência disto é que a restrição que se impõe ao número de “neurônios” da rede, a fim de se evitar superfícies artificialmente nervosas na aproximação, pode também impossibilitar que a rede expresse características que sejam próprias da função. Isto pode ser facilmente percebido pelo fato do referido equilíbrio ser estabelecido em razão do comportamento médio da função. Nas regiões do domínio onde a função é naturalmente mais volátil, a rede neural tende a atenuar a volatilidade real, podendo resultar em superfícies cujas taxas de variação estejam aquém do desejado. Esta limitação às variações súbitas da função é uma restrição que deve ser entendida e considerada na avaliação dos resultados.

Cabe, neste momento, esclarecer melhor uma dificuldade que é oriunda de uma característica intrínseca das redes neurais. A busca que a rede neural faz da superfície que melhor representa a função a ser aproximada é uma busca aleatória em um espaço infinito de possibilidades. Desta forma, a cada vez que se treina uma rede neural, ainda que se utilizem *os mesmos dados e parâmetros*<sup>9</sup> e que a rede esteja bem sintonizada, obtém-se uma superfície diferente como resposta. Isto incorpora um caráter probabilístico às aproximações geradas pelas redes. Evidentemente, superfícies diferentes geram estimativas diferentes para suas derivadas.

---

<sup>9</sup> Supondo que cada treinamento parte de valores aleatórios para a configuração inicial dos pesos das interligações entre os “neurônios” da rede. Caso contrário a rede apresenta resultados que se repetem a cada treinamento se forem utilizados os mesmos dados e parâmetros.

Estas diferenças podem ser até bastante acentuadas dependendo da sintonia estabelecida para a rede. Sendo assim, quando se diz que é esperado acurácia nas estimativas das derivadas da função não significa esperar que seus valores se repitam a cada treinamento distinto, mas sim esperar que elas tenham uma distribuição de frequência que seja unimodal e que tenha uma curtose o mais pontiaguda possível (leptocúrtica), conforme ilustrado na figura 2. Desta forma, é possível identificar uma medida de tendência central (média, moda ou mediana) que possa representar, com relativa significância, o valor da resposta da rede neural para cada ponto do treinamento.

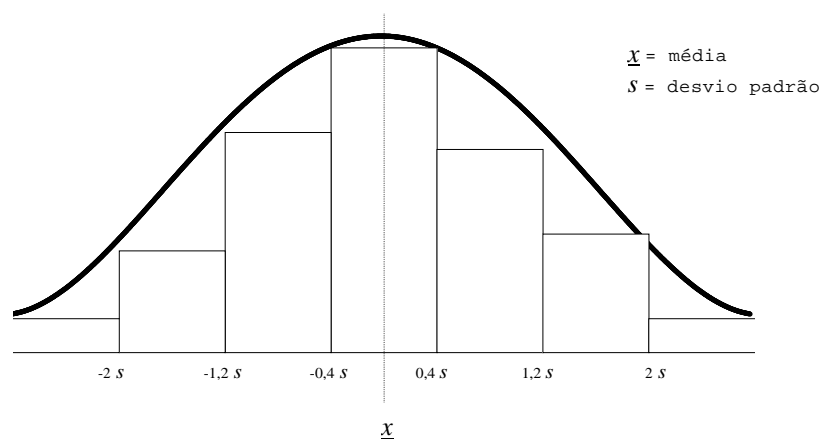


FIGURA 2: Típico histograma esperado para as respostas das redes neurais quando estas estão bem sintonizadas: uma distribuição unimodal e leptocúrtica.

Pode-se perceber que este caráter probabilístico do modelo neural confere pouca significância às estimativas de derivadas que sejam oriundas de uma única rede treinada. É necessário buscar maior representatividade para as estimativas, o que pode ser conseguido através de uma média de diversos valores provenientes de redes neurais que, ainda que possuam mesma topologia e parâmetros, tenham treinamento distintos para o mesmo conjunto de dados. Sendo assim, pode-se apresentar uma única resposta que seja resultante da unificação das diversas estimativas em direção aos valores mais prováveis para as derivadas da função segundo a distribuição de frequência dos valores apresentados pela rede.

Partindo dos dados apresentados na figura 1 e seguindo os critérios estabelecidos para guiar a sintonia da rede neural, efetuou-se inúmeros treinamentos na busca da melhor topologia

para a rede<sup>10</sup> - sempre procurando equilibrar uma aproximação em relação aos pontos apresentados com uma relativa acurácia para as estimativas das derivadas nestes pontos.

Como era esperado, estas estimativas apresentaram flutuações em diversas faixas de valores para cada topologia testada. Entretanto, observou-se que em algumas topologias elas se concentraram em faixas mais estreitas que noutras, apresentando, portanto, resultados mais acurados. Mesmo não significando precisão, já que as faixas mais freqüentes podem não incluir as derivadas verdadeiras, esta maior acurácia é sugestiva de melhores resultados, pelo menos em relação à sintonia da rede neural para o problema. Quanto mais pontiaguda for a curtose das distribuições de freqüência nos pontos de treinamento, melhor sintonizada está a rede para estimar as derivadas, observado o erro aceitável para a aproximação da função em relação a estes pontos.

A fim de conhecer as características da distribuição de freqüência das estimativas das derivadas, construiu-se, utilizando a topologia da rede de melhor sintonia<sup>11</sup>, um histograma para cada ponto do conjunto de treinamento. Na observação dos resultados, pôde-se verificar que as distribuições de freqüência eram em forma de sino e com bastante simetria para uma grande parte deles, apesar da heterocedasticidade dos dados<sup>12</sup>. Sendo assim, mesmo que o universo dos valores possíveis para as derivadas seja desconhecido, pode-se supor que as redes neurais, uma vez bem sintonizadas, geram amostras de dados cujos parâmetros estatísticos podem ser calculados, de forma aproximada, através das propriedades de uma distribuição quase-normal de dados.

A figura 3 ilustra os resultados alcançados para as estimativas das derivadas da função 1. Uma comparação entre as derivadas parciais verdadeiras da função, calculadas analiticamente, e os valores médios estimados para elas pelas redes neurais revela uma boa consistência entre eles: as curvas são semelhantes, seguem as mesmas oscilações e flutuam dentro da mesma grandeza de valores.

---

<sup>10</sup> Todos os experimentos aqui apresentados foram realizados com o software MatLab acrescido do toolbox de redes neurais do mesmo fabricante.

<sup>11</sup> A sintonia da rede foi alcançada com uma rede neural MLP com uma única camada escondida de 3 neurônios. A função de ativação da camada escondida foi a tangente-sigmoide, e a da camada de saída foi a puramente linear. O erro médio admitido para a aproximação da função nos pontos de treinamento foi menor que 1%, com ponto de parada em, no máximo, 100 iterações. As variáveis foram normalizadas para o intervalo [-1,1] por uma transformação linear, cujo impacto nas derivadas foi restabelecido após o treinamento.

<sup>12</sup> A variância das estimativas das derivadas parciais da função em cada ponto da série não foi constante. Certas regiões apresentaram maiores flutuações de resultados que outras.

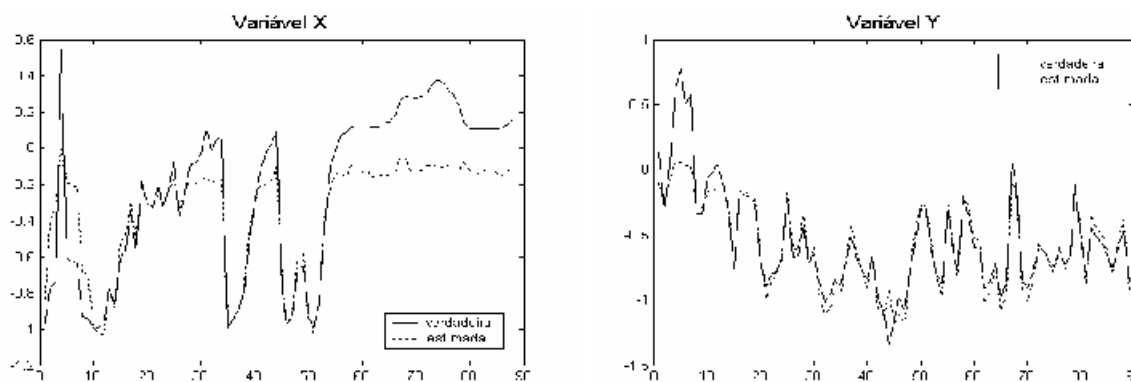


FIGURA 3: Comparativo entre as derivadas parciais verdadeiras e as estimadas pela rede neural para a função estabelecida.

Apesar disto, pode-se observar que em certos pontos, mais do que em outros, ocorreram maiores diferenças entre a derivada estimada pela rede e as derivadas verdadeiras, o que sugere a existência de algum erro sistemático da rede na moldagem da superfície nestes pontos. As explicações para isto nos remetem para as considerações matemáticas de modelagem das redes neurais, cujas investigações fogem do escopo deste artigo. O importante, da leitura que se faz dos resultados apresentados, é que é possível inferir que as redes neurais, ao aproximarem funções, o fazem de modo a permitir que se extraia estimativas das derivadas parciais destas funções. Estimativas estas que, apesar de carregar alguma imprecisão, podem oferecer uma boa aproximação de seus valores reais.

Cabe ressaltar ainda que, quanto mais complexo e não-linear for o comportamento da variável explicada em relação às variáveis explicativas do problema, maior a potencialidade das redes neurais, já que o esforço de modelagem é majoritariamente computacional. Mesmo porque, como este modelo não exige qualquer suposição sobre o comportamento das variáveis envolvidas, suas vantagens se acentuam a medida em que o problema ganha complexidade.

### 3 - Um Ensaio de Aplicabilidade

A metodologia e os resultados apresentados anteriormente abrem espaço para que se avalie a utilidade das redes neurais na explicação de fenômenos econômicos, permitindo que se investigue o papel e o peso de cada uma das variáveis, que se supõe serem explicativas, no comportamento da variável explicada. A fim de demonstrar estas possibilidades é apresentado uma análise de caso hipotético em que se aplica o modelo neural numa questão macroeconômica típica.

Uma variável importante em economia é a taxa de desemprego que, por conta disto, é objeto de muitos estudos. Vamos supor, nesta análise, que o desemprego seja explicado, no curto prazo, por uma função que contém simplesmente a taxa de juro nominal e o déficit público relativo ao PIB<sup>13</sup>. Vamos supor ainda que estas três variáveis se relacionem conforme as séries temporais mostradas na figura 4.

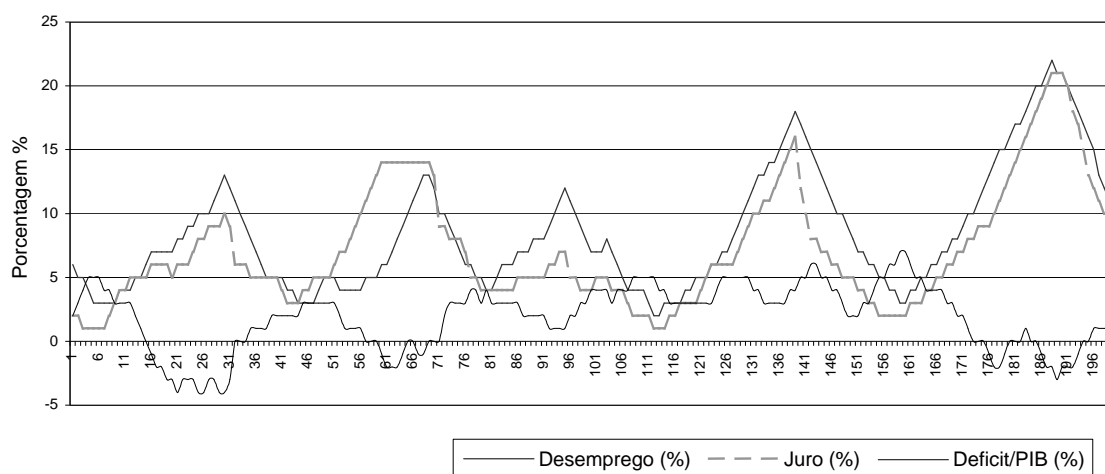


FIGURA 4: Séries temporais das variáveis do ensaio hipotético, onde se procura explicar o desemprego, no curto prazo, como função da taxa de juro nominal e do déficit público relativo ao PIB.

Como a taxa de juro nominal e o déficit público/PIB são variáveis administradas, é importante, do ponto de vista econômico, saber como o desemprego se relacionou com elas, ao longo do tempo, no período em questão. Desta forma, será possível identificar os momentos em que a variabilidade do desemprego foi explicada, predominantemente, pelo juro e momentos em que foram as variações da política fiscal de gastos que mais influíram na variação da taxa de desemprego. Confrontando estas informações com o comportamento das variáveis administradas, taxa de juro nominal e déficit/PIB, pode-se expor as políticas de controle destas variáveis à avaliações qualitativas e quantitativas.

<sup>13</sup> O desemprego no presente caso pode ser descrito por um simples modelo baseado nos escritos de Keynes e/ou nos desenvolvimentos pós-keynesianos. Para essa corrente teórica, uma política monetária expansionista de redução dos juros provoca uma variação positiva da renda e do nível de emprego. Ademais, consideram que um aumento dos gastos do governo (dada uma certa arrecadação), o que causa um aumento do déficit em relação ao PIB, implica uma variação positiva da renda e do nível de emprego. Para maiores detalhes sobre essa relação entre políticas econômicas ativas e nível de emprego, ver por exemplo, Cardim de Carvalho (1992, cap.12), Sicsú (1997) e Sicsú (2001).

Seguindo os aspectos metodológicos apresentados anteriormente para sintonia das redes neurais, ajustou-se<sup>14</sup> a topologia e os parâmetros da rede de forma a ela ser capaz de aproximar satisfatoriamente uma superfície, a partir do conjunto de pontos pertencentes à função buscando um equilíbrio entre duas condições básicas:

- O comportamento da rede, quando comparado à variável explicada nos pontos de treinamento, deve obedecer a um determinado erro médio quadrático mínimo estipulado como aceitável;
- Sucessivos treinamentos da rede sintonizada devem resultar em um conjunto de estimativas para as derivadas parciais, nos pontos de treinamento, que possuam distribuições de frequência unimodais e o mais leptocúrticas possíveis, permitindo que obtenha o melhor valor esperado para as derivadas da função com *relativa significância*<sup>15</sup>.

Uma vez sintonizada a rede, foram coletados dados oriundos de *dez treinamentos*<sup>16</sup> aleatórios. A figura 5 mostra o comportamento médio da rede em relação à variável explicada quando estimulada pelas devidas variáveis explicativas. Percebe-se que as flutuações deste comportamento são mínimas, resultando em uma pequena variância para o conjunto de dados. De certa forma, este resultado já era esperado, visto que os pontos do gráfico são os mesmos utilizados no treinamento, que atuam como âncoras no processo de aproximação da função.

---

<sup>14</sup> A sintonia da rede foi alcançada com uma rede neural MLP com uma única camada escondida de 5 neurônios. A função de ativação da camada escondida foi a tangente-sigmóide, e a da camada de saída foi a puramente linear. O erro médio quadrático admitido para a aproximação da função nos pontos de treinamento foi menor que 2% com ponto de parada em, no máximo, 500 iterações. As variáveis foram normalizadas para o intervalo [-1,1] por uma transformação linear, cujo impacto nas derivadas foi restabelecido após o treinamento.

<sup>15</sup> Só é possível conhecer a significância do resultado conhecendo a forma de distribuição dos dados. Valores aproximados podem ser obtidos supondo que os dados obedecem certas distribuições conhecidas.

<sup>16</sup> Do ponto de vista estatístico, quanto maior o número de amostras, mais significativo é o resultado. Mas devido ao caráter apenas ilustrativo desta análise, foram coletados dados de poucas amostras (dez). Acredita-se que uma quantidade de trinta amostras se mostre suficiente.

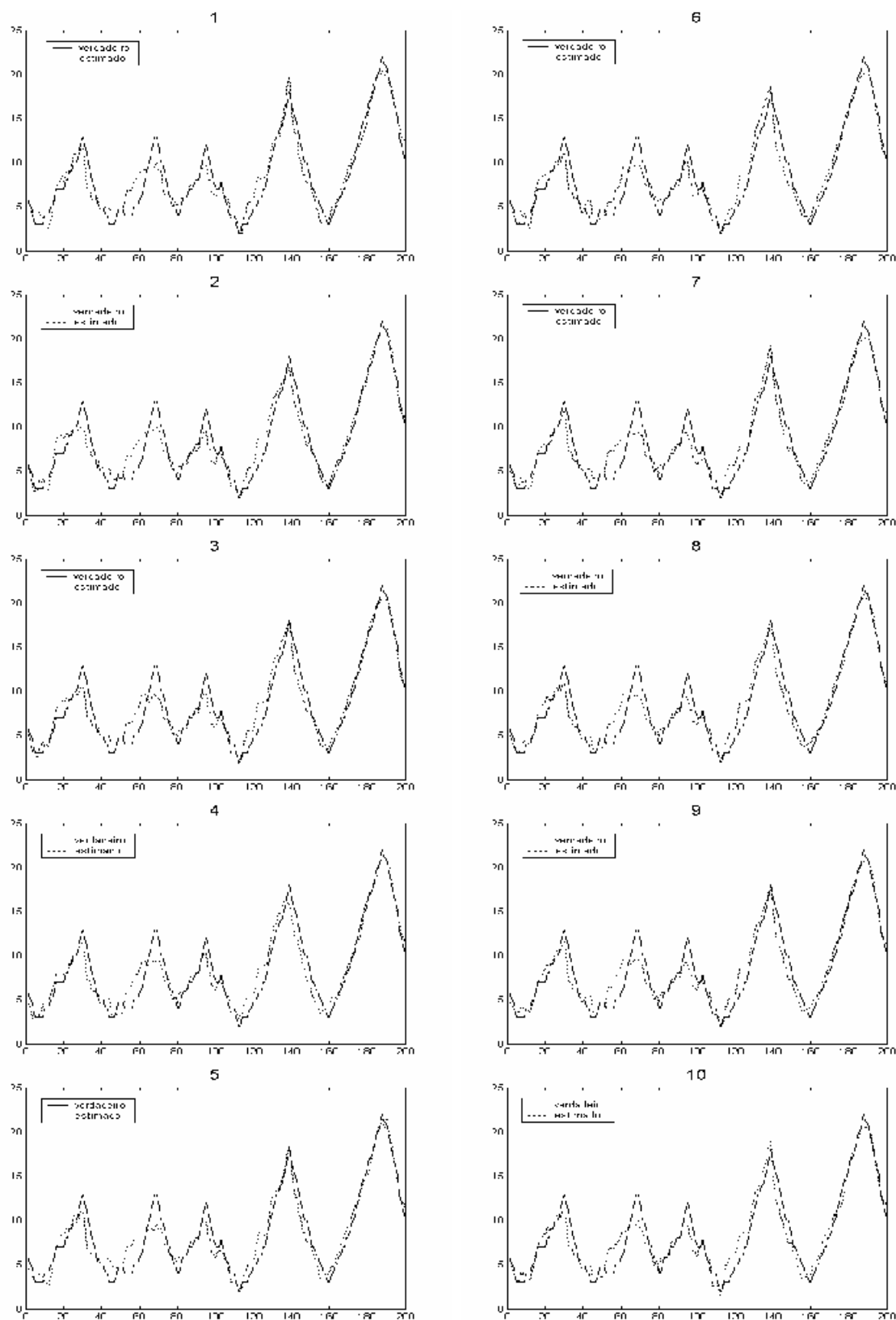


FIGURA 5: Amostras do comportamento da rede neural em relação à variável explicada, a partir de dados oriundos de sucessivos treinamentos da rede neural sintonizada.



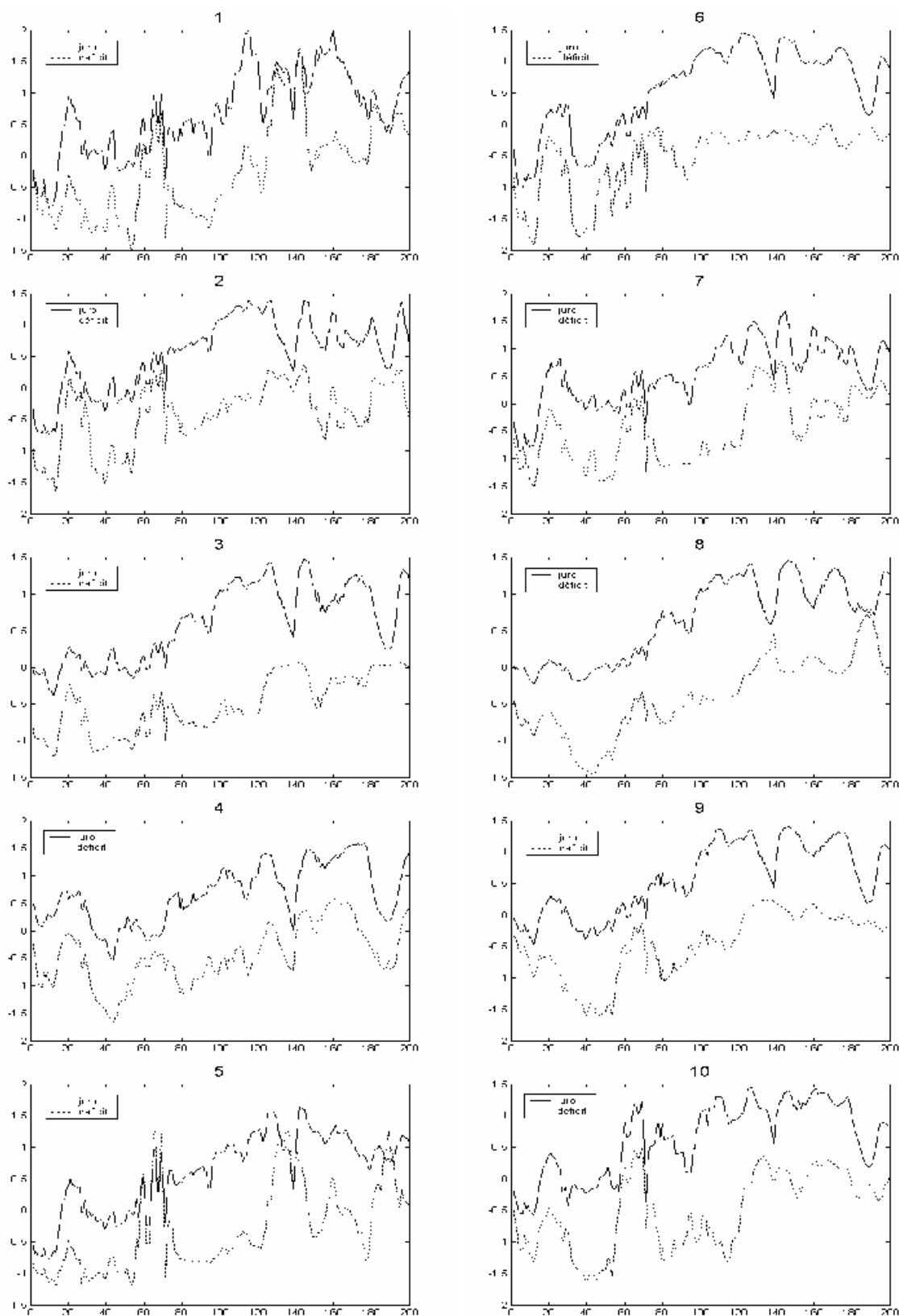


FIGURA 6: Amostras de derivadas parciais estimadas por métodos numéricos, a partir de dados oriundos de sucessivos treinamentos da rede neural sintonizada.

Entretanto, o mesmo já não se pode dizer das estimativas das derivadas parciais da função. Suas flutuações são bem mais significativas, apesar de apresentarem tendências parecidas, conforme mostrado na figura 6. Esta maior variância é decorrente da liberdade oferecida à rede neural nos pontos distintos daqueles usados no treinamento, que incluem as vizinhanças destes pontos, utilizadas no método numérico para estimar as derivadas parciais da função. Observando a figura 6, fica claro a importância de se buscar uma medida de tendência central, dentre as amostras coletadas, a fim de que se possa expressar com maior representatividade os resultados alcançados.

Do ponto de vista econômico, as derivadas parciais da função aproximada representam suas sensibilidades, tanto da taxa de juro nominal quanto do déficit público/PIB em relação ao desemprego, cujas estimativas médias são mostradas na figura 7, onde a média aritmética foi a medida de tendência central utilizada. Estas sensibilidades refletem as variações absolutas da variável explicada em função das variações, também absolutas, das variáveis explicativas. Pode-se perceber que, sendo grandezas absolutas, se não houver uniformidade nos valores definidos para as escalas das variáveis envolvidas, a interpretação gráfica dos resultados pode ser distorcida, já que as sensibilidades apuradas são influenciadas por estes valores.

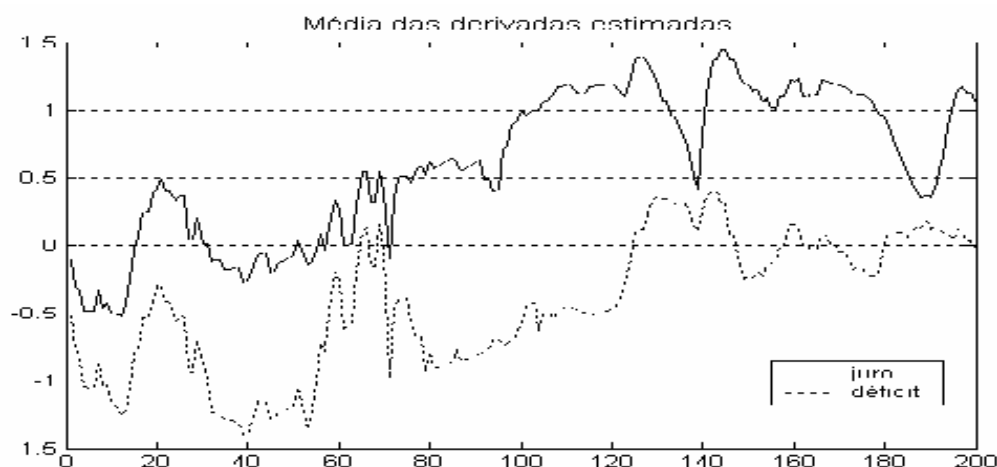


FIGURA 7: Derivadas médias do desemprego em relação a taxa de juro e ao déficit público relativo ao PIB.

Como esta uniformidade nem sempre é possível, devido a uma eventual diversidade das variáveis envolvidas, para que se possa fazer interpretações gráficas consistentes, independentemente das escalas, é preciso tornar relativas as medidas de sensibilidade. Este conceito de sensibilidade relativa é conhecido como coeficiente de elasticidade, expresso pela razão entre a variação percentual da variável explicada e a variação, também percentual, de uma

variável explicativa. A equação 3, mostrada a seguir e adaptada de GUJARATI (2000, pág. 159) define as elasticidades pontuais de uma função de múltiplas variáveis:

$$\varepsilon_i = \frac{\partial f}{\partial x_i}(a) \cdot \frac{x_i}{f}(a). \quad \text{<equação 3>}$$

Partindo das derivadas parciais nos pontos de treinamento, estimadas anteriormente, pode-se obter estimativas para as elasticidades desemprego-juro e desemprego-déficit nestes pontos. A figura 8 ilustra os resultados, onde percebe-se que no início do período em questão, fase 1, o desemprego foi, quase sempre, mais elástico em relação à variável fiscal do que em relação à variável monetária (taxa de juro), devido ao maior valor absoluto de suas elasticidades. Isto sugere que uma eventual política de redução do desemprego seria mais efetiva se feita através dos gastos públicos, relativo ao PIB, neste período.

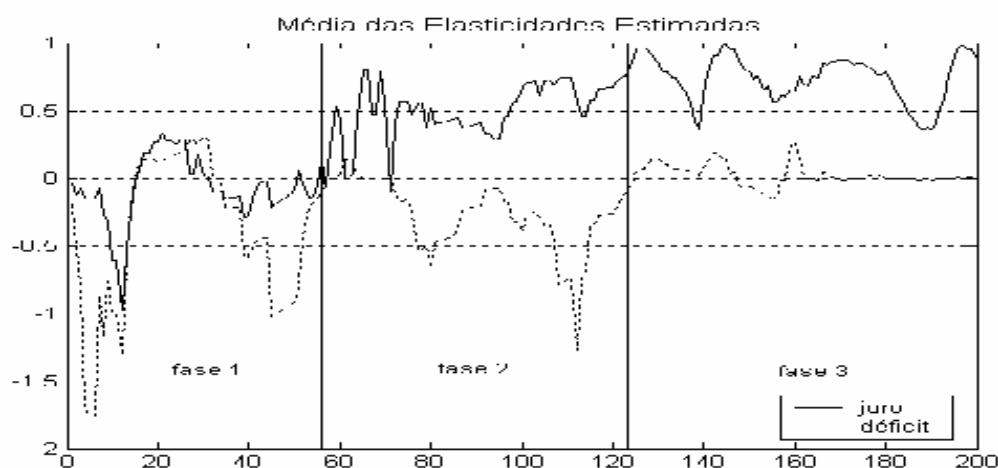


FIGURA 8: Elasticidades médias do desemprego em relação à taxa de juro e ao déficit público relativo ao PIB.

A partir da fase 2 esta situação começa a ser alterada. A elasticidade desemprego-juro cresce, fazendo com que as variações do desemprego sejam explicadas, equilibradamente, pelas variações tanto da taxa de juro nominal quanto do déficit público. Neste período pode-se inferir que as duas variáveis administradas foram eficazes para provocar alterações no nível de emprego. Na fase 3 a situação claramente se inverte. A elasticidade desemprego-déficit se reduz significativamente, fazendo da taxa de juro nominal o principal instrumento responsável pelas variações do desemprego.

Resumindo, conhecer a elasticidade significa conhecer em que sentido uma variável econômica influencia e quanto explica a outra. No limite, uma elasticidade nula significa

nenhuma influência. No final da fase 3 tivemos uma situação próxima a isto, onde a elasticidade da variável fiscal flutuou bem próximo da nulidade.

## **4 - Conclusões**

Neste artigo procurou-se expor os benefícios e as limitações das redes neurais na estimativa de elasticidades de funções econômicas. A primeira questão que precisou ser investigada foi a viabilidade desta ferramenta computacional para efetuar estas estimativas, procurando avaliar se as aproximações decorrentes do modelo eram suficientemente fiéis para permitir que, numericamente, se pudesse estimar as derivadas parciais da função aproximada. Como a resposta exigia que se soubesse o verdadeiro valor das derivadas parciais, para efeito de comparação, realizou-se um teste de avaliação. Partindo de uma expressão analítica qualquer estabelecida, gerou-se pontos, seqüenciados no tempo, para serem usados no treinamento da rede neural a fim de se aproximar a respectiva função e suas derivadas. O resultado, mediante a comparação entre as derivadas estimadas e as verdadeiras, foi sugestivo de que as redes neurais podem aproximar funções satisfatoriamente, apesar das limitações assinaladas.

Uma característica que é importante ressaltar é que cada vez que se treina uma rede neural, utilizando-se os mesmos dados e parâmetros, obtém-se valores diferentes como resposta para a superfície da função que se busca conhecer e, conseqüentemente, suas derivadas parciais. Esta flutuação das respostas para as estimativas das derivadas, que verificou-se obedecer a uma distribuição de freqüência quase-normal quando a rede está bem ajustada, é um ponto chave a ser considerado no equacionamento do problema. A sintonia da rede deve buscar a minimização destas flutuações, observando a distribuição de freqüência em cada ponto do treinamento. Quanto mais leptocúrtica a distribuição, mais estável é comportamento da rede.

Outro ponto importante é o erro médio quadrático do ajuste em relação aos pontos de treinamento. De nada adianta ter estimativas bem regulares para as derivadas parciais da função se a rede neural aproximar mal a variável explicada. E a única informação que se tem para quantificar este erro é a resposta da rede neural nos pontos usados no seu treinamento, já que neles se conhece o valor verdadeiro da variável explicada. A diferença entre a resposta da rede e o alvo (variável explicada), é o erro do ajuste. A sintonia deve atender ao nível de erro estabelecido como aceitável para o problema.

Vale ressaltar que o modelo neural não oferece regras firmes e seguras para se chegar à melhor resposta para o problema. A solução que ela apresenta é resultado de uma busca aleatória, num espaço infinito de possibilidade, através de um processo essencialmente de tentativa e erro. Também é preciso frisar que não é possível extrair da observação dos parâmetros da rede sintonizada, qualquer característica da relação de interdependência entre as variáveis explicativas e a explicada. As informações ali contidas são exclusivamente computacionais.

Quanto às limitações do modelo, a principal delas é que não há parâmetros objetivos para se avaliar a qualidade dos resultados alcançados pelas redes neurais. Basicamente, o que se pode inferir, baseado nos experimentos, é que as respostas podem ser estimativas coerentes, mas que precisam ser avaliadas por um conhecedor do problema para que possam ganhar relevância.

Resumindo, assim como qualquer outro instrumento aproximativo, não se pode extrair das redes neurais informações que sejam a expressão da verdade. Mas elas podem sim responder satisfatoriamente de modo a permitir que se obtenha informações relevantes para o esclarecimento de fenômenos econômicos. Além do mais, se considerarmos que muitos destes fenômenos são de grande complexidade analítica e que o esforço demandado pelas redes neurais para o equacionamento da solução é majoritariamente computacional, as informações por elas geradas ganham importância, se não para concluir, sim para a levantar novas hipóteses ou acumular evidências numa ou noutra direção.

Existem trabalhos futuros correlacionados, tanto na área econômica quanto na computacional, que se pretende realizar. Na economia, seria oportuno aplicar os aspectos metodológicos aqui apresentados a problemas econômicos reais, confrontando as soluções alcançadas com outros estudos e métodos conhecidos. Na área da computação, há espaço para perseguir um aprimoramento do modelo neural, definindo novos parâmetros ou topologias, a fim de se melhorar os resultados das redes neurais na aproximação de funções, fazendo de suas respectivas derivadas uma importante condição de contorno para o equacionamento do problema.

## Referências Bibliográficas

- CARDIM DE CARVALHO, F.; *Mr.Keynes and the Post Keynesians*. Cheltenham: Edward Elgar, 1992
- BRAGA, A. P., CARVALHO, A. P. L. F., LUDERMIR, T. B., *Fundamentos de Redes Neurais Artificiais*. 11<sup>a</sup> Escola de Computação, 1998.
- DIAZ, M. D. M., ARAÚJO, L. J., S.; *Aplicação de Redes Neurais à Economia: Demanda por Moeda no Brasil*. Economia Aplicada, v.2, n.2, 1998.
- FERNANDES, L. G. L., PORTUGAL, M. S., NAVAUX, P. O. A.; *O Problema da Escolha da topologia da Rede Neural na Previsão de Séries de Tempo*. III SBRN, 1996.
- GUJARATI, D. N.; *Econometria Básica*. 3<sup>a</sup> Edição. Makron, 2000.
- HYKIN, S. S.; *Redes Neurais – Princípios e Prática*. Bookman, 2000.
- LIMA, E. L.; Curso de Análise Volumes 1 e 2. Instituto de Matemática Pura e Aplicada, 1995.
- SICSÚ, J.; *Credible Monetary Policy*. Journal of Keynesian Economics, 23(4): pp.669-87, 2001
- SICSÚ, J.; *Reputação das Autoridades Monetárias e Credibilidade das sua Políticas: um abordagem pós-keynesiana*. Estudos Econômicos, 27(1): pp.5-27, 1997.
- SILVA, A. B. M., PORTUGAL, M. S., CHECHIN, A. L.; *Redes Neurais Artificiais e Análise de Sensibilidade: Uma aplicação à demanda de importações brasileira*. Revista de Economia Aplicada. São Paulo, v.5, n.4, p.645 - 693, 2001.
- ZURADA, J.M.; *Introduction to Artificial Neural Systems*, West Publishing Company, 1992.