



UNIVERSIDADE  
FEDERAL DO CEARÁ

25



SÉRIE  
ESTUDOS ECONÔMICOS CAEN

# Previsão da Demanda por Moeda no Brasil: *Machine Learning* vs Modelos Autorregressivos

**Maurício Benegas**  
**Bernardo Melo Volker**

FORTALEZA • MAIO • 2018



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA - CAEN**

**SÉRIE ESTUDOS ECONÔMICOS – CAEN**

**Nº 25**

**Previsão da Demanda por Moeda no Brasil:**  
***Machine Learning* vs Modelos Autorregressivos**

FORTALEZA – CE

MAIO – 2018



# PREVISÃO DA DEMANDA POR MOEDA NO BRASIL: *MACHINE LEARNING* VS MODELOS AUTORREGRESSIVOS

Mauricio Benegas

CAEN - UFC

Bernardo Melo Volker

CAEN - UFC

## RESUMO

Este trabalho propõe o uso de modelos de *machine learning* para prever a demanda por moeda no Brasil. Os resultados são comparados com os métodos econométricos atualmente utilizados para este fim. Considerando o equilíbrio no mercado monetário, foram utilizados os agregados M1 e M4 como medidas da demanda por moeda. Os modelos de *machine learning* utilizados foram as redes neurais autorregressivas cujos resultados foram comparados aos obtidos de um modelo autorregressivo de defasagens distribuídas. Para M1, encontrou-se considerável melhora na previsão da variável com os modelos de redes neurais, entretanto, o mesmo não ocorreu na previsão de M4, em que os modelos de rede neurais constantemente subestimam os valores reais.

**Palavras-chave:** Demanda por Moeda. Previsão. *Machine Learning*.

## ABSTRACT

This paper proposes the use of machine learning models to predict the demand for money in Brazil. The results are compared with the econometric methods currently used for this purpose. Considering the equilibrium in the money market, the aggregates M1 and M4 were used as measures of the demand for money. The machine learning models used were the autoregressive neural networks whose results were compared to those obtained from an autoregressive model of distributed lags. For M1, there was a considerable improvement in the prediction of the variable with the neural network models, however, the same did not occur in the M4 prediction, where the neural network models constantly underestimate the real values.

**Keywords:** Demand for money; Forecasting; Machine Learning

**JEL Codes:** C45, C53, E41

## 1 INTRODUÇÃO

A política monetária é peça fundamental como instrumento estabilizado e mesmo para a promoção do crescimento de uma economia. Historicamente, países em que a autoridade monetária não agiu de forma responsável tiveram suas economias levadas a instabilidade de preços ou a baixo crescimento. Entretanto, a manutenção de políticas monetárias sólidas não é uma tarefa fácil para os bancos centrais, pois o este apenas controla a oferta de moeda, necessitando de estimativas confiáveis para a demanda de moeda dos agentes. Com as inovações financeiras ocorridas principalmente nos séculos XX e XXI, a complexidade entre relações causais das variáveis macroeconômicas aumentou sem precedentes, limitando o poder de explicação dos antigos modelos econômicos.

A moeda nas ciências econômicas tem sua importância reconhecida ao longo do desenvolvimento das diversas escolas de economia, mesmo antes do *establishment* da teoria econômica clássica. Pensadores como Jean Bodin e David Hume teorizavam sobre a grande inflação que acontecera no continente europeu nos anos de 1500, argumentando que a repentina e descomunal elevação no volume de moeda – ouro e prata na época – causada pela exploração espanhola nas américas, influenciava diretamente o nível de preços da economia.

Mais tarde, os clássicos teorizaram sobre os determinantes da demanda por moeda, definindo dois motivos: realizar transações ordinárias no mercado e como segurança para eventos inesperados. Depois, Keynes (1936) ampliou a complexidade do mecanismo que leva os indivíduos a demandarem moeda. Em seus trabalhos o conceito de demanda para especulação foi introduzido a teoria econômica. Todos os futuros avanços na teoria monetária sobre demanda por moeda se utilizaram dos três conceitos

desenvolvidos pelos economistas clássicos e Keynes, seja desenvolvendo o modelo Keynesiano como feito por Baumol (1952) ou revisando o modelo quantitativista clássico em Friedman (1966).

Se por um lado a teoria econômica da moeda goza atualmente de relativo consenso entre os economistas, a questão de estimar e prever a demanda por este ativo ainda esbarra em várias dificuldades<sup>1</sup>. Uma destas dificuldades, e talvez a mais controversa, está na especificação dos modelos destinados à realizar previsões. Assim, o objetivo principal deste trabalho consiste em propor como método de previsão da demanda por moeda no Brasil os modelos de *Machine Learning*. O exercício proposto utiliza uma amostra de frequência mensal no período de 1995 a 2015. Por fim os resultados são comparados aos obtidos através dos modelos mais recentes utilizados na literatura para a estimação de demanda por moeda.

Devido a impossibilidade de se observar diretamente a demanda por moeda, suporemos que o mercado monetário está em equilíbrio, portanto, os agregados monetários de oferta de moeda do BACEN serão utilizados como *proxy* para demanda. No Brasil, o BCB em sua programação monetária – publicada trimestralmente – utiliza os meios de pagamentos restritos e a poupança financeira, M1 e M4 respectivamente, não estabelecendo valores fixos para esses agregados, mas projetando os valores esperados pelo BCB para períodos futuros. Dessa forma, iremos trabalhar apenas com esses dois conceitos.

Além desta introdução, este trabalho está distribuído da seguinte forma: na segunda seção será feita uma revisão da literatura; na terceira seção serão expostos os métodos utilizados neste trabalho; na quarta seção é feita uma descrição e análise das variáveis utilizadas; na quinta seção são apresentados os resultados obtidos; por fim a sexta seção apresenta as considerações finais.

## **2 REVISÃO DA LITERATURA: A dificuldade na estimação de modelos empíricos**

Chow (1966) estudou a demanda por moeda nos Estados Unidos entre 1897 e 1958 em curto e longo prazo. O autor tinha como foco estudar os determinantes da demanda por moeda: transação e especulação, e qual a sua importância relativa. Para tal, a variável renda corrente foi considerada como *proxy* para o motivo transação e a renda permanente para o motivo especulação. Ao final, as variações de curto prazo da demanda são causadas por variações na renda corrente, enquanto que no longo prazo a renda permanente é mais importante.

A princípio, devido ao baixo desenvolvimento do mercado monetário, as estimações de Chow (1966) se demonstraram satisfatórias. Entretanto, como demonstrado por Bernanke (2006), a partir da década de 1970 as previsões para a demanda por moeda nos Estados Unidos se tornaram menos precisas, constantemente sobrestimando o crescimento da demanda por M1. Na década de 1980, o FED mudou o foco da política monetária para M2, de modo a considerar as inovações que ocorreram para a formulação da política monetária. Novamente, os modelos tradicionais não foram eficientes em prever a demanda por M2, levando o FED a abandonar o estabelecimento de metas para os agregados monetários.

Tendo em vista as inovações monetárias que ocorriam nos Estados Unidos, Goldfeld, Fand e Brainard (1976) em seu trabalho intitulado “*The Case of the Missing Money*” estudaram as falhas presentes nos modelos tradicionais de previsão e porquê esses estavam sobre-estimando as predições. Apesar dos esforços, o trabalho não conseguiu, de fato, encontrar as falhas que levaram os modelos tradicionais a perderem poder de explicação. Concluindo que, possivelmente, houve uma quebra estrutural na função de demanda por moeda.

Este problema do “sumiço do dinheiro” na previsão de demanda por moeda não se observou apenas nos Estados Unidos, sendo estudado também no Reino Unido por Hendry e Ericsson (1991). Os autores, assim como Goldfeld, Fand e Brainard (1976), buscaram os motivos que levavam os métodos tradicionais a serem falhos a partir das inovações financeiras. Eles propuseram cinco possíveis problemas: inconstância das equações tradicionais, exogeneidade ou endogeneidade da moeda, inversibilidade do modelo, determinantes de longo e curto prazo e a causalidade entre as variáveis. Dessa forma, os autores concluem que a especificação errada dos modelos leva aos erros de previsão, e desenvolvem um modelo que corrige esses problemas, obtendo resultados satisfatórios.

---

<sup>1</sup> Na seção seguinte é feita uma revisão da literatura sobre trabalhos empíricos na área.

Até o trabalho de Henry e Ericsson (1991), os estudos passados na área consideravam linearidade entre a demanda por moeda e seus determinantes, o sucesso na estimação do modelo não linear de correção de erros proposto pelos autores levou ao desenvolvimento de outros trabalhos considerando a não linearidade da demanda por moeda. Primeiramente, Teräsvirta e Eliasson (2001) identificaram que o modelo desenvolvido por Henry e Ericsson (1991) é uma aproximação de uma regressão de transição suave (STR) e buscou generalizar o modelo de correção de erro usando uma estrutura STR. Novamente os resultados foram satisfatórios, com o modelo desenvolvido descrevendo de forma significativa a demanda por moeda no Reino Unido no período de 1878-1993.

Mais recentemente, Jawadi e Sousa (2013) também advogam em favor da não linearidade dos agregados monetários:

*“Na literatura, autores têm usado diferentes técnicas econométricas para estimar a demanda por moeda na zona do Euro, Estados Unidos e Reino Unido, mas todos os trabalhos possuem uma característica em comum: a demanda por moeda possui uma relação linear com os balanços monetários, o PIB real e a taxa de juros nominal [...] Apesar disso, nos anos 1970, alguns bancos centrais e vários economistas da academia começaram a afirmar que a função de demanda por moeda era ou instável ou nem mesmo existia” (JAWADI e SOUSA, 2013, tradução livre)*

Para tanto, fora utilizada uma regressão quantílica e uma versão exponencial do modelo STR para estimar a demanda por moeda na zona do Euro, Estados Unidos e Reino Unido. Os resultados dos autores para ambos os modelos indicaram estabilidade da função de demanda por moeda para as três áreas estudadas.

Na contramão da utilização de modelos não lineares, trabalhos como Daniele, Foresti e Napolitano (2017), Ange-Patrick e Hervé (2017) utilizaram modelos autoregressivos de defasagens distribuídas (ARDL). O primeiro trabalho estudou a estabilidade da demanda por  $M1$  e  $M2$  na Itália, os autores concluíram que para ambos os agregados apresentam demanda estável se considerado um modelo aumentado, utilizando, além das variáveis tradicionais, a taxa real de câmbio e uma medida de volatilidade da taxa real de câmbio. Dessa forma, os autores argumentam que as estimações de demanda por moeda as quais apresentam não estabilidade da função de demanda sofrem de variáveis omitidas. O segundo trabalho realiza o mesmo procedimento do primeiro trabalho para a Costa do Marfim e Gana, estimando um modelo aumentado com a taxa real de câmbio. Os resultados indicam que a função de demanda por moeda é linear para ambos os países.

Por fim, a estimação da demanda por moeda pelo método de redes neurais não é algo inédito. Shao (2013) comparou o poder de previsão por  $M0$  em Taiwan entre um modelo ARIMA, um modelo de Rede Neural Artificial (RNA) e um modelo híbrido de dois estágios, em que primeiro as variáveis eram selecionadas através de critérios de seleção e pelo fator de inflação da variância e, após a seleção, apenas as variáveis restantes eram usadas no modelo de RNA. Os resultados demonstraram que os modelos híbridos foram superiores a todas as outras especificações utilizadas.

No Brasil há certa escassez de estudos objetivando a previsão da demanda por moeda. Até os anos 2000, o país sempre apresentou instabilidades na economia, principalmente nas décadas de 80 e 90, quando a inflação se intensificou e levou a adoção de planos de estabilização. A especificação de um modelo de demanda por moeda com inflação elevada já apresenta grandes dificuldades, adicionadas as quebras estruturais presentes com os planos de estabilização, entendemos a pouca literatura na área.

Rossi (1988) utilizou dados trimestrais entre 1966 e 1985 em busca das variações que devem ter ocorrido na função de demanda por moeda com as inovações financeiras adotadas no país nessa época. Duas especificações são estimadas pelo autor: a primeira é uma simples regressão log-log, tendo a taxa nominal de juros, renda real e inflação como variáveis independentes; o segundo modelo utiliza a defasagem da demanda por moeda como variável explicada, de forma a comparar os resultados das estimações para o curto e longo prazo. Os resultados das previsões feitas pelo autor não foram

satisfatórios, o valor previsto ficou acima do valor real observado em até 120%, demonstrando exatamente o mesmo problema de sobre-estimação ocorrido nas economias Norte Americana e Inglesa.

O mesmo problema foi observado em Tourinho (1996) para o período entre 1974 e 1995. Diferente do trabalho realizado por Rossi (1988), o estudo utilizou uma especificação alternativa ao modelo log-log comumente utilizado, realizando uma transformação Box-Cox na variável dependente – o autor, diferentemente dos outros trabalhos, utilizou a Base Monetária como variável dependente – e mantendo as variáveis independentes na sua forma linear. Para os períodos antes Plano Real, as estimações do autor se mostraram satisfatórias, entretanto, ao prever o comportamento da demanda por moeda houve novamente sobre-estimação, sendo o valor previsto em média 50% maior que o real ocorrido.

### 3 METODOLOGIA

Nesse trabalho, utilizaremos um modelo ARDL como feito por Daniele, Foresti e Napolitano (2017) e através da raiz do erro quadrado médio (REQM) compararemos a qualidade de previsão com dos modelos de RNA.

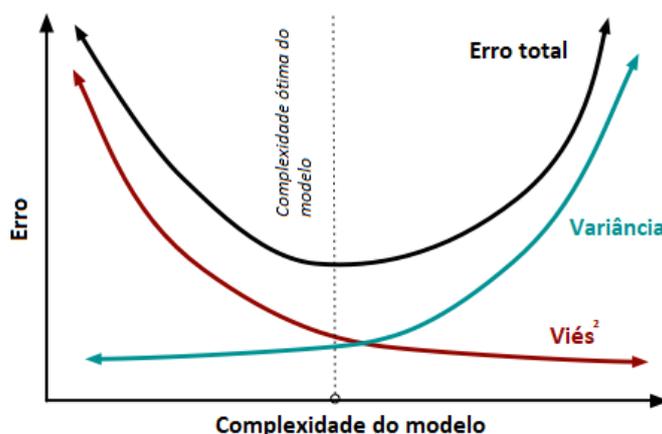
Os dados utilizados possuem periodicidade mensal, começando em janeiro de 1995 e se estendendo para dezembro de 2015. Todos os valores foram deflacionados pelo Índice de Preços ao Consumidor Amplo (IPCA) para valores de dezembro de 2015. De forma a compararmos os modelos igualmente, os modelos serão estimados no período de 1995 a 2014 e o ano de 2015 será utilizado para realizar as previsões.

Como citado, será utilizado o REQM como medida de performance na comparação dos métodos. Essa medida pode ser expressa na forma do bem-conhecido *tradeoff* entre viés e precisão (HASTIE,2009), de modo que, sendo  $y$  o verdadeiro valor da variável de interesse e  $\hat{y}$  o valor previsto para a mesma, tem-se:

$$REQM(\hat{y}) = \sqrt{\text{viés}(\hat{y})^2 + \text{Var}(\hat{y})}$$

Portanto, para se obter o melhor modelo de previsão possível devemos tentar minimizar o viés e a variância da estimação. Porém, viés e variância possuem relações opostas com a complexidade (número de parâmetros utilizados) do modelo, como demonstrado em James *et al* (2013). Caso vários parâmetros sejam adicionados ao modelo reduzindo o viés ao máximo, estamos uma situação de *overfitting*, isso significa que o modelo treinado tem fraco poder de generalização para observações fora da amostra de treinamento utilizado. Também pode ocorrer casos em que o modelo é parcimonioso, com poucos parâmetros e uma variância baixa do erro, entretanto o modelo é muito viesado, gerando *underfitting*, tendo os mesmos problemas de generalização de um modelo muito complexo. Graficamente:

Gráfico 1 - Relação entre viés e variância



Fonte: *Understanding bias and variance trade-off*<sup>2</sup>

<sup>2</sup> Disponível em: <http://scott.fortmann-roe.com/docs/BiasVariance.html>.

Para encontrarmos o melhor modelo, devemos então encontrar a relação ótima entre viés de variância. O método de validação cruzada será utilizado para esse fim, os dados serão divididos em blocos, sendo os primeiros 90% dos dados utilizados como conjunto de treinamento, os próximos 5% são utilizados como validação e os últimos 5%, correspondentes as doze últimas observações, são os valores a serem previstos. O modelo será definido ótimo quando o treinamento não conseguir mais melhorar a qualidade de previsão dos dados de validação e os dados de teste são mantidos fora desse processo, sendo uma métrica independente do processo de treinamento. Esse processo será repetido  $n$  vezes para se obter uma média da REQM no conjunto de teste.

### 3.1 Redes neurais artificiais

As redes neurais artificiais (RNA) buscam imitar o modo de processamento de dados do cérebro humano. O modelo aprende padrões complexos e não lineares dos dados fornecidos através da interação entre neurônios artificiais os quais são ativados por sinapses. Esse processo de aprendizado possui grande capacidade de adaptação e generalização, sendo um modelo eficaz para a previsão de problemas complexos. (HAYKIN, 2009)

Nesse trabalho será utilizada redes neurais dinâmicas, também conhecidas com redes neurais recursivas (RNR). A estrutura básica de uma RNR é dada por três camadas, sendo a primeira camada uma camada de entrada onde as variáveis exógenas são imputadas. A segunda camada intermediária possui os neurônios da rede e finalmente a terceira é a camada de saída. A recursividade da rede vem do fato que os valores emitidos pelas camadas de saídas voltam a rede para ajudar a previsão dos valores futuros.

Seguindo Haykin (2009), os neurônios artificiais na camada intermediária atuam através de uma média ponderada dos sinais recebidos (no presente caso as variáveis explicativas) e utilizando esse valor resultante em uma função de ativação. Formalmente, temos para um neurônio  $u_k$ :

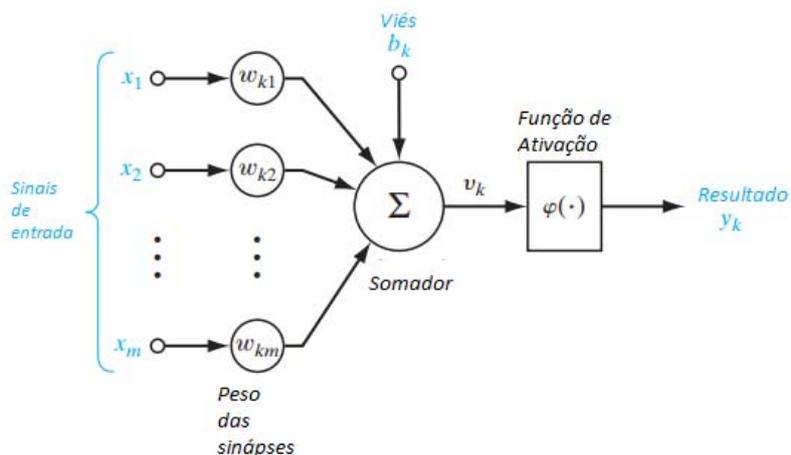
$$u_k = \sum_{j=1}^m w_{kj} x_j$$

Onde  $(w_{kj})_{k=1}^m$  são os respectivos pesos dados as sinapses e  $(x_k)_{k=1}^m$  são os valores das variáveis explicativas introduzidas no modelo. Dessa forma, temos que o resultado obtido no neurônio é a soma de todas as sinapses com seus respectivos valores e um termo de viés  $b_k$  aplicado em uma função de ativação  $\phi$ . O termo de viés é determinado de forma exógena, de modo a corrigir possíveis falhas de previsão. Nesse trabalho utilizaremos  $b_k = 0$ . Assim, temos que a resposta do neurônio é dada por:

$$\hat{y}_k = \phi(u_k)$$

Graficamente, um neurônio pode ser representado tal qual mostrado no gráfico 2 abaixo.

Gráfico 2: Um neurônio em uma RNA



Fonte: Adaptado de Haykin (2009)

Existe uma grande variedade funções de ativação, alguns exemplos podem ser vistos em Karlik e Olgac (2011). A função de ativação necessita ser continuamente diferenciável, essa propriedade é necessária para o treinamento da rede. Além disso, é desejável que ela seja não linear, propiciando a característica de não linearidade do modelo de RNA, como visto em Haykin (2009). Tal como em Shao (2013), neste trabalho será utilizada a função logística, tal que:

$$\phi(u_k) = \frac{1}{1 + e^{-u_k}}$$

Uma das vantagens no uso da função logística está na simplicidade de sua derivada (representando a contribuição do neurônio para a previsão da variável de interesse). Esta é dada por:

$$\phi'(u_k) = \phi(u_k)[1 - \phi(u_k)] \quad (1)$$

Por último, devemos definir uma função custo  $\psi_n$  a ser minimizada. Esta função compara o valor previsto pelo modelo com o valor real da variável para cada iteração  $n$ . Tal como no caso da função de ativação, a função custo deve ser diferenciável (HAYKIN, 2009). Uma formulação muito comum, e que será adotada neste trabalho, é a função custo quadrática, dada por:

$$\psi_n(\hat{y}_k) = \frac{(\hat{y}_k - y_k)^2}{2} \quad (2)$$

As funções de ativação e de custo são os elementos necessário para a implementação do algoritmo de *backpropagation*. Esse algoritmo busca minimizar a função custo atualizando o peso das sinapses através do gradiente da função custo. Especificamente, utilizando (1) e (2), a derivada de  $\psi_n$  em relação aos pesos  $(w_{kj})_{k=1}^m$  é dada por:

$$\frac{\partial \psi_n}{\partial w_{kj}} = (\hat{y}_k - y_k) \phi(u_k) [1 - \phi(u_k)] x_j, \quad j = 1, 2, \dots, m$$

Logo, dado um neurônio  $k$  fixado e denotando por  $\Delta w_{kj}^n$  a correção da  $j$ -ésima sinapse na  $n$ -ésima iteração do algoritmo, tem-se que:

$$\Delta w_{kj}^n = -\eta \frac{\partial \psi_n}{\partial w_{kj}^n}$$

onde  $\eta$  é um parâmetro de velocidade de aprendizado. Esse parâmetro é positivo de modo que a direção de atualização dos pesos reflita um mínimo da função  $\psi_n$ .

### 3.1.1 Rede neural artificial autoregressiva (NAR)

Diferentemente de um modelo RNA onde a variável de saída  $\hat{y}$  é função de um vetor de variáveis explicativas  $(x_1, \dots, x_m)$ , no modelo autorregressivo consideram-se os valores defasados de  $y$  como variáveis explicativas do modelo. Especificamente, sendo  $\hat{y}_t$  o predictor de  $y$  na data  $t$  a forma geral de um modelo NAR é dada por:

$$\hat{y}_t = \phi(y_{t-1}, \dots, y_{t-d})$$

Note que, para uma rede NAR, além do número ótimo de neurônios, o número de defasagens que minimiza o REQM também deve ser obtido através de validação cruzada. Para referência futura, será utilizada a notação  $NAR(d, n)$  para indicar um modelo de Rede Neural Artificial Autorregressivo com  $d$  defasagens e  $n$  neurônios.

### 3.1.2 Rede neural artificial autoregressiva com variáveis exógenas (NARX)

A adição de variáveis exógenas no modelo NAR dá ao modelo novas informações as quais podem ser úteis para melhorar o desempenho do previsor  $\hat{y}_t$ . Além dos valores defasados da variável explicada, as variáveis explicativas também são utilizadas em nível e com defasagens no treinamento do modelo. A forma geral para um modelo NARX é dada por:

$$\hat{y}_t = \phi(y_{t-1}, \dots, y_{t-d}, x_t, x_{t-1}, \dots, x_{t-f})$$

onde  $x_{t-s} = (x_{1t-s}, \dots, x_{mt-s})$  para  $s = 0, 1, \dots, f$ . Similarmente a rede NAR, o nível ótimo de defasagens para ambas as variáveis explicativas e explicada são definidos por validação cruzada. Para referência futura, será utilizada a notação NARX( $f, d, n$ ) para indicar um modelo de Rede Neural Artificial Autorregressivo com Variáveis Exógenas, com  $f$  defasagens nas variáveis exógenas,  $d$  defasagens na variável de interesse e  $n$  neurônios.

### 3.2 Modelo autorregressivo com defasagens distribuídas (ARDL)

O modelo ARDL vêm sendo comumente utilizado para a estimação e previsão da demanda por moeda como visto na revisão de literatura deste trabalho. Portanto, será utilizado como *benchmark* para a comparação entre os modelos de *machine learning* propostos. Este modelo autorregressivo de série temporal também considera valores defasados exógenos para a previsão.

Neste trabalho será utilizada a seguinte especificação do modelo ARDL para a demanda de moeda:

$$m_t = \alpha_0 + A(L)m_{t-1} + B(L)z_t + C(L)\varepsilon_t$$

onde,  $m_t = \ln M_t$  e  $z_t = (\ln i_t, \ln \pi_t, \ln Y_t)$  são, respectivamente, o logaritmo da demanda por moeda e o logaritmo das variáveis explicativas, quais sejam, a taxa de juros ( $i$ ), a taxa de inflação ( $\pi$ ) e o produto ( $Y$ ). Os termos  $A(L)$ ,  $B(L)$  e  $C(L)$  são polinômios no operador de defasagens  $L$ . Para referência futura utilizaremos a notação ARDL( $d_M, d_i, d_\pi, d_Y$ )

## 4 DADOS UTILIZADOS

As variáveis utilizadas, sua descrição, periodicidade e fonte são apresentadas na Tabela 1 abaixo.

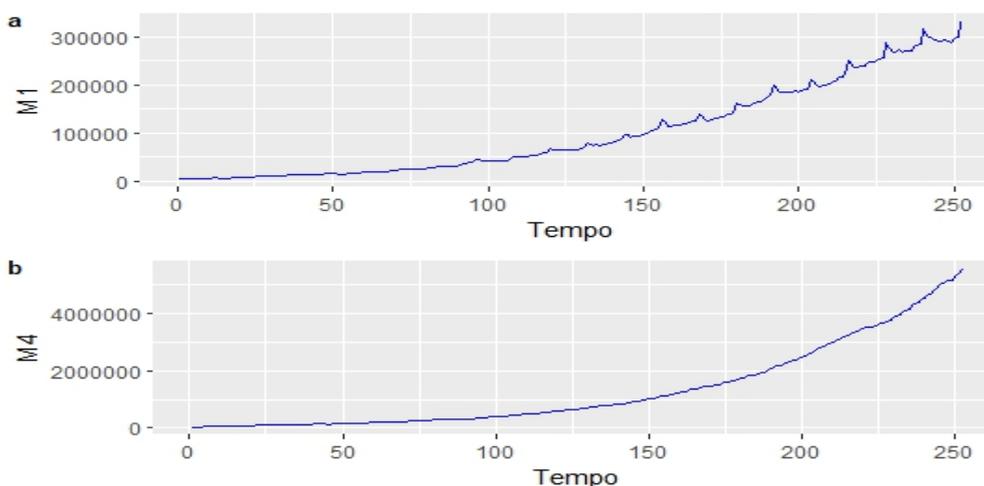
Tabela 1 - Variáveis utilizadas

Variável	Descrição	Periodicidade	Fonte
M1	Recursos prontamente disponíveis para pagamento de bens e serviços. O M1 é composto pelo papel-moeda em poder do público mais depósitos à vista. <sup>1</sup>	Mensal	BCB
M4	Conceito amplo de meios de pagamentos definido com base no critério de seus sistemas emissores. O M2 corresponde ao M1 e às demais emissões de alta liquidez realizadas primariamente no mercado interno por instituições depositárias. O M3 é composto pelo M2, pelas quotas dos fundos de renda fixa e pelas carteiras de títulos registrados no Selic. O M4 engloba o M3 e os títulos públicos de alta liquidez. <sup>1</sup>	Mensal	BCB
PIB	Produto interno bruto	Mensal	IBGE
Selic	Taxa básica de juros da economia	Mensal	BCB
IPCA	Índice de preço ao consumidor amplo	Mensal	IBGE

<sup>1</sup>Definições retiradas diretamente do BCB.

O Gráfico 2 ilustra o comportamento das variáveis M1 e M4. É possível notar certa sazonalidade na variável M1 nos últimos meses do ano, apresentando acelerado crescimento entre novembro e tendo um pico em dezembro, tendo forte retração em janeiro. Esse comportamento aparenta ganhar força a partir do ano de 2007, onde os picos passam a ser mais acentuados. A variável M4 apresenta um crescimento mais suave, entretanto, menos linear. Há uma notável aceleração no crescimento de M4 em períodos mais recentes, presume-se que a necessidade de financiamento do setor público com as recentes crises fiscais tenha aumentado a emissão de títulos públicos, causando esse comportamento.

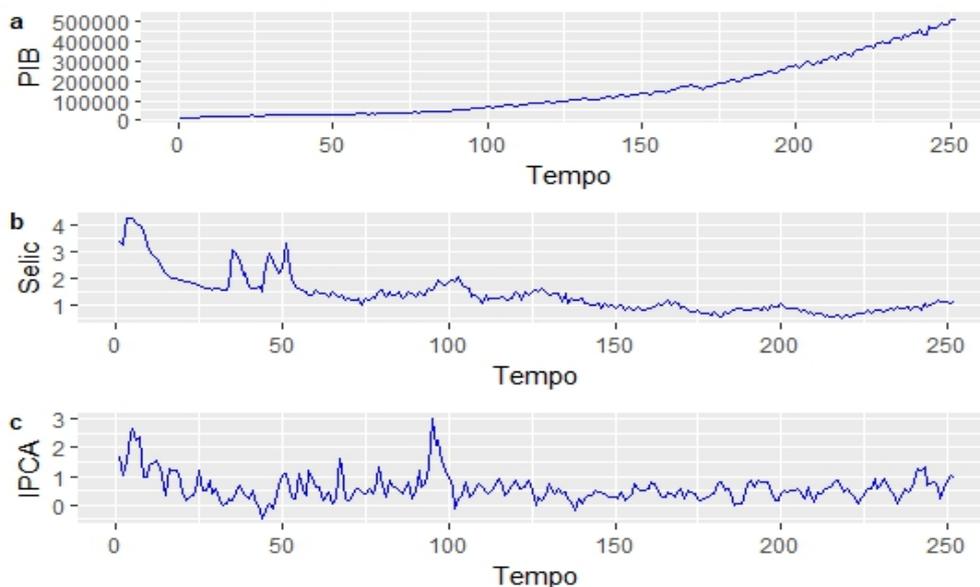
Gráfico 2 - (a) Variável M1. (b) Variável M4



Fonte: Elaboração própria

No Gráfico 3 é apresentado o comportamento das variáveis PIB, Selic e IPCA. O PIB apresenta fortes períodos de crescimento intercalados com períodos de recessão. O início do milênio e o final de 2010 apresentam o mais forte crescimento da economia, com as piores recessões ocorrendo em 2008 e a estagnação em 2015. A Selic apresenta constante tendência de baixa, no período, esse comportamento advém da estabilização da economia desde o Plano Real e da liberação do câmbio, que deixou de exigir um nível tão elevado de reservas internacionais para manutenção da paridade com o dólar, exigindo menos entrada de capital estrangeiro. Por fim, a inflação apresentou tendência de baixa nos primeiros períodos do Plano Real, porém, após a estabilização sua volatilidade continua consideravelmente forte, intercalando entre períodos de forte aceleração e períodos de baixos níveis inflacionários.

Gráfico 3 - (a) PIB. (b) Selic. (c) IPCA



Fonte: Elaboração própria

## 5 RESULTADOS

Para o modelo ARDL, não há problemas para a estimação se as variáveis utilizadas forem integradas de ordem 0 ou 1, entretanto há problemas se estas forem integradas de ordem 2. Por isso, testaremos as variáveis da especificação do modelo ARDL para a presença de integração de segunda ordem. Para o modelo de RNA, não há hipóteses necessárias à serem assumidas sobre o formato dos dados usados, entretanto, para acelerar o processo de treinamento e facilitar a convergência do algoritmo, a normalização dos dados é recomendável.

A Tabela 1 reporta o resultado dos testes de estacionariedade das variáveis utilizadas no modelo ARDL. Em nenhum dos testes houve a ocorrência de uma variável integrada de ordem 2, possibilitando a estimação desse modelo.

Tabela 2 - Testes de raiz unitária

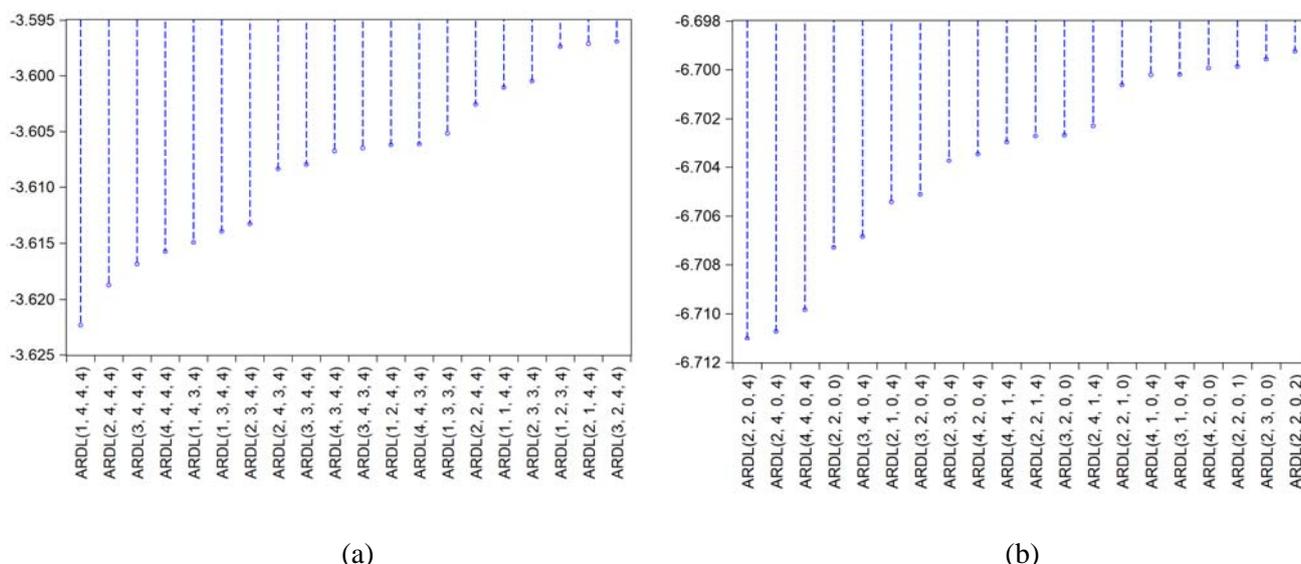
Variáveis	ADF		Philips-Perron	
	Constante	Constante e tendência	Constante	Constante e tendência
$\ln c$	$I(1)$	$I(1)$	$I(1)$	$I(1)$
$\ln \pi$	$I(0)$	$I(0)$	$I(0)$	$I(0)$
$\ln Y$	$I(1)$	$I(1)$	$I(1)$	$I(0)$
$\ln i$	$I(1)$	$I(1)$	$I(1)$	$I(1)$
$\ln M1$	$I(0)$	$I(1)$	$I(0)$	$I(1)$
$\ln M4$	$I(0)$	$I(0)$	$I(0)$	$I(0)$

Fonte: Estimativas dos autores

Nível de significância de 5% para todos os resultados

O número ótimo de defasagens para o modelo ARDL a ser escolhido será definido através do critério de Akaike (AIC), sendo escolhido a configuração que apresente o menor valor de AIC. Ao todo, foram testados todas as combinações possíveis com até quatro defasagens de todas as variáveis explicativas e o termo autorregressivo. Os Gráficos 2 e 3 apresentam os resultados dos vinte melhores modelos estimados para M1 e M4 respectivamente.

Gráfico 4: Melhor modelo ARDL: (a) para M1; (b) para M4



Fonte: Estimativas dos autores

Para M1, o ARDL(1,4,4,4) apresentou o menor AIC, indicando que o melhor modelo apresenta uma defasagem autorregressiva da variável explicada e quatro defasagens de todas as variáveis exógenas. Já para M4, o modelo ARDL(2,2,0,4) apresenta o menor AIC. A variável endógena entra com duas defasagens, sendo esse o mesmo número de defasagens para a taxa de juros; o PIB é defasado quatro vezes, enquanto o IPCA entra no modelo apenas em nível.

Para os modelos selecionados, devemos testar se os erros são não correlacionados serialmente e nem heterocedásticos. O teste de Breusch-Godfrey foi utilizado para a identificação de correlação serial e o teste de Breusch-Pagan-Godfrey para a presença de homocedasticidade. A Tabela 3 reporta os resultados de ambos os testes.

Tabela 3 – Testes de correlação serial e homocedasticidade

Test	$H_0$	$p$ -value (M1)	$p$ -value (M4)
Correlação Serial	Resíduos não correlacionados	0,9118	0,7502
Homocedasticidade	Resíduos Homocedasticos	0,0000	0,0014

Fonte: Elaboração própria

Com o  $p$ -value de 0,9118 e 0,7502, não podemos rejeitar a hipótese nula de não existência de correlação serial para os modelos de M1 e M4 respectivamente. Caso a hipótese nula fosse rejeitada, a estimação poderia ser viesada. O teste de homocedasticidade falhou em não rejeitar  $H_0$  para ambos os modelos. A presença de heterocedasticidade afeta a eficiência do estimador, entretanto, pela discussão do *trade-off* entre viés e variância, aceitaremos o modelo para a previsão apesar do fato dele não apresentar a menor variância possível.

Para os modelos de RNA, não há regra definida da forma de se obter um número mínimo ou máximo de neurônios para a RNA e um valor ótimo do parâmetro de velocidade de aprendizado. Um número muito grande de neurônios pode fazer com que a rede aprenda os dados do conjunto de treinamento muito bem, mas o poder de generalização para prever valores fora da amostra é prejudicado. Será utilizada a mesma metodologia de Shao (2013), o parâmetro de velocidade de aprendizado será  $\eta = 0,01$  e o número de neurônios testado será  $m = \pm 2$ , onde  $m$  é o número de variáveis de entrada. Além disso, como os modelos são autorregressivos, serão testadas com até quatro defasagens. De forma a validar os resultados, será feita validação cruzada de 10 repetições, utilizando-se a média da REQm para definir o melhor modelo.

Tabela 4 - Melhores modelos

	M1			M4		
	ARDL	NAR(1,1)	NARX(3,4,1)	ARDL	NAR(3,4)	NARX(1,2,4)
REQM	43147,45	10615,06	9627,09	68644,55	320508,02	187491,79

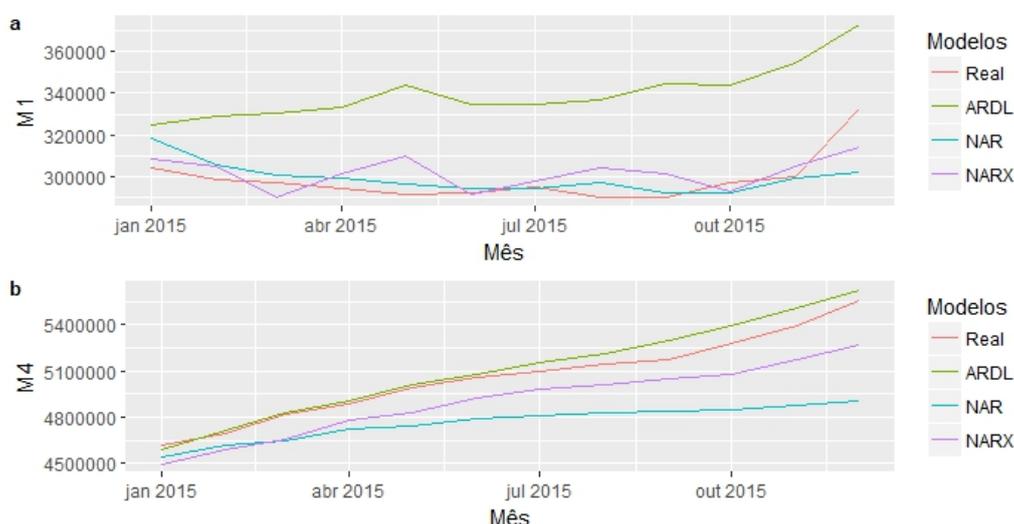
Fonte: Elaboração própria

A Tabela 4 apresenta os resultados da REQM dos melhores modelos estimados. Para M1 os dois modelos de RNA propostos apresentaram melhoras de até 77% na previsão em relação ao modelo ARDL usado como *benchmark*. Entretanto, os modelos propostos constantemente subestimaram o valor real em M4, levando uma previsão mais do que 100% pior em relação ao valor de referência.

Em momentos de relativa estabilidade econômica como no período aqui estudado, a variável M1 é bem-comportada, com seu componente estocástico não afetando seu processo de formação significativamente. Dessa forma, a modelagem pelos modelos de *machine learning* os quais “aprendem” com os dados foi mais eficiente que o método tradicional. Por outro lado, a variável M4 a qual considera títulos de diferente liquidez é influenciado por inúmeras variáveis da economia, aumentando o componente estocástico em M4. Portanto, ao se modelar essa variável com os modelos NAR e NARX, esses modelos não foram capazes de generalizar bem o processo de formação da série, obtendo resultados aquém dos esperados.

Para melhorar a previsão de M4, a utilização de métodos híbridos como o proposto por Zhang (2003) onde é utilizado um modelo ARIMA para a modelagem da parte linear e os modelos RNA modelam os resíduos possivelmente não lineares obtidos pelo modelo ARIMA, explorando as forças de cada método. Cadenas (2010) utilizando o modelo híbrido proposto obteve melhores previsões em relação à utilização de apenas modelos tradicionais. Além disso, a utilização do algoritmo de “*gradient boosting*” nas previsões de modelos ARIMA como proposto por Taieb (2014).

Gráfico 4 - (a) Série prevista para M1. (b) Série prevista para M4



Fonte: Elaboração própria

## 6 CONCLUSÃO

A proposta de utilizar modelos de *machine learning* para a previsão de demanda por moeda adveio das dificuldades de estimação desse parâmetro gerada pelas inovações financeiras, principalmente nas décadas de 70 e 80, as quais tornaram a função de demanda por moeda instável levando os bancos centrais pelo mundo a abandonar ou dar menos importância ao controle dos agregados monetários. Apesar do desenvolvimento e aplicação de modelos econométricos mais eficientes, como a regressão de transição

suave utilizada por Jawadi e Souza (2013), a demanda por moeda continua sendo uma variável a qual os bancos centrais continuam dando pouca importância.

Baseando-se na teoria, este trabalho utilizou a especificação tradicional da demanda por moeda, relacionando essa última ao produto, inflação e aos juros da economia. A opção por utilizar M1 e M4 parte diretamente da programação monetária publicada pelo BACEN. Após a definição do melhor modelo para previsão de ambas as variáveis, encontramos que os modelos dinâmicos de RNA melhoraram em até 77% a previsão de M1 em relação ao modelo ARDL utilizado como *benchmark* e considerando o comportamento aparentemente sazonal da variável, podemos inferir que os modelos de RNA têm maior facilidade na previsão de séries que apresentam sazonalidade. Por outro lado, ao tentar prever M4 que apresenta comportamento mais explosivo que M1, as RNAs constantemente subestimaram os valores reais sendo mais de 100% piores que o modelo ARDL. Não se pode, contudo, descartar a utilização desses modelos para a previsão de M4, devendo ser testadas novas especificações para o melhoramento das previsões geradas.

Finalmente, os modelos utilizados nesse trabalho são simples em comparação as possibilidades abertas pela introdução do *machine learning* nas ciências econômicas. A utilização para a previsão de demanda por moeda é só uma das várias áreas que podem se beneficiar da utilização desses modelos. A pouca relevância dos agregados monetários para a formulação da política monetária é questionada desde o começo do milênio, notadamente por Nelson (2003), Leeper e Roush (2003) e Goodhart (2007), e estudos como Azevedo e Pereira (2010) demonstram a relação de longo prazo entre os agregados monetários e a inflação. Portanto, a utilização de novos métodos para a previsão de demanda por moeda pode trazer de volta a atenção dos bancos centrais para este tema.

## REFERÊNCIAS

- ANGE-PATRICK, Yao Kouadio; HERVÉ, Drama Bédi Guy. Comparative Analysis of the Stability of Money Demand between Côte d'Ivoire And Ghana: An Application of ARDL Model. **International Journal of Economics and Finance**, v. 9, n. 11, p. 163, 2017.
- BAUMOL, William J. The transactions demand for cash: An inventory theoretic approach. **The Quarterly Journal of Economics**, p. 545-556, 1952.
- BERNANKE, Ben S. et al. Monetary aggregates and monetary policy at the federal reserve: A historical perspective. In: **Speech at the Fourth ECB Central Banking Conference, Frankfurt, Germany**. 2006.
- BREIMAN, Leo et al. **Classification and regression trees**. CRC press, 1984.
- CADENAS, Erasmo; RIVERA, Wilfrido. Wind speed forecasting in three different regions of Mexico, using a hybrid ARIMA-ANN model. **Renewable Energy**, v. 35, n. 12, p. 2732-2738, 2010.
- CARVALHO, Fernando; SICSU, João. **Economia monetária e financeira: teoria e política**. Elsevier Brasil, 2007.
- CHOW, Gregory C. On the long-run and short-run demand for money. **Journal of Political Economy**, v. 74, n. 2, p. 111-131, 1966.
- DANIELE, Vittorio; FORESTI, Pasquale; NAPOLITANO, Oreste. The stability of money demand in the long-run: Italy 1861-2011. **Cliometrica**, v. 11, n. 2, p. 217-244, 2017.
- ENDERS, Walter. **Applied econometric time series**. John Wiley & Sons, 2008.
- E AZEVEDO, João Valle et al. Forecasting inflation with monetary aggregates. **Economic Bulletin and Financial Stability Report Articles**, 2010.
- FRIEDMAN, Milton (Ed.). **Studies in the quantity theory of money**. University of Chicago Press, 1956.
- FRIEDMAN, Milton. Interest rates and the demand for money. **The Journal of Law and Economics**, v. 9, p. 71-85, 1966.
- GOLDFELD, Stephen M.; FAND, David I.; BRAINARD, William C. The case of the missing money. **Brookings Papers on Economic Activity**, v. 1976, n. 3, p. 683-739, 1976.
- GOODHART, Charles. Whatever became of the monetary aggregates?. **National Institute Economic Review**, v. 200, n. 1, p. 56-61, 2007.
- GREENE, William H. **Econometric analysis**. Pearson Education India, 2012.
- JAMES, Gareth et al. **An introduction to statistical learning**. New York: springer, 2013.

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. Springer series in statistics. **The elements of statistical learning: data mining, inference, and prediction**, 2009.

HAYKIN, Simon S. et al. **Neural networks and learning machines**. Upper Saddle River, NJ, USA:: Pearson, 2009.

HENDRY, David F.; ERICSSON, Neil R. Modeling the demand for narrow money in the United Kingdom and the United States. **European Economic Review**, v. 35, n. 4, p. 833-881, 1991.

JAWADI, Fredj; SOUSA, Ricardo M. Money demand in the euro area, the US and the UK: Assessing the role of nonlinearity. **Economic Modelling**, v. 32, p. 507-515, 2013.

KARLIK, Bekir; OLGAC, A. Vehbi. Performance analysis of various activation functions in generalized MLP architectures of neural networks. **International Journal of Artificial Intelligence and Expert Systems**, v. 1, n. 4, p. 111-122, 2011.

KEYNES, John Maynard et al. general theory of employment, interest, and money. 1936.

LEEPER, Eric M.; ROUSH, Jennifer E. **Putting'M'back in monetary policy**. National Bureau of Economic Research, 2003.

NELSON, Edward. The future of monetary aggregates in monetary policy analysis. **Journal of Monetary Economics**, v. 50, n. 5, p. 1029-1059, 2003.

ROSSETTI, José Pascoal; LOPES, João do Carmo. Economia monetária. **Atlas, São Paulo, 9º edição, 1987**, 2005.

ROSSI, José W. A demanda por moeda no Brasil: o que ocorreu a partir de 1980?. 1988.

SERLETIS, Apostolos. **The demand for money: Theoretical and empirical approaches**. Springer Science & Business Media, 2007.

SHAO, Yuehjen E. Prediction of currency volume issued in Taiwan using a hybrid artificial neural network and multiple regression approach. **Mathematical Problems in Engineering**, v. 2013, 2013.

TAIEB, Souhaib Ben et al. Boosting multi-step autoregressive forecasts. In: **ICML**. 2014. p. 109-117.

TERÄSVIRTA, Timo; ELIASSON, Ann - Charlotte. Non - linear error correction and the UK demand for broad money, 1878–1993. **Journal of Applied Econometrics**, v. 16, n. 3, p. 277-288, 2001.

TOURINHO, Octávio AF. A demanda por moeda no Brasil: 1974/95. 1996.

ZHANG, G. Peter. Time series forecasting using a hybrid ARIMA and neural network model. **Neurocomputing**, v. 50, p. 159-175, 2003.



---

Este trabalho propõe o uso de modelos de *machine learning* para prever a demanda por moeda no Brasil. Os resultados são comparados com os métodos econométricos atualmente utilizados para este fim. Considerando o equilíbrio no mercado monetário, foram utilizados os agregados M1 e M4 como medidas da demanda por moeda. Os modelos de *machine learning* utilizados foram as redes neurais autorregressivas cujos resultados foram comparados aos obtidos de um modelo autorregressivo de defasagens distribuídas. Para M1, encontrou-se considerável melhora na previsão da variável com os modelos de redes neurais, entretanto, o mesmo não ocorreu na previsão de M4, em que os modelos de rede neurais constantemente subestimam os valores reais.