Bournemouth University

BAFES –
Bournemouth Accounting, Finance & Economic Series

NO 04 / 2016

A copula-based clustering algorithm to analyse EU country diets

Di Lascio, F. Marta L. & Disegna Marta

A copula-based clustering algorithm to analyse EU country diets

F. Marta L. Di Lascio[a], Marta Disegna[b*]

[a]*Faculty of Economics and Management, Free University of Bolzano, Italy,*[b]*Faculty of Management, School of Tourism, Bournemouth University, UK*

Abstract

The aim of the paper is to suggest a novel clustering technique to explore the changes of the food diet in 40 European countries in accordance with common European policies and guidelines on healthy diets and lifestyles. The proposed clustering algorithm is based on copulas and it is called CoClust. The CoClust algorithm is able to find clusters according to the multivariate dependence structure of the data generating process. The database analysed contains information on the proportions of calories from 16 food aggregates in 40 European countries observed over 40 years by the Food and Agriculture Organization of the United Nations (FAO). The findings suggest that European country diets are changing, individually or as a group, but not in a unique direction. Central and Eastern European countries are becoming unhealthier, while the tendency followed by the majority of the remaining countries is to integrate the common European guidelines on healthy, balanced, and diversified diets in their national policies.

Keywords: Clustering, CoClust, Healthy diet, Convergence, Dietary energy, EU countries.

*Corresponding author
Email addresses: `marta.dilascio@unibz.it` (F. Marta L. Di Lascio), `disegnam@bournemouth.ac.uk` (Marta Disegna)

1 Introduction

In the literature there is substantial agreement regarding the idea that food consumption patterns, or diets, are changing over time in a non-uniform way, especially showing large spatial variation (Kastner et al., 2012; Naska et al., 2009; Sengul & Sengul, 2006; Traill, 1997). However, as regards European Union (EU) countries, Schmidhuber & Traill (2006) discovered an increased homogenisation of diets from 1961 to 2001, even though regional diet differences were still recognisable. This result can be partially attributed to the common food-based dietary guidelines (FBDG) adopted since World War II by EU governments in order to promote healthy diets ensuring adequate daily intakes of both macronutrients (proteins, carbohydrates and fats) and micronutrients (vitamins and minerals). In 1996 the Food and Agricultural Organisation (FAO) and the World Health Organisation (WHO) published guidelines for the creation of FBDG at the national level, accepted by the EU and subsequently published in 2001. Specifically, WHO/FAO are encouraging and supporting EU countries to develop and implement their own FBDG for healthy, diversified and balanced diets adapted to each country's specific needs (e.g. individual needs, cultural context, locally available foods and dietary customs). Diets are in fact complex combinations of different food products which do not merely represent regional food consumption patterns but which also describe more widely the social, cultural, political, economic and environmental situation of a country (Capone et al., 2014). As an example, the Mediterranean diet (MD) has been registered in the United Nations Educational, Scientific and Cultural Organisation (UNESCO) list of intangible Cultural Heritage in 2010 as a "lifestyle" describing the intimate relationship between population and nature (Capone et al., 2014). The necessity to reach a deep knowledge on regional food consumption and production patterns is undoubtedly recognised in the literature due to the strong direct effect of diet on both environment and health (Capone et al., 2014; de Ruiter et al., 2014; Tukker et al., 2011). No less important are the indirect effects of diet on individual/household income and on national healthcare budgets, due to the social and economic costs that arise to treat diet-related illnesses, climate change and distribution of welfare (Capone et al., 2014; Tukker et al., 2011; Wan, 2005).

This study aims to explore how EU diets have changed in accordance with the WHO/FAO guidelines on healthy diet between 1970 and 2011. In the literature, different food indexes have been proposed to evaluate the health of a country diet. In this paper, the Mediterranean Adequacy Index (MAI), developed by Alberti-Fidanza et al. (1999), has been adopted to assess how close each country diet is to the healthy MD over time. Interest in studying tendencies toward homogenisation in nutrient supply among different countries over time, i.e. convergence in food consumption, has increased over the years for a number of reasons (Healy, 2014; Kalawole, 2015; Wan, 2005). Consequently, besides the use of traditional measures of convergence, such as gamma and sigma convergence (Barro & Sala-i Martin, 1992; Regmi & Unnevehr, 2005; Sengul & Sengul, 2006; Wan, 2005), different econometric models (Connor, 1994; Elsner & Hartmann, 1998; Herrmann & Roder, 1995; Kalawole, 2015; Lyons et al., 2009; Wan, 2005), cluster analysis and data mining (Balanza et al., 2007; Blandford, 1984; Gil & Gracia, 2000; Gil et al., 1995; Healy, 2014; Lazaroua et al., 2012; Pérez-Ortiz et al., 2014; Petrovici et al., 2005; Sengul & Sengul, 2006; Tukker et al., 2011) have been adopted to detect and analyse period of convergence in food consumption. In this study, gamma convergence, as defined by Wan (2005), has been used to provide a first insight on the convergence in food consumption across EU countries over time. Moreover, a novel clustering algorithm based on copula function (Durante & Sempi, 2015; Nelsen, 2006; Sklar, 1959), called CoClust (Di Lascio & Giannerini, 2012, 2016), has been employed to identify sets of countries characterised by complex associations in their dietary structures.

The CoClust is a model-based clustering algorithm that assumes data are generated by a copula model (Nelsen, 2006). This means that this algorithm is able to discover complex multivariate relationships that are not possible to identify using more traditional dependence measures, like the linear correlation coefficient, which is only able to capture linear bivariate dependence relationships.

Since the seminal work of Grubel (1968), copulas have been adopted to study the dependence structure of financial markets, i.e. to measure the co-movements among financial time series, and nowadays there is a vast and growing literature on this topic (see for instance the following recent studies: Bartram & Wang, 2015; Berger, 2016; Durante et al., 2014, 2015; Ling et al., 2015; Min & Czado, 2014; Shahzad et al., 2016; Wang & Xie, 2016; Weng & Gong, 2016). Copulas have also been used in studies related to applied economics, such as tourism (Pérez-Rodríguez et al., 2015; Tang et al., 1959; Zhu et al., 2016) and agriculture. Focusing on economic agriculture, the field of this study, copulas have been adopted to study the co-movements between time series regarding prices for food (corn, soyabean, wheat, and rice) and either oil prices (Reboredo, 2012) or US dollar (USD) exchange rate (Reboredo & Ugando, 2014). Furthermore, at the micro-level copula models have been integrated to censored equation systems (Yen & Lin, 2008) and nonparametric median regression (Brakers & Van Keilegom, 2008) to study meat consumption and total food expenditure respectively. However, to the best of our knowledge, copulas have been never used to perform cluster analysis using food consumption as segmentation variables.

The paper is organised as follows. Section 2 describes the CoClust algorithm and illustrates pros and cons related to its use. In Section 3 data has been presented focusing on the description of the evolution of EU countries' diets towards the MD. Section 4 presents the results with a focus on which countries are evolving towards a (un)healthy diet. The paper concludes in Section by offering some final remarks.

2 The CoClust algorithm

2.1 Introduction

It is well known that clustering is a data-driven method that attempts to discover structures within data itself, grouping together objects into clusters. Moreover, cluster analysis is a useful exploratory technique for multivariate data as it allows the identification of potentially meaningful relationships between objects. The extensive literature of clustering includes both methods based on distance/dissimilarity measures and methods based on probability models (Everitt et al., 2001). Generally speaking, distance-based clustering techniques group objects into the same cluster on the basis of their similarity computed through a suitable distance or dissimilarity measure between two objects, like the Euclidean distance or the one minus the squared correlation coefficient. Hence, in this case clusters are generated in a way which maximises homogeneity within-cluster and the separation between-cluster. On the contrary, model-based clustering techniques (Fraley & Raftery, 1998) assume that data are generated by a finite mixture of probability distributions. This means that objects are grouped in the same $k-$th cluster if they come from the same specific density function f_k that is generally a Gaussian one. In this case the operational definition of clusters is based on the internal linear dependence among objects. In practice, both distance-based and model-based methods are able to cope only with pairwise and/or linear relationships between objects, but they are not suitable to model multivariate complex dependence. To overcome these limits, it is possible to adopt the CoClust algorithm, a model-based technique that assumes data are generated by a copula function.

2.2 Copula function background

Copula function is born in the probabilistic metric space with Sklar's theorem (Sklar, 1959) that states that every joint distribution function $F(\cdot)$ can be expressed in terms of K marginal distribution function F_k and the copula distribution function C as follows:

$$F(x_1,\ldots,x_k,\ldots,x_K) = C(F_1(x_1),\ldots,F_k(x_k),\ldots,F_K(x_K)) \tag{1}$$

for all $(x_1,\ldots,x_k,\ldots,x_K) \in \bar{\mathbb{R}}^K$ (where $\bar{\mathbb{R}}$ denotes the extended real line). According to this theorem we can split any joint probability function $f(\cdot)$ into the margins and a copula, so that the latter represents the association between variables, e.g. the multivariate dependence structure of a joint density function (Trivedi & Zimmer, 2005, for details):

$$f(x_1,\ldots,x_k,\ldots,x_K) = c(F_1(x_1),\ldots,F_k(x_k),\ldots,F_K(x_K))\prod_{k=1}^{K} f_k(x_k). \quad (2)$$

Such separation determines the modelling flexibility given by copulas since it is possible to decompose the estimation problem in two steps: in the first step margins are estimated; and in the second step the copula model is estimated. The most used estimation method is the two-stage inference for margins method (Joe & Xu, 1996) that employs the log-likelihood estimation method to estimates both the parameter(s) of each margin and the copula parameter θ. This method can be used in a semi-parametric approach (Genest et al., 1995) that does not require distributional assumptions on the margins since these are modelled through the empirical cumulative distribution functions $\hat{F}_k(X_{ki})$ with $k = 1,\ldots,K$. Then, the log-likelihood copula function is used to estimate θ as follows:

$$\hat{\theta} = \arg\max_{\theta} \sum_{i=1}^{n} \log c\left\{\hat{F}_1(X_{1i}),\ldots,\hat{F}_k(X_{ki}),\ldots,\hat{F}_K(X_{Ki});\theta\right\} \quad (3)$$

where n is the sample size. In the literature, many different copula models are available (Nelsen, 2006, for details) but it has been demonstrated that the Elliptical and the Archimedean families are the most useful in empirical modelling. The Elliptical family includes the Gaussian copula and the t-copula: both copulas are symmetric; they exhibit the strongest dependence in the middle of the distribution; and they can take into account both positive and negative dependence since $-1 \leq \theta \leq 1$. As usual, the t-copula is characterised by two parameters, the dependence parameter θ and the number of degrees of freedom, and it converges to a Gaussian copula as the number of degrees of freedom approaches infinity. The Archimedean family, by comparison, enables us to describe both left and right asymmetry as well as weak symmetry among the margins by employing Clayton's, Gumbel's and Frank's model, respectively. Clayton's copula has the parameter $\theta \in (0,\infty)$ and as θ approaches zero, the margins become independent. The dependence parameter θ of a Gumbel model is restricted to the interval $[1,+\infty)$ where the value 1 means to independence. Finally, the dependence parameter θ of a Frank copula may assume any real value and as θ approaches zero, the marginal distributions become independent. Figure 1 shows the contour plots of the bivariate density functions defined by the above five copula models with standard normal margins and a level of θ such that the Kendall's correlation coefficient is $\tau = 0.7$. According to the kind of copula model, the value of θ will have a specific meaning. However, it is always true that the greater the value of the dependence parameter, the stronger the association among the margins.

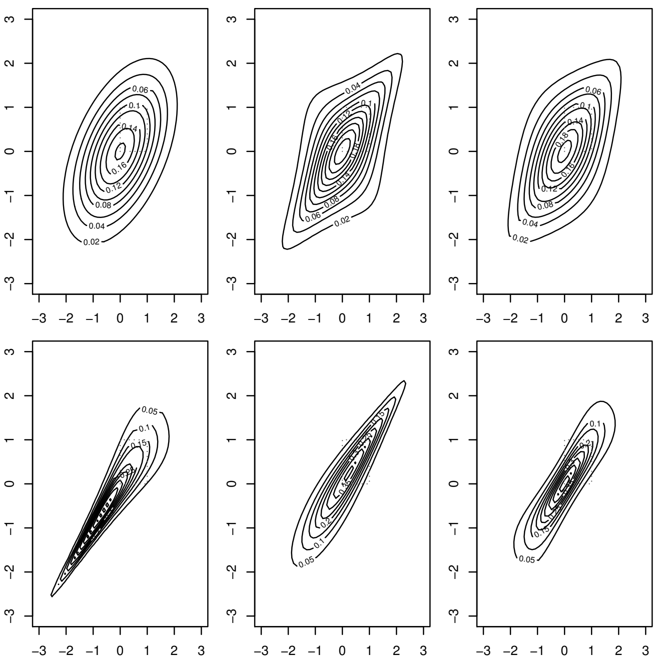


Figure 1: Contour plots of bivariate copula models with normal standard margins and dependence parameter θ such that the Kendall's correlation coefficient is $\tau = 0.7$; upper panel: Gaussian and t-Student copula models for two number of degrees of freedom: 2 and 4; lower panel: Clayton, Gumbel and Frank copula models.

2.3 Idea behind the CoClust algorithm

The CoClust algorithm assumes that data are generated by a K-dimensional copula function C where each margin F_k is the probability-integral transform of the density function f_k that generates the k-th cluster. Hence, a K-dimensional copula represents a clustering of K clusters; therefore, the copula model C describes the shape of the multivariate dependence structure among clusters (margins) and its parameter θ expresses the strength of the multivariate dependence. Consequently, each cluster can be viewed as the realization of a random variable and it is identified by one (univariate) margin. Having K clusters means having K dependent margins and a copula makes it possible to investigate this kind of dependence. Hence, objects in the same cluster are independent and identically distributed realizations from the same marginal distribution while objects across clusters, which can be called profiles, share an inter-cluster multivariate dependence structure, i.e. they have a mutually dependent relationship. Therefore, the CoClust aims at within-cluster independence and between-cluster dependence instead of within-cluster homogeneity and between-cluster separation, as in the more traditional clustering approaches.

The starting point of the CoClust algorithm is a standard $N \times q$ data matrix in which N/K are the objects to be grouped in K groups and q are the segmentation variables. The basic idea behind the CoClust and how the data are grouped and the final profiles are identified at the end of the clustering procedure is represented in Figure 2. The main steps of the CoClust algorithm are represented in Figure 3 (refer to Di Lascio & Giannerini, 2012, 2016, for more technical details) and described as follows:

1. for $k = 2, \ldots, K_{max}$, where $K_{max} \leq N$ is the maximum number of clusters to be tried, select a subset of n_k k-plets of rows in the data matrix on the basis of a sort of multivariate measure of dependence based on pairwise Spearman's ρ correlation coefficient (see Di Lascio & Giannerini, 2016, for more technical details);

2. fit a copula model (for details see Section 2.2) and select the subset of n_k k-plets of rows, say n_K K-plets, that maximizes the log–likelihood copula function; hence, the number of clusters K, that is the dimension of the copula, is automatically chosen;

3. select a K-plet on the basis of the Spearman's ρ-based measure of dependence used at step 1. and estimate $K!$ copulas by using the observations already clustered and a permutation of the candidate to the allocation;

4. allocate the permutation of the selected K-plet to the clustering by assign-

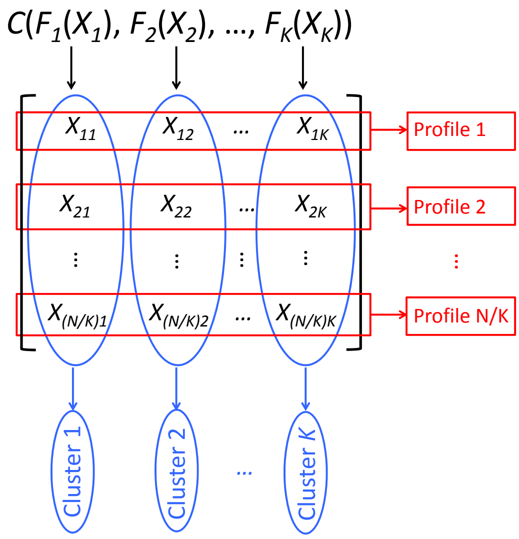


Figure 2: The basic idea of the CoClust algorithm.

ing each observation to the corresponding cluster if it increases the log-likelihood of the copula fit, otherwise drop the entire K-plet of rows;

5. repeat steps 3. and 4. until all the observations are evaluated (either allocated or discarded).

In summary, at the first two steps the algorithm selects the optimal number of clusters K; from the second step onwards, it evaluates a $K-$plet of rows at a time and it allocates the observations to the K clusters in a way that the complex dependence relationships among objects are represented by a K-dimensional copula function.

2.3.1 Selection of the number of clusters and the copula model

The CoClust algorithm selects automatically the number of clusters K on the basis of the log-likelihood of the copula function estimated on the subsets of k-plets allocated until a step predefined by the user. However, it is possible to select K post-clustering, that is, on the basis of the whole final clustering. In this case, the number of clusters can be selected by using an information criterion, such as the Bayesian information criterion (BIC, from now on) that, for a copula model m

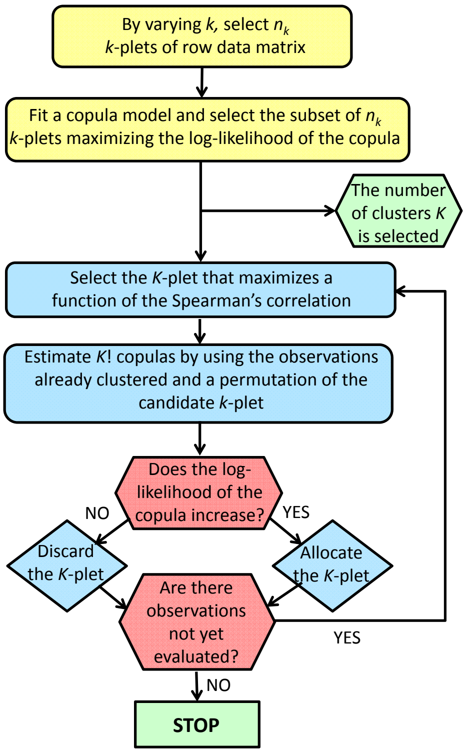


Figure 3: The CoClust algorithm procedure.

with single parameter, has the following expression:

$$\mathrm{BIC}_{K,m} = -2\log\Pi_{i=1}^{n} c_m\left\{\hat{F}_1(X_{1i}),\ldots,\hat{F}_k(X_{ki}),\ldots,\hat{F}_K(X_{Ki});\hat{\theta}\right\} - \log\left((N/K)q\right) \tag{4}$$

where $\hat{\theta}$ is in eq. (3) and $(N/K)q$ is the total number of observations allocated in each cluster (N/K q-dimensional vectors). According to Raftery & Nema (2006), we can compute K as follows:

$$K = \arg\max_{k,m}\left[\frac{\mathrm{BIC}_{k,m} - \mathrm{BIC}_{k-1,m}}{\mathrm{BIC}_{k-1,m}}\right] \tag{5}$$

where m indicates a specific copula model and varies in a predefined set of models and $k \in \{2,\ldots,K_{max}\}$. The copula model used in CoClust is estimated through the two-stage inference for margins method in its semi-parametric version (see Section 2.2). The selected number of clusters K and copula model are the ones that maximize the reduction of the BIC.

2.4 Pros and Cons of the CoClust algorithm

The main advantages of the CoClust with respect to more traditional clustering algorithms are as follows:

- it does not require a starting classification to be chosen;
- it does not require the number of clusters to be set a *priori*;
- it is able to capture multivariate and nonlinear dependence relationships underlying the observed data (see Di Lascio & Giannerini, 2012, 2016, for details);
- it does not require the marginal probability distributions to be set as Gaussian;
- it is able to discard irrelevant observations (see Di Lascio & Giannerini, 2016, for details).

On the other hand, this algorithm does not select automatically the copula model therefore a *posteriori* model selection criteria have to be employed (see Section 2.3.1). Furthermore, the algorithm can be slow when the number of clusters is more than 6 and/or the sample size is big since the permutations of the selected $k-$plet have to be computed.

Di Lascio & Giannerini (2012, 2016) have proved in several Monte Carlo studies that the CoClust algorithm is able to i) find almost always the correct number of clusters; *ii*) work well irrespectively of the number of clusters, the sample size and the copula model used; *iii*) distinguish objects coming from different data generating processes or dependent on independent objects. Finally, the CoClust algorithm has been implemented in the R package `CoClust` which is available on CRAN (Di Lascio & Giannerini, 2015).

3 Data

Annual data covering 1970 to 2011 from the 38 countries that constitute the continent of Europe (following the FAO list), plus Cyprus and Turkey, have been considered. Average calories per capita per day from different food aggregates have been analysed in this study, since calories have been considered in the literature as a good approximation of food consumption useful to analyse changes over space and time (Gil et al., 1995). Data have been obtained from the national food balance sheet of the FAO database (FAOSTAT, 2016). The following 16 food aggregates have been analysed: (1) animal fats; (2) eggs; (3) fish and seafood; (4) meat; (5) milk (excluding butter); (6) other animal; (7) alcoholic beverages; (8) cereals (excluding beer); (9) fruits (excluding wine); (10) potatoes; (11) pulses; (12) sugar and sweeteners; (13) soyabeans; (14) vegetable oils; (15) vegetables; (16) other vegetables. Overall, these food aggregates make up the diet of any country included in the study and can be grouped into different aggregates depending on the research objectives. In this study, two different classifications have been considered: animal (1-6) vs. vegetables (7-16); Mediterranean (3, 8-11, 13-16) vs. non-Mediterranean (1, 2, 4-6, 12). Alcoholic beverages have been excluded from the Mediterranean/non-Mediterranean classification since it was impossible to separate the calories from healthy and unhealthy food items.

3.1 Healthiness of the EU diet

The average food consumption in EU countries was 3380 calories/capita/day in 1970 while it was 3534 calories/capita/day in 2011. The former Soviet Union and the former Yugoslavia have been excluded from this analysis owing to the particular political situation experienced by these countries in the 1990s. The upward trend of the average EU food consumption is shown in Figure 4(a).

Even though specific recommendations will vary from country to country, WHO/FAO guidelines commonly recommend that people should eat plenty of fruits, vegetables, cereals, preferably whole grains, and fish; should choose foods low in sugar, salt and saturated fat; and should do regular physical activities. In this respect, the MD is considered a healthy prudent diet since it is plant-centered (i.e. it is characterised by a high consumption of legumes, whole grains, fruits and vegetables, nuts and seeds) and the consumption of meat and dairy products is moderately low. As we can observe in Figure 4(b), in the EU countries the average proportion of calories derived from animal products is lower than the proportion derived from vegetable products in all years. Over time, three major trends in food calorie consumption can be observed, confirming what was found by Gil et al. (1995) for the period 1970-1990. On average, the subgroup of EU countries considered shows an upward trend in the share of animal calories consumed over the period 1970 to 1981. After a short initial period of instability, a decline has been observed until 1994, while in the remaining years the share of animal calories has stabilised at around 29% of the total consumed calories. To perform an in-depth analysis of the adherence of the EU diet to the MD, and therefore to a healthy diet, the MAI has been computed over time as defined by Alberti-Fidanza et al. (1999). The MAI is easily obtained by dividing the sum of the percentages of the calories from Mediterranean food aggregates (M), by the sum of the percentages of the calories from Non-Mediterranean food aggregates $\left(\overline{M}\right)$, which is as follows:

$$\text{MAI}_{it} = \frac{\sum_{j \in M} y_{ijt}}{\sum_{j \in \overline{M}} y_{ijt}} \tag{6}$$

where y_{ijt} is the per capita per day calories from the j-th food aggregate observed in the t-th year for the i-th EU country, with $i = 1, \ldots, n,\ j = 1, \ldots, J$, and $t = 1970, \ldots, 2011$. The higher the MAI value, the higher the adherence to the MD. As it is possible to observe from Figure 4(c), two major trends characterised the MAI: a downward trend starting in 1970 and ending in 1983; and an upward trend from 1984 to 2011 that is partially described by the efforts made by EU governments to educate people towards the adoption of healthy diets and good lifestyle practices. It is particularly interesting to observe that the highest degree of adherence of the EU diet to the MD is observed in 2010-2011 and this value is close to the one observed 40 years before, in 1970.

3.2 Convergence among EU dietary structures

Convergence in food consumption has been defined by Kalawole (2015) as the tendency toward homogenization in nutrient supply over space and time. With the aim of exploring EU diet over time, in this study the coefficient of variation (CV) of total calories from food products has been computed as follows:

$$CV_{jt} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{ijt} - \overline{y}_{jt})^2}}{\overline{y}_{jt}} \tag{7}$$

where $\overline{y}_{jt}$ is the average per capita per day calories from the j-th food aggregate observed in the t-th year for all EU countries, with $i = 1, \ldots, n,\ j = 1, \ldots, J$, and

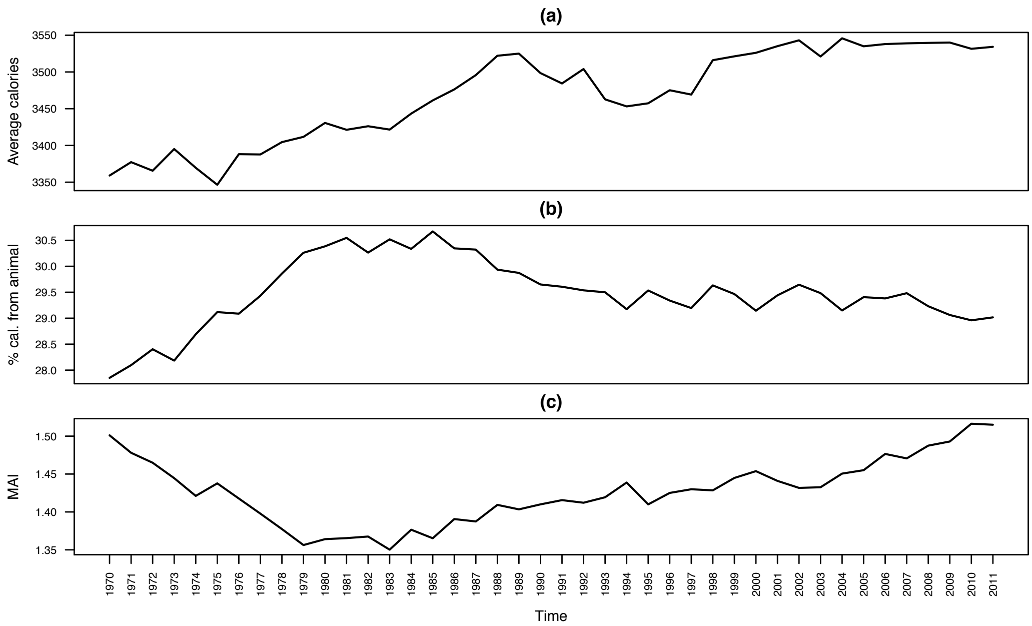


Figure 4: Evolution of average food consumption (daily calories), average proportions of calories from animal products (%) and MAI trend in EU countries.

$t = 1970, \ldots, 2011$. The evolution of the *CV* for the EU diet is plotted in Figure 5. As defined by Wan (2005), reductions in the *CV* in food consumption over time are identified as a period of convergence. Therefore, the period 1975-1984 is characterised by a gamma convergence among EU countries, while in 1985-1992 we observe an increase in the *CV* ending with a strong peak. Comparing the *CV* computed on all EU countries (black line figure 5) with the *CV* computed on the subgroup of countries in which the former Soviet Union/block and the former Yugoslavia are excluded (grey line figure 5), it seems clear that the main cause of this peak is the dissolution of the former Soviet Union/block (December, 1991) and of the former Yugoslavia (1991-1992). From 1992 to 2000 a period of instability is registered while from 2000 to 2009 we observe an overall tendency of gamma convergence. Finally, in 2011 a slight increase in the *CV* is observed and this can be considered as a sign that the trend is about to change towards a divergence in food consumption patterns among EU countries. It is interesting to note that, as for the MAI, the *CV* values observed in 1970 and in 2010-2011 are similar, meaning that the current EU situation regarding both homogenization in food consumption and adherence of the EU diet to the MD is close to the EU situation observed 40 years ago.

In the following Section 4, the discussion will focus on the most recent and

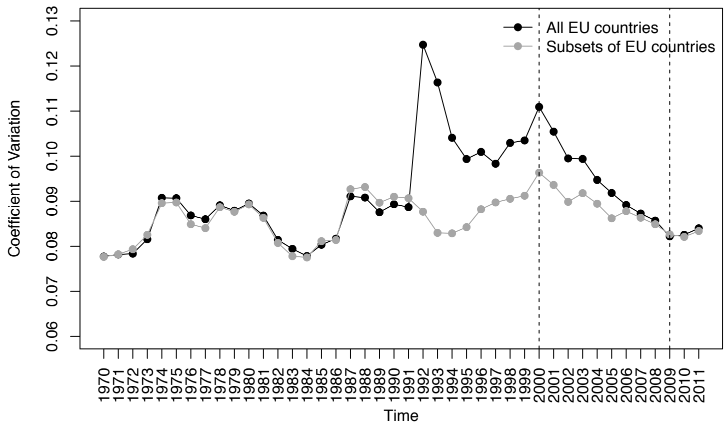


Figure 5: Trends of *CV* in food consumption.

long gamma convergence period in food consumption observed across EU countries, i.e. the period from 2000 to 2009. It is important to observe that also the MAI increase during this period, meaning that EU countries are converging towards a healthier diet.

More in detail, Figure 6 shows the evolution of the *CV* in Europe for the total calories from the main food aggregates. In the period 2000-2009, there is a substantial reduction in the *CV* of the calories from the two macro groups, animal and vegetable, with a faster decrease in animal-based calories. In particular, the *CV* of the calories from meat seems to decrease faster than the *CV* of the calories from fish, and the downward trend of the *CV* of the calories from fruits is similar to that of the calories from vegetables. Overall, the EU is converging towards a healthier diet but the speed of convergence is affected by the different speeds of convergence observed per food aggregate.

4 Results of clustering analysis

The CoClust algorithm has been run separately on data collected in 2000 and 2009 to have an in-depth understanding of the food dietary characteristics across EU countries during the identified convergence period.

In both years, the number of clusters K and the most suitable copula model have been selected by using the BIC as in eq. (5). Figure (7) shows the value of the BIC for any partition from 2 to 8 clusters and for 5 different copula models. In both years, when $k = 2$ the Gaussian copula is the best copula model, while for $k > 2$ the $t-$Student copula with 2 degrees of freedom is the copula model that

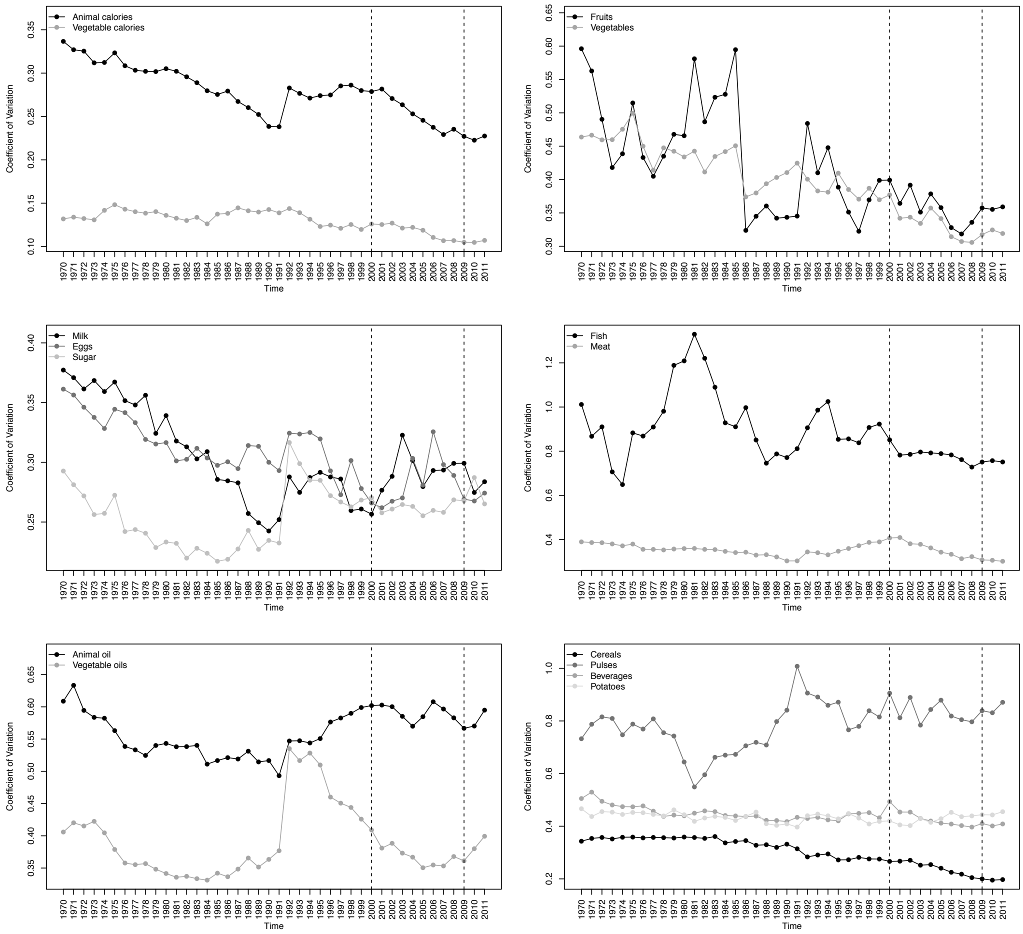
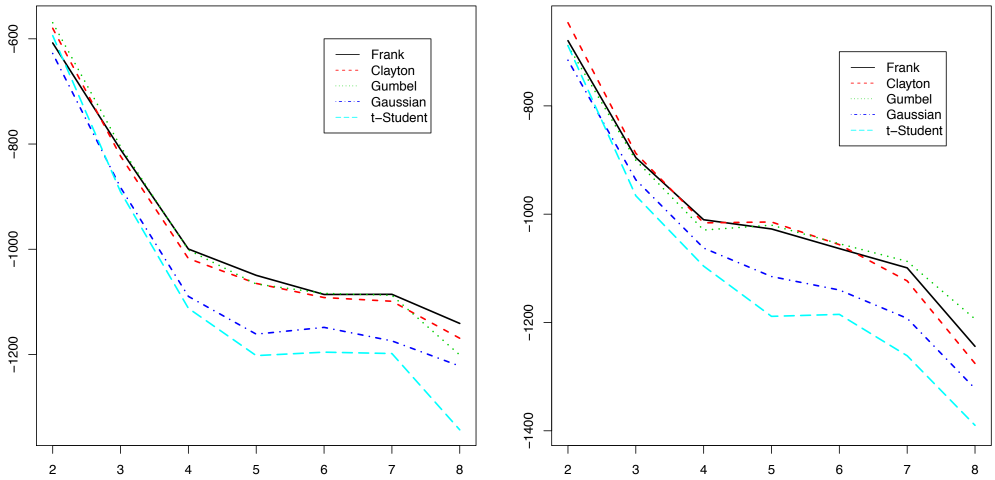


Figure 6: Gamma-convergence of the diet structure across EU.

allows us to obtain smaller values of the BIC. The lines of the $t-$Student show the typical elbow when $k = 5$ indicating that the decrement of the BIC is maximum. Hence, the selected number of clusters in each year is $K = 5$ and the most suitable copula is the t-Student with 2 degrees of freedom.

Figure 7: BIC value by varying copula models (y-axis) and number of clusters (x-axis) for the two years under investigation.



(a) 2000 (b) 2009

In both years, the CoClust algorithm allocates all countries, meaning that 8 profiles of 5 countries each have been identified. It is important to remember that the dietary structure of countries in the same cluster are independent and identically distributed while the dietary structure of countries in the same profile are dependent, i.e. countries in the same profile have a mutual multivariate structure of dependence. The 8 profiles obtained in each year are shown and compared in Table 1. Looking at the two-way table, it is possible to observe that profiles are made up of two parts: one part, called static from now on, comprises the countries characterized by common changes in dietary structure such that any country does not change profile from 2000 to 2009 (groups of countries located on the main diagonal of Table 1); the other part, called dynamic from now on, comprises countries with a dietary structure dependent on different countries in different years (single country or groups of countries located outside the main diagonal of Table 1).

Summing up, it is possible to identify 10 static aggregates of countries (for the sake of simplicity labelled SAs from now on). The remaining 11 countries, i.e. Hungary (H), Turkey (TR), Serbia-Montenegro (S_M), Czech Republic (CZ), Italy (I), Poland (PL), the United Kingdom (UK), Latvia (LV), Croatia (HR), Malta (M) and the Republic of Moldova (MD), constitute the dynamic part, moving from one profile to another and likely to embrace different diet compositions over time.

Table 1: EU countries profiles 2000-2009.

		2000							
		Profile 1	Profile 2	Profile 3	Profile 4	Profile 5	Profile 6	Profile 7	Profile 8
	Profile 1	Portugal Finland Norway Iceland Sweden (SA1)							
	Profile 2		France Austria Switzerland Germany (SA2)						Croatia
	Profile 3			Greece Slovenia Albania FYROM* (SA3)			Italy		
2009	Profile 4				Denmark Belgium Slovakia (SA4)		Poland		Malta
	Profile 5				Serbia- Montenegro	Luxembourg Cyprus Spain (SA5)		Latvia	
	Profile 6		Hungary		Czech Republic	Ireland Bulgaria (SA6)	UK		
	Profile 7						the Netherlands Romania (SA7)	Russian Federation Ukraine (SA8)	Republic of Moldova
	Profile 8			Turkey				Belarus Estonia (SA9)	Lithuania Bosnia- Herzegovina (SA10)

*The Former Yugoslav Republic of Macedonia.

Figure 8 geographically maps each element of Table 1, i.e. SAs and each country of the dynamic part, providing a spatial visualisation of the clustering results. It is interesting to note that, even though the geographical parameter is not always a valid criterion for grouping countries, some SAs (1, 2, 3, 8, and 9) are made up by neighbouring countries.

Since the set of countries that make up the profiles changes over time, it is not meaningful to study the evolution of profile diets. Therefore, the following analysis will focus on diet evolutions of SAs and each country that belongs to the dynamic part of the profiles.

4.1 Who converges towards a healthy diet?

Diet compositions per profiles, SAs and single countries belonging to the dynamic part over time are reported in Tables 1, 2, and 3 in Appendix A.

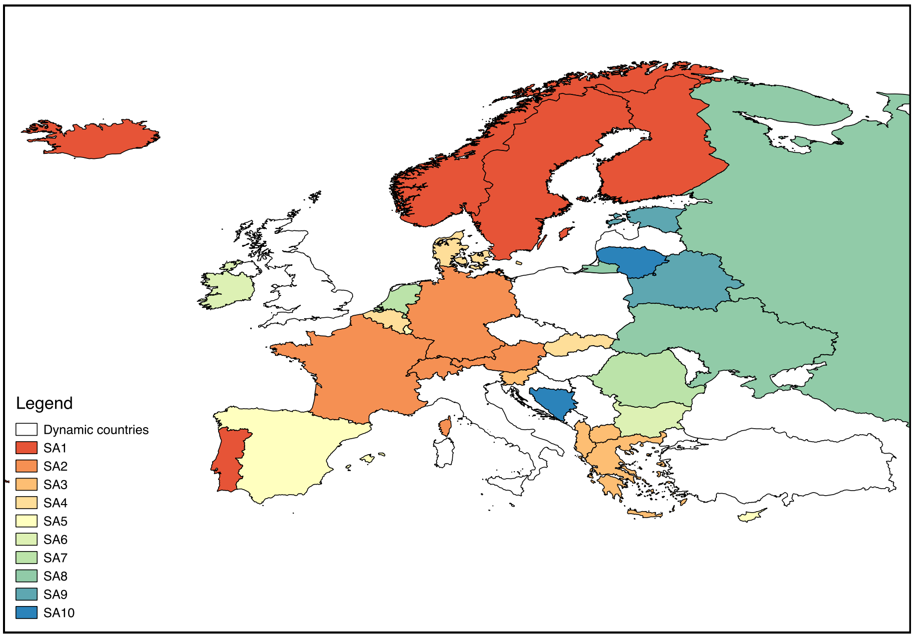


Figure 8: Geographical distribution of static and dynamic aggregates of countries.

To evaluate how healthy are the diets of SAs and single dynamic countries are and how each diet has evolved over time, the MAI has been computed and represented in Figure 9. Single countries and SAs above the main diagonal have experienced an increase in MAI and the higher the vertical distance to the diagonal, the healthier the diet has become over time. Tables 2 and 3 offer more information on changes in MAI respectively for SAs and countries that belong to the dynamic aggregate of countries. In particular, the proportions of calories from Mediterranean and non-Mediterranean aggregates have been computed and tests for proportions have been calculated to identify significant changes over time. Moreover, percentage changes in MAI (%$\triangle$MAI) have been calculated and included at the end of Tables 2 and 3.

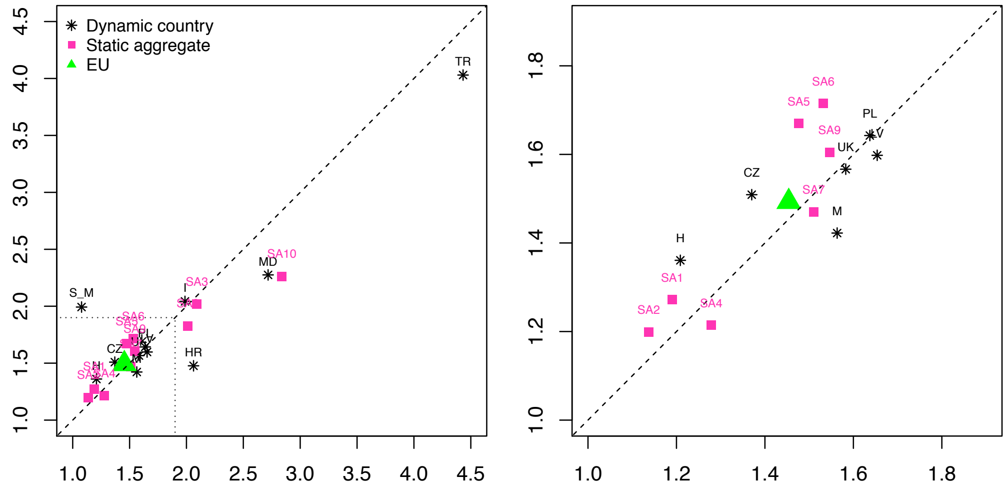
Among the SAs, the SA7 shows the lowest, in absolute value, percentage change in MAI and it is the only one that is characterised by significant changes in the proportions of calories from both Mediterranean and non-Mediterranean aggregates, meaning that this aggregate does not significantly change its diet towards either a healthier or unhealthier diet. All the remaining SAs experienced a significant change in their diet towards either a healthier diet (SA1, SA2, SA5, SA6, and SA9) or an unhealthier diet (SA3, SA4, SA8, and SA10). In particular, SA5 and SA6 experienced the highest percentage increase of the MAI (respectively 13% and 12%), mainly attributable to an increase in the proportion of calories from vegetable oils, cereals (excluding beer), and pulses. On the other hand, SA10 shows the highest deterioration of its diet (-20%), mainly due to an increase in the proportion of calories from meat, sugar and sweeteners. Moreover, it is possible to observe that, in both 2000 and 2009, SA1 and SA2 are characterised by the highest proportions of non-Mediterranean calories, while SA10 shows the highest proportions of Mediterranean calories.

Analysing the dynamic part (Table 3), it is possible to recognise a group of countries that did not experience significant movements over time towards or away from the MD, namely Italy, Latvia, Poland, and the United Kingdom. Czech Republic, Hungary, and Serbia and Montenegro changed their diets towards a healthier diet while the remaining countries experienced a decrease in MAI, i.e. their diets became less healthy over time. Serbia and Montenegro show the highest percentage increase of the MAI (85%) mainly thanks to an increase in the proportion of calories from cereals (excluding beer), fruits (excluding wine), pulses and other vegetables. Hungary also experienced a healthy change in its diet structure as well, but with a definitely lower intensity (13%). In particular, Hungary moved from a diet highly characterised by calories from animal products, in particular from meat, to a diet characterised by a higher proportion of calories from cereals (excluding beer) and vegetable oils. In contrast, reducing the proportion of calories from potatoes, other vegetables and cereals (excluding beer), Croatia moved from a vegetables-oriented diet towards an animal-oriented diet and its decrement in MAI is the highest observed among countries that belong to the dynamic part. Turkey was characterised by the highest proportion of calories from Mediterranean aggregates, in both 2000 and 2009, and therefore by the highest MAI values, but the diet composition of this country changed over the years and its diet became less healthy. In particular, even though the Turkish diet has been always characterised by a predominant consumption of vegetable products it moved from a fruit-pulses oriented diet towards a cereals-vegetables-oriented diet overall, reducing the proportion of calories from Mediterranean aggregates. Finally, Malta and the Republic of Moldova were used to following a diet rich in cereals and vegetables, but in 2009 they joined different groups of countries both characterised by more animal products-oriented diets, i.e. animal fats and meat.

5 Discussion and conclusions

The convergence of EU country diets towards a healthy diet, i.e. the Mediterranean diet (MD), have been analysed over a period of 10 years. The adherence of each country diet towards the MD has been computed by means of the Mediterranean Adequacy Index (MAI), while the convergence in food consumption has been investigated by making use of a novel clustering algorithm based on copula and called CoClust. In contrast with the more classical clustering techniques,

Figure 9: MAI value per country aggregate in 2000 (x-axis) and in 2009 (y-axis) for all aggregates (a) and for a subset of aggregates characterised by MAI values smaller than 1.9 in both years (b).



(a) (b)

Table 2: Mediterranean and Non-Mediterranean proportions over time and SAs.

	SA1	SA2	SA3	SA4	SA5	SA6	SA7	SA8	SA9	SA10
Mediterranean aggregates										
2000	51.84	50.17	65.29	52.91	55.21	56.17	57.48	64.66	58.19	70.46
2009	53.53	51.51	64.93	52.16	58.51	58.93	56.83	61.30	57.17	64.85
$p-$value χ^2_1	0.002^{***}	0.022^{**}	0.544	0.278	$<0.001^{***}$	0.001^{***}	0.448	$<0.001^{***}$	0.235	$<0.001^{***}$
Non-Mediterranean aggregates										
2000	43.54	44.12	31.21	41.38	37.39	36.66	38.05	32.12	37.63	24.81
2009	42.07	42.95	32.17	42.94	35.04	34.37	38.65	33.55	35.62	28.66
$p-$value χ^2_1	0.006^{***}	0.043^{**}	0.095^{*}	0.023^{**}	0.001^{***}	0.005^{***}	0.482	0.084^{*}	0.015^{**}	$<0.001^{***}$
$\%\triangle$MAI	6.88	5.44	-3.55	-5.00	13.09	11.93	-2.64	-9.22	3.81	-20.32

Notes: $^{***}p-$value$\leqslant 0.01$, $^{**}p-$value$\leqslant 0.05$, $^{*}p-$value$\leqslant 0.1$

Table 3: Mediterranean and Non-Mediterranean proportions over time and countries that belong to the dynamic part.

	CZ	H	HR	I	LV	M	PL	S_M	TR	UK	MD
Mediterranean aggregates											
2000	52.86	51.48	62.59	64.03	59.22	59.38	59.73	49.23	81.12	58.31	71.54
2009	55.26	54.17	56.76	65.24	57.20	56.94	59.21	63.19	79.60	58.39	67.24
$p-$value χ^2_1	0.050^{*}	0.029^{**}	$<0.001^{***}$	0.283	0.101	0.040^{**}	0.663	$<0.001^{***}$	0.097^{*}	0.965	0.001^{***}
Non-Mediterranean aggregates											
2000	38.58	42.59	30.35	32.22	35.81	37.98	36.48	45.67	18.30	36.84	26.32
2009	36.62	39.81	38.43	31.94	35.79	40.03	36.05	31.71	19.75	37.27	29.56
$p-$value χ^2_1	0.101	0.022^{**}	$<0.001^{***}$	0.809	1.000	0.080^{*}	0.713	$<0.001^{***}$	0.108	0.722	0.008^{***}
$\%\triangle$MAI	10.13	12.58	-28.39	2.79	-3.38	-9.01	0.33	84.89	-9.07	-1.02	-16.29

Notes: $^{***}p-$value$\leqslant 0.01$, $^{**}p-$value$\leqslant 0.05$, $^{*}p-$value$\leqslant 0.1$

which group units on the basis of either a suitable similarity/distance measure or an underlying linear probability model, the CoClust algorithm makes it possible to identify sets of units on the basis of the complex multivariate dependence structure of the data generating process. More specifically, CoClust allows us to identify sets of EU countries, called profiles, characterised by complex associations in their food consumption patterns.

The CoClust algorithm has been run using data on the proportion of calories from 16 different food aggregates collected by the Food and Agriculture Organization of the United Nations (FAO) on 40 EU countries. The clustering analysis has been performed separately on data observed in 2000 and 2009 since these years represent the beginning and the end of the more recent and long gamma convergence period in EU, as identified through the coefficient of variation (CV) computed from 1970 to 2011. In each year, a 5-dimensional t-Student copula model has been selected and all countries have been allocated to one cluster. Therefore, 8 profiles, each of which made up by 5 countries characterised by a multivariate dependence structure in their food consumption, have been detected and further analysed.

The 8 profiles have been made up of different countries over years but stable groups of countries are identified. Overall, 10 different sets of countries, called static aggregates (SAs), that are stable over time regarding countries composition, have been highlighted. Among the SAs, sets like the Nordic countries, the Western EU countries and the Balkans have been identified confirming the findings of Gil et al. (1995), although a different set of EU countries has been considered. Moreover, 11 countries, that belong over time to different profiles, have been identified. Most of the time, these countries changed their diets towards the (un)healthier diet of the SA that belongs to the profile towards which the country is going.

While the univariate descriptive analyses, jointly provided by the MAI and the CV, showed that from 2000 to 2009 EU countries experienced a convergence towards a common healthier food dietary structure, the multivariate descriptive analyses, provided by CoClust, suggests a different EU food dietary picture. Diets of EU countries are inevitably becoming more and more similar thanks to the adoption of common public policies (as for instance those regarding organic, local products and FBDG), multinational market strategies (with the creation of EU brands) and the internationalisation of food distribution. However, dietary differences within the EU still exist and, maybe linked to migration and globalisation issues, some countries, either individually or as a group, changed their dietary structure over years towards a (un)healthier diet as represented in Figure 10.

In particular, it is important to underline that SA10 (made up of Lithuania and Bosnia-Herzegovina), SA8 (made up of the Russian Federation and Ukraine), Malta, Republic of Moldova and Croatia are experiencing a worrisome increase in the consumption of high-calories and nutrient-poor foods (high in fats and sweet-

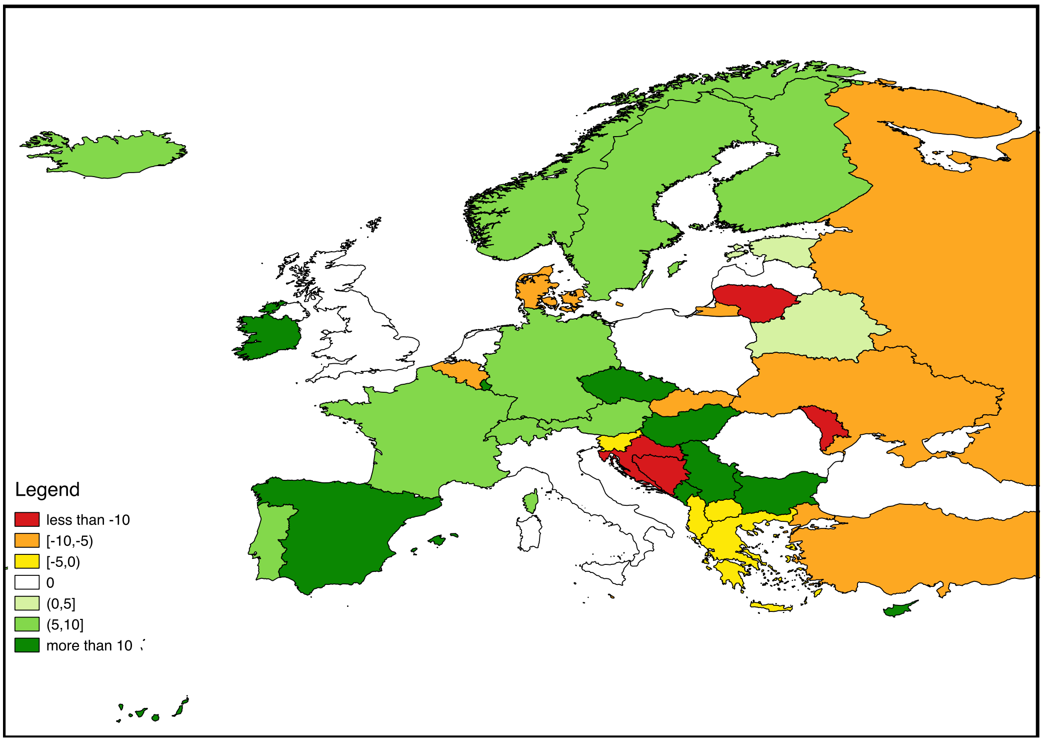


Figure 10: Geographical distribution of the percentage changes in MAI among European countries.

eners) that will lead to an increase in obesity and diet-related chronic disease. Conversely, Serbia and Montenegro is going towards a modern healthy diet (as defined by Healy, 2014) rich in vegetables and fruits. Among the countries that did not experience a significant change in their food composition, it is worth noting that SA7 (the Netherlands and Romania) and Italy present respectively the lowest and the highest MAI values in both years analysed. In particular, the Netherlands and Romania might introduce new or more powerful and persuasive food policies that encourage people to follow a healthier diet with lower consumption of meat and milk. On the other hand, Italy, together with SA3 (made up by Greece, Slovenia, Albania and FYROM), seem to be worthy ancestors of the Greek peasant farmers of the 1950s, from which the Mediterranean diet originates, embracing varied and healthy diets rich in cereals (excluding beer), fruits (excluding wine), vegetables and vegetable oils (especially Italy).

Finally, as it has been observed (see Figure 8), the geographical proximity among countries does not imply either a common food dietary or the convergence to a common diet over years but, looking at Figure 10, it seems that this is a relevant criterion in understanding times and modalities by which common guidelines and policies are implemented among EU countries.

Acknowledgment

We would like to thank professor Tim Lloyd for his valuable comments and suggestions which helped us to improve the paper. The first author acknowledges the support of Free University of Bozen-Bolzano, Faculty of Economics and Management, via the project "Correlation, Clustering, and Copulas: Detecting association in multivariate data" (COCCO).

References

Alberti-Fidanza, A., Fidanza, F., Chiuchiu, M., Verducci, G., & Fruttini, D. (1999). Dietary studies on two rural italian population groups of the seven countries study. 3. trend of food and nutrient intake from 1960 to 1991. *European Journal of Clinical Nutrition*, *53*, 854–860.

Balanza, R., Garcia-Lorda, P., Perez-Rodrigo, C., Aranceta, J., Bullo Bonet, M., & Salas-Salvado, J. (2007). Trends in food availability determined by the food and agriculture organization's food balance sheets in mediterranean europe in comparison with other european areas. *Public Health Nutrition*, *10*, 168–176.

Barro, R., & Sala-i Martin, X. (1992). Convergence. *The Journal of Political Economy*, *100*, 223–251.

Bartram, S., & Wang, Y.-H. (2015). European financial market dependence: An industry analysis. *Journal of Banking and Finance*, *59*, 146–163.

Berger, T. (2016). Forecasting based on decomposed financial return series: A wavelet analysis. *Journal of Forecasting*, *35*, 419–433.

Blandford, D. (1984). Changes in food consumption patterns in the oecd area. *European Review of Agricultural Economics*, *11*, 43–65.

Brakers, R., & Van Keilegom, I. (2008). Flexible modeling based on copulas in nonparametric median regression. *Journal of Multivariate Analysis*, *100*, 1270–1281.

Capone, R., El Bilali, H., Debs, P., Cardone, G., & Driouech, N. (2014). Mediterranean food consumption patterns sustainability: setting up a common ground for future research and action. *American Journal of Nutrition and Food Science*, *1*, 37–52.

Connor, J. (1994). North america as a precursor of changes in western european food-purchasing patterns. *European Review of Agricultural Economics*, *21*, 155–173.

Di Lascio, F., & Giannerini, S. (2012). A copula-based algorithm for discovering patterns of dependent observations. *Journal of Classification*, *29(1)*, 50–75.

Di Lascio, F., & Giannerini, S. (2015). Coclust: copula based cluster analysis. r package version 0.3-1. *http://CRAN.R-project.org/package=CoClust*, .

Di Lascio, F., & Giannerini, S. (2016). Clustering dependent observations with copula functions. *Statistical Papers*, (pp. 1–17). doi:`10.1007/s00362-016-0822-3`.

Durante, F., Foscolo, E., Jaworski, P., & Wang, H. (2014). A spatial contagion measure for financial time series. *Expert Systems with Applications*, *41*, 4023–4034.

Durante, F., Foscolo, E., Jaworski, P., & Wang, H. (2015). Connectedness measures of spatial contagion in the banking and insurance sector. *Advances in Intelligent Systems and Computing*, *315*, 217–224.

Durante, F., & Sempi, C. (2015). *Principles of Copula Theory*. CRC Press, Boca Raton.

Elsner, K., & Hartmann, M. (1998). Convergence of food consumption patterns between eastern and western europe. Germany: Discussion Paper No.13, Institute of Agricultural Development in Central and Eastern Europe.

Everitt, B. S., Landau, S., & Leese, M. (2001). *Cluster Analysis*. (4th ed.). London: Hodder Arnold.

FAOSTAT (2016). *Food supply on FAOSTAT webpage*. FAOSTAT. Http://faostat3.fao.org/download/FB/CC/E (accessed July, 2016).

Fraley, C., & Raftery, A. (1998). How many clusters? which clustering method? answers via model–based cluster analysis. *The Computer Journal*, *41*, 578–588.

Genest, C., Ghoudi, K., & Rivest, L. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika*, *82*, 543–552.

Gil, J., & Gracia, A. (2000). Regional food consumption patterns in the eu. *Économies et Sociétiés*, *24*.

Gil, J., Gracia, A., & Pérez, L. (1995). Food consumption and economic development in the european union. *European Review of Agricultural Economics*, *22*, 385–399.

Grubel, H. (1968). Internationally diversified portfolios: Welfare gains and capital flows. *American Economic Review*, *58*, 1299–1314.

Healy, A. (2014). Convergence or difference? western european household food expenditure. *British Food Journal*, *116*, 792–804.

Herrmann, R., & Roder, C. (1995). Does food consumption converge internationally? measurement, empirical tests and determinants. *European Review of Agricultural Economics*, *22*, 400–412.

Joe, H., & Xu, J. (1996). *The estimation method of inference functions for margins for multivariate models*. Technical Report Department of Statistics, University of British Columbia.

Kalawole, O. (2015). Convergence and determinants of change in nutrient supply. *British Food Journal*, *117*, 2880–2898.

Kastner, T., Rivas, M., Koch, W., & Nonhebel, S. (2012). Global changes in diets and the consequences for land requirements for food. *PNAS*, *109*, 6868–6872.

Lazaroua, C., Karaolisb, M., Matalasa, A.-L., & Panagiotakos, D. (2012). Dietary patterns analysis using data mining method. an application to data from the cykids study. *Computer methods and programs in biomedicine*, *108*, 706–714.

Ling, S., McAleer, M., & Tong, H. (2015). Frontiers in time series and financial econometrics: An overview. *Journal of Econometrics*, *189*, 245–250.

Lyons, S., Mayor, K., & Tol, R. (2009). Convergence of consumption patterns during macroeconomic transition: A model of demand in ireland and the oecd. *Economic Modelling*, *26*, 702–714.

Min, A., & Czado, C. (2014). Scomdy models based on pair-copula constructions with application to exchange rates. *Computational Statistics and Data Analysis*, *76*, 523–535.

Naska, A., Berg, M.-A., Cuadrado, C., Freisling, H., Gedrich, K., Gregorič, M., Kelleher, C., Leskova, E., Nelson, M., Pace, L., Remaut, A.-M., Rodrigues, S., Sekula, W., Sjöstrom, M., Trygg, K., Turrini, A., Volatier, J., Zajkas, G., & Trichopoulou, A. (2009). Food balance sheet and household budget survey dietary data and mortality patterns in europe. *British Journal of Nutrition*, *102*, 166–171.

Nelsen, R. (2006). *Introduction to copulas*. New York: Springer.

Pérez-Ortiz, M., de la Paz-Marín, M., Gutiérrez, P., & Hervás-Martínez, C. (2014). Classification of eu countries' progress towards sustainable development based on ordinal regression techniques. *Knowledge-Based Systems*, *66*, 178–189.

Pérez-Rodríguez, J., Ledesma-Rodríguez, F., & Santana-Gallego, M. (2015). Testing dependence between gdp and tourism's growth rates. *Tourism Management*, *48*, 268–282.

Petrovici, D., Ritson, C., & Ness, M. (2005). Exploring disparities and similarities in european food consumption patterns. *Cahiers d'économie et sociologie rurales*, *75*, 24–49.

Raftery, A., & Nema, D. (2006). Variable selection for model-based clustering. *Journal of the American Statistical Association*, *101*, 168–178.

Reboredo, J. (2012). Do food and oil prices co-move? *Energy policy*, *49*, 456–467.

Reboredo, J., & Ugando, M. (2014). US dollar exchange rate and food price dependence: implications for portfolio risk management. *North American Journal of Economics and Finance*, *30*, 72–89.

Regmi, A., & Unnevehr, L. (2005). Convergence or divergence in food demand: A comparison of trends in the eu and north america. Copenhagen, Denmark, Aug. 24-27: Paper presented at the 11th Congress of the EAAE, The Future of Rural Europe in the Global Agri-Food System.

de Ruiter, H., Kastner, T., & Nonhebel, S. (2014). European dietary patterns and their associated land use: Variation between and within countries. *Food Policy*, *44*, 158–166.

Schmidhuber, J., & Traill, W. (2006). The changing structure of diets in the european union in relation to healthy eating guidelines. *Public Health Nutrition*, *9*, 584–595.

Sengul, H., & Sengul, S. (2006). Food consumption and economic development in turkey and european union countries. *Applied Economics*, *38*, 2421–2431.

Shahzad, S., Kumar, R., Ali, S., & Ameer, S. (2016). Interdependence between greece and other european stock markets: A comparison of wavelet and vmd copula, and the portfolio implications. *Physica A: Statistical Mechanics and its Applications*, *47*, 8–33.

Sklar, A. (1959). Fonctions de répartition à *n* dimensions et leures marges. *Publications de l'Institut de Statistique de L'Université de Paris*, *8*, 229–231.

Tang, J., Sriboonchitta, S., Ramos, V., & Wong, W.-K. (1959). Modelling dependence between tourism demand and exchange rate using the copula-based garch model. *Current Issues in Tourism*, *19*, 876–894.

Traill, B. (1997). Convergence in US and EU food systems? the case of food consumption. In L. Caswell, & R. Cotterill (Eds.), *Strategy and Policy in the Food System: Emerging Issues. Part Three: Are food marketing Systems in North America and the European Union Converging?* chapter 11. (pp. 141–145). Amherst, MA: Food marketing policy center, University of Connecticut and University of Massachusetts.

Trivedi, P. K., & Zimmer, D. M. (2005). *Copula Modeling: An Introduction for Practitioners* volume 1. Foundations and Trends in Econometrics.

Tukker, A., Goldbohm, R., de Koning, A., Verheijden, M., Kleijn, R., Wolf, O., Pérez-Domínguez, I., & Rueda-Cantuche, J. (2011). Environmental impacts of changes to healthier diets in europe. *Ecological Economics*, *70*, 1776–1788.

Wan, G. (2005). Convergence in food consumption in rural China: evidence from household survey data. *China Economic Review*, *16*, 90–102.

Wang, G.-J., & Xie, C. (2016). Tail dependence structure of the foreign exchange market: A network view. *Expert Systems with Applications*, *46*, 164–179.

Weng, Y., & Gong, P. (2016). Modeling spatial and temporal dependencies among global stock markets. *Expert Systems with Applications*, *43*, 175–185.

Yen, S., & Lin, B.-H. (2008). Quasi-maximum likelihood estimation of a censored equation system with a copula approach: meat consumption by u.s. individuals. *Agricultural Economics*, *39*, 207–217.

Zhu, L., Lim, C., Xie, W., & Wu, Y. (2016). Modelling tourist flow association for tourism demand forecasting. *Current Issues in Tourism*, *In press*, 1–15.

Appendix

A Table

Table 1: Diet composition profiles in 2000 and 2009 (percentage).

Food categories	Profile 1	Profile 2	Profile 3	Profile 4	Profile 5	Profile 6	Profile 7	Profile 8
2000								
Animal fats	5.94	7.93	2.05	8.62	2.58	4.03	4.14	3.03
Eggs	1.07	1.39	1.15	1.38	1.15	1.33	1.33	1.07
Fish and seafood	2.99	0.85	0.48	0.77	1.27	0.89	0.94	0.99
Meat	11.90	12.33	6.71	9.34	12.37	10.38	6.88	5.60
Milk (excluding butter)	11.12	8.81	8.77	8.12	9.63	9.08	8.56	8.63
Other animals	1.43	1.74	1.48	1.92	1.75	1.59	1.76	1.37
Animal calories	34.45	33.05	20.63	30.14	28.76	27.31	23.61	20.69
Alcoholic beverages	4.62	5.75	2.81	6.17	7.30	4.25	3.95	4.23
Cereals (excluding beer)	24.77	22.98	36.24	24.44	25.25	28.14	32.48	40.38
Fruits (excluding wine)	3.13	3.24	4.81	2.65	3.37	2.98	1.80	2.59
Potatoes	3.46	3.28	3.35	4.25	3.25	4.91	7.71	5.21
Pulses	0.45	0.42	1.67	0.69	0.82	0.84	0.30	0.94
Sugar and sweeteners	12.08	11.63	8.01	12.17	9.61	9.83	12.45	9.40
Soyabeans	0.00	0.07	0.04	0.01	0.00	0.00	0.01	0.00
Vegetable oils	8.85	11.54	10.43	10.38	12.11	11.96	6.59	5.43
Vegetables	1.74	1.93	3.48	2.12	2.35	2.44	1.73	2.58
Other vegetables	6.44	6.11	8.52	6.98	7.17	7.35	9.37	8.54
Vegetable calories	65.55	66.95	79.37	69.86	71.24	72.69	76.39	79.31
2009								
Animal fats	5.76	6.74	3.77	8.01	2.74	5.09	2.95	2.66
Eggs	1.08	1.29	1.04	1.47	1.21	1.31	1.40	1.17
Fish and seafood	2.92	1.14	0.75	1.28	1.50	0.68	0.99	1.13
Meat	12.27	11.17	8.52	8.64	11.33	10.18	7.63	6.84
Milk (excluding butter)	11.58	8.87	9.87	7.53	8.95	8.12	10.61	7.81
Other animals	1.39	1.56	1.39	1.69	1.72	1.67	1.70	1.56
Animal calories	35.00	30.77	25.33	28.61	27.46	27.06	25.30	21.16
Alcoholic beverages	4.40	5.41	2.88	4.50	6.33	6.35	4.56	5.51
Cereals (excluding beer)	24.92	23.64	30.79	27.48	26.08	27.92	31.84	33.99
Fruits (excluding wine)	3.45	3.44	5.16	2.71	3.40	2.69	2.32	2.95
Potatoes	3.05	2.88	2.73	3.88	3.13	3.52	5.23	5.26
Pulses	0.58	0.30	1.36	0.62	1.28	0.84	0.45	1.29
Sugar and sweeteners	9.99	12.50	7.54	13.60	8.63	10.11	10.68	9.43
Soyabeans	0.00	0.13	0.03	0.00	0.00	0.01	0.00	0.09
Vegetable oils	9.57	12.72	12.64	8.88	12.83	12.73	9.67	7.97
Vegetables	2.01	1.94	3.30	2.72	2.56	1.81	2.57	2.95
Other vegetables	7.02	6.26	8.22	7.01	8.29	6.95	7.38	9.40
Vegetable calories	65.00	69.23	74.67	71.39	72.54	72.94	74.70	78.84

Table 2: Diet composition of static aggregates (SAs) in 2000 and 2009 (percentage).

Food categories	SA1	SA2	SA3	SA4	SA5	SA6	SA7	SA8	SA9	SA10
2000										
Animal fats	5.94	7.56	2.38	9.67	1.78	3.75	2.97	3.03	4.14	2.41
Eggs	1.07	1.30	1.15	1.35	1.28	0.97	1.70	1.32	1.38	0.94
Fish and seafood	2.99	0.98	0.50	0.98	1.64	0.74	0.78	1.02	0.87	1.13
Meat	11.90	12.50	8.04	7.49	14.20	9.71	10.02	5.72	8.45	5.21
Milk (excluding butter)	11.12	9.31	9.92	7.89	9.00	10.53	11.39	7.58	8.76	8.10
Other animals	1.43	1.67	1.55	1.94	1.74	1.76	1.96	1.86	1.73	1.30
Animal calories	34.45	33.33	23.54	29.31	29.65	27.46	28.82	20.54	25.34	19.08
Alcoholic beverages	4.62	5.71	3.50	5.71	7.39	7.17	4.47	3.22	4.18	4.73
Cereals (excluding beer)	24.77	22.65	33.59	23.87	21.93	30.10	27.67	37.20	29.17	44.72
Fruits (excluding wine)	3.13	3.34	5.13	2.56	4.18	2.19	2.93	1.40	1.90	1.78
Potatoes	3.46	3.24	3.35	4.74	2.82	3.88	5.04	7.11	8.54	5.83
Pulses	0.45	0.34	1.16	0.48	0.88	0.74	0.41	0.55	0.19	1.13
Sugar and sweeteners	12.08	11.77	8.17	13.05	9.39	9.94	10.02	12.61	13.17	6.86
Soyabeans	0.00	0.09	0.05	0.00	0.00	0.00	0.00	0.00	0.03	0.00
Vegetable oils	8.85	11.48	10.02	10.51	13.30	10.39	11.03	6.98	5.66	4.43
Vegetables	1.74	1.78	3.36	2.24	2.56	2.03	2.54	2.02	1.53	2.32
Other vegetables	6.44	6.27	8.11	7.53	7.91	6.10	7.07	8.37	10.29	9.13
Vegetable calories	65.55	66.67	76.46	70.69	70.35	72.54	71.18	79.46	74.66	80.92
2009										
Animal fats	5.76	7.07	3.69	9.84	1.87	4.04	2.67	2.77	3.46	2.67
Eggs	1.08	1.29	0.96	1.56	1.24	1.05	1.24	1.66	1.47	0.98
Fish and seafood	2.92	1.14	0.60	1.21	1.74	0.71	0.77	1.24	0.91	1.75
Meat	12.27	11.73	8.01	7.96	12.84	9.39	9.59	6.89	8.95	7.07
Milk (excluding butter)	11.58	8.88	10.62	8.21	9.00	8.48	13.54	7.70	8.04	8.63
Other animals	1.39	1.59	1.34	1.79	1.63	1.56	1.47	2.02	2.05	1.34
Animal calories	35.00	31.70	25.22	30.56	28.32	25.23	29.28	22.28	24.89	22.44
Alcoholic beverages	4.40	5.54	2.90	4.90	6.45	6.70	4.52	5.16	7.21	6.49
Cereals (excluding beer)	24.92	22.76	30.92	25.07	24.06	31.89	28.09	31.91	26.54	36.76
Fruits (excluding wine)	3.45	3.29	5.13	2.88	3.39	2.36	3.09	1.86	2.57	2.67
Potatoes	3.05	2.76	2.97	3.68	2.39	3.13	4.69	6.42	7.38	4.65
Pulses	0.58	0.30	1.37	0.42	1.26	1.00	0.51	0.43	0.03	1.73
Sugar and sweeteners	9.99	12.39	7.55	13.58	8.45	9.85	10.14	12.50	11.64	7.98
Soyabeans	0.00	0.16	0.04	0.00	0.00	0.01	0.00	0.00	0.01	0.17
Vegetable oils	9.57	13.35	11.40	9.81	14.84	11.36	9.27	9.24	7.38	4.96
Vegetables	2.01	1.86	3.42	2.48	2.53	1.71	2.79	2.33	2.57	2.94
Other vegetables	7.02	5.88	9.07	6.61	8.30	6.76	7.63	7.87	9.78	9.21
Vegetable calories	65.00	68.30	74.78	69.44	71.68	74.77	70.72	77.72	75.11	77.56

Table 3: Diet composition of countries belonging to the dynamic aggregates in 2000 and 2009 (percentage).

Food categories	CZ	H	HR	I	LV	M	PL	S_M	TR	UK	MD
2000											
Animal fats	4.78	9.54	3.17	4.21	6.48	5.31	6.07	9.19	0.98	3.78	1.30
Eggs	1.91	1.78	1.32	1.25	1.22	1.18	1.10	0.86	1.16	0.98	0.95
Fish and seafood	0.67	0.24	0.50	1.12	0.89	1.60	0.78	0.11	0.40	0.95	0.34
Meat	9.08	11.59	5.52	10.33	5.82	7.76	9.40	16.70	2.39	12.11	3.73
Milk (excluding butter)	8.02	6.61	8.55	7.15	10.16	8.77	6.79	9.12	5.04	9.06	9.87
Other animals	2.33	2.05	1.50	1.59	1.61	1.43	1.32	1.33	1.23	1.20	1.33
Animal Calories	26.78	31.80	20.56	25.64	26.18	26.06	25.45	37.31	11.20	28.08	17.52
Alcoholic beverages	8.56	5.93	7.05	3.75	4.97	2.64	3.79	5.10	0.58	4.84	2.13
Cereals (excluding beer)	24.39	24.42	28.07	30.22	29.99	31.18	32.03	26.64	44.84	22.91	55.31
Fruits (excluding wine)	2.54	2.79	3.60	4.71	2.40	3.23	1.64	3.12	3.75	2.61	2.63
Potatoes	4.36	3.47	6.88	1.98	7.14	3.82	6.58	2.26	3.32	6.01	3.77
Pulses	0.51	0.77	0.78	1.33	0.00	1.41	0.54	1.69	3.32	1.47	0.04
Sugar and sweeteners	12.47	11.03	10.30	7.70	10.52	13.52	11.81	8.47	7.50	9.71	9.14
Soyabeans	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Vegetable oils	12.07	11.77	10.08	16.54	7.83	5.26	7.97	7.90	11.78	12.95	3.16
Vegetables	1.57	2.61	2.60	3.02	1.58	3.43	2.31	2.30	3.88	1.77	2.06
Other vegetables	6.69	5.42	10.08	5.12	9.40	9.45	7.89	5.21	9.82	9.63	4.23
Vegetable calories	73.22	68.20	79.44	74.36	73.82	73.94	74.55	62.69	88.80	71.92	82.48
2009											
Animal fats	5.80	8.67	5.23	4.08	6.42	4.28	6.35	1.27	1.16	3.11	4.08
Eggs	1.67	1.72	1.27	1.31	1.53	1.48	1.19	0.74	0.98	1.09	1.19
Fish and seafood	0.60	0.33	1.12	1.28	1.78	1.76	1.00	0.36	0.39	1.01	0.95
Meat	9.39	10.51	8.68	10.34	8.20	9.01	10.23	10.04	2.59	12.09	4.66
Milk (excluding butter)	7.29	6.87	8.80	7.17	9.30	7.67	5.45	8.39	5.89	9.39	10.51
Other animals	2.07	1.90	1.45	1.55	2.06	1.71	1.38	1.62	1.06	1.31	1.50
Animal Calories	26.82	29.99	26.55	25.73	29.29	25.92	25.60	22.41	12.07	27.99	22.89
Alcoholic beverages	8.13	6.02	4.81	2.83	7.01	3.02	4.75	5.10	0.65	4.34	3.20
Cereals (excluding beer)	25.24	26.23	27.61	30.34	25.67	29.90	32.06	33.28	42.45	24.72	40.75
Fruits (excluding wine)	2.27	2.92	4.14	5.28	1.64	3.05	1.90	5.55	4.14	3.49	1.53
Potatoes	3.39	2.80	3.42	1.84	5.82	2.46	5.83	2.42	2.53	5.02	3.71
Pulses	0.78	0.69	0.30	1.33	0.00	1.37	0.43	2.86	2.77	0.74	0.37
Sugar and sweeteners	10.40	10.15	13.00	7.49	8.28	15.87	11.45	9.65	8.07	10.29	7.62
Soyabeans	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.00	0.00
Vegetable oils	14.47	13.82	9.89	17.09	11.11	6.49	8.52	8.19	14.53	12.61	11.70
Vegetables	1.49	2.35	2.30	2.88	2.37	3.89	2.28	2.90	3.67	1.83	2.62
Other vegetables	7.01	5.03	7.98	5.20	8.79	8.01	7.19	7.63	9.05	8.98	5.61
Vegetable calories	73.18	70.01	73.45	74.27	70.71	74.08	74.40	77.59	87.93	72.01	77.11