



The Multivariate k-Nearest Neighbor Model for Dependent Variables: One-Sided Estimation and Forecasting

Dominique Guegan, Patrick Rakotomarolahy

► To cite this version:

Dominique Guegan, Patrick Rakotomarolahy. The Multivariate k-Nearest Neighbor Model for Dependent Variables: One-Sided Estimation and Forecasting. 2009. halshs-00423871v1

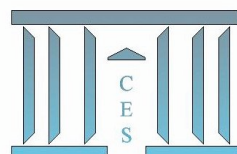
HAL Id: halshs-00423871

<https://shs.hal.science/halshs-00423871v1>

Submitted on 13 Oct 2009 (v1), last revised 5 Jan 2010 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



**The Multivariate k-Nearest Neighbor Model for Dependent
Variables : One-Sided Estimation and Forecasting**

Dominique GUEGAN, Patrick RAKOTOMAROLAHY

2009.50



The Multivariate k -Nearest Neighbor Model for Dependent Variables: One-Sided Estimation and Forecasting

Dominique Guégan*, Patrick Rakotomarolahy†

Abstract

Forecasting current quarter GDP is a permanent task inside the central banks. Many models are known and proposed to solve this problem. Thanks to new results on the asymptotic normality of the multivariate k -nearest neighbor regression estimate, we propose an interesting and new approach to solve in particular the forecasting of economic indicators, included GDP modelling. Considering dependent mixing data sets, we prove the asymptotic normality of multivariate k -nearest neighbor regression estimate under weak conditions, providing confidence intervals for point forecasts. We introduce an application for economic indicators of euro area, and compare our method with other classical ARMA-GARCH modelling.

Keywords: Multivariate k -nearest neighbor - Asymptotic normality of the regression - Mixing time series - Confidence intervals - Forecasts - Economic indicators - Euro area.

1 Introduction

The monetary policy decisions in real time are based on the forecast of future economic conditions and on the assessment of current economic conditions based on incomplete data. Since most data are released with a lag and are subsequently revised, the reconstruction of current quarter GDP, inflation and other key variables is an important task for central banks.

Monthly indicators are routinely used in judgmental forecasting to form a view of current economic conditions before GDP data are made available. Statistical models which can perform this

*Paris School of Economics, CES-MSE, Université Paris 1 Panthéon-Sorbonne, 106 boulevard de l'Hôpital 75647 Paris Cedex 13, France, e-mail: dguegan@univ-paris1.fr

†CES-MSE, Université Paris 1 Panthéon-Sorbonne, 106 boulevard de l'Hôpital 75647 Paris Cedex 13, France, e-mail: rakotopapa@yahoo.fr

exercise have been discussed a lot, mixing econometric models and judgmental analysis as well as informal weighting of information from diverse sources.

An important key to forecast GDP is the modelling or the forecasting of the different indicators used inside the GDP models. These last models are based on different approaches including the VAR models Marcellino *et al.* (2006), the dynamic factor models Bernanke and Boivin (2003), Forni *et al.* (2005), Kapetanios and Marcellino (2006), or the bridge equations Diron (2008). In this paper, we do not discuss these previous models. We focus on the forecasting of economic indicators using new methodology. Our aim is to be free of technical assumptions like Gaussianity, stationarity and linearity, which are often considered and which are not true in practice.

To obtain accurate forecasts, we consider techniques belonging to non-parametric methods. Non-parametric techniques have a long history in statistics: they have been extensively developed for density estimation Rosenblatt (1956), and then extended for regression smoothing by Nadaraya (1964) and Watson (1964). These techniques include for instance the kernel method, the nearest neighbors method, the radial basis function method, the wavelets and the genetic algorithms. There are numerous reasons to be interested in such approaches when studying time series. Indeed, these methods can be employed as a convenient and succinct means of displaying the features of a data set; they can be used to check diagnose of an estimated parametric model; they permit inference under very weak assumptions: they can be used to construct estimators for semi-parametric modelling.

In this paper, we re-exploit the k -nearest neighbors (k -NN) method in order to reconstruct multivariate time series for forecasting. We focus on estimation problems for regression under weak assumptions, in the multivariate setting. We provide new results which concern the asymptotic normality of our estimate for dependent variables, with respect to the bias-variance fit dilemma inherent in this kind of methodology. Our result permits a confidence interval, to be obtained which can be used to discriminate between classical methods.

Our proposal can be considered as a contribution of the general problem concerning the non-parametric estimate of a regression: we extend some results obtained in an independent setting to a dependent framework, permitting it to be applied to time series. Some references in nearest neighbors are Mack (1981) and Stute (1984), who obtain the asymptotic normality of the regression using the k -NN method in an independent multivariate setting. Other classes of convergence for independent samples have also been developed by Stone (1975, 1977), and Devroye

(1978, 1982). Some results have been obtained for dependent variables in an univariate setting: Collomb (1984) provides piecewise convergence for univariate dependent random variables and Yakowitz (1987) gets the quadratic mean error for uniformly weighted k -NN estimate for the same kind of samples. Here, working with multivariate time series, we control the bias of a general multivariate k -NN estimate, using several weights extending the Yakowitz (1987) result, and we establish the asymptotic normality of this estimate which permits exact confidence intervals to be constructed.

In order to illustrate the interest of our approach, for instance in the forecasting of economic indicators, we apply our results to economic indices used to forecast and nowcast GDP, proposing at the same time an interesting alternative to the work developed around the GDP modelling.

The paper is organized as follows. In Section 2, we recall the k -nearest neighbor method used to re-build a regression and we establish our main result: the asymptotic normality of the multivariate k -NN regression estimate for a mixing time series. In Section 3, we apply our theoretical result to a set of six economic indices. We describe two forecasting methods and we provide the forecasts and the associated confidence intervals. We compare our results with forecasts obtained using an ARMA modelling. Section 4 concludes and Section 5 is devoted to the proofs.

2 Estimation of a Regression Using k -Nearest Neighbors

Regression analysis concerns the estimation of a relationship between two random variables Y (the dependent variable) and X (the explanatory or regressor variable) and can be summarized by the expression $Y = m(X)$, where $m(\cdot)$ is the function to estimate. The random variables X can be any variable, and in particular can be past values of the random variables Y . The underlying principle that theoretical laws usually do not hold in every individual case but merely on average will be considered and is formalized as:

$$m(x) = E[Y|X = x]. \quad (2.1)$$

Now, if we observe (x_i, y_i) , $i = 1, \dots, n$, such that:

$$y_i = m(x_i) + \varepsilon_i, \quad i = 1, \dots, n, \quad (2.2)$$

where $(\varepsilon_i)_i$ is an independent random error satisfying $E[\varepsilon_i|X_i = x] = 0$ and $Var[\varepsilon_i|X_i = x] = \sigma^2(x)$, then equation (2.2) says that the relationship $Y = m(X)$ does not need to hold exactly

for the i th observation but is disturbed by the random variable ε . Yet, equation (2.1) says that the equation holds on average, i.e., the expectation of Y on the condition that $X = x$ is given by $m(x)$. Considering the set of previous observations, we can estimate $m(\cdot)$ in the following way:

$$\hat{m}_n(x) = \sum_{i=1}^n \omega_{i,n}(x) y_i, \quad (2.3)$$

where $\omega_{i,n}$ are weights to specify and to estimate and which arise from different motivations, while possessing different statistical properties, (see Silverman (1986), Guégan (2003) and Härdle *et al.* (2004) for instance).

The objective of this Section is twofold: to estimate $m(\cdot)$ using k -nearest neighbors while observing a time series $(X_n)_n$ embedded in a space of dimension d and to establish asymptotic normality of the estimate, providing a confidence interval used to discriminate between different approaches. We describe now the method and introduce our main results.

The classical univariate k -nearest neighbors method used to estimate $m(\cdot)$ works as follows. Considering real random variables $(X_n)_n \in \mathbb{R}$, we define a neighborhood around $x \in \mathbb{R}$: $N(x) = \{i \mid i = 1, \dots, k \text{ whose } X_{(i)} \text{ represents the } i^{th} \text{ nearest neighbor of } x \text{ in the sense of a given distance measure}\}$. Thus, the k -NN estimate of m can be represented as the weighted average of the $X_{(i)}$'s with an index in $N(x)$ and the relationship (2.3) becomes:

$$\hat{m}_n(x) = \sum_{i \in N(x)} w(x - X_{(i)}) X_{(i)+1}, \quad (2.4)$$

where $w(\cdot)$ is a weighting function associated to neighbors. We can distinguish three types of k -NN regression estimates for which it is noteworthy that the parameter k needs to be estimated.

1. The first one corresponds to the so called uniformly weighted k -NN regression:

$$m_n(x) = \frac{1}{k} \sum_{i \in N(x)} X_{(i)+1} \quad i.e. \quad w(x - X_{(i)}) = \frac{1}{k}. \quad (2.5)$$

2. The second one is a generalization of the previous one by considering a non-uniform weight, a non-function of $(X_n)_n$:

$$m_n(x) = \sum_{i \in N(x)} w_i X_{(i)+1}, \quad w_i \in \mathbb{R}, \quad \text{then} \quad w(x - X_{(i)}) = w_i. \quad (2.6)$$

3. The third one takes into account the distance of the point observation and the neighbors, and depends on $(X_n)_n$:

$$m_n(x) = \frac{\sum_{i \in N(x)} K\left(\frac{x - X_{(i)}}{R(k)}\right) X_{(i)+1}}{\sum_{i \in N(x)} K\left(\frac{x - X_{(i)}}{R(k)}\right)}, \quad (2.7)$$

where $R(k)$ corresponds to the distance between x and the further k -nearest neighbors and $K(\cdot)$ is a weighting function. Estimating $m(\cdot)$ using the k -NN method at a point x , when the data are sparse, might occur when the k -nearest neighbors are rather far away from x (and each other), consequently we end up with a wide neighborhood around x , for which an average of the corresponding values of $X_{(i)}$ is computed.

In the rest of the paper we work with vectors of dimension d . Assuming that we observe a time series observed in \mathbb{R} , we transform the original data set embedding it in a space of dimension d , building $\underline{X}_n \subset \mathbb{R}^d$. The embedding is interesting because it permits different features of the data to be taken into account which are not observed on the trajectory. We are interested to get an estimate of $m(\underline{x})$, $\underline{x} \in \mathbb{R}^d$, using the k closest vectors of $\underline{X}_n = (X_{n-d+1}, \dots, X_n)$ inside the training set $S \in \mathbb{R}^d$:

$$S = \{\underline{X}_{l+d} = (X_{l+1}, \dots, X_{l+d}) \mid l = 0, \dots, l = n - d - 1\}.$$

We keep the previous notations and $N(\underline{x})$ will again represent the set of indices of k -NN in S . We denote by $\underline{X}_{(i)}$ $i = 1, \dots, k(n)$, the k -nearest vectors around \underline{X}_n , then the k -NN estimate of $m(\underline{x})$, $\underline{x} \in \mathbb{R}^d$ is:

$$m_n(\underline{x}) = \sum_{\underline{X}_{(i)} \in S, i \in N(\underline{x})} w(\underline{x} - \underline{X}_{(i)}) X_{(i)+1}. \quad (2.8)$$

A general form for the weights is:

$$w(\underline{x} - \underline{X}_{(i)}) = \frac{\frac{1}{nR_n^d} K\left(\frac{\underline{x} - \underline{X}_{(i)}}{R_n}\right)}{\frac{1}{nR_n^d} \sum_{i=1}^n K\left(\frac{\underline{x} - \underline{X}_{(i)}}{R_n}\right)}, \quad (2.9)$$

where $K(\cdot)$ is a given weighting function vanishing outside the unit sphere in \mathbb{R}^d and R_n is the distance between the k th NN of \underline{x} and \underline{x} itself. We can still consider the exponential weighted function $K\left(\frac{\underline{x} - \underline{X}_{(i)}}{R_n}\right) = \exp(-\|\underline{x} - \underline{X}_{(i)}\|^2)$ or the uniform weighted $K(\underline{x} - \underline{X}_{(i)}) = \frac{1}{k}$.

Establishing asymptotic properties of the previous estimates is fundamental to justifying and applying the method. The knowledge of the bias and speed rate of variance provides consistent estimates, and the asymptotic normality of the estimates permits confidence intervals to be built. Thus, we assume that we observe a strictly stationary time series $(X_n)_n$ that is characterized by an invariant measure with density f , and a conditional density $f(y \mid \underline{x})$ for the random variable $X_{n+1} \mid (\underline{X}_n = \underline{x})$, and the invariant measure associated to the embedded time series $\{\underline{X}_n = (X_{n-d+1}, \dots, X_n)\}$ is denoted by h . We thus have the following main result.

Theorem 2.1. *Assuming that $\{X_n\}$ is a stationary time series, and that the following assumptions are verified:*

- (i) $(X_n)_n$ is ϕ -mixing.
 - (ii) $m(\underline{x})$, $f(y \mid \underline{x})$ and $h(\underline{x})$ are p continuously differentiable.
 - (iii) $f(y \mid \underline{x})$ is bounded,
 - (iv) the sequence $k(n) < n$ is such that $\sum_{i=1}^{k(n)} w_i \rightarrow 1$ as $n \rightarrow \infty$,
- then the k -NN regression function $m_n(\underline{x})$ defined in (2.8) verifies:

$$\sqrt{n^Q}(m_n(\underline{x}) - Em_n(\underline{x})) \rightarrow_{\mathcal{D}} \mathcal{N}(0, \sigma^2), \quad (2.10)$$

with

$$E[(m_n(\underline{x}) - m(\underline{x}))^2] = O(n^{-Q}), \quad (2.11)$$

$0 \leq Q < 1$, and $Q = \frac{2p}{2p+d}$. Moreover,

1. for uniform weight $w(\underline{x}) = \frac{1}{k(n)}$, $\sigma^2 = \text{Var}(X_{n+1} \mid \underline{X}_n = \underline{x})$.
2. for non-uniform weights $w(\underline{x}) \in \mathbb{R}$ depending (or not) on $(X_n)_n$,

$$\sigma^2 = \gamma^2(\text{Var}(X_{n+1} \mid \underline{X}_n = \underline{x}) + B^2),$$

where B is given in Lemma 3.1 and γ is a positive constant.

The proof of this theorem is postponed to the end of the article.

The conditions under which this result is established are quite weak in time series analysis. Indeed, the mixing condition permits data sets which are asymptotically independent to be considered. Parametric processes like bilinear models including ARMA models, related GARCH processes and Markov switching processes, are known to be mixing (Guégan (1983) and Carrasco

and Chen (2002)). The condition (iv) is in particular verified for the weights introduced in (2.5) - (2.7), and (2.9). This theorem is interesting because it provides asymptotic normality for the estimate $m_n(x)$ under quite regular conditions and thus permits confidence intervals to be built. Therefore, this result is fundamental in applications because it will permit, for instance, our comparison of several methods (parametric and non-parametric methods) to be enhanced in forecasting, in a large framework beyond point forecast. Indeed, no test is available to discuss the choice between the parametric and the non-parametric approaches.

In order to construct a confidence interval for the k -NN estimate regression, we establish the following result.

Corollary 2.1. *Under the assumptions of Theorem 2.1, a general form of the confidence interval for $m(\underline{x})$, for a given α , is :*

$$m(\underline{x}) \in [m_n(\underline{x}) - B - \frac{\hat{\sigma} z_{1-\frac{\alpha}{2}}}{\sqrt{k}}, m_n(\underline{x}) + B + \frac{\hat{\sigma} z_{1-\frac{\alpha}{2}}}{\sqrt{k}}] \quad (2.12)$$

where $z_{1-\frac{\alpha}{2}}$ is the $(1 - \frac{\alpha}{2})$ quantile of the Student law, $\hat{\sigma}$ is an estimate for σ and B is such that:

1. B is negligible, if $\frac{k(n)}{n} \rightarrow 0$, $n \rightarrow \infty$,
2. If not, $B = O(r^p)$, with $r = \left(\frac{k(n)}{(n-d)\hat{h}(\underline{x})^c} \right)^{\frac{1}{d}}$ where $c = \frac{\pi^{d/2}}{\Gamma((d+2)/2)}$, and $\hat{h}(\underline{x})$ is an estimate for the density $h(\underline{x})$.

The proof of this lemma is postponed to the end of the article.

3 Application: Forecasting of Some Economic Indices

In order to assess the quality of the forecasts obtained from the previous method we carry out a preliminary exercise using the following monthly macroeconomic indices of euro area. The Industrial Production Index (IPI) has 213 observations from January 1990 until September 2007. The Industrial Production Index in Construction (IPIC) also has 213 observations from January 1990 until September 2007. The Confidence Indicator in Industry (ICI), has 275 observations from January 1985 until November 2007. The Confidence Index in Retail Trade (RCI), has also 275 observations from January 1985 until November 2007. The Retail Sale Index (RI) has 153

observations from January 1995 to September 2007. The Consumer Confidence Indicator (CI) has 275 observations from January 1985 until November 2007. We provide a summary of these data sets with their descriptive statistics in Table 1.

Table 1

From Table 1, we notice that the index IPI is right skewed while the others indices are all left skewed. Furthermore, the IPI and RI have thinner tail compared with the Gaussian one while the other indexes have small excess of kurtosis compared to the Gaussian one. The graphs of these six macroeconomic indicators are given in figure 1.

Figure 1

Now we describe the two forecasting methods we use. The first one called a "recursive forecast" provides forecasts at different horizons: the parameters of the models (parametric and non-parametric modellings) are done once and used all along the forecasting modelling. The second approach called "Direct forecast" provides forecasts for several horizons: the parameters of the models are estimated for each horizon. In both approaches, the information set is the same, but the method to provide the forecasts is totally different.

3.1 Recursive forecast

We assume that we observe a data set X_1, \dots, X_n . We denote this information set $I_n^0 = \sigma(X_s, s \leq n)$: it is the sigma algebra generated by $X_s, s \leq T$. We are interested in forecasting X_{n+h} for a given horizon $h > 0$. We use here a recursive forecasting method. Having estimated once the parameters of the models using I_n^0 , we build \hat{X}_{n+1} . To predict \hat{X}_{n+2} , we use the information set $I_n^1 = I_n^0 \cup \hat{X}_{n+1}$. At the end, we predict \hat{X}_{n+h} using the information set $I_T^{h-1} = I_n^0 \cup (\hat{X}_{n+j})_{j < h}$. In order to measure the accuracy of the predictions, we use 24 observations out of sample (corresponding to two years).

3.1.1 Parametric modelling

In a first step, we consider an ARMA model for the six indicators, modelling the conditional variance with a GARCH process. We validate the modelling with back testings on the residuals. For all t , we define $Y_t = X_t - X_{t-1}$, with sample mean \bar{Y} . Then, the following ARMA(p,q)-GARCH(r,s) model is considered for $(Y_t)_t$.

$$\begin{cases} \Phi(B)(Y_t - \bar{Y}) = \Theta(B)\epsilon_t \\ \epsilon_t = \sqrt{h_t}\nu_t \\ \mathcal{B}(B)h_t = a_0 + \mathcal{A}(B)\epsilon_t^2 \end{cases}$$

where $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$, $\Theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$, $\mathcal{B}(B) = 1 - b_1 B - b_2 B^2 - \dots - b_r B^r$, $\mathcal{A}(B) = a_1 B + a_2 B^2 + \dots + a_s B^s$, $(\phi_i)_{i=1}^p$, $(\theta_j)_{j=1}^q$, $(a_k)_{k=0}^r$ and $(b_\ell)_{\ell=1}^s$ are real parameters, B is the lag operator $B^n Y_t = Y_{t-n}$, and $(\nu_t)_t$ is assumed to be a Gaussian strong white noise.

The estimated parameters are given in Table 2. In brackets we provide the standard deviation associated with each parameter. Table 3 provides information concerning the residuals of these models. The null assumption for non-correlation of the residuals can be retained for all the indexes, and we observe that the squared residuals are correlated for the IPIC and the RCI, thus we apply a GARCH model on the conditional variance of these indicators. The results for GARCH modelling are provided in Table 2. In summary, we retain the following models for the six indicators: an ARIMA(3,1,0) for IPI, an ARIMA(2,1,0)-ARCH(1) for IPIC, an ARIMA(3,1,0) for ICI, an ARIMA(2,1,0)-GARCH(1,1) for RCI, an ARIMA(2,1,0) for RI and an ARIMA(3,1,0) for CI.

Table 2

Table 3

3.1.2 Multivariate k -NN modelling

In this Section, we describe the multivariate k -NN algorithm to forecast $(X_t)_t$.

1. Given the data set X_1, \dots, X_n in \mathbb{R} .
2. We embed this set in a space of dimension d , for several values of d . We obtain a sequence of vectors in \mathbb{R}^d : $\{\underline{X}_d, \underline{X}_{d+1}, \dots, \underline{X}_n$, where $\underline{X}_i = (X_{i-d+1}, \dots, X_i)\}$.
3. We determine the k nearest vectors of \underline{X}_n inside the above vectors, and we denote $r_i = \|\underline{X}_n - \underline{X}_i\|$, $i = d, d+1, \dots, n-1$ the distance between these vectors, building a sequence $r_d, r_{d+1}, \dots, r_{n-1}$ that we order : $r_{(d)} < r_{(d+1)} < \dots < r_{(n-1)}$.
4. We detect the k vectors, denoted $\underline{X}_{(j)}$, corresponding to $r_{(j)}$, $j = d, d+1, \dots, d+k-1$ (the k smallest distance).
5. To compute $m_n(\underline{X}_n) = \hat{X}_{n+1}$, we use the following expression:

$$\hat{X}_{n+1} = \sum_{j=d}^{k+d-1} w(\|\underline{X}_n - \underline{X}_{(j)}\|) X_{(j)+1}. \quad (3.1)$$

It may be noted that we obtain the one step ahead forecast.

6. Then we consider the information set: $X_1, \dots, X_n, \hat{X}_{n+1}$ and redo Step 2 to Step 5, we get the two steps ahead of the forecast. We obtain the forecast of third step ahead in a similar way, as obtaining two steps ahead forecast, and so on.

Some remarks are important concerning the application of the previous algorithm. Sometimes we use large values for the embedding dimension d : this means that some structural behaviours exist inside the data set in a distant past referring to the present time t . These types of behaviours correspond to hidden seasonality, trends or persistence. In our applications, we have used $6 < d < 10$, reflecting the importance of the information captured in the past to obtain good forecasts in the future. We can observe that the bias B and the variance σ^2 of the forecasts increase and decrease with k , respectively. The predictions are sensitive to the choice of the weights and we have observed that in our applications the exponential weights performed the results, in the sense of forecast accuracy. We often need to make a trade-off for the choice of a couple (d, k) to get accurate results at the same time for a small horizon or a large horizon. Sometimes, it will be possible to retain another choice for the couple (d, k) . Indeed, given a couple (d, k) , we can get smaller errors for a small horizon, but larger errors for greater horizon. Thus no specific rule can be proposed and the method has to be adapted in practice for the sets on which it is applied:

the criteria used by practitioners will be determinant. The filtering of the data set is important, but in our method, even if we transform the data set, all the predictions are done on the raw data sets.

3.1.3 Forecasting criteria and results

We use two criteria to compare the forecasts obtained with the two previous methods :

$$RMSE = \sqrt{\frac{1}{h} \sum_{j=1}^h (X_{n+j}^i - \hat{X}_{n+j}^i)^2}, \quad (3.2)$$

and

$$Error(h) = |X_{n+h}^i - \hat{X}_{n+h}^i|, \quad (3.3)$$

where X_{n+h}^i is the true value, and \hat{X}_{n+h}^i is given by:

$$\hat{X}_{n+h}^i = X_n + \sum_{j=1}^h Y_{n+j-1}, \quad h = 1, \dots, 24, \quad (3.4)$$

with $Y_t = X_t - X_{t-1}$, and $i = 1, \dots, 6$ denotes the indicator.

We summarize the results obtained with the two methods (ARMA and NN- method) in Tables 4 and 5. The k -NN RMSEs and errors increase slowly with the forecasting horizon, showing the ability of the methods for a longer projection forecast. They are close to zero in the short run (horizons less than six) for IPI, IPIC, ICI and RI indices, and turn around 0.5 to 2 for the mean horizon. For the CI the RMSEs vary from 1 to 2, and are between zero and 3.5 for RCI, for both short and mean horizons.

Table 4

Table 5

Denoting RNN, the RMSE obtained with k -NN method and RAR the RMSE obtained with ARMA modelling, we can summarize the results as follows:

- For IPI and IPIC : for $h < 4$, $RNN < \frac{2}{3}RAR$, and for $h \geq 4$, $RNN < \frac{1}{2}RAR$.
- For ICI, for $h < 4$, $\frac{1}{3}RAR < RNN < \frac{1}{2}RAR$, and for $h \geq 4$ $RNN < \frac{1}{4}RAR$.
- For RCI, $\frac{1}{2}RAR < RNN < \frac{3}{4}RAR$.
- For RI, for $h < 4$, $RNN < \frac{1}{3}RAR$, and for $h \geq 4$, $\frac{2}{3}RAR < RNN < \frac{3}{4}RAR$.
- For CI, $h < 6$, $RNN \sim \frac{2}{3}RAR$, and for $h \geq 6$ $RNN < \frac{1}{2}RAR$.

We notice also that the forecast error for each horizon (Table (3.3)) provides similar results than with RMSEs. Nevertheless the errors are larger with this approach than using the RMSEs criteria.

In summary, the accuracy of the forecasts has been improved largely using multivariate k -NN method. The rate is almost half or two thirds of ARMA's RMSE, for $h < 4$.

3.1.4 Confidence interval

We provide now confidence intervals for the forecasts obtained previously using Corollary 2.1. We can use the relationship (2.12) to get these intervals. We assume also that the bias B is neglecting, and we consider uniform weights. The confidence intervals at 95% are given in Figure 2 and Figure 3 using the k -NN and ARIMA modellings.

Figure 2

Figure 3

Concerning the pointwise forecasts, when we compare the two methods (ARMA modelling and the k -NN method), we observe that the ARMA modelling provides the classical mean forecast (we get nearly a straight line around the mean). The k -NN forecasts follow the true trajectory, except for the ICI indicator. For this last indicator the predictions are under the true values and close to the mean value.

Now if we compare the confidence intervals obtained by the both methods (ARMA modelling and the k -NN method), the true trajectory is always inside the k -NN confidence intervals: it is not always the case looking at the ARMA confidence intervals. We can also notice that the bandwidth of the k -NN confidence interval is smaller than the bandwidth of the ARMA confidence interval, in particular for the small horizons (h corresponding to one year). Thus, the method that we have developed is more precise than the classical ARMA-GARCH modelling.

3.2 Direct forecast

We now present a direct multistep ahead forecast method using AR(p) modelling and the previous k -NN method, Bansali (1997), Marcellino (2006), for which we provide the forecast error at each horizon h , in Table 6.

We assume that we have a data set $(X_t)_t$, sample size n . We provide a direct h step ahead forecast for X_{n+h} using an AR(p):

$$\hat{X}_{n+h} = \hat{\rho}_0 + \sum_{i=1}^p \hat{\rho}_i X_{n-i}, \quad (3.5)$$

where ρ_i , $i=0, \dots, p$, are real parameters. BIC criteria is used to determine the order of the AR process, the parameters $(\rho_i)_{i=0}^p$ are estimated using the ordinary least square method. The estimation is done for each horizon h , for which we minimize the relationship:

$$CSS(\rho_0^h, \rho_1^h, \dots, \rho_p^h) = \sum_{t=p}^{n-h} (X_{t+h} - \rho_0^h - \sum_{i=1}^p \rho_i^h X_{t-i})^2. \quad (3.6)$$

For each horizon h , we consider a training set composed by the observations until time $n - h$, to get the direct forecast of X_{n+h} by the k -NN method:

$$\hat{X}_{T+h} = \sum_{\underline{X}_{(i)} \in S, i \in N(\underline{x})} w(\underline{x} - \underline{X}_{(i)}) X_{(i)+h}. \quad (3.7)$$

Here, we work with no filtering data, because with this approach, if we carry out transformation on the data set, it is difficult to come back except by using an h order difference, but this last transformation seems meaningless for all horizons. Thus, in our application, we exclude the two time series with trends, IPI and RI. Indeed for these series modelling an $AR(p)$ has no sense.

Table 6

We observe that for each horizon h , the error computed using k -NN is smaller than the error computed with AR(p).

It is important to remark that in the classical literature concerning forecasting methodology, it seems that the "recursive" forecast method gives better results than the "direct forecasts" method when AR modelling is used. In our application, we found that this recursive AR is dominated by recursive k -NN and we obtain the same kind of result using direct forecast.

Finally, we have noticed that the time for computing confidence interval forecasting is generally smaller with ARMA modelling than with k -NN modelling, because with this latter method, we need to do some learning at each step. But for point forecasts, the opposite holds. Computation time by the k -NN is less than using the ARMA-GARCH modelling.

4 Conclusion

We have provided theoretical results on asymptotic normality and the confidence interval for the k -NN regression function estimate using non-*i.i.d* samples. We provide an application with new results, showing that the non-parametric method improves greatly the accuracy of forecasting obtained with ARMA-GARCH modelling. These interesting results on point forecasts by the k -NN method have been confirmed by some very good results for confidence intervals. Moreover, this method is also interesting because it permits some hidden behaviours of a series to be modelled.

A natural extension of this work concerns the comparison of this method with the radial basis function method, also developed in a multivariate setting. Another interesting point will be the comparison of density forecasts. The best way to choose the number of neighbors, the embedding dimension and the weighting functions for the k -NN method remains to be discussed.

We can also extend this method for data with trends permitting the parameters to evolve over time, making it more consistent and attractive for applications, when non-stationarity is detected.

APPENDIX: Proof of Theorem 2.1 and Corollary 2.1

We start giving the proof of theorem 2.1. To prove this theorem, we first establish a lemma which provides the relationship (2.11).

Lemma 4.1. *Under the hypotheses of theorem 2.1, either the estimate $m_n(\underline{x})$ is asymptotically*

unbiased or

$$E[m_n(\underline{x})] = m(\underline{x}) + O(n^{-\beta}) \quad (4.1)$$

with $\beta = \frac{(1-Q)p}{d}$.

Proof 4.1. We denote R_n the distance between \underline{x} and the $k(n)^{th}$ nearest neighbor of \underline{x} and $B(\underline{x}, r_0) = \{z \in \mathbb{R}^d, \|\underline{x} - z\| \leq r_0\}$ the ball centered at \underline{x} with radius $r_0 > 0$. To be sure that $k(n)$ observations fall in the ball $B(\underline{x}, r)$, we specify r . Since the function $h(\cdot)$ is p -continuously differentiable, for a given i the probability q_i of an observation \underline{x}_i to fall in $B(\underline{x}, r)$ is:

$$q_i = P(\underline{x}_i \in B(\underline{x}, r)) \quad (4.2)$$

$$= \int_{B(\underline{x}, r)} h(\underline{x}_i) d\underline{x}_i = h(\underline{x}) \cdot \int_{B(\underline{x}, r)} d\underline{x}_i + \int_{B(\underline{x}, r)} (h(\underline{x}_i) - h(\underline{x})) d\underline{x}_i \quad (4.3)$$

$$= h(\underline{x}) c r^d + o(r^d), \quad (4.4)$$

where c is the volume of the unit ball and $\underline{x} = dx_1 dx_2 \cdots dx_d$. Thus, $q_i - q_j = o(r^d)$ for all $i \neq j$.

We consider now the k th-NN vectors $\underline{x}_{(k)}$ and we denote q the probability that they are in the ball $B(\underline{x}, r)$, $q = P(\underline{x}_{(k)} \in B(\underline{x}, r))$, then :

$$q_i = q + o(r^d). \quad (4.5)$$

Being given $N(r, n)$, the number of observations falling in the ball $B(\underline{x}, r)$, for a given $r > 0$, we consider r such that $k(n)$ observations fall in $B(\underline{x}, r)$. We proceed as follows: we denote \mathcal{S}_i^n all non ordered combinations of the i -uple indices from n indices, then:

$$\begin{aligned} E[N(r, n)] &= \sum_{i=0}^{n-d} i P(N(r, n) = i) = \sum_{i=0}^{n-d} i \sum_{(j_1, \dots, j_i) \in \mathcal{S}_i^n} \prod_{j=j_1}^{j_i} q_j \prod_{\substack{\ell=1 \\ \ell \notin \{j_1, \dots, j_i\}}}^{n-d} (1 - q_\ell) \\ &\geq \sum_{i=0}^n i \sum_{(j_1, \dots, j_i) \in \mathcal{S}_i^n} \underline{q}^i (1 - \bar{q})^{n-d-i} = \sum_{i=0}^n i \binom{n}{i} \underline{q}^i (1 - \bar{q})^{n-d-i} \\ &= \underline{q}(n-d)(1 + \underline{q} - \bar{q})^{n-d}, \end{aligned} \quad (4.6)$$

where \underline{q} and \bar{q} are respectively the smallest and largest probabilities q_i $i = 1, \dots, n-d$. Thus, we obtain a lower bound for $E[N(r, n)]$. If $E[N(r, n)] = k(n)$, using (4.4) - (4.6), we obtain:

$$r \leq \left(\frac{k(n)}{(n-d)} \right)^{\frac{1}{d}} D(\underline{x}), \quad (4.7)$$

with $D(\underline{x}) = \left(\frac{1}{h(\underline{x})c} \right)^{\frac{1}{d}}$.

Now, using the relationship (3.5), we know that:

$$E[m_n(\underline{x})] = \sum_{i \in N(\underline{x})} E[w(\underline{x} - \underline{X}_{(i)})Y_i], \quad (4.8)$$

where $Y_i = X_{(i)+1}$. We can remark that $E[w(\underline{x} - \underline{X}_{(i)})Y_i] = \int_{\mathbb{R}^d} \int_{\mathbb{R}} w(\underline{x} - \underline{x}_i) y f(y, \underline{x}_i) d\underline{x}_i dy$. Since $f(y, \underline{x}_i) = f(y | \underline{x}_i) h(\underline{x}_i)$, then we get $E[w(\underline{x} - \underline{X}_{(i)})Y_i] = \int_{\mathbb{R}^d} \int_{\mathbb{R}} w(\underline{x} - \underline{x}_i) y f(y | \underline{x}_i) h(\underline{x}_i) d\underline{x}_i dy$. Thus, as soon as the weighting function $w(\cdot)$ is vanishing outside the ball $B(\underline{x}, r)$:

$$E[w(\underline{x} - \underline{X}_{(i)})Y_i] = \int_{B(\underline{x}, r)} w(\underline{x} - \underline{x}_i) \left(\int_{\mathbb{R}} y f(y | \underline{x}_i) dy \right) h(\underline{x}_i) d\underline{x}_i \quad (4.9)$$

$$= \int_{B(\underline{x}, r)} w(\underline{x} - \underline{x}_i) m(\underline{x}_i) h(\underline{x}_i) d\underline{x}_i. \quad (4.10)$$

To compute the bias we need to evaluate: $E[m_n(\underline{x})] - m(\underline{x})$. We begin to evaluate :

$$\sum_{i \in N(\underline{x})} \int_{B(\underline{x}, r)} w(\underline{x} - \underline{x}_i) m(\underline{x}_i) g(\underline{x}_i) d\underline{x}_i = m(\underline{x}) E \left[\sum_{i \in N(\underline{x})} w(\underline{x} - \underline{X}_{(i)}) \right] = m(\underline{x}). \quad (4.11)$$

Then,

$$E[m_n(\underline{x})] - m(\underline{x}) = \sum_{i \in N(\underline{x})} \int_{B(\underline{x}, r)} w(\underline{x} - \underline{x}_i) (m(\underline{x}_i) - m(\underline{x})) h(\underline{x}_i) d\underline{x}_i. \quad (4.12)$$

The equation (4.12) holds because $\sum_{i \in N(\underline{x})} \int_{B(\underline{x}, r)} w(\underline{x} - \underline{x}_i) = 1$, (Assumption (iv) in Theorem 3.1). Then,

$$|E[m_n(\underline{x})] - m(\underline{x})| \leq \sum_{i \in N(\underline{x})} \int_{B(\underline{x}, r)} w(\underline{x} - \underline{x}_i) a \|\underline{x}_i - \underline{x}\|^p g(\underline{x}_i) d\underline{x}_i. \quad (4.13)$$

We get this last expression since a is a known constant and $m(\cdot)$ is p -continuously differentiable. Noticing that $\|\underline{x}_i - \underline{x}\|^p < r^p$, as soon as $\underline{x}_i \in B(\underline{x}, r)$, the previous inequality gives:

$$|E[m_n(\underline{x})] - m(\underline{x})| \leq ar^p E \left[\sum_{i \in N(\underline{x})} w(\underline{x} - \underline{X}_{(i)}) \right]. \quad (4.14)$$

Now, two cases can be considered:

1. When r is very small, the bias is negligible and $E[m_n(\underline{x})] = m(\underline{x})$.

2. If the bias is not negligible, using (4.7) and (4.14), we get:

$$|E[m_n(\underline{x})] - m(\underline{x})| \leq a \left(\frac{k(n)}{(n-d)} \right)^{\frac{p}{d}} D(\underline{x})^p. \quad (4.15)$$

If we chose $k(n)$ as an integer part of n^Q , and the fact that $\frac{k}{n-d} \sim \frac{k}{n}$, then $|E[m_n(\underline{x})] - m(\underline{x})| = O(n^{-\beta})$ with $\beta = \frac{(1-Q)p}{d}$.

The proof of lemma 4.1 is finished.

We now prove the theorem 2.1.

Proof 4.2. 1. We begin to prove the relationship (2.11). In the following, we denote $Y_i = X_{(i)+1}$. We rewrite the left part of (2.11) as:

$$E[(m_n(\underline{x}) - m(\underline{x}))^2] = \text{Var}(m_n(\underline{x})) + (E[m_n(\underline{x})] - m(\underline{x}))^2. \quad (4.16)$$

We first compute the variance of $m_n(\underline{x})$, considering two cases:

a) First case: The weights w_i , $i = 1, \dots, k$, are independent of $\{X_n\}$. In that case the variance of $m_n(\underline{x})$ is equal to:

$$\text{Var}(m_n(\underline{x})) = A + B, \quad (4.17)$$

where $A = \sum_{i=1}^{k(n)} w_i^2 \text{Var}(Y_i)$ and $B = \sum_{i=1}^{k(n)} \sum_{j \neq i} w_i w_j \text{cov}(Y_i, Y_j)$. Using the assumption (ii) of theorem 3.1, we get $|B| \leq \sum_{i=1}^{k(n)} \sum_{j \neq i} |\text{cov}(Y_i, Y_j)|$. This last term is negligible due to Yakowitz' result (1987) on the sum of covariances. Now, $A = \frac{1}{k(n)^2} \sum_{i=1}^{k(n)} (k(n)w_i)^2 (v(\underline{x}) + (E[Y_i] - m(\underline{x}))^2)$. Using the fact that the weights are decreasing with respect to the chosen distance, $w_k \leq \dots \leq w_1$, then we get:

$$\frac{1}{k(n)^2} \sum_{i=1}^{k(n)} (k(n)w_k)^2 (v(\underline{x}) + (E[Y_i] - m(\underline{x}))^2) \leq A \leq \frac{1}{k(n)^2} \sum_{i=1}^{k(n)} (k(n)w_1)^2 (v(\underline{x}) + (E[Y_i] - m(\underline{x}))^2). \quad (4.18)$$

As soon as $k(n) \rightarrow \infty$, the product $k(n)w_i$ converges to one, and there exist two positive constants c_0 and c_1 , such that (4.18) becomes :

$$\frac{c_1^2}{k(n)^2} \sum_{i=1}^{k(n)} (v(\underline{x}) + (E[Y_i] - m(\underline{x}))^2) \leq A \leq \frac{c_0^2}{k(n)^2} \sum_{i=1}^{k(n)} \text{var}(Y_i). \quad (4.19)$$

Under the assumption (iv) of Theorem 3.1, $E[Y_i] = E[m_n(\underline{x})]$, then using lemma 4.1, and if $k(n) = [n^Q]$, $A = O(n^{-Q})$ when $n \rightarrow \infty$. It follows that equation (4.17) gives:

$$\text{Var}(m_n(\underline{x})) = O(n^{-Q}), \quad (4.20)$$

and

$$(E[m_n(\underline{x}) - m(\underline{x})])^2 = O(n^{-2\beta}). \quad (4.21)$$

Plugging equations (4.20) and (4.21) inside equation (4.16), we get $2\beta = Q$ or $Q = \frac{2p}{2p+d}$ and the proof is complete.

b) Second case: the weights w_i , $i = 1, \dots, k$, depend on (X_n) . We use again the expression (4.17) with $A = \sum_{i=1}^{k(n)} \text{Var}(w(x - \underline{X}_{(i)})Y_i)$ and $B = \sum_{i=1}^{k(n)} \sum_{j \neq i} \text{cov}(w(x - \underline{X}_{(i)})Y_i, w(x - \underline{X}_{(j)})Y_j)$. Since the weights are less than one, we get $|B| \leq \sum_{i=1}^{k(n)} \sum_{j \neq i} |\text{cov}(Y_i, Y_j)|$ and B is negligible from Yakowitz' result (1987). Now we remark that $A = \sum_{i=1}^{k(n)} (E[(w(x - \underline{X}_{(i)})Y_i)^2] - (E[w(x - \underline{X}_{(i)})Y_i])^2)$, then

$$A = \sum_{i=1}^{k(n)} \left[\int_{\mathbb{R}^d} \int_{\mathbb{R}} w(\underline{x} - \underline{x}_i)^2 y_i^2 f(y_i, \underline{x}_i) d\underline{x}_i dy_i - \left(\int_{\mathbb{R}^d} \int_{\mathbb{R}} w(\underline{x} - \underline{x}_i) y_i f(y_i, \underline{x}_i) d\underline{x}_i dy_i \right)^2 \right]. \quad (4.22)$$

When k increases, the weights w_i decrease, and $k(n)w_i \sim \gamma$ where γ is a constant, and

$$A = \frac{\gamma^2}{k(n)^2} \sum_{i=1}^{k(n)} \left[\int_{\mathbb{R}^d} \int_{\mathbb{R}} y_i^2 f(y_i, \underline{x}_i) d\underline{x}_i dy_i - \left(\int_{\mathbb{R}^d} \int_{\mathbb{R}} y_i f(y_i, \underline{x}_i) d\underline{x}_i dy_i \right)^2 \right] \quad (4.23)$$

$$= \frac{\gamma^2}{k(n)^2} \sum_{i=1}^{k(n)} (E[Y_i^2] - E[Y_i]^2). \quad (4.24)$$

Assuming that (X_n) is stationary, the expression (4.24) becomes $A = \frac{\gamma^2}{k(n)} (E[Y_n^2] - E[Y_n]^2)$ and $A = \frac{\gamma^2}{k(n)} \text{Var}(Y_n)$. Finally expression (4.17) is equal to:

$$\text{Var}(m_n(\underline{x})) = \frac{\gamma^2}{k(n)} \text{Var}(Y_n). \quad (4.25)$$

Furthermore, $\text{Var}(Y_n)$ goes to $\text{Var}(Y_n | \underline{X}_n)$ when $n \rightarrow \infty$, for large d and negligible bias, and as soon as $k(n) = n^Q$, the relationship (4.25) becomes:

$$\text{Var}(m_n(\underline{x})) = O(n^{-Q}). \quad (4.26)$$

Plugging equations (4.26) and (4.21) in (4.16), we get $2\beta = Q$, and $Q = \frac{2p}{2p+d}$. Thus, the proof is complete.

2. We prove now the asymptotic normality. We assume that $\sigma_n = \text{var}[m_n(\underline{x})]$ exists and is non null, thus:

$$\frac{m_n(\underline{x}) - Em_n(\underline{x})}{\sigma_n} = \sum_{i=1}^{k(n)} \frac{w_i Y_i - Ew_i Y_i}{\sigma_n}. \quad (4.27)$$

To establish the asymptotic normality of $m_n(\underline{x})$, we distinguish three steps corresponding to three different weighting functions.

i) The weights are uniform, $w_i = \frac{1}{k(n)}$, then (4.27) becomes:

$$\frac{m_n(\underline{x}) - Em_n(\underline{x})}{\sigma_n} = \sum_{i=1}^{k(n)} \frac{1}{k(n)} Z_i, \quad (4.28)$$

where $Z_i = \frac{Y_i - EY_i}{\sigma_n}$. The asymptotic normality of (4.28) is obtained using theorem 2.2 in Peligrad and Utev (1997). To compute the variance, we use Yakowitz result (1987). $\text{var}(m_n(\underline{x})) = \frac{1}{k(n)^2} \text{var}(\sum_{i=1}^{k(n)} Y_i) = \frac{\text{var}(Y|X=\underline{x})}{k(n)}$, then (4.28) becomes,

$$\frac{m_n(\underline{x}) - Em_n(\underline{x})}{\sigma_n} = \sqrt{nQ} \sum_{i=1}^{k(n)} \frac{w_i Y_i - Ew_i Y_i}{\sigma}, \quad (4.29)$$

when $k(n) = [n^Q]$ and $\sigma^2 = \text{var}(Y | X = \underline{x})$. The proof is complete.

ii) The weights w_i are real numbers and do not depend on $(X_n)_n$, then

$$\frac{m_n(\underline{x}) - Em_n(\underline{x})}{\sigma_n} = \sum_{i=1}^{k(n)} w_i Z_i, \quad (4.30)$$

where $Z_i = \frac{Y_i - EY_i}{\sigma_n}$. Now, we apply again the theorem 2.2 Peligrad and Utev (1997) to get the asymptotic normality, remarking that $E[\sum_{i=1}^{k(n)} w_i Z_i] = 0$ and $\text{Var}[\sum_{i=1}^{k(n)} w_i Z_i] = 1$. To obtain $\sigma_n^2 = \text{Var}[m_n(x)]$, using the stationarity of $(X_n)_n$, we remark that :

$$\text{Var}[m_n(\underline{x})] = \sum_{i=1}^{k(n)} w_i^2 \text{Var}[Y_i] = \sum_{i=1}^{k(n)} w_i^2 [\text{Var}[Y_{n+1} | X_n = \underline{x}] + B^2],$$

where B is given in lemma 3.1. Remarking that $\frac{1}{k(n)^2} \sum_{i=1}^{k(n)} (k(n)w_i)^2$ is finite, then the serie $\sum_{i=1}^{k(n)} w_i^2$ is also finite and

$$\text{Var}[m_n(\underline{x})] = [\text{Var}[Y_i | X_i = \underline{x}] + B^2] \sum_{i=1}^{k(n)} w_i^2.$$

As soon as $\sum_{i=1}^{k(n)} w_i^2 \sim \frac{\gamma^2}{k(n)}$, and $k(n) = [n^Q]$, we get the result.

iii) Finally, we assume that the weights $w_i = \frac{w(\underline{x} - \underline{X}_{(i)})}{\sum_{i=1}^K w(\underline{x} - \underline{X}_{(i)})}$ where $w(\cdot)$ is a given function. In that latter case, the weights depend on the process $(X_n)_n$. In the following, we denote by $N(i)$ the order of the i^{th} neighbor. We rewrite the neighbor indices in an increasing order such that $M(1) = \min\{N(i), 1 \leq i \leq K\}$ and $M(k) = \min\{N(i) \notin \{M(j), \forall j < k\}, 1 \leq i \leq K\}$ for $2 \leq k \leq K$ and $K = k(n)$ is the number of neighbors. We introduce a real triangular sequence $\{\alpha_{Ki}, 1 \leq i \leq K \text{ and } \alpha_{Ki} \neq 0 \forall i\}$ such that

$$\sup_K \sum_{i=1}^K \alpha_{Ki}^2 < \infty \quad \text{and} \quad \max_{1 \leq i \leq K} |\alpha_{Ki}| \xrightarrow{n \rightarrow \infty} 0. \quad (4.31)$$

Now using the sequences $M(j), j = 1, \dots, K$ and $(\alpha_{Ki}), 1 \leq i \leq K$, we can rewrite (4.27) as:

$$\frac{m_n(\underline{x}) - Em_n(\underline{x})}{\sigma_n} = \sum_{i=1}^K \alpha_{Ki} S_i, \quad (4.32)$$

with $S_i = \frac{w_{M(i)} X_{M(i)+1} - E w_{M(i)} X_{M(i)+1}}{\alpha_{Ki} \sigma_n}$. The sequence (S_i^2) is uniformly integrable and each S_i is function only of $(X_j, j \leq M(i) + 1)$, thus if we denote $\mathcal{F}_i, \mathcal{G}_i, \mathcal{F}_i^j$ and \mathcal{G}_i^j , the sigma algebras generated, respectively, by $\{X_r\}_{r \leq i}, \{S_r\}_{r \leq i}, \{X_r\}_{r=i}^j$ and $\{S_r\}_{r=i}^j$, then $S_i \in \mathcal{F}_{M(i)+1}$, and $\mathcal{G}_i \subset \mathcal{F}_{M(i)+1}$. For a given integer ℓ , we have also $\mathcal{G}_{n+\ell}^\infty \subseteq \mathcal{F}_{n+M(\ell)+1}^\infty$ since $M(1) < M(1) + 1 \leq M(2) < \dots \leq M(n + \ell) < M(n + \ell) + 1 \leq M(n + \ell + 1)$. Finally, we get:

$$\sup_{\ell} \sup_{A \in \mathcal{G}_1^\ell, B \in \mathcal{G}_{n+\ell}^\infty, P(A) \neq 0} |P(B | A) - P(B)| \leq \sup_{\ell} \sup_{A \in \mathcal{F}_1^{M(\ell)+1}, B \in \mathcal{F}_{n+M(\ell)+1}^\infty, P(A) \neq 0} |P(B | A) - P(B)|. \quad (4.33)$$

Under the ϕ -mixing assumption on $(X_n)_n$, the right hand part of the expression (4.33) tends to zero as $n \rightarrow \infty$ and the left hand part of (4.33) converges to zero, hence the sequence (S_i) is ϕ -mixing. Moreover:

$$S_i \text{ is centered and } \text{var}\left(\sum_{i=1}^K \alpha_{Ki} S_i\right) = \text{var}\left(\frac{m_n(\underline{x})}{\sigma_n}\right) = 1. \quad (4.34)$$

Then using expressions (4.31) - (4.34) and the theorem 2.2 in Peligrad and Utev (1997), we get:

$$\frac{m_n(\underline{x}) - Em_n(\underline{x})}{\sigma_n} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad (4.35)$$

The variance of $m_n(x)$ is given by the relation (4.25). The proof of the theorem 3.1 is complete.

We provide now the proof of the Corollary 3.1.

Proof 4.3 (Proof of corollary 2.1). *From theorem 2.1, a confidence interval, for a given α can be computed:*

$$-z_{1-\frac{\alpha}{2}} \leq \frac{m_n(\underline{x}) - Em_n(\underline{x})}{\hat{\sigma}_n} \leq z_{1-\frac{\alpha}{2}} \quad (4.36)$$

where $z_{1-\frac{\alpha}{2}}$ is the $(1-\frac{\alpha}{2})$ quantile of Student law, (because we use for σ^2 an estimate). Previously, we have seen that $\hat{m}_n(\underline{x})$ is biased, thus the equation (4.36) becomes:

$$m_n(\underline{x}) + B - \hat{\sigma}_n z_{1-\frac{\alpha}{2}} \leq m(\underline{x}) \leq m_n(\underline{x}) + B + \hat{\sigma}_n z_{1-\frac{\alpha}{2}} \quad (4.37)$$

When the bias is negligible, the corollary is established. If this bias is not negligible, we bound it. Using (4.7) and (4.14), we obtain:

$$B = \left(\frac{k(n)}{(n-d)\hat{h}(\underline{x})c} \right)^{\frac{p}{d}} \quad (4.38)$$

with $c = \frac{\pi^{d/2}}{\Gamma((d+2)/2)}$, $\hat{h}(\underline{x})$ being an estimate of the density $h(\underline{x})$. Introducing this bound in (4.37) permits to complete the proof.

References

- Bernanke BS., Boivin J (2003) Monetary policy in a data-rich environment, *Journal of Monetary Economics*, 50, 525 - 546.
- Bhansali RJ. 1997. Direct Autoregressive predictions for Multistep Prediction: Order Selection and Performance Relative to the Plug In Predictors. *Statistica Sinica* **7** : 425 - 449.
- Carrasco M, Chen X. 2002. Mixing and moment properties of various GARCH and Stochastic Volatility models. *Econometric Theory* **18** : 17-39.
- Collomb G. 1984. Nonparametric time series analysis and prediction: Uniform almost sure convergence of the window and k-NN autoregression estimates. *Math Oper. Stat., Ser. Statistics* 1984.
- Devroye LP. 1978. The uniform convergence of nearest neighbor regression function estimators and their application in optimization. *IEEE Transaction Information Theory* **61** : 467-481.

- Devroye LP. 1982. Necessary and sufficient conditions for the pointwise convergence of nearest neighbor regression function estimates. *Z. Wahrscheinlichkeitstheorie verw Gebiete* **61** : 467-481.
- M. Diron (2008) Short-term forecasts of Euro area real GDP growth: an assessment of real-time performance based on vintage data, *Journal of Forecasting*, Vol. 27, Issue 5, pp. 371-390.
- M. Forni, D. Giannone, M. Lippi, L. Reichlin (2005), Opening the black box: structural factor models with large cross-sections, European Central Bank, WP 571.
- G. Kapetanios, M. Marcellino (2006), A parametric estimation method for dynamic factors models of large dimensions, IGIER WP No 305, Bocconi University, Italy.
- Guégan D. 1983. Une condition d'ergodicité pour des modèles linéaires à temps discret. *CRAS Série 1* : 297-301.
- Guégan D. 2003. *Les Chaos en Finance: Approche Statistique*. Economica Série Statistique Mathématique et Probabilité: Paris.
- Hardle W, Muller M, Sperlich S, Werwatz A. 2004. *Non-parametric and semi-parametric models*. Springer Verlag.
- Mack YP. 1981. Local properties of k -NN regression estimates. *S.I.A.M. J. Alg. Disc. Meth.* **2** : 311-323.
- Marcellino M, Stock JH, Watson MW. 2006. A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series. *Journal of Econometrics* **135** : 499-526.
- Nadaraya EA. 1964. On estimating regression. *Theory of probability and its applications* **10** : 186-190.
- Peligrad M, Utev SA. 1997. Central Limit Theorem for Linear Processes. *The Annals of Probability* **25** : 443-456.
- Rosenblatt M. 1956. Remarks on some nonparametric estimates of a density function. *Annals of mathematical Statistics* **27** : 642-669.
- Silverman BW. 1986. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall: London.

Stone (1977)

Stone C. 1977. Consistent non parametric regression. *Annals of Statistics* **5** : 595-645.

Stute W. 1984. Asymptotic normality of nearest neighbor regression function estimates. *Annals of Statistics* **12** : 917-926.

Watson GS. 1964. Smooth regression analysis. *Sankhya Series A* **26** : 359-372.

Yakowitz S. 1987. Nearest neighbors method for time series analysis. *Journal of Time Series Analysis* **8** : 235-247.

Table 1: Descriptive statistics for the six economic indicators

	N°obs	Mean	Std dev	Skewness	Kurtosis	Min	Max
IPI	213	93.77	9.27	0.16	1.75	78.98	113.37
IPIC	213	96.75	5.56	-0.26	3.27	78.06	111.75
ICI	275	-6.10	8.27	-0.66	3.45	-30.20	6.60
RCI	275	-6.41	6.82	-0.30	2.61	-27	7.4
RI	153	100.50	6.82	-0.12	1.71	90.63	110.64
CI	275	-11.07	6.72	-0.39	3.03	-29	2

Table 2: ARIMA-GARCH estimated coefficients for the six indices.

ARIMA(p,d,q)	ARIMA coefficients					GARCH coefficients		
	ϕ_0	ϕ_1	ϕ_2	ϕ_3	θ_1	a_0	a_1	b_1
IPI(3,1,0)	0.11	-0.33	0.03	0.29				
Std error	(0.04)	(0.07)	(0.07)	(0.06)				
IPIC(2,1,0)	0.06	-0.57	-0.20			0.37	0.88	
Std error	(0.10)	(0.07)	(0.07)			(0.04)	(0.18)	
ICI(3,1,0)	0.02	0.43	0.13	0.16				
Std error	(0.24)	(0.06)	(0.07)	(0.06)				
RCI(2,1,1)	0.08	-0.69	-0.31		0.42	0.12	0.14	0.73
Std error	(0.11)	(0.36)	(0.08)		(0.39)	(0.06)	(0.05)	(0.09)
RI(2,1,1)	0.12	-0.48	-0.34		-0.27			
Std error	(0.01)	(0.18)	(0.11)		(0.17)			
CI(3,1,0)	0.01	0.13	0.06	0.21				
Std error	(0.14)	(0.06)	(0.06)	(0.06)				

Table 3: p-value of Ljung-Box test for residuals of the six modellings.

Portmanteau test p-value for residuals						
Serie\lag	1	2	3	4	5	squared residual
IPI	0.51	0.77	0.89	0.24	0.23	0.82
IPIC	0.60	0.59	0.16	0.15	0.20	0.00
ICI	0.90	0.98	0.99	0.93	0.94	0.49
RCI	0.85	0.88	0.97	0.69	0.25	0.02
RI	0.95	0.99	0.98	0.97	0.97	0.69
CI	0.75	0.93	0.98	0.68	0.75	0.57

These are the p-values associated to Ljung-Box statistic to test the null non correlation of residuals for different lag values from one to five. The last column corresponds to the p-values associated to the test of non correlation of order one of squared residuals.

Table 4: ARMA and kNN RMSE forecasts over two years for the six indices.

Serie	Model\h	RMSE over horizons h											
		1	2	3	4	5	6	9	12	15	18	21	24
IPI	kNN(4,2)	0.44	0.41	0.36	0.42	0.38	0.47	1.09	1.32	1.36	1.35	1.36	1.66
	ARIMA(3,1,0)	0.58	0.71	0.66	0.95	1.05	1.26	2.16	2.69	3.22	3.77	4.22	4.85
IPIC	kNN(6,2)	0.74	0.75	0.67	1.44	1.59	1.65	1.65	1.59	1.70	1.86	2.32	3.19
	ARIMA(2,1,0)	1.17	1.10	1.02	1.18	1.12	1.81	2.50	3.80	5.33	5.20	5.23	5.25
	-ARCH(1)												
ICI	kNN(10,3)	0.96	0.73	0.78	1.14	1.69	1.67	1.65	1.58	1.71	1.82	1.75	1.83
	ARIMA(3,1,0)	1.76	2.15	3.07	3.98	5.09	5.87	7.41	8.77	9.58	10.33	10.47	10.67
RCI	kNN(10,4)	0.04	1.91	1.65	1.74	2.67	2.45	3.15	3.79	3.56	3.51	3.62	3.66
	ARIMA(2,1,0)	2.30	1.74	2.14	2.78	4.03	4.15	5.37	6.32	6.13	6.05	6.20	6.12
	-GARCH(1,1)												
RI	kNN(6,5)	0.15	0.13	0.13	0.49	0.57	0.53	0.74	0.93	1.03	0.98	0.92	0.90
	ARIMA(2,1,0)	0.65	0.49	0.40	0.60	0.67	0.61	0.94	1.10	1.24	1.37	1.42	1.51
CI	kNN(9,2)	1.38	1.75	1.56	1.37	1.23	1.17	1.04	1.02	1.12	1.73	1.98	2.03
	ARIMA(3,1,0)	2.07	2.02	2.22	2.15	2.20	2.56	3.15	3.76	4.60	5.77	6.49	6.60

These RMSEs of the six indicators have been computed over two years observations. For k -NN, we use exponential weight and in the notation $kNN(\cdot, \cdot)$ the first coordinate corresponds to the embedding dimension d and the second to the number of neighbors k .

Table 5: ARMA and kNN recursive forecast errors for the six indices.

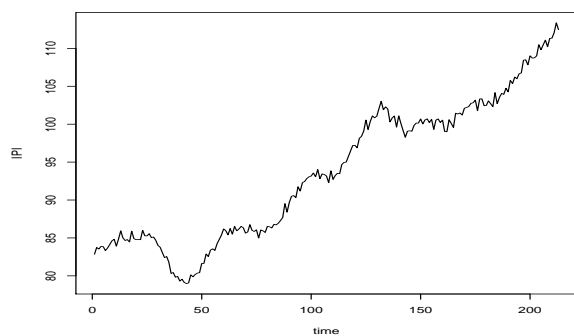
Serie	Model\h	Forecast error for each horizon h											
		1	2	3	4	5	6	9	12	15	18	21	24
IPI	kNN(4,2)	0.44	0.39	0.22	0.55	0.15	0.81	1.79	1.65	2.02	1.72	1.99	2.84
	ARIMA(3,1,0)	0.58	0.81	0.57	1.52	1.37	2.00	3.79	4.03	5.78	6.39	6.60	7.77
IPIC	kNN(6,2)	0.74	0.77	0.48	2.63	2.08	1.97	1.17	2.58	1.77	3.77	3.91	4.42
	ARIMA(2,1,0)	1.17	1.03	0.87	1.35	1.16	1.62	3.01	3.32	6.06	9.28	4.49	4.68
	-ARCH(1)												
ICI	kNN(10,3)	0.96	0.23	0.93	1.82	3.01	1.52	2.44	1.24	2.53	1.54	1.07	3.86
	ARIMA(3,1,0)	1.76	2.18	3.06	4.01	5.12	5.76	7.38	8.31	9.11	10.51	10.51	10.30
RCI	kNN(10,4)	0.04	2.17	0.91	1.98	4.65	0.63	6.05	3.87	1.57	3.77	3.61	3.41
	ARIMA(2,1,0)	2.30	0.87	2.73	4.41	7.12	4.84	8.44	4.38	6.67	7.17	2.35	6.37
	-GARCH(1,1)												
RI	kNN(6,5)	0.15	0.11	0.13	0.96	0.81	0.33	0.93	1.07	1.34	1.08	0.22	0.35
	ARIMA(2,1,0)	0.65	0.27	0.06	0.94	0.91	0.01	1.27	1.08	2.19	2.06	1.63	2.07
CI	kNN(9,2)	1.38	2.04	1.11	0.48	0.07	0.82	0.35	1.61	2.07	4.74	3.58	3.78
	ARIMA(3,1,0)	2.07	1.98	2.56	1.95	2.39	3.87	4.25	5.64	8.04	8.84	9.14	4.84

These errors have been computed over two year observations. The k -NN model is denoted by kNN(d,k) where d is the embedding dimension and k the number of neighbors and it is with exponential weight.

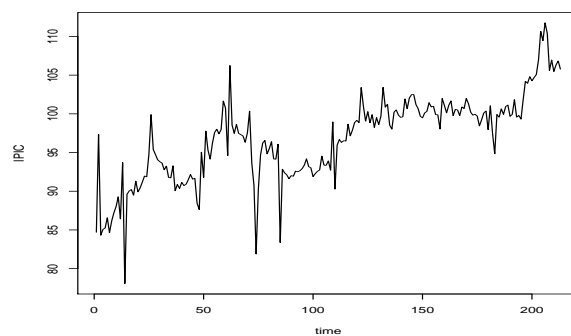
Table 6: ARMA and kNN direct forecast errors for the six indices.

Serie	Model\h	Forecast error for each horizon h											
		1	2	3	4	5	6	9	12	15	18	21	24
IPIC	(d,k)	(2,2)	(6,2)	(5,6)	(6,2)	(4,4)	(6,2)	(5,2)	(4,2)	(7,6)	(4,2)	(5,2)	(7,4
	kNN(d,k)	0.23	0.54	0.29	1.14	0.04	4.61	4.80	4.21	5.22	3.85	4.77	5.26
	p	6	6	6	6	5	6	5	4	5	4	6	5
	AR(p)	0.72	0.43	1.71	0.44	0.18	0.64	4.52	5.51	11.57	11.44	6.52	6.65
ICI	(d,k)	(10,3)	(2,8)	(2,3)	(9,3)	(6,2)	(3,5)	(10,4)	(9,3)	(9,3)	(9,3)	(6,4)	(10,2)
	kNN	0.92	0.008	0.93	1.14	1.07	0.78	0.13	0.27	1.19	0.25	0.56	0.65
	p	4	5	5	5	5	6	6	6	6	6	6	6
	AR(p)	1.59	2.19	3.97	5.47	7.68	7.91	8.84	13.34	13.32	13.50	12.50	9.36
RCI	(d,k)	(2,3)	(2,10)	(3,5)	(2,10)	(2,10)	(5,2)	(6,2)	(3,2)	(2,4)	(7,7)	(5,4)	(5,3)
	kNN(d,k)	0.02	2.02	0.25	0.03	0.008	0.02	6.13	1.26	0.35	0.49	0.08	0.05
	p	6	6	6	6	6	6	6	6	6	6	6	6
	AR(p)	2.04	1.40	1.97	3.45	5.94	3.48	7.03	6.77	3.54	7.60	8.79	8.06
CI	(d,k)	(2,2)	(2,3)	(2,2)	(3,2)	(2,2)	(8,4)	(2,9)	(4,9)	(3,3)	(2,2)	(4,3)	(3,3)
	kNN(d,k)	1.33	0.04	0.18	2.68	2.67	2.40	0.44	0.04	0.93	0.70	0.15	0.10
	p	5	5	4	5	5	5	6	6	6	6	6	6
	AR(p)	1.79	1.53	2.02	1.05	1.40	2.63	2.72	3.79	5.92	9.52	7.54	3.03

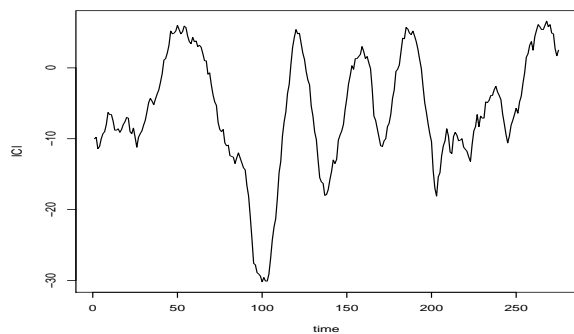
ARMA and kNN direct forecast errors for the six indices over last two years. The notation (d,k) corresponds to the embedding dimension d and the number of neighbors k of an exponentially weighted k -NN and p is for the autoregressive order of AR.



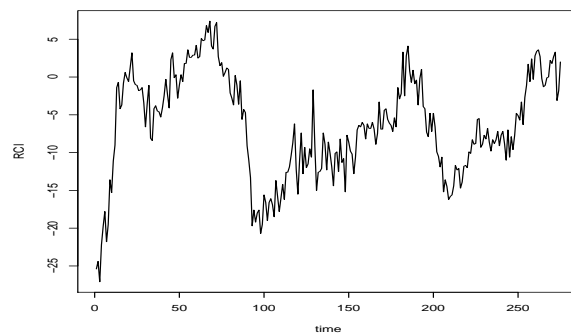
(a) Monthly IPI from 01-1990 to 09-2007



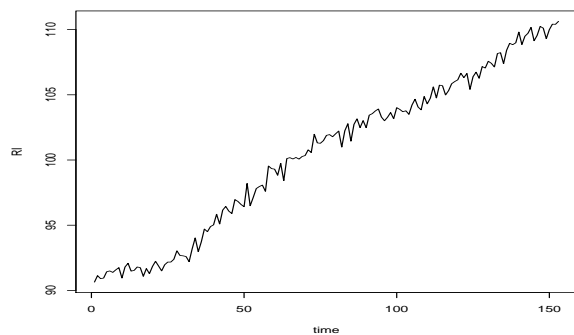
(b) Monthly IPIC from 01-1990 to 09-2007



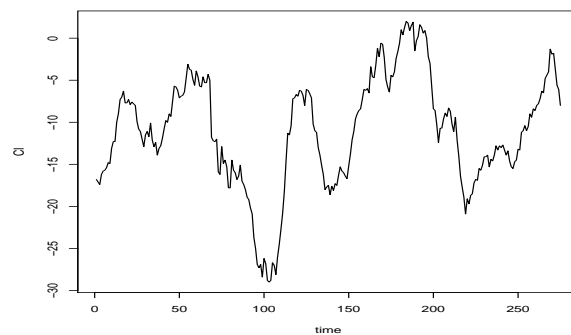
(c) Monthly ICI from 01-1985 to 11-2007



(d) Monthly RCI from 01-1985 to 11-2007

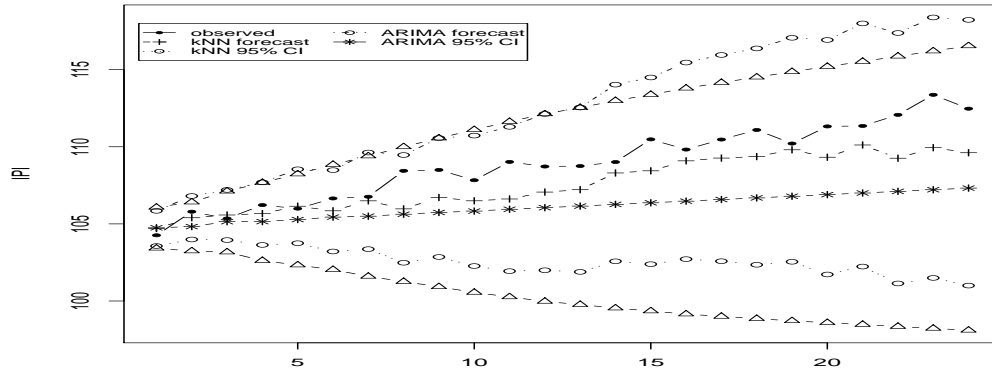


(e) Monthly RI from 01-1995 to 09-2007

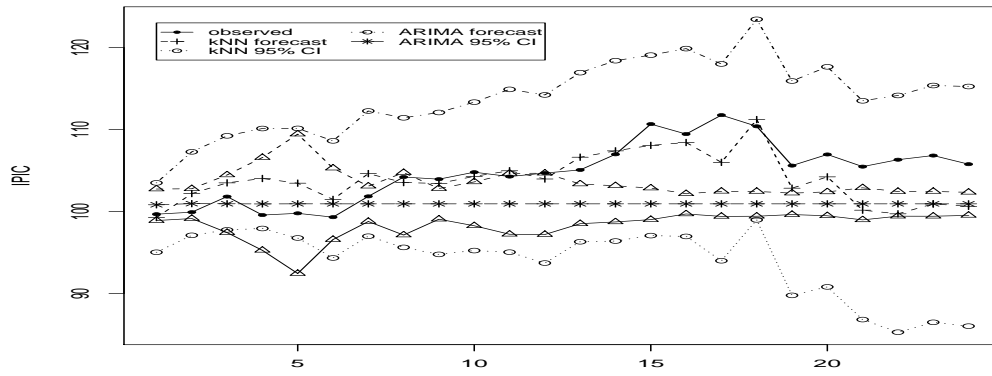


(f) Monthly CI from 01-1985 to 11-2007

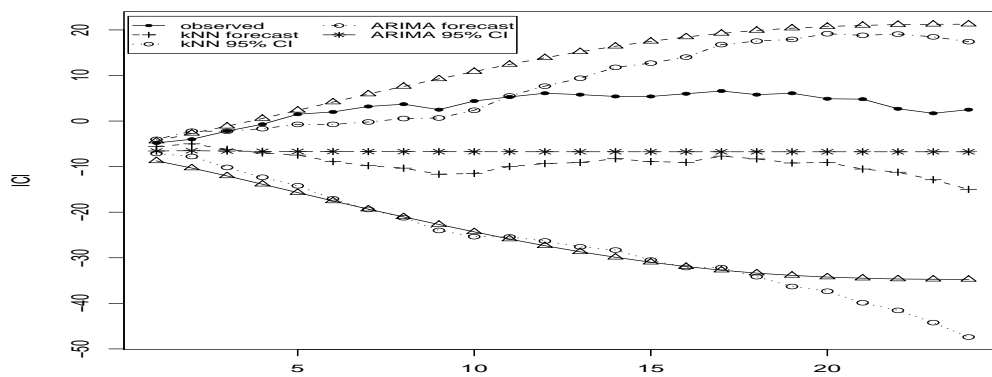
Figure 1: Six macroeconomic indicators of the Euro Area.



(a) 95% confidence interval of IPI forecast by k -NN and ARMA-GARCH from 10-2005 to 09-2007.

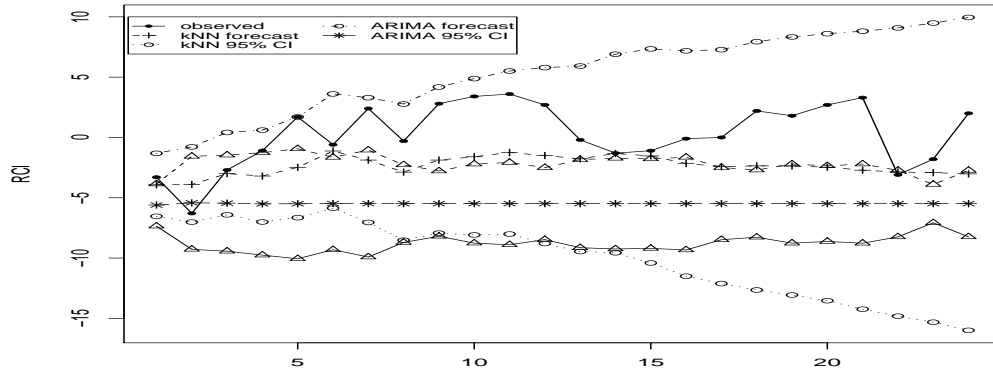


(b) 95% confidence interval of IPIC forecast by k -NN and ARMA-GARCH from 10-2005 to 09-2007.

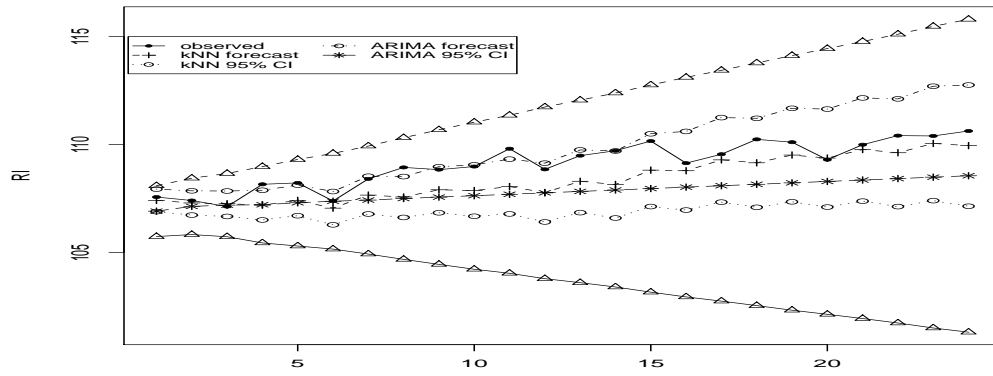


(c) 95% confidence interval of ICI forecast by k -NN and ARMA-GARCH from 12-2005 to 11-2007.

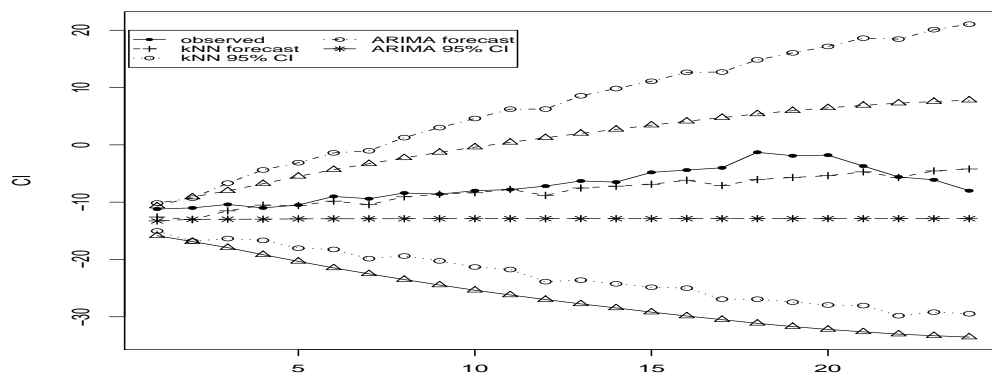
Figure 2: IPI, IPIC and ICI observed values (dashed) and forecasted values by k -NN (plus sign) and by ARMA-GARCH (star) with their 95% Confidence intervals (circle) and (triangle) respectively.



(a) 95% confidence interval of RCI forecast by k -NN and ARMA-GARCH from 12-2005 to 11-2007.



(b) 95% confidence interval of RI forecast by k -NN and ARMA-GARCH from 10-2005 to 09-2007.



(c) 95% confidence interval of CI forecast by k -NN and ARMA-GARCH from 12-2005 to 11-2007.

Figure 3: RCI, RI and CI observed values (dashed) and forecasted values by k -NN (plus sign) and by ARMA-GARCH (star) with their 95% confidence intervals (circle) and (triangle) respectively.