Economics Letters 124 (2014) 104–107

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

A note on the representative adaptive learning algorithm

Michele Berardi [a], Jaqueson K. Galimberti [b,*]

[a] Economics, School of Social Sciences, The University of Manchester, M13 9PL, Manchester, UK

[b] KOF Swiss Economic Institute, ETH Zurich, 8092 Zurich, Switzerland

HIGHLIGHTS

- Learning algorithms are assumed to represent agents learning-to-forecast behavior.
- Main algorithms in the literature: Least Squares (LS) and Stochastic Gradient (SG).
- We compare the forecasts associated with the LS and the SG algorithms.
- We use US real-time data on inflation and output growth.
- Our results favor the use of the Least Squares algorithm as representative.

ARTICLE INFO

Article history:
Received 20 February 2014
Received in revised form
17 April 2014
Accepted 26 April 2014
Available online 4 May 2014

JEL classification:
C53
D83
D84
E03
E37

Keywords:
Expectations
Learning algorithms
Forecasting
Learning-to-forecast
Least squares
Stochastic gradient

ABSTRACT

We compare forecasts from different adaptive learning algorithms and calibrations applied to US real-time data on inflation and growth. We find that the Least Squares with constant gains adjusted to match (past) survey forecasts provides the best overall performance both in terms of forecasting accuracy and in matching (future) survey forecasts.

1. Introduction

Adaptive learning algorithms have been proposed to provide an alternative to, and a justification for, rational expectations (RE) equilibria in macroeconomics (Evans and Honkapohja, 2001). Going beyond the RE hypothesis, however, comes at the cost of introducing another degree of freedom in macroeconomic modeling, since one has to be specific about which algorithm is assumed to represent agents behavior.

The usual choice for this purpose has been the Least Squares (LS) algorithm (Branch and Evans, 2006; Markiewicz and Pick, 2014), possibly due to its widespread popularity between econometricians. A computationally simpler alternative is offered by the Stochastic Gradient (SG) algorithm (Barucci and Landi, 1997; Evans and Honkapohja, 1998). We argue that the previous literature has neglected the need of a realistic justification in the choice of the representative learning algorithm.

Importantly, theoretical analyses of learning convergence have shown that these learning algorithms may lead to different learnability conditions of RE equilibria (Heinemann, 2000; Giannitsarou, 2005). The LS dominance also has been challenged in previous

* Correspondence to: KOF Swiss Economic Institute, ETH Zurich, WEH D4, Weinbergstrasse 35, 8092 Zurich, Switzerland. Tel.: +41 44 632 8529; fax: +41 44 632 1150.

E-mail addresses: jakaga2002@yahoo.com.br, galimberti@kof.ethz.ch (J.K. Galimberti).

http://dx.doi.org/10.1016/j.econlet.2014.04.028

applied studies (see Bullard and Eusepi, 2005; Carceles-Poveda and Giannitsarou, 2007). Hence, it remains open the question of which algorithm should be taken as representative from an empirical standpoint. Our main contribution is therefore an attempt to fill that gap, and we do this by comparing the quality and fit to surveys of the forecasts associated with each of these learning algorithms.

We estimate Vector Autoregressions (VARs) recursively with real-time quarterly data on US inflation and output growth, and then compare the associated multi-horizon forecasts over an evaluation sample from 1981q1 to 2011q4. Details of our approach are presented in Section 2. Our results, presented in Section 3, are favorable to the use of the LS as representative of agents learning-to-forecast behavior for the growth variable, whereas for inflation we obtained mixed evidence depending on the calibration of the learning gains. Namely, the LS dominance is weakened when the learning gains are calibrated so as to minimize observed squared forecasting errors rather than their distance to survey forecasts. We discuss these results in Section 4.

2. Approach

Our approach is based on learning-to-forecast exercises that mimic the real-time environment faced by an economic agent when forming expectations on inflation (π_t) and output growth (g_t). We assume that this agent attempts to construct inferences about these variables estimating a VAR of the form

$$y_{i,t} = \mathbf{x}_t' \boldsymbol{\theta}_{i,t} + \varepsilon_{i,t}, \tag{1}$$

where $y_{1,t} = \pi_t$ and $y_{2,t} = g_t$, $\mathbf{x}_t = (1, \pi_{t-1}, \dots, \pi_{t-p}, g_{t-1}, \dots, g_{t-p})'$, $\boldsymbol{\theta}_{i,t} = \left(\theta_{0,i,t}, \theta_{1,i,t}, \dots, \theta_{p,i,t}, \theta_{p+1,i,t}, \dots, \theta_{2p,i,t}\right)'$, p denotes the VAR lag order, and $\varepsilon_{i,t}$ is a white noise disturbance. To estimate each equation's vector of coefficients, $\boldsymbol{\theta}_{i,t}$, we follow the adaptive learning literature and adopt the LS and the SG specifications.

Algorithm 1 (*LS*)**.** Under the estimation context of (1), the *LS* algorithm assumes the form of

$$\hat{\boldsymbol{\theta}}_{i,t}^{LS} = \hat{\boldsymbol{\theta}}_{i,t-1}^{LS} + \gamma_t \mathbf{R}_t^{-1} \mathbf{x}_t \left(y_{i,t} - \mathbf{x}_t' \hat{\boldsymbol{\theta}}_{i,t-1}^{LS} \right), \tag{2}$$

$$\mathbf{R}_t = \mathbf{R}_{t-1} + \gamma_t \left(\mathbf{x}_t \mathbf{x}_t' - \mathbf{R}_{t-1} \right), \tag{3}$$

where γ_t is a learning gain parameter, and $\mathbf{R}_t$ stands for an estimate of regressors matrix of second moments, $E\left[\mathbf{x}_t \mathbf{x}_t'\right]$.

Algorithm 2 (*SG*)**.** Under the estimation context of (1), the *SG* algorithm is given by

$$\hat{\boldsymbol{\theta}}_{i,t}^{SG} = \hat{\boldsymbol{\theta}}_{i,t-1}^{SG} + \mu_t \mathbf{x}_t \left(y_{i,t} - \mathbf{x}_t' \hat{\boldsymbol{\theta}}_{i,t-1}^{SG} \right), \tag{4}$$

with μ_t standing for the learning gain parameter.

We then use the LS and the SG algorithms to obtain recursive estimates of the parameters of VAR model specifications (1–4 lag orders) applied to real-time quarterly data on US real GNP/GDP and its price index from 1947q2 to 2011q4. Our data on these series come from the Philadelphia's Fed Real-Time Data Research Center and consist of vintages from 1966q1 to 2012q1, i.e., a total of 185 snapshots of what was known on these variables by a market participant in real-time (see Stark and Croushore, 2002). For the purpose of comparing the algorithms' forecasts to those provided by survey respondents, we use data from the Survey of Professional Forecasters (SPF). Here we use the median of the individual forecasts made for a total of five horizons, namely from t (nowcast) to $t+4$. The SPF data is available from 1968q4 onwards, and, consistent to our data on actuals, the last survey data we use is that of 2010q4, which contain forecasts up to 2011q4.

Operation of these algorithms requires the specification of a (sequence of) gain value(s) determining how quickly some given information is incorporated into the algorithm's coefficients estimates. Recognizing the prominent role that the learning gains have in determining the statistical properties of the estimates associated with each algorithm (see, e.g., Benveniste et al., 1990), here we follow the calibration approach proposed in Berardi and Galimberti (2014). Particularly, we distinguish between two gain determination rationales: as a *choice* of rational agents, selecting the gains that minimize the (average) squared forecasting errors over a given window of observations; and as a *primitive* parameter of agents behavior, where the gains are selected so as to minimize the distance of the algorithms' forecasts to those collected through survey forecasts. Regarding the windows used to select the gains according to the criteria above we adopt two alternatives: a *fixed* and a *time-varying* gain calibration. Under the *fixed* calibration we pick the gain evaluating the corresponding criterion over the full sample of forecasts that we have computed, and keep this gain fixed throughout our exercise. For the *time-varying* calibration, in contrast, we use a rolling window sample of 60 forecasts to evaluate each gain determination criterion, hence selecting a new gain for every iteration on the real-time learning process. In both cases, the set of admissible gains is based on a grid of 100 values constructed taking an upper bound, experimentally computed to ensure algorithms' stability, as reference.

Our design unfolds into a three-stages routine to generate the forecasts associated with each learning algorithm: initialization, estimation and forecasting, and evaluation. The first 75 observations in our sample (up to 1965q4) are used for the smoothing-based initialization of the algorithms (following Berardi and Galimberti, 2012). The next 60 observations (from 1966q1 to 1980q4) are used for the algorithms' (first) *time-varying* calibration. Therefore, our evaluation sample corresponds to the period from 1981q1 to 2010q4. To match the timing of information in the SPF data set, we compute and evaluate forecasts over five horizons, each of these with its own instance of gain calibrations.

3. Results

We start looking over the forecasts associated with each algorithm and gain value included in the grid computations. Fig. 1 presents surfaces of average past performance for each algorithm and variable, showing their evolution through time and for the different gain values. Two main observations arise: (i) the behavior of each algorithm depends on the variable being forecasted, whereas for a given variable the LS and SG algorithms behave differently; (ii) the magnitudes of forecast errors were relatively higher during the first decade in our sample, irrespective of the variable forecasted and the algorithm used, an observation that can be associated with the period of greater volatility that preceded the Great Moderation in the US economy (see Stock and Watson, 2003).

To obtain a relative assessment of the learning mechanisms we now conduct two evaluation exercises comparing the forecasts associated with the learning algorithms with respect to their *accuracy* and their *resemblance* to the survey forecasts. We check for the statistical significance of these paired comparisons using tests common to the literature on forecast evaluation: the Diebold and Mariano (1995) (DM) test for equal (unconditional) predictive ability, and its more recently developed conditional counterpart test of Giacomini and White (2006) (GW). Our coverage of multiple forecasting horizons and VAR lag order specifications, for robustness, requires performing a high quantity of such comparisons.[1] Hence, to synthesize these evaluations we adopt

[1] To be specific, 40 for each pair of algorithms/calibrations: 5 horizons $\times$ 4 VARs $\times$ 2 evaluation criteria. An online Appendix (see Appendix A) is provided with the individual comparison results, and some descriptive statistics for each series of forecasts.