Metadata of the chapter that will be visualized in SpringerLink

Book Title	Artificial Life and Evolutionary Computation
Series Title	
Chapter Title	Calibrating Dynamic Factor Models with Genetic Algorithms
Copyright Year	2018
Copyright HolderName	Springer International Publishing AG, part of Springer Nature

Corresponding Author	Family Name	**Della Marra**
	Particle	
	Given Name	**Fabio**
	Prefix	
	Suffix	
	Role	
	Division	
	Organization	European Centre for Living Technology
	Address	Ca' Minich, S. Marco 2940, 30124, Venice, Italy
	Division	Department of Environmental Sciences, Informatics and Statistics
	Organization	Ca' Foscari University of Venice
	Address	Cannaregio 873, 30121, Venice, Italy
	Email	fabio.dellamarra@unive.it
	ORCID	http://orcid.org/0000-0001-7976-1387

Abstract

In this work, we address the problem of calibrating dynamic factor models for macroeconomic forecasting. The variables upon which the forecasts are computed are the logarithm of the Industrial Production (IP) and the yearly change of the logarithm of the Consumer Price Index (CPI). Our purpose is to provide a contribution to the model identification by proposing a new kind of calibration of static and dynamic factor models. The innovative part of our work consists of building a genetic algorithm for calibrating three dynamic factor models. We first analyse a dataset of 176 EU macroeconomic and financial time series and then we conduct the same study on a dataset of 115 US macroeconomic and financial time series. In both studies, the employment of genetic algorithm in the calibration procedure produces very good results and more significant than those achieved in similar studies, such as [1, 2].

Keywords (separated by '-')

Macroeconomic time-series forecasting - Genetic algorithms - Dynamic factor models

Calibrating Dynamic Factor Models with Genetic Algorithms

Fabio Della Marra[1,2](✉)

[1] European Centre for Living Technology, Ca' Minich, S. Marco 2940, 30124 Venice, Italy
fabio.dellamarra@unive.it
[2] Department of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, Cannaregio 873, 30121 Venice, Italy

Abstract. In this work, we address the problem of calibrating dynamic factor models for macroeconomic forecasting. The variables upon which the forecasts are computed are the logarithm of the Industrial Production (IP) and the yearly change of the logarithm of the Consumer Price Index (CPI). Our purpose is to provide a contribution to the model identification by proposing a new kind of calibration of static and dynamic factor models. The innovative part of our work consists of building a genetic algorithm for calibrating three dynamic factor models. We first analyse a dataset of 176 EU macroeconomic and financial time series and then we conduct the same study on a dataset of 115 US macroeconomic and financial time series. In both studies, the employment of genetic algorithm in the calibration procedure produces very good results and more significant than those achieved in similar studies, such as [1,2].

Keywords: Macroeconomic time-series forecasting
Genetic algorithms · Dynamic factor models

1 Introduction

In this work, we propose a novel approach to the calibration of three selected large-dimensional dynamic factor models for macroeconomic forecasting by means of a genetic algorithm. Some insights about the three selected dynamic factor models are reported below:

(i) *Stock and Watson (SW) model.* This time-domain method was introduced in [3,4]. The factors are estimated by computing static principal components of the variables in the dataset. Let y_{it} be the variable of the dataset to be forecasted at time t, its h-step-ahead prediction equation (also called *Diffusion Forecast Index*) is obtained by regressing y_{it+h} on the factors and on y_{it} itself. Lags of the factors and of y_{it} may be added.

(ii) *Forni, Hallin, Lippi and Reichlin (FHLR) model.* This frequency-domain method was proposed in [5,6] and requires the computation of two steps. In a

M. Pelillo et al. (Eds.): WIVACE 2017, CCIS 830, pp. 1–15, 2018.
https://doi.org/10.1007/978-3-319-78658-2_15

first step, the common component χ_t, the idiosyncratic component ξ_t and their covariances are estimated using a frequency-domain method introduced in [5] named *Dynamic Principal Component*. In the second step, the factors are estimated by computing Generalized Principal Components.

(iii) *Forni, Hallin, Lippi and Zaffaroni (FHLZ) model.* This frequency-domain method was proposed in [7,8]. Here, the underlying assumption in (i) and (ii) that the common components span a finite-dimensional space as n tends to infinity is relaxed.

There exists some literature comparing the forecasting performances of SW and FHLR, but universal consensus still does not seem to have been reached. Theoretically, time-domain methods (as FHLR and FHLZ) consider only relations among the variables at the same time, whereas frequency-domain methods (as SW) exploit leaded and lagged relations among the variables. However time-domain methods require less parameters to be calibrated. Hence they are more robust to misspecification than frequency-domain methods. Instead, a systematic comparison of the forecasting performances of SW, FHLR and FHLZ can be found only in [1,2]. [2] conducted a forecasting exercise on a US macroeconomic dataset, taking an autoregressive process of order 4 as a benchmark. They showed that FHLZ outperforms SW, FHLR and the benchmark both for the Industrial Production and the CPI during the Great Moderation (1982–2007). In the Great Recession (2007–2012), the forecasting performances of the Industrial Production change dramatically: all factor models are outperformed by the benchmark. SW and FHLR outperform FHLZ. Hence, Forni et al. concluded that, due to its more dynamical structure, FHLZ tends to be the best performing method in "stationary periods", but it loses ground during regime changes. [1] conducted a forecasting exercise on an EU macroeconomic dataset. The global settings of his exercise are basically the same as in [2], but also the length of the rolling window is suboptimally selected during the calibration process. He found that, on the proper sample, FHLZ is the most performing for the CPI. However, mixed evidences appear over the proper sample for the Industrial Production. Since each model is characterized by several parameters to estimate, an exhaustive exploration of the parameter space would be computationally infeasible. In order to give a partial solution to this issue, in [1] and in [2] the calibration procedure is carried out in a *naif* fashion, i.e. an initial configuration for each parameter is randomly selected and then, for each parameter at a time, a predetermined range of values is tested while keeping the other parameters fixed. As all the parameters have been tested, the configuration of the parameters with the lowest mean-squared forecast error (MSFE) is selected. The drawback of this procedure is that the final configuration selected may depend on the order on which the parameters have been processed in the calibration process. The novelty introduced in this paper is the employment of a genetic algorithm to explore the parameter space. In fact, the genetic algorithm allows us to select a suboptimal configuration of the parameters without imposing any order on the parameters to be estimated. In this work, we also compare the macroeconomic forecasting performance of the three selected dynamic factor models on two datasets. The former (an EU macroeconomic and financial dataset) is the same employed in [1]. Instead, the latter (an US macroeconomic and financial dataset) is the same employed in [2]. The paper is structured as follows. In Sect. 2, the calibration process of the models with a genetic algorithm is described. In Sect. 3, the results achieved on the EU dataset are discussed and, the same analysis is developped in Sect. 4 for the US dataset. In Sect. 5, some concluding remarks are presented.

2 The Calibration Process with a Genetic Algorithm

Both datasets contain real variables (import/export price indexes, employment, Industrial Production) and nominal variables (money aggregates, consumer price indexes, wages), asset prices (stock prices and exchange rates) and surveys. To achieve stationarity, several series are deseasonalized and transformed. No treatment for outliers is applied. In addition to SW, FHLR, FHLZ, the forecasts of an autoregressive process (AR) are computed. The order p of the AR process is determined in the calibration process. As in [1,4], to assess the forecasting performances, the variables which are taken into account are the level of the logarithm of the Industrial Production (IP) and the yearly change of the logarithm of the Consumer Price Index (CPI). Forecasts are computed h-months ahead, with $h \in \{1, 3, 6, 12, 24\}$. For each methods, we employ a rolling-window scheme $[t - l, t]$, whose size l is determined in the calibration sample.

As to the calibration process, the observations of the EU dataset ranging from February 1986 to December 2000 will be used to calibrate the methods SW, FHLR, FHLZ and the benchmark. For this reason, we will refer to this range of the EU dataset as the *calibration sample*. Instead, the calibration sample in the US dataset will range from March 1960 to December 1984. At each epoque, the population of the genetic algorithm is a subset of the strings containing all the possibile configurations of the parameters. We set the MSFE as the objective function to be minimised by the genetic algorithm. For each method, we iterate the genetic algorithm ten times on the calibration sample of the two datasets. The fitness of each individual is stored in a data structure. Eventually, for each method we select as the most performing configuration the one endowed with the lowest MSFE. More precisely, we select the configuration with the lowest objective function value that has been assessed during each of the ten runs of the genetic algorithms, independently from the final solutions obtained at each run. The parameters of each run of the genetic algorithm are the following:

(i) *Population size of the genetic algorithm at each generation = 100;*
(ii) *Crossover fraction = 0.6;*
(iii) *Number of individuals who passes to the next generation = 25;*
(iv) *Mutation = Gaussian model (adds a random number chosen from a Gaussian distribution, to each entry of the parent vector).*

The stopping criteria of each run of the genetic algorithm are the following:

(i) *Maximum number of generations = 1000;*
(ii) *Maximum number of generations in which the difference between the average MSFE is less than the threshold* $10^{-7} = 5$;
(iii) *MSFE of an individual in the last generation tending to zero.*

The same procedure is described in [9] and in [10], but the main purpose of these articles is to achieve suboptimal variable selection in a regressive setting.

3 Results on the EU Dataset

In this chapter, we will use the same notation as in [1].

3.1 Calibration of SW Model

In SW, the following parameters must be calibrated:

(i) *The number of static factors* r: ranging from 1 to 10. Also, a comparison with Bai & Ng criterium (BN) with maximum 12 factors has been made.
(ii) *The degree* α *of* $\mathbf{a}(L)$: ranging from 1 to 10.
(iii) *The degree* β *of* $b(L)$: ranging from 0 to 10.
(iv) *The size* l *of the rolling window*: ranging from 5 to 12 years.

After the ten runs of the genetic algorithms in the calibration process, the individual granted with the minimum objective function value for the IP is the following:

$$(r, \alpha, \beta, l) = (5, 1, 0, 11). \tag{3.1}$$

Instead, the individual granted with the minimum objective function value for the CPI is the following:

$$(r, \alpha, \beta, l) = (1, 0, 1, 7). \tag{3.2}$$

3.2 Calibration of FHLR Model

In FHLR, the following parameters must be calibrated:

(i) *The number of static factors* r: ranging from 1 to 10. Also, a comparison with Bai & Ng criterium (BN) with maximum 12 factors has been carried out.
(ii) *The number of dynamic factors* q: ranging from 0 to 10. Also, a comparison with Hallin-Liska criterium (HL) with maximum 12 factors has been carried out.
(iii) *The type of kernel* k: ranging in the set {Triangular, Rectangular, Parzen, Gaussian, Exponential, Cosine, Tukey, Hann}.

(iv) *The lag window d for spectral density estimation*: ranging in the set {25, 35, 40}.
(v) *The size l of the rolling window*: ranging from 5 to 12 years.

After the ten runs of the genetic algorithms in the calibration process, the individual granted with the minimum objective function value for the IP is the following:

$$(r, q, k, d, l) = (10, 3, Cosine, 35, 11). \tag{3.3}$$

Instead, the individual granted with the minimum objective function value for the CPI is the following:

$$(r, q, k, d, l) = (8, 4, Hann, 25, 7). \tag{3.4}$$

3.3 Calibration of FHLZ Model

In FHLZ, the following parameters must be calibrated:

(i) *The number of dynamic factors q*: ranging from 1 to 5. Also, a comparison with Hallin-Liska criterium has been carried out.
(ii) *The type of kernel k*: ranging in the set {Triangular, Rectangular, Parzen, Gaussian, Exponential, Cosine, Tukey, Hann}.
(iii) *The lag window d for spectral density estimation*: ranging in the set {25, 35, 40}.
(iv) *The maximum lag ml for the matrix $\mathbf{A}^k(L)$*: ranging from 1 to 5.
(v) *The size l of the rolling window*: ranging from 5 to 12 years.

After the ten runs of the genetic algorithms in the calibration process, the individual granted with the minimum objective function value for the IP is the following:

$$(q, k, d, ml, l) = (4, Parzen, 25, 4, 11). \tag{3.5}$$

Instead, the individual granted with the minimum objective function value for the CPI is the following:

$$(q, k, d, ml, l) = (2, Parzen, 35, 1, 7). \tag{3.6}$$

3.4 Calibration of the Benchmark

To calibrate the benchmark $AR(p)$, the only parameter that needs to be fixed is the order p. In our exercise, we let p range from 1 to 13. By selecting the values of the parameter p which guarantee the lowest mean rMSFE, the chosen configuration for the IP is the following:

$$p = 2. \tag{3.7}$$

Instead, the chosen configuration for the CPI is the following:

$$p = 1. \tag{3.8}$$

3.5 Empirical Proof of the Convergence of the Runs of the Genetic Algorithm

To give an empirical proof of the convergence of the genetic algorithm, in Fig. 1 the boxplots of the results of the ten runs of each selected dynamic factor models for the IP (on the left) and for the CPI (on the right) are reported.

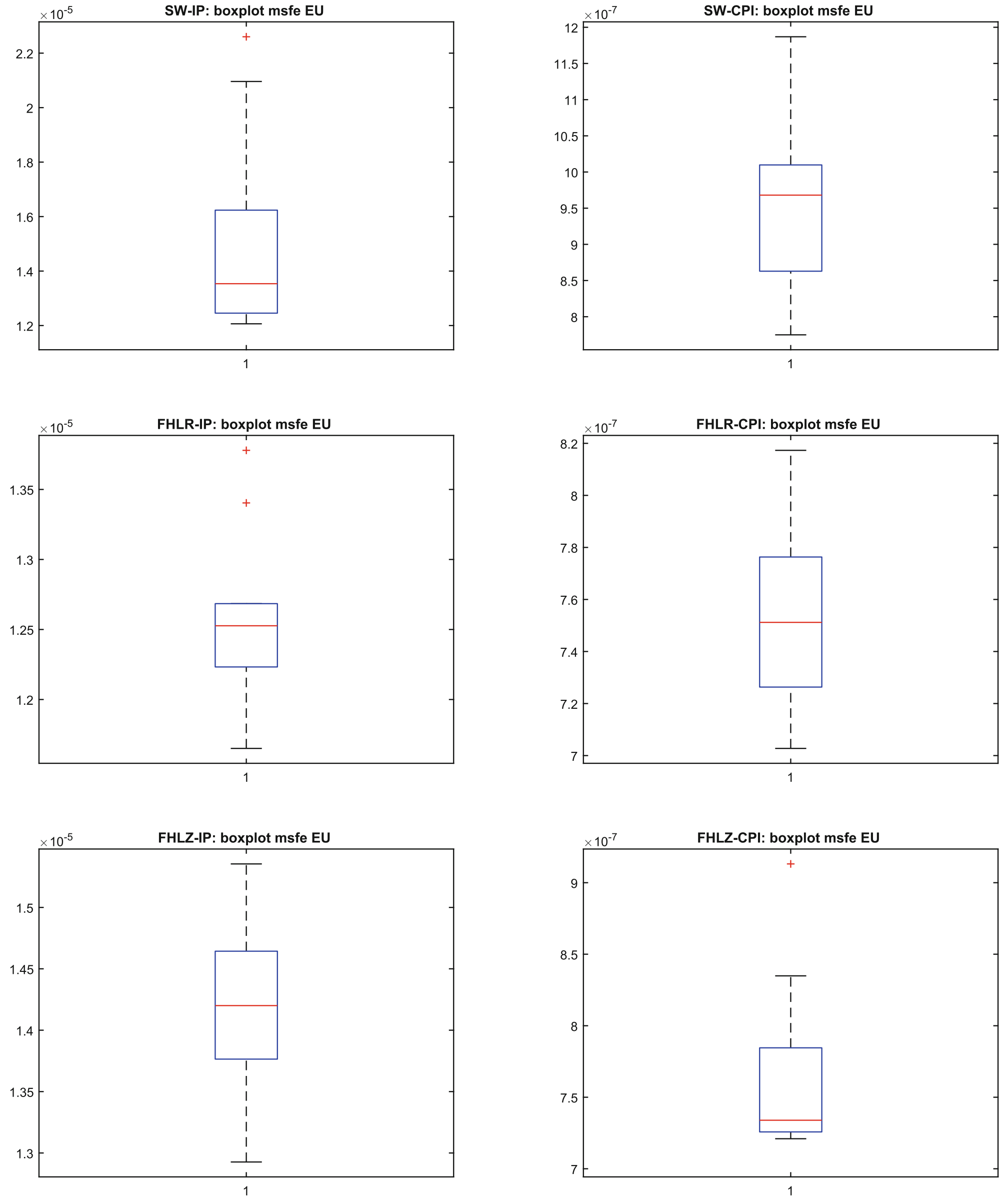

Fig. 1. Boxplot of the results on the EU dataset of the ten runs of the genetic algorithm for SW, FHLR and FHLZ over the IP (on the left) and over the CPI (on the right).

Since the results achieved for all dynamic factor models span a narrow region, we can conclude that the ten runs of the genetic algorithms for all methods have reached convergence. In addition, we can see that, over the IP, the dynamic methods show better results since the ten runs span a narrower region than

SW. Over the CPI, FHLZ shows better results since its ten runs of the genetic algorithm span a narrower region than the other methods. Moreover, the boxplot of FHLR covers a smaller region than SW.

3.6 Forecasting of the Industrial Production and the CPI

The forecasting performances of the three dynamic factor models over the IP and CPI are compared on the proper sample, which starts on January 2001 and ends on November 2015. The common benchmark for the factor models is the autoregressive process (AR) of order $p = 2$ for the IP and $p = 1$ for the CPI. However, as reported by CEPR, during the proper sample, the european economy faces two crisis periods: the first starts on May 2008 and ends on January 2009. The second starts on September 2011 and ends on March 2013. Hence, it is reasonable to assess whether the relative forecasting performances of the three dynamic factor models present a relevant change during the crisis periods. As in [2] and in [1], to assess the forecasting performance of each couple of methods locally, each time series of the dataset is smoothed by a centered moving average of length $m = 61$ (with coefficients equal to $1/m$) and then the Fluctuation test is run, at 5% significance level. Further details about this test can be found in [11]. The results for the IP at horizons $h \in \{6, 12, 24\}$ are reported in Fig. 2. All factor models outperform significantly the benchmark from the first crisis on at all horizons. SW tends to outperform the dynamic methods between the two crises. Instead, outside the period between the two crises, the dynamic methods show significantly better performances than SW. As to the performance of the dynamic methods, FHLR outperforms FHLZ between the two crises. To sum up, FHLR tends to outperform the other methods. However, this does not hold true in the period between the two crises, in which SW seems to be the most performing method. These results are similar to those obtained in [1], but in our exercise the relative performance of FHLR in comparison with SW are neater. The results for the CPI at horizons $h \in \{6, 12, 24\}$ are reported in Fig. 3. All methods perform better than AR significatively from the first crisis on. At horizon $h \in \{12, 24\}$, FHLR and FHLZ outperform SW on average on the whole sample, except between the two crises. As to the comparison of dynamic methods, at horizons $h \in \{6, 12\}$ FHLZ globally outperforms FHLR from the first crisis on. Instead, at horizon $h = 6$, FHLR globally outperforms FHLZ from the first crisis on. In comparison with [1], FHLR shows slightly better forecasting performance in comparison with other methods. In addition, SW seems to be the most performing method between the two crises.

FHLR and FHLZ tend to outperform SW at all horizons, except FHLR at horizon $h = 6$ during the first crisis. FHLR and FHLZ outperform AR at horizons $h \in \{6, 12\}$. At horizon $h = 24$, AR outperforms FHLR and FHLZ from the first crisis on. SW outperforms AR during the two crisis periods at horizons $h \in \{6, 12\}$. At horizons $h \in \{12, 24\}$, AR outperforms SW from the second crisis on. At all horizons, FHLZ outperforms FHLR during the first crisis. At horizons $h \in \{6, 12\}$, this behaviour seems to be persistent. Instead, at horizon $h = 24$, FHLR outperforms FHLZ from the second crisis on.

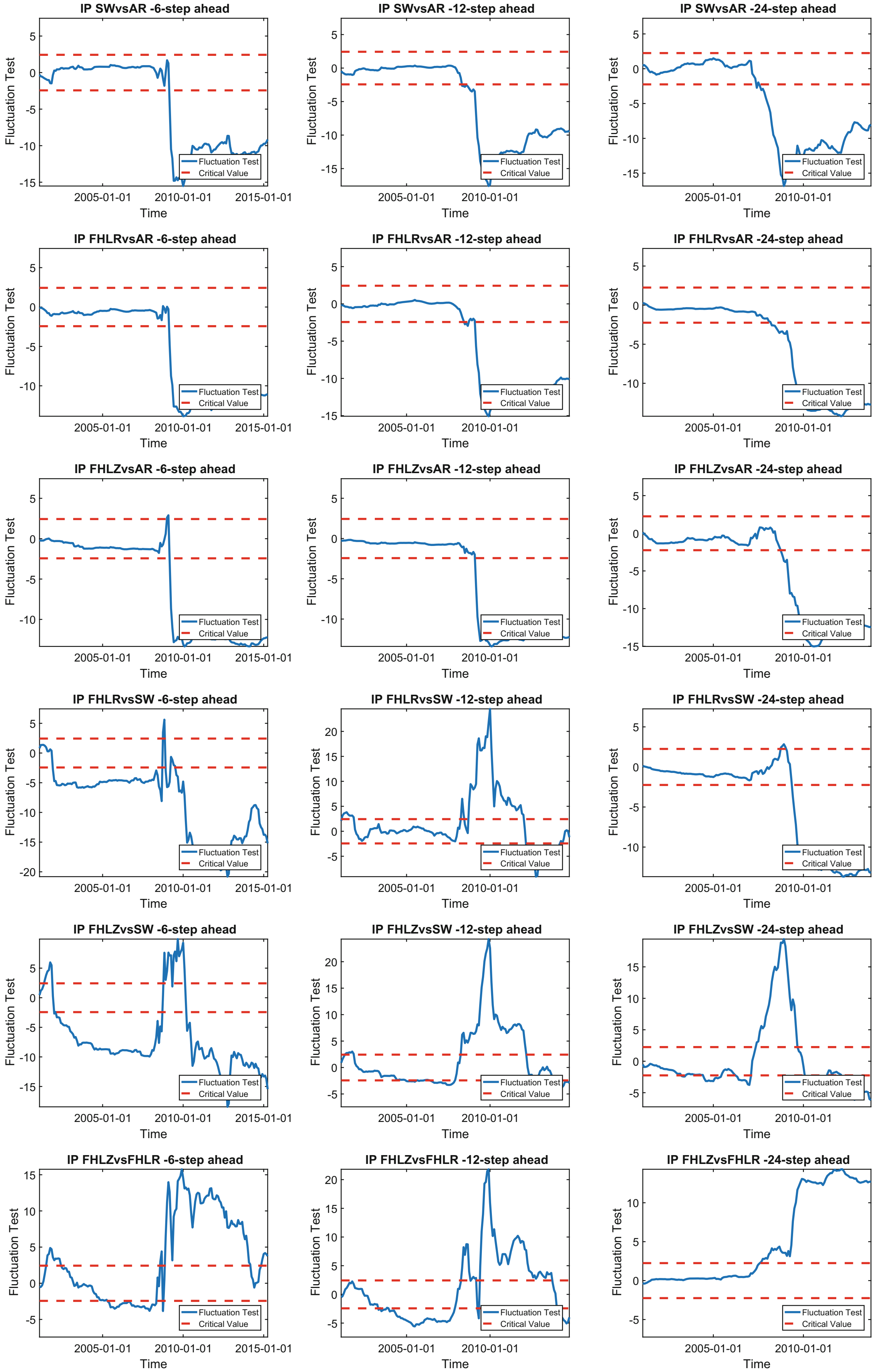

Fig. 2. Fluctuation test for the IP on the EU dataset.

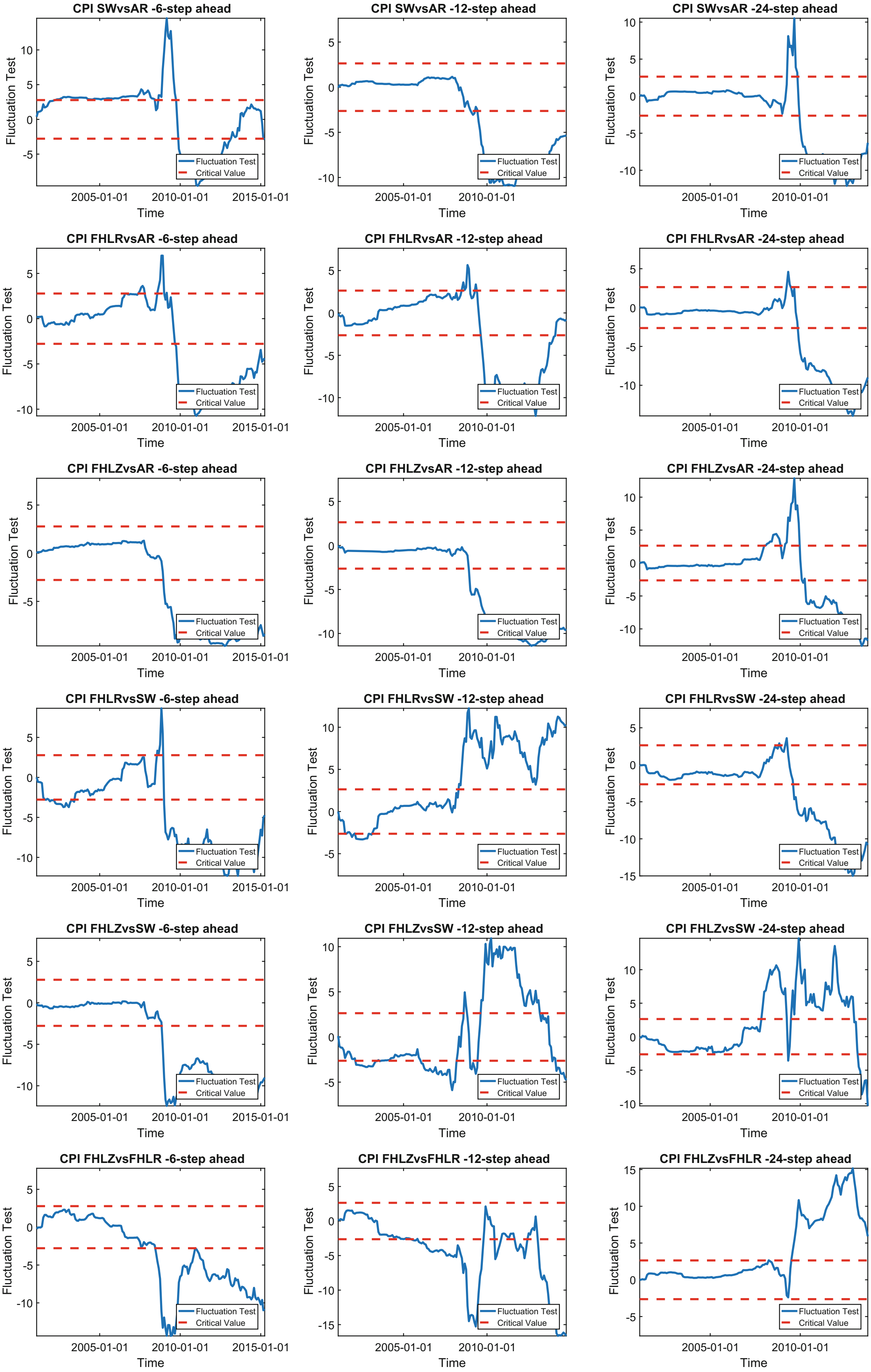

Fig. 3. Fluctuation test for the CPI on the EU dataset.

4 Results on the US Dataset

In this chapter, we will use the same notation as in [2].

4.1 Calibration of SW Model

As in Subsect. 3.1, by selecting the values of the parameters which guarantee the lowest mean rMSFE, the chosen configuration for the IP is the following:

$$(r, \alpha, \beta, l) = (BN, 0, 0, 12). \tag{4.1}$$

Instead, the chosen configuration for the CPI is the following:

$$(r, \alpha, \beta, l) = (3, 1, 10, 15). \tag{4.2}$$

4.2 Calibration of FHLR Model

As in Subsect. 3.2, by selecting the values of the parameters which guarantee the lowest mean rMSFE, the chosen configuration for the IP is the following:

$$(r, q, k, d, l) = (9, 2, Exponential, 40, 12). \tag{4.3}$$

Instead, the chosen configuration for the CPI is the following:

$$(r, q, k, d, l) = (6, 1, Hann, 25, 15). \tag{4.4}$$

4.3 Calibration of FHLZ Model

As in Subsect. 3.3, by selecting the values of the parameters which guarantee the lowest mean rMSFE, the chosen configuration for the IP is the following:

$$(q, k, d, ml, l) = (5, Triangular, 40, 2, 12). \tag{4.5}$$

Instead, the chosen configuration for the CPI is the following:

$$(q, k, d, ml, l) = (5, Triangular, 25, 5, 15). \tag{4.6}$$

4.4 Calibration of the Benchmark

As in Subsect. 4.4, by selecting the values of the parameter p which guarantee the lowest mean rMSFE, the chosen configuration for the IP is the following:

$$p = 2 \tag{4.7}$$

Instead, the chosen configuration for the CPI is the following:

$$p = 9 \tag{4.8}$$

4.5 Empirical Proof of the Convergence of the Runs of the Genetic Algorithm

To give an empirical proof of the convergence of the genetic algorithm, in Fig. 4 the boxplots of the results of the ten runs of each selected dynamic factor model for the IP (on the left) and for the CPI (on the right) are reported.

Since the results achieved for all dynamic factor models span a narrow region, we can conclude that the ten runs of the genetic algorithms for all methods over IP and over CPI have reached convergence. We can see that, over both the IP and the CPI, FHLZ shows better results since its ten runs of the genetic algorithm span a narrower region than the other methods. Moreover, the boxplot of FHLR covers a smaller region than SW.

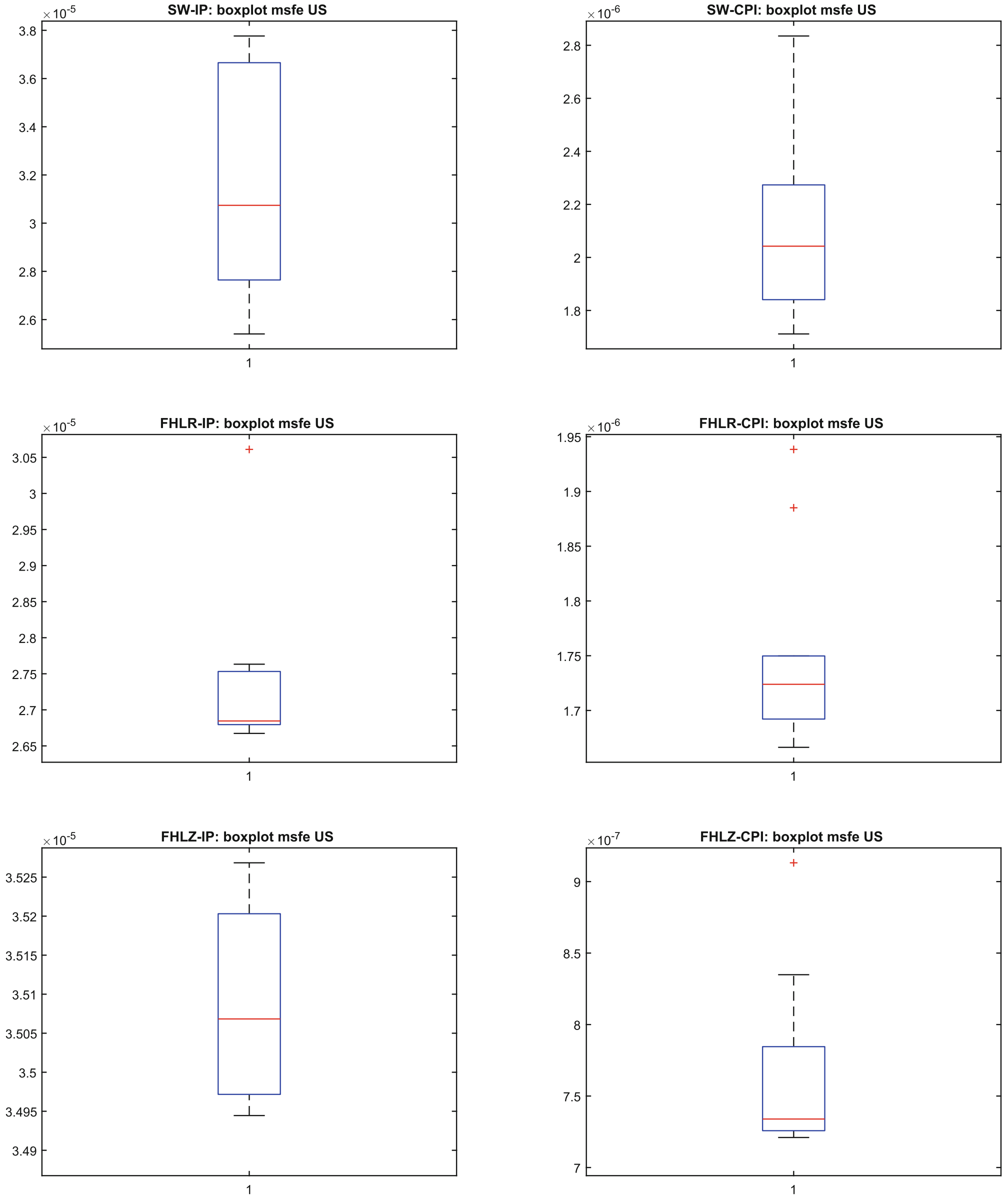

Fig. 4. Boxplot of the results on the US dataset of the ten runs of the genetic algorithm for SW, FHLR and FHLZ over the IP (on the left) and over the CPI (on the right).

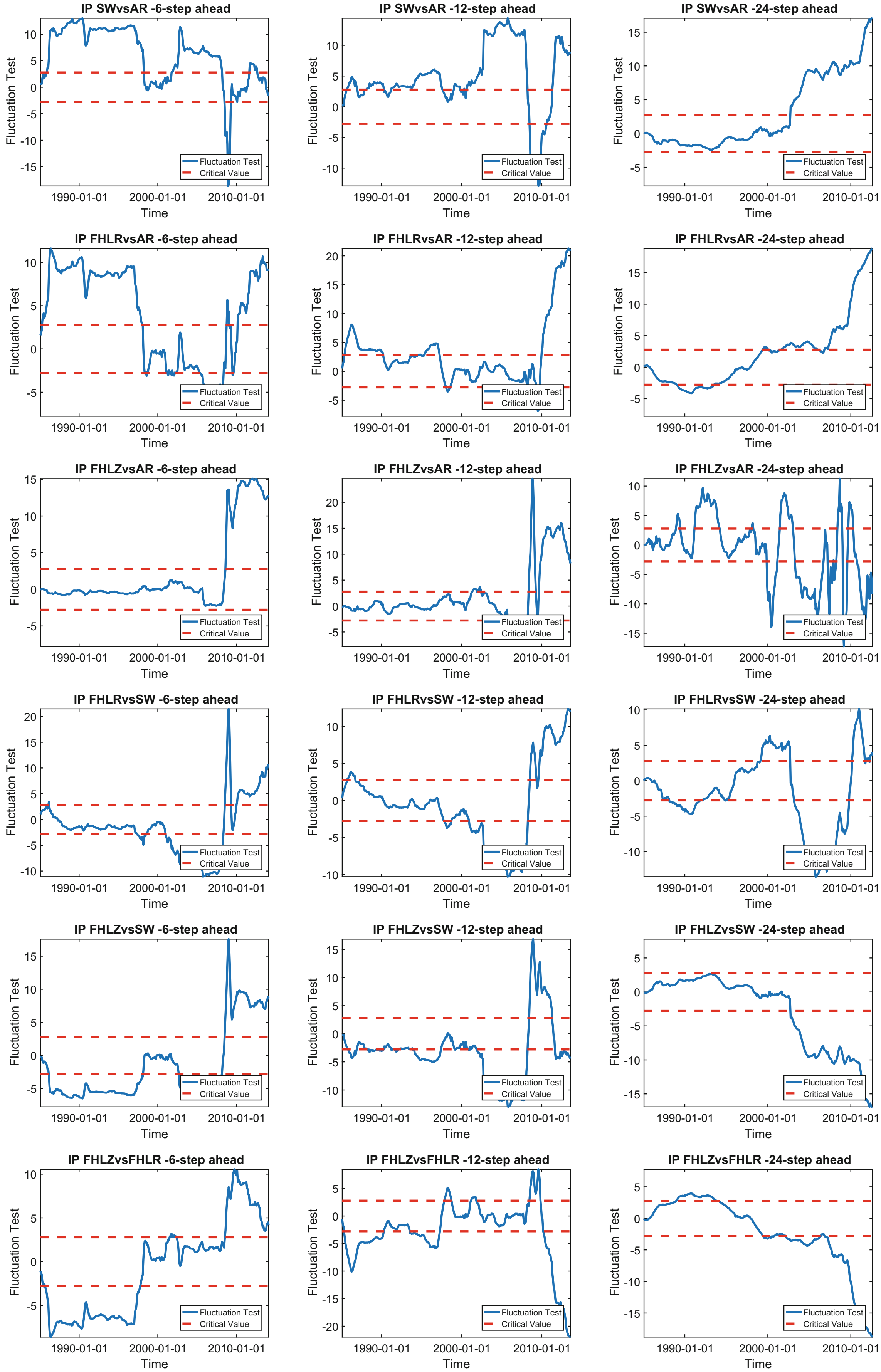

Fig. 5. Fluctuation test for the IP on the US dataset.

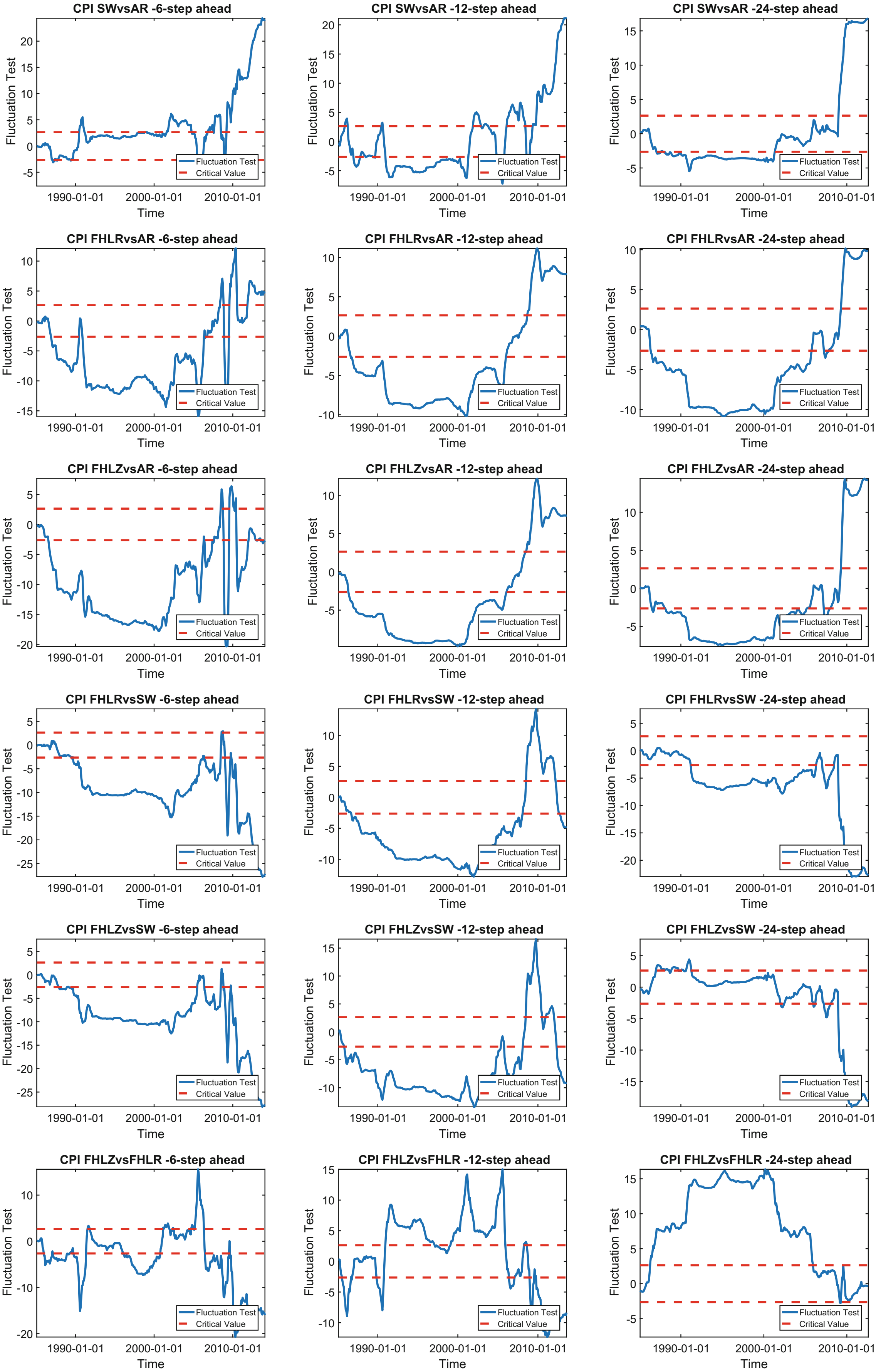

Fig. 6. Fluctuation test for the CPI on the US dataset.

4.6 Forecasting of the Industrial Production and the CPI

The forecasting performances of the three dynamic factor models over the IP and CPI are compared on the proper sample, which starts on March 1960 and ends on October 2014. The common benchmark for the factor models is the autoregressive process (AR) of order $p = 2$ for the IP and $p = 9$ for the CPI. However, as reported by NBER, during the proper sample, the american economy faces a crisis period which starts on December 2007 and ends on June 2009. Hence, it is reasonable to assess whether the relative forecasting performances of the three dynamic factor models present a relevant change during the crisis period. As in Subsect. 4.6, to assess the forecasting performance of each couple of methods locally, each time series of the dataset is smoothed by a centered moving average of length $m = 61$ (with coefficients equal to $1/m$) and then the Fluctuation test is run, at 5% significance level. The results for the IP at horizons $h \in \{6, 12, 24\}$ are reported in Fig. 5. The benchmark globally shows significantly better results than the factor models from the Great Recession on. However, this does not hold for SW at horizons $h \in 6, 12$ and for FHLZ at horizon $h = 24$ during the Great Recession. SW tends to outperform the dynamic methods from the Great Recession on, apart from FHLZ at horizon $h = 24$. As in [2], FHLR outperforms FHLZ from the Great Recession on. The results for the CPI at horizons $h \in \{6, 12, 24\}$ are reported in Fig. 6. No factor model seems to perform better than the benchmark from the Great Recession on. Instead, before the Great Recession, the contrary seems to hold. Dynamic methods show significant better performances than SW at all horizons, except for $h = 12$. FHLZ outperforms FHLR outside the Great Recession. Apart from this period, mixed evidences appear as far as the comparison between dynamic methods is concerned. Hence, as to the performances of dynamic methods, we can draw less clear conclusions than in [2].

5 Concluding Remarks

In this paper, we address the problem of calibrating dynamic factor models for macroeconomic forecasting. The novelty in this study consists in having designed and built a genetic algorithm for calibration. In this paper, we have empirically shown that the genetic algorithm in the calibration process plays a crucial role in this study, since a more efficient exploration of the parameter space allows us to empirically prove the superiority of frequence-domain dynamic factor models against time-domain factor model in a macroeconomic forecasting setting. We also notice that the time-domain factor model performs much better that the frequency-domain models considered in this paper. We eventually stress that our novel calibration approach has produced very good results in prediction.

Acknowledgements. We would like to thank Alessandro Giovannelli, Viviana Doldi, Marco Lippi, Valentina Mameli, Irene Poli, Simona Sanfelici, Debora Slanzi, Stefano Soccorsi and three anonymous referees for their support and comments on an earlier version of the manuscript.

References

1. Della Marra, F.: A forecasting performance comparison of dynamic factor models based on static and dynamic methods. Commun. Appl. Ind. Math. **8**(1), 44–66 (2017)
2. Forni, M., Giovannelli, A., Lippi, M., Soccorsi, S.: Dynamic factor model with infinite dimensional factor space: forecasting. CEPR discussion paper (2017)
3. Stock, J.H., Watson, M.W.: Forecasting using principal components from a large number of predictors. J. Am. Stat. Assoc. **97**(460), 1167–1179 (2002)
4. Stock, J.H., Watson, M.W.: Macroeconomic forecasting using diffusion indexes. J. Bus. Econ. Stat. **20**(2), 147–162 (2002)
5. Forni, M., Hallin, M., Lippi, M., Reichlin, L.: The generalized dynamic factor model: identification and estimation. Rev. Econ. Stat. **82**(4), 540–554 (2000)
6. Forni, M., Hallin, M., Lippi, M., Reichlin, L.: The generalized dynamic factor model: one-sided estimation and forecasting. J. Am. Stat. Assoc. **100**, 830–840 (2005)
7. Forni, M., Hallin, M., Lippi, M., Zaffaroni, P.: Dynamic factor model with infinite dimensional factor space: representation. J. Econom. **185**, 359–371 (2015)
8. Forni, M., Hallin, M., Lippi, M., Zaffaroni, P.: Dynamic factor model with infinite dimensional factor space: asymptotic analysis. In: EIEF Working Paper (2016)
9. Kapetanios, G.: Variable selection in regression models using nonstandard optimisation of information criteria. Comput. Stat. Data Anal. **52**(1), 4–15 (2007)
10. Kapetanios, G., Marcellino, G.M., Papailias, F.: Variable selection for large unbalanced datasets using non-standard optimisation of information criteria and variable reduction methods. In: Quantf Working Paper (2014)
11. Giacomini, R., Rossi, B.: Forecast comparisons in unstable environments. J. Appl. Econom. **25**(4), 595–620 (2010)