ON THE USE OF THE FIRST PRINCIPAL COMPONENT AS A CORE INFLATION INDICATOR

José Ramos Maria

WP 3-04 January 2004

The analyses, opinions and findings of these papers represent the views of the authors, they are not necessarily those of the Banco de Portugal.

Please address correspondence to José Ramos Maria, Economic Research Department, Banco de Portugal, Av. Almirante Reis nº. 71, 1150-165 Lisboa, Portugal; Tel: 351 213130879; Fax: 351 213107804; email: jfrmaria@bportugal.pt
On the use of the first principal component as a \textit{core} inflation indicator

José Ramos Maria\(^*\)

Banco de Portugal

January 2004

Abstract

This paper investigates if the (OLS-scaled) first principal component ($PC_1$), extracted from standardized yearly rates of change of basic items of the CPI, represents a reasonable option for a \textit{core} inflation indicator. The evaluation is carried out by (i) confronting alternative linear transformations of the original variables; (ii) analyzing the impact of stacking lagged variables to the original database; and (iii) exploring the contents of the remaining principal components. An orthogonal factor model framework will also be introduced so as to fully reproduce any variable that can be expressed as a linear combination of the original input variables, such as, in this case, the overall inflation rate. The model incorporates the following properties: (i) the results are not conditional on the eigenvectors length; (ii) the variance of the CPI accounted for by each component is unique, and (iii) the outcome is equivalent to an OLS regression between the CPI and the $PC_1$. Along with empirical evidence for the Portuguese case, it will be claimed that the above-mentioned (OLS-scaled) $PC_1$ does capture the general movement of the overall inflation rate, however, no OLS regression would have to be implemented if the core indicator is fully aligned with the orthogonal factor model.

\textbf{KEYWORDS:} Principal components, factor models, core inflation indicators.

\textbf{JEL Classification:} C43, E31.

\textbf{Acknowledgements:} I am indebted to Francisco Dias and to Maximiano Pinheiro for helpful comments and suggestions. The assistance of Teresa Nascimento was also appreciated. The usual disclaimer applies.

\(^*\)E-mail: jfrmaria@bportugal.pt
1 Introduction

Principal components analysis continues on being one of the most used techniques in multivariate analysis. Within the price development dimension, several authors have used this technique to estimate “summary indicators” and to classify those time series as trend inflation indicators. Recent work for the euro area includes, for instance, Angelini, Henry and Mestre (2001b), who use a time domain approach to estimate a non-stationary factor that intends to represent a common trend in the inflation measures (along the lines suggested by Stock and Watson (1998); the forecasting performance was address in Angelini, Henry and Mestre (2001a)); and Cristadoro, Forni, Reichlin and Veronese (2001), who use a frequency domain approach to extract a medium and long-run common component for consumer prices. Recent work with Portuguese data includes, for instance, Machado, Marques, Neves and Silva (2001).

This paper investigates if the (OLS-scaled) first principal component ($PC_1$), extracted in the time domain from standardized yearly rates of change of basic items of the consumer price index ($CPI$), represents a reasonable option for a core inflation indicator. Currently, the Banco de Portugal uses this procedure to measure (and regularly publish) a core inflation indicator for the Portuguese case. An objective of this paper is to evaluate this option, which was proposed by Coimbra and Neves (1997), while looking for possible improvements. Machado et al. (2001) suggested, for instance, that to take into account the non-stationarity feature of the input variables, a specific linear transformation could be implemented, instead of making use of standardized variables, as in Coimbra and Neves. In both studies, the core indicator is only derived from the information contained in the basic items of the Portuguese CPI, which are mostly non-stationary data. Not surprisingly, the need of having to compare the core measure with some observable variable raised the question of having to find an “appropriate scaling” for the $PC_1$. In what has been classified as just an “ad-hoc” procedure, both studies solved this issue by running an OLS regression between the inflation rate and the $PC_1$ (and, therefore, depend on this final parameter to scale the $PC_1$). It will be shown that this decision has some important implications. The current paper confronts the results obtained by these studies and, furthermore, investigates if alternative linear transformations of the observed variables, within the same course of action, produce structurally different outcomes. It will be claimed
that none of the transformations have an obvious superiority or advantage; nor can
the non-stationarity nature of the data be used to distinguished the overall results.
Other possibilities of improvement, such as (i) stacking lagged variables to the original
database, along the lines suggested for instance by Stock and Watson (1998) or (ii)
including more principal components for the derivation of the core inflation indicator
are also analyzed.

The rest of the paper is organized as follows. Section 2 confronts the results obtained
from alternative linear transformations of the original data. Section 3 analyzes the
use of lagged variables in the database. The use of more than the first principal
component is considered in section 4. Finally, section 5 introduces a simple theoretical
orthogonal factor model that captures the general underlying conceptual framework
of the proposed indicator. The factor model fully reproduces the overall CPI, and
embodies several appealing properties, in which the so called “ad-hoc” procedure has
a well defined interpretation. Without any assumption on the data generation process
of the different input variables, it will be shown that the model represents a solution
for the fundamental eigenvectors length indeterminacy, it uniquely determines the
variance of the CPI accounted for by each $PC_i$ and it is fully equivalent to an OLS
regression, using the $CPI$ as the endogenous variable and the $PC_i$ as the exogenous
variables.

Section 6 concludes. It will be claimed that unless a smoother core inflation indicator
is being envisaged on a priori grounds, no obvious gain seems to be achieved by
changing the currently used procedure at the Banco de Portugal. However, no OLS
regression would have to be implemented if the core inflation indicator is fully aligned
with the factor model introduced in section 5.

2 On the use of transformed variables

Let the matrix $X$ be the initial information set, with $N$ observable variables $x_1, x_2,$
..., $x_N$ (in columns), and $T$ observations (in lines). The principal components are just
special linear combinations of those initial variables and if none of the $N$ variables is
an exact linear combination of the others, then there will exist as many distinct $PC_i$
as variables.

The principal components methodology can be applied on any second moment matrix
of the initial information set and therefore a decision has always to be made on whether the database justifies certain transformations prior to the calculation of the PC.\textsuperscript{1} Given that the second moment matrix may be non-centered, the first common transformation of the original data is to subtract the mean of each \( x_i \). If one does not wish to have a \( PC_1 \) dominated by those variables who have the largest variance, then a second common transformation is standardization and instead of deriving the \( PC_1 \) from the variance-covariance structure of \( X \) (from now on denoted as \( \Sigma_X \)), the \( PC_1 \) can be derived from the correlation matrix (from now on denoted as \( \rho_X \)). Using year-on-year rates of change of basic items of the CPI, which are basically non-stationary data, this was the procedure followed by Coimbra and Neves (1997). Let this linear transformation be denoted as LT1.

The general conceptual framework of Coimbra and Neves (1997) can be written down as

\[
\begin{align*}
\quad x_1 &= \overline{x}_1 + l_{11}^* F_1^* + \varepsilon_1 \\
\cdots &\quad \cdots \\
\quad x_N &= \overline{x}_N + l_{N1}^* F_1^* + \varepsilon_N
\end{align*}
\]

where \( \overline{x}_1, \overline{x}_2, \ldots, \overline{x}_N \) correspond to the average values of \( x_{1t}, x_{2t}, \ldots, x_{Nt} \); \( F_1^* \) represents a common factor to all variables and \( F_1^* = PC_1 \); \( l_{ij}^* = a_{j1} \), \( j = 1 \ldots N \), where \( a_{j1} \) represents the scalars defining the eigenvector (scaled to unity) associated to the largest eigenvalue (extracted from \( \rho_X \)); and, finally, \( \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N \) correspond to specific factors of each variable. Note that the \( N \) variables \( x_1, \ldots, x_N \) are just being linearly expressed in terms of their mean plus \( (1 + N) \) unobservable variables: \( F_1^* \) and \( \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N \).\textsuperscript{2} Presumably, the behaviour of each year-on-year rate of change is not only reflecting specific factors but also the “inflation trend”.

It is well known that if the original variables are subject to LT1, the eigenvalues and resulting eigenvectors are in general not the same as the ones derived from \( \Sigma_X \), or even a simple function of them.\textsuperscript{3} If the eigenvectors are collected in \( A = [ a_1 \ a_2 \ldots \ a_N ] \), this implies that, in general, the matrix \( A \) will depend on the transformation.

Other transformations of the original (centered) variables are also available. For

\textsuperscript{1}See Jackson (1991) and Jolliffe (2002).

\textsuperscript{2}This model is usually expressed in terms of deviations, i.e. \( (x_i - \overline{x}) \), and to satisfy the usual assumption of unit variances of the common factors, the \( F_i \) should have been defined as \( F_i = PC_i/\sqrt{\text{Var}(PC_i)} \) and, therefore, \( F_i = PC_i/\sqrt{\lambda_i} \Rightarrow PC_i = \sqrt{\lambda_i} F_i \), with \( l_{ij}^* = \sqrt{\lambda_i} a_{ji} \), instead of assuming \( PC_i = F_i^* \). See, for instance, Afifi (1984).

\textsuperscript{3}See, in particular, Krzanowski (1996), pp 65-66.
instance, Machado et al. (2001) suggested that to take into account the non-stationary
feature of the original variables, the $x_i$ should be scaled by $(\sigma_{\Delta x_i})^{-1}$, where \(\sigma_{\Delta x_i}\)
represents the standard error of the first difference of $x_i$. By defining the smoothness
of the integrated variable as the variance of the first differences, Machado et al. (2001)
argue that a core inflation indicator should take into account the degree of smoothness
of the $PC_1$ by looking at linear combinations of the year-on-year rates of change of
the basic CPI items “with a large signal (variance) and not to much volatility” (p.7).
Let this linear transformation be denoted as LT2. Note that both possibilities can
be represented by a diagonal matrix $H$, with $h_{ii}$ in the main diagonal, $i = 1, ..., N$,
where $H$ is alternatively associated with LT1 or LT2, i.e.,

$$PC = (X - \bar{X})H^{-1}A$$

(2)

The principal components (extracted from unit length eigenvectors) are being gath-
ered in a matrix named $PC$, where the first principal component ($PC_1$) is placed
in the first column of $PC$, the $PC_2$ in the second, etc; and $\bar{X}$ corresponds to the
average values of $x_{1t}$, $x_{2t}$, ..., $x_{Nt}$. If $h_{ii} = \sigma_i$ in (2), $H$ is associated with LT1. If
$h_{ii} = \sigma_{\Delta x_i}$, and $\sigma_{\Delta x_i}$ is the standard error of the first difference of $x_i$, then $H$
is associated with LT2. This section will directly confront the results obtained under
this two transformations. However, other transformations (purely arbitrary and
motivated by no reason, except for comparison purposes against LT1 or LT2) will
also be implemented. Alternative linear transformations of the same type of LT1 or
LT2 may be given, for instance, by $h_{ii} = \max(x_i) - \min(x_i)$, abbreviated to LT3;
or simply $h_{ii} = \max(x_i)$, abbreviated to LT4. Note that for each possibility, the
importance of each original (centered) variable $(x_i - \bar{x})$ is just being changed by a
scaling constant $h_{ii}^{-1}$, in particular when computing the second moment matrix from
which the scalars of the eigenvectors are extracted. In the case of LT1, for instance,
the importance of the variables with high standard deviations will be highly affected
and the same follows for a higher $\sigma_{\Delta x_i}$, $[\max(x_i) - \min(x_i)]$ or $\max(x_i)$ in the case of
LT2, LT3 or LT4, respectively.

After having decided the relevant second moment matrix from which the matrix $A$
is derived, there is still the need to find comparable scores to those of the observed

\[\text{Eq. 2}\]

\[\text{Eq. 2}\]

\[\text{Eq. 2}\]

\[\text{Eq. 2}\]
overall inflation rate. To such purpose, Coimbra and Neves (1997) or Machado et al. (2001) computed the OLS fitted values of the following regression

$$CPI = \beta_0 + \beta_1 PC_1 + \varepsilon$$

(3)

where $CPI$ denotes the observed inflation rate.

The database used by Coimbra and Neves (1997) and by Machado et al. (2001) has 90 variables ($N = 90$), with monthly year-on-year rates of change of basic items of the $CPI$. Following the EUROSTAT classification, 14 of those variables refer to unprocessed food items, 24 are processed food items, 3 are energy items, 26 are non-energy industrial goods items and 23 are services items.\(^6\) The current paper makes use of 141 observations, covering the period 1992:01-2003:09, and the same procedures will be extended to the other loosely defined transformations (LT3 and LT4). Therefore, the principal components methodology will be just used as a mechanical device to decompose, through alternative $N \times N$ matrices obtained from different transformations, the original matrix of observed variables.\(^7\)

Using the full sample period, the left panel of figure (1) contains all OLS-scaled $PC_1$ from LT1, LT2, LT3 and LT4. As it can be seen, all transformations seem to capture the general driving behaviour of the observed inflation rate, and thus, these final results cannot be used to distinguish between the transformations. The results are obviously not identical, nevertheless, no outcome is systematically above or below the

\(^6\)A complete list may be found in Annex 1.

\(^7\)An overview of matrix eigenstructures may be seen in Carroll and Green (1997).
remaining ones. The differences, in percentage points, against the results obtained under LT1 were highlighted in the right panel of the same figure; during most of the time, the differences evolved within a small range (the interval $[-0.15; +0.15]$ is highlighted); over the sample period, the average is virtually nil for all cases. In the case of LT2, the major differences against LT1 were registered in the beginning of the sample period. In mid-1998 and in the last part of the sample period, the differences have also exceeded the lower limit of the reported interval.

In the face of such similar results, it was then investigated if non-stationarity effects could be used so as to distinguish between the linear transformations. The use of centered data has no effect on $\Sigma_X$ or $\rho_X$, however, this may raise an issue with non-stationarity data, given that (3) was derived after having unfolded the second moment matrices and their associated eigenvector structure. For instance, with one additional observation, the mean of the $CPI$ may change from $CPI^{(T)}$ to $CPI^{(T+1)}$.

To evaluate the impact of additional observations on all scalars that produce the OLS-scaled $PC_1$, either from LT1, LT2, LT3 or LT4, the following recursive procedure was implemented: the sample period was (arbitrarily) shortened in 90 observations, to 1992:01-1996:3, and a first set of estimates was computed; one observation was then added and a second set of estimates was computed for the sample period 1992:01-1996:4, and so on, until $T = 141$. These ninety one sets of estimates give rise to ninety one core inflation indicators for each transformation. When the full sample period is used, the last OLS-scaled $PC_1$ are equal to the ones presented above, which was already seen not to be very different. Those estimates may be analyzed so as to answer several questions: are there disruptive effects, given that most of the input variables are non-stationarity?; are the results dramatically different in nature, given that the transformations are themselves very different from each other?; given that the introduction of LT3 and LT4 were not motivated by any specific reason and may be seen as rather loosely defined, can the final outcome be used to distinguish among the linear transformations?

After having compiled these sets of core inflation indicators, two general procedures were then followed. Given that the level recorded for each month may obviously change within each sample period, a first general approach was just to evaluate if along with another additional observation, the final OLS-scaled $PC_1$ emerges as
fundamentally unstable against the previous scaled result.

A second general approach was to evaluate if, along with another additional observation, the inner products under which the final OLS-scaled $PC_1$ are being computed incorporates some abnormal or unacceptable behaviour in time; for this purpose, the partial derivatives of $PC_1$ with respect to $x_i$, $i = 1, 2, ..., 90$ were also disclosed and analyzed.

Within the first general approach, the maximum and minimum values recorded in each month were firstly retrieved and plotted in figure (2). This was done for all transformations. For each month, the maximum and minimum values define the upper and lower limit of the grey region. The differential between those limits acts as an indication of how much the core inflation level has changed (for each month) within the 91 sets of final OLS-scaled $PC_1$. The core inflation indicator, using the full sample period, and the evolution in time of $\beta_0$ of equation (3), during the recursive computation, were also depicted.

The following table contains the average (Av), the standard deviation (StDev) the
maximum (Max) and the sum (Sum) of the differential between the upper and lower core inflation levels of each month (in basis points). As it can be seen, the differences cannot be considered very substantial and once again cannot be used to clearly distinguish between the transformations. The exception is perhaps LT4. Using the full set of recursive estimates, the reported statistics have higher levels vis-à-vis the other transformations.

<table>
<thead>
<tr>
<th></th>
<th>Av</th>
<th>Stdev</th>
<th>Max</th>
<th>Sum</th>
<th></th>
<th>Av</th>
<th>Stdev</th>
<th>Max</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>LT1</td>
<td>21</td>
<td>12</td>
<td>47</td>
<td>2928</td>
<td>LT3</td>
<td>22</td>
<td>13</td>
<td>53</td>
<td>3149</td>
</tr>
<tr>
<td>LT2</td>
<td>21</td>
<td>13</td>
<td>47</td>
<td>2979</td>
<td>LT4</td>
<td>34</td>
<td>25</td>
<td>104</td>
<td>4801</td>
</tr>
</tbody>
</table>

It should be clear that one of the reasons behind such close results derives from the fact that all core inflation indicators have the same mean. If the overall mean is non-stationary and changes from $\overline{CPI(T)}$ to $\overline{CPI(T+1)}$, it will always be $\overline{CPI(T+1)}$ that will be the relevant mean underlying the systematic use of equation (3), and not $\overline{CPI(T)}$, implying that the mean of the core inflation indicator will not diverge, by construction, and over any sample period, from the mean of the observed inflation rate. Given that $\beta_0$ is added in the end of the process, as additional observations are being disclosed, the use of (3) simply adapts the fact that in the Portuguese case the mean of the inflation rate has decreased over time (as registered by $\beta_0$ in the graph). With no differences in the mean, the only possible way to differentiate between the linear transformations has to be based on the inner products under which the OLS-scaled $PC_1$ are being derived. The inner products depend upon three types of scalars: (i) the scaling constant $\beta_1$, (ii) the scalars defining the first eigenvector and (iii) the scalars defining the transformation of the original variables. With one additional observation, any one of these scalars may change. Let $[PC_1]_{x_i}$ be the partial derivative of the $PC_1$ with respect to the original variable $x_i$. From (2) and (3),

$$[PC_1]_{x_i} = \beta_1 \ a_{i1} \ h_{ii}^{-1}, \ i = 1, 2, ..., N.$$  

The scaling constants $\beta_1$ of LT1, LT2, LT3 and LT4 were also computed and plotted in the left panel of figure (3). For comparison purposes, the eigenvectors were scaled to the inverse of their roots. As it can be seen, $\beta_1$ has not remained unchanged over time. However, although the transformations are different in nature, the results are
The inner products of \([PC_1]'x_i\) were further investigated and the following course of action was implemented. In a first step, the results obtained during the computation of the ninety one sets of core inflation indicators were separated into (i) \(N\) scalars defining the first eigenvector (the \(a_1\)) and (ii) \(N\) multiplying scalars of the original variables associated to each transformation (the \(h_{i1}^{-1}\)). In a second step, for comparison purposes, the \(a_1\) and the \(h_{i1}^{-1}\) obtained in (i) and (ii) were scaled under the restriction that their sum would equal one, respectively, i.e. \(\sum_{i=1}^{N} a_{i1} = \sum_{i=1}^{N} h_{i1}^{-1} = 1\).

As a third step, the disaggregated results obtained in the second step were aggregated (added) according to the Eurostat classification of unprocessed food, processed food, energy, non-energy industrial goods and services prices. Without any scaling, the same aggregation was implemented for the \(N\) partial derivatives \([PC_1]'x_i\) (i.e., the
results are just equal to the sum of each \( \beta_1 a_{i1} h_{ii}^{-1} \) belonging to the same Eurostat aggregate).

The behaviour over time of the Eurostat aggregates stemming from LT1, LT2, LT3 and LT4 were plotted in the next 5 rows of figure (4). In each figure, the graphs on the left correspond to the results for the first eigenvector (the scaled \( a_{i1} \)); the graphs on the center correspond to the results for the scalars with an effect on the importance of the original variables (the scaled \( h_{ii}^{-1} \)); the graphs on the right correspond to the results for the final partial derivatives affecting the original variables (the \( \beta_1 a_{i1} h_{ii}^{-1} \)).

Several conclusions may be drawn from those figures. First, energy prices are treated in a rather indistinguishable way in all cases. Second, all transformations share the common internal features, in the end of the process, of (i) attaching a lower importance to those original variables belonging to the same aggregates (Energy and Unprocessed Food), and (ii) favour those of non-energy industrial goods. Third, LT4 does effectively seem to incorporate a higher variability then the remain transformations, but specially when the \( PC_1 \) is derived with fewer observations. In the case of LT2, it should be mentioned that the partial derivatives associated with processed food prices sum up to a lower level, as compared to the remaining transformations; and services items have received a growing importance over time. Using the full sample period, the differences against LT1 are almost non-existent in the case of unprocessed food, energy and services prices. Finally, the non-stationarity feature of the original data is not having any noticeable undesirable and distinctive effect on the final partial derivatives of any of the transformations (including under LT1\(^8\)). Nor can it be used to clearly distinguish the overall results.

The empirical evidence therefore suggests that, under all the transformations considered, the importance of each original variable is being relatively changed in such a way that the final results do empirically emerge as similar and hard to distinguish (which is particularly striking, given that no justification was given to implement LT3 or LT4). Although each transformation may qualitatively change the eigenvalues/eigenvectors solution of the original maximization problem, the use of structurally different linear transformations do not produce, clearly, dissimilar final results. In the case of LT1,

---

\(^8\)It may also be the case that the non-stationarity of the original data is basically creating a correlation matrix with a large number of positive elements, as the variables are somehow moving in the same directions, and to some extent the evolution of the first eigenvector over the different sample periods may only be reflecting what is usually referred to as a "size effect". See Chatfield and Collins (1996) and Jollife (2002). During the different sample periods, the number of positive elements of the second moment matrix stood always between 75 and 80 per cent.
Figure 4 - Aggregation in accordance with Eurostat classification
the sum of the partial derivatives belonging to the each aggregate has remained rela-
tively stable in time and no obvious gain is apparently achieved by changing the core
inflation indicator, as proposed in Coimbra and Neves (1997), for any other of the
remaining possibilities. On the contrary, under LT2, LT3 or LT4, the quantity under
maximization ceases to be a standard and well perceived statistic, i.e, without a clear
advantage, the input variables cease to be treated as “equally important” variables,
as in the case of LT1. It may be argued that standardization has only an immediate
statistical interpretation when the input variables are stationary.\(^9\) However, all four
transformations herein applied do not change the non-stationarity feature of the orig-
inal database. It is only a matter of different scaling of the original variables and of
different second moment matrices. Whereas under LT2, LT3 or LT4, these matrices
are somehow more difficult to interpret, under LT1 the second moment matrix is the
rather straightforward correlation matrix. By conveying a well known information,
the overall process becomes apparently more easily perceptible and intuitive.

3 On the use of non-contemporaneous variables

The common factors can be derived not only from contemporaneous but also from
non-contemporaneous values of \(X\). So far, the \(PC_i\) has only been derived from
contemporaneous data and therefore a possible improvement of this core measure
could be to allow for some dynamics, along the lines suggested for instance by Stock

Following the same recursive procedure introduced in the previous section, the core
inflation indicator can now be derived from equation (3), but after having stacked
non-contemporaneous figures to the original database. Figure (5) reports the results
for all OLS-scaled \(PC_1\), conditional on LT1, where the grey region is limited, as in
the previous section, by the maximum and minimum values obtained for each month.
The left panel of figure (5) considers contemporaneous variables and variables lagged
by one period, the right panel also considers variables lagged by two periods.\(^{10}\)

As before, these results do not seem highly different from the ones obtained by only
using contemporaneous variables in the database, except that the inflation indicator

\(^9\)Alexander (2001) goes even further and declares that the data input of Principal Component Analysis “must be
stationary” (p. 145). On this issue, see also Machado et al. (2001).

\(^{10}\)With contemporaneous variables and variables lagged by one and two periods, the number of input data is now
given by a matrix \([(T - 2) \times 3N]\) and \(A\) is a matrix \((3N \times 3N)\).
is now smoother (see figure (6)). The standard deviation of the first difference of the OLS-scaled $PC_1$, using the full sample period, falls from 0.17 in the contemporaneous case, to 0.14 with one lag and to 0.13 with two lags. The following table contains the average, the standard deviation, the maximum and the sum of the differentials, as in the previous section. In all cases, the sum of the differentials obtained for each month is also lower than in the contemporaneous case.

<table>
<thead>
<tr>
<th></th>
<th>Av</th>
<th>Stdev</th>
<th>Max</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contemporaneous and lag1</td>
<td>21</td>
<td>11</td>
<td>43</td>
<td>2881</td>
</tr>
<tr>
<td>Contemp., lag1 and lag2</td>
<td>21</td>
<td>11</td>
<td>40</td>
<td>2863</td>
</tr>
<tr>
<td>Contemp., lag1, lag2 and lag3</td>
<td>21</td>
<td>11</td>
<td>41</td>
<td>2863</td>
</tr>
</tbody>
</table>

Nevertheless, unless a smoother core inflation indicator is being envisaged on \textit{a priori} grounds, no obvious gain seems to be achieved by changing the currently used procedure at the Banco de Portugal (as in the previous section), specially if the shocks that hit the economy are themselves not smooth or if the transmission of the effects of these shocks onto prices is changing. Furthermore, this possible strategy requires that some type of non-contemporaneous dynamic structure as to be superimposed on the original data. In the above examples, one, two or three lags were just arbitrarily considered and analyzed. On the other hand, if leads are also to be considered as

\footnote{Although the empirical results reported in this section are all conditional on LT1, it should be mentioned that the differences between LT1 and the transformations analyzed in the previous section are again very small. In the case of LT2, LT3 and LT4, the results were 0.135, 0.159 and 0.168, respectively. Using $h_{ii} = \sigma_{\Delta^2x_i}$, where $\sigma_{\Delta^2x_i}$ represents the standard deviation of the second difference of $x_i$, the result would have been 0.138.}

\footnote{See Mankikar and Paisley (2002). On the desirable properties of a measure of “core” inflation, see, for instance, Marques, Neves and Sarmento (2000).}
input variables and unless some projections can be used, it is clear that the core inflation indicator ceases to be a real time index (not computable until \( t = T \)).

4 On the use of more than one principal component

Within the principal components methodology, i.e., behind \( PC = (X - \overline{X})H^{-1}A \), the objective is to find “clean”, orthogonal (uncorrelated) variables, which are extracted from noisy and possibly highly correlated original variables. It may be the case that a large percentage of the total variance present in the system might be retained by a few \( PC_i \) and, in this sense, the effective dimensionality of the original information set may be substantially reduced. Within factor analysis, the objective may be seen as having the inverse direction. It may be the case that the main internal features of a given set of variables may be captured by a small number of unobservable variables - the common factors. It is therefore crucial to the correct specification of the factor model the use of an adequate number of factors. Several proposals may be found in the literature to properly determine, from the observed data, the number of factors.\(^\text{13}\)

A classical way to determine the number of principal components that should be retain as factors simply relies on the contribution of each \( PC_i \) to the total variation present in the system. If the eigenvalues and eigenvectors (scaled to unity) are extracted from the correlation matrix \( \rho_X \), the appealing feature of \( \sum_{i=1}^{N} \text{Var}[PC_i] = \sum_{i=1}^{N} \lambda_i = tr(\Lambda) = tr(\rho) = N \) is valid, where a descending order of the eigenvalues has\(^\text{13}\)

\(^{13}\)A recent proposal may be found in Bai and Ng (2002).
a direct link with a descending order of variance accounted for by their respective \( PC_i \). The results of up to 6 \( PC \) were gathered and plotted in figure (7). From the graph on the left, the percentage variance accounted for by the \( PC_1 \) has never reached 60\%, during the computation of the 91 core inflation indicators (already introduced in the previous sections), and has somehow evolved along a downward trend.\(^{14}\) In the end of the sample period (see graph on the right), with \( T = 141 \), it stood at 49.9\%.

As expected, the inclusion of more \( PC_i \) increases this percentage, with marginally decreasing contributions (for instance, using all observations, \( PC_2 \) accounts for 8.7\% of the total variance present in the system; \( PC_3 \) accounts for 6.3\%; \( PC_4 \) for 4.7\%).

With the aim of capturing more variance of the observed inflation rate, it might therefore be suggested that the core inflation indicator should be derived not only from the first, but probably from two or more \( PC_i \). Coimbra and Neves (1997) or Machado et al. (2001) simply seem to assume the need for one factor and, in fact, there are several reasons to maintain this option. A first reason is based on the relevant information criteria under which the number of retained \( PC_i \) is determined. In fact, although a slight downward trend was also detected, the OLS-scaled \( PC_1 \) has always captured a high percentage of the variance of the overall inflation rate, as illustrated in the left panel of figure (8). In the end of the sample period (see right panel), it stood at 91.5\%. Therefore, the general driving behaviour of the \( CPI \) is being captured, to a large extent, by one single \( PC \), and the contribution of the remaining ones is negligible. Moreover, a descending order of \( \lambda_i \) has no direct link with a descending

\(^{14}\)The percentage of total variance accounted for by the \( i \)th \( PC \) can be sought as \( \frac{\lambda_i}{\sum(\lambda)} = \frac{\lambda_i}{N} \), which evolves in time according to the sample period; the first two \( PC \) will account \( (\lambda_1 + \lambda_2)/N \), and so on...
order of the variance accounted for by the respective OLS-scaled $PC_i$.\textsuperscript{15} For instance, the variance of the OLS-scaled $PC_4$ is higher than the variance of the OLS-scaled $PC_2$ or $PC_3$. Therefore, marginal gains would be obtained from including, successively, the $PC_i$ with $i = 1, 2$ and 3, and a sudden larger gain is obtained when the OLS-scaled $PC_4$ is also included. This result, which simply implies a clear approximation of the “core” inflation to the observed inflation, is basically explained by the fact that a higher variance accounted for by a specific $PC_i$ (obtained from eigenvectors scaled to unity) may be abruptly reduced or increased by its respective OLS-scaling.

Another related reason to reject the possibility of including more than the $PC_1$ in the computation of the core inflation indicator is based on the analysis of the eigenvectors. Using the full sample period, the scalars defining the first six eigenvectors (scaled to unity) were plotted in figure (9). With few exceptions, the scalars associated to the first eigenvector have all basically the same sign (non-negative), whereas the scalars associated to the remaining eigenvectors oscillate quite substantially between positive and negative signs.\textsuperscript{16} This implies that the different $PC_i$ may be capturing different phenomena and therefore some additional explanation will have to be put forward so as to include more the first principal component.\textsuperscript{17} By taking advantage of the positive correlations found in the database, positive $a_{i1}h_{ii}^{-1}$ simply implies that higher year-on-year rates of change of specific items of the CPI are associated

\textsuperscript{15}This is not surprising given that the variance of the $PC_i$ is dependent upon the scaling of the eigenvectors. For instance, under eigenvectors scaled to the inverse of their roots, the link with the descending order of the (respective) $\lambda_i$ does not exists, since the variances of each $PC_i$ are always equal to one.

\textsuperscript{16}The negative signs in the case of $PC_1$ were already mentioned in Machado et al. (2001).

\textsuperscript{17}An example where the $PC_1$, the $PC_2$ and the $PC_3$ are used to capture three different effects may be seen in Alexander (2001).
Eigenvectors scaled to unity

For the computation of PC1

For the computation of PC2

For the computation of PC3

For the computation of PC4

For the computation of PC5

For the computation of PC6

Figure 9 - Scalars defining the first six eigenvectors (scaled to unity)
with higher scores of the principal component. Moreover, it will be seen in the next section that the small number of negative signs in the case of the $PC_1$ can also assure that $\beta_1$ of (3) will most probably be positive, whereas the sign and the magnitude in the remaining cases is largely unknown.

5 The OLS scaling of the principal components solution

The estimation of factor scores, common to all variables present in the system, using unweighted ordinary least has been a customary procedure and the previous section has in fact computed the OLS fitted values of (3) in order to have a core inflation indicator with comparable scores to those of the CPI. Along with other properties, this so called “ad-hoc” procedure can now have a well defined interpretation if the system of equations (1) also incorporates the fact that the CPI is also a $[T \times 1]$ vector where $CPI = \sum_{i=1}^{N} \alpha_i x_i = X\alpha$. Derived from some household budget survey, $\alpha_i$ are the consumption basket weights of each item and $\Sigma_{i=1}^{N} \alpha_i = 1$. In this case, note that $x_i$ is a price index and not a price change. If the $x_i$ are previously centered: $(CPI - \sum_{i=1}^{N} \alpha_i x_i) = (CPI - \overline{CPI}) = (X - \overline{X})\alpha$. Let $\overline{CPI}$ represent the average of the CPI index.

For simplicity reasons, assume that (2) was written down with price indices and that $H$ is equal to the identity matrix, i.e. the principal components were derived from $\Sigma_X$. It is well known that from $PC = (X - \overline{X})A$, the full $[T \times N]$ matrix $X$ can be reproduced without error through a principal components representation

$$X = \overline{X} + PCA^{-1}$$

(5)

If the eigenvectors $a_i$ included in $A$ were scaled to unity, the variance of their respective $PC_i$ will be given by $\lambda_i$ and $A^{-1}$ of (5) is equal to $A'$. However, the variance of each $PC_i$ is fundamentally undetermined as each $a_i$ may be scaled to any constant $c_i$, i.e., to $a'_i a_i = c_i$, which implies that $Var(PC_i) = a'_i \Sigma_X a_i = c_i \lambda_i$. Unless other considerations are brought into the computational process, choosing $a'_i a_i = 1$ and $Var(PC_i) = \lambda_i$ is just one of the possible options. Let $PC_i^U$ represent the $i$th principal component obtained from unit length eigenvectors.

After having unfolded the matrix $A$ and using, for simplicity reasons, unit length eigenvectors.

---

18 See, for instance, Johnson and Wichern (2002).
19 See Coimbra and Neves (1997, p.31).
eigenvectors, it can now be showed that both the CPI, and its overall variance \( (\sigma_{\text{CPI}}^2) \) can also be reproduced without error through a principal components structure. Algebraically, the initial linear combination of the original variables, i.e., the overall CPI, is just going to be expressed as a linear combination of another set of vectors. From (5), the entire \([T \times N]\) matrix \(X\) collapses to a single equation for the CPI and, with a principal component solution, the following orthogonal factor model emerges

\[
\text{CPI} = X\alpha = (\overline{X} + PC^U A')\alpha = \overline{CPI} + \sum_{i=1}^{N} PC_i U_i a_i' \alpha
\]

\[
\text{Var}[\text{CPI}] \equiv \sigma_{\text{CPI}}^2 = \text{Var}[PC^U A'\alpha] = [\alpha' A PC^U][PC^U A'\alpha] = [\alpha' A \Lambda A'\alpha] = \alpha' \Sigma_X \alpha
\]

In short, the overall CPI index and its variance can be exactly reproduced within an orthogonal factor model framework with as many common factors (principal components) as variables. Thus, without any assumption on the data generation process of each variable, the CPI is not only a product of the aggregation of all its basic items; it is also a product the all its scaled principal components. Moreover, if only \( p < N \) principal components are retained within this framework, this will produce an index (by construction) with comparable scores as those of the observed overall CPI (it will have, namely, the same average), but with a smaller variance. Those retained \( PC \) will now account for a certain percentage of the overall variance of the CPI, which is a different quantity against the accounted percentage of the total variance present in the system.

It turns out that equation (6) not only represents a solution for the fundamental eigenvectors length indeterminacy, but also uniquely determines the variance of the CPI accounted for by each \( PC_i \).\(^{20}\) Model (6) surpasses the problem of having to choose a particular eigenvector length to solve the initial problem \( \Sigma_X a = \lambda a \), given that all possible choices can be proved to collapse to (6). Changing the length of an eigenvector will change the variance of the \( PC_i \) but this will also functionally change the scaling constant and the inner product of both will, most conveniently, remain unchanged. Under this result, the percentage variance of the CPI accounted for by these products is not conditional on the eigenvectors length.\(^{21}\) Using unit length

\(^{20}\) The proofs are included in Annex 2.

\(^{21}\) Each product \( PC_i U_i a_i' \) is therefore independent of the eigenvector’s length (\( a_i'\alpha \) is the result of an inner product of dimensions \([1 \times N] \times [N \times 1]\)). Note also that to accommodate alternative linear transformation, such as the
eigenvectors, $\text{Var}[\text{CPI}]$ can be further subjected to the following breakdown.

$$\text{Var}[\text{CPI}] \equiv \sigma_{\text{CPI}}^2 = \sum_{i=1}^{n} \text{Var}[\text{PC}_i^U](\alpha'a_i\alpha') = \sum_{i=1}^{n} \lambda_i(\alpha'a_i\alpha')$$

Within this framework, the use of one or several $\text{PC}_i^U$ can be easily scrutinized and naturally allows to classify any core inflation indicator stemming from this specification to be, by construction, a “trimmed variance” index (to use an expression from the trimmed mean literature). Finally, it is relevant to point out that equation (6) is fully equivalent to an OLS scaling. A regression between the CPI and the $\text{PC}_i$ structure, i.e. $\text{CPI} = \beta_0 + \beta_1 \text{PC}_1^U + \ldots + \beta_N \text{PC}_N^U + \varepsilon$ will pre-determine that $\beta_0, \beta_1, \ldots, \beta_N$ will be equal to the above results $\overline{\text{CPI}}, a'_1\alpha, \ldots, a'_N\alpha$, respectively, and, in this case, $\varepsilon = 0$. By construction,

$$(\text{PC}_i^U \text{PC}_i^U)^{-1} \text{PC}_i^U \text{CPI} = (\text{PC}_i^U \text{PC}_i^U)^{-1} \text{PC}_i^U(\text{PC}_i^U A')\alpha = [ a'_1\alpha \ a'_2\alpha \ \ldots \ a'_N\alpha ]'$$

Assume for now that the core indicator ($\text{CPI}^T$) was defined as the OLS fitted values of $\text{CPI} = \beta_0 + \beta_1 \text{PC}_1^U + \varepsilon$ (i.e. it only uses the first $\text{PC}$). It is now clear that this specification is not ad hoc and, instead, has the following consequences: (i) the $\text{CPI}^T$ is pre-determined and is equal to $\overline{\text{CPI}} + a'_1\alpha \text{PC}_1$, as the PC are orthogonal; (ii) $\text{Var}[\text{CPI}^T]$ is pre-determined and is equal to $\lambda_1(\alpha'a_1\alpha')$, where $\text{Var}[\text{CPI}^T]$ represents part of the variance of the overall $\text{CPI}$ that is being captured; (iii) $\varepsilon = \Sigma_{i=2}^{n} a'_i\alpha \text{PC}_i$ is ignored; and, as a consequence, (iv) $\text{Var}[\varepsilon] = \Sigma_{i=2}^{n} \text{Var}[\text{PC}_i]$ ($\alpha'a_i\alpha'$) is also ignored. The $\text{CPI}^T$ will be, by construction, smoother than the CPI by an amount given by $\text{Var}[\varepsilon]$ as the $\text{CPI}^T$ is “trimming” the variance of the CPI by a percentage given by $\Sigma_{i=2}^{n} \lambda_i(\alpha'a_i\alpha')/(\sigma_{\text{CPI}}^2)$. $\text{Var}[\varepsilon]$ can be seen as the ignored variability of the overall CPI against the variability of the $\text{CPI}^T$.

After having established in the previous section that the core inflation indicator derived from the correlation matrix does emerge as a reasonable option, namely under non-stationary input variables, the OLS scaling can now be simply interpreted as the appropriate linear combination of $\text{PC}_i$ that replicates the observed inflation rate. Given that the database that has been used so far is not made of indices but
of $N$ year-on-year rates of changes, the restriction for the overall $CPI$ index has to be adapted for the rate of change of the overall $CPI$ index. With monthly data, this can be implemented by changing the database from $x_i \equiv (I_{it} - I_{it-12})/I_{it-12}$, to $x_i \equiv \omega_i \times (I_{it} - I_{it-12})/I_{it-12}$, i.e., to contributions for the year-on-year rate of change to the $CPI$. $I_i$ is a specific price index, $i = 1 \ldots N$ and $\omega_i = (I_{it-12}/CPI_{t-12}) \times \alpha_i$. In this case, an OLS regression such as $CPI = \beta_0 + \beta_1 PC_1$ would simply collapse to

$$CPI^T = CPI + (\Sigma_{i=1}^{N} a_{i1} \sigma_i)PC_1^U$$

(7)

where the data has been standardized, $a_{i1}$ represents the first eigenvector scaled to unity, $\sigma_i$ is the standard deviations of the new $x_i$, $PC_1^U$ is the first principal component obtained from the first eigenvector scaled to unity, and $CPI$ is the average inflation rate. To put it differently, no OLS regression would have to be implemented and no parameter has to be estimated as a final stage for the determination of the core inflation indicator if the length of the $j$th eigenvector is not scaled to unity, but to $(\Sigma_{i=1}^{N} a_{i1} \sigma_i)^2$. The percentage of the overall variance of the rate of change of the $CPI$ that is being captured is given by $\text{Var}[CPI^T]/\sigma_{CPI}^2$, where $\text{Var}[CPI^T] = (\beta_1^2 \lambda_1) = (\Sigma_{i=1}^{N} a_{i1} \sigma_i)^2 \lambda_1$. Under LT1 and with unit length eigenvectors, model (7) can easily be expanded to

$$CPI = \Sigma_{i=1}^{N} x_i = CPI + (\Sigma_{i=1}^{N} a_{i1} \sigma_i)PC_1^U + \ldots (\Sigma_{i=1}^{N} a_{iN} \sigma_i)PC_N^U.$$  

Moreover, as already mentioned, if the $a_i$ are extracted from the correlation matrix, which is the case under LT1, this choice involves an arbitrary decision to make the variables “equally important”. The transformed variables will be indistinguishable from a “variance and location point of view”, as the diagonal elements of the correlation matrix are all unity. Using the new database, it also seems conceptually appealing to make the contributions and not the rates of change the “equally important variables”. An implementation of (7) may be found in figure (10).

The new database incorporates the weighting schemes associated to the Household Budget Surveys of 1989-90, 1994-95 and 2000, used for the calculation of the CPI(1991=100), CPI(1997=100) and the CPI (2002=100), respectively. As an improvement against the previous core inflation indicators and with the objective of widening the information contained in the $PC_1$, the database was also expanded with

24 If the $PC_i$ are extracted from standardized data and the consumption basket weights remain unchanged over time, note also that the use of indices or weighted indices is a totally irrelevant issue, as it has no effect on the correlation matrix. Nevertheless, equation (6) can only be seen as an approximation for the specification in changes.
two additional variables related with housing expenditures (housing rents and prices of maintenance and repair of the dwelling). As expected, the overall behaviour of the observed inflation rate is once again being captured by this factor model.

6 Conclusions

This paper has investigated if the (OLS-scaled) first principal component ($PC_1$), extracted from standardized yearly rates of change of basic items of the CPI, represents a reasonable option for a core inflation indicator. Currently, the Banco de Portugal uses this procedure to measure (and regularly publish) a core inflation indicator for the Portuguese case. A special focus was placed on the final stage of the process of finding the core indicator, in which the $PC_1$ is subject to an OLS scaling, through a regression of the CPI inflation rate on the $CP_1$. The fitted values of this regression - a so called ad-hoc procedure - determines the core inflation level.

From the confrontation of structurally different linear transformations of the original data, including the one suggested by Machado et al. (2001), it was concluded that no obvious gain is clearly achieved by changing the standardization procedure for any other of the remaining possibilities. The overall final results of all transformations were not easily distinguishable, which may be seen as particularly striking given that two of those transformations were purely arbitrary and not motivated by any

25During the period 1992-1997, only annual observations are available. In line with the monthly behaviour that has been observed since 1997, it was therefore assumed that those annual figures were basically determined in the beginning of each year.

26See Annex 3 for more information.
reason (except for comparison purposes). Even though most input variables are non-stationary, which was the main reason underlying the suggested transformation of Machado et al. (2001), the results were not find to be structurally dissimilar.

Secondly, if the objective is the find a smoother core inflation indicator then the one currently in use, then lagged variables can be stacked to the original database. However, it was also argued that unless other reasons are brought into the decision process, this increased smoothness does not seem to represent per se a clear superiority feature. Some type of non-contemporaneous dynamic structure as to be superimposed on the original data and if the shocks that hit the economy are not smooth, why should the core inflation indicator be smooth? In addition, the inclusion of leads would prevent the indicator to be updated until the last period (abstracting from the possible use of some type of projections).

Finally, within a classical approach, the contents of the remaining principal components were also explored. Given that the main objective in the current analysis is to capture the general driving behaviour of the overall inflation rate, and not the total variance present in the system, the conclusion was that a single component (the $PC_1$) seems sufficient. In the last 91 estimates, the OLS-scaled $PC_1$ has always captured more than 90% of the total variance of the $CPI$. Moreover, it was claimed that the other eigenvectors may be capturing other effects rather than the “trend component”. The sign and the magnitude of the OLS scaling of the remaining $PC_i$ is also largely unknown.

In general, the empirical evidence incorporated throughout this paper does suggest that the OLS-scaled $PC_1$, extracted from standardized yearly rates of change of basic items of the CPI, does represent a reasonable option for a core inflation indicator. Moreover, it was showed that no OLS parameter has to be estimated as a final stage of the process of finding the core indicator. Instead of being seen as an ad-hoc procedure, the OLS scaling can simply be interpreted as an appropriate linear combination of $PC_i$ that replicates the observed inflation rate. To such purpose, this paper introduced an orthogonal factor model that fully reproduces the $CPI$, in which the following properties apply: (i) the results are not conditional on the eigenvectors length; (ii) the variance of the $CPI$ accounted for by each component is unique, and (iii) the outcome is equivalent to an OLS scaling of the components, using the $PC_i$
as explanatory variables. The model was written down in price levels and in price changes (yearly rates) and this has given a clear interpretation to the OLS scaling. To achieve these results, it is only necessary to respect the fact that the CPI can be written down as a linear combination of the input data. In the latter case, under standardization, it would be the contributions and not the rates of change that could be made “equally important variables” (which implies that the weights of the CPI have to be fully taken into account for the determination of the $CP_i$) and no OLS regression has to be implemented (given that the final results are fully equivalent to a well defined scaling of the first eigenvector).

References


Bai, J. and Ng, S. (2002), ‘Determining the number of factors in approximate factor models’, *Econometrica* 70, 191–221.


Annex 1

List of CPI items

<table>
<thead>
<tr>
<th>Item</th>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>UnpF</td>
<td>x1</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>x2</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>x3</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>x4</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>x5</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>x6</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>x7</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>x8</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>x9</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>x10</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>x11</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>x12</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>x13</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>x14</td>
</tr>
<tr>
<td>1</td>
<td>Neig</td>
<td>x42</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>x43</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>x44</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>x45</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>x46</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>x47</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>x48</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>x49</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>x50</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>x51</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>x52</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>x53</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>x54</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>x55</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>x56</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>x57</td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>x58</td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>x59</td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>x60</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>x61</td>
</tr>
<tr>
<td>21</td>
<td></td>
<td>x62</td>
</tr>
<tr>
<td>22</td>
<td></td>
<td>x63</td>
</tr>
<tr>
<td>23</td>
<td></td>
<td>x64</td>
</tr>
<tr>
<td>24</td>
<td></td>
<td>x65</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>x66</td>
</tr>
<tr>
<td>26</td>
<td></td>
<td>x67</td>
</tr>
<tr>
<td>1</td>
<td>Serv</td>
<td>x68</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>x69</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>x70</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>x71</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>x72</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>x73</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>x74</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>x75</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>x76</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>x77</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>x78</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td>x79</td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>x80</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td>x81</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>x82</td>
</tr>
<tr>
<td>16</td>
<td></td>
<td>x83</td>
</tr>
<tr>
<td>17</td>
<td></td>
<td>x84</td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>x85</td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>x86</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td>x87</td>
</tr>
<tr>
<td>21</td>
<td></td>
<td>x88</td>
</tr>
<tr>
<td>22</td>
<td></td>
<td>x89</td>
</tr>
<tr>
<td>23</td>
<td></td>
<td>x90</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item</th>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Enrg</td>
<td>x39</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>x40</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>x41</td>
</tr>
</tbody>
</table>
Annex 2

Using unit length eigenvectors, equation (6) respects the following propositions:

1. It represents a solution for the fundamental eigenvectors length indeterminacy.

Principal components were said to be special linear combinations of \( x_1, x_2, \ldots, x_N \), as they provide a solution to the problem of \( \text{Max} \{ \text{Var}(a_1x_1 + a_2x_2 + \ldots + a_Nx_N) \} = \text{Max} \{ \text{Var}(\text{PC}_i) \} \) = \( \text{Max}(a'_1\Sigma x a_1) \), where the constants \( a_i \) have to be found. With the aim of finding uncorrelated \( \text{PC}_i \) whose variances are as large as possible, it is common to restrict the analysis to eigenvectors of unit length. However, a fundamental eigenvector length indeterminacy does exist, and other commonly used cases are the scaling of the eigenvectors to their roots or to the inverse of their roots. Therefore, in general, the eigenvectors may be scaled to any constant \( c_i \), implying that the matrices \( A \) and \( \text{PC} \) will change in accordance to such scaling. For notation purposes, assume that (i) \( C \) is a diagonal matrix with \( c_i \) in the main diagonal, \( i = 1, \ldots, N \), (ii) the matrices \( A^U \) and \( \text{PC}^U \) have been derived from eigenvectors scaled to unity and that (iii) the matrices \( A^C \) and \( \text{PC}^C \) have been derived from eigenvectors scaled to \( c_i \). In the general case, therefore, \( a'_ia_i = c_i \) and \( a'_i\Sigma x a_i = c_i\lambda_i \). In a matrix representation

\[
A^C = A^U C^{1/2} \\
\text{PC}^C = \text{PC}^U C^{1/2}
\]

\[
\text{CPI} = X\alpha = (X + \text{PC}^C(A^C)^{-1})\alpha = (X + (\text{PC}^U C^{1/2})(C^{-1/2}(A^U)^{-1}))\alpha \\
(\bar{X} + \text{PC}^U A)\alpha = \text{CPI} + \Sigma_{i=1}^N \text{PC}^U a'_i\alpha, \text{ which is equal to (6)}.
\]

2. It uniquely determines the variance of the CPI accounted for by each \( \text{PC} \).

The fundamental eigenvector length indeterminacy has a full equivalence on the level of the variance of each \( \text{PC}_i \), which is therefore an open issue left for the analyst to decide. Note that although the matrix \( A \) changes within each scaling, the eigenvalues coming from the digitalization process of \( \Sigma_X \) do not change \( (\Lambda^{-1}\Sigma_X A = \Lambda) \), remains unchanged for all possibilities, where \( \Lambda \) is a diagonal matrix with \( \lambda_i \) in the main diagonal. Nevertheless, from the first eigenvalue/eigenvector problem, i.e. from \( \Sigma_X a_i = \lambda_i a_i \), it should be noted that \( a'_i\Sigma_X a_i = a'_i\lambda_i a_i = \lambda_i a'_i a_i \). Choosing \( a'_ia_i = 1 \) and therefore \( \text{Var}[\text{PC}_i] = \text{Var}[a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{iN}x_N] = a'_i\Sigma_X a_i = \lambda_i \) is, again, just one possible option. For instance, if the eigenvector is scaled to its root, i.e., \( a'_ia_i = \lambda_i \), then \( a'_i\Sigma_X a_i = \lambda_i^2 \). If the eigenvector is scaled to the inverse of its root, i.e. \( a'_ia_i = 1/\lambda_i \), then \( a'_i\Sigma_X a_i = 1 \). Within equation (6) this is no longer the case and the intuition can be put forward in the following way. Given that \( a'_ia_i = 1 \) is not the sole solution to the eigenvalue/eigenvector problem, the eigenvector can effectively be scaled to any constant \( c_1 \), which will, automatically, change the variance of \( \text{PC}_1 \) to \( c_1\lambda_1 \). This scaling issue represents an arbitrary decision and it will always be left to the analyst to decide which is the appropriate length of the eigenvector (and, therefore, the appropriate variance of \( \text{PC}_1 \)). However, any “new” \( \text{PC}_1 \) is within this factor model framework just the “old one” (that comes from \( a'_ia_i = 1 \)), multiplied by a particular constant, i.e. \( \sqrt{c_1}PC_1^U \). Therefore, the “old” \( a'_1\alpha \) of (6), is just transformed into \( (1/\sqrt{c_1})a'_1\alpha \), where \( a'_ia_i = 1 \), making it clear that the old \( (PC_1^U) \times (a'_1\alpha) \) will be given by the “new” \( (\sqrt{c_1}PC_1^U) \times [(1/\sqrt{c_1})a'_1\alpha] \). Most conveniently, this outcome is nothing more then the old \( PC_1^U a'_1\alpha \), again. On the other hand, the “old” \( \text{Var}[\text{PC}_1] (a'_1a'_1\alpha) = \lambda_1(a'_1a'_1\alpha) \) will be transformed into \( (c_1\lambda_1)(a'(1/\sqrt{c_1})a_1(1/\sqrt{c_1})a'_1\alpha) = \lambda_1(a'_1a'_1\alpha) \), i.e., it remains unchanged.
Figure 11 - Some results associated to the “new” core inflation indicators (recursive procedure).

<table>
<thead>
<tr>
<th></th>
<th>Av</th>
<th>Stdev</th>
<th>Max</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>LT1</td>
<td>21</td>
<td>10</td>
<td>47</td>
<td>2952</td>
</tr>
</tbody>
</table>
WORKING PAPERS

2000

1/00 UNEMPLOYMENT DURATION: COMPETING AND DEFECTIVE RISKS
   — John T. Addison, Pedro Portugal

2/00 THE ESTIMATION OF RISK PREMIUM IMPLICIT IN OIL PRICES
   — Jorge Barros Luís

3/00 EVALUATING CORE INFLATION INDICATORS
   — Carlos Robalo Marques, Pedro Duarte Neves, Luís Morais Sarmento

4/00 LABOR MARKETS AND KALEIDOSCOPIC COMPARATIVE ADVANTAGE
   — Daniel A. Traça

5/00 WHY SHOULD CENTRAL BANKS AVOID THE USE OF THE UNDERLYING INFLATION INDICATOR?
   — Carlos Robalo Marques, Pedro Duarte Neves, Afonso Gonçalves da Silva

6/00 USING THE ASYMMETRIC TRIMMED MEAN AS A CORE INFLATION INDICATOR
   — Carlos Robalo Marques, João Machado Mota

2001

1/01 THE SURVIVAL OF NEW DOMESTIC AND FOREIGN OWNED FIRMS
   — José Mata, Pedro Portugal

2/01 GAPS AND TRIANGLES
   — Bernardino Adão, Isabel Correia, Pedro Teles

3/01 A NEW REPRESENTATION FOR THE FOREIGN CURRENCY RISK PREMIUM
   — Bernardino Adão, Fátima Silva

4/01 ENTRY MISTAKES WITH STRATEGIC PRICING
   — Bernardino Adão

5/01 FINANCING IN THE EUROSYSTEM: FIXED VERSUS VARIABLE RATE TENDERS
   — Margarida Catalão-Lopes

6/01 AGGREGATION, PERSISTENCE AND VOLATILITY IN A MACROMODEL
   — Karim Abadir, Gabriel Talmain

7/01 SOME FACTS ABOUT THE CYCLICAL CONVERGENCE IN THE EURO ZONE
   — Frederico Belo

8/01 TENURE, BUSINESS CYCLE AND THE WAGE-SETTING PROCESS
   — Leandro Arozamena, Mário Centeno

9/01 USING THE FIRST PRINCIPAL COMPONENT AS A CORE INFLATION INDICATOR
   — José Ferreira Machado, Carlos Robalo Marques, Pedro Duarte Neves, Afonso Gonçalves da Silva

10/01 IDENTIFICATION WITH AVERAGED DATA AND IMPLICATIONS FOR HEDONIC REGRESSION STUDIES
     — José A.F. Machado, João M.C. Santos Silva

2002

1/02 QUANTILE REGRESSION ANALYSIS OF TRANSITION DATA
     — José A.F. Machado, Pedro Portugal

2/02 SHOULD WE DISTINGUISH BETWEEN STATIC AND DYNAMIC LONG RUN EQUILIBRIUM IN ERROR CORRECTION MODELS?
     — Susana Betas, Carlos Robalo Marques
<table>
<thead>
<tr>
<th>Date</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>3/02</td>
<td>MODELLING TAYLOR RULE UNCERTAINTY</td>
<td>Fernando Martins, José A. F. Machado, Paulo Soares Esteves</td>
</tr>
<tr>
<td>4/02</td>
<td>PATTERNS OF ENTRY, POST-ENTRY GROWTH AND SURVIVAL: A COMPARISON BETWEEN DOMESTIC AND FOREIGN OWNED FIRMS</td>
<td>José Mata, Pedro Portugal</td>
</tr>
<tr>
<td>6/02</td>
<td>AN “ART”, NOT A “SCIENCE”? CENTRAL BANK MANAGEMENT IN PORTUGAL UNDER THE GOLD STANDARD, 1854-1891</td>
<td>Jaime Reis</td>
</tr>
<tr>
<td>7/02</td>
<td>MERGE OR CONCENTRATE? SOME INSIGHTS FOR ANTITRUST POLICY</td>
<td>Margarida Catalão-Lopes</td>
</tr>
<tr>
<td>8/02</td>
<td>DISENTANGLING THE MINIMUM WAGE PUZZLE: ANALYSIS OF WORKER ACCESSIONS AND SEPARATIONS FROM A LONGITUDINAL MATCHED EMPLOYER-EMPLOYEE DATA SET</td>
<td>Pedro Portugal, Ana Rute Cardoso</td>
</tr>
<tr>
<td>9/02</td>
<td>THE MATCH QUALITY GAINS FROM UNEMPLOYMENT INSURANCE</td>
<td>Mário Centeno</td>
</tr>
<tr>
<td>10/02</td>
<td>HEDONIC PRICES INDEXES FOR NEW PASSENGER CARS IN PORTUGAL (1997-2001)</td>
<td>Hugo J. Reis, J.M.C. Santos Silva</td>
</tr>
<tr>
<td>11/02</td>
<td>THE ANALYSIS OF SEASONAL RETURN ANOMALIES IN THE PORTUGUESE STOCK MARKET</td>
<td>Miguel Balbina, Nuno C. Martins</td>
</tr>
<tr>
<td>12/02</td>
<td>DOES MONEY GRANGER CAUSE INFLATION IN THE EURO AREA?</td>
<td>Carlos Robalo Marques, Joaquim Pina</td>
</tr>
<tr>
<td>13/02</td>
<td>INSTITUTIONS AND ECONOMIC DEVELOPMENT: HOW STRONG IS THE RELATION?</td>
<td>Tiago V.de V. Cavalcanti, Álvaro A. Novo</td>
</tr>
<tr>
<td>2003</td>
<td>FOUNDRING CONDITIONS AND THE SURVIVAL OF NEW FIRMS</td>
<td>P.A. Geroski, José Mata, Pedro Portugal</td>
</tr>
<tr>
<td>1/03</td>
<td>THE TIMING AND PROBABILITY OF FDI: An Application to the United States Multinational Enterprises</td>
<td>José Brandão de Brito, Felipa de Melo Sampayo</td>
</tr>
<tr>
<td>2/03</td>
<td>OPTIMAL FISCAL AND MONETARY POLICY: EQUIVALENCE RESULTS</td>
<td>Isabel Correia, Juan Pablo Nicolini, Pedro Teles</td>
</tr>
<tr>
<td>3/03</td>
<td>FORECASTING EURO AREA AGGREGATES WITH BAYESIAN VAR AND VECM MODELS</td>
<td>Ricardo Mourinho Félix, Luís C. Nunes</td>
</tr>
<tr>
<td>4/03</td>
<td>CONTAGIOUS CURRENCY CRISES: A SPATIAL PROBIT APPROACH</td>
<td>Álvaro Novo</td>
</tr>
<tr>
<td>5/03</td>
<td>THE DISTRIBUTION OF LIQUIDITY IN A MONETARY UNION WITH DIFFERENT PORTFOLIO RIGIDITIES</td>
<td>Nuno Alves</td>
</tr>
<tr>
<td>6/03</td>
<td>COINCIDENT AND LEADING INDICATORS FOR THE EURO AREA: A FREQUENCY BAND APPROACH</td>
<td>António Rua, Luís C. Nunes</td>
</tr>
<tr>
<td>7/03</td>
<td>WHY DO FIRMS USE FIXED-TERM CONTRACTS?</td>
<td>José Varejão, Pedro Portugal</td>
</tr>
<tr>
<td>8/03</td>
<td>NONLINEARITIES OVER THE BUSINESS CYCLE: AN APPLICATION OF THE SMOOTH TRANSITION AUTOREGRESSIVE MODEL TO CHARACTERIZE GDP DYNAMICS FOR THE EURO-AREA AND PORTUGAL</td>
<td>Francisco Craveiro Dias</td>
</tr>
<tr>
<td>Date</td>
<td>Title</td>
<td>Authors</td>
</tr>
<tr>
<td>------</td>
<td>----------------------------------------------------------------------</td>
<td>----------------------------------------------</td>
</tr>
<tr>
<td>10/03</td>
<td>WAGES AND THE RISK OF DISPLACEMENT</td>
<td>Anabela Carneiro, Pedro Portugal</td>
</tr>
<tr>
<td>11/03</td>
<td>SIX WAYS TO LEAVE UNEMPLOYMENT</td>
<td>Pedro Portugal, John T. Addison</td>
</tr>
<tr>
<td>12/03</td>
<td>EMPLOYMENT DYNAMICS AND THE STRUCTURE OF LABOR ADJUSTMENT COSTS</td>
<td>José Varejão, Pedro Portugal</td>
</tr>
<tr>
<td>13/03</td>
<td>THE MONETARY TRANSMISSION MECHANISM: IS IT RELEVANT FOR POLICY?</td>
<td>Bernardino Adão, Isabel Correia, Pedro Teles</td>
</tr>
<tr>
<td>14/03</td>
<td>THE IMPACT OF INTEREST-RATE SUBSIDIES ON LONG-TERM HOUSEHOLD DEBT:</td>
<td>Nuno C. Martins, Ernesto Villanueva</td>
</tr>
<tr>
<td></td>
<td>EVIDENCE FROM A LARGE PROGRAM</td>
<td></td>
</tr>
<tr>
<td>15/03</td>
<td>THE CAREERS OF TOP MANAGERS AND FIRM OPENNESS: INTERNAL VERSUS EXTERNAL</td>
<td>Francisco Lima, Mário Centeno</td>
</tr>
<tr>
<td></td>
<td>LABOUR MARKETS</td>
<td></td>
</tr>
<tr>
<td>16/03</td>
<td>TRACKING GROWTH AND THE BUSINESS CYCLE: A STOCHASTIC COMMON CYCLE</td>
<td>João Valle e Azevedo, Siem Jan Koopman, António</td>
</tr>
<tr>
<td></td>
<td>MODEL FOR THE EURO AREA</td>
<td>Rua</td>
</tr>
<tr>
<td>17/03</td>
<td>CORRUPTION, CREDIT MARKET IMPERFECTIONS, AND ECONOMIC DEVELOPMENT</td>
<td>António R. Antunes, Tiago V. Cavalcanti</td>
</tr>
<tr>
<td>18/03</td>
<td>BARGAINED WAGES, WAGE DRIFT AND THE DESIGN OF THE WAGE SETTING SYSTEM</td>
<td>Ana Rute Cardoso, Pedro Portugal</td>
</tr>
<tr>
<td>19/03</td>
<td>UNCERTAINTY AND RISK ANALYSIS OF MACROECONOMIC FORECASTS:</td>
<td>Álvaro Novo, Maximiano Pinheiro</td>
</tr>
<tr>
<td></td>
<td>FAN CHARTS REVISITED</td>
<td></td>
</tr>
</tbody>
</table>

2004

<table>
<thead>
<tr>
<th>Date</th>
<th>Title</th>
<th>Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/04</td>
<td>HOW DOES THE UNEMPLOYMENT INSURANCE SYSTEM SHAPE THE TIME PROFILE OF</td>
<td>John T. Addison, Pedro Portugal</td>
</tr>
<tr>
<td></td>
<td>JOBLESS DURATION?</td>
<td></td>
</tr>
<tr>
<td>2/04</td>
<td>REAL EXCHANGE RATE AND HUMAN CAPITAL IN THE EMPIRICS OF ECONOMIC</td>
<td>Delfim Gomes Neto</td>
</tr>
<tr>
<td></td>
<td>GROWTH</td>
<td></td>
</tr>
<tr>
<td>3/04</td>
<td>ON THE USE OF THE FIRST PRINCIPAL COMPONENT AS A CORE INFLATION</td>
<td>José Ramos Maria</td>
</tr>
<tr>
<td></td>
<td>INDICATOR</td>
<td></td>
</tr>
</tbody>
</table>