ISSN 1518-3548

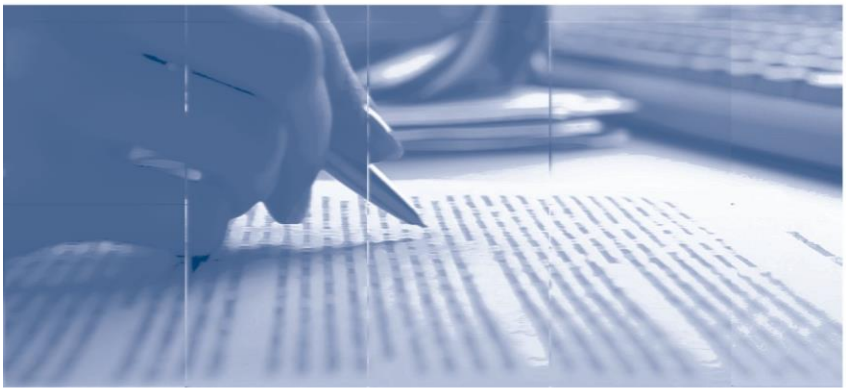
BANCO CENTRAL DO BRASIL

Behavioral Models of the Foreign Exchange Market: is there any empirical content?

João Barata R. B. Barroso

September, 2014

Working Papers



364

ISSN 1518-3548
CGC 00.038.166/0001-05

Working Paper Series	Brasília	n. 364	September	2014	p. 1-20

Working Page Series

Edited by Research Department (Depep) – E-mail: workingpaper@bcb.gov.br

Editor: Francisco Marcos Rodrigues Figueiredo – E-mail: francisco-marcos.figueiredo@bcb.gov.br
Editorial Assistant: Jane Sofia Moita – E-mail: jane.sofia@bcb.gov.br
Head of Research Department: Eduardo José Araújo Lima – E-mail: eduardo.lima@bcb.gov.br

The Banco Central do Brasil Working Papers are all evaluated in double blind referee process.

Authorized by Carlos Hamilton Vasconcelos Araújo, Deputy Governor for Economic Policy.

General Control of Publications

Banco Central do Brasil

Comun/Dipiv/Coivi

SBS – Quadra 3 – Bloco B – Edifício-Sede – 14º andar

Caixa Postal 8.670

70074-900 Brasília – DF – Brazil

Phones: +55 (61) 3414-3710 and 3414-3565

Fax: +55 (61) 3414-1898

E-mail: editor@bcb.gov.br

Citizen Service Division

Banco Central do Brasil

Deati/Diate

SBS – Quadra 3 – Bloco B – Edifício-Sede – 2º subsolo

70074-900 Brasília – DF – Brazil

Toll Free: 0800 9792345

Fax: +55 (61) 3414-2553

Internet: <http//www.bcb.gov.br/?CONTACTUS>

Behavioral Models of the Foreign Exchange Market: is there any empirical content?

João Barata R. B. Barroso*

Abstract

Behavioral models of the foreign exchange market explore the bias of economic agents towards forecasting rules with good recent performance. We propose an empirical framework to study such models without imposing restrictions on the set of forecasting rules or performance metrics. In particular, we propose a significance test for the constraints imposed by behavioral models relative to a very general non parametric alternative based on neural networks. We apply the framework to a unique dataset for the Brazilian foreign exchange market with full records of net order flow intermediated by the financial system, therefore connecting behavioral models to market microstructure models. The results support tightening constraints by 96% in the direction of behavioral models and this result is robust to assumptions regarding private order flow information.

Keywords: exchange rate dynamics, behavioral finance, neural network, order flow, market micro structure.
JEL Classification: F31, G02, C45, C58

* Research Department, Central Bank of Brazil. E-mail: joao.barroso@bcb.gov.br.

1. Introduction

Behavioral models of the foreign exchange market explore the bias of economic agents towards forecasting rules with good recent performance. This feature tends to generate complex dynamics in the population of active forecasting rules and disconnect from fundamentals in exchange rate behavior (Brock and Hommes (1997), De Grauwe and Grimaldi (2006a, 2006b)). Therefore, it is an interesting class of models with the potential to capture important features of foreign exchange markets.

Empirical specifications of such models have to incorporate both forecast heterogeneity and the associated population dynamics. The usual approach is to start from a set forecasting rules, generally based on the concept of chartists and fundamentalists, following the survey evidence of forecast heterogeneity (e.g. Froot (1990), Cheung and Chinn (2001), Ellen et.al (2013)). The forecasting rules are then estimated by first stage regressions and the population dynamics on a second stage based on some performance metric for the forecasting rules (e.g. De Grauwe and Grimaldi (2006a), Manan and Werterholff (2007), Reiner (2009), Jong et. al, (2013) and Ellen et.al (2013).

In this paper we propose a much less restricted approach. It works without imposing any significant restrictions on the set of forecasting rules or performance metrics. The population dynamics is estimated in a single step together with the set of forecasting rules. The estimator is computationally efficient and designed to capture robust nonlinear features in the data generating process. Moreover, we propose a significance test for the constraints imposed by behavioral models relative to a very general non parametric alternative. The null hypothesis is not contaminated by untested assumptions based on survey evidence and intuitive definitions of forecasting rules and performance metrics. In summary, we are able to investigate the empirical content of behavioral models in a very general and flexible framework.

We implement our nonparametric methodology for a unique dataset with order flow information. Therefore, we are able to connect with the important microstructure approach to foreign exchange markets (e.g. Evans and Lyons (2002), Vitale (2007)), but from a behavioral perspective that has not been explored in that literature. Relative to the empirical behavioral model literature, we also contribute by incorporating into the data a major close determinant of foreign exchange rate dynamics, the exclusion of which may bias any inference on the empirical content of behavioral features. The dataset refers to the Brazilian foreign exchange market and order flow information, due to regulatory requirements, covers all spot transactions. We show the dataset supports strong behavioral constraints on the data generating process. We also show the conclusions from significance tests of behavioral constraints are robust to alternative assumptions for the relation between aggregate order flow and private information. Both empirical results are important and showcase the relevance of our methodology.

The methodology is based on a nonparametric framework defined in terms of neural networks, a rich class of adaptive learning models. The main intuition from the paper is the formal similarities between behavioral models and neural network models. The central idea is to start from a very weak concept of behavioral models which is actually equivalent to neural networks and increasingly impose constraints on top of this concept. This allows us to explore classic nonparametric approximation results from neural networks and implement nonparametric testing procedures.

First, we show behavioral models, in a weak sense, approximate any data generating process, a corollary of White (1990). Second, we show that constrained models also approximate the process, but up to a neighborhood the size of which we estimate from the data. Third, we propose a significance test of the constraints based on a nonparametric wild bootstrap procedure (Kreiss, Neumann and Yao (2008)). Fourth, we consider a grid of increasingly tighter constraints and formulate a joint test controlling for false discovery rate (Benjamini and Hochberg (1995)). Finally, we propose a computationally efficient estimator for behavioral models which requires at most least squares optimization with a quadratic constraint. This estimator is based on the algorithm proposed by White (2006) and is designed to capture the most nonlinearity while also controlling for over-fitting. The computational efficiency estimation allows one to implement the testing procedure in a feasible timeframe.

The paper is structured as follows. In the following section we explore the approximation properties of behavioral models. The third section proposes a test for the behavioral constraints. In the fourth section we propose computationally efficient single step estimators for behavioral models. The next section presents the results of applying our methodology to Brazilian foreign exchange data. We summarize the results in the final section.

2. Approximation properties of behavioral models

A *behavioral model* for the foreign exchange market is

$$y_t = \sum_j \mu_{j,t} F_{j,t} + \varepsilon_t \tag{1}$$

where $\varepsilon_t = y_t - E(y_t|x_t)$ is the forecast error, $\mathrm{y_t}$ is the current period exchange rate, measured as deviation from fundamental or in first difference; x_t are exchange rate covariates; $F_{j,t} = F_j(x_t)$ is the forecast type j at period t for next period exchange rate; and $\mu_{j,t}$ is the measure of type j forecasters which follows a multinomial logit:

$$\mu_{j,t} = \frac{e^{\rho\pi_{j,t}}}{\sum_i e^{\rho\pi_{i,t}}} \tag{2}$$

with $\rho > 0$ the sensitivity to past performance of forecasting rules and $\pi_{j,t} = \pi_j(x_t)$ the realized profit from having used the forecasting rule of type j in the previous period. Regarding the set of conditioning variables x_t, it may include lags of the exchange rate itself and current values and lags of variables that forecast the exchange rate and are available to the agents at the time of the forecast. Notice this is a behavioral model in a weak sense, because there are no reasonable restrictions on profit function and forecasting rules suggested by economic theory. We address this as follows.

A *reasonable behavioral model* is such that (i) the forecast are related to the realized profit by the functional relation $\pi_{j,t} = \varphi_y(y_{t-1})\varphi_{Ey}(F_{i,t-1})$ with $\mathrm{\varphi_y}$ and $\mathrm{\varphi_{Ey}}$ the sign, the identity or other simple function, (ii) the forecasts are sensible in the sense that $F_{i,t} = argmin\, E(f(x_t) - y_{t+1})^2$ for f in a set which may not include $E(y_t|x_t)$. Without loss of generality, we consider reasonable behavioral models such that

$$\pi_{j,t} = (y_{t-1})^+ F_{i,t-1} \tag{3}$$

The constraint on the realized profit function that characterizes reasonable behavioral model is consistent with most of the literature, e.g. De Grauwe and Grimaldi (2006b) for a book length treatment. It measures if the forecast captures the correct sign, and the size of any economic gains, which are proportional to the actual forecast. The multinomial logit specification was originally proposed by Brock and Hommes (1997) in connection to a discrete rational choice between competing forecasting rules and the resulting complex system behavior. The requirement that forecasts are sensible is not emphasized, but has been the rule in empirical work, where first stage regressions implement sensible forecasting rules in restricted parametric sets. Except for the forecast dynamics, the model is classical, with optimal decision and market clearing.

Intuition

Our results follow from a somewhat trivial observation: single hidden layer neural networks may be written as a behavioral model, and vice versa. Consider the following neural network model $y_t = \beta_0 + \sum_{j=1..q} e^{x_t'\lambda_j}\beta_j$. We actually have to consider a threshold exponential, but let's focus on the exponential for the intuition. Dividing and multiplying the right hand side by $\sum_i e^{x_t'\lambda_i}$ we have a behavioral model with $\rho > 0$ arbitrary, $\pi_{j,t} = x_t'\lambda_j/\rho$ and $F_{j,t} = \beta_j \sum_i e^{x_t'\lambda_i} + \beta_0$. Of course, nothing guarantees $\pi_{j,t}$ and $F_{j,t}$ are related as required in a reasonable behavioral model, or that $F_{j,t}$ is a sensible forecast for that matter. Our estimation and testing procedure will try to make the neural network as close as possible to a reasonable model.

Notice that forecasts belong to a restricted set of neural networks and therefore cannot perform as well as the conditional expectation, since it is unrestricted. The approximation of the profit function by a linear function may be rather poor as well. Yet, the same argument above would apply to a two hidden layers neural network $y_t = \sum_{j=1..q} \exp\left(\beta_{j,0} + \sum_{i=1..q} e^{x'\lambda_{j,i}}\beta_{j,i}\right)\beta_j + \beta_0$. In this case, the forecasts would be restricted two hidden layers networks $\beta_j \sum_{j'=1..q} exp\left(\beta_{j',0} + \sum_{i=1..q} e^{x'\lambda_{j',i}}\beta_{j',i}\right) + \beta_0$ and the realized profits one hidden layer networks $\left(\beta_{j,0} + \sum_{i=1..q} e^{x'\lambda_{j,i}}\beta_{j,i}\right)/\rho$. Therefore, by increasing the number of layers, we allow for more flexibility in each of the components associated with the behavioral interpretation of the model. But we still have no guarantee the model is reasonable.

White (1990) shows that if we let network complexity (the integer q defining the number of hidden units) and coverage (norm of the βs and λs) grow at a well calibrated deterministic rate then the least squares estimator of the neural network converges in probability to the true conditional expectation. Since each neural network corresponds to a behavioral model, the same result applies. White (1990) also shows that if there is an error in obtaining the least squares estimator which is persistent even asymptotically, than the estimated neural network will be within a neighborhood of the true model. We show bellow that a restricted least squares estimator which imposes reasonable constraints, to the extent that it imposes an error from the point of view of the unrestricted problem, also defines a neighborhood of approximation.

Results

Assume (Y_t, X_t) is a strictly stationary mixing process in the bounded spaces $Y \times X$ of dimensions 1 and r respectively. The mixing condition can be either α or ϕ mixing with geometric rate (e.g., White (1990); see also Lindner (2009) for the case of conditional heteroscedasticity). Let Θ be the set of measurable functions $\theta: X \rightarrow R$ bounded in the norm $\int \theta(x)^2 \mu(dx)$, with μ the measure defined over X, and let ρ be the distance function with this norm. Suppose there is a unique $\theta_0 \in \Theta$ such that $E(Y_t|X_t) = \theta_0(X_t)$. Let $\theta \in \Theta(q, \Delta) \subset \Theta$ when

$$\theta(x) = \beta_0 + \sum_{j=1..q} \min\left(e^{(1,x)'\lambda_j}, 1\right)\beta_j \ \ with \sum_{j=0}^{q} |\beta_j| \leq \Delta, \sum_{j=0}^{q}\sum_{i=0}^{r} |\lambda_{ji}| \leq q\Delta \tag{4}$$

For any given $\theta \in \Theta(q, \Delta)$,

$$\theta(x) = \sum_{j=1..q} \frac{e^{\rho \pi_j(x)}}{\sum_{j=1..q} e^{\rho \pi_j(x)}} F_j(x)$$

$$\pi_j(x) = \left(I_{((1,x)'\lambda_j \leq \phi)}(1,x)'\lambda_j + I_{((1,x)'\lambda_j > \phi)}\phi\right)/\rho \tag{5}$$

$$F_j(x) = \beta_j \sum_i \min\left(e^{(1,x)'\lambda_j}, e^{\phi}\right) + \beta_0$$

define the associated behavioral model (θ, π, F, ρ) for some $\rho, \phi > 0$. The parameter ρ measures the sensitivity of agents to past profit when setting their forecasts. The parameter ϕ introduces an upper bound on the flexible function used to model the forecasting rules, which is a necessary feature to obtain our approximation results. Later in the paper we show how to estimate the parameter ρ, but the parameter ϕ is calibrated to ensure a good range of positive and negative values for the profit function.

Consider the *unconstrained estimator*

$$\hat{\theta}_T(q,\Delta) = \underset{\theta\in\Theta(q,\Delta)}{\operatorname{argmin}} \frac{1}{T}\sum_{t=1..T}\left(y_t - \theta(x_t)\right)^2 \tag{6}$$

According to Theorem 2.2 in White (1990), if we allow q_{T} and Δ_{T} to grow at a well calibrated deterministic rate in relation to the sample size, then $\rho\left(\hat{\theta}_T(q_T,\Delta_{\mathrm{T}}),\theta_0\right) \xrightarrow{p} 0$. This is true, in particular, if $\Delta_{\mathrm{n}} \to \infty, q_T \to \infty$ when $\mathrm{T} \to \infty$, such that $\Delta_{\mathrm{T}} = \mathrm{O}(\log \mathrm{T})$ and $q_T = o(T^{\alpha})$, $\alpha < 1/2$. In words, the least squares estimator consistently estimates the conditional expectation function. The least square estimator has associated parameter estimators $\hat{\beta}(q,\Delta)$ and $\hat{\lambda}(q,\Delta)$. From definitions in the previous paragraph, it also has an *associated behavioral model estimator* $\left(\hat{\theta}_T, \hat{\pi}_{\mathrm{T}}, \hat{F}_T, \rho, \phi\right)(q,\Delta)$ for an arbitrary $\rho, \phi > 0$. Therefore, we can restate White theorem in terms of behavioral models:

Proposition 1 [White (1990), Theorem 2.2]: If q_{T} and Δ_{T} grow at a well calibrated deterministic rates in relation to the sample size, then the sequence of associated behavioral models $\left(\hat{\theta}_T, \hat{\pi}_{\mathrm{T}}, \hat{F}_T, \rho, \phi\right)(q_T,\Delta_{\mathrm{T}})$ is such that $\rho\left(\hat{\theta}_T(q_T,\Delta_{\mathrm{T}}),\theta_0\right) \xrightarrow{p} 0$.

Remark 1. The proposition is a direct consequence of the stated theorem once we show that the threshold exponential $\psi(z) = \min(\exp(z), \exp(\phi))$ is a bounded squashing function that satisfies a Lipchitz condition. A squashing function is monotonic function from the real numbers to [0,1] with finite limits at minus and plus infinity, and this is evident in the threshold exponential. The exponential function is also known to respect a Lipchitz condition in intervals, and the threshold guarantees the condition generally.

Now consider the following *constrained estimator*

$$\tilde{\theta}_T(q,\Delta,\delta) = \underset{\theta\in\Theta(q,\Delta),\rho>0}{\operatorname{argmin}} \frac{1}{T}\sum_{t=1..T}\left(\hat{\theta}_T(x_t) - \theta(x_t)\right)^2 \; s.a.: \\ \frac{1}{2Tq}\sum_{t=1..T,j=1..q}\left[\left(\pi_{\mathrm{j}}(x_t) - (y_{t-1})^{+}F_j(x_{t-1})\right)^2 + \left(y_t - F_j(x_{t-1})\right)^2\right] \le \delta, \tag{7}$$

The idea here is to estimate an *approximately reasonable* behavioral model, using the sum of squared deviation of the exact conditions defining a reasonable model. As we decrease δ we impose tighter constraints. From the perspective of the unconstrained estimator, the constrained estimator includes an error, the average size of which is

$$d_T(\delta) = \frac{1}{T}\sum_{t=1..T}\left(\hat{\theta}_T(x_t) - \tilde{\theta}_T(x_t)\right)^2 \tag{8}$$

The error is bounded above by $T^{-1}\sum_{t=1..T}\hat{\theta}_T(x_t)^2$ and bellow by zero. We may choose the sequences q_T and Δ_T so that $d_T(\delta)$ has a limit. Unless the true model satisfies the constraint, we have $\lim d_T(\delta) = d(\delta) > 0$. By an immediate application of a theorem in White (1990) we have the following result:

Proposition 2 [White (1990), Theorem 3.5]: If $\hat{\theta}_T$ is a consistent estimator, then $\tilde{\theta}_T(\delta)$ will be within neighborhood of size $d(\delta)$ of the true conditional expectation, using the distance metric ρ defined before.

Remark 2. The proposition is a direct consequence of the stated theorem, since the sequence $d_T(\delta)$ built in the previous paragraph satisfies the necessary assumptions (Assumption B.4 in White (1990)).

For a sufficiently large sample, $d_T(\delta)$ estimates the size of the neighborhood. Since $d_T(\delta)$ increases for tighter constraints, they are associated with less reasonable models. The important empirical question is: how far should we go in the direction of reasonable models? We propose to select the most reasonable model compatible with the data. Therefore we must have a procedure to test the constraint, which is the subject of the next section.

It is important to observe that the behavioral model associated with a neural network imposes strong functional form restrictions. As argued in the Appendix, we may extend our results to neural networks with more than one hidden layer, therefore increasing the flexibility of these functions. But it is clear we cannot rule out that still more flexible functional forms would provide better approximating properties[1].

[1] For example, if forecasts are single layer neural networks and profits are reasonable, the conditional expectation is a two layer network with time varying coefficients, which appears very flexible but has unknown approximation properties.

Also notice we may impose stronger constraints. For instance, if $x_t = (z_t, z_{t-1}, y_{t-1}, y_{t-2})$, we may consider restricted forecasts conditional on (z_t, y_{t-1}), and lagged restricted forecasts on (z_{t-1}, y_{t-2}). This would give the unrestricted model a better chance of approximating the restricted model as far as the information structure is concerned. Allowing more network complexity in the constrained estimator may have the same effect. It is also possible to introduce several constraints instead of an average of the reasonable constraints. This would likely result in less reasonable models and wider neighborhoods. The constraints might also receive different weights if the researcher believes some properties defining reasonable models to be more important than others for theoretical or empirical reasons.

3. Testing reasonable constraints

In this section we propose a test for the null hypothesis that the conditional expectation respects reasonable constraints, that is $H_0(\delta): d(\delta) = 0$, against the alternative $H_1(\delta): d(\delta) > 0$. Following Kreiss, Neumann and Yao (2008), we consider the L_2-distance between the unconstrained and constrained estimator. That is, the distance $d_T(\delta)$ defined in equation (8) above is the proposed test statistic. We consider the following nonparametric wild bootstrap algorithm to evaluate the p-value:

1. Generate the wild bootstrap residuals $\{\varepsilon_t^*\}_{t=1}^T$ from $\varepsilon_t^* = \hat{\varepsilon}_t \eta_t$, where η_t is a sequence of i.i.d. random variables with zero mean and unit variance, $\hat{\varepsilon}_t = \ y_t - \hat{\theta}_T(x_t),$ and such that $y_t^* = \ \tilde{\theta}_T(x_t) + \varepsilon_t^*$.
2. Calculate the bootstrap test statistic $d_T^*(\delta)$ on the sample $\{y_t^*, x_t\}_{t=1}^T$.
3. Reject the null hypothesis $H_0(\delta)$ if $d_T(\delta)$ is greater than the upper-α point of the conditional distribution of $d_T^*(\delta)$ given $\{y_t, x_t\}_{t=1}^T$.

Note that the bootstrap algorithm uses the residuals from the unconstrained nonparametric fit. This is in accordance with Hall and Wilson (1991), since it results in consistent estimates of residuals under both the null and the alternative hypothesis. Also note that the wild bootstrap is able to account for dependency and conditional heteroscedasticity common in exchange rate data.

To investigate how far should we go in the direction of reasonable models, we consider multiple tests over a grid $\{\delta_n\}_{n=1}^{N}$ indexing different null hypothesis. For example, $\delta_n = \delta_0 - n\delta_0/N$ with δ_0 of the constraint evaluated at the unconstrained estimator. To control for false discoveries rates, we propose to use the Benjamini and Hochberg (1995) testing procedure at level α. That is, let $p_{(1)} \leq p_{(2)} \leq .. \leq p_{(N)}$ be the ordered p-values. Find the highest n^* such that $p_{(n^*)} \leq n^*\alpha/N$, and reject the null for all the ranks $n \leq n^*$.

4. Estimation of behavioral models

As mentioned in the introduction, estimation behavioral models has been usually performed in two stages, first estimating reasonable forecasting rules and then substituting into the full model for non linear estimation. Our proposed estimator proceeds in a single step by imposing approximately reasonable constraints. The estimator also avoids the severe numerical problems associated with least squares optimization of neural networks by relying on an approximate optimization, as described in this section.

The *unconstrained neural network* and the *associated behavioral model* may be estimated by the approximate optimization algorithm proposed by White (2006). The algorithm proceeds from specific to general as follows: (i) start from a random sample of hidden units (λs) in a bounded support; (ii) include hidden units one at a time so as to minimize the chance of rejection in a misspecification test based on neglected non linearity (Bierens (1990), Stinchcombe and White (1998)); (iii) for each additional unit, set the βs coefficients by least squares; (iv) select the final number of hidden units based on the minimum cross validated mean squared error in a “hv-block” design (Racine (2000)). The algorithm maximizes the chance of capturing relevant non linear effects but also minimizes the chance of over-fitting. Notice that the data driven cross-validation procedure also provides consistent approximation of bounded continuous functions (White (1990), Theorem 3.4), and is therefore equivalent to the deterministic setting of network complexity growth rates.

To estimate the *constrained behavioral model* we first define the optimization over the β and ρ coefficients. It is convenient to adopt a matrix notation. Let

$$X = \left(x, \psi(\Gamma_1), \dots, \psi(\Gamma_q)\right)$$

$$W_j = (Ly^{+}x, \dots, Ly^{+}\Sigma, \dots) \text{ with } Ly^{+}\Sigma \text{ in colum j + n, and zero otherwise}$$

$$V_j = (Lx, \dots, L\Sigma, ..) \text{ with } L\Sigma \text{ in colum j + n, and zero otherwise}$$

where x is either the unity column vector or the matrix of stacked x_t in case there are common linear terms in the forecasts (see Section 5), n is the number of columns in x, y is the vector of stacked y_t, L is the lag operator, $\Gamma_j = \left(\Gamma_{j,1}, .., \Gamma_{j,T}\right)'$, $\Gamma_{j,t} = (1, x_t)'\lambda_j$, $\Sigma = \sum_j \psi(\Gamma_j)$, and ψ is defined in Remark 1. Then the constrained optimization over β and ρ for a given network structure and hidden units is equivalent to the following penalized unconstrained optimization, as long as λ is set appropriately:

$$\min_{\beta} (\hat{y} - X\beta)^2 + \lambda/N \sum_j \left[\left(\pi_j - W_j\beta\right)^2 + \left(y - V_j\beta\right)^2\right]$$

with N the number of items in the sum so as to express it as an average, $\hat{y}$ the unconstrained fit, $\pi_j = \left(I_{(\Gamma_{j,t} \le \phi)}\Gamma_{j,t} + I_{(\Gamma_{j,t} > \phi)}\phi\right)/\rho \triangleq \bar{\bar{\pi}}_j/\rho$ the vector with the evaluated profits. The first order condition for β and is ρ are, respectively,

$$X'\breve{y} + \lambda/N \sum_j \left[W_j'\bar{\bar{\pi}}_j/\rho + V_j'y\right] = \left(X'X + \lambda/N \sum_j \left[W_j'W_j + V_j'V_j\right]\right)\beta,$$

and

$$\frac{1}{\rho} = \frac{\sum_j \bar{\bar{\pi}}_j'W_j}{\sum_j \bar{\bar{\pi}}_j'\bar{\bar{\pi}}_j}\beta.$$

Substituting the second on the first, we have

$$X'\breve{y} + \lambda/N \sum_j V_j'y = \left(X'X + \lambda/N \sum_j \left[W_j'W_j + V_j'V_j\right] - \lambda/N \frac{\left[\sum_j \bar{\bar{\pi}}_j'W_j\right]^2}{\sum_j \bar{\bar{\pi}}_j'\bar{\bar{\pi}}_j}\right)\beta$$

such that we may obtain the solution for β when the right hand side matrix inside the parenthesis is invertible. The solution for ρ is then easy to obtain - it will be positive for proper values of the parameter ϕ defining the range of the profit function.

With this definition, a simple estimator for the *constrained behavioral model* is to proceed as in the unconstrained case, but with the following additional step: (v) perform constrained or penalized optimization over β and ρ. One may also consider searching over the sample of λs for further gains, but we do not recommend it since hidden unit loadings are rough approximations anyway. By fixing network complexity and hidden units loadings we focus the comparison on the constraints.

Although we could have used nonlinear least squares on the full set of parameters of the sieve, such an optimization is computationally hard and the fit of the candidate solution obtained by such method is often worse than the approximate solution of the method proposed here (see White (2006) for Monte Carlo experiments). The algorithm proposed here could be used as a generator of starting values for optimization on the full parameter space, since this is the major difficulty for this class of objective functions. In the context of testing for reasonable models and the large bootstrap experiments necessary to implement it, this additional step would probably be too costly to be feasible. Moreover, notice that the consistency results from Section 2 would still be valid under approximate optimization, although the neighborhood around the true model should be extended by the size of the optimization errors. To the extent that such errors are of similar magnitudes in constrained and unconstrained estimation the distance $d(\delta)$ in Proposition 2 can be interpreted as the *increase* in the size of the neighborhood of approximation implied by optimization errors.

5. Application: Brazilian foreign exchange market and net order flow

Our database begins in January 2002 and ends in November 2012. The series are sampled at a daily frequency. The order flow variable is from the Central Bank of Brazil electronic records of private spot transactions intermediated by financial institutions and covers the entire spot foreign exchange market. The set of conditioning variables is taken from the empirical literature of the BRL/USD market (e.g. Kolsheen (2013) and Barroso (2014)), and includes the CRB commodity price index, the VIX implicit volatility index, the DOL dollar index and the EMBI spread. Except for order flow, all the variables are measured in first difference.

Due to regulatory constraints, the set of intermediaries allowed to participate in the spot market is very restricted and includes only major financial institutions. Information assumptions are therefore particularly important. Each institution collects partial information on their clients order flow. However, from this and other partial information they may infer, to some degree of accuracy, the aggregate net order flow in the market. For this reason, it is not straight forward how to incorporate aggregate order flow in the model. Our baseline specification includes net order flow as a linear term. The extended specification includes it into the non linear approximation which is supposed to capture expectation formation dynamics. This captures two limiting assumptions of no information and full information of aggregate order flow. As for the timing of covariates, market participants usually have early information of international exogenous variables from other markets, so we include them contemporaneously.

For our application, instead of a scalar β_0 as in (4)-(5), we substitute it by $x_t{}'\beta_0$ in the formulas, with vector β_0 and a constant included in x_t. Accordingly, we write $x'\lambda_i$ in place of $(1,x)'\lambda_i$. This does not influence at all the approximation theorems, but avoids the cumbersome approximation of a linear part by a complex network. It does imply forecasts have a common linear term, and therefore puts all the weight of disagreement on the flexible nonlinear part of the expectation formation process.

To implement the computationally efficient estimators of Section 4, we fix $\phi = 4$, and sample 2000 vectors of λs from the uniform distribution on (-1,1), except for the constant, which we sample from (1,2) to ensure a good range for the profit function[2]. The interval seems appropriate given the sample of 2480 daily observations and the scale of the variables. Following recommendations in Racine (2000), the hv-block cross validation uses $h = 7$ and $v = 700$. For simplicity, the specification selected for the unconstrained case is the same for all in sample and bootstrap estimators. By fixing network complexity we focus the comparison on the constraints.

Baseline model. The baseline model excludes order flow from the nonlinear term. The best cross validation performance in the unconstrained case is obtained with network complexity $q = 30$. That is, there are 30 types of forecasting rules in use by economic agents. The results are summarized in Table 1. To investigate how far should we go in the direction of reasonable models, we consider multiple tests over a grid of null hypothesis. Each null is defined by the δ defining the constraint in (7) and we report this as a percentage of the constraint evaluated at the unconstrained estimation. The test statistic is the L2 distance between the estimators defined in (8). The bootstrap p-values were obtained by the procedure described in Section 3. Boldface p-values are significant at 5% controlling for false discovery rate as described in Section 3.

[2] We experimented with several intervals and values for phi. The final choice was selected by comparing the performance, given we select the number of hidden units for each alternative by cross validation.

According to the results reported in Table 1, the constraints can be tightened at most by 95.7% in the direction of reasonable models. The value of the constraint evaluated at the unconstrained estimator is a first estimate of how far the data is from a reasonable model. That we can tighten this by 95.7% without compromising the fit of the model is an economically significant result.

Table 1. Testing reasonable behavioral constraints: baseline/1

Covariates: CRB, DOL, VIX, EMBI, OFLOW

100-delta /2	21.76	79.89	87.26	95.24	95.69	99.91
test statistic /3	0.004	0.008	0.010	0.020	0.022	0.123
p-value /4	0.822	0.480	0.392	0.124	0.066	**0.006**

Notes:
/1 Excludes order flow inside hidden units; 30 hidden units selected by cross validation.
/2 Delta measured as % of the constraint evaluated at unconstrained estimator
/3 Defined as the L2 distance between unconstrained and constrained estimator
/4 Bold if rejects constrained model at 5% controlling for false discovery rate

Extended model. The extended model has order flow in both linear and non linear terms, so as to capture possible knowledge of aggregate order flow in the expectation formation process. The best cross validation performance in the unconstrained case is obtained with network complexity of $q = 17$. That is, there are 17 types of forecasting rules used by economic agents. The results are summarized in Table 2. As before, boldface p-values are significant at 5% controlling for false discovery rate as described in Section 3. According to the results, the constraints can be tightened at most by 95.6% in the direction of reasonable models, which is close to the baseline.

Table 2. Testing reasonable behavioral constraints: extended/1

Covariates: CRB, DOL, VIX, EMBI, OFLOW

100-delta /2	30.89	43.92	74.66	83.97	95.61	96.33
test statistic /3	0.004	0.005	0.010	0.017	0.041	0.079
p-value /4	0.281	0.435	0.588	0.270	0.098	**0.002**

Notes:
/1 Includes order flow inside hidden units; 17 hidden units selected by cross validation.
/2 Delta measured as % of the constraint evaluated at unconstrained estimator
/3 Defined as the L2 distance between unconstrained and constrained estimator
/4 Bold if rejects constrained model at 5% controlling for false discovery rate

When testing for reasonable constraints, we rely on the approximation properties of the individual components in the associated behavioral model, that is, on the approximation properties of profit and forecasting functions. Our framework allows for more flexibility in such components by increasing the number of hidden units. There are other alternatives, such as allowing for some time variation in the coefficients in the neural network. In any case, our empirical application did not explore more flexibility in the representation of forecasts and profit functions beyond the level afforded by single layer networks. In this sense, our results provide only a lower bound on how far we may go in the direction of reasonable models in this particular foreign exchange market.

6. Conclusion

We devise approximation, estimation and testing results for behavioral models without restricting the set of forecasting rules beyond some very weak reasonable constraints. This is made possible by the analogy between behavioral models and neural networks, with some adaptation to incorporate reasonable properties of the forecasting rules and performance evaluation metrics used by economic agents. In this context, the important empirical question is: how far should we go in the direction of reasonable models? We propose to select the most reasonable model compatible with the data by multiple tests over a grid of such reasonable models. To implement such testing procedure we also develop computationally efficient estimators for constrained and unconstrained behavioral models which do not require first step estimation of different forecasting models for the exchange rate process.

We apply the proposed testing framework to a unique dataset for the Brazilian foreign exchange market with full records of net order flow intermediated by the financial system. The results support tightening of constraints in the direction of reasonable models by 96%. This result is robust to alternative assumptions regarding private information of economic agents with respect to order flow. As noted in the previous section, our empirical application explores only single layer networks, although more layers would increase the flexibility. In this sense, the result is a lower bound on how far we may go in the direction of reasonable models.

References

Barroso, João Barata. 2014. Realized Volatility as an Instrument to Official Intervention. *Banco Central do Brasil Working Papers Series*

Bénassy-Quéré, Agnès; Larribeau, Sophie and Macdonald, Ronald. 2003, Models of exchange rate expectations: how much heterogeneity? *Journal of International Financial Markets, Institutions and Money*, 13 (2)

Beine, Michael; De Grauwe, Paul and Grimaldi, Marianna. 2009. The impact of FX central bank intervention in a noise trading framework. *Journal of Banking & Finance* 33 (7)

Benjamini, Yoav and Yosef Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Joural of the Royal Statistical Society, B, 57

Brock, William and Cars Hommes. 1997. A Rational Route to Randomness. *Econometrica*, 68 (5)

Chen, Xiaohong and Halbert White. 1999. Improved Rates and Asymptotic Normality for Nonparametric Neural Network Estimators. *IEEE Transactions on Information Theory*, 45, 682-691

Cheung, Yin-Wong and Menzie Chinn. 2001. Currency traders and exchange rate dynamics: a survey of the US market. *Journal of International Money and Finance* 20(4)

Chiarella C, He XZ, Zheng M. 2011. Heterogeneous expectations and exchange rate dynamics. *The European Journal of Finance*

De Grauwe, Paul and Grimaldi, Marianna, 2006a. Exchange rate puzzles: A tale of switching attractors. *European Economic Review,* 50

De Grauwe, Paul and Marianna Grimaldi. 2006b. The Exchange Rate in a Behavioral Finance Framework. *Princeton University Press*, Princeton, New Jersey.

Ellen, Saskia, Willem Vershoor and Remco Zwinkels. 2013. Dynamic expectation formation in the foreign exchange market. *Journal of International Money and Finance* 37

Evans, Martin and Richard Lyons. 2002. Order Flow and Exchange Rate Dynamics. *Journal of Political Economy,* 110 (11)

Frankel, Jeffrey and Kenneth Froot. 1990. Chartist, Fundamentalists, and Trading in the Foreign Exchange Market. *The American Economic Review*, 80 (2)

Frydman, Roman, and Michael D. Goldberg. 2007. Imperfect knowledge economics: Exchange rates and risk. *Princeton University Press*, Princeton, New Jersey.

Kohlsheen, Emanuel. 2013. Order Flow and the Real: Indirect Evidence of the Effectiveness of Sterilized Interventions. BIS Working Paper Series 426.

Hall, Peter and Susan Wilson. 1991. Two guidelines for bootstrap hypothesis testing. *Biometrics* 47

Humpage, Owen. 2003. Government intervention in the foreign exchange market. *Federal Reserve Bank of Cleveland. Working Paper* 03-15.

Ito, Takatoshi. 1990. Foreign exchange rate expectations: Micro survey data. *NBER Working paper* 2679

Jong, Eelke; Verschoor, Willem and Zwinkels, Remco. 2010. Heterogeneity of agents and exchange rate dynamics: Evidence from EMS. *Journal of International Money and Finance*, 29 (8)

Kreiss, Jens-Peter; Michael Neumann and Qiwei Yao. 2008. Bootstrap tests for simple structures in nonparametric time series regression. *Statistics and Its Interface*, 1 (2)

Kuan, Chung-Ming and Halbert White. 1994. Artificial neural networks: an econometric perspective. *Econometric Reviews*, 13 (1)

White, Halbert. 1990. Connectionist Nonparametric Regression: Multilayer Feedforward Networks Can Learn Arbitrary Mappings. *Neural Networks*, 3.

White, Halbert. 2006. Approximate Nonlinear Forecasting Methods. *Handbook of Economic Forecasting*, 1, 459-512

Lindner, Alexander. 2009. Stationarity, mixing, distributional properties and moments of GARCH (p,q) processes. *Handbook of financial time series*, 43-69

Manzan, Sebastiano and Frank Werterholff. 2007. Heterogeneous expectations, exchange rate dynamics and predictability. *Journal of Economic Behavior & Organization, 64 (1), 111-128*

Menkloff, Lukas and Taylor, Mark. 2007. The obstinate passion of foreign exchange professionals: technical analysis. *Journal of Economic Literature* 45

Reiner, Franke. 2009. Applying the method of simulated moments to estimate a small agent-based asset pricing model. *Journal of Empirical Finance* 16 (5), 804-815

Vitale, Paolo. 2007. A guided tour of the market microstrucutre approach to exchange rate determination. *Journal of Economic surveys*, 21 (5).

Appendix

In this appendix we show how to extend the propositions to many hidden layers neural networks. From Theorem 4.1 in White (1990), consistency results apply for a two hidden layer neural network, for well defined deterministic rates. Let the threshold exponential be denoted by $\psi(z) = \min(\exp(z), \exp(\phi))$. Assume for simplicity the same complexity q and coverage Δ for both hidden layers. Let $\theta \in \Theta'(q, \Delta) \subset \Theta$ when

$$\theta \in \Theta: \theta(x) = \beta_{0,0} + \sum_{j=1..q} \psi\left(\beta_{j,0} + \sum_{i=1..q} \psi\big((1,x)'\lambda_{j,i}\big)\beta_{j,i}\right)\beta_{0,j}$$

$$with \sum_{i=0}^{q} |\beta_{j,i}| \leq \Delta, \sum_{i=0}^{q}\sum_{h=0}^{r} |\lambda_{j,i,h}| \leq q\Delta \text{ for j} = 0..\text{J}$$

For a given $\theta \in \Theta'(q, \Delta)$, we can always write

$$\theta(x) = \sum_{j=1..q} \frac{e^{\rho\pi_j(x)}}{\sum_{j=1..q} e^{\rho\pi_j(x)}} F_j(x)$$

$$\pi_j(x) = \left[I_j(x)\left(\beta_{j,0} + \sum_{i=1..I} \psi\big((1,x)'\lambda_{j,i}\big)\beta_{j,i}\right) + \big(1 - I_j(x)\big)\phi\right]/\rho$$

$$F_j(x) = \beta_{0,0} + \beta_{0,j} \sum_{j'=1..q} \psi\left(\beta_{j',0} + \sum_{i=1..I} \psi\big((1,x)'\lambda_{j',i}\big)\beta_{j',i}\right)$$

$$I_j(x) = I_{\left(\beta_{j,0} + \sum_{i=1..I} \psi((1,x)'\lambda_{j,i})\beta_{j,i} \leq \phi\right)}$$

The unconstrained and constrained estimators are defined as before but now over $\Theta'(q, \Delta)$. The test statistic for the constraints is also the same as before. For estimation we could still sample the lambdas randomly, but we would still face non linear optimization in the beta parameters, with associated numerical problems. A possible procedure would be to sample second layer betas as well, maybe around the one hiddn layer estimates, and then select the set of second layer parameters to capture the most non linearity. For the two layer estimator to have a more meaningful difference from the one layer case, one may consider a final stage non linear least squares optimization in the enlarged set of betas.