EUROPEAN CENTRAL BANK

WORKING PAPER SERIES

NO 627 / MAY 2006

BCE ECB EZB EKT EKP

EURO AREA BANKING SECTOR INTEGRATION USING HIERARCHICAL CLUSTER ANALYSIS TECHNIQUES

by Christoffer Kok Sørensen
and Josep Maria Puigvert Gutiérrez

EUROPEAN CENTRAL BANK

WORKING PAPER SERIES

NO 627 / MAY 2006

EURO AREA BANKING SECTOR INTEGRATION USING HIERARCHICAL CLUSTER ANALYSIS TECHNIQUES [1]

by Christoffer Kok Sørensen [2]
and Josep Maria Puigvert Gutiérrez [2]

In 2006 all ECB publications will feature a motif taken from the €5 banknote.

1 Comments from F. Drudi, J. Fortiana, D. Marques Ibañez, A. Colangelo and an anonymous referee are gratefully acknowledged. The views expressed in this paper are those of the authors and do not necessarily represent those of the European Central Bank.

2 Both European Central Bank, Postfach 160319, 60066 Frankfurt am Main, Germany; e-mails: christoffer.kok_sorensen@ecb.int; josep_maria.puigvert@ecb.int

Address
Kaiserstrasse 29
60311 Frankfurt am Main, Germany

Postal address
Postfach 16 03 19
60066 Frankfurt am Main, Germany

Telephone
+49 69 1344 0

Internet
http://www.ecb.int

Fax
+49 69 1344 6000

Telex
411 144 ecb d

ISSN 1561-0810 (print)
ISSN 1725-2806 (online)

CONTENTS

Abstract

In this study we apply cluster analysis techniques, including a novel smoothing method, to detect some basic patterns and trends in the euro area banking sector in terms of the degree of homogeneity of countries. We find that in the period 1998-2004 the banking sectors in the euro area countries seem to have become somewhat more homogeneous, although the results are not unequivocal and considerable differences remain, leaving scope for further integration. In terms of clustering, the Western and Central European countries (like Germany, France, Belgium, and to some extent also the Netherlands, Austria and Italy) tend to cluster together, while Spain and Portugal and more recently also Greece usually are in the same distinct cluster. Ireland and Finland form separate clusters, but overall tend to be closer to the Western and Central European cluster.

JEL classification: C49; F36; G21

Keywords: financial integration; cluster analysis; banking sector

Non-technical summary

In this study we apply cluster analysis techniques to examine the degree of financial integration in the euro area, focusing in particular on the banking industry. Using cluster analysis we develop an alternative tool to the more traditional measures of financial integration, in particular "Law of One Price"-based indicators, typically applied in this strand of the economic literature. Basing our analysis on a number of banking, financial and economic indicators for the euro area countries and applying some newly developed cluster analysis techniques, we examine two basic questions: 1) to what extent do euro area countries "cluster" together and which countries tend to be in the same clusters (i.e. what is the degree of cross-country homogeneity); 2) how does the clustering of countries evolve over time. That is, we attempt to answer the question whether the countries have become more similar during the period under consideration (1998-2004) – in other words, has financial integration proceeded or not?

We focus in particular on banking sector integration, as this is arguably the financial market segment of the euro area which is the least integrated. However, despite this apparent lack of banking integration, the introduction of the euro and the ongoing completion of the single market for financial services may have had a beneficial effect on the degree of integration. As banks remain major players in the euro area financial system and hence play a key role in the transmission of monetary policy impulses to the real economy, a more homogeneous and integrated banking sector should help ensuring a uniform and effective monetary policy transmission mechanism in the common currency area.

Cluster analysis may be seen as a complementary tool to traditional regression analysis where the relation between exogenous and endogenous variables is determined from the outset. In cluster analysis, the researcher let the data speak for themselves without imposing any a priori restrictions. While the derived clusters provide information on the often complex interrelationships between related variables, cluster analysis does not produce any definitive results or causality prescriptions. The results are more diagnostic in nature and may provide some insights into the underlying interlinkages between a set of variables (and countries) that common econometric techniques would not be able to detect. In our paper, we employ both a classical hierarchical cluster method and a newly developed smoothing method that is especially suited for analysing changes in the clustering over time.

Focusing on a data set consisting mainly of banking-related variables, we find that the euro area countries overall have become more homogenous since the introduction of the euro, although significant differences still remain leaving scope for further integration in the years ahead. In terms of the clustering of countries, the Western and Central European countries (i.e. Germany, France, Belgium, and to some extent also Austria, Italy and the Netherlands, and more recently also Ireland) tend to form distinct clusters, and similarly countries like Spain and Portugal, and recently also Greece, tend to form another set of clusters.

1. Introduction

The topic of European financial integration has been at the forefront of economic research in recent years, in particular sparked by the advent of Economic and Monetary Union and the endeavours to create a Single Market for Financial Services. There is not one single agreed measure of financial integration and the empirical literature has applied many different approaches, often but not exclusively based on the so-called "law of one price" and using various types of convergence and dispersion measures.[3] In this study, we deviate somewhat from the main strands of the literature on financial integration by using hierarchical cluster analysis with the objective of detecting some basic patterns in the euro area financial system in terms of the degree of homogeneity of countries.

Cluster analysis is a useful tool to examine complex relations among national characteristics and international linkages without imposing any a priori restrictions on the interrelationships. That is, when linkages between related variables and across countries are too complex to model under a single-equation framework (assuming causal relations), it might be preferable to let the data guide themselves rather than a priori imposing a test equation upon them. Cluster analysis may hence be seen as a complementary analysis to regression-style studies where exogenous and endogenous variables are designed at the outset. It is important to note that cluster analysis does not produce results that are definitive in nature. Cluster analysis is more diagnostic in nature and may be used as a data reduction technique, which could eventually provide input to other types of statistical analysis of the data.

In our study, we focus on the degree of homogeneity (and hence implicitly the degree of integration) of the banking sector in the euro area countries and its development over time in the period 1998-2004. The banking sector is usually found to be the least integrated segment of the euro area financial system[4] and therefore we should a priori not expect to find a very tight clustering of our data. However, it can not be ruled out

---

[3] For some recent European-oriented studies, see e.g. Galati and Tsatsamoris (2001), Fratzscher (2001), Giannetti et al. (2002), London Economics (2002), Kleimeier and Sander (2002), Cabral et al. (2002), Adam et al. (2002), Hartmann et al. (2003), Adjouté and Danthine (2003), Manna (2004) and Baele et al. (2004).

[4] See in particular Baele et al. (2004), Cabral et al. (2002) and Gropp and Corvoisier (2001).

that the introduction of the euro may have fuelled cross-border competition and interlinkages (despite the limited number of cross-border mergers) among euro area banks thereby setting up a process of structural convergence of the banking sectors in the euro area countries. This is, in fact, what we set out to investigate using cluster analysis techniques. Hence, we analyse i) which countries tend to form clusters together (and are therefore relatively similar in terms of structures) and ii) whether the clustering changes over time both in terms of which countries cluster together and in terms of whether the clusters in general become more homogenous (that would indicate an increasing similarity over time between the countries of our study).

Our main results are that the euro area countries seem to have become more homogeneous since the introduction of the euro, although the results are not unequivocal and considerable differences remain, leaving scope for further integration in the coming years. The Western and Central European countries like Germany, France, Belgium, and to some extent also Austria, Italy and the Netherlands, tend to cluster together. Likewise, Spain and Portugal usually form a separate cluster. In the beginning of the sample period (c. 1999-2001), Ireland and Greece tend to form distinct clusters of their own, but over time become more closely related to the other clusters (Ireland converges towards the Western and Central European one and Greece towards the Spanish-Portuguese one). Finally, the Finnish financial system seems to show most similarities with the Dutch one, but also (perhaps somewhat surprisingly) shows some relation to the Spanish-Portuguese one.

The paper is structured as follows: Section 2 describes the general methodologies applied in deriving the clusters. In Section 3, the data and underlying theoretical foundation for the choice of variables are explained. The results are presented in Section 4 and Section 5 concludes and outlines areas for further research. The detailed results are presented in the Appendix.

2. Methodology

The objective of cluster analysis is to search in the data for groups of countries in which countries belonging to that group would have their attributes closer to each other, but that at the same time would differ from countries belonging to the other groups. This would allow the researcher to classify the data in different groups so that each one would contain countries with similar economic typologies.[5] The researcher would then have a better and more accurate description of the observations with a minimal loss of information. Cluster analysis imposes no a priori restrictions on the structure of the data and requires no assumptions about the probabilistic nature (or independence) of the observations. However, the application of cluster analysis involves some limitations. It may be difficult to determine (1) the correct number of clusters, and (2) whether the clusters formed from the data significantly represent different groupings or randomly occurring concentrations of observations within an original distribution (see Korobow and Stuhr, 1991). Hence, although cluster analysis is very useful to describe the data, it can be merely characterized as a statistical exploratory technique (see Hair et al., 1998; for cluster analysis caveats).

At the same time, by using cluster analysis in different time periods it is feasible to analyze how the different countries evolve over time. Our objective is precisely to detect whether all countries remain stable over time or whether they evolve with a particular trend or characteristic. As mentioned in the introduction, we would expect some groups of countries to remain stable, but also a reduction in the distance between the different groups would be desired, because it would imply that over time more countries have the same characteristics. This could be interpreted as a gradually more homogenous and integrated banking sector in the euro area.

2.1. Selecting the appropriate clustering technique

When doing a cluster analysis it is important to know: (1) how the variables have to be selected, (2) which type of distance or similarity measure is the most appropriate, and 3) which kind of clustering method to employ. The selection of the variables has been done taking into regard theoretical and conceptual considerations related to the structure of the euro area banking sector (see Section 3). Besides, each variable has been standardized using its own maximum and minimum value over all the periods, by applying the formula: I-Imin/|Imax- Imin|. Without the standardization those variables with a larger scale would have had a greater impact in each cluster than other

[5] In geometrical terms, the cluster analysis techniques describe the objects (i.e. the countries) as points in a m-dimensional space, with each of the m-variables represented by one of the axes of the space. In the words of Dillon and Goldstein (1984), "…a [m]-dimensional space is now defined in the space by the values of the variables for each object. We can describe the clusters as continuous regions appearing in the space having relatively large mass, that is, a high density of points, which are separated from other regions by regions having relatively little mass…"

variables and hence would have dominated and potentially biased the results. The formula I-Imin/|Imax- Imin| is, in this case, a more robust measure than the normal standardization method (observations minus the mean, divided by standard deviation) because its denominator is more sensible to observations far away from the centre. As far as the second step about the different types of distances is concerned, the most typical and well-known distances that might be used are the Euclidean and squared Euclidean distance, the Manhattan or city block distance, the Mahalanobis distance or the Chebychev distance, among others.[6] The final choice among them depends on the data and the type of variables collected. The standardization methodology already defined in the previous step is shown to be more robust and appropriate and could somehow discard the use of the Mahalanobis distance, since it would mean standardizing again through the classical method of standardization. Moreover, the variables finally used are relatively weakly correlated, once standardized. In fact, the variables have been selected to avoid undue multicollinearity. Thus certain potential variables that showed persistently high correlation coefficients over time have been regrouped or excluded from the study. The lack of correlation between the variables would be a good reason for using the Euclidean or squared Euclidean distance (see Everitt, 1993). Furthermore, squared Euclidean measurements place greater emphasis on outliers to generate distance patterns. For that reason in particular, we decided to use the squared Euclidean measurement in this study, since we presume that the grouping of countries should be based on a great deal of similarity across all variables and that distinctions should be formed on the basis of outliers.[7]

Finally, a cluster analysis algorithm has to be chosen. Clustering algorithm techniques are mostly divided into two main groups: partitioning techniques and hierarchical

[6] Some measures of distance are special cases of the Minkowski metric defined by $d_{ij} = (\sum_{k=1}^{P} \left| X_{ik} - X_{jk} \right|^r)^{1/r}$ where d_{ij} denotes the distance between two objects i and j. If we set r=2, then we have the familiar Euclidean distance between objects i and j. If we set r=1, then we have what is referred to as the absolute, Manhattan or city-bloc metric. Another legitimate distance measure is the Mahalanobis distance given by $(X_i - X_j)^1 S^{-1} (X_i - X_j)$ where S is the pooled within-group covariance matrix and X_i and X_j are the respective vectors of measurements on objects i and j. This distance measure has the advantage of explicitly accounting for any correlations that might exist between the variables. Finally, the Chebychev distance is defined as $d_{ij} = \max(\left| X_{ik} - X_{jk} \right|)$.

[7] Wolfson et al. (2004) in a study of similar nature argue that the "Squared Euclidean measurement places greater emphasis on outliers to generate distance patterns. Since it was believed that grouping of countries should be based on a great deal of similarity across all variables and that distinctions should be formed based on outliers, it was decided to use Squared Euclidean measurement in this study".

techniques. The partitioning techniques usually assume a certain number of final clusters in advance, while the hierarchical techniques do not have any a priori assumption on the final number of clusters. The latter are basically characterised by the fact that once an object joins a cluster it is never removed nor fused with other objects belonging to some other clusters.

The hierarchical techniques are again divided into two main methods: the agglomerative methods and the divisive methods. The output from both methods may be represented by a two-dimensional treelike diagram known as a dendrogram which illustrates the fusions or partitions made at each successive stage of the analysis. The dendrogram also shows the distance between the clusters, once they have been fused. We chose to apply hierarchical techniques, since the number of final clusters was unknown, and the agglomerative methods were preferred to the divisive ones because they are widely implemented in software. Agglomerative methods start by placing each country in its own cluster. At the next level, or step, the two closest countries are fused into a cluster by the linkage method previously selected. At the third level, either a new object is added to the cluster or another two-country cluster is formed. The process continues until all the countries are agglomerated into a single cluster. We have calculated the final clusters using the most common agglomerative algorithms, such as the single linkage and the average linkage techniques.[8] In order to capture the underlying structural characteristics of the data and their development over time, and to reduce the impact of temporary factors a recently developed smoothing method has also been applied to complement the classical cluster analysis (i.e. without smoothing). We describe below in more detail the classical hierarchical and the smooth cluster methods over a fixed time period as well as the selected agglomerative algorithms.

2.2. The classical hierarchical cluster method

The *classical hierarchical cluster method* over a fixed J time-period considers an ordered paired list $\{t_j, W_j; j = 1,........, J\}$, t_j being the different time periods and W_j being m row-matrices of the observed variables for the m individuals in each

[8] We also derived clusters using the "complete linkage" approach. The results according to this approach were somewhat more volatile, though qualitatively similar to the other two techniques.

t_j period. In our case, the t_j periods are the different quarters and m represents the 11 euro area countries (excluding Luxembourg[9]). A description of the selected variables in W_j is presented in Section 3. In each t_j time-period the hierarchical cluster method is applied to the W_j variable matrix. From each W_j matrix we obtain a D_j squared $m \times m$ distance matrix representing the dissimilarity or distance between each pair of individuals or objects based on the squared Euclidean distance previously selected. For a particular t_j the initial D_j matrix is a symmetric matrix represented as

$$D_j = \begin{pmatrix} 0 & d_{12} & \cdots & \cdots & d_{1i} & \cdots & \cdots & d_{1m} \\ d_{21} & \ddots & \ddots & & & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & d_{ji} & & & \vdots \\ \vdots & & \ddots & 0 & \ddots & & & \vdots \\ d_{j1} & & d_{ij} & \ddots & 0 & \ddots & & \vdots \\ \vdots & & & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & & & \ddots & \ddots & d_{m-1m} \\ d_{m1} & \cdots & \cdots & \cdots & \cdots & \cdots & d_{mm-1} & 0 \end{pmatrix},$$

where d_{ij} represents the distance between the individuals i and j. From this D_j matrix we obtain the dendrogram treelike diagram based on the agglomerative algorithms. In order to obtain the final dendrogram, different linkage methods have been described in the literature. The most common ones are: the single linkage method, the complete linkage method and the average linkage method. For the *single linkage* method in the first step we fused into one cluster the two closest single individuals of the D_j matrix. In the next step, we defined a new distance matrix including the new two-cluster and derive a new minimum distance between the clusters. Hence, either a third cluster joins the first two-cluster to form a three-cluster or a new two-cluster is formed. The process finishes when all the objects are fused into a single cluster containing all the m initial objects. The *complete linkage* method follows the opposite approach. It considers in each step the maximum distance between the clusters instead of the minimum distance. This is why these two methods are also known as the “nearest neighbour” and the “furthest neighbour”, respectively. Finally the *average linkage* method calculates in each step the average distance

[9] Owing to an only limited set of data for this country.

between pairs of clusters (see Dillon et al. 1984 for numerical examples of these three algorithms). In each step the fusion between countries is represented in the dendrogram that has been previously defined. Other linkage methods like the Ward's, the median or the centroid method have also been described in the literature. But the first three methods since they are more common and broadly known in most of the statistical packages were used to obtain the final dendrograms. Overall, the single linkage method and the average method led to the most consistent and stable results. We have therefore based our discussion on the two former methods, and the dendrograms for each time period showing the cluster-relation between the different countries are presented in a condensed form in the appendix. The statistical package S-Plus 6.2 was applied to carry out the calculations.

2.3. The smoothed hierarchical cluster method

In the standard cluster method, the clusters are derived for each time period without taking into regard the clustering in preceding periods. This may, in some cases, result in some volatility in the clustering over time. In order to mitigate some of this volatility, which for example may be driven by certain one-off events (outliers), we use a smoothed cluster method to complement the results obtained in the classic cluster method.

The smoothed cluster method was introduced by Esteve et al. (2004) and can be inscribed within the framework of the distance-based prediction methods introduced by Cuadras (1989). It is based on the distances between individuals over different periods. The aim of the smoothing method is not solely the detection of outliers but also the linking of isolated and fragmented descriptions or snapshots of the same reality. As a consequence, the smoothing method also takes into account the possible misclassification of an individual into a cluster in a certain period by observing a distance matrix between the individuals over all the periods. Like in the classic cluster method, it also considers an ordered paired list $\{t_j, W_j; j = 1,........, J\}$, being t_j the different time periods, and W_j m-row matrices, with the observed variables for the m individuals for the periods t_j. Using a moving time window of width $\pm k$, in each interval centered at j_0, $I(j_0) = \{t_j, j = j_0 - k,......., j_0 + k\}$ is the projector method, described in Esteve *et al.* (2004). Using the $I(j_0)$ projector, we attempt to incorporate into each period t the way the Euclidean configuration of t+1, t-1 etc. is seen from t. Thus, for each period *t* we have the following set, {configuration of t seen from t+j;j= -k,....,0,........k}, that can be averaged.

From the initial paired list, $\{t_j, W_j; j = 1,........, J\}$, we obtain $G_{ij} = W_i \times W_j'$ and compute:

$$D_{i,j} = g_i \times 1' + 1 \times g_j - 2 \times G_{i,j} \quad \forall i, j. \quad (1),$$

where $D_{i,j}$ represents the inter-distance matrix from one period to another, g_i represents the $m \times 1$ column vector where each element is the squared element of Wi, i.e. $g_i = \|w_i\|^2$, and 1 is the $m \times 1$ column vector of ones.

In particular, the elements of each of the distance matrices are calculated as follows:

For each pair s, t:

$$D_{1,1}(s,t) = \|w_1(s) - w_1(t)\|^2 = w_1(s)w_1(s)' + w_1(t)w_1(t)' - 2w_1(s)w_1(t)'$$
$$D_{1,2}(s,t) = \|w_1(s) - w_2(t)\|^2 = w_1(s)w_1(s)' + w_2(t)w_2(t)' - 2w_1(s)w_2(t)' =$$

$G_{11}(s,s) + G_{22}(t,t) - 2 \times G_{12}(s,t)$ and so on.

The notation in (1) is inspired by the one used in the Multidimensional Scaling problem. Since entries in a distance matrix must be positive and subjected to triangular inequalities, they cannot be directly smoothed. Hence, the need for a detour through their Euclidean configurations is required.

From each of the $D_{i,j}$ we obtain *D*, an inter-distance matrix between all periods. *D* is an $m \cdot J \times m \cdot J$ matrix, or a $J \times J$ matrix of $m \times m$ blocks of inter-distance matrices defined as above. In our case, since the number of periods is not very big, we have used a window of width *k* equal to the number of periods[10].

[10] There is not a mandatory value for the time window length. We have also observed that small widths values, i.e. for *k* equal to 2, 3 and 4, have produced unstable results as expected. In general, one could choose different window lengths depending on the total number of periods. It might be argued that the use of a time window length equal to the number of periods could lead to a parsimonious solution.

It is sufficient to fix $k \le J$ and to consider D as a $(2k+1)$ - diagonal block matrix:

$$D = \begin{pmatrix} D_{1,1} & D_{1,2} & \cdots & \cdots & D_{1,k} & * & \cdots & * \\ D_{2,1} & D_{2,2} & \ddots & & D_{2,k} & D_{2,k+1} & * & \vdots \\ \vdots & \ddots & \ddots & \ddots & & & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & & & \vdots \\ D_{k,1} & D_{k,2} & \cdots & \cdots & D_{k,k} & \ddots & & \vdots \\ * & D_{k+1,2} & & & \ddots & \ddots & \ddots & \vdots \\ * & * & D_{k+2,3} & & & \ddots & \ddots & \\ & \cdots & * & \cdots & \cdots & \cdots & & \end{pmatrix}$$

The * represent those distance matrices that do not contribute any new information because the time lag is too big.

From the distance-matrix D we define the matrix $G_0 = -\frac{1}{2}\tilde{J} * D * \tilde{J}$, where $\tilde{J}$ is the centering matrix of dimension $m \times J$ defined as $\tilde{J} = Id_{m \times J} - \frac{1}{m \times J} Id_{m \times J}$. G_0 is the inner product matrix, the symmetric matrix solution to the exact Euclidean representation of the distance matrix D.

We then project the matrix G_0:

$$\hat{G} = P * G_0 * P \text{ and } \tilde{G} = Q * G_0 * Q, \qquad (2)$$

where P and Q are the projectors defined as $P = diag(K_1, \ldots\ldots, K_{2k+1})$ of dimension $m \times J$, $Q = Id - P$ of dimension $m \times J$ and $K_i = \frac{1}{m_i} 1 \times 1'$ of dimension $m \times m$.

The matrices $\hat{G}$ and $\tilde{G}$ described in (2) by the projection of P and Q are the so-called inner product matrix "between", $\hat{G}$, and the "within" matrix, $\tilde{G}$, of dimension $((2k+1) \times m) \times ((2k+1) \times m)$. The eigenvectors of the matrices $\hat{G}$ and $\tilde{G}$ constitute an orthogonal base of vectors that can be easily computed and used to obtain the Euclidean "between" and "within" coordinates.

The “between” and “within¨ matrices are based on the following underlying idea:

For each period, a moving window $j_0 = -k,.....,1,......k$ is constructed. The same individual j_0 is then observed for the different periods using the moving window and hence 2k + 1 groups are formed. The centroids of each of these groups constitute the “between” matrix, and the “within” matrix is based on the distances within each of these groups to the centroid. From $\hat{G}(j_0)$ and $\tilde{G}(j_0)$ the respective Euclidean configurations $\hat{X}(t)$ and $\tilde{X}(t)$ can also be obtained for $t \in [t_1, t_J]$.

From the matrix $\tilde{G}$, a smoothed distance matrix $\tilde{D}$ for each period is obtained. The procedure to obtain the smoothed distance matrix is based again on multidimensional scaling metrics.

The matrix $\tilde{D}$ is for each period an $m*m$ matrix defined as follows:

$$\tilde{D} = \tilde{g}_i \times 1' + 1 \times \tilde{g}_j - 2 \times \tilde{G}_{i,j}$$

where $\tilde{G}_{i,j} = u \times \Lambda \times u'$, Λ and u are the eigenvalues and eigenvectors of $\tilde{G}$, the inner product matrix “within”, $\tilde{g}_i$, represents the $m \times 1$ column vector where each elements is the squared element of $u \times \Lambda^{1/2}$, i.e. $g_i = \left\| u_i \times \lambda_i^{1/2} \right\|^2$, and 1 is the $m \times 1$ column vector of ones.

The final matrix $\tilde{D}$ is a smoothed distance matrix that takes into account the different periods. From this smoothed distance matrix $\tilde{D}$ the dendrogram for each time period is calculated. In this case, the dendrogram would take into account the possible outliers of an individual or country in a particular period.

In our case, we have calculated the matrices $\tilde{G}(j_0)$, $\hat{G}(j_0)$ and $\hat{D}$ using a MatLab function developed by Esteve and Fortiana, and from the matrix $\tilde{D}$, the final dendrograms using S-Plus 6.2, as before. Further description of the methodology can be found in Esteve *et al.* (2004).

The results that we have obtained for the two methods are described in turn below in Section 4.

3. Data and selection of variables

The focus of the analysis is to examine the degree of homogeneity among the banking sectors of the euro area countries using a number of harmonised banking indicators and some more cyclical indicators that may be expected to have a direct, or indirect, impact on bank behaviour. The basic idea is that by applying cluster analysis techniques on a wide variety of price and quantity variables related to the banking sector and including a range of potentially relevant macro-variables we may be able to detect some patterns and developments in the structure of the banking sector across the euro area countries. In this light, increasingly larger and more closely tied clusters would, all things being equal, indicate that the euro area banking sector is becoming more integrated.

As already mentioned, cluster analysis implies that no restrictions or stipulated structures are imposed upon the data *ex ante*. The selection of variables to be included in the cluster analysis is therefore highly important, since it is the data themselves that structure the results. Leaving out or adding an important variable may hence alter the results significantly.

The variables we use have been selected with the aim of capturing to the extent possible the behaviour and structure of the banking sector in the euro area countries taking into account i) factors affecting the supply of and demand for credit/deposits ("cyclical indicators"), ii) factors expressing the prevailing banking structures ("structural indicators") and iii) factors expressing banks' pricing behaviour ("price indicators", see Table 1). The three groups of factors are interrelated as, for example, banks' prices (here measured as bank margins) are determined by both structural factors (such as the degree of competition, bank capital and liquidity levels, credit and interest rate risks, etc.) and by more cyclical factors affecting the demand for credit and supply of deposits (such as GDP growth, consumer confidence, house price developments, etc.).[11] Likewise, banks' prices are likely to influence the demand for credit and supply of deposits, while structural factors (e.g. banks' degree of risk

[11] On the literature concerning bank margin determinants see e.g. Ho and Saunders (1981), McShane and Sharpe (1985), Allen (1988), Angbazo (1997), Wong (1997), Saunders and Schumacher (2000) and Maudos and de Guevera (2004).

aversion, capital ratios and liquidity risk) are likely to have an impact on bank supply of credit and demand for deposits.

By selecting this confluence of variables we aim to capture a considerable part of the factors determining the behaviour and development of the banking sector in the euro area without imposing any causality links or structures a priori. As mentioned above, in the selection process we have strived to exclude variables that were too highly correlated with other variables and that could therefore have biased (or created too much noise in) the results. Naturally, it cannot be excluded that we have included some irrelevant variables in the sample or have left out some potentially relevant variables.

For example, it might be argued that the cyclical variables have a substantial impact on the results and hence that the observed clustering may to some extent reflect cyclical variations/similarities rather than structural developments in the banking sector. A preliminary analysis, comparing the averages coefficient of variation across countries of the structural and cyclical indicators, respectively, shows that the dispersion is less pronounced with respect to the cyclical measures. At the same time, however, it is notable that the average coefficient of variations decline (by 13-14%) over the sample period with regard to both the cyclical and the structural indicators. This, at least, suggests that our finding of some convergence among the euro area countries in the sample period most likely is due to both types of measures.

Table 1 List of variables

Description	**Definition**	**Source**
Price indicators		
Overall bank margin	Difference between weighted averages of loan and deposit rates	ECB
10-year government bond yield		ECB
Term structure	Difference between 10-year government bond yield and 3-month Euribor	ECB
Structural indicators		
Liquidity risk (bank excess funds)	Cash holdings plus securities holdings to total assets	ECB
Degree of risk aversion	Capital and reserves to total assets	ECB
Credit risk vis-à-vis non-financial corporations	Loans to non-financial corporations to total assets	ECB

Credit risk vis-à-vis households	Loans to households to total assets	ECB
Pool of bank deposits	Ratio of interbank deposits and debt securities issued by MFIs to non-bank deposits	ECB
Average size of banks	Total assets to number of MFIs (banks)	ECB
Debt securities issued by MFIs	Annual percentage change	ECB
Herfindahl index of bank concentration	Sum of squared market shares to total assets in MFI population	ECB
Cyclical indicators		
Real GDP growth	Annual percentage change	Eurostat
Real private consumption	Annual percentage change	Eurostat
Real gross fixed capital formation	Annual percentage change	Eurostat
MFI loans to households	Annual percentage change	ECB
MFI loans to non-financial corporations	Annual percentage change	ECB
Industrial confidence indicator	Percentages of the balances	EC
Consumer confidence indicator	Percentages of the balances	EC
Index of house prices	Index	National Sources

The data have been collected for 11 euro area countries (excluding Luxembourg) on a quarterly basis for the period Q3 1998 to Q2 2004 (i.e. 24 quarters). The series have been standardized (as described in Section 2). For some countries there are missing variables with respect to some of the series (i.e. we are working with an unbalanced panel).

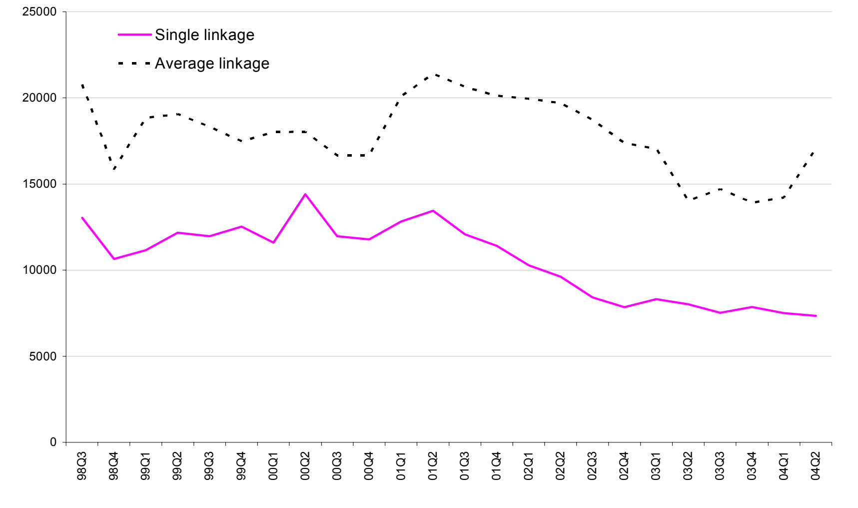
4. Results

4.1 Clusters using the standard method

Maximum height

As regards the maximum distance ("height") between the clusters we observe that it overall was reduced between Q3 1998 and Q2 2004 for both the "Single linkage" approach (-44%) and the "Average linkage" approach (-17%), see Chart 1. However, some variations over the sample period are observed. Thus, the maximum height peaked in Q2 2001 with respect to the "Average linkage" approach and declined afterwards until late 2003. In 2004, some reversal of the trend was observed. The maximum height as derived from the "Single linkage" approach shows a somewhat different pattern: peaking Q2 2000 and declining more or less steadily afterwards.

Chart 1. Maximum height; standard method



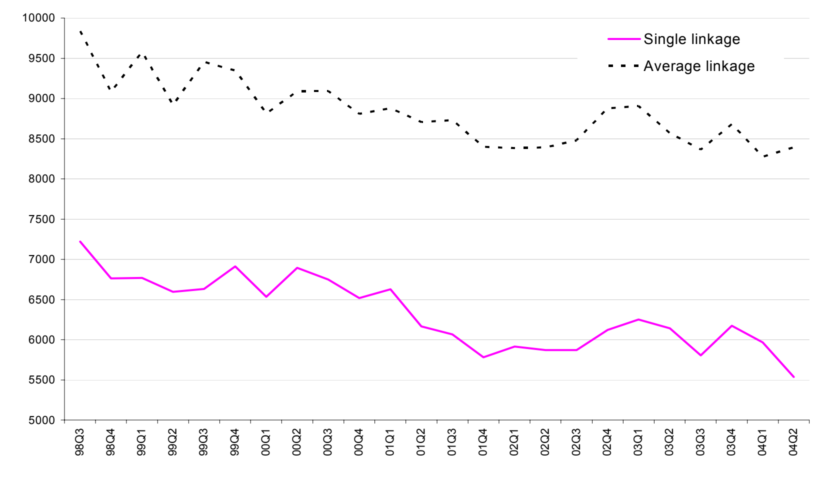
Source: Own calculations.

Average height

Another measure of the distance between the clusters is the average height of all the derived clusters, which should not only capture the distance to the most extreme cluster but also the degree of closeness between the intermediate clusters. Chart 2 illustrates that measured by the average height of the clusters there was an overall decline over the sample period with respect to both approaches:

"Single linkage" (-23%) and "Average linkage" (-15%). This suggests that the countries tended to become more homogeneous over time.

Chart 2. Average height; standard method



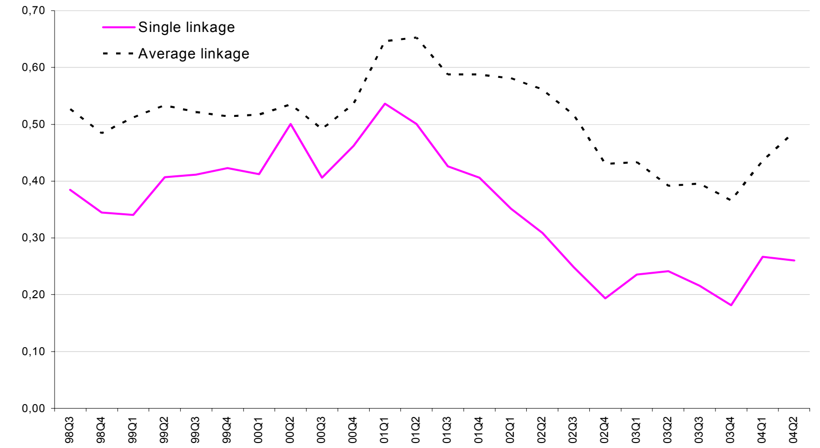
Source: Own calculations.

Coefficient of variation

Yet another measure of the distance between the different clusters is the coefficient of variation (measured as the standard deviation of the "height" divided by its mean), which to a greater extent takes into account the overall distribution of the clusters – both taking into account the variation and the mean. A smaller coefficient of variation would tend to imply a lower variation around the mean and hence a more homogeneous population.

Chart 3 illustrates that overall in the sample period the coefficient of variation declined with respect to both the "Single linkage" (-32%) and "Average linkage" (-8%) approaches. It reached its lowest level during 2003 and reversed to some extent during 2004.

Chart 3. Coefficient of variation; standard method *(standard deviation divided by mean height)*



Source: Own calculations.

Clustering of countries

As illustrated in Tables A1 and A2 in the Appendix, the clustering of countries is broadly similar across the two different approaches, although the clusters formed in the "Single linkage" approach appear to be comparatively less stable.[12]

[12] The clustering dendrograms describing the detailed results in each quarter are available upon request from the authors. The Appendix tables (Tables A1-A4) provide the clustering results in a condensed manner. The upper matrix in each table shows the bilateral distances between each pair of countries averaged over the whole sample. For each country the average of the bilateral distances are then calculated and compared with the overall average distance. A value below zero indicates that the relevant country overall tends to belong to clusters which are more homogenous than the average. The middle two matrices show the bilateral distances from the average height of each country, and thus

Focusing on the "Average linkage" approach, which results in the most distinct clusters, the following patterns are observed:

- Austria and Italy are almost always in the same and most homogeneous cluster, which in Table A2 is reflected in a distance in the average height of clusters involving the two countries which is 68% below the overall average height (vis-à-vis all other countries) of each country. In the beginning and towards the end of the sample period the French (-12%, on average) and the German (-6%) clusters tend to be closely related to the Austrian-Italian cluster, while in the middle of the sample period the latter is more closely related to the Spanish (-3-5%) and the Portuguese (+1-2%) clusters;
- Spain and Portugal are almost always in the same cluster in the period Q3 1998-Q2 2002, and in mostly in neighbouring clusters in the subsequent period. This is illustrated by a distance in the average height of clusters involving the two countries is 37-39% below the overall average height (vis-à-vis all other countries) of each of the two countries;
- Belgium and France are always in the same cluster from Q3 1998-Q2 2000. In the subsequent period, France and Germany are almost always (one exception) in the same cluster. Prior to 2001, Ireland tends to be a relatively long distance from the main clusters. From 2001 onwards, Belgium and Ireland are almost always in the same cluster or close to each other, and both countries are also close to the French-German cluster. Thus, in terms of figures the average height of clusters involving France and Germany is 33-35% below the overall averages, while French-Belgian clusters have an average height 21-25% below the overall averages;
- Finland and the Netherlands are almost always (two exceptions) in the same cluster, and both countries are usually also close to the Spanish-Portuguese cluster. Thus, clusters involving Finland and the Netherlands have an average height 44% below the overall average of each of the two countries. Similarly, clusters involving Finland and either Portugal or Spain have an average height 26% respectively 27% below the overall average of Finland, while clusters

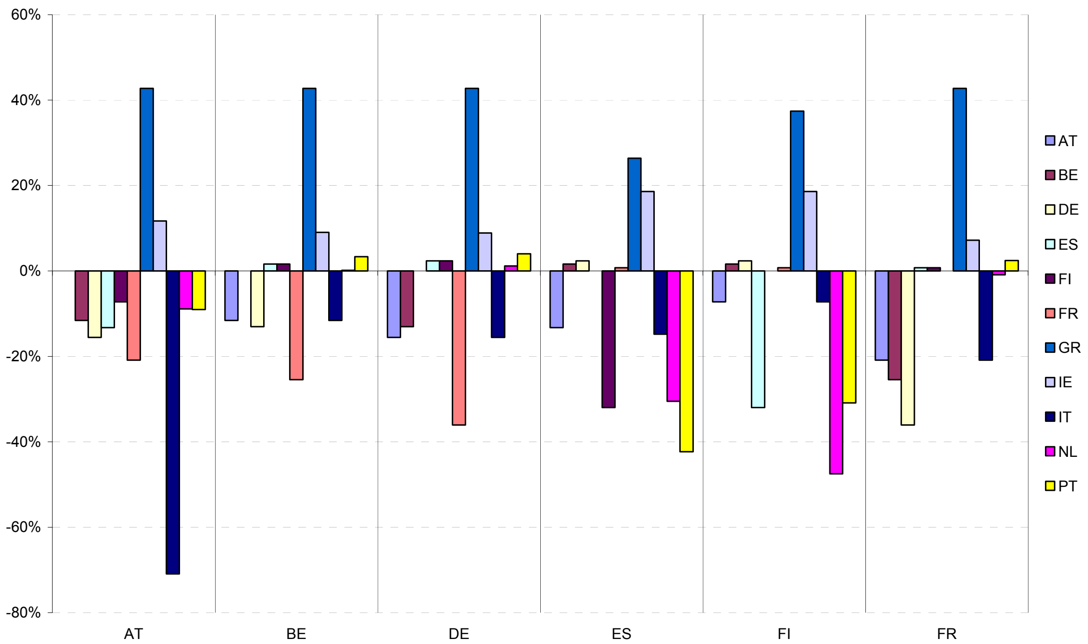
describes the countries with which a particular country tends to cluster (indicated by a high negative value), while the lower two matrices display in a similar fashion the bilateral distances compared to the overall average height of the whole sample.

involving the Netherlands and either Portugal or Spain have an average height of 26% and 25%, respectively, below the overall average of the Netherlands;

- Until early 2003, Greece also tended to be a relatively long distance from the main clusters, but by Q2 2003 onwards always formed a cluster with Spain.

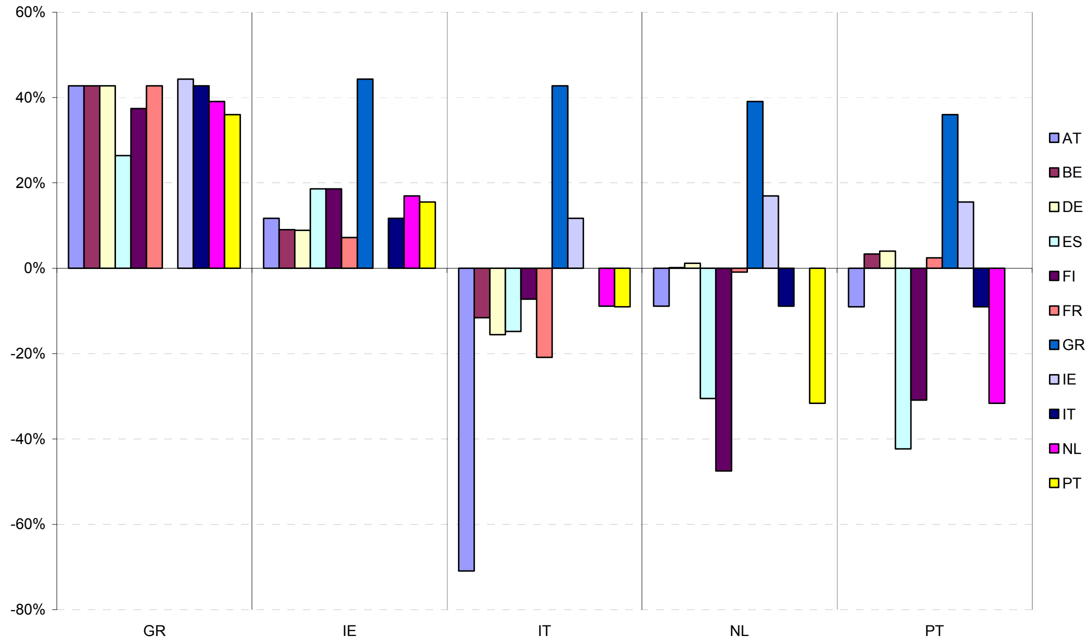
Charts 4a-b illustrate these findings by displaying the relative height of clusters involving two countries as a percentage to overall average height of the sample. It is thus clearly illustrated that on average over the sample period Austria and Italy form the most homogenous clusters (compared to the overall average of between-cluster heights). At the other extreme, Ireland and Greece, in particular, have been far from the euro area core for most of the sample period.

Chart 4a. Country-by-country distances to overall average height of sample; according to the standard method (average linkage approach)



Note: A large negative value indicates that the two countries concerned are relatively homogenous – and vice versa.

Chart 4b. Country-by-country distances to overall average height of sample; according to the standard method (average linkage approach)



Note: A large negative value indicates that the two countries concerned are relatively homogenous – and vice versa.

4.2 Clusters using the smoothing method

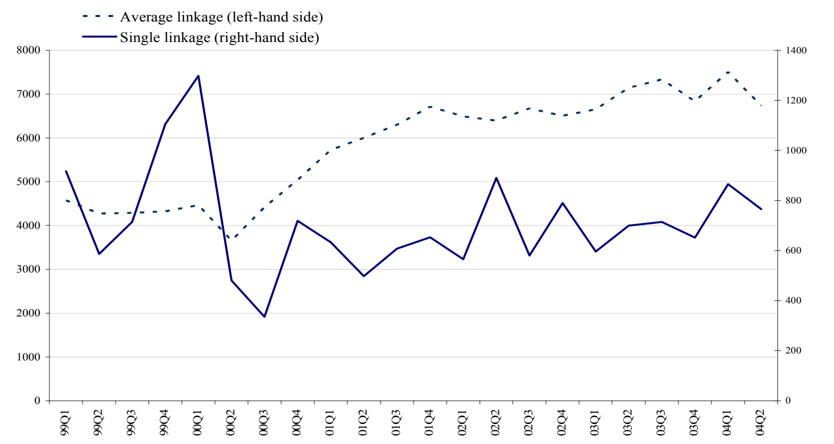
Using the smoothing method described in Section 2.3, we obtain the following results:

Maximum height

Applying the smoothing method produces somewhat different results compared with the standard method. The possible differences that might arise between both methods, as explained in Section 2.3, relate in particular to the smoothing out of temporary deviations from the general clustering patterns. As a result, the dispersion of the average distances between countries/clusters is significantly lower under the smoothing method. In our analysis, the maximum height using the smoothing method follows a different pattern than under the standard method (Chart 5). In particular, according to the "average linkage" approach we find that the maximum distance reaches a peak towards the end of the period and overall between Q1 1999[13] and Q2 2004 increased by 47%. By contrast, the maximum height according to the "single linkage" approach declined overall by (-12%) although it reached its lowest point already in Q3 2000.

As expected, the maximum height measures derived using the smoothing method are significantly lower compared with the standard method owing to the mitigated effect of temporary deviations from the clustering over the longer-run.[14]

Chart 5. Maximum height; smoothing method



Source: Own calculations.

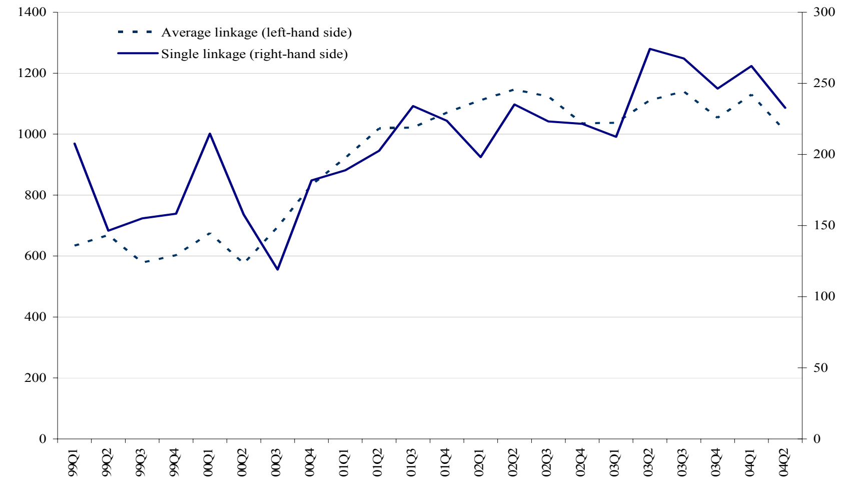
Average height

With respect to the average height measure we observe an overall increase in the sample period for both approaches (Chart 6) – i.e. contrary to the standard method. Thus, the average height using the "single linkage" approach trended upwards during most of the time and overall increased by 12% between Q1 1999 and Q2 2004. Likewise, the average height measure using the "average linkage" approach increased by almost 60% in the same period, although since early 2001 it has been relatively stable. As was the case for the maximum height measure, the average height measures using the smoothing method are significantly lower compared to the standard method.

[13] The starting period for the smoothing method is Q1 1999 as data for Greece were not available prior to this date.

[14] Since, under the smoothing method, the whole sample (k=J) is taken into account when deriving the clusters, this may also partly explain why we do not experience the same degree of reduction in the distances over time. The reason is that the information of clustering in the first part of the sample is taken into account when deriving the clusters in the latter part of the sample – and vice versa.

Chart 6. Average height *(smoothing method)*



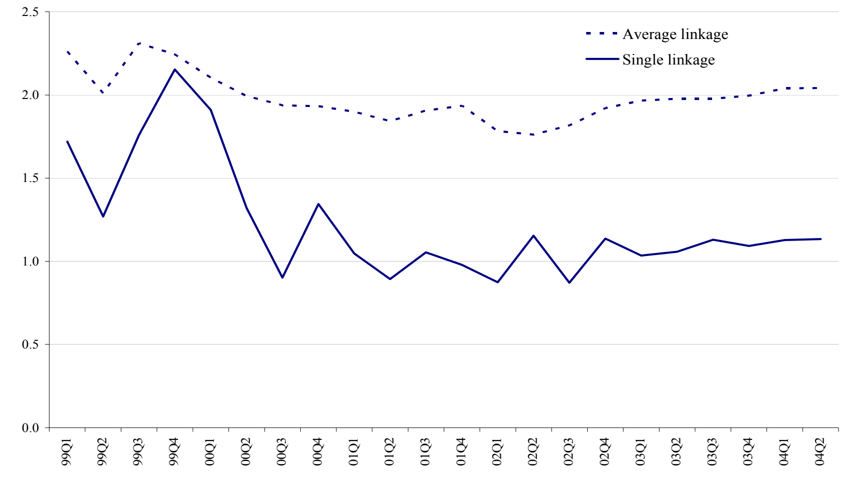
Source: Own calculations.

Coefficient of variation

According to the "average height" indicator, the clusters derived by applying the smoothing method do not seem to have become more homogeneous over time. However, taking into account the variation around the mean, as reflected by the coefficient of variation, we observe on the contrary that the distances between the euro area countries have been reduced (Chart 7) – as was also the case for the standard method. This overall reduction applies to both approaches: -34% for the "single linkage" approach and -10% for the "average linkage" approach. The dynamics differed somewhat between the two approaches, but both stabilised at a lower level around early 2001 onwards.

The size of the coefficients of variation using the smoothing method is considerably larger than under the standard method, which is explained by the generally lower average height derived using the former method.

Chart 7. Coefficient of variation *(smoothing method)*



Source: Own calculations.

Clustering of countries

As illustrated in Tables A3-A4 in the Appendix, the clustering of countries using the smoothing method is broadly similar for the "single linkage" and "average linkage" approaches, and with some exceptions also comparable to the results found using the standard clustering method.

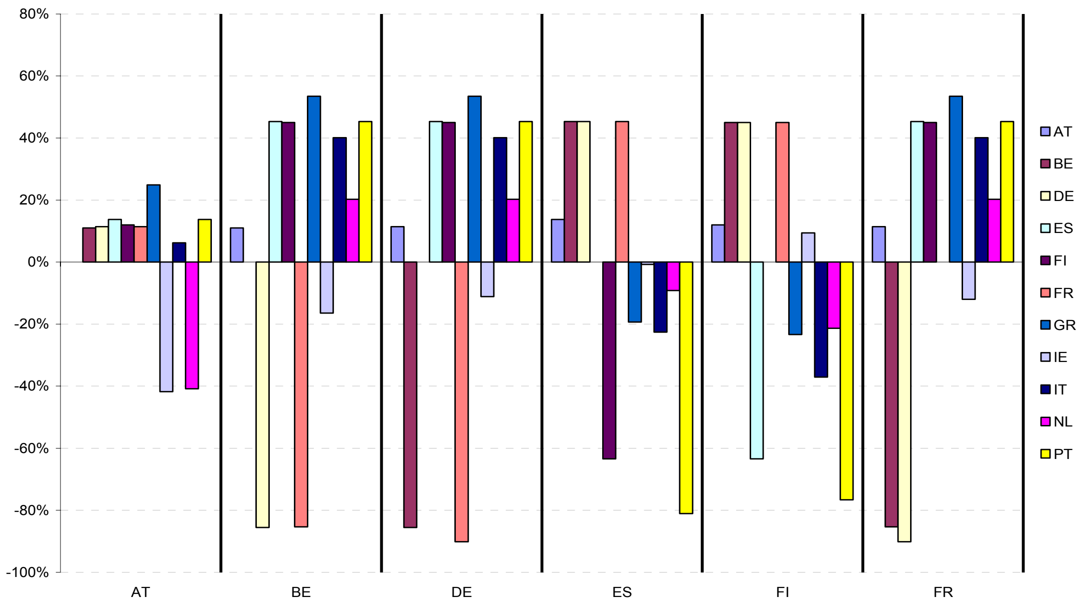
Focusing on the "Average linkage" approach the following patterns are observed:

- Germany, Belgium and France are almost always in the same cluster (its average distance being more than 90% below the country averages), and most of the time this cluster is related to clusters of Ireland and Austria, especially in the latter part of the sample.
- Spain and Portugal are in the same homogenous cluster (or at least neighbouring cluster) throughout the sample. Thus the average height of clusters involving the two countries is more than 90% below their respective country averages;
- The Netherlands and Italy are often in the same clusters (thus, the average height of clusters involving the two countries is 46-47% below the averages of each country). In addition, the Netherlands often clusters with Austria and Ireland belonging to the French-German cluster, while Italy more often clusters with the Finnish-Spanish-Portuguese group of countries;

- In the first part of the sample period Finland mostly seem to cluster with the Netherlands and Italy, but in the latter part of the period (Q1 2001 onwards) tend to be in the same cluster as Portugal and Spain (and Greece);
- Greece is often in a cluster of its own and for most of the sample period it is the most distant cluster. Towards the end of the sample period, Greece seems to become more closely related with the Spanish cluster;

These findings are illustrated in Charts 8a-b.[15] One notable difference from the results of the standard method is that Greece and Ireland to a lesser degree form clusters that are highly distinct from the other clusters. The intuition behind this result is that the k-period projector applied in the smoothing method makes use of information over the entire sample period. Thus, under the smoothing method, the fact that Greece and Ireland in the latter part of the sample converge somewhat to the more homogenous clusters is taken into account also when forming the clusters in the earlier part of the sample, and hence overall the two countries are not found to be distinct to the same extent as under the standard method (as illustrated in Charts 4b and 8b).

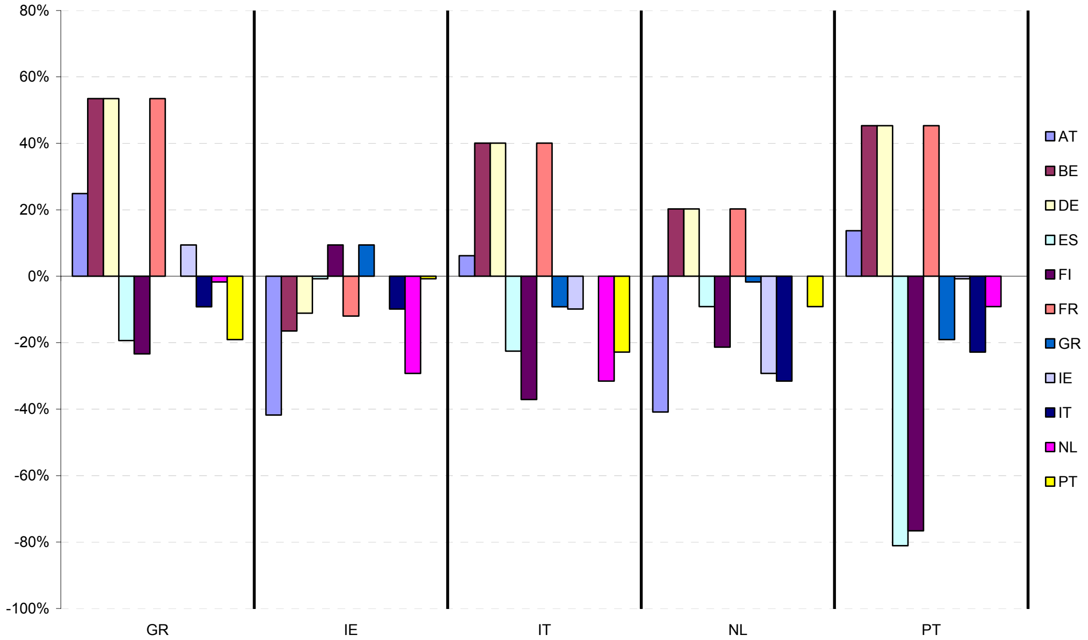
Chart 8a. Country-by-country distances to overall average height of sample; according to the smoothing method (single linkage approach)



Note: A large negative value indicates that the two countries concerned are relatively homogenous – and vice versa.

[15] In Charts 8a-b the ’ single linkage’ approach has been illustrated (instead of the ’average linkage’ approach). The reason being simply that the results according to this approach provide a clearer illustration of the findings (which are qualitatively comparable across the two approaches).

Chart 8b. Country-by-country distances to overall average height of sample; according to the smoothing method (single linkage approach)



Note: A large negative value indicates that the two countries concerned are relatively homogenous – and vice versa.

5. Conclusion

To sum up, while there is some degree of discrepancy across the two methods (smoothing and non-smoothing) and across the two approaches ("single linkage" and "average linkage") the general picture which emerges is that the clustering of countries have become more dense over the sample period. This development may tentatively suggest that the euro area countries have become more homogenous in terms of economic and financial structures since the beginning of EMU. Yet differences remain and some distinct clusters are discernible (most notably: DE, FR and BE; AT, IT and NL; ES and PT; IE in-between the two former, FI in-between the two latter, and GR mainly related to the latter). Indeed, while some progress in terms of integration seems to have taken place among the euro area countries since the introduction of the euro, the reduction in the distances of the clusters has only been gradual and distances remain substantial pointing to further scope of integration in the future. The finding that the financial structures of the euro area countries are not yet fully harmonised is perhaps not surprising since the major part of our selected variables relate to the banking sector, which is most likely less integrated than other market segments.[16]

Compared to the non-smoothing ("classical") method the clusters formed applying the smoothing method are somewhat different as it has already been explained in section 4. The smoothing method usually results in clusters containing more countries (or having fewer distinct clusters), and are generally more stable over time (as expected) and have fewer persistent "outliers".[17] A few common characteristics emerging from both methods are worth emphasising:

- Spain and Portugal are mostly in the same cluster suggesting similar structural characteristics in the two countries, which may partly reflect the geographical and cultural proximity of the two countries and the fact that they have followed a broadly similar economic and financial development, including some cross-border bank mergers, since joining the EU in 1986;
- Greece and Ireland are both in a cluster of their own, respectively, during some part of the sample period: Ireland is mainly distinctive in the beginning and subsequently seems to get closer to the structures of the other countries, notably Belgium (Germany and France). Greece remains a cluster of its own throughout the major part of the sample period and only seems to significantly converge to other euro area countries (i.e. Spain/Portugal) from 2003 onwards;
- Belgium is generally in the same cluster as France and Germany, or close to the French cluster, which may point to some structural similarities (probably reflecting geographical and structural proximity);
- Finland and the Netherlands are in the same or related clusters for most (but not all) of the sample period. This may reflect the somewhat more Anglo-Saxon type financial systems of these two countries compared to the rest of the euro area countries.[18]
- While being in the same cluster in the results based on the non-smoothing method, Austria and Italy are never in the same cluster in the results based on the smoothing method (although often in neighbouring clusters). Overall,

[16] This is confirmed by other studies, most notably Cabral et al. (2002) and Baele et al. (2004).
[17] Q4 2000 stands out from preceding as well as subsequent periods under the "average linkage" approach using the smoothing method. This "outlier" remains unexplained.
[18] See e.g. ECB (2002) Report on Financial Structures, December.

Italy mostly seems to display broadly the same structural features as Austria and to some extent also the Netherlands. In addition, these three countries often appear rather closely related to Belgium, France and Germany, which might point to some broad similarities between the Western and Central European Member States; having converged in real economic terms and being based on generally similar financial systems.

To sum up, the present study applies an alternative method to study financial and economic integration in the euro area. The results point to some (though limited) progress in the degree of integration and indicate some plausible patterns in the clustering of countries. The method, however, is silent about which factors drive the results and what their statistical significance is. This study should therefore mainly be regarded as complementary to other studies on the subject and generally seems to confirm the findings of previous studies. One extension of the study might be to investigate which factors are the most important in terms of differences and similarities. Such an analysis might be conducted employing principal component or factor analysis methods. Another alternative analysis to the above mentioned multivariate analysis could be the use of partial least squares regression considering, for example, the overall bank margin as the dependent variable.

References

Adam, K., T. Jappelli, A.M. Manichini, M. Padula and M. Pagano (2002), *Analyse, Compare and Apply Alternative Indicators and monitoring Methodologies to Measure the Evolution of Capital Market Integration in the European Union*, Report to the European Commission.

Adjaouté, K. and J.-P. Danthine (2003), "European Financial Integration and Equity Returns: A Theory-based Assessment", in Gaspar, V. et al. *The Transformation of the European Financial System*, ECB, Frankfurt.

Allen, L. (1988), "The determinants of bank interest margins: A note", The Journal of Quantitative Analysis, Vol. 23, No. 2 (June).

Angbazo, L. (1997), "Commercial bank net interest margins, default risk, interest-rate risk, and off-balance sheet banking", Journal of Banking & Finance, 21, 55-87.

Baele, L., A. Ferrando, P. Hördahl, E. Krylova and C. Monnet (2004), "Measuring Financial Integration in the Euro Area", ECB Occasional Paper Series No. 14, April 2004.

Batista-Foguet J.M., Fortiana J., Currie C., Villalbí J.R., "Socio-economic Indexes in Surveys for Comparisons between Countries", Social Indicators Research 65, 1–18.

Cabral, I., F. Dierick and J. Vesala (2002), "Banking Integration in the Euro Area" ECB Occasional Paper Series No. 6, December 2002.

Corvoisier, S. and R. Gropp (2001), "Bank Concentration and Retail Interest Rates", ECB Working Paper No. 72.

Cuadras, C.M (1989), "Distance Analysis in discrimination and classification using both continuous and categorical variables", in: Y. Dodge (ed.), Statistical Data Analysis and Inference, Amsterdam: North Holland Publishing Co., pp. 459–473.

Dillon, W., Goldstein M. (1984), "Multivariate Analysis", John Wiley and Sons.

European Central Bank (2002), Report on Financial Structures, December 2002.

Esteve A., Fortiana J., Batista-Foguet J.M. (2004), "Analysis of Stratified Data as a Tool in Designing Socio-Economic Indexes", Working Paper.

Esteve A. and Fortiana J. (2004), "Tratamiento basado en distancias para datos longitudinales", XXVIII Congreso Nacional de Estadística e Investigación Operativa. Cadiz Octubre 2004.

Everitt, Brian (1993), "Cluster analysis", John Wiley & Sons Inc, April 1993.

Fratzscher, M. (2001), "Financial Market Integration in Europe: on the effects of EMU on Stock Markets", International Journal of Finance and Economics Vol. 7 (3), pp 165-194, July 2002.

Galati, G. and K. Tsatsaronis (2001), "The Impact of the Euro on Europe's Financial Markets", BIS Working Paper No. 100.

Giannetti, M., L. Guiso, T. Jappelli and M. Pagano (2002), "Financial Market Integration, Corporate Financing and Economic Growth", European Economy, Economic Papers No. 179.

Hair, J., Anderson R. , Tatham, R., Black W. (1998), "Multivariate Data Analysis", Fifth Edition, Prentice Hall.

Hartmann, P., A. Maddaloni and S. Manganelli (2003), "The Euro Area Financial System: Structure, Integration and Policy Initiatives", Oxford Review of Economic Policy, Spring 2003, Vol. 19 (1), pp. 180-213.

Ho, T. and Saunders, A. (1981), "The Determinants of Banks Interest Margins: Theory and Empirical Evidence", Journal of Financial and Quantitative Analysis, XVI, 581-600.

Kleimeier, S. and H. Sander (2002), European Financial Market Integration, Mimeo.

Korobow, L. and Stuhr D.P. (1991), "Using cluster analysis as a tool for economic and financial analysis", Federal Reserve Bank of New York, Research Paper No. 9132.

London Economics (2002), *Quantification of the Macro-economic Impact of Integration of EU Financial Markets*, Report to the European Commission.

Manna, M. (2004), "Developing Statistical Indicators of the Integration of the Euro Area Banking System", ECB Working Paper No. 300.

Maudos, J. and J.F. de Guevara (2004), "Factors Explaining the Interest Margin in the Banking Sectors of the European Union", Journal of Banking and Finance, 28, 2259-2281.

McShane, R.W. and I.G. Sharpe (1985), "A Time Series/Cross-Section Analysis of the Determinants of Australian Trading bank Loan/Deposit Interest Margins: 1962-1981", Journal of Banking & Finance, 9, 115-136.

Saunders, A. and L. Schumacher (2000), "The Determinants of Bank Interest Rate Margins: An International Study", Journal of International Money and Finance, 19, 813-832.

Wolfson, M., Zagros M., James P. (2004), "Identifying National Types: A Cluster Analysis of Politics, Economics, and Conflict", 2004 Journal of Peace Research, vol.41 n0. 5, pp. 607-623.

Wong, K.P. (1997), "On the Determinants of Bank Interest Margins under Credit and Interest Rate Risks", Journal of Banking & Finance, 21, 251-271.

APPENDIX

Table A1. Single linkage (standard clustering method)

Average distance between clusters	AT	BE	DE	ES	FI	FR	GR	IE	IT	NL	PT
AT	0										
BE	7791	0									
DE	6713	8203	0								
ES	4768	7783	6997	0							
FI	5939	7831	7223	5930	0						
FR	6341	7637	6713	6420	6698	0					
GR	9719	10010	9660	9425	9838	9642	0				
IE	8487	9333	8582	8547	8680	8507	10168	0			
IT	3365	7783	6970	4774	5916	6340	9633	8547	0		
NL	5868	7794	7176	5916	5690	6624	9776	8666	5868	0	
PT	5747	7840	7341	5456	6169	6768	9966	8778	5747	6079	0
Average height	**6474**	**8201**	**7558**	**6602**	**6991**	**7169**	**9784**	**8830**	**6494**	**6946**	**6989**
Distance from overall average	-984	743	100	-856	-467	-289	2326	1372	-964	-512	-469
Overall average height	**7458**										

Distance from country average	AT	BE	DE	ES	FI	FR	GR	IE	IT	NL	PT
AT		-409	-845	-1834	-1053	-828	-64	-342	-3130	-1078	-1243
BE	1318		645	1182	839	468	226	503	1289	849	851
DE	239	2		395	231	-456	-124	-247	476	231	352
ES	-1706	-417	-561		-1061	-749	-359	-283	-1720	-1030	-1533
FI	-535	-370	-335	-672		-471	55	-149	-579	-1256	-820
FR	-133	-564	-845	-181	-293		-142	-322	-154	-322	-221
GR	3246	1810	2102	2823	2847	2473		1338	3138	2830	2977
IE	2013	1132	1025	1945	1689	1338	384		2053	1720	1789
IT	-3109	-417	-588	-1827	-1076	-829	-151	-283		-1078	-1243
NL	-606	-406	-381	-686	-1302	-545	-8	-164	-626		-910
PT	-727	-361	-217	-1145	-822	-400	183	-52	-748	-867	

%-distance from country average	AT	BE	DE	ES	FI	FR	GR	IE	IT	NL	PT
AT		-5%	-11%	-28%	-15%	-12%	-1%	-4%	-48%	-16%	-18%
BE	20%		9%	18%	12%	7%	2%	6%	20%	12%	12%
DE	4%	0%		6%	3%	-6%	-1%	-3%	7%	3%	5%
ES	-26%	-5%	-7%		-15%	-10%	-4%	-3%	-26%	-15%	-22%
FI	-8%	-5%	-4%	-10%		-7%	1%	-2%	-9%	-18%	-12%
FR	-2%	-7%	-11%	-3%	-4%		-1%	-4%	-2%	-5%	-3%
GR	50%	22%	28%	43%	41%	34%		15%	48%	41%	43%
IE	31%	14%	14%	29%	24%	19%	4%		32%	25%	26%
IT	-48%	-5%	-8%	-28%	-15%	-12%	-2%	-3%		-16%	-18%
NL	-9%	-5%	-5%	-10%	-19%	-8%	0%	-2%	-10%		-13%
PT	-11%	-4%	-3%	-17%	-12%	-6%	2%	-1%	-12%	-12%	

Distance from overall average	AT	BE	DE	ES	FI	FR	GR	IE	IT	NL	PT
BE	333										
DE	-745	745									
ES	-2690	326	-461								
FI	-1519	373	-235	-1528							
FR	-1117	179	-745	-1038	-760						
GR	2261	2552	2202	1967	2381	2184					
IE	1029	1875	1124	1089	1223	1050	2710				
IT	-4093	326	-488	-2684	-1542	-1117	2175	1089			
NL	-1590	337	-282	-1542	-1768	-834	2318	1208	-1590		
PT	-1711	382	-117	-2001	-1289	-689	2509	1320	-1711	-1379	

%-distance from overall average	AT	BE	DE	ES	FI	FR	GR	IE	IT	NL	PT
AT											
BE	4%										
DE	-10%	10%									
ES	-36%	4%	-6%								
FI	-20%	5%	-3%	-20%							
FR	-15%	2%	-10%	-14%	-10%						
GR	30%	34%	30%	26%	32%	29%					
IE	14%	25%	15%	15%	16%	14%	36%				
IT	-55%	4%	-7%	-36%	-21%	-15%	29%	15%			
NL	-21%	5%	-4%	-21%	-24%	-11%	31%	16%	-21%		
PT	-23%	5%	-2%	-27%	-17%	-9%	34%	18%	-23%	-18%	

Note: In the distance matrices the values of the vertical columns reflect the distance of the average height between country X and country Y from the overall average height of country Y (vis-à-vis all the other countries). Conversely, the values of the horizontal rows reflect the distance of the average height between country X and country Y from the overall height of country X.

Table A2. Average linkage (standard clustering method)

Average distance between clusters	AT	BE	DE	ES	FI	FR	GR	IE	IT	NL	PT
AT	0										
BE	10777	0									
DE	10297	10605	0								
ES	10571	12391	12479	0							
FI	11313	12391	12479	8293	0						
FR	9647	9092	7795	12283	12283	0					
GR	17406	17406	17406	15410	16754	17406	0				
IE	13622	13297	13272	14465	14465	13074	17598	0			
IT	3545	10777	10297	10386	11313	9647	17406	13622	0		
NL	11111	12212	12335	8472	6398	12080	16956	14262	11111	0	
PT	11088	12599	12681	7030	8424	12491	16579	14084	11088	8335	0
Average height	**10938**	**12155**	**11964**	**11178**	**11411**	**11580**	**17033**	**14176**	**10919**	**11327**	**11440**
Distance from overall average	-1255	-38	-228	-1015	-782	-613	4840	1983	-1274	-866	-753
Overall average height	**12193**										

Distance from country average	AT	BE	DE	ES	FI	FR	GR	IE	IT	NL	PT
AT		-1377	-1668	-607	-98	-1933	373	-554	-7375	-216	-351
BE	-160		-1359	1213	980	-2488	373	-879	-142	885	1159
DE	-641	-1549		1301	1067	-3785	373	-904	-623	1007	1241
ES	-367	236	514		-3118	703	-1622	289	-533	-2855	-4410
FI	376	236	514	-2885		703	-278	289	394	-4930	-3016
FR	-1291	-3063	-4169	1105	872		373	-1102	-1272	753	1051
GR	6468	5251	5441	4232	5343	5826		3422	6486	5629	5139
IE	2685	1142	1308	3287	3054	1494	566		2703	2935	2644
IT	-7393	-1377	-1668	-792	-98	-1933	373	-554		-216	-351
NL	173	58	370	-2706	-5014	501	-77	86	192		-3105
PT	151	444	717	-4148	-2988	911	-454	-92	169	-2992	

%-distance from country average	AT	BE	DE	ES	FI	FR	GR	IE	IT	NL	PT
AT		-11%	-14%	-5%	-1%	-17%	2%	-4%	-68%	-2%	-3%
BE	-1%		-11%	11%	9%	-21%	2%	-6%	-1%	8%	10%
DE	-6%	-13%		12%	9%	-33%	2%	-6%	-6%	9%	11%
ES	-3%	2%	4%		-27%	6%	-10%	2%	-5%	-25%	-39%
FI	3%	2%	4%	-26%		6%	-2%	2%	4%	-44%	-26%
FR	-12%	-25%	-35%	10%	8%		2%	-8%	-12%	7%	9%
GR	59%	43%	45%	38%	47%	50%		24%	59%	50%	45%
IE	25%	9%	11%	29%	27%	13%	3%		25%	26%	23%
IT	-68%	-11%	-14%	-7%	-1%	-17%	2%	-4%		-2%	-3%
NL	2%	0%	3%	-24%	-44%	4%	0%	1%	2%		-27%
PT	1%	4%	6%	-37%	-26%	8%	-3%	-1%	2%	-26%	

Distance from overall average	AT	BE	DE	ES	FI	FR	GR	IE	IT	NL	PT
BE	-1415										
DE	-1896	-1587									
ES	-1622	198	286								
FI	-879	198	286	-3900							
FR	-2546	-3101	-4398	90	90						
GR	5213	5213	5213	3217	4561	5213					
IE	1430	1104	1079	2272	2272	881	5405				
IT	-8648	-1415	-1896	-1807	-879	-2546	5213	1430			
NL	-1082	20	142	-3721	-5795	-113	4763	2069	-1082		
PT	-1104	406	488	-5163	-3769	298	4386	1891	-1104	-3857	

%-distance from overall average	AT	BE	DE	ES	FI	FR	GR	IE	IT	NL	PT
AT											
BE	-12%										
DE	-16%	-13%									
ES	-13%	2%	2%								
FI	-7%	2%	2%	-32%							
FR	-21%	-25%	-36%	1%	1%						
GR	43%	43%	43%	26%	37%	43%					
IE	12%	9%	9%	19%	19%	7%	44%				
IT	-71%	-12%	-16%	-15%	-7%	-21%	43%	12%			
NL	-9%	0%	1%	-31%	-48%	-1%	39%	17%	-9%		
PT	-9%	3%	4%	-42%	-31%	2%	36%	16%	-9%	-32%	

Note: In the distance matrices the values of the vertical columns reflect the distance of the average height between country X and country Y from the overall average height of country Y (vis-à-vis all the other countries). Conversely, the values of the horizontal rows reflect the distance of the average height between country X and country Y from the overall height of country X.

Table A3. Single linkage (smoothing method)

Average distance between clusters	AT	BE	DE	ES	FI	FR	GR	IE	IT	NL	PT
AT	0										
BE	514	0									
DE	516	67	0								
ES	526	673	673	0							
FI	518	671	671	169	0						
FR	516	68	46	673	671	0					
GR	578	710	710	373	355	710	0				
IE	269	387	411	459	506	407	506	0			
IT	491	648	648	358	291	648	420	417	0		
NL	274	557	557	420	364	557	455	327	317	0	
PT	526	673	673	87	108	673	375	459	357	420	0
Average height	**473**	**497**	**497**	**441**	**432**	**497**	**519**	**415**	**460**	**425**	**435**
Distance from overall average	10	34	34	-22	-30	34	57	-48	-3	-38	-28
Overall average height	**463**										

Distance from country average	AT	BE	DE	ES	FI	FR	GR	IE	IT	NL	PT
AT		17	19	85	86	19	59	-146	32	-151	91
BE	41		-430	231	238	-429	191	-28	188	132	237
DE	43	-430		231	238	-451	191	-4	188	132	237
ES	53	176	176		-263	176	-146	44	-101	-4	-348
FI	45	174	174	-272		174	-165	91	-168	-61	-327
FR	43	-429	-451	231	238		191	-8	188	132	237
GR	105	214	213	-68	-78	214		91	-40	30	-60
IE	-203	-110	-86	18	74	-89	-13		-43	-97	24
IT	19	152	151	-83	-141	151	-99	2		-108	-78
NL	-199	60	60	-21	-68	60	-64	-88	-143		-15
PT	53	176	176	-354	-324	176	-145	44	-102	-4	

%-distance from country average	AT	BE	DE	ES	FI	FR	GR	IE	IT	NL	PT
AT		3%	4%	19%	20%	4%	11%	-35%	7%	-36%	21%
BE	9%		-87%	52%	55%	-86%	37%	-7%	41%	31%	55%
DE	9%	-87%		52%	55%	-91%	37%	-1%	41%	31%	55%
ES	11%	35%	35%		-61%	35%	-28%	11%	-22%	-1%	-80%
FI	10%	35%	35%	-62%		35%	-32%	22%	-37%	-14%	-75%
FR	9%	-86%	-91%	52%	55%		37%	-2%	41%	31%	55%
GR	22%	43%	43%	-15%	-18%	43%		22%	-9%	7%	-14%
IE	-43%	-22%	-17%	4%	17%	-18%	-2%		-9%	-23%	6%
IT	4%	31%	30%	-19%	-33%	30%	-19%	1%		-25%	-18%
NL	-42%	12%	12%	-5%	-16%	12%	-12%	-21%	-31%		-3%
PT	11%	35%	35%	-80%	-75%	35%	-28%	11%	-22%	-1%	

Distance from overall average	AT	BE	DE	ES	FI	FR	GR	IE	IT	NL	PT
AT											
BE	51										
DE	53	-396									
ES	63	210	210								
FI	55	208	208	-294							
FR	53	-395	-417	210	208						
GR	115	247	247	-89	-108	247					
IE	-193	-76	-52	-4	44	-55	44				
IT	29	185	185	-104	-172	185	-43	-46			
NL	-189	94	94	-42	-99	94	-8	-135	-146		
PT	63	210	210	-375	-355	210	-88	-4	-106	-42	

%-distance from overall average	AT	BE	DE	ES	FI	FR	GR	IE	IT	NL	PT
AT											
BE	11%										
DE	11%	-86%									
ES	14%	45%	45%								
FI	12%	45%	45%	-63%							
FR	11%	-85%	-90%	45%	45%						
GR	25%	53%	53%	-19%	-23%	53%					
IE	-42%	-16%	-11%	-1%	9%	-12%	9%				
IT	6%	40%	40%	-23%	-37%	40%	-9%	-10%			
NL	-41%	20%	20%	-9%	-21%	20%	-2%	-29%	-32%		
PT	14%	45%	45%	-81%	-77%	45%	-19%	-1%	-23%	-9%	

Note: In the distance matrices the values of the vertical columns reflect the distance of the average height between country X and country Y from the overall average height of country Y (vis-à-vis all the other countries). Conversely, the values of the horizontal rows reflect the distance of the average height between country X and country Y from the overall height of country X.

Table A4. Average linkage (smoothing method)

Average distance between clusters	AT	BE	DE	ES	FI	FR	GR	IE	IT	NL	PT
AT	0										
BE	1684	0									
DE	1593	106	0								
ES	5403	5820	5820	0							
FI	5308	5820	5820	462	0						
FR	1684	107	61	5820	5820	0					
GR	5360	5820	5820	893	926	5820	0				
IE	1072	1540	1552	4988	4878	1542	4940	0			
IT	3602	4405	4405	2547	2256	4405	2659	3387	0		
NL	2568	3224	3224	3825	3617	3224	3806	2157	1682	0	
PT	5403	5820	5820	176	310	5820	943	4988	2502	3825	0
Average height	**3368**	**3434**	**3422**	**3575**	**3522**	**3430**	**3699**	**3104**	**3185**	**3115**	**3561**
Distance from overall average	-34	33	21	174	120	29	297	-297	-217	-286	159
Overall average height	**3401**										

Distance from country average	AT	BE	DE	ES	FI	FR	GR	IE	IT	NL	PT
AT		-1751	-1829	1828	1787	-1746	1661	-2033	417	-547	1843
BE	-1684		-3316	2245	2298	-3323	2121	-1564	1220	109	2259
DE	-1774	-3329		2245	2298	-3369	2121	-1552	1220	109	2259
ES	2035	2385	2398		-3059	2390	-2806	1883	-638	709	-3385
FI	1941	2385	2398	-3113		2390	-2773	1774	-929	501	-3251
FR	-1684	-3327	-3361	2245	2298		2121	-1563	1220	109	2259
GR	1992	2385	2398	-2683	-2596	2390		1836	-526	691	-2617
IE	-2296	-1894	-1870	1412	1356	-1888	1241		203	-958	1427
IT	234	970	982	-1028	-1266	974	-1040	283		-1433	-1059
NL	-800	-210	-198	249	95	-206	108	-947	-1503		264
PT	2035	2385	2398	-3400	-3212	2390	-2755	1883	-683	709	
%-distance from country average	AT	BE	DE	ES	FI	FR	GR	IE	IT	NL	PT
AT		-51%	-53%	51%	51%	-51%	45%	-65%	13%	-18%	52%
BE	-50%		-97%	63%	65%	-97%	57%	-50%	38%	4%	63%
DE	-53%	-97%		63%	65%	-98%	57%	-50%	38%	4%	63%
ES	60%	69%	70%		-87%	70%	-76%	61%	-20%	23%	-95%
FI	58%	69%	70%	-87%		70%	-75%	57%	-29%	16%	-91%
FR	-50%	-97%	-98%	63%	65%		57%	-50%	38%	4%	63%
GR	59%	69%	70%	-75%	-74%	70%		59%	-17%	22%	-74%
IE	-68%	-55%	-55%	40%	39%	-55%	34%		6%	-31%	40%
IT	7%	28%	29%	-29%	-36%	28%	-28%	9%		-46%	-30%
NL	-24%	-6%	-6%	7%	3%	-6%	3%	-31%	-47%		7%
PT	60%	69%	70%	-95%	-91%	70%	-74%	61%	-21%	23%	
Distance from overall average	AT	BE	DE	ES	FI	FR	GR	IE	IT	NL	PT
AT											
BE	-1718										
DE	-1808	-3296									
ES	2002	2418	2418								
FI	1907	2418	2418	-2939							
FR	-1718	-3294	-3340	2418	2418						
GR	1958	2418	2418	-2509	-2476	2418					
IE	-2330	-1861	-1849	1586	1477	-1860	1539				
IT	201	1003	1003	-854	-1146	1003	-743	-14			
NL	-833	-177	-177	423	215	-177	405	-1244	-1719		
PT	2002	2418	2418	-3226	-3091	2418	-2458	1586	-900	423	
%-distance from overall average	AT	BE	DE	ES	FI	FR	GR	IE	IT	NL	PT
AT											
BE	-50%										
DE	-53%	-97%									
ES	59%	71%	71%								
FI	56%	71%	71%	-86%							
FR	-50%	-97%	-98%	71%	71%						
GR	58%	71%	71%	-74%	-73%	71%					
IE	-68%	-55%	-54%	47%	43%	-55%	45%				
IT	6%	29%	29%	-25%	-34%	29%	-22%	0%			
NL	-24%	-5%	-5%	12%	6%	-5%	12%	-37%	-51%		
PT	59%	71%	71%	-95%	-91%	71%	-72%	47%	-26%	12%	

Note: In the distance matrices the values of the vertical columns reflect the distance of the average height between country X and country Y from the overall average height of country Y (vis-à-vis all the other countries). Conversely, the values of the horizontal rows reflect the distance of the average height between country X and country Y from the overall height of country X.

European Central Bank Working Paper Series

For a complete list of Working Papers published by the ECB, please visit the ECB's website (http://www.ecb.int)

585 "Are specific skills an obstacle to labor market adjustment? Theory and an application to the EU enlargement" by A. Lamo, J. Messina and E. Wasmer, February 2006.

586 "A method to generate structural impulse-responses for measuring the effects of shocks in structural macro models" by A. Beyer and R. E. A. Farmer, February 2006.

587 "Determinants of business cycle synchronisation across euro area countries" by U. Böwer and C. Guillemineau, February 2006.

588 "Rational inattention, inflation developments and perceptions after the euro cash changeover" by M. Ehrmann, February 2006.

589 "Forecasting economic aggregates by disaggregates" by D. F. Hendry and K. Hubrich, February 2006.

590 "The pecking order of cross-border investment" by C. Daude and M. Fratzscher, February 2006.

591 "Cointegration in panel data with breaks and cross-section dependence" by A. Banerjee and J. L. Carrion-i-Silvestre, February 2006.

592 "Non-linear dynamics in the euro area demand for M1" by A. Calza and A. Zaghini, February 2006.

593 "Robustifying learnability" by R. J. Tetlow and P. von zur Muehlen, February 2006.

594 "The euro's trade effects" by R. Baldwin, comments by J. A. Frankel and J. Melitz, March 2006.

595 "Trends and cycles in the euro area: how much heterogeneity and should we worry about it?" by D. Giannone and L. Reichlin, comments by B. E. Sørensen and M. McCarthy, March 2006.

596 "The effects of EMU on structural reforms in labour and product markets" by R. Duval and J. Elmeskov, comments by S. Nickell and J. F. Jimeno, March 2006.

597 "Price setting and inflation persistence: did EMU matter?" by I. Angeloni, L. Aucremanne, M. Ciccarelli, comments by W. T. Dickens and T. Yates, March 2006.

598 "The impact of the euro on financial markets" by L. Cappiello, P. Hördahl, A. Kadareja and S. Manganelli, comments by X. Vives and B. Gerard, March 2006.

599 "What effects is EMU having on the euro area and its Member Countries? An overview" by F. P. Mongelli and J. L. Vega, March 2006.

600 "A speed limit monetary policy rule for the euro area" by L. Stracca, April 2006.

601 "Excess burden and the cost of inefficiency in public services provision" by A. Afonso and V. Gaspar, April 2006.

602 "Job flow dynamics and firing restrictions: evidence from Europe" by J. Messina and G. Vallanti, April 2006.

603 "Estimating multi-country VAR models" by F. Canova and M. Ciccarelli, April 2006.

604 "A dynamic model of settlement" by T. Koeppl, C. Monnet and T. Temzelides, April 2006.

605 "(Un)Predictability and macroeconomic stability" by A. D'Agostino, D. Giannone and P. Surico, April 2006.

606 "Measuring the importance of the uniform nonsynchronization hypothesis" by D. A. Dias, C. Robalo Marques and J. M. C. Santos Silva, April 2006.

607 "Price setting behaviour in the Netherlands: results of a survey" by M. Hoeberichts and A. Stokman, April 2006.

608 "How does information affect the comovement between interest rates and exchange rates?" by M. Sánchez, April 2006.

609 "The elusive welfare economics of price stability as a monetary policy objective: why New Keynesian central bankers should validate core inflation" by W. H. Buiter, April 2006.

610 "Real-time model uncertainty in the United States: the Fed from 1996-2003" by R. J. Tetlow and B. Ironside, April 2006.

611 "Monetary policy, determinacy, and learnability in the open economy" by J. Bullard and E. Schaling, April 2006.

612 "Optimal fiscal and monetary policy in a medium-scale macroeconomic model" by S. Schmitt-Grohé and M. Uribe, April 2006.

613 "Welfare-based monetary policy rules in an estimated DSGE model of the US economy" by M. Juillard, P. Karam, D. Laxton and P. Pesenti, April 2006.

614 "Expenditure switching vs. real exchange rate stabilization: competing objectives for exchange rate policy" by M. B. Devereux and C. Engel, April 2006.

615 "Quantitative goals for monetary policy" by A. Fatás, I. Mihov and A. K. Rose, April 2006.

616 "Global financial transmission of monetary policy shocks" by M. Ehrmann and M. Fratzscher, April 2006.

617 "New survey evidence on the pricing behaviour of Luxembourg firms" by P. Lünnemann and T. Y. Mathä, May 2006.

618 "The patterns and determinants of price setting in the Belgian industry" by D. Cornille and M. Dossche, May 2006.

619 "Cyclical inflation divergence and different labor market institutions in the EMU" by A. Campolmi and E. Faia, May 2006.

620 "Does fiscal policy matter for the trade account? A panel cointegration study" by K. Funke and C. Nickel, May 2006.

621 "Assessing predetermined expectations in the standard sticky-price model: a Bayesian approach" by P. Welz, May 2006.

622 "Short-term forecasts of euro area real GDP growth: an assessment of real-time performance based on vintage data" by M. Diron, May 2006.

623 "Human capital, the structure of production, and growth" by A. Ciccone and E. Papaioannou, May 2006.

624 "Foreign reserves management subject to a policy objective" by J. Coche, M. Koivu, K. Nyholm and V. Poikonen, May 2006.

625 "Sectoral explanations of employment in Europe: the role of services" by A. D'Agostino, R. Serafini and M. Ward-Warmedinger, May 2006.

626 "Financial integration, international portfolio choice and the European Monetary Union" by R. A. De Santis and B. Gérard, May 2006.

627 "Euro area banking sector integration: using hierarchical cluster analysis techniques" by C. Kok Sørensen, J. M. Puigvert Gutiérrez, May 2006.

ISSN 1561081-0
9 771561 081005