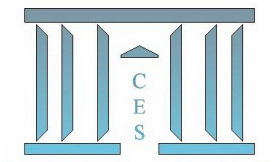
Documents de Travail du
Centre d'Economie de la Sorbonne

CES Working Papers




# A non-parametric method to nowcast the Euro Area IPI

Laurent FERRARA, Thomas RAFFINOT

**2008.33**

CENTRE NATIONAL
DE LA RECHERCHE
SCIENTIFIQUE

Maison des Sciences Économiques, 106-112 boulevard de L'Hôpital, 75647 Paris Cedex 13
http://ces.univ-paris1.fr/cesdp/CES-docs.htm
ISSN : 1955-611X

# A non-parametric method to nowcast the Euro Area IPI

**Laurent FERRARA* and Thomas RAFFINOT****

**April 2008**

**Abstract:**

Non-parametric methods have been empirically proved to be of great interest in the statistical literature in order to forecast stationary time series, but very few applications have been proposed in the econometrics literature. In this paper, our aim is to test whether non-parametric statistical procedures based on a Kernel method can improve classical linear models in order to nowcast the Euro area manufacturing industrial production index (IPI) by using business surveys released by the European Commission. Moreover, we consider the methodology based on bootstrap replications to estimate the confidence interval of the nowcasts.

**Key Words**: Non-parametric, Kernel, nowcasting, Bootstrap, Euro area IPI.

**JEL Classification**: C22, C51, E66

* Banque de France and CES, University Paris 1 – Panthéon - Sorbonne. Email: laurent.ferrara@banque-france.fr.
The views expressed herein are those of the author and do not necessarily reflect those of the Banque de France.

** CPR Asset Management. Email: thomas.raffinot@cpr-am.fr

## 1. Introduction

Accurate and timely information on the current state of economic activity is of great interest for market analysts and policy-makers. However, diffusion delays of the main economic indicators do not allow a true real-time assessment. For example, first estimates of the quarterly national accounts in the Euro area are released by Eurostat around 45 days after the end of the quarter. The same delay applies for the industrial production index (IPI hereafter), around 45 days after the end of the reference month. Thus, during this interval of time there is an uncertainty about the economic conditions. In this respect, several statistical procedures have been developed to provide a consistent picture of the current state of the economy by using more timely information conveyed by indicators (see, among others, Sédillot and Pain, 2005, Diron, 2006, or Angelini et al., 2008). This exercise is often referred to as *nowcasting*. Among those indicators, business and consumer surveys constitute very useful information set for analysts. Moreover, they are generally released the last working day of the reference month and they are not revised, or very slightly.

In this paper, we focus on the one-month-ahead forecasting (or nowcasting) of the aggregated Euro area manufacturing industrial production index (IPI, hereafter) by using the industrial opinion surveys, released monthly by the of the European Commission. Several studies have empirically proved that those surveys are strongly linked to the fluctuations of the IPI (see Bengoechea *et al.*, 2006 or Ferrara, 2005). Instead of using a parametric model to calibrate IPI with surveys, we propose here a non-parametric tool, which has never been considered for such exercise. Non-parametric methods have been empirically proved to be of great interest in order to forecast stationary time series (see for instance Matzner-Lober, Gannoun and De Gooijer, 1998, or Thomakos and Guerard, 2004), mainly because they are easy to implement and provide competitive results. Especially, when comparing with classical ARMA processes, non-parametric methods seem to give better results in the presence of non-linearities in time series

(see Cao and Vilar-Fernandez, 2005). However, very few applications have been proposed in the econometrics literature, see for instance Thomakos and Guerard (2004). We propose here a non-parametric bivariate procedure based on a Kernel method (see for instance Härdle and Vieu, 1992) in order to nowcast the euro area IPI and we compare the results with benchmark parametric models as regards both the root mean squared errors and the directional forecasts. Moreover, we provide bootstrapped confidence intervals for the nowcasts. In the next section, we briefly present the non-parametric forecasting method and section three contains the main results.

## 2. Non-Parametric forecasting

Let $(Y_1,\ldots,Y_T)$ denote an observed covariance-stationary time series stemming from a $p$-order Markovian process $(Y_t)_t$, with $p \geq 1$. Our aim is to forecast the value $Y_{T+k}$, $k \geq 1$ being the horizon, by a non-parametric predictor $\hat{Y}_T(k)$. Among the diverse non-parametric forecasting methods, we focus in this paper on the Kernel-type $k$-step-ahead predictor. We refer to Bosq (1998) for further details on non-parametric prediction. In a first step, we present the univariate predictor, then we describe how to integrate information brought by another variable into a bivariate approach. Last, we deal with the empirical issues linked to this forecasting method and we present the bootstrap resampling method to compute confidence intervals.

### 2.1. Univariate forecasting

In this subsection, we assume that the whole information set is contained in the past values of the series of interest. The non-parametric predictor is strongly related to the notion of smoothing and can be seen as a kind of weighted average of the past values where the weights are adequately computed.

First, we split the original series into two sub-series defined, for $t = p,\ldots,T-k$, by $(Y_{t+k})_t$, and $(V_t)_t = (Y_t, Y_{t-1} \ldots, Y_{t-p+1})_t$. The first sub-series contains the $k$-step-ahead value we aim to predict *in fine*

(for $t = T$) and the second, $(V_t)_t$, contains the information set related to the nature of the $p$-order process that will be useful to predict future values. In fact, due to $p$-order Markovian nature of the process, we assume that any value $Y_t$ can be explained by the $p$ preceding values $(Y_{t-1},\ldots,Y_{t-p})$. The $k$-step-ahead Kernel-type non-parametric predictor is thus given by the following formula:

$$\hat{Y}_T(k) = \sum_{t=p}^{T-k} w_t(V_T) Y_{t+k} \quad ,$$

where :

$$w_t(x) = \frac{K((x - V_t)/h)}{\sum_{t=p}^{T-k} K((x - V_t)/h)},$$

where $h$ is a positive sequence, referred to as the bandwidth, which controls the size of the local neighbourhood and where $K$ is a $p$-multivariate kernel function, defined as being continuous, positive, symmetric, integrable and integrates to one. Generally, the $p$-multivariate kernel function is the product of $p$ univariate kernel $(K_j)_{j=1,\ldots,p}$ defined on the interval ] $-\infty$, $+\infty$ [, such that:

$$K(x_1,\ldots,x_n) = \prod_{j=1}^{p} K_j(x_j)$$

As regards the choice of the kernel, we refer to subsection 2.3.

In fact, it can be shown that the kernel-type predictor minimizes the least-squares criteria and thus $\hat{Y}_T(k)$ is a non-parametric estimate of the conditional expectation of $Y_{T+k}$ knowing all the past, namely $E(Y_{T+k} / Y_T, Y_{T-1}, \ldots)$. Thus, this predictor is often referred to as the conditional mean predictor. Other non-parametric predictor have been similarly proposed to estimate the location parameter of the conditional distribution of $Y_{T+k}$, such as the conditional median and the conditional mode (see Matzner-Lober, Gannoun and De Gooijer, 1998, for a comparison of these approaches). As regards the asymptotic properties of this predictor, Collomb (1984) proved that $\left|\hat{Y}_T(k) - Y_{T+k}\right|$ tends to zero almost surely as T$\rightarrow \infty$.

As a special case of this predictor, we consider the one-step-ahead predictor ($k=1$) and we assume that the generating process is a first order Markovian process ($p=1$). Hence, the kernel-type predictor is defined as follows:

$$\hat{Y}_T(1) = \sum_{t=1}^{T-1} \frac{K((Y_T - Y_t)/h)}{\sum_{t=1}^{T-1} K((Y_T - Y_t)/h)} Y_{t+1} .$$

In this simple case, we clearly observe that the predictor $\hat{Y}_T(1)$ is a weighted average of past values where a higher weight is given to values $Y_{t+1}$ for which the preceding value $Y_t$ is close to the last observed value $Y_T$.

### 2.2. Bivariate forecasting

Let us turn to the case where another variable $(X_t)_t$ may contain useful information to predict more efficiently $Y_{T+k}$. Thus, we observe two trajectories $(Y_1,\ldots,Y_T)$ and $(X_1,\ldots,X_T)$ stemming from covariance-stationary Markovian processes of order, respectively, $p$ and $q$. Here again, we split the time series $(Y_t)_t$ into two sub-series, defined, for $t = max(p-1,q), \ldots, T-k$, by: $(Y_{t+k})$ and $(V_t)_t = (Y_t, Y_{t-1} \ldots, Y_{t-p+1}, X_t, X_{t-1} \ldots, X_{t-q})_t$. In this case, the information set is represented by the past values of the variable of interest $(Y_t)_t$ and by the coincident and lagged values of the explanatory variable $(X_t)_t$. The $k$-step-ahead Kernel-type non-parametric predictor is thus given by:

$$\hat{Y}_T(k) = \sum_{t=\max(p-1,q)}^{T-k} w_t(V_T) Y_{t+k} ,$$

where:

$$w_t(x) = \frac{K((x - V_t)/h)}{\sum_{t=\max(p-1,q)}^{T-k} K((x - V_t)/h)} ,$$

where $K$ is a $p+q+1$-multivariate kernel defined as the product of $p+q+1$ univariate kernels. As in the application presented in section 3, it may happen in practice that the explanatory variable $(X_t)_t$ is known at date $T+s$, where $s \geq 1$. In this case, the vector of information $(V_t)_t$ is replaced by the vector $(V'_t)_t$ defined, for $t = max(p-1,q), \ldots, T-k$, by:

$(V'_t)_t = (Y_t, Y_{t-1} \ldots, Y_{t-p+1}, X_{t+s}, X_{t+s-1} \ldots, X_{t+s-q})_t$.

For instance, consider the special case $p=1$ and $q=0$ and assume that the explanatory variable is known with an advance of $s=1$. The one-step-ahead predictor is thus given by:

$$\hat{Y}_T(1) = \sum_{t=1}^{T-1} \frac{K((Y_T - Y_t)/h)K((Y_T - X_{t+1})/h)}{\sum_{t=1}^{T-1} K((Y_T - Y_t)/h)K((Y_T - X_{t+1})/h)} Y_{t+1}$$

Obviously, this bivariate approach can be extended to the multivariate case by considering several explanatory variables $X_t^1, X_t^2, \ldots$

**2.3. Empirical issues**

When dealing with parametric forecasting, the two main empirical issues are the specification of the adequate parametric model, supposed to reproduce the stylized facts of the series, and the parameter estimation. In the non-parametric framework, the corresponding issue is the choice of the smoothing parameters involved in the forecast procedure, namely the bandwidth $h$, the kernel $K(.)$ and the order $p$.

Several kernels have been proposed in the statistical literature to be implemented in non-parametric forecasting. The most often used kernel is the Gaussian kernel defined as the density function of a standardized Gaussian distribution, i.e., for $u \in ]-\infty, +\infty[$:

$$K_G(u) = \frac{1}{\sqrt{2\pi}} \exp(-u^2/2)$$

Other used kernels are the Epanechnikov and the triangle respectively defined as follows, for $u \in ]-\infty, +\infty[$:

$$K_E(u) = \frac{3}{4}(1-u^2) I_{\{|u| \le 1\}},$$

$$K_T(u) = (1-|u|) I_{\{|u| \le 1\}},$$

where $I_{\{.\}}$ is the indicator function. Generally, the Gaussian kernel is preferred because an

optimal bandwidth *h* can be exhibited under this assumption.

The choice of the bandwidth *h* is crucial in the non-parametric framework, because it controls the smoothing degree of the non-parametric estimate. Thus, the value of *h* is proportional to the bias and inversely proportional to the variance of the estimate. Therefore, the classical statistical trade-off between bias and variance arises. According to the classical literature on non-parametric density estimation, a bandwidth which assures an optimal rate of convergence has been proposed by Deheuvels (1977), such as: $h = \hat{\sigma}_T \times T^{-1/(p+4)}$, where $\hat{\sigma}_T$ is the empirical standard error of the variable of interest and *p* the Markovian order. In this paper, we consider this latter bandwidth. It is noteworthy that data-driven approaches, like cross-validation (see Hardle and Vieu, 1992) are widely used in empirical studies to determine the optimal bandwidth in a, for example, forecasting sense.

As regards the choice of Markovian orders *p* and *q*, there is, to our knowledge, no optimal theoretical approach. In this respect we advocate a data-driven approach by choosing the best orders that minimize a forecasting criterion (for example the root mean squared error, RMSE) over a learning set.

**2.4. Confidence intervals**

We consider now the construction of a confidence interval around the predictor $\hat{Y}_T(1)$. In opposition to parametric forecasting, no assumption is done for the residuals distribution. Therefore, we focus on the re-sampling smoothed bootstrap technique that provides confidence intervals when the underlying distribution is unknown. Let consider the learning set $(Y_t\ X_t)_t$ from $t=1$ to $t=t_0$ with $t_0 \leq t_1 \leq t_2 \leq \ldots \leq T$. For $t=1,\ldots,T$ we get $n = T - t_1 + 1$ residuals $\varepsilon_t$, $t=t_1,\ldots,T$, by :

$$\varepsilon_t = Y_t - \hat{Y}_{t-1}(1).$$

As regards the confidence interval for $\hat{Y}_T(1)$, we compute *B* replications of the smoothed bootstrap residuals (see for instance Cao and Vilar-Fernandez, 2005) defined by, for $i=1,\ldots,B$:

$$\varepsilon_i^* = \varepsilon_{I_i} + g\xi_i ,$$

where $I_i$ is a discrete random variable with uniform distribution in *{1,…,n}*, $\xi_i$ is a standard Normal random variable and $g$ is the smoothing parameter such as $g = (4/3n)^{1/5}s$ , with $s$ being the sample standard deviation. The smoothed bootstrap methodology seems to be more adequate than the classical bootstrap in the sense that it reduces the effects of outliers in the empirical residuals distribution.

Once the smoothed bootstrap residuals are computed, they are sorted from $\varepsilon_{(1)}^*$ to $\varepsilon_{(B)}^*$; and the quantiles of order $[(\alpha/2)B]$ and $[(1-\alpha/2)B]$ are selected for a given confidence level $1-\alpha$. Thus, the confidence prediction interval for the one-step-ahead predictor, at the level $1-\alpha$, is given by:

$$[\hat{Y}_T(1) + \varepsilon^*_{([(\alpha/2)B])}, \hat{Y}_T(1) + \varepsilon^*_{([(1-\alpha/2)B])}].$$

## 3. Application

In this section, we apply our methodology to nowcast the manufacturing IPI for the Euro area (seasonally adjusted and corrected from working days[1]) one month in advance by using the industrial opinion surveys. Besides, surveys at the euro area aggregated level published by the EC are stemming from national statistical institutes and are harmonised in Brussels (more information can be obtained from the DG-EcFin web site[2]). The monthly EC industrial survey collects the following 7 balances of opinions (covering about 22 000 enterprises in the euro area):

1) Production trend observed in recent months
2) Assessment of order-book levels
3) Assessment of export order-book levels
4) Assessment of stocks of finished products

[1] We use the official IPI data set released by Eurostat the 17th of March 2006
[2] http://europa.eu.int/comm/economy-finance/index-en.htm

5) Production expectations for the months ahead

6) Selling price expectations for the months ahead

7) Employment expectations for the months ahead

To summarise the information contained in those balances of opinions at the euro aggregated level, the EC also computes two composite indices. First, the Industrial Confidence Indicator (ICI) is computed by simply averaging items 2, 4 and 5. Second, the Business Climate Indicator (BCI) is computed as the first common factor of a static Principal Component Analysis applied to items 1, 2, 3, 4 and 5.

As the manufacturing IPI series appears to be non-stationary (the mean is clearly non-constant and the usual unit root tests tend to reject the stationarity hypothesis), we consider the annual growth rate of the series. It turns out that the annual IPI growth rate is the most strongly correlated with the items of the industrial survey. Thus, we work on the series:

$$Y_t = 100 \times \frac{IPI_t - IPI_{t-12}}{IPI_{t-12}}$$

We focus on the years 2002-2005 as period of interest. Those years are interesting for economists because the euro area economic activity experienced a lot of uncertainty during this period, just after the industrial recession of 2001. 2002 was a year of recovery, but, 2003 experienced a strong decay during the first 6 months, followed by a recovery starting in July. Then, the annual growth rate went down from 4.1 percent in May 2004 to -2.0 percent in June 2005. Last, since July 2005, the growth rate has been increasing. We compare four diverse approaches to predict the annual IPI growth rate:

- Naive: the naive forecast for a period *t* is simply the observed value at *t-1*.

- ARMA: we apply an univariate ARMA model

- Non-parametric univariate forecasting: we apply the method presented in 2.1.

- Non-parametric bivariate forecasting: we apply the method presented in 2.2, with various exogenous variables, namely the seven items plus BCI and ICI.

As regards the choice of the Markovian orders $p$ and $q$ the ability to make useful *ex ante* forecasts is the real test. Therefore, we perform a selection of parameters $p$ and $q$ via one-step-ahead predictions over the years 2002-2005. For each related time series, we compute dynamic one-step-ahead predictions with a range of parameters $p$ and $q$ from 1 to 4. It means that the nowcasts are computed recursively each month by integrating the new data and by re-estimating dynamically the models with fixed parameters $p$ and $q$. Thus, for each related time series, sixteen various models are tested. The performances of nowcasts are evaluated through RMSE and MAE criteria, defined by:

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(\hat{Y}_{T+i-1}(1) - Y_{T+i})},$$

and

$$MAE = \frac{1}{m}\sum_{i=1}^{m}\left|\hat{Y}_{T+i-1}(1) - Y_{T+i}\right|,$$

where $m$ is the number of observations retained to evaluate the one-step ahead forecasts $\hat{Y}_{T+i-1}(1)$ generated from models fitted to the first $T+i-1$ observations of $(Y_t)_t$. In our application, $m$ equals to 48.

Prediction criteria for the IPI annual growth rate for years 2002-2005 are presented in the Table 1. According to those prediction criteria, it turns out that surveys improve the nowcasting of the IPI, especially the Industrial Confidence Indicator (see Figure 1). The modified Diebold-Mariano test (1995) points out that the inclusion of the ICI into a bivariate non-parametric approach provides significatively better results than the three other methods (at the classical level of 5%). Note also that the univariate non-parametric forecast outperforms the naive forecast and also the forecasts based on some items, less correlated with the IPI annual growth rate (items 3, 4 and 5).

From Table 2, we observe that 2004 is the most difficult year to nowcast. It seems that the turning-point in the mid-2004, was hard to catch for industrials. Conversely, the 2002 recovery

was easier to predict. This point underlies an asymmetric behaviour of the cycle.

Unfortunately, it is well known from experts that the IPI annual growth rate is too volatile to be accurately forecasted on a monthly basis. Nevertheless, it may be interesting for economists, to evaluate whether the time series is going to accelerate or to decelerate, in order to asses the current phase of the acceleration cycle. Therefore, the ability to guess the sign of the future evolution is also a widely used criterion. For this purpose, we consider the so-called success ratio (SR) defined by:

$$SR = \frac{1}{m}\sum_{j=1}^{m} I_j\left\{(\hat{Y}_{T+j-1}(1) - Y_{T+j-1})(Y_{T+j} - Y_{T+j-1}) > 0\right\},$$

where $I_j\{.\}$ is the indicator function. Note that the SR criterion is simply the fraction of times the sign of $(Y_{T+j} - Y_{T+j-1})$ is predicted correctly. According to the results presented in table 3, it turns out that using ICI in a bivariate non-parametric approach is more efficient in sign prediction. Moreover, Pesaran and Timmermann (1992) developed a sign predictability test. Using the ICI, we obtain a test statistic of 9.2, which is well above the 95% critical value of a standard normal variable. There is thus a clear evidence of predictive power in ICI for the IPI movements.

Last, we compute prediction intervals as presented in section 2.4. We choose 1995-2002 as learning period and compute $B=1000$ replications of the smoothed bootstrap residuals. To get narrower confidence intervals, we opt for a confidence level $1 - \alpha$ equal to 70%. The dynamic one-step-ahead confidence intervals are shown in figure 2.

## Conclusion

In this paper, we introduce a non-parametric Kernel-based method in order to predict the one-month-ahead manufacturing Euro area IPI using industrial opinion surveys. Moreover, we compute confidence intervals for the predictor based on the re-sampling Bootstrap

methodology. We show that this non-parametric method allows to improve the forecasting performances with respect to more classical techniques. This approach could be fruitfully used to asses the current state of the industrial activity during the 45 days between releases of IPI and opinion surveys.

## References


Angelini, E, G. Camba-Mendez, D. Giannone, L. Reichlin and G. Rünstler (2008), Short-term forecasts of Euro area GDP growth, *Discussion Paper Series No. 6746, CEPR.*

Bengoechea, P., M. Camacho and G. Perez-Quiros (2006), A useful tool for forecasting the Euro area business cycle phases, *International Journal of Forecasting*, 22, 735-749.

Cao, R. and Vilar-Fernández, J.M. (2007), Nonparametric forecasting in time series. A comparative study, *Communications in Statistics / Simulation and Computation*, 36, 3043-3057.

Collomb, G. (1984), Propriétés de convergence presque complète du prédicteur à noyau, *Zeitschrift für Wahrseinlichkeitstheorie*, 66, 441-460.

Diron, M. (2006), Short-term forecasts of euro area real GDP growth: An assessment of real time performance based on vintage data, *Working Paper No. 622, European Central Bank.*

Diebold, F.X. and R.S. Mariano (1995), Comparing predictive accuracy, *Journal of Business and Economic Statistics*, 13, 253-263.

Ferrara, L. (2007), Point and interval nowcasts of the euro area IPI, *Applied Economics Letters*, 14, 2, 115-120.

Härdle, W., and Vieu, P. (1992) Kernel regression smoothing of time series, *Journal of the Time Series Analysis*, 13, 209-232.

Harvey, D.I., Leybourne, S.J. and P. Newbold (1997), Testing the equality of prediction mean squared errors, *International Journal of Forecasting*, 13, 281-291.

Matzner-Lober, E., Gannoun, A. and De Gooijer, J.G. (1998), Nonparametric forecasting: Comparison of three kernel-based methods, *Theory and Methods*, 27 , 1593-1617.

Pesaran, H. and Timmermann, A. (1992), A Simple Nonparametric Test of Predictive Performance, *Journal of Business & Economic Statistics*, 10, 4, 561-65.

Sédillot, F., and N. Pain (2005), Indicator models of real GDP growth in the major OECD countries, *OECD Economic Studies*, No. 40, 2005/1, 167-217.

Thomakos, D. and Guerard, J. (2004), Naive, ARIMA, nonparametric, transfer function and VAR models: A comparison of forecasting performance, *International Journal of Forecasting*, vol. 20(1), 53-67.

Vieu, P. (1994). Order choice in nonlinear autoregressive models, *Statistics*, 24, 1-22.

**Table 1. Prediction criteria for the IPI annual growth rate (period 2002-2005)**

| | (p,q) | MAE | RMSE |
|---|---|---|---|
| **item 1** | (2,1) | 0.91 | 1.13 |
| **item 2** | (3,1) | 1.00 | 1.23 |
| **item 3** | (2,1) | 1.17 | 1.46 |
| **item 4** | (2,2) | 1.09 | 1.32 |
| **item 5** | (3,1) | 1.02 | 1.20 |
| **item 6** | (3,2) | 1.02 | 1.24 |
| **item 7** | (2,1) | 1.06 | 1.24 |
| **BCI** | (2,4) | 0.97 | 1.17 |
| **ICI** | (3,2) | 0.88 | 1.09 |
| **Naive** | | 1.14 | 1.38 |
| **Univariate NP** | 3 | 1.03 | 1.23 |
| **ARMA** | (3,0) | 1.06 | 1.25 |

**Table 2. Prediction criteria for the IPI annual growth rate (period 2002-2005)**

| | 2002 | | 2003 | | 2004 | | 2005 | |
|---|---|---|---|---|---|---|---|---|
| | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE |
| **ICI** | 0.67 | 0.90 | 0.87 | 1.08 | 1.10 | 1.30 | 0.90 | 1.08 |
| **Naive** | 0.95 | 1.15 | 1.12 | 1.38 | 1.25 | 1.49 | 1.26 | 1.50 |
| **Univariate NP** | 0.94 | 1.17 | 0.99 | 1.14 | 1.30 | 1.47 | 0.92 | 1.15 |

**Table 3. Prediction criteria: Success Ratio (period 2002-2005)**

| | Success Ratio |
|---|---|
| **ICI** | 0.78 |
| **Univariate NP** | 0.68 |
| **ARMA** | 0.71 |

**Figure 1. IPI annual growth rate (in %) and the non-parametric kernel predictor wit h ICI (2002-2005)**

**Figure 2. Bootstrapped confidence interval at the 0.70 confidence level (2002-2005)**